

On deriving most of mathematics from first principles

J.A.Pyne

Abstract

In this volume, we aim to build a strong mathematical foundation that will be used in the

Contents

1 Foundations	1
1.1 Mathematical logic (To add to as needed)	1
1.1.1 Defining a definition	1
1.1.2 What is truth?	1
1.1.2.1 Logical statements and logical connectives	1
1.1.3 Logical propositions	4
1.1.4 Proof	5
1.1.4.1 Direct Proof	5
1.1.4.2 Proof by contradiction	6
1.1.4.3 Proof by contra-position	7
1.2 Sets and mappings	9
1.2.1 Sets	9
1.2.1.1 Introduction and basic definitions	9
1.2.1.2 Subsets and universal quantifiers	10
1.2.1.3 Operations on sets	13
1.2.1.3.1 The union, the intersection and set inclusion	13
1.2.1.3.2 The complement of a set	16
1.2.1.3.3 Cartesian Product	20
1.2.1.3.4 Power Set	25
1.2.1.4 Set Partitions	25
1.2.1.5 A brief look at Zermelo–Fraenkel set theory	27
1.2.2 Mappings	28
1.2.2.1 Introduction and basic definitions	28
1.2.2.2 The image and pre-image	30
1.2.2.3 Injective, surjective and bijective mappings	37
1.2.2.4 Compositions of maps	42
1.2.2.5 Inverse mappings	49
1.3 The Natural numbers	59
1.3.1 Constructing the Natural numbers	59
1.3.2 Properties of the natural numbers	65
1.3.2.1 Equality of natural numbers	65
1.3.2.2 Inequality of natural numbers	66
1.3.2.3 Defining addition and multiplication on the Natural numbers	67
1.3.2.4 Closure properties of addition and multiplication	70
1.3.2.5 Commutativity of addition and multiplication	72

1.3.2.6	Associativity of addition	74
1.3.2.7	Multiplication distributes over addition	75
1.3.2.8	Associativity of multiplication	76
1.3.2.9	The Zero and Identity laws	76
1.3.2.10	The cancellation laws	77
1.3.2.11	Summation and product notation	79
1.3.2.12	Exponentiation	87
1.3.2.13	Subtraction	90
1.3.2.14	The principle of strong induction	92
1.3.2.15	The well-ordering principle	93
1.3.2.16	Rules for the inequality operators	94
1.4	Cardinality, countability, relations	98
1.4.1	Cardinality	98
1.4.2	Countability	104
1.4.3	Relations	106
1.4.3.1	Definition of a relation	106
1.4.3.2	Reflexive Relation	107
1.4.3.3	Equivalence Relations	111
1.5	Construction of the Integers	114
1.5.1	Defining the Integers	114
1.5.2	Extending equality to the integers	117
1.5.3	Extending inequality operators to the integers	118
1.5.4	Extending addition to the integers	118
1.5.5	Extending multiplication to the integers	119
1.5.6	Closure properties of addition and multiplication	119
1.5.7	Associativity of integer addition and multiplication	121
1.5.8	Commutativity of integer addition and multiplication	122
1.5.9	Multiplication distributes over addition	123
1.5.10	The Zero and Identity laws	124
1.5.11	Extending subtraction to the integers	124
1.5.12	The cancellation laws	126
1.5.13	Extending the summation and product notations to integers	128
1.5.14	Extending the rules for inequalities to the integers	134
1.5.15	The absolute value function	141
1.5.16	Extending exponentiation to the integers	147
1.6	Construction of the Rationals	151
1.6.1	Defining the Rationals	151
1.6.2	Extending equality to the rationals	152
1.6.3	Extending inequality operators to the rationals	153
1.6.4	Extending addition to the rationals	153
1.6.5	Extending multiplication to the rationals	154
1.6.6	Closure properties of addition and multiplication	155
1.6.7	Associativity of rational addition and multiplication	156
1.6.8	Commutativity of rational addition and multiplication	157
1.6.9	The Zero and Identity laws	158
1.6.10	Multiplication distributes over addition	159
1.6.11	Extending subtraction to the rationals	160
1.6.12	The cancellation laws	162
1.6.13	Defining multiplicative inverses and division	163
1.6.14	Extending the summation and product notations to the rationals	167
1.6.15	Extending the rules for inequalities to the integers	171

1.6.1	Extending exponentiation to the rational numbers	181
1.6.1	Extending the absolute value function	186
2	Elementary Number Theory	192
2.1	Introduction	192
2.2	Divisibility	192
2.2.1	Definition of divisibility of integers	192
2.2.2	The greatest common divisor and the least common multiple	196
2.3	Prime and co-prime numbers	211
2.3.1	The divisor function	211
2.3.2	Prime numbers	211
2.4	The integers modulo n	229
2.4.1	Remainders after division	229
2.4.2	Congruences and residues (Modular arithmetic)	230
2.5	Diophantine equations and Polynomials	240
2.5.1	Linear Diophantine equations	241
2.5.1.1	Linear equations with two variables	241
2.5.1.2	Linear equations with more than two variables	244
2.5.2	Polynomials	266
2.5.2.1	Defining addition between two polynomials	268
2.5.2.2	Defining multiplication between two polynomials	273

Part 1

Foundations

1.1 Mathematical logic (To add to as needed)

There are no facts, only interpretations.

Friedrich Nietzsche

In this section, we will introduce mathematical logic. This will give us the tools and basic building blocks to be able to talk about mathematics formally. What do we mean by ‘in a formal way’? Modern mathematics is built on a bedrock of logic, that is to say, given some statements which we will take to be true or have already been proven true, what can we logically deduce must also be true, and what is also false. As an example, we are familiar with the idea of positive whole numbers, also called positive integers; we are also familiar with the idea of a positive whole number being prime when the only other positive whole numbers that divide it are 1 and itself, for example, 2 is prime. From the facts that the positive whole numbers exist and there is at least one prime, we can logically deduce there must be infinitely many primes. We will see the proof of this later.

In this document we won’t be needing the full tools of mathematical logic, doing so will take us too far afield, instead, we will only cover the key fundamentals we will need as well as define some terms which will be used throughout.

1.1.1 Defining a definition

What is a definition? What does it mean to define something? Definitions are at the heart of mathematics, without them we wouldn’t be able to do anything at all. A definition is a declaration that gives a formal name to an object, class of objects, ideas, etc. For example, we can define prime numbers, such a definition might look something like this.

Definition. *Definition of a prime number*

Consider a positive whole number, we say that this positive whole number is a prime number if the only other positive whole numbers that divide it are itself and the number 1.

With this definition whenever we refer to the idea of a prime number, we know that this prime number must satisfy that it only has two distinct numbers that divide it, itself and 1. As we will say throughout this document, we can use a definition when making logical arguments. Definitions are the backbone of defining the setup to logical arguments, if we don’t know about the objects we are arguing about then we can’t make any logical deductions, or deduce the truth of mathematical statements. Now that we know what a definition is, we can start using it to lay the foundation for the rest of the document. For formality, we will make, somewhat paradoxically, a formal definition of a definition

Definition 1.1.1. *Definition*

A definition is a statement which gives a formal name to a concept.

1.1.2 What is truth?

What is truth? In particular, what is mathematical truth? Loosely speaking truth and mathematical truth is based on the idea of does the premise entail this conclusion. That is to say, if we assume that a few statements are true, then the conclusion we are trying to reach is also true. This is rather vague at the moment because we haven’t defined what we mean by true.

1.1.2.1 Logical statements and logical connectives

We will need a few definitions.

Definition 1.1.2. *Declarative logical statement*

We define a Declarative logical statement to be either true or false. Here we are using the intuitive definition of true and false.

We need to make the definition of declarative logical statements to define what we mean by true and false, again somewhat paradoxically we need a definition of true and false to define what we mean when a declarative logical statement is true. We shall ignore the paradoxical nature of these definitions.

Definition 1.1.3. *Assignment of truth*

Let P be a declarative logical statement, an assignment of truth is an interpretation of the statement P that sees P as either true or false. We write this as $\delta(P)$.

If this assignment of truth δ sees P as true we write $\delta(P) = 1$ and we say that δ interprets P as true. If this assignment of truth sees P as false we write $\delta(P) = 0$ and we say that δ interprets P as false.

These two definitions will allow us to build the foundations that we will need. It is first important to note that an assignment of truth is not an absolute assignment of the truth of a declarative logical statement. Different assignments of truth, and thus different interpretations, can give rise to different values of P being true or false. Now, we have a building blocks to build more complex logical statements.

A first natural question is when does one the truth of one logical statement imply the truth or falseness of another? Thinking about how this should work gives us a sense that something true should never imply that something false is true, whereas something false can imply anything at all. Using this we define the logical implication operator.

Definition 1.1.4. *Logical implication*

Let P and Q be logical statements. We define the logical implication of the statements P and Q , written as $P \Rightarrow Q$, to have the following logical values

P	Q	$P \Rightarrow Q$
1	1	1
1	0	0
0	1	1
0	0	1

Table 1: The truth table for the logical implication operator.

We read this as P implies Q , or if P then Q .

Example 1.1.1. Let P = “The sky is overcast” and let Q = “The sun is not visible”. We have by the truth table of logical implication that $P \Rightarrow Q$ is true when

1. P is true and Q is true
2. P is false and Q is true
3. Both P and Q are false.

In words we have $P \Rightarrow Q$ is true in these circumstances

1. If it is true the sky is overcast then the sun is not visible.
That is, if the sky is overcast then the sun is not visible
2. If it is false that the sky is overcast then the sun is not visible.
That is, if the sky is not overcast then the sun is not visible.
3. If it is false that the sky is overcast then the sun is visible.
That is, if the sky is not overcast then the sun is visible.

In particular case two could be true say when it is nightttime, if it is nightttime the sun is clearly not visible¹.

Lets look at these statements the other way, $Q \Rightarrow P$. We have that is is true when

1. Q is true and P is true
2. Q is false and P is true
3. Both Q and P are false.

In words that is we have $Q \Rightarrow P$ is true in these circumstances

1. If the sun is not visible then the sky is overcast
2. If the sun is visible then the sky is overcast
3. If the sun is visible then the sky not is overcast

There is one definition that arises from logical implication that is occasionally useful in proving other statements.

Definition 1.1.5. *Vacuous truth*

Let P and Q be statements such that we have $P \Rightarrow Q$. Suppose that P is false, then by the definition of logical implication we have that $P \Rightarrow Q$ is true. We say that $P \Rightarrow Q$ is vacuously true.

Example 1.1.2. The statement “All my children are goats” is vacuously true for someone who doesn’t have any children.

It is often the case we have theorems in mathematics which are of the form P implies Q and Q implies P , that is two separate logical sentences can imply each other. This is the logical bi-conditional.

Definition 1.1.6. *Logical Bi-conditional*

Let P and Q be logical statements. We define the logical Bi-conditional of the statements P and Q , written $P \Leftrightarrow Q$, to have the following logical values

P	Q	$P \Leftrightarrow Q$
1	1	1
1	0	0
0	1	0
0	0	1

Table 2: The truth table for the logical Bi-conditional operator.

We read this as P if and only Q , meaning P implies Q and Q implies P .

Example 1.1.3. Let P = “A number is even” and let Q = “It is divisible by 2”. By the truth table of the logical bi-conditional that $P \Leftrightarrow Q$ is true when

1. Both P and Q are true.
2. Both P and Q are false.

That is in words we have $P \Leftrightarrow Q$ when

1. A number is even if and only if it is divisible by 2
2. A number is not even if and only if it is not divisible by 2

Now that we have the logical implication and logical bi-conditional, we can start defining more complex logical connectives. These are the logical conjunction, logical disjunction and logical negation

¹We are clearly not talking about sunsets

Definition 1.1.7. *Logical conjunction*

Suppose we have two logical statements P and Q . We define logical conjunction, written as $P \wedge Q$, to be true if and only if P and Q are both true, that is to say the logical conjunction connective has the following truth table

P	Q	$P \wedge Q$
1	1	1
1	0	0
0	1	0
0	0	0

Table 3: The truth table for the logical conjunction operator.

Informally, we call this logical AND rather than logical conjunction.

Example 1.1.4. Let $P = "x > 2"$ and $Q = "x < 10"$ and suppose that P and Q are true, then $P \wedge Q$ is true and represents the expression $2 < x < 10$.

Example 1.1.5. Let $P \wedge Q$ be the expression "Adam likes apples and oranges". We can break down $p \wedge Q$ into the two separate logical sentences, $P = "Adam likes apples"$ and $Q = "Adam likes oranges"$.

Definition 1.1.8. *Logical disjunction*

Suppose we have two logical statements P and Q . We define logical disjunction, written as $P \vee Q$, to be true when either one of P and Q are true or both P and Q are true. This is to say the logical disjunction connective has the following truth table

P	Q	$P \vee Q$
1	1	1
1	0	1
0	1	1
0	0	0

Table 4: The truth table for the logical disjunction operator.

Informally, we call this logical OR rather than logical disjunction.

1.1.3 Logical propositions

Now that we have an idea about logical connectives we can consider more complex logical statements, in particular we now start to consider statements whose truth values can depend on a variable.

Definition 1.1.9. *Variable*

A variable is something that has a value that can change.

How can a statement whose truth value change depending on a variable. To answer this we need to introduce the idea of all the possible values that this variable can take.

Definition 1.1.10. *Domain of Discourse*

We define the Domain of Discourse to be the collection of all values that a variable can take. We will denote the domain of Discourse by \mathbb{D} .

Definition 1.1.11. *Logical proposition*

Let \mathbb{D} be a Domain of Discourse. We define a logical proposition, denoted by $P(n_1, n_2, \dots, n_k)$ be a proposition that is based on the variables n_1, n_2, \dots, n_k are variables from the domain of discourse.

Example 1.1.6. Let $P(n)$ be the proposition denoted by

$$P(n) = n \text{ is a even number}$$

where the domain of discourse $D = \mathbb{N}$ where \mathbb{N} denotes the positive whole numbers, i.e $1, 2, 3, \dots$

We have for even n , say $2, 4, 6, 8, \dots$ that $P(n)$ is true. and for odd n , say $1, 3, 5, 7, \dots$ that $P(n)$ is false.

Example 1.1.7. Let $P(n, m)$ be the proposition denoted

$$P(n, m) = n > m$$

that is n is greater than m , where the domain of discourse $D = \mathbb{N}$ is again the positive whole numbers $1, 2, 3, \dots$

Suppose that $n = 2$ and $m = 3$, then $P(n, m)$ is false, if $n = 45$ and $m = 7$ then $P(n, m)$ is true.

We see that logical propositions allow us to construct more complex logical statements and are the building blocks for the more complex Mathematical statements that we will be using.

1.1.4 Proof

Logic and truth are two of the corner stones of Mathematics, the third is proof. Without proof we are unable to verify the truth of any mathematical statements. So what exactly is a proof?

Definition 1.1.12. *Mathematical proof*

Suppose we have some logical statements which are known or assumed to be true, and suppose we wish to see if some conclusion is true given this assumption. We define a Mathematical proof is where we start from these assumptions and at each step logically deduce additionally true statements until we have proven the conclusion. In other words a Mathematical proof can be broken down into a simple question. Do the assumptions entail this conclusion?

This isn't a truly rigorous definition of a mathematical proof, and one can define this rigorously in a course on mathematical logic. To do so here would be too much of a diversion, instead we will just keep in our minds that a proof is a series of logical deductions from assumptions to a conclusion. When we have reached the conclusion we use a special symbol. We use the symbol \square at the end of a proof to show that we are done.

There are many different types of proof that we will invoke throughout the rest of this document.

1.1.4.1 Direct Proof

The first type of proof we define is direct proof. We define a direct proof as follows.

Definition 1.1.13. *Direct Proof*

In a direct proof, the conclusion is logically established by using axioms, definitions and previously proven theorems.

We will give an example of direct proof.

Example 1.1.8. In this example we will breakdown each step of a direct proof.

Here we will give the definitions we will be using and any assumptions which we will be using in the proof (i.e previously proven theorem):

1. We say a number is an integer if it is a whole number, such as $-4, -3, 54, 8, 0, 2, 7$ and so on.
2. We will also assume that adding and multiplying integers works as we would have taught in school, for example $5 + 7 = 12$, $2 * 14 = 28$ etc.
3. We say that an integer is an even integer if it can be written as $x = 2 * m$ where m is any integer.

We now move to the proof.

Suppose we have two such even integers, say x and y . We will use direct proof to show that $x + y$ must also be even.

Proof:

Suppose we have two even integers x and y . By the definition of an even integer we have that $x = 2 * n$ and $y = 2 * m$ for some integers n and m . Now consider $x + y$, we have

$$x + y = 2 * n + 2 * m = 2 * (n + m)$$

Now, $n + m$ is adding two integers together and is an integer. say $k = n + m$, hence we have that $x + y = 2 * k$, but by definition of an even we have that $x + y$ is even. This concludes the proof. \square

1.1.4.2 Proof by contradiction

The second type of proof we define is proof by contradiction. This is a very powerful tool.

Definition 1.1.14. *Proof by contradiction*

Suppose we have a logical statement P that we wish to find the truth of. If we suppose that $\neg P$ is true and then assuming $\neg P$ we can derive another logical statement Q which is known to be false, or we can derive both Q and $\neg Q$. Then we must have that $\neg P$ is false and P is true.

In other words, proof by contradiction states that if, when making an assumption, we can derive a false statement, then the assumption itself must have been invalid. We can justify proof by contradiction using the following truth table.

P	$\neg P$	$\neg\neg P$	$\neg\neg P \Rightarrow P$
1	0	1	1
0	1	0	1

Table 5: The truth table for proof by contradiction.

Example 1.1.9. *Like with the example using direct proof. We will break down each step of proof by contradiction.*

Here we will give the definitions we will be using and any assumptions which we will be using in the prove(i.e previously proven theorem):

1. We say a number is a rational number if it is the ration of two integers a and b where $b \neq 0$. Examples of rational numbers are $\frac{1}{2}$, $\frac{2}{3}$, $-\frac{15}{8}$ and so on. We say a number is irrational if it is not rational.
2. We say that a rational number $\frac{a}{b}$ is in simplest form if the only number that divides both a and b is 1.
3. Any rational number has a simplest form.
4. We will also assume that adding and multiplying rational numbers works as we would have taught in school, that is we have for two rational numbers $\frac{a}{b}$ and $\frac{c}{d}$ that

$$\frac{a}{b} + \frac{c}{d} = \frac{a * d + b * c}{b * d}, \quad \frac{a}{b} * \frac{c}{d} = \frac{a * c}{b * d}$$

5. We say $\sqrt{2}$ is the number which satisfies $\sqrt{2} * \sqrt{2} = 2$
6. We assume the definition of an even integer from the previous example
7. If $a * a = a^2$ is an even integer, then so is a

We now move to the proof.

We have that $\sqrt{2}$ is an irrational number. This is to say that $\sqrt{2}$ is not the ratio of two whole numbers a and b where $\frac{a}{b}$ is in simplest form.

Proof:

Aiming for a proof by contradiction, suppose that $\sqrt{2}$ is a rational number that is in simplest form. This is to say we have that $\sqrt{2} = \frac{a}{b}$ for some integers a, b . We have by assumption that $\sqrt{2}$ is the number such that $\sqrt{2} * \sqrt{2} = 2$. Hence we have that

$$\sqrt{2} * \sqrt{2} = \frac{a}{b} * \frac{a}{b} = \frac{a^2}{b^2} = 2$$

Where $a^2 = a * a$ and $b^2 = b * b$. We can multiply the above expression by b^2 on both sides to get

$$a^2 = 2 * b^2$$

By definition of an even integer we have that a^2 is even and so a must be even, that is $a = 2 * k$ for some integer k . Hence we have that

$$a^2 = (2 * k)^2 = 4 * k^2 = 2 * b^2$$

That is $4 * k^2 = 2 * b^2$ which implies that $b^2 = 2 * k^2$, that is b^2 is even and so b must be even. This is a contradiction, as we have that a is even and b is even and so there share a divisor of 2, contradicting the fact we assumed that $\sqrt{2} = \frac{a}{b}$ was in simplest form.

Therefore, $\sqrt{2}$ must be irrational. \square

1.1.4.3 Proof by contra-position

Another type of proof that we define is proof by contra-position, sometimes called proof by contra-positive.

Definition 1.1.15. *Proof by contra-position*

Suppose we have a logical statement P and we wish to show that P implies some other statement Q . We are able to show that $P \Rightarrow Q$ if we can show that $\neg Q \Rightarrow \neg P$.

Proof by contra-position states that proving a statement of the form $P \Rightarrow Q$ is the same as showing that $\neg Q \Rightarrow \neg P$. It is easier to see this from the truth table.

P	Q	$\neg P$	$\neg Q$	$P \Rightarrow Q$	$\neg Q \Rightarrow \neg P$
1	0	0	1	0	0
1	1	0	0	1	1
0	0	1	1	0	1
0	1	1	0	1	1

Table 6: The truth table for proof by contra-positive.

Maybe, to make it even clearer, we can use a worded example. Let P denote the statement “It is raining” and Q denote the statement “I wear my coat”. We have that $P \Rightarrow Q$ ². The contra-positive would be $\neg Q \Rightarrow \neg P$. In words this would be “If I don’t wear my coat” then “It is not raining”.

Example 1.1.10. *A more mathematical example can be seen now. We will let x be an integer and we will show that if x^2 is even then x is even. We will use proof by contra-position. So We will show that if x is not even then x^2 is not even.*

1. So, x not being even means x is odd. This means that $x = 2n + 1$ for some integer n .

2. Now, we have $x^2 = (2n + 1)^2 = 4n^2 + 4n + 1 = 2(2n^2 + 2n) + 1$.

²If we are being logical and don’t want to get soaked before we get to our destination.

3. Hence, we have shown that x^2 is of the form $2k + 1$ for some integer k .
4. Therefore x^2 is odd.

Concluding the proof by contra-positive.

1.2 Sets and mappings

No one shall expel us from the paradise that Cantor has created for us.

David Hilbert

1.2.1 Sets

1.2.1.1 Introduction and basic definitions

We start with the most elementary definition, a Set or less formally, a collection of ‘objects’. This notion of an object is not very rigorous, what do we mean by an object? Do these objects really exist?³ In what way can one collection of objects differ from another?

These questions are at the foundation of Mathematics and to justify the notions and hence tools we need would require a significant detour into the realm of Mathematical logic. The interested reader would find so-called Zermelo–Fraenkel set theory to be of interest in formalising the notion of a set, we will give a brief overview at the end of the section. To avoid the trip into Mathematical logic, we will instead define sets with a more ‘hands on’ approach

Definition 1.2.1. *Naive definition of a Set*

A set is a collection of objects. We list the elements surrounded by curly brackets $\{ \}$.

This definition will make sense after we see some examples

Example 1.2.1. *Let $S = \{1, 2, 3, \text{Dogs}, \text{Cats}, \text{Apples}, \text{Pears}\}$. Then S is a set.*

Example 1.2.2. *Let $S = \{ \text{“Foo”}, \{1, 2, 3, \text{Dogs}, \text{Cats}, \text{Apples}, \text{Pears}\}, \text{Apples}, \text{Pears} \}$. Then S is a set. We note that the set from the previous example is now in this set.*

It would be useful to talk about a particular object in some set S . For example we can say that 1 is in the set from example 2.1. above. We formalise this idea

Definition 1.2.2. *Element of a set*

An object in a set is called an element of the set.

Definition 1.2.3. *Set membership*

Let S be a set and let x be an element of the set S . We say that x is a member of the set S and write $x \in S$. If y is some object which is not in the set S we write that $y \notin S$.

Example 1.2.3. *Let $S = \{1, 2, 3, \text{Dogs}, \text{Cats}, \text{Apples}, \text{Pears}\}$. We have that $1 \in S$ and $\text{Dogs} \in S$ but we have that $\text{Blue} \notin S$.*

The above example shows a few interesting points. Dogs in English is used when we wish to talk about multiple dogs at once, so it would be absurd to deny that *Dogs* could itself be a set, for example $\text{Dogs} = \{\text{Lassie}, \text{Scooby} - \text{Doo}, \text{Snoopy}, \text{Blue}\}$. So we have that

$$S = \{1, 2, 3, \{\text{Lassie}, \text{Scooby} - \text{Doo}, \text{Snoopy}, \text{Blue}\}, \text{Cats}, \text{Apples}, \text{Pears}\}$$

Does this now mean that $\text{Blue} \in S$? The answer is no, *Blue* is not any one of the objects in S , however there is an object in S that does contain *Blue*, namely *Dogs*. This shows that \in only looks at most one layer deep of $\{ \dots \}$.

One might wonder if it can ever be the case that a set contains itself, that is a set like $S = \{S\}$? Again the answer is no, to see why we need to define a new way of making sets, where the elements of the set are conditioned on some statement being true.

³By exist we mean in the abstract sense.

Example 1.2.4. Suppose we want the set of all even integers then we have

$$S = \{x : x \text{ is an even integer}\}$$

The $:$ symbol stands for such that, so S reads the elements x such that x is an even integer.

Returning to the question of can a set contain itself. Consider the set

$$S = \{R : R \text{ is a set and } R \notin R\}$$

That is S is the set of all sets R such that R is a set and R does not contain itself. Now suppose that $S \in S$. By definition of S we must conclude that $S \notin S$. Conversely if $S \notin S$ then by definition of S we have that $S \in S$. This is an issue, and shows the flavour of the issues of allowing a set to contain itself, so we shall revise our definition to not allow for a set to contain itself.

Definition 1.2.4. Set

A set is a collection of objects such that none of the objects in the collection is the set itself.

1.2.1.2 Subsets and universal quantifiers

Given a set, we can talk about a smaller collection of the elements of the set, which we call a subset.

Definition 1.2.5. Subset

Let S be a set. If K is also a set such that for every $x \in K$ we also have that $x \in S$ then we say that K is a subset of S , and write $K \subseteq S$. We say that K is a proper subset of S if we have that $S \subseteq T$ and $S \neq T$, we denote a proper subset by \subset , hence \subseteq allows for the possibility that $K = S$. We call \subseteq and \subset the set inclusion operators.

Conversely can also define the notion of a super-set, this isn't too useful for what we are doing but it does sometimes appear in other text so it worth mentioning it now.

Definition 1.2.6. Super-set

Let $S \subseteq T$. We say that T is a super-set of the set S and we write this as $T \supseteq S$.

Example 1.2.5. Let $S = \{1, 2, 3, 4, 5, 6\}$ then some subsets of S are $\{1, 2\}$, $\{4\}$ and $\{1, 2, 6\}$

With the idea of a subset we have our first proposition

Proposition 1.2.1. Two sets are equal if and only if they are subsets of each other

Let X and Y be sets. We have that $X = Y$ if and only if $X \subseteq Y$ and $Y \subseteq X$.

Proof:

This is an if and only if proposition so we have to prove that given $X = Y$ then $X \subseteq Y$ and $Y \subseteq X$ and then we need to show that given $X \subseteq Y$ and $Y \subseteq X$, that $X = Y$.

(\Rightarrow) : Suppose that $X = Y$ then we have that X and Y have the same elements, in particular we have that every $x \in X$ is also in Y so that $X \subseteq Y$. Likewise $Y \subseteq X$.

(\Leftarrow) : Suppose that $X \subseteq Y$ and $Y \subseteq X$. $X \subseteq Y$ means that for every $x \in X$ we have that $x \in Y$. Likewise $Y \subseteq X$ means that for every $x \in Y$ we have that $x \in X$. Hence we must have that the elements of X and Y are the same, that is $X = Y$. \square

There is also another property of subsets that is useful.

Proposition 1.2.2. Set inclusion transitivity property

Let R, S and T be sets such that $R \subseteq S$ and $S \subseteq T$. We have that $R \subseteq T$

Proof:

Let R, S and T be sets such that $R \subseteq S$ and $S \subseteq T$. Suppose that $x \in R$. By assumption we have that $R \subseteq S$ and so $x \in S$. Likewise by assumption we have that $S \subseteq T$ and so $x \in T$. Hence $R \subseteq T$.

The result follows. \square

A similar result holds if we replace subsets with proper subsets.

Proposition 1.2.3. *Proper set inclusion transitivity property*

Let R, S and T be sets such that $R \subset S$ and $S \subset T$. We have that $R \subset T$

Proof:

Let R, S and T be sets such that $R \subset S$ and $S \subset T$. Suppose that $x \in R$. By assumption we have that $R \subset S$ and so $x \in S$. Likewise by assumption we have that $S \subset T$ and so $x \in T$. Hence $R \subset T$.

We must show that it is not possible for $R = T$. As $R \subset S$ then by definition we have that $R \neq S$, likewise as $S \subset T$ then $S \neq T$. As $R \neq S \neq T$ we conclude that $R \neq T$ and so $R \subset T$.

The result follows. \square

We can also make the following observation.

Proposition 1.2.4. *Proper set inclusion and subset inclusion is not transitive*

Let R, S and T be sets such that $R \subseteq S$ and $S \subset T$. We have that $R \subset T$

Proof:

Let R, S and T be sets such that $R \subseteq S$ and $S \subset T$.

If $R \neq S$ then $R \subset S$ and so proposition 1.2.3 applies. So suppose that $R = S$ then $R \subseteq S$ and so $\forall x \in R$ we have that $x \in S$. Now as $S \subset T$ we have that $S \neq T \implies R \neq T$ as $R = S$.

The result follows. \square

We will define what we truly mean by transitivity in the next chapter, right now it is more important to know that sets satisfy this property than why this property is named the way it is. As set inclusion is transitive, so is set equality.

Proposition 1.2.5. *Set equality transitivity property*

Let R, S and T be sets such that $R = S$ and $S = T$. We have that $R = T$.

Proof:

Let R, S and T be sets such that $R = S$ and $S = T$. We have that $R = T$. By equality of sets we have that $R \subseteq S$ and $S \subseteq R$, likewise we also have that $S \subseteq T$ and $T \subseteq S$. Now as $R \subseteq S$ and $S \subseteq T$ then we must have by transitivity of set inclusion that $R \subseteq T$. Moreover as $T \subseteq S$ and $S \subseteq R$ we again have by transitivity that $T \subseteq R$. The result follows by equality of sets. \square

Definition 1.2.7. *The empty-set*

The empty-set is the set that contains no elements. It is denoted by \emptyset .

To make our lives a little easier we will introduce some notation

Definition 1.2.8. *Universal and existential quantifiers*

Let S be any set. The universal quantifier \forall , meaning for all, allows us to talk about every element S . We can condition the universal quantifier with a such that \therefore , in order to pick all the elements that satisfy a given condition.

The existential quantifier \exists tells us of the existence of an element in S . Just saying an element in a set exists is not particularly usual and so we normally combine \exists with a condition.

Some examples will help us here.

Example 1.2.6. Consider the set $\{1, 2, 3, 4, 5, \dots\} = \mathbb{N}$, we call \mathbb{N} the natural numbers. Moreover, consider $S = \{1, 2, 3, 4, 5, 6\}$

1. We have that $\forall x \in S$ that $x \in \mathbb{N}$, that is every element of S is also an element of \mathbb{N} .
2. We can apply the universal quantifier multiple times in a statement, for example

$$\forall a \in \mathbb{N}, \forall b \in \mathbb{N}, \exists c \in \mathbb{N} : a + b = c$$

3. Let $a, b \in \mathbb{N}$ that is let $a \in \mathbb{N}$ and let $b \in \mathbb{N}$. Then we can construct the following set. We say that a is divisible by b if $\exists c \in \mathbb{N}$ such that $a = bc$, we write this as $b \mid a$. The set of all such c can be expressed by

$$C = \{c \in \mathbb{N} : a = bc\}$$

The empty set has the interesting property that it is a subset of any set.

Proposition 1.2.6. *The empty-set is contained in every set*

Let S be any set. Then $\emptyset \subseteq S$

Proof:

We have that $\emptyset \subseteq S$ means that every element of \emptyset is also contained in S . The definition of the empty set means that there are no elements in \emptyset . We can phrase this to the following statement

$$\forall x : x \in \emptyset \Rightarrow x \in S$$

But $x \in \emptyset$ is not true for any x so

$$\forall x : x \in \emptyset \Rightarrow x \in S$$

is vacuously true. It hence follows the empty-set is contained in any set. \square

Proposition 1.2.7. *The empty-set is unique*

The empty-set is unique, that is there is only one distinct set which is the empty-set.

Proof:

Suppose that \emptyset and \emptyset' are two empty sets. By proposition 1.2.6 we have that $\emptyset \subseteq \emptyset'$, likewise $\emptyset' \subseteq \emptyset$. So by proposition 1.2.1 we have that $\emptyset = \emptyset'$. Hence the empty-set is unique. \square

It would be nice to have more ways to construct sets. Two key ways to do this are with the union operation and intersection operation.

Definition 1.2.9. *Union and intersection of sets*

Let S and T be any two sets. We define the union of S and T , denoted by $S \cup T$, is the set

$$S \cup T = \{x : x \in S \text{ or } x \in T\}$$

The intersection of S and T , denoted by $S \cap T$, is the set

$$S \cap T = \{x : x \in S \text{ and } x \in T\}$$

If we have a finite number of sets, given by A_1, A_2, \dots, A_n then the union of all of these sets is denoted by

$$\bigcup_{i=1}^n A_i$$

and the intersection is denoted by

$$\bigcap_{i=1}^n A_i$$

Sometimes it is useful to define a union or intersection of multiple sets given some condition or multiple conditions, usually when the conditions involve other previously defined sets, this is denoted as

$$\bigcup_{\substack{\text{Condition 1 for } A \\ \text{Condition 2 for } A \\ \dots}} A$$

for the union and for the intersection

$$\bigcap_{\substack{\text{Condition 1 for } A \\ \text{Condition 2 for } A \\ \dots}} A$$

Example 1.2.7. Let $S = \{1, 2, 3, 4, 5, 6\}$ and let $T = \{2, 4, 5, 6, 7, 8\}$, we have that

$$\begin{aligned} S \cup T &= \{1, 2, 3, 4, 5, 6\} \cup \{2, 4, 5, 6, 7, 8\} = \{1, 2, 3, 4, 5, 6, 2, 4, 5, 6, 7, 8\} = \{1, 2, 3, 4, 5, 6, 7, 8\} \\ S \cap T &= \{1, 2, 3, 4, 5, 6\} \cap \{2, 4, 5, 6, 7, 8\} = \{1, 2, 3, 4, 5, 6, 2, 4, 5, 6, 7, 8\} = \{2, 4, 5, 6\} \end{aligned}$$

We note that in the union we have multiple elements, for example we have two 2's. Repeated elements in a set are considered to be the same element so we don't write them, i.e $\{2, 2\} = \{2\}$

Example 1.2.8. Let $A_1 = \{1, 2, 3\}$, $A_2 = \{1, 2, 7, 9\}$ and $A_3 = \{1, 4, 8, 12\}$. We have that the union of these sets is given by

$$\begin{aligned} \bigcup_{i=1}^n A_i &= A_1 \cup A_2 \cup A_3 \\ &= \{1, 2, 3\} \cup \{1, 2, 7, 9\} \cup \{1, 4, 8, 12\} \\ &= \{1, 2, 3, 4, 7, 8, 9, 12\} \end{aligned}$$

The intersection of these sets is given by

$$\begin{aligned} \bigcap_{i=1}^n A_i &= A_1 \cap A_2 \cap A_3 \\ &= \{1, 2, 3\} \cap \{1, 2, 7, 9\} \cap \{1, 4, 8, 12\} \\ &= \{1\} \end{aligned}$$

We make one useful definition about intersections

Definition 1.2.10. *Disjoint sets*

Let X and Y be sets. If we have that $X \cap Y = \emptyset$ then we say that X and Y are disjoint sets.

1.2.1.3 Operations on sets

1.2.1.3.1 The union, the intersection and set inclusion Before we continue we introduce three new ideas that will play a role throughout the rest of this paper.

Definition 1.2.11. *Operation*

An operation \circ acts on some inputs to produce an output or some outputs.

Example 1.2.9. The union \cup and intersection \cap are examples of operations. These operators operate on two sets to produce a third.

Definition 1.2.12. *Commutative operation*

Let \circ be an operation that accepts two inputs, i.e we have $A \circ B$ for valid inputs A and B . We say that \circ is commutative if and only if $A \circ B = B \circ A$

Example 1.2.10. Consider $\mathbb{N} = \{1, 2, 3, 4, 5, \dots\}$. We are familiar with the idea of addition of positive numbers, say $1 + 2 = 3$. It is clear that the addition operation is commutative for \mathbb{N} , e.g. $1 + 2 = 3 = 2 + 1$

Definition 1.2.13. *Associative operation*

Let \circ be an operation that accepts two inputs, i.e we have $A \circ B$ for valid inputs A and B . We say that \circ is associative if and only if $(A \circ B) \circ C = A \circ (B \circ C)$ where the operation in the brackets should be computed first.

Example 1.2.11. Again consider $\mathbb{N} = \{1, 2, 3, 4, 5, \dots\}$. The addition operator for \mathbb{N} is associative, e.g. $(1 + 2) + 3 = 3 + 3 = 6 = 1 + 5 = 1 + (2 + 3)$

We note that we have not defined a rigorous notion of addition, to do so will require us to consider mappings which we do later.

We have the following proposition about the properties of intersections, unions and set inclusions.

Proposition 1.2.8. *Properties of intersection, union and set inclusion*

Let A, B, C be sets. Then we have that the following properties are true

1. $A \cap B = B \cap A$
2. $A \cup B = B \cup A$
3. $A \cap B \subseteq A$
4. $A \subseteq A \cup B$
5. $A \subseteq B \Rightarrow A \cap B = A$
6. $A \subseteq B \Rightarrow A \cup B = B$
7. $A \subseteq B \Rightarrow A \cap C \subseteq B \cap C$
8. $A \subseteq B \Rightarrow A \cup C \subseteq B \cup C$
9. $A \cap A = A$
10. $A \cup A = A$
11. $A \cap (B \cap C) = (A \cap B) \cap C$
12. $A \cup (B \cup C) = (A \cup B) \cup C$
13. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
14. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

Proof:

1. $A \cap B = B \cap A$:

Let $x \in A \cap B$ then $x \in A$ and $x \in B$ by the definition of the intersection. It is hence clear that $x \in B \cap A$. So we have $A \cap B \subseteq B \cap A$. Likewise if $x \in B \cap A$ then $x \in B$ and $x \in A$, so that $x \in A \cap B$. So $B \cap A \subseteq A \cap B$. It hence follows by proposition 1.2.1 that $A \cap B = B \cap A$.

2. $A \cup B = B \cup A$:

Let $x \in A \cup B$ then $x \in A$ or $x \in B$ by the definition of the union. We hence have that $x \in B \cup A$. So we have $A \cup B \subseteq B \cup A$. Likewise if $x \in B \cup A$ then $x \in B$ and $x \in A$, so that $x \in A \cup B$. So $B \cup A \subseteq A \cup B$. It hence follows by proposition 1.2.1 that $A \cup B = B \cup A$.

3. $A \cap B \subseteq A$:

Let $x \in A \cap B$, then by the definition of the intersection $x \in A$ and $x \in B$. Hence $x \in A \cap B$ means that $x \in A$ so that $A \cap B \subseteq A$.

4. $A \subseteq A \cup B$:

Let $x \in A$. By the definition of the union of two sets we have that $y \in A \cup B$ if and only if $y \in A$ or $y \in B$. Hence it follows that $x \in A \cup B$.

5. $A \subseteq B \Rightarrow A \cap B = A$:

Let $A \subseteq B$ and suppose that $x \in A$, then we have that $x \in B$ as $A \subseteq B$. Hence $x \in A \cap B$. This holds for any choice of $x \in A$. We conclude that if $A \subseteq B$ then $A \cap B = A$.

6. $A \subseteq B \Rightarrow A \cup B = B$:

Let $A \subseteq B$. Observe that $B \subseteq B$ so that $A \cup B \subseteq B \cup B = B$, that is to say $A \cup B \subseteq B$. Now $B \subseteq A \cup B$. Hence $A \cup B = B$.

7. $A \subseteq B \Rightarrow A \cap C \subseteq B \cap C$:

Suppose that $A \subseteq B$ and let $x \in A \cap C$, then by definition $x \in A$ and $x \in C$. Also we have that as $A \subseteq B$ that $x \in A$ gives $x \in B$. Hence $x \in B \cap C$. It follows that $A \cap C \subseteq B \cap C$.

8. $A \subseteq B \Rightarrow A \cup C \subseteq B \cup C$:

Suppose $A \subseteq B$ and let $x \in A \cup C$. We have that $x \in A$ or $x \in C$. If $x \in A$ then as $A \subseteq B$ we have that $x \in B$ so that $x \in B \cup C$. If $x \in C$ then clearly $x \in B \cup C$. Either way we have that $A \cup C \subseteq B \cup C$.

9. $A \cap A = A$:

Let $x \in A$, then by the definition of the intersection we have that $y \in A \cap A$ if and only if $y \in A$ and $y \in A$, hence $x \in A \cap A$. So that $A \subseteq A \cap A$. Now If $x \in A \cap A$ we have by definition of the intersection of two sets that $x \in A$ and $x \in A$, so the force of deductive logic then drives one to the conclusion that $x \in A$. So $A \cap A \subseteq A$. Hence $A \cap A = A$.

10. $A \cup A = A$:

Let $x \in A$, then by the definition of the union of two sets, we have that $y \in A \cup A$ if and only if $y \in A$ or $y \in A$, hence $x \in A \cup A$ so that $A \subseteq A \cup A$. Now suppose that $x \in A \cup A$, then again by the definition of the union we have that $x \in A$ so that $A \cup A \subseteq A$. Hence $A = A \cup A$.

11. $A \cap (B \cap C) = (A \cap B) \cap C$:

Let A, B and C be sets. Consider $A \cap (B \cap C)$, we have that $x \in A \cap (B \cap C)$ means that $x \in A$ and $x \in B \cap C$, likewise $x \in B \cap C$ means that $x \in B$ and $x \in C$. Now as $x \in A$ and $x \in B$ and $x \in C$ so we have that $x \in A \cap B$ and $x \in C$. Finally we have that $x \in (A \cap B) \cap C$ so that $A \cap (B \cap C) \subseteq (A \cap B) \cap C$.

Now consider $(A \cap B) \cap C$, if $x \in (A \cap B) \cap C$ then $x \in A \cap B$ and $x \in C$, also $x \in A \cap B$ means that $x \in A$ and $x \in B$. As $x \in A$ and $x \in B$ and $x \in C$ so we have that $x \in A$ and $x \in B \cap C$ so that $x \in A \cap (B \cap C)$. Hence $(A \cap B) \cap C \subseteq A \cap (B \cap C)$.

Hence $A \cap (B \cap C) = (A \cap B) \cap C$

12. $A \cup (B \cup C) = (A \cup B) \cup C$:

Let A, B and C be sets. Consider $A \cup (B \cup C)$ and let $x \in A \cup (B \cup C)$, we have that either $x \in A$ or $x \in (B \cup C)$. If $x \in A$ then we have that $x \in A \cup B$ so that $x \in (A \cup B) \cup C$. If $x \in B \cup C$ then either $x \in B$ or $x \in C$. If $x \in B$ then $x \in A \cup C$ so that $x \in (A \cup B) \cup C$. Otherwise $x \in C$ and we have that $x \in (A \cup B) \cup C$. Hence we have that $A \cup (B \cup C) \subseteq (A \cup B) \cup C$

Conversely let $x \in (A \cup B) \cup C$. We have that either $x \in (A \cup B)$ or $x \in C$. If $x \in (A \cup B)$ then either $x \in A$ or $x \in B$, in either case we have that $x \in A \cup (B \cup C)$. If $x \in C$ then $x \in A \cup (B \cup C)$. So that $(A \cup B) \cup C \subseteq A \cup (B \cup C)$.

Hence $A \cup (B \cup C) = (A \cup B) \cup C$

13. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$:

Let $x \in A \cap (B \cup C)$, then we have that $x \in A$ and $x \in B \cup C$. We have $x \in B \cup C$ gives us that $x \in B$ or $x \in C$. If $x \in B$ then $x \in A \cap B$ and so $x \in (A \cap B) \cup (A \cap C)$. Likewise if $x \in C$ then $x \in A \cap C$ so $x \in (A \cap B) \cup (A \cap C)$. Hence $A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$.

For the opposite inclusion, let $x \in (A \cap B) \cup (A \cap C)$ then we have that either $x \in A \cap B$ or $x \in A \cap C$. If $x \in A \cap B$ then $x \in A$ and $x \in B$, so we hence have that $x \in B \cup C$ so that $x \in A \cap (B \cup C)$. Likewise if we have $x \in A \cap C$ then $x \in A$ and $x \in C$, so $x \in B \cup C$ and $x \in A \cap (B \cup C)$. Hence $(A \cap B) \cup (A \cap C) \subseteq A \cap (B \cup C)$

So $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

$$14. A \cup (B \cap C) = (A \cup B) \cap (A \cup C):$$

Let $x \in A \cup (B \cap C)$ then either $x \in A$ or $x \in B \cap C$. If $x \in A$ then $x \in A \cup B$ and $x \in A \cup C$, which is to say $x \in (A \cup B) \cap (A \cup C)$. If $x \in B \cap C$ then $x \in B$ and $x \in C$, so it follows that $x \in A \cup B$ and $x \in A \cup C$ which is to say $x \in (A \cup B) \cap (A \cup C)$. Hence $A \cup (B \cap C) \subseteq (A \cup B) \cap (A \cup C)$.

Now, suppose that $x \in (A \cup B) \cap (A \cup C)$. We then have that $x \in A \cup B$ and $x \in A \cup C$. Now $x \in A \cup B$ gives $x \in A$ or $x \in B$, also $x \in A \cup C$ means that $x \in A$ or $x \in C$. This gives us two possible outcomes. If $x \in A$ then $x \in A \cup (B \cap C)$ so that $(A \cup B) \cap (A \cup C) \subseteq A \cup (B \cap C)$. Suppose that $x \notin A$ then we must have that $x \in B$ and $x \in C$ as $x \in A \cup B$ and $x \in A \cup C$. Hence $x \in B \cap C$ so $x \in A \cup (B \cap C)$. Hence $(A \cup B) \cap (A \cup C) \subseteq A \cup (B \cap C)$.

So we have that $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

The proposition now follows. \square

Theorem 1.2.1. *Equivalence of Subsets with union and intersection*

Let A, B be sets. The following are equivalent

1. $A \subseteq B$
2. $A \cap B = A$
3. $A \cup B = B$

Proof:

Suppose $A \subseteq B$. By proposition 1.2.8 we have that

$$A = A \cap A \subseteq A \cap B \subseteq A$$

Hence $A = A \cap B$.

Now suppose that $A \cap B = A$, then $A \subseteq B$. This shows 1 and 2 are equivalent.

Suppose $A \subseteq B$. Let $x \in A$ then $x \in B$. Then as $x \in B$ we have that $x \in A \cup B$ so that $B \subseteq A \cup B$. Suppose that $x \in A \cup B$, then either $x \in A$ or $x \in B$. If $x \in B$ we are done and we have that $A \cup B \subseteq B$. If $x \in A$ then as $A \subseteq B$ we have that $x \in B$ so that $A \cup B \subseteq B$.

Hence $A \cup B = B$.

Now suppose that $A \cup B = B$. Suppose that $x \in A$ then $x \in A \cup B = B$ so $x \in B$, hence $A \subseteq B$.

This shows the equivalence of 1 and 3.

The equivalence of 2 and 3 now follows. Indeed, suppose that $A \cap B = A$ then by the equivalence of 1 and 2 we know that $A \subseteq B$, also by the equivalence of 1 and 3 we know that $A \cup B = B$. \square

1.2.1.3.2 The complement of a set It sometimes becomes useful to talk about the elements that are not in some set S . This only makes sense if S is contained inside some larger set.

Definition 1.2.14. *Complement of a set*

Let S be a set such that $S \subseteq U$ for some set U . We define the complement of S , denoted by S^C as the following set

$$S^C = \{x \in U : x \notin S\}$$

We can alternatively write $S^C = U \setminus S$, where \setminus is the set difference operation.

Moreover we can also consider the complement of a set A with respect to some other set B , again occurring inside some larger set U which is to say $A \subseteq U$ and $B \subseteq U$. We have that

$$A \setminus B = \{x \in A : x \notin B\}$$

We call

Example 1.2.12. Let $U = \{1, 2, 3, 4, 5, 6\}$, $S = \{1, 2, 3, 4, 6\}$ and $T = \{2, 4, 6\}$. We have that $S \subseteq U$ so that

$$\begin{aligned} S^C &= \{x \in U : x \notin S\} = \{5\} \\ T^C &= \{x \in U : x \notin T\} = \{1, 3, 5\} \end{aligned}$$

Also

$$\begin{aligned} S \setminus T &= \{x \in S : x \notin T\} = \{1, 3\} \\ T \setminus S &= \{x \in T : x \notin S\} = \emptyset \end{aligned}$$

An immediate result follows from the previous definitions of the complement of a set and set difference.

Theorem 1.2.2. *De-Morgan's laws*

Let A and B be subsets of some universal set U . We have the complement laws

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

We also have the set difference laws

1. $U \setminus (A \cap B) = (U \setminus A) \cup (U \setminus B)$
2. $U \setminus (A \cup B) = (U \setminus A) \cap (U \setminus B)$

Proof:

We first prove the complement laws.

1. $(A \cap B)^C = A^C \cup B^C$:

Let $x \in (A \cap B)^C$, by the definition of the set complement we have that $x \notin (A \cap B)$. So by the definition of the intersection and x not being an element of $A \cap B$ we have that $x \notin A$ or $x \notin B$. Suppose that $x \notin A$, then by the definition of set complement we have that $x \in A^C$ so that $x \in A^C \cup B^C$. Likewise if $x \notin B$ then $x \in B^C$ so that $x \in A^C \cup B^C$. Hence we have that $(A \cap B)^C \subseteq A^C \cup B^C$.

Now suppose $x \in A^C \cup B^C$, then $x \in A^C$ or $x \in B^C$. Suppose $x \in A^C$ then $x \notin A$ so that $x \notin A \cap B$ hence $x \in (A \cap B)^C$. Likewise if $x \in B^C$ then $x \notin B$ so $x \notin A \cap B$ so that $x \in (A \cap B)^C$. Thus $A^C \cup B^C \subseteq (A \cap B)^C$

Hence $(A \cap B)^C = A^C \cup B^C$.

2. $(A \cup B)^C = A^C \cap B^C$:

Let $x \in (A \cup B)^C$, then we have that $x \notin A \cup B$ so $x \notin A$ and $x \notin B$. This means that $x \in A^C$ and $x \in B^C$ which is to say $x \in A^C \cap B^C$. So $(A \cup B)^C \subseteq A^C \cap B^C$.

Suppose $x \in A^C \cap B^C$ then $x \in A^C$ and $x \in B^C$. $x \in A^C$ means that $x \notin A$ and $x \in B^C$ means that $x \notin B$, so $x \notin A$ and $x \notin B$ hence $x \notin A \cup B$. Thus $x \in (A \cup B)^C$. Hence $A^C \cap B^C \subseteq (A \cup B)^C$

Thus $(A \cup B)^C = A^C \cap B^C$

It is left to prove the set difference laws.

1. $U \setminus (A \cap B) = (U \setminus A) \cup (U \setminus B)$:

Let $X \in U \setminus (A \cap B)$ then by definition we have that $x \in U$ and $x \notin A \cap B$, which is to say that $x \notin A$ or $x \notin B$ with the possibility of being in neither. If $x \notin A$ then $x \in (U \setminus A)$ and we clearly have

$x \in (U \setminus A) \cup (U \setminus B)$. Likewise if $x \notin B$ and both cases clearly hold in the case where $x \notin A$ and $X \notin B$. It follows that in every case that $x \in (U \setminus A) \cup (U \setminus B)$. Hence $U \setminus (A \cap B) \subseteq (U \setminus A) \cup (U \setminus B)$

Now suppose that $x \in (U \setminus A) \cup (U \setminus B)$ then by definition we have that $x \in U \setminus A$ or $x \in U \setminus B$ with the possibility of being in both. If $x \in U \setminus A$ then $x \in U$ and $X \notin A$. Hence $x \notin A \cap B$, likewise if $X \in Y \setminus B$ then we again conclude that $X \notin A \cap B$. However as $x \in U$ then we have by definition that $x \in U \setminus (A \cap B)$. We conclude that $(U \setminus A) \cup (U \setminus B) \subseteq U \setminus (A \cap B)$

It follows that $U \setminus (A \cap B) = (U \setminus A) \cup (U \setminus B)$

2. $U \setminus (A \cup B) = (U \setminus A) \cap (U \setminus B)$:

Suppose that $U \setminus (A \cup B)$ then $x \in U$ and $x \notin A \cup B$ so $x \notin A$ and $x \notin B$. Clearly then $x \in U \setminus A$ and $x \in U \setminus B$ so that $x \in (U \setminus A) \cap (U \setminus B)$. So we have that $U \setminus (A \cup B) \subseteq (U \setminus A) \cap (U \setminus B)$.

Let $x \in (U \setminus A) \cap (U \setminus B)$ then $x \in U \setminus A$ and $x \in U \setminus B$ which is to say that $x \in U$ and $x \notin A$ and $x \notin B$. Clearly $x \notin A$ and $x \notin B$ gives us that $x \notin A \cup B$ and so $x \in U \setminus (A \cup B)$ by definition. This allows us to conclude that $(U \setminus A) \cap (U \setminus B) \subseteq U \setminus (A \cup B)$

Hence $U \setminus (A \cup B) = (U \setminus A) \cap (U \setminus B)$

This proves the theorem. \square

Proposition 1.2.9. Additional properties of set complements and set differences

Let A, B and C be a sets such that $A \subseteq U$, $B \subseteq U$ and $C \subseteq U$. Moreover suppose U is not contained in any other set. Then we have that

1. $A \cup A^C = U$
2. $A \cap A^C = \emptyset$
3. $\emptyset^C = U$
4. $U^C = \emptyset$
5. If $A \subseteq B$ then $B^C \subseteq A^C$
6. $(A^C)^C = A$
7. $A \setminus B = A \cap B^C$
8. $(A \setminus B)^C = A^C \cup B$
9. $A^C \setminus B^C = B \setminus A$
10. $(A \setminus B) \cap C = (A \cap C) \setminus (B \cap C)$
11. $A \setminus (B \setminus C) = (A \cap B) \setminus (A \cap C)$
12. $(A \setminus B) \cap B = \emptyset$
13. $(A \setminus B) \cap (A \cap B) = \emptyset$

Proof:

1. $A \cup A^C = U$:

Let $x \in A \cup A^C$ then $x \in A$ or $x \in A^C$. If $x \in A$ then as $A \subseteq U$ we have that $x \in U$. If $x \in A^C$ then by the definition of set complements we have that $x \in A^C$ if and only if $x \in U$. Hence $A \cup A^C \subseteq U$.

Conversely suppose that $x \in U$. We know that $A \subseteq U$ so if $x \in A$ we clearly have $x \in A \cup A^C$. So suppose $x \notin A$ then by definition of the set complement we have that $x \in A^C$ so that $x \in A \cup A^C$. Hence $U \subseteq A \cup A^C$.

So $A \cup A^C = U$.

2. $A \cap A^C = \emptyset$:

Let $x \in A \cap A^C$, then $x \in A$ and $x \in A^C$, however $x \in A^C$ means that $x \notin A$. This contradicts the fact that $x \in A$, hence there are no elements $x \in U$ so that $x \in A$ and $x \in A^C$, this is to say $A \cap A^C = \emptyset$.
Hence $A \cap A^C = \emptyset$.

3. $\emptyset^C = U$:

By the definition of the empty set we have that \emptyset has no elements. The complement of the empty-set is

$$\emptyset^C = \{x \in U : x \notin \emptyset\}$$

Hence every $x \in U$ is such that $x \notin \emptyset$. So $\emptyset^C \subseteq U$.

Conversely let $x \in U$, then $x \notin \emptyset$ as \emptyset has no elements. so $x \in \emptyset^C$ hence $U \subseteq \emptyset^C$.

It follows that $\emptyset^C = U$.

4. $U^C = \emptyset$:

Let $x \in U^C$, by the definition of set complement we have that

$$U^C = \{y \in U : y \notin U\}$$

This is clearly empty as no such y can satisfy $y \in U$ and $y \notin U$.

Hence $U^C = \emptyset$.

5. If $A \subseteq B$ then $B^C \subseteq A^C$:

Suppose that $A \subseteq B$. We have by proposition 1.2.8 property 5 we have that $A \cap B = A$. It follows that $(A \cap B)^C = A^C$. Now by De-Morgan's laws we have that $(A \cap B)^C = A^C \cup B^C$. Hence $A^C \cup B^C = A^C$. Finally by theorem 1.2.1 we know that $X \cup Y = Y$ if and only if $X \subseteq Y$ for sets X and Y . Hence $B^C \subseteq A^C$.

6. $(A^C)^C = A$:

Let $x \in (A^C)^C$. By definition we have that

$$(A^C)^C = \{x \in U : x \notin A^C\}$$

Hence $x \in (A^C)^C$ if and only if $x \notin A^C$. However $x \notin A^C$ means that $x \in A$. Hence $(A^C)^C \subseteq A$

Suppose that $x \in A$, then $x \notin A^C$, moreover by definition $x \notin A^C$ if and only if $x \in (A^C)^C$, hence $A \subseteq (A^C)^C$.

Hence $(A^C)^C = A$

7. $A \setminus B = A \cap B^C$:

Let $x \in A \setminus B$, then by definition we have that $A \setminus B$ is the set

$$A \setminus B = \{y \in A : y \notin B\}$$

Hence $x \in A \setminus B$ means that $x \in A$ and $x \notin B$. We have that $x \notin B$ means that $x \in B^C$. So that $x \in A \cap B^C$. It follows that $A \setminus B \subseteq A \cap B^C$.

Let $x \in A \cap B^C$, then $x \in A$ and $x \in B^C$. $x \in B^C$ means that $x \notin B$, so by definition $x \in A$ and $x \notin B$ means that $x \in A \setminus B$. Hence $A \cap B^C \subseteq A \setminus B$.

Hence $A \setminus B = A \cap B^C$.

8. $(A \setminus B)^C = A^C \cup B$:

We know that $A \setminus B = A \cap B^C$ by the previous property. Now by De-Morgan's laws we have that

$$(A \setminus B)^C = (A \cap B^C)^C = A^C \cup (B^C)^C = A^C \cup B$$

9. $A^C \setminus B^C = B \setminus A$:

We know that $A^C \setminus B^C = A^C \cap (B^C)^C$. Now, $(B^C)^C = B$ hence $A^C \cap (B^C)^C = A^C \cap B = B \cap A^C$. Finally we know that $B \cap A^C = B \setminus A$ by property 7.

Hence $A^C \setminus B^C = B \setminus A$.

10. $(A \setminus B) \cap C = (A \cap C) \setminus (B \cap C)$:

11. $A \setminus (B \setminus C) = (A \cap B) \setminus (A \cap C)$

12. $(A \setminus B) \cap B = \emptyset$

13. $(A \setminus B) \cap (A \cap B) = \emptyset$

The proposition now follows. \square

1.2.1.3.3 Cartesian Product We now look to another method of constructing a set. This method differs from the union and intersection as it allows us to construct a set where the elements come in pairs, in particular these pairs are ordered.

Definition 1.2.15. *Ordered pair*

Let S and T be sets. Let $s \in S$ and $t \in T$. We say that the tuple (s, t) is an ordered pair of an element in S and an element in T .

Definition 1.2.16. *Cartesian product of two sets*

Let S and T be sets. We define the Cartesian product of S and T , denoted $S \times T$ to be the set of all ordered pairs of the form (s, t) where $s \in S$ and $t \in T$. This is to say that

$$S \times T = \{(s, t) : s \in S, t \in T\}$$

Example 1.2.13. Let $S = \{1, 2, 3\}$ and $T = \{4, 5, 6\}$. We have that

$$S \times T = \{(1, 4), (1, 5), (1, 6), (2, 4), (2, 5), (2, 6), (3, 4), (3, 5), (3, 6)\}$$

$$T \times S = \{(4, 1), (4, 2), (4, 3), (5, 1), (5, 2), (5, 3), (6, 1), (6, 2), (6, 3)\}$$

This example shows that $S \times T \neq T \times S$ in general.

We can make repeated uses of this idea, we just need to defined an ordered n -tuple.

Definition 1.2.17. *Ordered n -tuple*

Let S_1, S_2, \dots, S_n be sets. Let $s_1 \in S_1, s_2 \in S_2, \dots, s_n \in S_n$. We say that (s_1, s_2, \dots, s_n) is an ordered n -tuple of an elements in S_1, S_2, \dots, S_n .

Definition 1.2.18. *Cartesian product of n sets*

Let S_1, S_2, \dots, S_n be sets. We define the Cartesian product of S_1, S_2, \dots, S_n , denoted $S_1 \times S_2 \times \dots \times S_n$ to be the set of all ordered pairs of the form (s_1, s_2, \dots, s_n) where $s_1 \in S_1, s_2 \in S_2, \dots, s_n \in S_n$. This is to say that

$$S_1 \times S_2 \times \dots \times S_n = \{(s_1, s_2, \dots, s_n) : s_1 \in S_1, s_2 \in S_2, \dots, s_n \in S_n\}$$

If all the sets are the same we denote this by S^n .

We make the following observations

Lemma 1.2.1. *Cartesian product is empty if and only if at least one of the sets in the product is empty*
Let A and B be sets. We have that $A \times B = \emptyset$ if and only if $A = \emptyset$ or $B = \emptyset$.

Proof:

We argue as follows. Suppose that $A \times B \neq \emptyset$ then we have by definition of a non-empty Cartesian product that $A \times B \neq \emptyset$ if and only if $\exists (a, b) \in A \times B$. Now, by the definition of a Cartesian product we have that as $(a, b) \in A \times B$ if and only if $\exists a \in A$ and $\exists b \in B$, which is to say $A \neq \emptyset$ and $B \neq \emptyset$.

This proves the result as assuming $A \times B \neq \emptyset$ gives us $A \neq \emptyset$ and $B \neq \emptyset$. \square

Proposition 1.2.10. *Criterion for commutativity of the Cartesian product*

Let A and B be sets. We have that $A \times B = B \times A$ only if at least one of the following holds.

1. $A = B$
2. $A = \emptyset$ or $B = \emptyset$ or $A = B = \emptyset$

Proof:

Let A and B be sets.

1. $A = B$:

Suppose that $A = B$ then without loss of generality⁴ consider

$$A \times B = A \times A = \{(a, a) : a \in A\}$$

Moreover

$$B \times A = A \times A = \{(a, a) : a \in A\}$$

Hence, varying over every $a \in A$ we have that $A \times B = B \times A$.

2. $A = \emptyset$ or $B = \emptyset$ or $A = B = \emptyset$:

By lemma 1.2.1 we have that if $A = \emptyset$ or $B = \emptyset$ or $A = B = \emptyset$ then $A \times B = \emptyset = B \times A$.

The proposition follows. \square

We have seen that the Cartesian product is not commutative, but what can we say about associativity.

Example 1.2.14. *Let $A = \{1\}$. Consider*

$$\begin{aligned} A \times (A \times A) &= A \times \{(1, 1)\} = \{(1, (1, 1))\} \\ (A \times A) \times A &= \{(1, 1)\} \times A = \{((1, 1), 1)\} \end{aligned}$$

Hence $A \times (A \times A) \neq (A \times A) \times A$. So in general the Cartesian product is not associative.

We have the following criterion for the associativity of the Cartesian product.

Proposition 1.2.11. *Criterion for associativity of the Cartesian product*

Let A, B and C be sets. We have that $A \times (B \times C) = (A \times B) \times C$ if and only if $A = \emptyset$ or $B = \emptyset$ or $C = \emptyset$.

Proof:

Suppose that $A \times (B \times C) = (A \times B) \times C$, we need to show one of A, B or C is empty.

Consider $A \times (B \times C)$, we have that

⁴Without loss of generality means we have made a choice in the proof which allows us to consider a single case as the other cases have the same argument just with the notation changed to reflect the different choice.

$$A \times (B \times C) = A \times \{(b, c) : b \in B, c \in C\} = \{(a, (b, c)) : a \in A, (b, c) \in B \times C\}$$

Now consider $(A \times B) \times C$, we have that

$$(A \times B) \times C = \{(a, b) : a \in A, b \in B\} \times C = \{((a, b), c) : (a, b) \in A \times B, c \in C\}$$

Hence for equality we need that $a = (a, b)$ and $(b, c) = c$. However this is not possible as $(a, b) \notin A$ and $(b, c) \notin C$. Hence one of the products must be empty, which implies that one of A, B or C is empty.

Now suppose that one of A, B or C is empty. Without loss of generality suppose that $A = \emptyset$, then by lemma 1.2.1 we know that one of $A \times B = \emptyset$ and $A \times (B \times C) = \emptyset$. Also $(A \times B) \times C = \emptyset \times C = \emptyset$.

Hence we have that $(A \times B) \times C = \emptyset = A \times (B \times C)$. is associative. \square

It is left to see how the Cartesian product interacts with unions, intersections and complements.

Proposition 1.2.12. *Properties of Cartesian products, unions, intersections and complements*

Let A, B, C and D be sets. We have the following properties

1. $(A \cap B) \times (C \cap D) = (A \times C) \cap (B \times D)$
2. $A \times (B \cap C) = (A \times B) \cap (A \times C)$
3. $(A \times B) \cap (B \times A) = (A \cap B) \times (A \cap B)$
4. $(A \cup B) \times (C \cup D) = (A \times C) \cup (B \times D) \cup (A \times D) \cup (B \times C)$
5. $A \times (B \cup C) = (A \times B) \cup (A \times C)$
6. $(B \cup C) \times A = (B \times A) \cup (C \times A)$
7. If $A \subseteq B$ and $C \subseteq D$ then $A \times C \subseteq B \times D$. Moreover if $A \neq \emptyset$ and $C \neq \emptyset$ then

$$A \times C \subseteq B \times D \iff A \subseteq B \text{ and } C \subseteq D$$

8. If $A \subseteq B$ then $A \times C \subseteq B \times C$
9. If $C \subseteq D$ then $A \times C \subseteq A \times D$
10. $A \times (B \setminus C) = (A \times B) \setminus (A \times C)$
11. $(A \setminus B) \times C = (A \times C) \setminus (B \times C)$
12. $(A \times B) \setminus (C \times D) = (A \times (B \setminus D)) \cup ((A \setminus B) \times C)$
13. Suppose $A \subseteq C$ and $B \subseteq D$ and consider $C \setminus A$ and $D \setminus B$. We have

$$\begin{aligned} (C \setminus A) \times D &= (C \times D) \setminus (A \times D) \\ C \times (D \setminus B) &= (C \times D) \setminus (C \times B) \end{aligned}$$

Proof:

1. $(A \cap B) \times (C \cap D) = (A \times C) \cap (B \times D)$:

Let $(x, y) \in (A \cap B) \times (C \cap D)$, then by definition of the Cartesian product we have that $(x, y) \in (A \cap B) \times (C \cap D)$ if and only if $x \in A$ and $x \in B$ and $y \in C$ and $y \in D$. $x \in A$ and $x \in B$ and $y \in C$ and $y \in D$ means that $(x, y) \in A \times C$ and $(x, y) \in B \times D$, finally this happens if and only if $(x, y) \in (A \times C) \cap (B \times D)$.

2. $A \times (B \cap C) = (A \times B) \cap (A \times C)$:

We know that $A \cap A = A$. By the previous property we have that

$$A \times (C \cap D) = (A \cap A) \times (B \cap C) = (A \times B) \cap (A \times C)$$

3. $(A \times B) \cap (B \times A) = (A \cap B) \times (A \cap B)$:

By property 1 we have

$$(A \times B) \cap (B \times A) = (A \cap B) \times (B \cap A) = (A \cap B) \times (A \cap B)$$

4. $(A \cup B) \times (C \cup D) = (A \times C) \cup (B \times D) \cup (A \times D) \cup (B \times C)$:

Let $(x, y) \in (A \cup B) \times (C \cup D)$, then by definition of Cartesian product and the union of sets we have that $(x, y) \in (A \cup B) \times (C \cup D)$ if and only if $x \in A$ or $x \in B$ and $y \in C$ or $y \in D$.

$x \in A$ or $x \in B$ and $y \in C$ or $y \in D$ will occur if and only if $(x \in A$ or $x \in B$ and $y \in C)$ or $(x \in A$ or $x \in B$ and $y \in D)$.

$(x \in A$ or $x \in B$ and $y \in C)$ or $(x \in A$ or $x \in B$ and $y \in D)$ occurs if and only if $(x \in A$ and $y \in C)$ or $(x \in B$ and $y \in C)$ or $(x \in A$ and $y \in D)$ or $(x \in B$ and $y \in D)$.

By the definition of the Cartesian product we have that $(x \in A$ and $y \in C)$ or $(x \in B$ and $y \in C)$ or $(x \in A$ and $y \in D)$ or $(x \in B$ and $y \in D)$ if and only if $(x, y) \in A \times C$ or $(x, y) \in A \times D$ or $(x, y) \in B \times C$ or $(x, y) \in B \times D$. Hence by the definition of the union of two sets, $(x, y) \in A \times C$ or $(x, y) \in A \times D$ or $(x, y) \in B \times C$ or $(x, y) \in B \times D$ occurs if and only if $(x, y) \in (A \times C) \cup (B \times D) \cup (A \times D) \cup (B \times C)$.

5. $A \times (B \cup C) = (A \times B) \cup (A \times C)$:

We know $A = A \cup A$ and so by the previous property we have that

$$\begin{aligned} A \times (B \cup C) &= (A \cup A) \times (B \cup C) \\ &= (A \times B) \cup (A \times C) \cup (A \times C) \cup (A \times B) \\ &= (A \times B) \cup (A \times C) \end{aligned}$$

6. $(B \cup C) \times A = (B \times A) \cup (C \times A)$:

Again $A = A \cup A$ and so by property 4 we have

$$\begin{aligned} (B \cup C) \times A &= (B \cup C) \times (A \cup A) \\ &= (B \times A) \cup (B \times A) \cup (C \times A) \cup (C \times A) \\ &= (B \times A) \cup (C \times A) \end{aligned}$$

7. If $A \subseteq B$ and $C \subseteq D$ then $A \times C \subseteq B \times D$. Moreover if $A \neq \emptyset$ and $C \neq \emptyset$ then

$$A \times C \subseteq B \times D \iff A \subseteq B \text{ and } C \subseteq D$$

:

Let $A \subseteq B$ and $C \subseteq D$. If $A = \emptyset$ or $C = \emptyset$ then by lemma 1.2.1 we have $A \times C = \emptyset$ and by proposition 1.2.6 we have $A \times C = \emptyset \subseteq B \times D$.

So suppose that $A \neq \emptyset$ and $C \neq \emptyset$ then lemma 1.2.1 gives $A \times C \neq \emptyset$. Then we have that $(x, y) \in A \times C$ if and only if $x \in A$ and $y \in C$. We have $A \subseteq B$ so $x \in B$ and $C \subseteq D$ so $y \in D$, hence $(x, y) \in B \times D$. Hence $A \times C \subseteq B \times D$.

It is left to prove that if $A \neq \emptyset$ and $C \neq \emptyset$ and $A \times C \subseteq B \times D$, then $A \subseteq B$ and $C \subseteq D$. Suppose $A \times C \subseteq B \times D$. If $A = \emptyset$ then $A \times C = \emptyset$ by lemma 1.2.1 and $A \times C = \emptyset \subseteq B \times D$ irrespective of C ,

so C need not be a subset of D . Likewise if $C = \emptyset$ then $A \times C = \emptyset \subseteq B \times D$ irrespective of A so A need not be a subset of B .

So suppose that $A \neq \emptyset$ and $C \neq \emptyset$ then $\exists x \in A$ and $\exists y \in C$ such that $(x, y) \in A \times C$, we have that $A \times C \subseteq B \times T$ and so $(X, y) \in B \times D$ so $x \in B$ and $y \in D$.

Hence for $A \neq \emptyset$ and $C \neq \emptyset$, we have that $A \subseteq B$ and $C \subseteq D$ gives $A \times C \subseteq B \times D$ and $A \times C \subseteq B \times D$ gives $A \subseteq B$ and $C \subseteq D$. Hence we have

$$A \times C \subseteq B \times D \iff A \subseteq B \text{ and } C \subseteq D$$

8. If $A \subseteq B$ then $A \times C \subseteq B \times C$:

Let A be such that $A \subseteq B$. We have for any set C that $C \subseteq C$, hence by the previous property we know that

$$A \subseteq B \text{ and } C \subseteq C \Rightarrow A \times C \subseteq B \times C$$

9. If $C \subseteq D$ then $A \times C \subseteq A \times D$:

Let C be such that $C \subseteq D$. We have that $A \subseteq A$ and so by property 7 we have that

$$A \subseteq A \text{ and } C \subseteq D \Rightarrow A \times C \subseteq A \times D$$

10. $A \times (B \setminus C) = (A \times B) \setminus (A \times C)$:

Let $(x, y) \in A \times (B \setminus C)$ then we have that $(x, y) \in A \times (B \setminus C)$ if and only if $x \in A$ and $y \in B \setminus C$. $y \in B \setminus C$ means that $y \in B$ and $y \notin C$. Thus, $x \in A$ and $y \in B$ and $y \notin C$ happens if and only if $(x, y) \in A \times B$ and $(x, y) \notin A \times C$. Hence by definition of the difference of two sets we have that $(x, y) \in A \times B$ and $(x, y) \notin A \times C$ if and only if $(x, y) \in (A \times B) \setminus (A \times C)$.

11. $(A \setminus B) \times C = (A \times C) \setminus (B \times C)$:

Let $(x, y) \in (A \setminus B) \times C$ then we have that $(x, y) \in (A \setminus B) \times C$ if and only if $x \in A \setminus B$ and $y \in C$, moreover $x \in A \setminus B$ means that $x \in A$ and $x \notin B$. Hence $x \in A$ and $x \notin B$ and $y \in C$ occurs if and only if $(x, y) \in A \times C$ and $(x, y) \notin B \times C$. Hence by definition we have that $(x, y) \in A \times C$ and $(x, y) \notin B \times C$ if and only if $(x, y) \in (A \times C) \setminus (B \times C)$.

12. $(A \times B) \setminus (C \times D) = (A \times (B \setminus D)) \cup ((A \setminus C) \times B)$:

Let $(x, y) \in (A \times B) \setminus (C \times D)$, then we have that $(x, y) \in A \times B$ and $(x, y) \notin C \times D$, which happens if and only if $x \in A$ and $y \in B$ and $x \notin C$ and $y \notin D$. Now, $x \in A$ and $y \in B$ and $x \notin C$ and $y \notin D$ means that either $x \in A$ and $y \in B$ and $x \notin C$ or $x \in A$ and $y \in B$ and $y \notin D$. In the first case, $x \in A$ and $y \in B$ and $x \notin C$, we have that $x \in A \setminus C$ and $y \in B$, in the second case, $x \in A$ and $y \in B$ and $y \notin D$ we have $x \in A$ and $y \in B \setminus D$.

$x \in A$ and $y \in B$ and $x \notin C$ or $x \in A$ and $y \in B$ and $y \notin D$ occurs if and only if $x \in A \setminus C$ and $y \in B$ or $x \in A$ and $y \in B \setminus D$. Now by the definition of the Cartesian product we have that $x \in A \setminus C$ and $y \in B$ gives us that $(x, y) \in (A \setminus C) \times B$ and $x \in A$ and $y \in B \setminus D$ gives us $(x, y) \in A \times (B \setminus D)$.

Hence $x \in A \setminus C$ and $y \in B$ or $x \in A$ and $y \in B \setminus D$ occurs if and only if $(x, y) \in (A \setminus C) \times B$ or $(x, y) \in A \times (B \setminus D)$, from which we deduce that $(x, y) \in (A \setminus C) \times B$ or $(x, y) \in A \times (B \setminus D)$ if and only if $(x, y) \in (A \setminus C) \times B \cup A \times (B \setminus D)$.

13. Suppose $A \subseteq C$ and $B \subseteq D$ and consider $C \setminus A$ and $D \setminus B$. We have

$$(C \setminus A) \times D = (C \times D) \setminus (A \times D)$$

$$C \times (D \setminus B) = (C \times D) \setminus (C \times B)$$

Recall that $C \setminus A = \{x : x \in C \text{ and } x \notin A\}$. Now we have by property 11. that

$$(C \setminus A) \times D = (C \times D) \setminus (A \times D)$$

Likewise, by property 10. we have that

$$C \times (D \setminus B) = (C \times D) \setminus (C \times B)$$

Hence the result has been shown. \square

1.2.1.3.4 Power Set We make one final definition of an elementary operation for sets.

Definition 1.2.19. *Power set*

Let S be a set. We define the power set of the set S , denoted $P(S)$ to be the set which contains all of the possible subsets of S .

Example 1.2.15. Let $S = \{1, 2, 3\}$ then we have that

$$P(S) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, S\}$$

1.2.1.4 Set Partitions

Recall the idea of disjoint sets, that is if X and Y are sets then X and Y are disjoint if $X \cap Y = \emptyset$. This is saying that X and Y have no elements in common. Now suppose we have a set S such that $X \cup Y = S$ but $X \cap Y = \emptyset$. Then S is made of two distinct pieces. Of course there is nothing special about S being made of only two pieces, and could be made of many many pieces. We capture this idea in the next definition.

Definition 1.2.20. *Partition of a set*

Let S be a set and define \mathbb{S} to be the set of subsets of S . We say that \mathbb{S} is a partition of S if the following hold.

1. $\forall S_1, S_2 \in \mathbb{S}$ we have $S_1 \cap S_2 = \emptyset$ whenever $S_1 \neq S_2$
2. Taking the union of every $T \in \mathbb{S}$ gives us S that is

$$S = \bigcup_{T \in \mathbb{S}} T$$

3. $\forall T \in \mathbb{S}$ we have that $T \neq \emptyset$.

If the number of sets in \mathbb{S} is finite with say n elements then we call \mathbb{S} an n -component partition

Example 1.2.16. Let $S = \{1, 2, 3, 4\}$ and let $S_1 = \{2, 4\}$ and $S_2 = \{1, 3\}$. Then S_1 and S_2 partition S . Interestingly we have that $S_1^C = S_2$ and $S_2^C = S_1$, so the complements of these sets still forms a partition

If instead we have $S_3 = \{1\}$ and $S_4 = \{2, 3, 4\}$ then we also have a partition where the complements are also a partition. Now if $S_5 = \{2\}$, $S_6 = \{1, 3\}$ and $S_7 = \{4\}$ then S_5, S_6 and S_7 is a partition of S .

The fact in the first two examples we had two sets partitioning S where the complements also partitioned S is not a coincidence.

Proposition 1.2.13. *Complements of 2-component partition is partition*

Let S be a set such that $A \subseteq S$ and $B \subseteq S$ is a 2-component partition for S . We have that A and B partition S if and only if A^C and B^C partition S .

Proof:

(\Rightarrow) : Suppose that $A \subseteq S$ and $B \subseteq S$ partition S . By definition we have that

1. $A \cap B = \emptyset$
2. $A \cup B = S$

3. $A \cap B = \emptyset$ and $B \neq \emptyset$

We need to show that A^C and B^C is a partition that is

1. $A^C \cap B^C = \emptyset$

2. $A^C \cup B^C = S$

3. $A^C \neq \emptyset$ and $B^C \neq \emptyset$

1. $A^C \cap B^C = \emptyset$:

As $A \cup B = S$ we have on taking the complement of both sides that

$$\begin{aligned} A \cup B &= S \\ (A \cup B)^C &= S^C \\ A^C \cap B^C &= \emptyset \end{aligned}$$

So $A^C \cap B^C = \emptyset$.

2. $A^C \cup B^C = S$:

Likewise as $A \cap B = \emptyset$ then on taking the complement of both sides we have that

$$\begin{aligned} A \cap B &= \emptyset \\ (A \cap B)^C &= \emptyset^C \\ A^C \cup B^C &= S \end{aligned}$$

So $A^C \cup B^C = S$.

3. $A^C \neq \emptyset$ and $B^C \neq \emptyset$:

Suppose that $A^C = \emptyset$ then by taking the complement of both sides we have that $A = S$ which implies $B = \emptyset$, which is a contradiction as A and B partition S . Likewise if we suppose that $B^C = \emptyset$ we will have to conclude that $B = S$ which will be a contradiction. It thus follows that neither A^C or B^C can be empty.

Hence $A^C \neq \emptyset$ and $B^C \neq \emptyset$.

It follows that A^C and B^C is a partition of S

(\Leftarrow): Suppose that A^C and B^C is a partition of S . We have that $A^C \subseteq S$ and $B^C \subseteq S$. By the previous part we have that $(A^C)^C$ and $(B^C)^C$ is a partition of S . However $(A^C)^C = A$ and $(B^C)^C = B$. Thus A and B is a partition of S

The result now follows. \square

There are some additional results we can state about partitions that relate to the operations we can do on sets. We will require the following lemma.

Lemma 1.2.2. *Set difference and intersection are disjoint sets*

Let S and T be two sets. We have that $S \setminus T$ and $S \cap T$ are disjoint sets, which is to say that

$$(S \setminus T) \cap (S \cap T) = \emptyset$$

Proof:

Suppose that $x \in (S \setminus T) \cap (S \cap T)$ then by definition $x \in S \setminus T$ and $x \in S \cap T$. As $x \in S \setminus T$ then we have that $x \in S$ and $x \notin T$, likewise as $x \in S \cap T$ then $x \in S$ and $x \in T$. It is clear that no such x can exist hence $(S \setminus T) \cap (S \cap T) = \emptyset$.

1.2.1.5 A brief look at Zermelo–Fraenkel set theory

At the start of this section we introduced the idea of Zermelo–Fraenkel set theory. This is the complete formalisation of set theory and the true bedrock of mathematics. The Zermelo–Fraenkel set theory axioms, hence now referred to as ZF, are given as follows.

Definition 1.2.21. *Zermelo–Fraenkel set theory axioms*

The Zermelo–Fraenkel set theory axioms are the following.

1. *The axiom of extensionality:*

The axiom of extensionality asserts that two sets are equal if and only if they contain the same elements.

2. *The axiom of the empty-set:*

The axiom of the empty-set asserts that there exists a set which contains no elements

3. *The axiom of pairing:*

The axiom of pairing asserts that given any set A and any set B , there is a set C such that, given any set D , D is a member of C if and only if D is equal to A or D is equal to B . This is to say, given two sets, there is a set whose members are exactly the two given sets.

4. *The axiom of specification:*

The axiom of specification asserts that we can construct a set which satisfies a given condition, so long as this condition is not inherently contradictory.

5. *The axiom of unions:*

The axiom of unions asserts that we can perform the union of two sets A and B

6. *The axiom of powers:*

The axiom of powers asserts that for any set S we can construct a set $P(S)$ whose elements are all the possible subsets of S .

7. *The axiom of infinity:*

The axiom of infinity asserts that there is at least one infinite set A , that is at least one set with infinitely many elements. That is we have a set A such that the $\emptyset \in A$ and if $x \in A$ then the set $x \cup \{x\}$ is also in A .

8. *The axiom of replacement:*

We will need the next section to fully understand this axiom, however informally asserts that for some set S , and form another set by replacing the elements of S by other sets according to any definite rule.

9. *The axiom of foundation:*

The axiom of foundation asserts that for every non-empty set S , there exists an element $x \in S$ such that x and S are disjoint. This also asserts that no set can contain itself.

There is also one axiom which we have left off. This is the controversial axiom of choice.

Definition 1.2.22. *The axiom of choice*

Let S be a set of non-empty sets. The axiom of choice asserts that there is a way to pick an element of each of the sets in S .

With the axiom of choice we have the following

Definition 1.2.23. *ZFC axioms*

The axioms of ZF along with the axiom of choice gives us the ZFC axioms

We can already see that our “hands-on” approach to set theory has somewhat indirectly captured the essence of the ZF axioms. We can use the ZF axiom to prove in a truly rigorous way what we did with out “hands-on” approach. Although an interesting field of study itself, we will not really need to use the ZF axioms, although occasionally we may rely on choice.

There is one other thing that needs bringing up, ZFC has one more component, the axioms alone are not enough to prove anything. We need the notion of inclusion, that is being an element of a set. That is we include the symbol \in along with the axioms, where \in takes on the meaning we defined earlier. With this we can in theory use ZFC to start proving and building up mathematics from the bedrock.

1.2.2 Mappings

1.2.2.1 Introduction and basic definitions

Now that we have the of a set what can we use it for? Many areas of mathematics can be broken down into the theory of sets, in particular how we can get from one set to another. Without this idea we wouldn’t be able to get very far at all. As an example, you may have seen, in a calculus course for example, the idea of a function $f(x)$, say $f(x) = x^2$ where x can be any number we choose. Say $x = 2$ then $f(2) = 4$. You may have also seen functions where we are not allowed to use any number we wish for example, if we take $f(x) = \sqrt{x}$ then we are only allowed positive numbers if we want a to find an answer using the numbers we are familiar with, such as $1, 88.125, \pi, \sqrt{2}$ etc. This set we will denote by \mathbb{R} . The alert reader may now see how sets will come into play, to define in a rigorous way the ideas of $f(x) = x^2$ and other such functions, we need to consider what are the allowable inputs which once done will give us the possible outputs. That is if we have a set whose elements are inputs and we define some form of function, which we will now call a map, then we will get another set whose elements are what inputs will be ‘mapped’ to.

Definition 1.2.24. Mapping

Let X and Y be sets. Suppose we have some rule or description, which we will denote by f , by which for each $x \in X$ there is some element $f(x) \in Y$. We say that the rule (description) is a mapping or map or function from X to Y . We denote a mapping with the following notation

$$\begin{aligned} f : X &\rightarrow Y \\ x &\mapsto f(x) \end{aligned}$$

where the first line tells us what sets the mapping is between, and the bottom line tells us where each element $x \in X$ gets mapped to

Definition 1.2.25. Domain

Let $f : X \rightarrow Y$ be a mapping between two sets X and Y . We say that the set X is the domain of the mapping f . The domain contains the elements which the map can act on. We can write this as

$$\text{Dom}(f) = X$$

Definition 1.2.26. Co-Domain

Let $f : X \rightarrow Y$ be a mapping between two sets X and Y . We say that the set Y is the Co-domain of the mapping f . The co-domain contains the possible elements that the map can send elements of X to. We can write this as

$$\text{Cdm}(f) = Y$$

We have some examples of mappings.

Example 1.2.17. Let $X = \{1, 2, 3\}$ and let $Y = X$. Define the map

$$\begin{aligned} f : X &\rightarrow Y \\ x &\mapsto f(x) = x \end{aligned}$$

To see what f does we will take each element of X one at a time. Starting with 1 we have that $1 \mapsto f(1) = 1$, for 2 we have $2 \mapsto f(2) = 2$ and finally $3 \mapsto f(3) = 3$. Hence the map f takes an element of X and leaves it alone. A map which takes every element of its domain and leaves it alone is called an identity map, or if you prefer the do nothing at all map.

Example 1.2.18. Let $X = Y = \mathbb{N}$. Let f be the map given by

$$\begin{aligned} f : X &\rightarrow Y \\ x &\mapsto f(x) = 2x \end{aligned}$$

It is clear to see that every element in the domain gets doubled, i.e $f(1) = 2$, $f(2) = 4$, $f(3) = 6$ and so on.

A map does not need to be given by an explicit mathematical formulae

Example 1.2.19. Let $A =$ The set of all humans currently alive on planet earth, from which it should be clear to see that $You \in A$ ⁵. Let $B = \{0, 1\}$. Let f be the mapping given by

$$\begin{aligned} f : A &\rightarrow B \\ a &\mapsto f(a) = \begin{cases} 1, & \text{If } a \text{ has hair on their head} \\ 0, & \text{If } a \text{ does not have hair on their head} \end{cases} \end{aligned}$$

Then f is a map which indicates if a given person has hair on their head or not.

The above definition of a mapping can be made more general

Definition 1.2.27. Piecewise mapping

Let $f : X \rightarrow Y$ be a mapping. We say that f is a piecewise mapping if we need multiple rules or descriptions to fully describe f . That we wish to define the mapping using different rules based on the input. If for each of this input ranges we define a mapping g_1, g_2, g_3, \dots then we can write the piecewise function as follows

$$\begin{aligned} f : X &\rightarrow Y \\ x &\mapsto f(x) = \begin{cases} g_1(x), & \text{Condition for } g_1 \\ g_2(x), & \text{Condition for } g_2 \\ g_3(x), & \text{Condition for } g_3 \\ \dots \end{cases} \end{aligned}$$

Example 1.2.20. Let $f : \mathbb{N} \rightarrow \mathbb{N}$ be defined by

$$\begin{aligned} f : \mathbb{N} &\rightarrow \mathbb{N} \\ x &\mapsto f(x) = \begin{cases} 2x, & \text{If } x \nmid 5 \\ 5x, & \text{Otherwise} \end{cases} \end{aligned}$$

We have that $f(1) = 2$, $f(2) = 4$ and so on up to $f(4) = 8$, then $f(5) = 25$ and so on.

We make one more useful definition that will be useful throughout the rest of the text,

Definition 1.2.28. Closure of a mapping

Let X be a set. If we have a mapping such that $f : X^n \rightarrow X$. We say the mapping has closure on the set X , or we say that f is a closed mapping.

⁵Unless you are either not a human or somehow reading this in some unknown form of existence

1.2.2.2 The image and pre-image

We now define a more technical notion of how a mapping f maps an element in the domain to the co-domain.

Definition 1.2.29. *Image of an element*

Let $f : X \rightarrow Y$ be a mapping of between two sets X and Y , and let $x \in X$ be an element of the domain. We say that $f(x) \in Y$ is the image of the element x .

Which in turn allows us to define a subset of the co-domain for which every element $x \in X$ gets mapped to

Definition 1.2.30. *Image of a mapping*

Let $f : X \rightarrow Y$ be a mapping of between two sets X and Y . We define the set

$$\text{Image}(f) = f(X) = \{f(x) : x \in X\} \subseteq Y$$

To be the image of the domain, sometimes called the range of f . That is the image is the set of all possible outputs of the mapping f with the domain X .

Moreover, suppose that $A \subseteq X$ then we define the image of the subset A to be

$$f(A) = \{f(x) : x \in A\} \subseteq f(X) \subseteq Y$$

That is we can consider the image of subsets of X .

Example 1.2.21. Consider the mapping in example 1.2.18, we have that $X = Y = \mathbb{N}$ and is f the map

$$\begin{aligned} f : X &\rightarrow Y \\ x &\mapsto f(x) = 2x \end{aligned}$$

then we have that $\text{Image}(f) = f(\mathbb{N}) = \{2x : x \in \mathbb{N}\}$

Example 1.2.22. Let f be an arbitrary mapping such that $f : \emptyset \rightarrow Y$ for some set Y . What is $\text{Image}(f)$? We have by the definition of a mapping 1.2.30, we have that

$$\text{Image}(f) = \{f(x) : x \in \emptyset\}$$

However, we know that the empty set has no elements, so there are no elements that f can send anything to, so $\text{Image}(f) = \emptyset$.

Likewise we can define how a mapping is mapped to from the domain to the co-domain. This is called the pre-image.

Definition 1.2.31. *Pre-image of an element*

Let $f : X \rightarrow Y$ be a mapping of between two sets X and Y , and let $y \in Y$ be an element of the co-domain. If $f(x) = y$ then we say that $f(x) \in X$ is the pre-image of the element y and we denote this $f^{-1}(y)$.

Which in turn allows us to define a subset of the domain for which every element $y \in Y$ gets mapped to

Definition 1.2.32. *Pre-image of a mapping*

Let $f : X \rightarrow Y$ be a mapping of between two sets X and Y . We define the set

$$\text{PreImage}(f) = f^{-1}(Y) = \{x \in X : f(x) \in Y\} \subseteq X$$

To be the pre-image of the co-domain. That is the pre-image is the set of all possible inputs that give the given outputs.

Moreover, suppose that $B \subseteq Y$ then we define the pre-image of the subset B to be

$$f^{-1}(B) = \{x \in X : f(x) \in B\} \subseteq f^{-1}(Y) \subseteq X$$

Example 1.2.23. Consider the mapping $f : \mathbb{N} \rightarrow \mathbb{N}$ given by

$$\begin{aligned} f : \mathbb{N} &\rightarrow \mathbb{N} \\ x &\mapsto f(x) = \frac{x}{2} \end{aligned}$$

We have that $\frac{x}{2}$ is defined in the naturals only when x is an even number, hence the pre-image must consist of the even numbers.

$$\text{PreImage}(f) = f^{-1}(\mathbb{N}) = \left\{ x \in \mathbb{N} : \frac{x}{2} \in \mathbb{N} \right\} = \{0, 2, 4, 6, 8, \dots\}$$

Example 1.2.24. Consider the mapping $f : \mathbb{N} \rightarrow \mathbb{N}$ given by

$$\begin{aligned} f : \mathbb{N} &\rightarrow \mathbb{N} \\ x &\mapsto f(x) = x^2 \end{aligned}$$

We have that the pre-image is given by

$$\text{PreImage}(f) = \{x \in \mathbb{N} : x^2 \in \mathbb{N}\} = \{0, 1, 2, 3, 4, \dots\} = \mathbb{N}$$

With these definitions we can make the following observations

Proposition 1.2.14. *Properties of the image and pre-image*

Let $f : X \rightarrow Y$ be a mapping and let $A \subseteq X$ and $B \subseteq Y$. We have that the following properties hold for the image and pre-image

1. $f(X) \subseteq Y$
2. $f(f^{-1}(Y)) = f(X)$
3. $f(f^{-1}(B)) \subseteq B$
4. $f(f^{-1}(B)) = B \cap f(X)$
5. $f(f^{-1}(f(A))) = f(A)$
6. $f(A) = \emptyset \iff A = \emptyset$
7. $B \subseteq f(A) \iff \exists C \subseteq A : f(C) = B$
8. $f(X \setminus A) \subseteq f(A) \iff f(A) = f(X)$
9. $f(X) \setminus f(A) \subseteq f(X \setminus A)$
10. $f(A \cup f^{-1}(B)) \subseteq f(A) \cup B$
11. $f(A \cap f^{-1}(B)) = f(A) \cap B$

Likewise the following properties hold for the pre-image

1. $f^{-1}(Y) = X$
2. $f^{-1}(f(X)) = X$
3. $A \subseteq f^{-1}(f(A))$

4. Suppose that instead of the mapping $f : X \rightarrow Y$ we consider a new mapping based on f , which we call \bar{f} . We define \bar{f} to be the mapping

$$\begin{aligned}\bar{f} : A &\rightarrow Y \\ x &\mapsto \bar{f}(x) = f(x)\end{aligned}$$

that is \bar{f} maps every element of $A \in A$ to what $f(a)$ does. With this new mapping we have the following property

$$(\bar{f})^{-1}(B) = A \cap f^{-1}(B)$$

5. $f^{-1}(f(f^{-1}(B))) = f^{-1}(B)$
6. $f^{-1}(B) = \emptyset \iff B \subseteq Y \setminus f(X)$
7. $A \subseteq f^{-1}(B) \iff f(A) \subseteq B$
8. $f^{-1}(Y \setminus B) \subseteq f^{-1}(B) \iff f^{-1}(B) = X$
9. $f^{-1}(Y \setminus B) = X \setminus f^{-1}(B)$
10. $A \cup f^{-1}(B) \subseteq f^{-1}(f(A) \cup B)$
11. $A \cap f^{-1}(B) \subseteq f^{-1}(f(A) \cap B)$

Proof:

We start with the properties of the image.

1. $f(X) \subseteq Y$:

This holds by definition of the image.

2. $f(f^{-1}(Y)) = f(X)$:

Let $x \in f(f^{-1}(Y))$ and recall the definition of the image and pre-image.

$$\begin{aligned}f(A) &= \{f(x) : x \in A\} \subseteq f(X) \subseteq Y \\ f^{-1}(B) &= \{x \in X : f(x) \in B\} \subseteq f^{-1}(Y) \subseteq X\end{aligned}$$

We have that

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

Hence $x \in f(f^{-1}(Y))$ means that $x = f(y)$ for some $y \in f^{-1}(Y)$, additionally we conclude that $y \in X$. Moreover by the definition of the pre-image we have that $f^{-1}(Y) \subseteq X$. It thus follows that $x \in f(X)$ and so $f(f^{-1}(Y)) \subseteq f(X)$.

Now suppose that $x \in f(X)$, that is $x = f(x')$ for some $x' \in X$. Now by definition of the pre-image as $x' \in X$ with $f(x') \in Y$ we have that $x' \in f^{-1}(Y)$. Hence by definition of the set $f(f^{-1}(Y))$ we must conclude that $f(x') \in f(f^{-1}(Y))$, which is to say $x \in f(f^{-1}(Y))$. Hence $f(X) \subseteq f(f^{-1}(Y))$.

It follows that $f(f^{-1}(Y)) = f(X)$.

3. $f(f^{-1}(B)) \subseteq B$:

Suppose that $x \in f(f^{-1}(B))$ where $B \subseteq Y$. We hence have that $x = f(b)$ for some $b \in f^{-1}(B)$, hence $b \in X$ giving us $f(b) \in B$ and so $f(f^{-1}(B)) \subseteq B$.

4. $f(f^{-1}(B)) = B \cap f(X)$:

Let $x \in f(f^{-1}(B))$ then by property 3 we have that $x \in B$. Additionally as $x \in f(f^{-1}(B))$ and $B \subseteq Y$ then $f(f^{-1}(B)) \subseteq f(f^{-1}(Y))$ and so $x \in f(f^{-1}(Y))$. Now by property 2 we have that $f(f^{-1}(Y)) = f(X)$ thus $x \in f(X)$ and so $x \in B \cap f(X)$. It follows that $f(f^{-1}(B)) \subseteq B \cap f(X)$.

Now suppose that $x \in B \cap f(X)$. By definition of $f(X)$ we have $x \in f(X)$ gives us that $x = f(x')$ where $x' \in X$, moreover we also have that $x \in B$. Now we have the set $f(f^{-1}(B))$ is given by

$$f(f^{-1}(B)) = \{f(b) : b \in f^{-1}(B)\}$$

We have that $x = f(x')$ and so $x' \in f^{-1}(B)$, hence clearly by definition of the image we have that $x \in f(f^{-1}(B))$. It follows that $B \cap f(X) \subseteq f(f^{-1}(B))$.

Hence the result $f(f^{-1}(B)) = B \cap f(X)$.

5. $f(f^{-1}(f(A))) = f(A)$:

By property 4 we have that

$$f(f^{-1}(f(A))) = f(A) \cap f(X)$$

as $f(A) \subseteq Y$. Finally $f(A) \cap f(X) = f(A)$ as $f(A) \subseteq f(X)$. The result follows.

6. $f(A) = \emptyset \iff A = \emptyset$:

(\Leftarrow): Suppose that $f(A) = \emptyset$. By definition of the image we have that

$$f(A) = \{f(x) : x \in A\}$$

By set equality we must have that $f(A) = \{f(x) : x \in A\} = \emptyset$. Hence there can be no elements $f(x)$ where $x \in A$ which can only occur if $A = \emptyset$ for if not then $f(A)$ has at least one element for some $x' \in A$, contradicting the fact that $f(A) = \emptyset$. It follows that $A = \emptyset$.

(\Rightarrow): Suppose that $A = \emptyset$, we have that the image of the empty set is given by

$$f(A) = f(\emptyset) = \{f(x) : x \in \emptyset\} = \emptyset$$

It follows that $f(A) = \emptyset$.

7. $B \subseteq f(A) \iff \exists C \subseteq A : f(C) = B$:

(\Rightarrow): Suppose that $B \subseteq f(A)$. We show that $\exists C \subseteq A : f(C) = B$. So, suppose that $x \in B$ then we have that $x \in f(A)$ by assumption. By definition of the image we have that

$$f(A) = \{f(x) : x \in A\}$$

Hence we have $x \in f(A)$ gives us that $x = f(x')$ for some $x' \in A$. We define the required set C as follows.

$$C = \bigcup_{\substack{x' \in A \\ f(x') \in B}} x'$$

That is C is defined to be those elements $x' \in A$ such that $f(x') \in B$ which is a subset of $f(A)$. Clearly $C \subseteq A$ as each $x' \in C$ is by construction an element of A . Additionally we also have $f(C) = B$ by construction of C .

(\Leftarrow): Suppose that $\exists C \subseteq A : f(C) = B$. As $f(C) = B$ we have by the definition of the image that

$$f(C) = \{f(x) : x \in C\}$$

that is $x \in f(C)$ gives $x = f(c)$ for some $c \in C$ and additionally $x \in B$ by assumption. Now $C \subseteq A$ so $c \in A$. Hence $x \in f(A)$, hence we must conclude that $B \subseteq f(A)$, possibly being equal if $C = A$.

The result follows.

8. $f(X \setminus A) \subseteq f(A) \iff f(A) = f(X)$:

(\Rightarrow): Suppose that $f(X \setminus A) \subseteq f(A)$ and recall the definition of the complement of sets. We have that

$$X \setminus A = \{x \in X : x \notin A\}$$

Now, $A \subseteq X$ by hypothesis of the proposition. So if $x \in f(X \setminus A)$ then by definition of the image we have that

$$f(X \setminus A) = \{f(x) : x \in X \setminus A\} = \{f(x) : x \in X \text{ and } x \notin A\}$$

but then if $x \notin A$ then $x \notin f(A)$. However if $A = X$ then we have that $X \setminus A = \emptyset$ from which it follows by property 6 that $f(X \setminus A) = \emptyset$ and so as the empty set is a subset of any set we conclude that $\emptyset \subseteq f(A)$, that is we must have $f(A) = f(X)$.

(\Leftarrow): Suppose that $f(A) = f(X)$, by definition of the image we have that

$$f(A) = \{f(a) : a \in A\} = \{f(x) : x \in X\} = f(X)$$

Now consider $f(X \setminus A)$ this set is given by

$$f(X \setminus A) = \{f(x) : x \in X \setminus A\} = \{f(x) : x \in X \text{ and } x \notin A\}$$

But as all such $x \in A$ must also be $x \in X$ by assumption we conclude that $f(X \setminus A) = \emptyset$ and the empty set is clearly contained in any other set. Hence $f(X \setminus A) \subseteq f(A)$. The result has now been shown.

9. $f(X) \setminus f(A) \subseteq f(X \setminus A)$:

Let $x \in f(X) \setminus f(A)$. By definition we have that

$$f(X) \setminus f(A) = \{x \in f(X) : x \notin f(A)\}$$

Hence $x \in f(X) \setminus f(A)$ gives us that $x \in f(X)$ and $x \notin f(A)$. That is $\exists y \in X$ with $y \notin A$ such that $x = f(y)$, this is $y \in X \setminus A$. Hence it follows that $x \in f(X \setminus A)$. That is $f(X) \setminus f(A) \subseteq f(X \setminus A)$.

10. $f(A \cup f^{-1}(B)) \subseteq f(A) \cup B$:

Let $x \in f(A \cup f^{-1}(B))$. This is our first usage of the pre-image of a set so we recall the definition, we have that

$$f^{-1}(B) = \{x \in X : f(x) \in B\} \subseteq X$$

Hence the image $f(A \cup f^{-1}(B))$ is given by

$$\begin{aligned} f(A \cup f^{-1}(B)) &= \{f(y) : y \in A \cup f^{-1}(B)\} \\ &= \{f(y) : y \in A \text{ or } y \in f^{-1}(B)\} \\ &= \{f(y) : y \in A \text{ or } y \in X : f(y) \in B\} \end{aligned}$$

Now, $x \in f(A \cup f^{-1}(B))$ gives us that either $\exists y \in A$ with $x = f(y)$ or $\exists y \in X$ with $f(y) \in B$. In the first case where $\exists y \in A$ with $x = f(y)$ then by definition of the image we have that $x \in f(A)$ and

so is clearly in the union $f(A) \cup B$. Now for the second case we have that $x \in B$ as $y \in X$ such that $x = f(y) \in B$, likewise it is in the union $f(A) \cup B$.

Hence $x \in f(A) \cup B$ and we have that $f(A \cup f^{-1}(B)) \subseteq f(A) \cup B$. Hence the result.

11. $f(A \cap f^{-1}(B)) = f(A) \cap B$:

Let $x \in f(A \cap f^{-1}(B))$, the image of $A \cap f^{-1}(B)$ is given by

$$\begin{aligned} f(A \cap f^{-1}(B)) &= \{f(y) : y \in A \cap f^{-1}(B)\} \\ &= \{f(y) : y \in A \text{ and } y \in f^{-1}(B)\} \\ &= \{f(y) : y \in A \text{ and } y \in X : f(y) \in B\} \end{aligned}$$

Now $x \in f(A \cap f^{-1}(B))$ gives us that $\exists y \in A$ with $x = f(y)$ and $\exists y \in X$ with $f(y) \in B$. Hence we clearly have that $x \in f(A)$ and $x \in B$ and so is in the intersection $f(A) \cap B$. Hence we have that $f(A \cap f^{-1}(B)) \subseteq f(A) \cap B$.

Now suppose that $x \in f(A) \cap B$. We have that $x \in f(A)$ and $x \in B$, from the first of these having $x \in f(A)$ means that $\exists y \in A$ such that $x = f(y)$. Now as $x \in B$ means there is some $y' \in X$ with $x = f(y')$. However as $f(A) \cap B$ then we must have that $f(y') \in f(A)$ hence $y' \in A$. Hence both y and y' are in the set $A \cap f^{-1}(B)$ and so we have $x \in f(A \cap f^{-1}(B))$ and therefore $f(A) \cap B \subseteq f(A \cap f^{-1}(B))$.

The result $f(A \cap f^{-1}(B)) = f(A) \cap B$ follows.

We now turn our attention to the results for the pre-image.

1. $f^{-1}(Y) = X$:

By definition of the pre-image we have that

$$f^{-1}(Y) = \{x \in X : f(x) \in Y\} \subseteq X$$

Clearly $f^{-1}(Y) \subseteq X$ by definition. Now if $x \in X$ then we must also clearly have $f(x) \in Y$ and so $X \subseteq f^{-1}(Y)$. Hence $f^{-1}(Y) = X$.

2. $f^{-1}(f(X)) = X$:

Let $y \in f^{-1}(f(X))$, we have that the set $f^{-1}(f(X))$ is given by

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

It is hence clear that for any $x \in f^{-1}(f(X))$ we have clearly have $x \in X$, that is $f^{-1}(f(X)) \subseteq X$. Likewise if $x \in X$ then clearly $x \in f(X)$ and so by the definition of $f^{-1}(f(X))$ we have that $x \in f^{-1}(f(X))$. That is $X \subseteq f^{-1}(f(X))$. The result follows.

3. $A \subseteq f^{-1}(f(A))$:

Suppose that $x \in A \subseteq X$. By property 2. of the pre-image we have that $f^{-1}(f(X)) = X$. Hence $x \in A \subseteq f^{-1}(f(X)) = X$ giving the result.

4. Suppose that instead of the mapping $f : X \rightarrow Y$ we consider a new mapping based on f , which we we call \bar{f} . We define \bar{f} to be the mapping

$$\begin{aligned} \bar{f} : X &\rightarrow Y \\ x &\mapsto \bar{f}(x) = f(x) \end{aligned}$$

that is \bar{f} maps every element of $a \in A$ to what $f(a)$ does. With this new mapping we have the following property

$$(\bar{f})^{-1}(B) = A \cap f^{-1}(B) :$$

Let $x \in (\bar{f})^{-1}(B)$. We have that $(\bar{f})^{-1}(B)$ is given by

$$(\bar{f})^{-1}(B) = \{x \in A : \bar{f}(x) \in B\}$$

So $x \in (\bar{f})^{-1}(B)$ gives that $x \in A$, moreover as $\bar{f}(x) \in B$ and \bar{f} maps every $x \in A$ to $f(x)$ then $\bar{f}(x) = f(x) \in B$. It follows that $x \in f^{-1}(B)$ and so $x \in A \cap f^{-1}(B)$. Thus $(\bar{f})^{-1}(B) \subseteq A \cap f^{-1}(B)$.

Now, suppose that $x \in A \cap f^{-1}(B)$, by definition of \bar{f} we have that $\bar{f}(x) = f(x)$. Now $x \in f^{-1}(B)$ means that $f(x) \in B$, now as $\bar{f}(x)$ maps any $x \in A$ to $f(x)$ we have that $\bar{f}(x) = f(x)$ and so $x \in (\bar{f})^{-1}(B)$

Hence $(\bar{f})^{-1}(B) = A \cap f^{-1}(B)$

5. $f^{-1}(f(f^{-1}(B))) = f^{-1}(B)$:

This follows by property 2. $f^{-1}(f(X)) = X$. Indeed we have

$$f^{-1}(f(f^{-1}(B))) = f^{-1}(B)$$

6. $f^{-1}(B) = \emptyset \iff B \subseteq Y \setminus f(X)$:

(\Rightarrow) : Suppose $f^{-1}(B) = \emptyset$, by definition of the pre-image we have

$$f^{-1}(B) = \{x \in X : f(x) \in B\} = \emptyset$$

Hence the pre-image being empty means that there are no elements $x \in X$ with $f(x) \in B$. Now the set $Y \setminus f(X)$ is given

$$Y \setminus f(X) = \{y \in Y : y \notin f(X)\}$$

Thus as there are no $x \in X$ with $f(x) \in B$, then $Y \setminus f(X)$ will not remove any $f(x) \in B$, that is $B \subseteq Y \setminus f(X)$.

(\Leftarrow) : Suppose that $B \subseteq Y \setminus f(X)$. We Have that $Y \setminus f(X)$ is precisely the set of $y \in Y$ with $y \notin f(X)$, therefore the set $B \subseteq Y \setminus f(X)$ means that if $f(b) \in B$ then we have have that $b \notin f(X)$ and hence $b \notin X$. This holds for any $f(b) \in B$ and hence we must have that the pre-image of B is empty. This is to say $f^{-1}(B) = \emptyset$.

7. $A \subseteq f^{-1}(B) \iff f(A) \subseteq B$:

(\Rightarrow) : Suppose that $A \subseteq f^{-1}(B)$. Recall the definition of the image

$$f(A) = \{f(x) : x \in A\}$$

Now for some $a \in A$ we have that $a \in f^{-1}(B)$ and so there is some $x \in X$ such that $f(x) \in B$, in particular $a = x$ and so $x \in A$ which gives $f(A) \subseteq B$.

(\Leftarrow) : Now, suppose that $f(A) \subseteq B$ we have that for some $y \in f(A)$ that $y \in B$ and in particular by definition there is some $x \in A$ such that $f(x) = y \in f(A)$. Hence as $A \subseteq X$ we have that $x \in X$ and so by definition of the pre-image we have that $x \in f^{-1}(B)$. This is to say we conclude that $A \subseteq f^{-1}(B)$.

8. $f^{-1}(Y \setminus B) \subseteq f^{-1}(B) \iff f^{-1}(B) = X$:

Suppose that $f^{-1}(Y \setminus B) \subseteq f^{-1}(B)$. We have that pre-image of $Y \setminus B$ is given by

$$f^{-1}(Y \setminus B) = \{x \in X : f(x) \in Y \setminus B\} = \{x \in X : f(x) \in Y \text{ and } f(x) \notin B\}$$

Hence by definition $y \in f^{-1}(Y \setminus B)$ gives us that $y = x$ for some $x \in X$ with $f(x) \in Y$ and $f(x) \notin B$, but then we can't have $y \in f^{-1}(B)$ by the definition of the pre-image on B . Hence we conclude that $f^{-1}(Y \setminus B) \subseteq f^{-1}(B)$ holds if and only if $Y \setminus B = \emptyset$ from which $B = Y$ and so by property 1. we have that $f^{-1}(B) = X$.

9. $f^{-1}(Y \setminus B) = X \setminus f^{-1}(B)$:

Suppose that $x \in f^{-1}(Y \setminus B)$ then by definition we have that $f(x) \in Y$ and $f(x) \notin B$ for some $x \in X$, but this is clearly the definition of $X \setminus f^{-1}(B)$ and so $x \in X \setminus f^{-1}(B)$.

Conversely if $x \in X \setminus f^{-1}(B)$ then $f(x) \notin B$ but by definition of f we have that $f(x) \in Y$ and so $x \in f^{-1}(Y \setminus B)$.

It follows that $f^{-1}(Y \setminus B) = X \setminus f^{-1}(B)$.

10. $A \cup f^{-1}(B) \subseteq f^{-1}(f(A) \cup B)$:

Let $x \in A \cup f^{-1}(B)$. We have that either $x \in A$ or $x \in f^{-1}(B)$. If $x \in A$ then $f(x) \in f(A)$ and so $f(x) \in f(A) \cup B$, the result follows on taking the pre-image as

$$f^{-1}(f(A) \cup B) = \{x \in X : f(x) \in f(A) \cup B\}$$

This is to say that $x \in f^{-1}(f(A) \cup B) = \{x \in X : f(x) \in f(A) \cup B\}$.

Now if $x \in f^{-1}(B)$ then we have by definition that $f(x) \in B$ and by a similar argument to above we conclude that $f(x) \in f(A) \cup B$ so that on taking the pre-image we conclude that $x \in f^{-1}(f(A) \cup B) = \{x \in X : f(x) \in f(A) \cup B\}$.

Hence it follows that $A \cup f^{-1}(B) \subseteq f^{-1}(f(A) \cup B)$.

11. $A \cap f^{-1}(B) \subseteq f^{-1}(f(A) \cap B)$:

Suppose that $x \in A \cap f^{-1}(B)$ then $x \in A$ and $x \in f^{-1}(B)$ and so $f(x) \in B$. As $x \in A$ then $f(x) \in f(A)$ and hence as $f(x) \in f(A)$ and $f(x) \in B$ then $f(x) \in f(A) \cap B$. The result follows on taking the pre-image.

Hence $A \cap f^{-1}(B) \subseteq f^{-1}(f(A) \cap B)$

The proposition now follows. \square

1.2.2.3 Injective, surjective and bijective mappings

Armed with the examples we have seen we can make a few comments about mappings. Consider example 1.2.18 where we have that $X = Y = \mathbb{N}$ and is f the map

$$\begin{aligned} f : X &\rightarrow Y \\ x &\mapsto f(x) = 2x \end{aligned}$$

We have that for every $x, y \in X$ with $f(x) = f(y)$ that $x = y$, which is to say if the image of two different elements agree, then the elements are in-fact the same. This is clear to see, suppose that $x, y \in X$ with $f(x) = f(y)$, then we have that

$$\begin{aligned} f(x) &= f(y) \\ 2x &= 2y \\ x &= y \end{aligned}$$

Another way of expressing this idea is that two distinct elements in the domain will have distinct images, we say a mapping with this property is an injective mapping. Now, if we consider $\text{Image}(f) \subseteq Y$ and consider the map

$$g : X \rightarrow \text{Image}(f)$$

$$x \mapsto g(x) = 2x$$

Then, for every $y \in \text{Image}(f)$, we have that there exists some element $x \in X$ such that $y = g(x)$. Again, we can show this. Let $y \in \text{Image}(f)$, then we need to show that $\exists x \in X$ such that $g(x) = y$. Now

$$y = g(x)$$

$$y = 2x$$

$$\frac{y}{2} = x$$

We hence will need to take $x = \frac{y}{2}$, however we first then to verify that $x = \frac{y}{2} \in X$. We note that $y \in \text{Image}(f)$ means that $y = 2k$ for some $k \in \mathbb{N}$, so

$$x = \frac{y}{2}$$

$$x = \frac{2k}{2}$$

$$x = k$$

as $x \in X = \mathbb{N}$ and $k \in \mathbb{N}$ then we can rest safe in the knowledge that our choice for x indeed works. As a sanity check we have that

$$g(x) = 2x = 2\frac{y}{2} = y$$

This choice of x works for any choice of y . Another way to express this idea is that every element in the image of the mapping is the image of some element in the domain, we say a mapping with this property is a surjective mapping.

It is worth noting that the mapping g is both injective and surjective, this makes g a special type of mapping. If we take an element in the domain x and consider its image $g(x) \in \text{Image}(f)$, then as g is injective we know that $g(x)$ is a distinct element in $\text{Image}(f)$. Moreover, as g is surjective then there is an element in the domain, say a with the property that $g(a) = g(x)$, but as g is injective then we know that $a = x$. This means that we can go between elements of the domain and elements of the image in a distinct way, a mapping with this property is called a bijective mapping and the domain and image are said to be in bijection with each other.

We formalise these ideas now to a mapping between any two sets.

Definition 1.2.33. *Injective, surjective and bijective maps*

Let $f : X \rightarrow Y$ be a mapping between two sets X and Y .

1. *We say that f is an injective mapping, sometimes called a one-to-one mapping, if*

$$\forall x, y \in X, f(x) = f(y) \Rightarrow x = y$$

That is we have that $f(x) = f(y)$ for $x, y \in X$ then $x = y$. If we know that f is injective we can write the mapping as

$$f : X \hookrightarrow Y$$

which is read as f is an injective mapping from X to Y .

2. We say that f is a surjective mapping, sometimes called a onto mapping, if

$$\forall y \in Y, \exists x \in X : y = f(x)$$

That is we have that for each $y \in Y$, there exists some $x \in X$ such that $f(x) = y$. If we know that f is a surjective then we can write the mapping as

$$f : X \twoheadrightarrow Y$$

which is read as f is a surjective mapping from X to Y

3. We say that f is a bijective mapping, sometimes called a one-to-one and onto mapping, if f is both injective and surjective. If we know that f is a bijection then we can write the mapping as

$$f : X \xrightarrow{\sim} Y$$

which is read as f is a bijective mapping from X to Y .

We will look for additional examples of each type of mapping.

Example 1.2.25. Let $f : \mathbb{N} \rightarrow \mathbb{N}$ where $f(x) = x$. We will prove that f is a bijective mapping.

Proof:

To show f is bijective we show that f is injective and surjective. To see that f is an injection, suppose that $f(x) = f(y)$ where $x, y \in \mathbb{N}$, the domain. then we have that

$$\begin{aligned} f(x) &= f(y) \\ x &= y \end{aligned}$$

This shows f is injective as this holds for any choice of $x, y \in \mathbb{N}$. To see that f is surjective consider $y \in \mathbb{N}$, the co-domain, we show there exists an $x \in \mathbb{N}$, the domain, so that $f(x) = y$. We have

$$\begin{aligned} y &= f(x) \\ y &= x \end{aligned}$$

so we take $x = y$. This works for every $y \in \mathbb{N}$, the co-domain, so f is surjective.

As f is both injective and surjective it is by definition a bijective map, that is $f : \mathbb{N} \xrightarrow{\sim} \mathbb{N}$. \square

Example 1.2.26. Let $f : \mathbb{N} \rightarrow \mathbb{N}$ where

$$f(x) = \begin{cases} x, & \text{If } x \text{ is odd} \\ \frac{x}{2}, & \text{If } x \text{ is even} \end{cases}$$

Is f injective? To see if it is we would need to show that $f(x) = f(y)$ with $x, y \in \mathbb{N}$ means that $x = y$. It becomes clear that there are $x, y \in \mathbb{N}$ where this does not hold, for example $f(1) = 1$ and $f(2) = 1$ so $f(1) = f(2)$ but $1 \neq 2$. This shows that f is not injective. Is f surjective? To see if it is we would need to show that $\forall y \in \mathbb{N}, \exists x \in \mathbb{N}$ such that $y = f(x)$. Note that for every even input $x = 2k$ we have that $f(x) = \frac{2k}{2} = k$. So for any $y \in \mathbb{N}$ if we take $x = 2y$ then every $y \in \mathbb{N}$ gets mapped to to by $2y$. So f is surjective.

As f was not injective we have that f is not a bijection, so we have $f : \mathbb{N} \twoheadrightarrow \mathbb{N}$.

Example 1.2.27. Let $X = \{1, 2\}$ and $Y = \{3, 4, 5\}$ and define the map $f : X \rightarrow Y$ by

$$f(1) = 3, f(2) = 4$$

Then it is clear that f is injective, as each input is mapped to a distinct output. More formally suppose that $f(x) = f(y)$ where $x, y \in X$. We have that by the definition of the mapping $f(1) = 3, f(2) = 4$. In the first case we have $f(x) = f(y) = 3$ and so $x = y = 1$, likewise in the second case we have that $f(x) = f(y) = 4$ and so $x = y = 2$. This proves injectivity.

To see that f is not surjective, consider the image $\text{Image}(f) = \{f(x) : x \in X\} = \{3, 4\} \neq Y$. So $\exists y \in Y$ such that $\nexists x \in X$ with $y = f(x)$.

It hence follows that f is not bijective, that is $f : \{1, 2\} \hookrightarrow \{3, 4, 5\}$.

Example 1.2.28. Let $X = \{1, 2, 3\}$ and $Y = \{4, 5\}$ and define the map $f : X \rightarrow Y$ by

$$f(1) = 4, f(2) = 4, f(3) = 5$$

We have that f is not injective as $f(1) = f(2) = 4$ but $1 \neq 2$. However we have that f is surjective as the image of f is $\text{Image}(f) = \{f(x) : x \in X\} = \{4, 5\} = Y$.

By definition f is not bijective, hence $f : \{1, 2, 3\} \rightarrow \{4, 5\}$.

We note that we can always construct a mapping g from $f : X \rightarrow Y$ such that $g : X \rightarrow \text{Image}(f)$ is a surjection.

Proposition 1.2.15. The restriction of a mappings co-domain to its image is a surjective mapping

Let $f : X \rightarrow Y$ be a mapping and consider $\text{Image}(f) = \{f(x) : x \in X\}$. Consider the following mapping

$$\begin{aligned} g : X &\rightarrow \text{Image}(f) \\ x &\mapsto f(x) \end{aligned}$$

Then g is a surjective map.

Proof:

Let $f : X \rightarrow Y$ and consider $\text{Image}(f) = \{f(x) : x \in X\}$. By the definition of the image of a mapping 1.2.30 we have that $\text{Image}(f) \subseteq Y$. Moreover, by the definition of the image of a map we have that $y \in \text{Image}(f)$ if and only if $\exists x \in X$ such that $y = f(x)$. This will hold for all $y \in \text{Image}(f)$ so g is a surjection. \square .

In the proof we used the idea of restricting the co-domain of the function so that it was the image $\text{Image}(f)$ rather than Y , while leaving the domain X unchanged. In actuality we didn't restrict the co-domain at all but instead only considered those elements of the co-domain that actually get mapped to. It should be clear that the image $\text{Image}(f)$, the elements that actually get mapped to, only depends on the allowable inputs for the function, that is only depend on the domain X . In many fields of mathematics it is sometimes desirable to restrict the domain X that is being worked with to a smaller subset of the domain $A \subseteq X$. As a quick example of why this is useful, and which we will see later, is for inverse mappings. For now the key idea of an inverse map is to be able to create a bijection between a mapping and its domain and co-domain to enable us to unambiguously go between the two. Why is this useful?

For an example, suppose that you wanted to go on holiday abroad then you'll need to convert your currency to the currency that is in use where you go to. Suppose that you use gold coins where as the contry you vist only uses silver coins. The exchange rate from gold coins to silver coins is given by the following mapping $E(x) = Ax^2$ where the domain is the set of all the numbers that we are familiar with, that is \mathbb{R} , and A is some positive number which is greater than 0.

Suppose we wish to convert 50 gold coins into the new currency, then we will have $E(50) = A * 50^2 = 2500A$ silver coins. Finally suppose that after our holiday we have some silver coins left over that we wish to convert back to gold coins, say $2500A - y$ where $0 < y < 2500A$, how many gold coins will we get back?

To work this out we will need to find a way to go backwards from $\text{Image}(E)$ back to the domain. To do this we will solve $g = Ax^2$ for x , we have that

$$g = Ax^2$$

$$\frac{g}{A} = x^2$$

$$x = \pm \sqrt{\frac{g}{A}}$$

You may wonder where \pm came from and what it means. \pm stands for plus or minus and is used when we are unsure whether the number is positive or negative. It occurs here because for the numbers we are familiar with there are two possible answers when taking the square root of a number, for example if we wanted to find the square root of 2 we have that $\sqrt{2} * \sqrt{2} = 2$ or $(-\sqrt{2})(-\sqrt{2}) = 2$.

So, going back to the currency problem. When we wish to convert our remaining silver coins back into gold coins we will get back

$$x = \pm \sqrt{\frac{2500A - y}{A}}$$

This is a problem, because the domain of E was any of the usual numbers we don't know whether we should get back the positive or negative value, as both will have given us the silver coins we had remaining; perhaps on a more relatable note, we would find it very annoying if we got back from holiday and converted our positive money back only to end up with a negative amount of money. To overcome this problem we should ensure that the domain of E consists of only positive numbers, rather than any value, by doing so the negative square root value is no longer valid and we hence get back the correct amount of money.

Although a simple example, this shows the importance sometimes having to restrict the domain of a mappings. We define the idea of a restriction of the domain now.

Definition 1.2.34. *Restriction of a mapping*

Let $f : X \rightarrow Y$ be a mapping between two sets X and Y . Let $A \subseteq X$ be any subset of X . We define the restriction of f to A , denoted by $f|_A$, by the mapping

$$f|_A : A \rightarrow Y$$

$$x \mapsto f|_A(x)$$

In particular, restricting a mapping will cause the image to change so that $\text{Image}(f|_A) \subseteq \text{Image}(f)$

Now that we have the idea of restricting a mapping we can see the following

Proposition 1.2.16. *Restriction of an injective mapping is injective*

Let $f : X \rightarrow Y$ be a mapping between two sets X and Y such that f is an injective mapping. Let $A \subseteq X$ be any subset of X . We have that the restriction $f|_A : A \rightarrow Y$ is an injective map. In particular we have that $f|_A : A \rightarrow \text{Image}(f|_A)$ is an injection.

Proof:

To show that $f|_A : A \rightarrow Y$ is an injective we show that $f|_A(x) = f|_A(y)$ for $x, y \in A$ means that $x = y$. Suppose that $f|_A$ is not an injective map, then we have that $\exists x, y \in A$ with $x \neq y$ such that $f|_A(x) = f|_A(y)$. However $A \subseteq X$ and so $x, y \in X$ but f is an injective map so $f(x) = f(y)$ with $x \neq y$, contradicting the fact that f is an injection.

We conclude that the restriction map $f|_A$ must be injective.

Finally, by definition of $\text{Image}(f|_A)$ we have that

$$\text{Image}(f|_A) = \{f|_A(x) : x \in A\} \subseteq Y$$

that is the image is all the elements $f|_A$ will map elements of A to, as $f|_A : X \rightarrow Y$ is an injection we must conclude that $f|_A : A \rightarrow \text{Image}(f|_A)$ is an injection, for if not then the original restriction map could not have been an injection. \square

Example 1.2.29. Let $f : X \rightarrow Y$ be a mapping where $x = \{1, 2, 3, 4, 5, 6\}$ and $Y = \{7, 8, 9, 10, 11, 12\}$ where

$$x \mapsto f(x) = x + 6$$

Consider $A \subseteq X$ where $A = \{1, 2, 3\}$ and $B \subseteq X$ with $B = \{1, 2, 3, 4, 5\}$, then $A \subseteq B$. We have that $f|_A : A \rightarrow Y$ has the image $\text{Image}(f|_A) = \{7, 8, 9\}$ and we have that $f|_B : B \rightarrow Y$ has the image $\text{Image}(f|_B) = \{7, 8, 9, 10, 11\}$

Hence under the two different restrictions we observe that $\text{Image}(f|_A) \subseteq \text{Image}(f|_B)$.

From this example we have the following result.

Proposition 1.2.17. The image of a subset is a subset of the image

Let $f : X \rightarrow Y$ be a mapping of sets and let $A, B \subseteq X$ where $A \subseteq B$, we have that

$$\text{Image}(f|_A) \subseteq \text{Image}(f|_B) \subseteq \text{Image}(f)$$

Proof:

Let $f : X \rightarrow Y$ be a mapping of sets and let $A, B \subseteq X$ where $A \subseteq B$. Consider the restriction mappings $f|_A : A \rightarrow Y$ and $f|_B : B \rightarrow Y$. Let $y \in \text{Image}(f|_A)$, then by definition we have that $\exists x \in A$ such that $f|_A(x) = y$. As $A \subseteq B$ we have that $x \in A \Rightarrow x \in B$ and so $f|_B(x) = y$, hence $y \in \text{Image}(f|_B)$. This shows that $\text{Image}(f|_A) \subseteq \text{Image}(f|_B)$. To see the final inclusion note that $A \subseteq B \subseteq X$ so $x \in A \Rightarrow x \in B \Rightarrow x \in X$ and so $f(x) = y$, hence $y \in \text{Image}(f)$.

This shows the result. \square

We conclude with the following observation.

Proposition 1.2.18. Injective mapping to image is a bijection

Let $f : X \rightarrow Y$ be an injective map between two sets X and Y . Let $A \subseteq X$ be any subset of X possibly being X itself. We have that the mapping $g : A \rightarrow \text{Image}(f|_A)$ is a bijection.

Proof:

Let $f : X \rightarrow Y$ be an injective mapping and let $A \subseteq X$. By proposition 1.2.16 we have that the mapping $f|_A : A \rightarrow \text{Image}(f|_A)$ is an injection. Also, by proposition 1.2.15 that $f|_A : A \rightarrow \text{Image}(f|_A)$ is a surjection. Hence by definition we have that $f|_A : A \rightarrow \text{Image}(f|_A)$ is a bijection. \square

1.2.2.4 Compositions of maps

We have seen how a mapping f takes elements in one set, the domain X , and sends them to the elements of another set, the image $\text{Image}(f) \subseteq Y$ of some co-domain Y . We can extend this idea so that the image $\text{Image}(f)$ and more generally the co-domain Y act as the domain for some other mapping g . This will allow us to consider some more interesting examples of mappings in general.

Definition 1.2.35. Composition of two mappings

Let $f : X \rightarrow Y$ and $g : Y \rightarrow Z$ be two mappings for some sets X, Y and Z . We define the composition map by

$$g \circ f : X \rightarrow Z$$

$$x \mapsto g(f(x))$$

That is, the mapping f is done first and then we apply the mapping g .

Additionally, let $h : X \rightarrow X$ be a mapping from X to X then if h is composed with itself we write $h \circ h = h(h(x)) = h^2(x)$. If h is composed with itself n times we write $h^{n+1}(x)$. This is sometimes called the $n + 1$ -fold composition of h with itself.

Example 1.2.30. Let $f : \mathbb{N} \rightarrow \mathbb{N}$ and $g : \mathbb{N} \rightarrow \mathbb{N}$ be maps such that

$$\begin{aligned}
f : \mathbb{N} &\rightarrow \mathbb{N} \\
x &\mapsto f(x) = x^2 \\
g : \mathbb{N} &\rightarrow \mathbb{N} \\
x &\mapsto g(x) = x^3
\end{aligned}$$

Then we have that, for some arbitrary $x \in \mathbb{N}$ that

$$\begin{aligned}
g \circ f(x) &= g(f(x)) = g(x^2) = (x^2)^3 = x^6 \\
f \circ g(x) &= f(g(x)) = f(x^3) = (x^3)^2 = x^6
\end{aligned}$$

In this case $g \circ f = f \circ g$, and it does not matter in which way we compose the two mappings.

The ideas of injectivity and surjectivity also apply to compositions of maps. We will see if $g \circ f$ is injective.

Recall that a mapping $h : X \rightarrow Y$ is injective if $h(x) = h(y)$ for $x, y \in X$ means that $x = y$. So let $x, y \in \mathbb{N}$ and consider $g \circ f(x) = g \circ f(y)$. Then we have that

$$\begin{aligned}
g \circ f(x) &= g \circ f(y) \\
x^6 &= y^6 \\
x &= y
\end{aligned}$$

This makes sense as $x^6, y^6 \in \mathbb{N}$ as $x^6 = x * x * x * x * x * x$ which is multiplication in \mathbb{N} , also We can take the sixth-root of x^6 without issue. Likewise for y . It is clear that the composition is not surjective for example $2 \in \mathbb{N}$ does not have an element $x \in \mathbb{N}$ such that $x^6 = 2$. If we were to include any possible positive number we would have $x = \sqrt[6]{2} \approx 1.1224620483094$.

Hence $g \circ f$ is not bijective as it is not surjective. Likewise for $g \circ f$.

Example 1.2.31. Consider the mappings $f : \mathbb{N} \rightarrow \mathbb{N}$ and $g : \mathbb{N} \rightarrow \mathbb{N}$ given by

$$\begin{aligned}
f : \mathbb{N} &\rightarrow \mathbb{N} \\
x &\mapsto f(x) = 4x + 2 \\
g : \mathbb{N} &\rightarrow \mathbb{N} \\
x &\mapsto g(x) = \sqrt{x}
\end{aligned}$$

We have that

$$\begin{aligned}
g \circ f(x) &= g(f(x)) = g(4x + 2) = \sqrt{4x + 2} \\
f \circ g(x) &= f(g(x)) = f(\sqrt{x}) = 4\sqrt{x} + 2
\end{aligned}$$

Unlike last time we have that $g \circ f \neq f \circ g$. Now is $g \circ f$ injective? Let $x, y \in \mathbb{N}$ and consider

$$\begin{aligned}
g \circ f(x) &= g \circ f(y) \\
\sqrt{4x + 2} &= \sqrt{4y + 2} \iff 4x + 2 = 4y + 2 \\
4x + 2 &= 4y + 2 \\
x &= y
\end{aligned}$$

So we have injectivity. We do not have surjectivity as, for example with $y = 1 \in \mathbb{N}$ then

$$\begin{aligned} 1 &= \sqrt{4x+2} \\ 1 &= 4x+2 \\ -1 &= 4x \\ x &= -\frac{1}{4} \notin \mathbb{N} \end{aligned}$$

What about $f \circ g$? For injectivity let $x, y \in \mathbb{N}$ then

$$\begin{aligned} f \circ g(x) &= f \circ g(y) \\ 4\sqrt{x} + 2 &= 4\sqrt{y} + 2 \\ \sqrt{x} &= \sqrt{y} \iff x = y \end{aligned}$$

hence we have injectivity. We do not have surjectivity, for example with $y = 1 \in \mathbb{N}$ we have that

$$\begin{aligned} 1 &= 4\sqrt{x} + 2 \\ -1 &= 4\sqrt{x} \\ -\frac{1}{4} &= \sqrt{x} \Rightarrow x \notin \mathbb{N} \end{aligned}$$

Example 1.2.32. Consider $X = \{1, 2, 3\}$, $Y = \{4, 5\}$ and $Z = \{6\}$ and the mappings $f : X \rightarrow Y$ and $g : Y \rightarrow Z$ given by

$$f(1) = 4, f(2) = 4, f(3) = 5$$

$$g(4) = 6, g(5) = 6$$

Finally, consider the composition map given by

$$g \circ f : X \rightarrow Z \tag{1}$$

$$x \mapsto g(f(x)) \tag{2}$$

Clearly $g \circ f$ is not injective as $g(f(1)) = 6$ and $g(f(2)) = 6$ but $1 \neq 2$. However the compositing map is surjective as $\text{Image}(g \circ f) = \{6\} = Z$.

We deduce an immediate result.

Proposition 1.2.19. Domain of composition mapping equals the domain of the first function

Let $f : X \rightarrow Y$ and $g : Y \rightarrow Z$ be mappings. Consider the composite mapping $g \circ f : X \rightarrow Z$. We have that

$$\text{Dom}(g \circ f) = \text{Dom}(f)$$

Proof:

Let $f : X \rightarrow Y$ and $g : Y \rightarrow Z$ be mappings and consider the composite mapping $g \circ f : X \rightarrow Z$. We need to show that $\text{Dom}(g \circ f) = \text{Dom}(f)$.

Let $x \in \text{Dom}(g \circ f)$, then $g(f(x))$ is well-defined with say $z = g(f(x))$ for some $z \in Z$. Hence for this to be well-defined we have that $\exists y \in Y$ such that $y = f(x)$ is well-defined. But then $x \in \text{Dom}(f)$, hence $\text{Dom}(g \circ f) \subseteq \text{Dom}(f)$.

For the inverse inclusion, let $x \in \text{Dom}(f)$ then $f(x) = y$ for some $y \in Y$. As $g : Y \rightarrow Z$ is a mapping with domain Y , then $\exists z \in Z$ such that $g(y) = z$. Hence we have that $g(y) = g(f(x)) = z$. Hence $\text{Dom}(f) \subseteq \text{Dom}(g \circ f)$.

As we have that $\text{Dom}(g \circ f) \subseteq \text{Dom}(f)$ and $\text{Dom}(f) \subseteq \text{Dom}(g \circ f)$, then we conclude by proposition 1.2.1 that $\text{Dom}(g \circ f) = \text{Dom}(f)$ as required. \square

These examples show something interesting. In the first example we note that f and g are both injective. Indeed we have for $x, y \in \mathbb{N}$ that

$$\begin{aligned}x^2 = y^2 &\Rightarrow x = y \\x^3 = y^3 &\Rightarrow x = y\end{aligned}$$

and the composition mappings $g \circ f$ and $f \circ g$ where both injective, in the last example we had that both f and g where surjective as

$$\begin{aligned}\text{Image}(f) &= \{f(x) : x \in \{1, 2, 3\}\} = \{4, 5\} = Y \\ \text{Image}(g) &= \{g(x) : x \in \{4, 5\}\} = \{6\} = Z\end{aligned}$$

and the composition map $g \circ f$ was also surjective. This is not a coincidence which we prove now

Proposition 1.2.20. *Injectivity, surjectivity and bijectivity of composition mappings*

Let $f : X \rightarrow Y$ and $g : Y \rightarrow Z$ be mappings.

1. *If f and g are injective maps then so is $g \circ f$*
2. *If f and g are surjective maps then so is $g \circ f$*
3. *If f and g are bijective maps then so is $g \circ f$*

Proof:

1. *If f and g are injective maps then so is $g \circ f$:*

Let $f : X \hookrightarrow Y$ and $g : Y \hookrightarrow Z$ be injective mappings, then by definition we have that

$$\begin{aligned}\forall a, b \in X, \quad f(a) = f(b) &\Rightarrow a = b \\ \forall c, d \in Y, \quad g(c) = g(d) &\Rightarrow c = d\end{aligned}$$

Consider the composition map

$$\begin{aligned}g \circ f : X &\rightarrow Z \\ x &\mapsto g(f(x))\end{aligned}$$

Let $x, y \in X$ then we have that

$$\begin{aligned}g(f(x)) &= g(f(y)) \\ f(x) &= f(y), \text{ As } g \text{ is an injective map} \\ x &= y, \text{ As } f \text{ is an injective map}\end{aligned}$$

As this works for every $x, y \in X$ we have that $g \circ f$ is injective.

2. *If f and g are surjective maps then so is $g \circ f$:*

Let $f : X \rightarrow Y$ and $g : Y \rightarrow Z$ be surjective mappings, then by definition we have that

$$\begin{aligned}\forall b \in Y, \exists a \in X : f(a) &= b \\ \forall d \in Z, \exists c \in Y : g(c) &= d\end{aligned}$$

Consider the composition map

$$g \circ f : X \rightarrow Z$$

$$x \mapsto g(f(x))$$

Let $z \in Z$, then $\exists y \in Y$ such that $g(y) = z$, also $\exists x \in X$ such that $f(x) = y$ as both f and g are surjective, then we have that

$$g(f(x)) = g(y) = z$$

As this works for every $z \in Z$ we have that $g \circ f$ is surjective.

3. If f and g are bijective maps then s is $g \circ f$:

Let $f : X \hookrightarrow Y$ and $g : Y \hookrightarrow Z$ be bijective mappings, then by definition we have that f is an injection and a surjection so f satisfies

$$\forall a, b \in X, f(a) = f(b) \Rightarrow a = b$$

$$\forall d \in Y, \exists c \in X : f(c) = d$$

Also g is an injection and surjection and so satisfies

$$\forall a, b \in Y, g(a) = g(b) \Rightarrow a = b$$

$$\forall d \in Z, \exists c \in Y : g(c) = d$$

Consider the composition map

$$g \circ f : X \rightarrow Z$$

$$x \mapsto g(f(x))$$

By part 1. we have that $g \circ f$ is an injection as f and g are injections, by part 2. we have that $g \circ f$ is a surjection as f and g are surjections. Hence by definition as $g \circ f$ is both injective and surjective it is bijective.

□

In a sense we can also deduce properties about the mappings f and g if we know something about the composition map $g \circ f$

Proposition 1.2.21. *Properties of mappings from composite map*

Let $f : X \rightarrow Y$ and $g : Y \rightarrow Z$ be mappings and consider the composite map $g \circ f : X \rightarrow Z$.

1. If $g \circ f : X \hookrightarrow Z$ is an injective map, then $f : X \rightarrow Y$ is an injective map.
2. If $g \circ f : X \twoheadrightarrow Z$ is a surjective map, then $g : Y \rightarrow Z$ is a surjective map.

Proof:

1. If $g \circ f : X \hookrightarrow Z$ is an injective map, then $f : X \rightarrow Y$ is an injective map:

Let $g \circ f : X \hookrightarrow Z$ is an injective composite mapping, then $g(f(x)) = g(f(y))$ for all $x, y \in X$, we need to show that $\forall x, y \in X$ that $f(x) = f(y) \Rightarrow x = y$.

So indeed, suppose that for some $x, y \in X$ that $f(x) = f(y)$, then we have that

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

Now, as $g \circ f$ is an injective map we conclude that $x = y$, hence $f(x) = f(y) \Rightarrow x = y$. Hence $f : X \rightarrow Y$ is an injection.

2. If $g \circ f : X \twoheadrightarrow Z$ is a surjective map, then $g : Y \rightarrow Z$ is a surjective map:

Let $g \circ f : X \twoheadrightarrow Z$ is a surjective composite mapping, then $\forall z \in Z, \exists x \in X : z = g \circ f(x)$, we need to show that $\forall z \in Z, \exists y \in Y : z = g(y)$. Let $z \in Z$ then as $g \circ f$ is surjective there is some $x \in X$ such that $z = g \circ f(x)$.

Now, we have by proposition 1.2.19 that $\text{Dom}(g \circ f) = \text{Dom}(f)$ and so $x \in \text{Dom}(f)$, so that $f(x) \in \text{Image}(f)$. This is to say $\exists y \in Y : y = f(x)$ and hence $z = g(y)$. As this can be done for any $z \in Z$ we conclude that $g : Y \twoheadrightarrow Z$ is a surjection.

□

The examples also allow us to deduce something about the image of composition mappings

Proposition 1.2.22. *The image of a composite mapping*

Let $f : X \rightarrow Y$ and $g : Y \rightarrow Z$ be mappings for sets X, Y and Z . Consider the composition mapping given by

$$\begin{aligned} g \circ f : X &\rightarrow Z \\ x &\mapsto g(f(x)) \end{aligned}$$

We have that $\text{Image}(g \circ f) = g(\text{Image}(f))$ where we recall the notation $f(X) = \{f(x) : x \in X\}$

Proof:

We have that

$$\text{Image}(g \circ f) = \{g(f(x)) : x \in X\}$$

also, we have that

$$g(\text{Image}(f)) = \{g(y) : y \in \text{Image}(f)\}$$

Now, $y \in \text{Image}(f)$ means that $y \in \{f(x) : x \in X\}$, hence $y = f(x)$ for some $x \in X$, hence

$$g(\text{Image}(f)) = \{g(f(x)) : x \in X\}$$

Hence the two definitions agree, that is $\text{Image}(g \circ f) = g(\text{Image}(f))$.

We do need to check the case of $Y = \emptyset$. If $Y = \emptyset$ then we note that $\text{Image}(g) = \emptyset$ by the remark after the definition of the image of a function. So $g : \emptyset \rightarrow \emptyset$, i.e g takes has no elements in its domain and no elements in its co-domain and so is a mapping that maps nothing to nothing. Also $f : \emptyset \rightarrow \emptyset$, we prove this

Lemma 1.2.3. Mapping from empty set to some co-domain set is valid if and only if co-domain is empty
Let Y be some set, then $f : \emptyset \rightarrow Y$ is a mapping if and only if $Y = \emptyset$

Proof:

(\Rightarrow) : Suppose that $Y \neq \emptyset$ then $\exists s \in Y$, that is there is at least one element in Y , but the domain is empty so there are no elements that could be mapped to s , hence f is not a well-defined mapping, so we conclude that $Y = \emptyset$.

(\Leftarrow) : Suppose that $Y = \emptyset$, then $f : \emptyset \rightarrow \emptyset$ holds as a mapping, mapping nothing to nothing. \square

So the lemma shows $f : \emptyset \rightarrow \emptyset$. Hence

$$\text{Image}(g \circ f) = \emptyset = g(\emptyset) = g(\text{Image}(f))$$

As required. \square

From the proposition we also deduce the following

Proposition 1.2.23. Image of composite mapping is a subset of the image of the second function
Let $f : X \rightarrow Y$ and $g : Y \rightarrow Z$ be mappings. Consider the composite mapping $g \circ f : X \rightarrow Z$. We have that

$$\text{Image}(g \circ f) \subseteq \text{Image}(g)$$

Proof:

We know by proposition 1.2.22 that $\text{Image}(g \circ f) = g(\text{Image}(f))$ where

$$g(\text{Image}(f)) = \{g(y) : y \in \text{Image}(f)\}$$

Now, observe that $\text{Image}(f) \subseteq Y$, and in particular $\text{Image}(f) \subseteq \text{Dom}(g)$. Hence, with proposition 1.2.17, we deduce that

$$g(\text{Image}(f)) = \{g(y) : y \in \text{Image}(f)\} \subseteq g(\text{Dom}(g)) = \{g(y) : y \in \text{Dom}(g)\} = \text{Image}(g)$$

Hence we have $\text{Image}(g \circ f) = \{g(y) : y \in \text{Image}(f)\} \subseteq \text{Image}(g)$, which is to say $\text{Image}(g \circ f) \subseteq \text{Image}(g)$ as required. \square

We have seen earlier that function composition need not be commutative, for example when $f : \mathbb{N} \rightarrow \mathbb{N}$ with $f(x) = 4x + 2$ and $g : \mathbb{N} \rightarrow \mathbb{N}$ with $g(x) = \sqrt{x}$. We saw that

$$\begin{aligned} g \circ f(x) &= \sqrt{4x + 2} \\ f \circ g(x) &= 4\sqrt{x} + 2 \end{aligned}$$

What can we say about associativity and function composition?

Example 1.2.33. Let $f : \mathbb{N} \rightarrow \mathbb{N}$, $g : \mathbb{N} \rightarrow \mathbb{N}$ and $h : \mathbb{N} \rightarrow \mathbb{N}$ where

$$\begin{aligned} f : \mathbb{N} &\rightarrow \mathbb{N} \\ x &\mapsto f(x) = 4x + 2 \\ g : \mathbb{N} &\rightarrow \mathbb{N} \\ x &\mapsto g(x) = x^2 \\ h : \mathbb{N} &\rightarrow \mathbb{N} \\ x &\mapsto h(x) = \sqrt{x} \end{aligned}$$

Consider the following

$$h \circ (g \circ f)(x) = h(g(f(x))) = h((4x+2)^2) = \sqrt{(4x+2)^2} = 4x+2$$

$$(h \circ g) \circ f(x) = h(g(x)) \circ f(x) = \sqrt{x^2} \circ (4x+2) = x \circ 4x+2 = 4x+2$$

In this case the function composition is associative.

This is not a coincidence

Proposition 1.2.24. *Function composition is associative*

Let $f : W \rightarrow X$, $g : X \rightarrow Y$ and $h : Y \rightarrow Z$ be mappings. We have that

$$h \circ (g \circ f) = (h \circ g) \circ f$$

Proof:

Let $f : W \rightarrow X$, $g : X \rightarrow Y$ and $h : Y \rightarrow Z$ and consider the composite mappings $h \circ (g \circ f) : W \rightarrow Z$ and $(h \circ g) \circ f : W \rightarrow Z$.

Let $w \in W$, then we have that as $f : W \rightarrow X$ is a mapping then $\exists x \in X$ such that $x = f(w)$, likewise as $g : X \rightarrow Y$, then $\exists y \in Y$ such that $y = g(x) = g(f(w))$. Finally as $h : Y \rightarrow Z$ is a mapping then $\exists z \in Z$ such that $z = h(y) = h(g(f(w)))$.

Likewise, for the same $w \in W$ we have $x = f(w)$, now as $(h \circ g) \circ f(w) = (h \circ g)(f(w))$ then we need to see where $h \circ g$ maps $f(w)$. As $h \circ g(x) = h(g(x))$ then we have that $(h \circ g)(f(w)) = h(g(f(w)))$, We know that $g(f(w)) = y$ and $h(g(f(w))) = z$.

Hence $h \circ (g \circ f) = (h \circ g) \circ f$ as w as an arbitrary element of W . \square

1.2.2.5 Inverse mappings

With the theory of composite mappings now understood, we are in a position to try and understand how to undo a given map $f : X \rightarrow Y$. Why did we need to develop a theory of composite mappings? The idea comes from the fact that undoing a mapping should somehow be the same as never doing anything in the first place. This is to say, if we denote the inverse map by f^{-1} then we should expect that $f^{-1} \circ f(x) = x$, likewise the original mapping f somehow undoes f^{-1} i.e $f \circ f^{-1}(y) = y$ where y is in the co-domain of f . As always in mathematics, examples will help to understand whats going on.

You may recall from a course in physics that an object thrown in a vacuum so that there is no air resistance, where only gravity acts has the following equation for its height

$$H(t) = V_0 \sin(\theta)t - \frac{1}{2}gt^2$$

where V_0 is the objects launch velocity in metres per second m/s , θ is the angle that the projectile is launched at from the horizontal, g is gravity in metres per second² m/s^2 and t is time in seconds s . Suppose the particle is launched with a velocity of $45m/s$ at an angle of 45 degrees to the horizontal and we take $g = 9.8m/s^2$, then for example, the height above the origin of the projectile at $t = 1s$ is

$$H(1) = 10 * \sin(45) * \frac{1}{2}9.8 = 5\sqrt{2} - \frac{49}{10} \approx 2.17m$$

Now suppose you are told that the maximum height is achieved at a time of $t = \frac{25\sqrt{2}}{49} \approx 0.721s$ which is $h = \frac{125}{49} \approx 2.551m$. Considering time values $0 < t < \frac{25\sqrt{2}}{49} \approx 0.721$, find the time that the projectile was first at $2m$ above the ground. In essence we need to take $h(t)$ and somehow undo the process to find some t such that $h(t) = 2$. How do we do this? Well set $h(t) = h$ then solve for t as follows

$$h = 10 \sin(45) t - \frac{1}{2} (9.8) t^2$$

$$h = 5\sqrt{2}t - \frac{49}{10}t^2$$

$$\frac{49}{10}t^2 - 5\sqrt{2}t + h = 0$$

$$49t^2 - 50\sqrt{2}t + 10h = 0$$

Now, from school we have learnt the quadratic formula, applying this here we will get two answers for t

$$t = \frac{-(-50\sqrt{2}) \pm \sqrt{(-50\sqrt{2})^2 - 4(49)(10h)}}{2(49)}$$

$$t = \frac{50\sqrt{2} \pm \sqrt{5000 - 1960h}}{98}$$

Hence when $h = 2$ we get the following times

$$t = \frac{50\sqrt{2} \pm \sqrt{5000 - 1960(2)}}{98}$$

$$t = \frac{50\sqrt{2} \pm \sqrt{5000 - 3920}}{98}$$

$$t = \frac{50\sqrt{2} \pm \sqrt{1080}}{98}$$

$$t = \frac{50\sqrt{2} \pm 6\sqrt{30}}{98}$$

$$t = \frac{25\sqrt{2} \pm 3\sqrt{30}}{49}$$

That is, $t = \frac{25\sqrt{2} + 3\sqrt{30}}{49} \approx 1.057s$ or $t = \frac{25\sqrt{2} - 3\sqrt{30}}{49} \approx 0.386s$. This example illustrates a key point about inverse maps, when we undo a given map we should get back the original input. Thankfully in this case we were told when the ball reaches its maximum height and the time it does so which was about $0.721s$ hence we have that the value we are looking for is the smaller $t = \frac{25\sqrt{2} - 3\sqrt{30}}{49} \approx 0.386s$. In fact if we want to find the time the projectile was first at hm above the ground we will always take the smaller of the two values for t found. That is, defining a new map T given by

$$t(h) = \frac{50\sqrt{2} - \sqrt{5000 - 1960h}}{98}$$

So that when the particle is launched with a velocity of $45m/s$ at an angle of 45 degrees to the horizontal with $g = 9.8m/s^2$ and using our knowledge of the fact that the maximum height is achieved at a time of $t = \frac{25\sqrt{2}}{49} \approx 0.721s$ which is $h = \frac{125}{49} \approx 2.551m$, then the mapping

$$T(h) = \frac{50\sqrt{2} - \sqrt{5000 - 1960h}}{98}$$

is the inverse to

$$H(t) = 10 \sin(45)t - \frac{1}{2}(9.8)t^2$$

Indeed, for example we have that

$$\begin{aligned}
H \circ T(t) &= H(T(h)) \\
&= H\left(\frac{50\sqrt{2} - \sqrt{5000 - 1960h}}{98}\right) \\
&= 10 \sin(45) \left(\frac{50\sqrt{2} - \sqrt{5000 - 1960h}}{98}\right) - \frac{1}{2}(9.8) \left(\frac{50\sqrt{2} - \sqrt{5000 - 1960h}}{98}\right)^2 \\
&= 5\sqrt{2} \left(\frac{50\sqrt{2} - \sqrt{5000 - 1960h}}{98}\right) - \frac{1}{2}(9.8) \frac{(50\sqrt{2} - \sqrt{5000 - 1960h})^2}{98^2} \\
&= \frac{500}{98} - \frac{5\sqrt{2}\sqrt{5000 - 1960h}}{98} - \frac{1}{2}(9.8) \frac{(5000 - 100\sqrt{10000 - 3920h} + (5000 - 1960h))}{98^2} \\
&= \frac{250}{49} - \frac{5\sqrt{10000 - 3920h}}{98} - \frac{1}{2} \frac{(5000 - 100\sqrt{10000 - 3920h} + (5000 - 1960h))}{980} \\
&= \frac{250}{49} - \frac{5\sqrt{10000 - 3920h}}{98} - \frac{1}{2} \frac{(10000 - 100\sqrt{10000 - 3920h} - 1960h)}{980} \\
&= \frac{250}{49} - \frac{5\sqrt{10000 - 3920h}}{98} - \frac{(5000 - 50\sqrt{10000 - 3920h} - 980h)}{980} \\
&= \frac{250}{49} - \frac{5\sqrt{10000 - 3920h}}{98} - \frac{5000}{980} + \frac{50\sqrt{10000 - 3920h}}{980} + \frac{980h}{h} \\
&= \frac{250}{49} - \frac{5\sqrt{10000 - 3920h}}{98} - \frac{250}{49} + \frac{5\sqrt{10000 - 3920h}}{98} + h \\
&= h
\end{aligned}$$

Again, we have this idea that inverse functions should somehow return any mapping back to where it started.

We can start to express this idea in terms of a so-called identity mapping.

Definition 1.2.36. Let $\text{id}_X : X \rightarrow X$ be a mapping from X to itself, so that

$$\begin{aligned}
\text{id} : X &\rightarrow X \\
x &\mapsto \text{id}(x) = x
\end{aligned}$$

We say that id is the identity mapping on the set X . Suppose we also have a mapping $\text{id}_Y : Y \rightarrow Y$, then id_Y is the identity map on the set Y , so it is clear that $\text{id}_X \neq \text{id}_Y$.

Indeed we can prove that these identity maps do nothing under function composition.

Proposition 1.2.25. Composition with the identity mapping does nothing

Let $f : X \rightarrow Y$ be a mapping and consider the identity maps $\text{id}_X : X \rightarrow X$ and $\text{id}_Y : Y \rightarrow Y$. We have that

1. $f \circ \text{id}_X = f$
2. $\text{id}_Y \circ f = f$

Proof:

We simply need to compose the maps to see the desired results.

1. $f \circ \text{id}_X = f$:

Let $x \in X$ then $f \circ \text{id}_X (x) = f (\text{id}_X (x)) = f (x) = f$.

2. $\text{id}_Y \circ f = f$:

Let $x \in X$ then $\text{id}_Y \circ f (x) = \text{id}_Y (f (x)) = f (x) = f$

Hence the result follows.

For completeness we will prove some trivial properties about the identity mapping.

Proposition 1.2.26. *Properties of the identity mapping*

Let $\text{id}_X : X \rightarrow X$ be the identity map on X . Then the following hold

1. id_X is an injective map

2. id_X is a surjective map

3. id_X is a bijective map

4. $\text{id}_X \circ \text{id}_X = \text{id}_X$

Proof:

1. id_X is an injective map:

Let $x, y \in X$ then we have that $\text{id}_X (x) = \text{id}_X (y) \Rightarrow x = y$, hence id_X is injective.

2. id_X is a surjective map:

Let $y \in X$ be such that $y = \text{id}_X (x)$ for some $x \in X$, then $y = x$ as this works for every $y \in X$ then id_X is surjective.

3. id_X is a bijective map:

By parts 1. and 2. we have that id_X is injective and surjective and thus by definition is bijective.

4. $\text{id}_X \circ \text{id}_X = \text{id}_X$:

Let $x \in X$ and consider $\text{id}_X \circ \text{id}_X (x) = \text{id}_X (\text{id}_X (x)) = \text{id}_X (x) = x = \text{id}_X (x)$.

□.

The identity mapping will allow us to define the idea of a left and right inverse of a mapping.

Definition 1.2.37. *Left inverse*

Let $f : X \rightarrow Y$ be a mapping. We define $g : Y \rightarrow X$ to be the left inverse of f if

$$g \circ f = \text{id}_X$$

Definition 1.2.38. *Right inverse*

Let $f : X \rightarrow Y$ be a mapping. We define $g : Y \rightarrow X$ to be the right inverse of f if

$$f \circ g = \text{id}_Y$$

Example 1.2.34. Let $f : \mathbb{N} \rightarrow \mathbb{N}$ be such that $f (x) = x + 1$. Define the mapping $g : \mathbb{N} \rightarrow \mathbb{N}$ by

$$g (x) = \begin{cases} x - 1, & \text{If } x \neq 0, \\ 0, & \text{Otherwise} \end{cases}$$

Then g is a left inverse of f . Indeed we have that

$$g \circ f (x) = g (f (x)) = g (x + 1) = (x + 1) - 1 = x$$

as $x + 1 > 0$ for every $x \in \mathbb{N}$. Observe also that f is an injective map, indeed let $x, y \in \mathbb{N}$ and suppose $f(x) = f(y)$ then

$$\begin{aligned} f(x) &= f(y) \\ x + 1 &= y + 1 \\ x &= y \end{aligned}$$

It is also worth noting that g is not injective as we have $g(1) = 0 = g(0)$ but $1 \neq 0$. We note that f is the right inverse of g as the calculation above shows.

Example 1.2.35. Let $X = \mathbb{R}$ and define $Y = \mathbb{R}^+ = \{x \in \mathbb{R} : x \geq 0\}$, the set of familiar numbers. Let $f : \mathbb{R} \rightarrow \mathbb{R}^+$ be defined by $f(x) = x^2$. We can define two possible right inverses of f . The first is given by $g_1 : \mathbb{R}^+ \rightarrow \mathbb{R}$ where $g_1(x) = \sqrt{x}$. Indeed

$$f \circ g_1(x) = f(g_1(x)) = f(\sqrt{x}) = (\sqrt{x})^2 = x = \text{id}_{\mathbb{R}}(x)$$

The second, as you may have guessed, is given by $g_2 : \mathbb{R}^+ \rightarrow \mathbb{R}$ where $g_2(x) = -\sqrt{x}$ where likewise we have

$$f \circ g_2(x) = f(g_2(x)) = f(-\sqrt{x}) = (-\sqrt{x})^2 = x = \text{id}_{\mathbb{R}}(x)$$

We note that f is surjective. Let $y \in \mathbb{R}^+$ then $f(x) = y \Rightarrow x^2 = y \Rightarrow x = \pm\sqrt{y} \in \mathbb{R}$, hence every output of f is mapped to by some input. It is clear that f is not injective as $f(2) = 4 = f(-2)$.

Does f have a left inverse? By the definition of a left inverse we will need to find some $g : \mathbb{R}^+ \rightarrow \mathbb{R}$ such that $g \circ f = \text{id}_{\mathbb{R}}$. So for each input of f , g will have to send $f(x)$ back to x , hence we might require that f be injective, for if not then $\exists x, y \in \mathbb{R}$ such that $f(x) = f(y)$ with $x \neq y$ and we have the problem where g could send $f(x)$ back to either x or y , and if it is sent back to y then we don't have the identity mapping!

Now, f is not injective as we have seen that $f(2) = 4 = f(-2)$, so if there were a left inverse g it wouldn't know where to send 4 back to, it could have been either 2 or -2.

Example 1.2.36. Let $X = \mathbb{R}$ and $Y = \mathbb{R} \setminus \{0\} = \{x \in \mathbb{R} : x \neq 0\}$. You may have seen the function e^x before, we shall consider this mapping, that is the mapping $f : \mathbb{R} \rightarrow \mathbb{R} \setminus \{0\}$ given by $f(x) = e^x = \exp(x)$. We can define a left inverse to f by $g : \mathbb{R} \setminus \{0\} \rightarrow \mathbb{R}$ where $g(x) = \ln(x)$, where $\ln(x)$ is the natural logarithm, the logarithm to the base e . We will discuss logarithms in more detail later but for now we can think of $\ln(x) = y$ as asking the question $e^y = x$, that is value of y do we need to raise e to to get x . This g is indeed a left inverse of f as

$$g \circ f(x) = g(f(x)) = g(e^x) = \ln(e^x) = x = \text{id}_{\mathbb{R}}$$

Like in the previous example, we can ask the question does f have a right inverse? By definition for f to have a right inverse, there needs to be a mapping $g : \mathbb{R} \setminus \{0\} \rightarrow \mathbb{R}$ such that $f \circ g = \text{id}_{\mathbb{R} \setminus \{0\}}$. So for each $g(y)$ with $y \in \mathbb{R} \setminus \{0\}$ we have that f will send $g(y)$ back to y . This will happen if every output of f has some input that generates it, that is f is a surjection. If this not the case then there is some element $y \in \mathbb{R} \setminus \{0\}$ that is not mapped to by $f(x)$ for some $x \in \mathbb{R}$.

For example we have that $\nexists x \in \mathbb{R}$ such that $e^x = -1$ for example. So f is not surjective in this case we are not able to define a right inverse that makes sense.

We can generalise the last two examples to the next two propositions.

Proposition 1.2.27. Condition for the existence of a left inverse

Let $f : X \rightarrow Y$ be a mapping with $X \neq \emptyset$. We have that f has a left inverse $g : Y \rightarrow X$ such that $g \circ f = \text{id}_X$ if and only if f is an injective mapping.

Proof:

(\Rightarrow): Suppose that f has a left inverse $g : Y \rightarrow X$ such that $g \circ f = \text{id}_X$. We know by proposition 1.2.26 that id_X is an injective mapping, moreover we know by proposition 1.2.21 that if a composite map $g \circ f$ is injective then so is f . Hence as $g \circ f = \text{id}_X$ and id_X is injective, we conclude that f is an injective map.

(\Leftarrow): Suppose that f is an injective map, then $\forall x, y \in X$ we have that $f(x) = f(y) \Rightarrow x = y$. Let $x \in X$, we need to construct a map which acts as a left inverse to f .

Consider the following map $h|_{\text{Image}(f)}: \text{Image}(f) \rightarrow X$, where we send $y \in \text{Image}(f)$ back to the element that it was mapped from. Now, define g as follows

$$g: Y \rightarrow X$$

$$y \mapsto g(y) = \begin{cases} x, & \text{If } y \in Y \setminus \text{Image}(f) \\ h(y), & \text{If } y \in \text{Image}(f) \end{cases}$$

We note that if $\text{Image}(f) = Y$ then we do not need to consider the first case x , If $y \in Y \setminus \text{Image}(f)$, however if $\text{Image}(f) \subset Y$ then there exists at least one x for this case.

Now with this g we have that

$$g \circ f(x) = g(f(x)) = h(f(x)) = x = \text{id}_X$$

Hence g is indeed a left inverse of f .

The proposition now follows. \square

Proposition 1.2.28. Condition for the existence of a right inverse

Let $f: X \rightarrow Y$ be a mapping with $X \neq \emptyset$. We have that f has a right inverse $g: Y \rightarrow X$ such that $f \circ g = \text{id}_Y$ if and only if f is a surjective mapping.

Proof:

(\Rightarrow): Suppose that f has a right inverse $g: Y \rightarrow X$ such that $f \circ g = \text{id}_Y$. We know by proposition 1.2.26 that id_X is a surjective mapping, moreover we know by proposition 1.2.21 that if a composite map $f \circ g$ is surjective then so is f . Hence as $f \circ g = \text{id}_Y$ and id_Y is surjective, we conclude that f is a surjective map.

(\Leftarrow): Suppose that f is a surjective map, then $\forall y \in Y, \exists x \in X: f(x) = y$. We need to construct a $g: Y \rightarrow X$ such that $f \circ g = \text{id}_Y$. As f is surjective we have that $\forall y \in Y, \exists x \in X: f(x) = y$, in particular we know that there maybe more than one such x so that $f(x) = y$, if this is the case we pick for that y one of the possible choices of x . Hence we can define $g(y) = x$ for every $y \in Y$ then we have that $f \circ g(y) = f(g(y)) = f(x) = y = \text{id}_Y$

The proposition now follows. \square

These two propositions give the following immediate results

Proposition 1.2.29. Left inverse of injective mapping is a surjection

Let $f: X \rightarrow Y$ be an injection with left inverse $g: Y \rightarrow X$. We have that g is a surjection.

Proof let f and g be as stated. Then by definition of a left inverse we have that $g \circ f = \text{id}_X$. Moreover we have the identity mapping id_X is an injection as it is bijective. We then have by proposition 1.2.21 that g is a surjection. \square

Proposition 1.2.30. Right inverse of surjective mapping is an injection

Let $f: X \rightarrow Y$ be a surjection with right inverse $g: Y \rightarrow X$. We have that g is an injection.

Proof let f and g be as stated. Then by definition of a right inverse we have that $f \circ g = \text{id}_Y$. Moreover we have the identity mapping id_Y is a surjection as it is bijective. We then have by proposition 1.2.21 that g is an injection. \square

The ideas of a left and right inverse will allow us to construct the idea of a so-called two-sided inverse, that is an inverse which is both a left inverse and a right inverse. this will allow us to consider when a mappings can be inverted without regards to how we compose the mappings. However there is one final result about left and right inverse that will be required in order to pave the way.

Proposition 1.2.31. Bijection has a left and right inverse

Let $f: X \rightarrow Y$ be a bijective mapping. We have that there exists a left inverse $g: Y \rightarrow X$ and there exists a right inverse $h: Y \rightarrow X$ such that

$$\begin{aligned}g \circ f &= \text{id}_X \\ f \circ h &= \text{id}_Y\end{aligned}$$

Proof:

Let $f : X \rightarrow Y$ be a bijection. We have that as f is a bijection then we know that f is both injective and surjective. Now by proposition 1.2.27 that a left inverse exists if and only if f is an injective mapping. Likewise by proposition 1.2.28 we have that a right inverse exists if and only if f is a surjective mapping. Hence we have the existence of a left and right inverse. As required. \square

Proposition 1.2.32. *The existence of a left and right inverse implies a bijection*

Let $f : X \rightarrow Y$ be a mapping such that $\exists g_1 : Y \rightarrow X$ such that $g_1 \circ f = \text{id}_X$ and $\exists g_2 : Y \rightarrow X$ such that $f \circ g_2 = \text{id}_Y$. We have that f is a bijection.

Proof:

Let $f : X \rightarrow Y$ be a mapping such that $\exists g_1 : Y \rightarrow X$ such that $g_1 \circ f = \text{id}_X$ and $\exists g_2 : Y \rightarrow X$ such that $f \circ g_2 = \text{id}_Y$. We have by proposition 1.2.27 that as g_1 is a left inverse of f then f must be injective. Likewise by proposition 1.2.28 that as g_2 is a right inverse of f then f must be surjective. It hence follows by definition that f is a bijective mapping. \square

These propositions are useful in proving the following.

Proposition 1.2.33. *Bijection if and only if left and right inverses exist*

Let $f : X \rightarrow Y$ be a mapping. We have that f is bijective if and only if $\exists g_1 : Y \rightarrow X$ such that $g_1 \circ f = \text{id}_X$ and $\exists g_2 : Y \rightarrow X$ such that $f \circ g_2 = \text{id}_Y$.

Proof:

(\Rightarrow): Let $f : X \rightarrow Y$ be a bijective mapping. We have by proposition 1.2.31 we have that f being a bijection gives the existence of a left and right inverse.

(\Leftarrow): Suppose we have a mapping $f : X \rightarrow Y$ such that $\exists g_1 : Y \rightarrow X$ such that $g_1 \circ f = \text{id}_X$ and $\exists g_2 : Y \rightarrow X$ such that $f \circ g_2 = \text{id}_Y$. Then f has both a left inverse and a right inverse, hence by proposition 1.2.32 we have that f is a bijection.

The result is shown. \square

We have seen that if $f : X \rightarrow Y$ is a bijection then f has both a left and a right inverse, likewise if these two inverses exist then we have that f is a bijection. This property is key to defining what we mean by the inverse to a bijective mapping.

Definition 1.2.39. *Inverse*

Let $f : X \rightarrow Y$ be a mapping. We say that the mapping $g : Y \rightarrow X$ is an inverse⁶ of f if we have that g is both a left inverse and a right inverse for f . This is to say, g is an inverse of f if we have that

$$\begin{aligned}g \circ f &= \text{id}_X \\ f \circ g &= \text{id}_Y\end{aligned}$$

We sometimes use the notation f^{-1} to denote the inverse.

Example 1.2.37. Let $f : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ be such that $f(x) = x^2$, then we have that $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ with $g(x) = \sqrt{x}$ is an inverse of f . Indeed

$$\begin{aligned}g \circ f(x) &= g(f(x)) = g(x^2) = \sqrt{x^2} = x = \text{id}_{\mathbb{R}^+} \\ f \circ g(x) &= f(g(x)) = f(\sqrt{x}) = (\sqrt{x})^2 = x = \text{id}_{\mathbb{R}^+}\end{aligned}$$

⁶We will first need to prove that in order to speak of the inverse of a mapping that we will need the left and right inverses to be equal

Example 1.2.38. The identity mapping $\text{id}_X : X \rightarrow X$ with $\text{id}_X(x) = x, \forall x \in X$ is its own inverse, indeed

$$\text{id}_X \circ \text{id}_X = \text{id}_X (\text{id}_X(x)) = \text{id}_X(x) = x = \text{id}_X$$

Example 1.2.39. Let $f : \{1, 2\} \rightarrow \{a, b\}$ be such that $f(1) = a$ and $f(2) = b$. We have that $g : \{a, b\} \rightarrow \{1, 2\}$ with $g(a) = 1$ and $g(b) = 2$ is an inverse to f . Indeed we have that

$$\begin{aligned} f(g(a)) &= f(1) = a \\ f(g(b)) &= f(2) = b \end{aligned}$$

It also follows that g is an inverse to f , indeed

$$\begin{aligned} g(f(1)) &= g(a) = 1 \\ g(f(2)) &= g(b) = 2 \end{aligned}$$

Example 1.2.40. Let $f : \mathbb{R} \rightarrow \mathbb{R}^+$ be given by $f(x) = e^x$. We have that $g : \mathbb{R}^+ \rightarrow \mathbb{R}$ where $g(x) = \ln(x)$ is an inverse of f .

We shall prove that the composition of a mapping and its inverse gives the identity mapping. Firstly, we will need to show the following propositions.

Proposition 1.2.34. Mapping is injective and surjective if and only if the inverse is a mapping

Let $f : X \rightarrow Y$ be a mapping. We have that f is a bijection if and only if f^{-1} , the inverse of f , is a mapping.

Proof:

(\Rightarrow) : Let $f : X \rightarrow Y$ be a bijection, then f is both surjective and injective. Let $y \in Y$, then as f is surjective we have that $\exists x \in X$ such that $f(x) = y$, moreover by injectivity of f we have that there is only one such x which does this. Define $g : Y \rightarrow X$ by

$$g(y) = x$$

As $y \in Y$ is an arbitrary element, it follows that

$$\forall y \in Y : \exists x \in X : g(y) = x$$

such that x is unique for a given y . That is g is a mapping. Now by the definition of g we have that

$$\forall y \in Y : f(g(y)) = y$$

Now, let $x \in X$ and let

$$x' = g(f(x))$$

then

$$f(x') = f(g(f(x))) = f(x)$$

by the above. However, f is an injection so we have that $x' = x$ and thus $x = g(f(x))$.

It follows that f and g are inverse mappings of each other.

(\Leftarrow) : Suppose that $f : X \rightarrow Y$ is a mapping, moreover suppose that $f^{-1} : Y \rightarrow X$ is also a mapping which is the inverse of f . We show that f must be a bijection.

1. f is injective:

Let $x, y \in X$ and suppose that $f(x) = f(y)$

$$\begin{aligned} f(x) &= f(y) \\ f^{-1}(f(x)) &= f^{-1}(f(y)) \\ \Rightarrow x &= y, \text{ As } f^{-1} \text{ is the inverse of } f \end{aligned}$$

Hence we have that f is injective.

2. f is surjective:

Suppose that $y \in Y$. We then have that

$$\begin{aligned} y &\in Y \\ \Rightarrow f^{-1}(y) &\in X, \text{ As } f^{-1} \text{ is the inverse of } f \\ \Rightarrow f(f^{-1}(y)) &= y, \text{ By definition of an inverse mapping} \\ \Rightarrow \exists x \in X : f(x) &= y, \text{ Where } x = f^{-1}(y) \end{aligned}$$

Hence we have that f is surjective.

As f is both injective and surjective it is a bijection. \square

We can now show that the inverse of a bijective mapping is also a bijective mapping.

Proposition 1.2.35. *Inverse of a bijective mapping is a bijective mapping*

Let $f : X \rightarrow Y$ be a bijective mapping. We have that $f^{-1} : Y \rightarrow X$, the inverse of f , is also a bijection.

Proof:

Let $f : X \rightarrow Y$ be a bijective mapping. By definition of being a bijection we have that f is both injective and surjective. By proposition 1.2.34 we have that f^{-1} is a mapping. Now it is clear that the inverse of the inverse is the original mapping that is.

$$(f^{-1})^{-1} = f$$

Now, f is a bijection and thus is a mapping. But as f is a mapping we have that by proposition 1.2.34 we have that f^{-1} is a bijection. As required. \square

We can now see that the composition of a bijective mapping with its inverse must be the identity map.

Proposition 1.2.36. *Composition of bijective mapping with the inverses is the identity mapping*

Let $f : X \rightarrow Y$ be a bijective mapping, and let $f^{-1} : Y \rightarrow X$ be the inverse mapping of f . We have that

$$\begin{aligned} f \circ f^{-1} &= \text{id}_Y \\ f^{-1} \circ f &= \text{id}_X \end{aligned}$$

Proof:

Let $f : X \rightarrow Y$ be a bijective mapping, with inverse given by $f^{-1} : Y \rightarrow X$. As f is bijective we have that by proposition 1.2.35 we have that f^{-1} is a bijection. Let $x \in X$, then we have that

$$\exists y \in Y : f(x) = y \Rightarrow f^{-1}(y) = x$$

Hence, we have that

$$\begin{aligned}
f^{-1} \circ f(x) &= f^{-1}(f(x)), \text{ By function composition} \\
&= f^{-1}(y), \text{ By above} \\
&= x, \text{ By above} \\
&= \text{id}_X(x), \text{ By the definition of the identity map of } X
\end{aligned}$$

We have that the domain of $f^{-1} \circ f$ is clearly X , likewise the co-domain is X , which is the same as id_X . Moreover $\forall x \in X$ we have $f^{-1} \circ f(x) = x = \text{id}_X(x)$. So the mappings are equal. Likewise, let $y \in Y$, then we have that

$$\exists x \in X : f^{-1}(y) = x \Rightarrow f(x) = y$$

Hence, we have that

$$\begin{aligned}
f \circ f^{-1}(y) &= f(f^{-1}(y)), \text{ By function composition} \\
&= f(x), \text{ By above} \\
&= y, \text{ By above} \\
&= \text{id}_Y(y), \text{ By the definition of the identity map of } Y
\end{aligned}$$

We have that the domain of $f \circ f^{-1}$ is clearly Y , likewise the co-domain is Y , which is the same as id_Y . Moreover $\forall y \in Y$ we have $f \circ f^{-1}(y) = y = \text{id}_Y(y)$. So the mappings are equal.

In both cases the composition yields the required identity mappings, as required. \square

1.3 The Natural numbers

The natural numbers are the work of God.
All the rest is the work of mankind.

Leopold Kronecker (Paraphrased)

1.3.1 Constructing the Natural numbers

We now have enough tools and core theory to start building up from the foundations of mathematics. We do this using the ZFC axioms, although perhaps not with the complete rigour we should be using. We touched on these briefly in section 1.2.1.5. We will state them again.

1. The axiom of extensionality:

The axiom of extensionality asserts that two sets are equal if and only if they contain the same elements.

2. The axiom of the empty-set:

The axiom of the empty-set asserts that there exists a set which contains no elements

3. The axiom of pairing:

The axiom of pairing asserts that given any set A and any set B , there is a set C such that, given any set D , D is a member of C if and only if D is equal to A or D is equal to B . This is to say, given two sets, there is a set whose members are exactly the two given sets.

4. The axiom of specification:

The axiom of specification asserts that we can construct a set which satisfies a given condition, so long as this condition is not inherently contradictory.

5. The axiom of unions:

The axiom of unions asserts that we can perform the union of two sets A and B

6. The axiom of powers:

The axiom of powers asserts that for any set S we can construct a set $P(S)$ whose elements are all the possible subsets of S .

7. The axiom of infinity:

The axiom of infinity asserts that there is at least one infinite set A , that is at least one set with infinitely many elements. That is we have a set A such that the $\emptyset \in A$ and if $x \in A$ then the set $x \cup \{x\}$ is also in A .

8. The axiom of replacement:

We will need the next section to fully understand this axiom, however informally asserts that for some set S , and form another set by replacing the elements of S by other sets according to any definite rule.

9. The axiom of foundation:

The axiom of foundation asserts that for every non-empty set S , there exists an element $x \in S$ such that x and S are disjoint. This also asserts that no set can contain itself.

We also recall that we include the symbol \in in the ZFC axioms, which allows us to talk about element inclusions in sets. In other words, ZFC defines a set of axioms that allow us to talk about sets and elements of sets. Next, we have that, formally speaking, ZFC is allowed to make statements about mappings. Finally, we will ZFC has the power to prove the results in the previous two sections we made on sets and mappings,

so we will assume these as well. We will use this as the building blocks for building the natural numbers. How can we do this from the ZFC axioms?

As it stands right now ZFC only gives us the existence of the empty set, and there is at least a set which contains infinitely many elements. We start with the empty set, a set which contains no elements, we can use the ZFC axioms to build a new set which contains the empty set.

Our ultimate goal is to identify each natural number with the number of elements in some corresponding set. Hence naturally the empty set containing no elements would be identified with the number 0, and so on. The question is given that we only have the empty set, how can we build a new set? We can use the axiom of powers. This states that we can take any set S and construct a new set $P(S)$ whose elements are the possible subsets of S . Applying this to the empty-set, a set which contains no elements and thus has no subsets except for itself, must give us $P(\emptyset) = \{\emptyset\}$. This is sufficient for what we need to do.

So, we have two sets, \emptyset and $\{\emptyset\}$. We shall identify \emptyset with 0 and $\{\emptyset\}$ with 1.

Definition 1.3.1. Zero

We define the number zero to be \emptyset . That is, we say Zero is a set that contains no elements.

Definition 1.3.2. One

We define the number one to be $\{\emptyset\}$. That is, we say One is the set whose only element is \emptyset .

How do we define any more numbers? We can use the axiom of unions. This raises the question why not use the axiom of powers again? If we apply the axiom of powers to $\{\emptyset\}$ we get the set

$$P(\{\emptyset\}) = \{\emptyset, \{\emptyset\}\}$$

If we assume we already know what the natural numbers are, we could identify this with the number 2. However, a repeated application of the axiom of powers would give us

$$P(\{\emptyset, \{\emptyset\}\}) = \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \{\emptyset, \{\emptyset\}\}\}$$

Which we would identify with the number 4. Another application would give us a set that we would identify with the number 8. Clearly, we are skipping numbers such as 3, 5, 7, 9 etc. We can't get additional numbers that aren't powers of 2. Instead, we can define an operation that will allow us to construct each number one at a time.

This operation uses the axiom of unions, and starts of with the numbers 0 and 1, which we recall are the sets \emptyset , and $\{\emptyset\}$ respectively. Applying the axiom of unions to these two sets gives us

$$\emptyset \cup \{\emptyset\} = \{\emptyset, \{\emptyset\}\}$$

This is in agreement with $P(\{\emptyset\})$, so we can identify this with the number 2. Now, the axiom of pairing allows us to create a set that contains as elements any two sets that have already been created. Applying this to $\{\emptyset, \{\emptyset\}\}$ with itself allows us to create the set $\{\{\emptyset, \{\emptyset\}\}\}$. Hence we can now apply this operation again on the set $\{\emptyset, \{\emptyset\}\}$ to get

$$\{\emptyset, \{\emptyset\}\} \cup \{\{\emptyset, \{\emptyset\}\}\} = \{\emptyset, \{\emptyset\}, \{\{\emptyset, \{\emptyset\}\}\}\}$$

A set of 3 elements so we identify this with the number 3. We can keep doing this to build the Natural numbers. Lets make some definitions

Definition 1.3.3. The successor operation

Let x be a set. We define the successor operation, denoted by S to be given by

$$S(x) = x \cup \{x\} \tag{3}$$

We call this the successor function, as it is clear in the context of the Natural numbers that $S(n) = n+1$, but we shall prove this later.

This definition allows us to essentially make any finite number. This leads us to our first potential definition for the Natural numbers. We first need to define the idea of recursion.

We have the following proposition

Proposition 1.3.1. *Equality of successor operation*

Let a, b be sets. We have that $S(a) = S(b)$ if and only if $a = b$.

Proof:

(\Rightarrow) : Suppose that a, b are sets and $S(a) = S(b)$. By definition of S we have that

$$a \cup \{a\} = b \cup \{b\}$$

Now, as $a \in S(a)$ and $S(a) = S(b)$ then we have that $a \in b \cup \{b\}$ and so $a \in b$ or $a = b$. Similarly, as $b \in S(b)$ we get that $b \in a \cup \{a\}$ and so $b \in a$ or $b = a$.

Now, if $a = b$ we are done, so suppose $a \neq b$, then we have that $a \in b$ and $b \in a$. Consider the set given by

$$X = \{a, b\}$$

which can be constructed by the Axiom of pairing. Now as $a \in b$ we have that $b \cap \{a, b\} \neq \emptyset$ and likewise as $b \in a$ we have $a \cap \{a, b\} \neq \emptyset$. This contradicts the Axiom of Foundation, X does not contain an element that is disjoint from it. It follows that we can't have $a \neq b$ and conclude that $a = b$.

(\Leftarrow) : This is trivial by the definition of S . \square

There are a few extra properties about the successor function that we shall make use of

Corollary 1.3.1. *Successor mapping is injective*

Let a, b be sets. We have that the successor function is injective, that is for all sets a, b we have that

$$S(a) = S(b) \Rightarrow a = b$$

Proof:

Suppose that a, b are arbitrary sets and that $S(a) = S(b)$, by proposition 1.3.1 this holds if and only if $a = b$. Hence we have injectivity. \square

Corollary 1.3.2. *Empty-set is not the successor of any set*

We have that $\emptyset \neq S(a)$ for all sets a .

Proof:

Consider the definition of $S(a)$ and suppose for contradiction that $\emptyset = S(a)$. We have by definition of the successor mapping that

$$\emptyset = S(a) = a \cup \{a\}$$

This is a contradiction, as $a \cup \{a\}$ is a set of two elements, namely a and $\{a\}$ but the empty-set by definition has no elements. \square

Definition 1.3.4. *Recursive definition of a set*

A set S is defined recursively if the elements of S are defined in terms of other elements $x \in S$. Moreover we have that there is some initial element x_0 which is used to define the other elements of the set.

Definition 1.3.5. *First definition of the Natural numbers*

We define the set \mathbb{N} , called the set of natural numbers, to be the set given by

$$\mathbb{N} = \{x : x = \emptyset \text{ or } x = S(y) \text{ for some } y \in \mathbb{N}\} \quad (4)$$

We have defined \mathbb{N} recursively in terms of elements of \mathbb{N} . As an example $2 \in \mathbb{N}$ as $2 = S(1)$ and likewise $1 = S(0)$ and we know that 0 is really the same as \emptyset , which is the initial element of \mathbb{N} as defined above. This definition allows us to get any $x \in \mathbb{N}$, however it is not quite enough to get every element of \mathbb{N} at the same time. We know that there should be infinitely many natural numbers, indeed for any $n \in \mathbb{N}$ we have also that $n + 1 \in \mathbb{N}$. In other words we have a chain of sets of increasing size, that is we have

$$\begin{aligned}
\mathbb{N}_0 &= \emptyset = 0 \\
\mathbb{N}_1 &= \{\emptyset\} = 1 \\
\mathbb{N}_2 &= \{\emptyset, \{\emptyset\}\} = 2 \\
\mathbb{N}_3 &= \{\emptyset, \{\emptyset\}, \{\{\emptyset, \{\emptyset\}\}\} = 3
\end{aligned}$$

Which satisfy $\mathbb{N}_0 \subset \mathbb{N}_1 \subset \mathbb{N}_2 \subset \mathbb{N}_3 \subset \dots$. So we see at each stage \mathbb{N}_n is a finite set of size n and so ultimately our current definition of \mathbb{N} can ultimately only ever reach a finite n . although we can make this n arbitrarily large. To ensure we get every possible n at once we need to invoke the axiom of infinity.

1. The axiom of infinity:

The axiom of infinity asserts that there is at least one infinite set A , that is at least one set with infinitely many elements. That is we have a set A such that the $\emptyset \in A$ and if $x \in A$ then the set $x \cup \{x\}$ is also in A .

There is a useful definition that we can extract from the axiom of infinity.

Definition 1.3.6. *Inductive set*

Let A be a set and let $f : A \rightarrow A$ be a mapping. We say that A is an inductive if it satisfies the following two properties

1. $\emptyset \in A$
2. If $x \in A$ then $f(x) \in A$

For now, we will be focused on the case where $f = S$, the successor mapping.

In light of the axiom of infinity we have a set that contains the infinitely many Natural numbers. This is nearly what we want, although it won't be the set of Natural numbers. This set could clearly have many, many more things than just the Natural numbers.

We can make a new definition, which will allow us to define \mathbb{N} . We will also be able to show the fact this definition defines \mathbb{N} to be the smallest such inductive set that contains all of the Natural numbers.

Definition 1.3.7. *The set \mathbb{N}_S*

Let S be an inductive set. We define \mathbb{N}_S as follows

$$\mathbb{N}_S = \bigcap_{\substack{A \subseteq S \\ A \text{ is inductive}}} A \quad (5)$$

This is well-defined by the axiom of specification, being an inductive step is definable and the collection of all subsets of S is a set we can define.

We have that all of these sets \mathbb{N}_S are the same.

Theorem 1.3.1. *Every \mathbb{N}_S set is the same set*

Let S and T be inductive sets. Define the sets \mathbb{N}_S and \mathbb{N}_T . We have that

$$\mathbb{N}_S = \mathbb{N}_T \quad (6)$$

Proof:

By the axiom of extensionality we know that two sets are equal if and only if they contain the same elements. To see that \mathbb{N}_S and \mathbb{N}_T have the same elements consider the new set given by

$$C = \mathbb{N}_S \cap \mathbb{N}_T$$

We recall from proposition 1.2.8 that for two sets A and B we have $A \cap B \subseteq A$. Hence it follows that

$$C = \mathbb{N}_S \cap \mathbb{N}_T \subseteq \mathbb{N}_S$$

That is, $C \subseteq \mathbb{N}_S$, that is to say every element of C is also an element of \mathbb{N}_S . Now recall the definition of \mathbb{N}_S ,

$$\mathbb{N}_S = \bigcap_{\substack{A \subseteq S \\ A \text{ is inductive}}} A$$

We know that $C \subseteq \mathbb{N}_S$, hence as \mathbb{N}_S is the intersection of all subsets of S we must conclude that $C \subseteq S$.

Now, we know that S is an inductive set. Hence S satisfies the following

1. $\emptyset \in S$
2. If $x \in S$ then $S(x) \in S$

If we can show that C is an inductive set we know that C was one of the sets we used to construct \mathbb{N}_S and hence $\mathbb{N}_S \subseteq C$, which will give the equality $C = \mathbb{N}_S$.

Now, to show that C is an inductive set we must show that

1. $\emptyset \in C$
2. If $x \in C$ then $S(x) \in C$

1. $\emptyset \in C$:

We have that $C = \mathbb{N}_S \cap \mathbb{N}_T$ and we have that

$$\begin{aligned} \mathbb{N}_S &= \bigcap_{\substack{A \subseteq S \\ A \text{ is inductive}}} A \\ \mathbb{N}_T &= \bigcap_{\substack{A \subseteq T \\ A \text{ is inductive}}} A \end{aligned}$$

In the definitions of both \mathbb{N}_S and \mathbb{N}_T we have that these are the intersections of inductive sets and so $\emptyset \in \mathbb{N}_S$ and $\emptyset \in \mathbb{N}_T$. It hence follows that as $C = \mathbb{N}_S \cap \mathbb{N}_T$ we must have $\emptyset \in C$.

2. If $x \in C$ then $S(x) \in C$:

Now suppose that $x \in C$. Like before we know that $C = \mathbb{N}_S \cap \mathbb{N}_T$, and by the definition of the intersection of two sets, it follows that $x \in \mathbb{N}_S$ and $x \in \mathbb{N}_T$. Now we have that

$$\mathbb{N}_S = \bigcap_{\substack{A \subseteq S \\ A \text{ is inductive}}} A$$

hence as $x \in \mathbb{N}_S$ we have we must have that $x \in A$ for every subset A of S . Moreover each such A is an inductive set and so by definition of an inductive set we have that $S(x) \in A$ for every subset A of S . Hence $S(x) \in \mathbb{N}_S$ and likewise a similar argument shows that $S(x) \in \mathbb{N}_T$. It thus follows that $S(x) \in C$.

As $x \in C$ was arbitrary we must conclude that this holds for any $x \in C$.

Hence C is an inductive set.

Now, we know that $C \subseteq S$ and C is an inductive set then it follows that C is one of the inductive sets in the definition of \mathbb{N}_S . It hence follows that $\mathbb{N}_S \subseteq C$. It follows by the axiom of extensionality that as \mathbb{N}_S and C contain the same elements then $C = \mathbb{N}_S$.

Likewise the a similar argument shows that $C = \mathbb{N}_T$. So it follows that $\mathbb{N}_S = \mathbb{N}_T$. \square

In light of this theorem we can now truly define \mathbb{N} .

Definition 1.3.8. *The Natural numbers \mathbb{N}*

Let S be an inductive set, and construct the set \mathbb{N}_S . The set \mathbb{N}_S is the set of Natural numbers and by theorem 1.3.1 no matter the inductive set S we have that all such \mathbb{N}_S are the same. Hence we simply refer to the natural numbers by \mathbb{N} .

We identify the elements of \mathbb{N} not in terms of \emptyset , and sets of sets containing \emptyset , but instead by the more usually numerals that we use. We have already defined Zero and One, by definitions 1.3.1 and 1.3.2. The other numbers follow likewise, i.e

$$\begin{aligned} 0 &= \emptyset \\ 1 &= S(0) = \{\emptyset\} \\ 2 &= S(1) = \{\emptyset, \{\emptyset\}\} \\ 3 &= S(2) \\ 4 &= S(3) \\ &\dots \\ n+1 &= S(n) \end{aligned}$$

We said that we can prove that \mathbb{N} is the smallest such inductive set that contains all the natural numbers, this is to say if $A \subseteq \mathbb{N}$ is an inductive set we must have that $A = \mathbb{N}$. We thankfully do not need to prove this as the previous theorem gives this for free. This also gives us the following definition for a minimally inductive set, we make the definition in such a way that we argue about sets of inductive sets.

Definition 1.3.9. *Minimally inductive set of sets*

Let S be a set whose elements are also sets satisfying some condition, and let $f : S \rightarrow S$ be a mapping. We say that S is minimally inductive if and only if the following holds

1. S is an inductive set under the mapping g
2. No proper subset of S is inductive under the mapping g

One of the most powerful properties of the natural numbers is the principle of Induction. This tool is powerful in proving many statements on the Natural numbers. It works in a similar way to how an inductive set works⁷. We show that the statement works for some base case, usually $n = 0$, then we assume that if it holds true for some n then it holds true for $S(n) = n + 1$.

Theorem 1.3.2. *The principle of induction*

Suppose we have a proposition $P(n)$ about a Natural number $n \in \mathbb{N}$. Moreover, suppose that

1. $P(0)$ is true
2. $P(n)$ being true implies $P(S(n)) = P(n + 1)$ is true for any Natural number n .

If these two statements are true, we have that $P(n)$ is true for any natural number n , and we say the proposition $P(n)$ holds by the principle of mathematical induction.

Moreover we call $P(0)$ the base case for induction and $P(n)$ being true implies $P(n + 1)$ being true is the inductive step.

Proof:

Let $P(n)$ be a proposition about a Natural number $n \in \mathbb{N}$ such that $P(n)$ satisfies

1. $P(0)$ is true
2. $P(n)$ being true implies $P(S(n)) = P(n + 1)$ is true for any Natural number n .

⁷Hence the similar names.

Consider the set given by

$$Q = \{n : P(n) \text{ is true}\}$$

That is, Q is defined as the set of Natural numbers such that that $P(n)$ is true, clearly $Q \subseteq \mathbb{N}$. By hypothesis we know that $P(0)$ is true, so $0 \in Q$. Also by hypothesis we know that if $P(n)$ is true for some $n \in \mathbb{N}$. then we have that $P(S(n)) = P(n+1)$ is also true, hence we have that every $n \in \mathbb{N}$ is also in Q , hence $\mathbb{N} \subseteq Q$ and so by the axiom of extensionality we have that $Q = \mathbb{N}$. Hence $P(n)$ is true for every Natural number $n \in \mathbb{N}$. \square .

Now that we have induction we can make a final definition that will be useful. This definition combines a few previously proven results into a convenient package, this package has the strength to prove the usual properties of the natural numbers and perhaps are an easy way to remember the basis for deducing properties about the natural numbers.

Definition 1.3.10. *The Peano axioms*

We define the Peano axioms as follows. Let A be a set and consider the successor mapping on A , $S : A \rightarrow A$. If we have that

1. A is an inductive set
 - (a) $\emptyset \in A$
 - (b) If $x \in A$ then $S(x) \in A$
2. S is an injective mapping.
3. $\forall x \in S$ we have that $\emptyset \neq S(x)$
4. $\forall B \subseteq A$. If $0 \in B$ and $S(n) \in B$ for all $n \in B$ then $B = A$

If A satisfies all of the above, then we say that A satisfies the Peano axioms and induces Peano arithmetic.

1.3.2 Properties of the natural numbers

Although we have constructed \mathbb{N} we haven't defined what we can do with this set. We know from our intuitions that we can define addition, a form of subtraction, multiplication and in some cases division. We also know that there is some notion of a Natural number being larger or smaller than another, when two Natural numbers are equal and so. We will explore some of these properties so that we can start doing some form of Mathematics.

1.3.2.1 Equality of natural numbers

Firstly, it is important to define when two Natural numbers are equal, again as we have defined the natural numbers in terms of Sets, this just comes down to the axiom of extensionality.

Definition 1.3.11. *Equality of natural numbers*

Let $n, m \in \mathbb{N}$ be two natural numbers. We define that two natural numbers are equal, denoted $n = m$ if and only if $n \subseteq m$ and $m \subseteq n$. This is simply the axiom of extensionality.

If we do not have $n = m$ then we say that n and m are not equal and we denote this $n \neq m$.

This definition clearly makes sense as each natural number is a set.

Example 1.3.1. We have that $1 = 1$. Indeed by definition $0 = \emptyset$ and $1 = \{\emptyset\}$. It is clear that $\{\emptyset\} \subseteq \{\emptyset\}$ hence the axiom of extensionality gives us that $\{\emptyset\} = \{\emptyset\}$. That is $1 = 1$

Example 1.3.2. We have that $3 = 3$. Indeed by construction we have that $3 = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}$ It is clear that $\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\} \subseteq \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}$ hence the axiom of extensionality gives us that $\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\} = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}$. i.e $3 = 3$

Example 1.3.3. We have $1 \neq 2$. We have that $1 = \{\emptyset\}$ and $2 = \{\emptyset, \{\emptyset\}\}$. Now $\{\emptyset\} \subseteq \{\emptyset, \{\emptyset\}\}$ but $\{\emptyset, \{\emptyset\}\} \not\subseteq \{\emptyset\}$.

In particular in light of the definition of equality on the natural numbers if $n = m$ and $m = k$ we must have that $n = k$.

1.3.2.2 Inequality of natural numbers

We can define also define what it means for natural numbers to not be equal. We make use of the notion of set inclusion. Recall that a set S is a subset of the set T , written $S \subseteq T$, if for every $s \in S$ we have that $s \in T$ and that S is a proper subset of T , written $S \subset T$ if $S \subseteq T$ and $S \neq T$. We will use the proper subset notation to define the so-called less than operator. This operation comes naturally from the definition of the natural numbers by the successor mapping. The successor function has the following chain of definitions for each $n \in \mathbb{N}$

$$\begin{aligned} 0 &= \emptyset \\ 1 &= S(0) = \{\emptyset\} \\ 2 &= S(1) = \{\emptyset, \{\emptyset\}\} \\ 3 &= S(2) = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} \\ 4 &= S(3) = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}\} \\ &\dots \\ n+1 &= S(n) \end{aligned}$$

From this chain of definitions and the axiom of foundation, $0 = \emptyset$ is the set element minimal element of \mathbb{N} , so every natural number is contained in one that comes after. We can make the following definition which defines when one natural number is smaller than another.

Definition 1.3.12. *Less than Operator*

Let $n, m \in \mathbb{N}$. The less than operator, denoted by $n < m$ and read as n is less than m , is defined as follows.

We have $n < m$ if and only if $n \subset m$. The set that denotes the number n is an element of the set m . In the language of mathematical logic, we have that that $<$ is actually a logical proposition, given by

$$<(n, m) = \begin{cases} 1, & \text{If } n \subset m \\ 0, & \text{Otherwise} \end{cases}$$

Recall that for predicates 0 indicates that the predicate is false and 1 indicates that the predicate is true.

Example 1.3.4. We have that $2 < 3$. Indeed $2 = \{\emptyset, \{\emptyset\}\}$ and $3 = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$, and clearly

$$\{\emptyset, \{\emptyset\}\} \subset \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$$

We can combine the less than operator with the equality operator.

Definition 1.3.13. *Less than or equal to operator*

Let $n, m \in \mathbb{N}$. The less than or equal to operator, denoted by $n \leq m$, and read as n is less than or equal to m , is defined the same as $n < m$ except we now allow for the situation that $n = m$. This is to say \leq is a logical proposition given by

$$\leq(n, m) = \begin{cases} 1, & \text{If } n < m \\ 1, & \text{If } n = m \\ 0, & \text{Otherwise} \end{cases}$$

Where on the right-hand side of the definition we are talking about sets, and on the left-hand side we are talking about natural numbers, although we know these are the same thing.

Example 1.3.5. We have that $2 \leq 3$. From the previous example, we know that $2 < 3$. Moreover, we have that $3 \leq 3$ as $3 = 3$.

We have defined less than and less than or equal to, we can define a similar notation of greater than and greater than then equal to, we can do this by considering when $n \not\subset m$.

Definition 1.3.14. *Greater than operator*

Let $n, m \in \mathbb{N}$. The greater than operator, denoted by $n > m$ and is read as n is greater than m , is defined as follows.

We have $n > m$ if and only if $n \notin m$. That is, the set that denotes the number n is not an element of the set m . That is to say that $>$ is a logical proposition, given by

$$> (n, m) = \begin{cases} 1, & \text{If } n \notin m \\ 0, & \text{Otherwise} \end{cases}$$

Likewise, we can define the greater than or equal to operator.

Definition 1.3.15. *Greater than or equal to operator*

Let $n, m \in \mathbb{N}$. The greater than or equal to operator, denoted by $n \geq m$, and read as n is greater than or equal to m , is defined the same as $n > m$ except we now allow for the situation that $n = m$. This is to say \geq is a logical proposition given by

$$\geq (n, m) = \begin{cases} 1, & \text{If } n > m \\ 1, & \text{If } n = m \\ 0, & \text{Otherwise} \end{cases}$$

1.3.2.3 Defining addition and multiplication on the Natural numbers

We can use the principle of induction to make definitions as well as a proof technique. We shall use induction now to make two definitions, in particular, we define two mappings that will allow us to start manipulating Natural numbers as we expect them to. To do so it is enough to specify what the mapping does when 0 is given as an argument, and then do define what the mapping does when given $S(n)$ as an argument, hence defining it in terms of n for each $n \in \mathbb{N}$. This will make sense when we define these operations.

We first recall the Cartesian product of two sets. Let S and T be sets, the Cartesian product of S and T , denoted $S \times T$ is the set of all ordered pairs of the form (S, t) where $s \in S$ and $t \in T$. This is to say that

$$S \times T = \{(s, t) : s \in S, t \in T\}$$

If $S = T$ then we denote $S \times T$ by S^2 .

Definition 1.3.16. *Addition on the Natural numbers*

We define addition on the Natural numbers by the following mapping. Let $+: \mathbb{N}^2 \rightarrow \mathbb{N}$ be such that for all $(m, n) \in \mathbb{N}^2$ we have the following

$$+ : \mathbb{N}^2 \rightarrow \mathbb{N} \tag{7}$$

$$(m, n) \mapsto + (m, n) = \begin{cases} m + 0 = m, & \text{If } n = 0 \\ m + S(n) = S(m + n), & \text{If } n \neq 0 \end{cases} \tag{8}$$

We will write $+(m, n)$ as $m + n$.

In light of this definition, we can prove that $1 + 1 = 2$

Theorem 1.3.3. *1+1=2*

We have that $1 + 1 = 2$.

Proof:

We know that $1 = S(0)$ and $2 = S(S(0))$. Hence, we are proving

$$S(0) + S(0) = S(S(0))$$

By the definition of the addition mapping, we know that $\forall (m, n) \in \mathbb{N}^2$ that

$$m + S(n) = S(m + n)$$

In particular if $n = 0$ we have $\forall m$ that

$$m + S(0) = S(m + 0)$$

and

$$S(0) + S(0) = S(S(0) + 0) \quad (9)$$

Moreover, by the definition of addition, we know that $\forall m$ that if $n = 0$ then

$$m + 0 = m$$

Hence

$$\begin{aligned} S(0) + 0 &= S(0) \\ \Rightarrow S(S(0) + 0) &= S(S(0)) \\ \Rightarrow S(0) + S(0) &= S(S(0)) \end{aligned}$$

This is to say. $1 + 1 = 2$. As required. \square .

Definition 1.3.17. *Multiplication on the Natural numbers*

We define multiplication on the Natural numbers by the following mapping. Let $*$: $\mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$ be such that for all $(m, n) \in \mathbb{N} \times \mathbb{N}$ we have the following

$$* : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N} \quad (10)$$

$$(m, n) \mapsto * (m, n) = \begin{cases} m * 0 = 0, & \text{If } n = 0 \\ m * S(n) = m * n + m, & \text{If } n \neq 0 \end{cases} \quad (11)$$

We will write $* (m, n)$ as $m * n$, or more compactly just as the juxtaposition mn

As with addition we provide a proof that $2 * 2 = 4$

Theorem 1.3.4. $2 * 2 = 4$

We have $2 * 2 = 4$.

Proof:

We know that $S(1) = 2$ and so by definition of multiplication we have that

$$2 * 2 = 2 * S(1) = 2 * 1 + 2$$

Likewise we know that $S(0) = 1$ and so by another application of the definition of multiplication we have that

$$2 * 1 + 2 = 2 * S(0) + 2 = 2 * 0 + 2 + 2$$

Now $2 * 0 = 0$ by definition as so we have that

$$2 * 2 = 2 * 0 + 2 + 2 = 0 + 2 + 2 = 2 + 2$$

It is left to show that $2 + 2 = 4$. We use a similar proof to $1 + 1 = 2$. As $4 = S(S(2)) = S(S(S(S(0))))$ and $2 = S(S(0))$ we need to show that

$$S(S(0)) + S(S(0)) = S(S(S(S(0))))$$

By the definition of addition we have that $\forall (m, n) \in \mathbb{N}^2$ that

$$m + S(n) = S(m + n)$$

In particular we have that if $n = 0$ and $\forall n \in \mathbb{N}$ that

$$m + S(0) = S(m + 0)$$

So that

$$\begin{aligned} S(S(0)) + S(S(0)) &= S(S(S(0)) + S(0)) \\ &= S(S(S(S(0)) + 0)) \\ &= S(S(S(S(0)))) \end{aligned}$$

That is $2 + 2 = 4$ and so the theorem is proved. \square

These two definitions are enough to prove every elementary property of addition and multiplication that we are familiar with. However to do so will require an upgrade to the idea of induction. This will allow us to perform induction on both the addition and multiplication mappings. Once we have done this we will have put the natural numbers on a firm logical basis. This idea is called double induction, or more clearly induction on two variables.

For example, we know from school that $n + m = m + n$ for all natural numbers n and m . To show that this is true, we start by induction on n , so we have to show that $m + 0 = 0 + m$ and then that $(m + n = n + m)$ implies that $(m + S(n) = S(n) + m)$, each of these will be proved by induction on m . This is the idea of double induction.

Theorem 1.3.5. Double induction

Let $P(m, n)$ be a proposition about a pair of natural numbers $m, n \in \mathbb{N}$. Moreover suppose that

1. $P(0, 0)$ is true.
2. $P(0, n)$ being true implies that $P(0, S(n))$ is true.
3. $P(m, 0)$ being true implies that $P(S(m), 0)$ is true
4. For a given $m \in \mathbb{N}$, from the truth that $P(m, x)$ is true for all x , and also that of $P(S(m), n)$ for some n , we can infer that $P(S(m), S(n))$ is true.

If these statements are true, we have that $P(m, n)$ is true for any natural numbers $m, n \in \mathbb{N}$ and we say that the proposition $P(m, n)$ hold by the principle of mathematical double induction.

Proof:

Let $P(m, n)$ be a proposition about a pair of natural numbers $m, n \in \mathbb{N}$, which satisfies

1. $P(0, 0)$ is true.
2. $P(0, n)$ being true implies that $P(0, S(n))$ is true.
3. $P(m, 0)$ being true implies that $P(S(m), 0)$ is true
4. For a given $m \in \mathbb{N}$, from the truth that $P(m, x)$ is true for all x , and also that of $P(S(m), n)$ for some n , we can infer that $P(S(m), S(n))$ is true.

Statements 1 and 2 are the base case and the inductive step for the proof of $P(0, n)$ for all $n \in \mathbb{N}$. Likewise statements 1 and 3 are the base case and the inductive step for the proof of $P(m, 0)$ for all $m \in \mathbb{N}$.

Finally, the statements 3 and 4 is the base case and inductive step for a proof, by induction on n for a proof of the statement that if $P(m, n)$ holds for all n , then $P(S(m), n)$ holds for all n , and thus by induction we have that $P(m, n)$ is true for all m . \square .

We can start proving the basic properties of \mathbb{N} that we are familiar with.

1.3.2.4 Closure properties of addition and multiplication

We show that addition and multiplication on the natural numbers produces a natural number.

Theorem 1.3.6. *The addition and multiplication mappings on the natural numbers are closed*

For all $n, m \in \mathbb{N}$. We have that

1. $n + m \in \mathbb{N}$.

2. $nm \in \mathbb{N}$.

Proof:

1. $n + m \in \mathbb{N}$:

Let $n, m \in \mathbb{N}$. We need to show that

(a) $0 + 0 \in \mathbb{N}$

(b) $0 + n \in \mathbb{N}$ implies $0 + S(n) \in \mathbb{N}$

(c) $m + 0 \in \mathbb{N}$ implies $S(m) + 0 \in \mathbb{N}$

(d) *For some $m \in \mathbb{N}$. Suppose that $m + x \in \mathbb{N}$ for all $x \in \mathbb{N}$, and $S(m) + n \in \mathbb{N}$ for some $n \in \mathbb{N}$ implies that $S(m) + S(n) \in \mathbb{N}$*

(a) $0 + 0 \in \mathbb{N}$:

We have by the definition of addition that

$$0 + 0 = 0$$

which is clearly in \mathbb{N} .

(b) $0 + n \in \mathbb{N}$ implies $0 + S(n) \in \mathbb{N}$:

Now, suppose that $0 + n \in \mathbb{N}$ for some n , we show that $0 + S(n) \in \mathbb{N}$.

By the definition of addition we have that

$$0 + S(n) = S(0 + n)$$

Now $0 + n \in \mathbb{N}$ by assumption, therefore we have that $S(0 + n) \in \mathbb{N}$. Hence $0 + S(n) \in \mathbb{N}$.

(c) $m + 0 \in \mathbb{N}$ implies $S(m) + 0 \in \mathbb{N}$:

Now, suppose that $m + 0 \in \mathbb{N}$ for some m , we show that $S(m) + 0 \in \mathbb{N}$.

By the definition of addition we have that

$$S(m) + 0 = S(m) = S(m + 0)$$

Now $m + 0 \in \mathbb{N}$ by assumption, therefore $S(m + 0) \in \mathbb{N}$. Hence $S(m) + 0 \in \mathbb{N}$

(d) *For some $m \in \mathbb{N}$. Suppose that $m + x \in \mathbb{N}$ for all $x \in \mathbb{N}$, and $S(m) + n \in \mathbb{N}$ for some $n \in \mathbb{N}$ implies that $S(m) + S(n) \in \mathbb{N}$*

Now suppose that $m + x \in \mathbb{N}$ for all $x \in \mathbb{N}$ and some fixed $m \in \mathbb{N}$, and suppose that $S(m) + n \in \mathbb{N}$ where n is some fixed value, we show that $S(m) + S(n) \in \mathbb{N}$.

So, we have that $S(m) \in \mathbb{N}$ and $S(n) \in \mathbb{N}$ we can use the definition of addition, doing so gives

$$S(m) + S(n) = S(S(m) + n)$$

By assumption $S(m) + n \in \mathbb{N}$, hence as we have that $m + x \in \mathbb{N}$ for all $x \in \mathbb{N}$, then we have that $S(S(m) + n) \in \mathbb{N}$. Therefore we must conclude that $S(m) + S(n) \in \mathbb{N}$.

Hence by the principle by double induction we have that $m + n \in \mathbb{N}$ for all $m, n \in \mathbb{N}$. That is, addition is closed.

2. $nm \in \mathbb{N}$:

Let $n, m \in \mathbb{N}$. We need to show that

- (a) $0 * 0 \in \mathbb{N}$
- (b) $0 * n \in \mathbb{N}$ implies $0 * S(n) \in \mathbb{N}$
- (c) $m * 0 \in \mathbb{N}$ implies $S(m) * 0 \in \mathbb{N}$
- (d) For some $m \in \mathbb{N}$. Suppose that $m * x \in \mathbb{N}$ for all $x \in \mathbb{N}$, and $S(m) * n \in \mathbb{N}$ for some $n \in \mathbb{N}$ implies that $S(m) * S(n) \in \mathbb{N}$

(a) $0 * 0 \in \mathbb{N}$:

We have by the definition of multiplication that

$$0 * 0 = 0$$

which is clearly in \mathbb{N} .

(b) $0 * n \in \mathbb{N}$ implies $0 * S(n) \in \mathbb{N}$:

Now, suppose that $0 * n \in \mathbb{N}$ for some n , we show that $0 * S(n) \in \mathbb{N}$.

By the definition of multiplication we have that

$$0 * S(n) = 0 * n + 0$$

Now $0 * n \in \mathbb{N}$ by assumption, moreover we have proved that addition is closed, so $0 * n + 0 \in \mathbb{N}$ therefore we have that $0 * S(n) \in \mathbb{N}$

(c) $m * 0 \in \mathbb{N}$ implies $S(m) * 0 \in \mathbb{N}$:

Now, suppose that $m * 0 \in \mathbb{N}$ for some m , we show that $S(m) * 0 \in \mathbb{N}$.

By the definition of addition we have that

$$S(m) * 0 = 0$$

Where $S(m) * 0 = 0$ by definition of multiplication. Hence as $0 \in \mathbb{N}$ we have that $S(m) * 0 \in \mathbb{N}$.

(d) For some $m \in \mathbb{N}$. Suppose that $m * x \in \mathbb{N}$ for all $x \in \mathbb{N}$, and $S(m) * n \in \mathbb{N}$ for some $n \in \mathbb{N}$ implies that $S(m) * S(n) \in \mathbb{N}$:

Now suppose that $m * x \in \mathbb{N}$ for all $x \in \mathbb{N}$ and some fixed $m \in \mathbb{N}$, and suppose that $S(m) * n \in \mathbb{N}$ where n is some fixed value, we show that $S(m) * S(n) \in \mathbb{N}$.

So, we have that $S(m) \in \mathbb{N}$ and $S(n) \in \mathbb{N}$ we can use the definition of multiplication, doing so gives

$$S(m) * S(n) = S(m) * n + S(m)$$

By assumption $S(m) * n \in \mathbb{N}$, moreover as $m * x \in \mathbb{N}$ for all $x \in \mathbb{N}$ we must have $S(m) * n + S(m) \in \mathbb{N}$ as addition is closed.

Hence $S(m) * S(n) \in \mathbb{N}$.

Hence by the principle by double induction we have that $m * n \in \mathbb{N}$ for all $m, n \in \mathbb{N}$. That is, multiplication is closed.

Hence, we have that the addition and multiplication mappings are closed. \square

1.3.2.5 Commutativity of addition and multiplication

This will prove that for all $a, b \in \mathbb{N}$ that $a + b = b + a$ and $ab = ba$.

Theorem 1.3.7. *Addition and multiplication are commutative*

For all $a, b \in \mathbb{N}$ we have that

1. $a + b = b + a$

2. $ab = ba$

Proof:

1. $a + b = b + a$:

We argue by double induction. We need to show that

- (a) $0 + 0 = 0 + 0$

- (b) $0 + n = n + 0$ implies $0 + S(n) = S(n) + 0$

- (c) $m + 0 = 0 + m$ implies $S(m) + 0 = 0 + S(m)$

- (d) If $m + x = x + m$ for all $x \in \mathbb{N}$ and $S(m) + n = n + S(m)$ for some $n \in \mathbb{N}$, then we have that $S(m) + S(n) = S(n) + S(m)$

- (a) $0 + 0 = 0 + 0$:

This is trivial by definition of addition.

- (b) $0 + n = n + 0$ implies $0 + S(n) = S(n) + 0$:

Suppose that $0 + n = n + 0$, we show that $0 + S(n) = S(n) + 0$. By the definition of addition we have that

$$0 + S(n) = S(0 + n)$$

We know by assumption that $0 + n = n + 0$. Hence

$$S(0 + n) = S(n + 0) = S(n) + 0$$

- (c) $m + 0 = 0 + m$ implies $S(m) + 0 = 0 + S(m)$:

Suppose that $m + 0 = 0 + m$, we show that $S(m) + 0 = 0 + S(m)$. By the definition of addition we have that

$$S(m) + 0 = S(m) = S(m + 0)$$

We know by assumption that $n + 0 = 0 + n$. Hence

$$S(m + 0) = S(0 + m) = 0 + S(m)$$

- (d) If $m + x = x + m$ for all $x \in \mathbb{N}$ and $S(m) + n = n + S(m)$ for some $n \in \mathbb{N}$, then we have that $S(m) + S(n) = S(n) + S(m)$:

Suppose $m + x = x + m$ for all $x \in \mathbb{N}$ and that $S(m) + n = n + S(m)$ for some $n \in \mathbb{N}$, we show that $S(m) + S(n) = S(n) + S(m)$.

We have

$$S(m) + S(n) = S(S(m) + n)$$

Now we have by assumption that $S(m) + n = n + S(m)$, for some $n \in \mathbb{N}$, hence

$$S(S(m) + n) = S(n + S(m)) = S(S(n + m))$$

Likewise a similar chain of reasoning gives

$$S(n) + S(m) = S(S(n) + m) = S(m + S(n)) = S(S(m + n))$$

Finally, we have that $m + n = m + n$ by assumption, and so $S(S(n + m)) = S(S(m + n))$

Hence by the principle of double induction we have that $a + b = b + a$ for all $a, b \in \mathbb{N}$. That is addition is commutative.

2. $ab = ba$:

We need to show that

$$(a) \ 0 * 0 = 0 * 0$$

$$(b) \ 0 * n = n * 0 \text{ implies } 0 * S(n) = S(n) * 0$$

$$(c) \ m * 0 = 0 * m \text{ implies } S(m) * 0 = 0 * S(m)$$

$$(d) \text{ If } m * x = x * m \text{ for all } x \in \mathbb{N} \text{ and } S(m) * n = n * S(m) \text{ for some } n \in \mathbb{N}, \text{ then we have that } S(m) * S(n) = S(n) * S(m)$$

$$(a) \ 0 * 0 = 0 * 0:$$

This is trivial by the definition of multiplication.

$$(b) \ 0 * n = n * 0 \text{ implies } 0 * S(n) = S(n) * 0:$$

Suppose that $0 * n = n * 0$, we show that $0 * S(n) = S(n) * 0$. We have by definition of multiplication that

$$\begin{aligned} 0 * S(n) &= 0 * n + 0 \\ &= n * 0 + 0, \text{ By assumption} \\ &= 0 + 0, \text{ By definition of multiplication} \\ &= 0, \text{ By definition of addition} \\ &= S(n) * 0, \text{ By definition of multiplication} \end{aligned}$$

$$(c) \ m * 0 = 0 * m \text{ implies } S(m) * 0 = 0 * S(m):$$

Suppose that $m * 0 = 0 * m$, we show that $S(m) * 0 = 0 * S(m)$. We have by definition of multiplication that

$$\begin{aligned} 0 * S(m) &= 0 * m + 0 \\ &= m * 0 + 0, \text{ By assumption} \\ &= 0 + 0, \text{ By definition of multiplication} \\ &= 0, \text{ By definition of addition} \\ &= S(m) * 0, \text{ By definition of multiplication} \end{aligned}$$

$$(d) \text{ If } m * x = x * m \text{ for all } x \in \mathbb{N} \text{ and } S(m) * n = n * S(m) \text{ for some } n \in \mathbb{N}, \text{ then we have that } S(m) * S(n) = S(n) * S(m):$$

Suppose that $m * x = x * m$ for all $x \in \mathbb{N}$ and $S(m) * n = n * S(m)$ for some $n \in \mathbb{N}$, we show $S(m) * S(n) = S(n) * S(m)$. By definition of multiplication we have that

$$S(m) * S(n) = S(m) * n + S(m) = n * S(m) + S(m) = n * m + n + S(m) = n * m + S(n + m)$$

Likewise, we have that

$$S(n) * S(m) = S(n) * m + S(n) = m * S(n) + S(n) = m * n + m + S(n) = m * n + S(m + n)$$

Now, we know that addition is commutative so we have that $S(m + n) = S(n + m)$, moreover by assumption we have that $n * m = m * n$. Hence

$$n * m + S(n + m) = m * n + S(m + n)$$

Hence by the principle of double induction we have that $ab = ba$ for all $a, b \in \mathbb{N}$. That is multiplication is commutative.

The result now follows. \square

We can also now deduce the following property of multiplication

1.3.2.6 Associativity of addition

This will prove that for all $a, b, c \in \mathbb{N}$ that $a + (b + c) = (a + b) + c$

Theorem 1.3.8. *Addition is associative*

For all $a, b, c \in \mathbb{N}$ we have that

$$a + (b + c) = (a + b) + c$$

Proof: We can show this by induction. Let $x, y \in \mathbb{N}$ be arbitrary, and let $P(n)$ be the proposition given by

$$(x + y) + n = x + (y + n)$$

For the base case we have $n = 0$ and so

$$\begin{aligned} (x + y) + 0 &= x + y, \text{ By definition of addition} \\ &= x + (y + 0) \end{aligned}$$

Hence $P(0)$ is true.

Now, suppose that $P(n)$ is true, that is

$$(x + y) + n = x + (y + n)$$

We show that $P(S(n))$ is also true, that is

$$(x + y) + S(n) = x + (y + S(n))$$

Now, we have that

$$\begin{aligned} (x + y) + S(n) &= S((x + y) + n), \text{ By definition of addition} \\ &= S(x + (y + n)), \text{ By the induction hypothesis} \\ &= x + (S(y + n)), \text{ By definition of addition} \\ &= x + (y + S(n)), \text{ By definition of addition} \end{aligned}$$

Hence $P(S(n))$ is true.

It follows by mathematical induction that $\forall a, b, c \in \mathbb{N}$ we have that $a + (b + c) = (a + b) + c$, that is addition is associative. \square

1.3.2.7 Multiplication distributes over addition

This will prove that for all $a, b, c \in \mathbb{N}$ we have that $a(b + c) = ab + ac$ and $(a + b)c = ac + bc$.

Theorem 1.3.9. *Multiplication distributes over addition*

For all $a, b, c \in \mathbb{N}$ we have that

1. $a(b + c) = ab + ac$
2. $(b + c)a = ba + ca = ab + ac$

Proof:

We can be quick, and solve both problems nearly simultaneously, as we have shown that multiplication is commutative.. To do this we show that for all $a, b, c \in \mathbb{N}$ we have that $a(b + c) = ab + ac$.

Let $a, b \in \mathbb{N}$ be arbitrary and we argue by induction on the proposition $P(n)$ given by

$$a(b + n) = ab + an$$

For the base case $n = 0$ we have that

$$\begin{aligned} a(b + 0) &= a(b), \text{ By definition of multiplication} \\ &= ab \\ &= ab + 0, \text{ By definition of addition} \\ &= ab + a * 0, \text{ By definition of multiplication} \end{aligned}$$

Hence $P(0)$ is true.

Now suppose that $P(n)$ is true, that is to say

$$a(b + n) = ab + an$$

We show that $P(S(n))$ is true, that is

$$a(b + S(n)) = ab + aS(n)$$

Indeed, we have that

$$\begin{aligned} a(b + S(n)) &= a(S(b + n)), \text{ By definition of addition} \\ &= a(b + n) + a, \text{ By definition of multiplication} \\ &= ab + an + a, \text{ By assumption} \\ &= ab + aS(n)0, \text{ By definition of multiplication} \end{aligned}$$

Hence $P(S(n))$ is true.

It hence follows by the principle of mathematical induction that $\forall a, b, c \in \mathbb{N}$ we have that $a(b + c) = ab + ac$.

Now, we have shown that $a(b + c) = ab + ac$, to see that $(b + c)a = ba + ca = ab + ac$ we simply observe that

$$\begin{aligned} (b + c)a &= a(b + c), \text{ Multiplication is commutative} \\ &= ab + ac, \text{ By part 1 of the theorem} \\ &= ba + ca, \text{ Multiplication is commutative} \end{aligned}$$

As required. \square

1.3.2.8 Associativity of multiplication

This will prove that for all $a, b, c \in \mathbb{N}$ that $a(bc) = (ab)c$

Theorem 1.3.10. *For all $a, b, c \in \mathbb{N}$ we have that $a(bc) = (ab)c$*

Proof:

We again show this by induction. Let $x, y \in \mathbb{N}$ be arbitrary, and let $P(n)$ be the proposition given by

$$(xy)n = x(yn)$$

For the base case we have $n = 0$ and so

$$\begin{aligned} (xy)0 &= 0, \text{ By definition of multiplication} \\ &= x(0), \text{ By definition of multiplication} \\ &= x(y * 0), \text{ By definition of multiplication} \end{aligned}$$

Hence $P(0)$ is true.

Now, suppose that $P(n)$ is true, that is

$$(xy)n = x(yn)$$

We show that $P(S(n))$ is also true, that is

$$(xy)S(n) = x(yS(n))$$

Now, we have that

$$\begin{aligned} (xy)S(n) &= (xy)n + xy, \text{ Definition of multiplication} \\ &= x(yn) + xy, \text{ By assumption} \\ &= xy + x(yn), \text{ Addition is commutative} \\ &= x(y + (yn)), \text{ Multiplication is distributive over addition} \\ &= x((yn) + y), \text{ Addition is commutative} \\ &= x(yS(n)), \text{ Addition is commutative} \end{aligned}$$

Hence $P(S(n))$ is true.

Hence, it follows by the principle of mathematical induction that for all $a, b, c \in \mathbb{N}$ we have that $a(bc) = (ab)c$. \square

1.3.2.9 The Zero and Identity laws

These two laws allow us to note that adding zero to any natural number n gives back n and multiplying n by 1 gives n .

Theorem 1.3.11. *The zero and Identity laws*

Let $n \in \mathbb{N}$. We have that

$$1. \ n + 0 = n = 0 + n$$

$$2. \ 1 * n = n = n * 1$$

Proof:

By commutativity, it is enough to only prove

1. $n + 0 = n$

2. $n * 1 = n$

1. $n + 0 = n$:

This is true by the definition of addition.

2. $n * 1 = n$:

We have by the definition of multiplication that

$$n * 1 = n * S(0) = n * 0 + n = 0 + n = n$$

Where the last equality comes from the zero law and the fact addition is commutative.

The result follows. □

1.3.2.10 The cancellation laws

These laws allow us to deduce that if $a + b = a + c$ then we must have $b = c$, and if $a \neq 0$ that $ab = ac$ gives $b = c$

Theorem 1.3.12. *The cancellation laws*

Let $a, b, c \in \mathbb{N}$. We have that

1. *If $a + b = a + c$ then we have $b = c$.*

2. *For $a \neq 0$, if $ab = ac$ then we have that $b = c$*

Proof:

1. *If $a + b = a + c$ then we have $b = c$:*

We argue by induction, let $b, c \in \mathbb{N}$ be arbitrary and let $P(n)$ be the proposition given by

$$n + b = n + c \Rightarrow b = c$$

For the base case $P(0)$ this holds trivially. Now suppose the proposition $P(n)$ holds that is

$$n + b = n + c \Rightarrow b = c$$

We show that $P(S(n))$ holds, that is

$$S(n) + b = S(n) + c \Rightarrow b = c$$

Now, we have that

$$\begin{aligned}
S(n) + b &= S(n) + c \\
S(n+0) + b &= S(n+0) + c \\
n + S(0) + b &= n + S(0) + c \\
n + (S(0) + b) &= n + (S(0) + c), \text{ By associativity} \\
(S(0) + b) &= (S(0) + c), \text{ By hypothesis, as } P(n) \text{ has } b, c \text{ being arbitrary} \\
b + S(0) &= c + S(0), \text{ By commutativity} \\
S(b+0) &= S(c+0) \\
S(b) &= S(c)
\end{aligned}$$

Hence we have $b = c$ by proposition 1.3.1. So $P(S(n))$ is true.

Hence by mathematical induction we have that if $a + b = a + c$ we must have that $b = c$.

2. For $a \neq 0$, if $ab = ac$ then we have that $b = c$:

We again argue by induction, let $b, c \in \mathbb{N}$ be arbitrary and let $P(n)$ be the proposition given by

$$nb = nc \Rightarrow b = c$$

Moreover, we do induction starting at $n = 1$ as the case $n = 0$ is vacuously true. So for $P(1)$ we have that this holds trivially. Now suppose that $P(n)$ holds. that is

$$nb = nc \Rightarrow b = c$$

We show that $P(S(n))$ is true

$$S(n)b = S(n)c \Rightarrow b = c$$

Indeed we have that

$$\begin{aligned}
S(n)b &= S(n)c \\
bS(n) &= cS(n), \text{ By commutativity} \\
bn + b &= cn + c, \text{ By commutativity} \\
a + b &= a + c, \text{ } nb = nc \text{ by assumption, so let } nb = nc = a \text{ for some } a \\
b &= c, \text{ By the cancellation law for addition}
\end{aligned}$$

Hence $P(S(n))$ is true.

Hence by mathematical induction we have that for $a \neq 0$ if $ab = ac$ we must have that $b = c$.

As required. \square .

1.3.2.11 Summation and product notation

Now that we have a well-defined notion of addition and multiplication we can define a shorthand to can be useful in avoiding writing out longer chains of additions (or multiplications) in certain situations. We will require the following mapping. Let $s \in \mathbb{N}^{n+1}$ be an ordered $n + 1$ -tuple of Natural numbers where $s = (s_0, s_1, s_1, s_2, \dots, s_n)$ and define $\mathbb{N}_n = \{0, 1, 2, 3, \dots, n\}$. Let $f : \mathbb{N}_n \rightarrow \mathbb{N}$ be a mapping defined by

$$\begin{aligned} f : \mathbb{N}_n &\rightarrow \mathbb{N} \\ i &\mapsto f(i) = s_i \end{aligned}$$

This is to say that f simply gets the value of s_i which is an element of the ordered tuple s .

Definition 1.3.18. *Summation notation*

Let $s \in \mathbb{N}^{n+1}$ be an ordered $n + 1$ -tuple of Natural numbers where $s = (s_0, s_1, s_1, s_2, \dots, s_n)$ and define $\mathbb{N}_n = \{0, 1, 2, 3, \dots, n\}$. Let $f : \mathbb{N}_n \rightarrow \mathbb{N}$ be a mapping defined by

$$\begin{aligned} f : \mathbb{N}_n &\rightarrow \mathbb{N} \\ i &\mapsto f(i) = s_i \end{aligned}$$

We define the summation notation by

$$\sum_{i=0}^n f(i) = f(0) + f(1) + f(2) + \dots + f(n)$$

This can also be written as

$$\sum_{i=0}^n s_i = s_0 + s_1 + s_2 + \dots + s_n$$

We call i the index of the summation and that $i = 0$ as the starting index of the summation for some $a \in \mathbb{N}$ and that n is the ending index of the summation. In the case that $s \in \emptyset$ then we define the summation to be 0 and call such a summation an empty sum.

We can also define the summation over a subset of \mathbb{N}_n which allows for starting the summation at a starting point other than $i = 0$. Let $T \subseteq \mathbb{N}$. We can define the summation over the set T by

$$\sum_{i \in T} s_i$$

If we have a mapping $g : \mathbb{N} \rightarrow \mathbb{N}$ for some mapping g then we can define a summation over g by

$$\sum_{i \in T} g(s_i)$$

Finally, we can define a summation over a predicate $P(i)$ for $i \in T$ giving

$$\sum_{P(i)} g(s_i)$$

which means to take the sum of the $g(s_i)$ where i satisfies the predicate P . If the predicate is not satisfied by any i then the summation is also said to be an empty summation and given a value of 0.

In light of definition a summation of a predicate we have that if $a > n$ where a is the index lower of summation and n the upper point of summation then the sum would be by definition equal to 0. That is to say

$$\sum_{i=a}^n s_i = 0, \text{ If } a > n$$

Example 1.3.6. Let $s = (2, 3, 4, 8) \in \mathbb{N}^4$ then we have that

$$\sum_{i=0}^3 s_i = 2 + 3 + 4 + 8 = 17$$

Example 1.3.7. Let $g(n) = n$ and let $k = 4$ then we have that

$$\sum_{i=0}^4 -1g(i) = \sum_{i=0}^3 i = 1 + 2 + 3 + 4 = 10$$

Example 1.3.8. Let $s_1 \in \mathbb{N}$ then we have

$$\sum_{i=1}^1 s_1 = s_1$$

Example 1.3.9. Let $g(n) = n * n$ and let $T = \{2, 6, 11\} \subseteq \mathbb{N}^{11}$ then

$$\sum_{i \in T} g(i) = g(2) + g(6) + g(11) = 2 * 2 + 6 * 6 + 11 * 11 = 4 + 36 + 121 = 161$$

Example 1.3.10. Let $g(n) = n$, let $P(n)$ be the predicate such that

$$P(n) = \begin{cases} 1, & \text{If } n = 2, 4, 6 \\ 0, & \text{Otherwise} \end{cases}$$

Let $T = \{2, 6, 11\} \subseteq \mathbb{N}^{11}$ then we have for the $i \in T$ that satisfies $P(i)$ is given by

$$\sum_{P(i)} i = 2 + 4 = 6$$

Example 1.3.11. Let $f(n) = n + 5$. Consider the sum

$$\sum_{i=3}^6 n + 5 = (3 + 5) + (4 + 5) + (5 + 5) + (6 + 5) = 8 + 9 + 10 + 11 = 38$$

We can re-express this sum as

$$\sum_{i=0}^3 n + 5 = ((0 + 3) + 5) + ((1 + 3) + 5) + ((2 + 3) + 5) + ((3 + 3) + 5) = 38$$

We have re-indexed the sum into an equivalent form.

We can make some observations about summation notation.

Proposition 1.3.2. *Properties of summation notation*

Let $n, m \in \mathbb{N}$ such that $m < n$. Let $s, t \in \mathbb{N}^n$ and let $c \in \mathbb{N}$. In addition define $A = \mathbb{N}_m$ and $B = \mathbb{N}_n \setminus A = \{m + 1, m + 2, \dots, n\}$ so that $A \cup B = \mathbb{N}_n$. Let $a \in \mathbb{N}$ be the lower index summation. We have that the following properties hold.

1. $\sum_{i=0}^n s_i = \sum_{i \in A} s_i + \sum_{i \in B} s_i = \sum_{i=0}^m s_i + \sum_{i=m+1}^n s_i$
2. $\sum_{i=a}^n s_i = \sum_{i=a}^m s_i + \sum_{i=m+1}^n s_i$

$$3. \sum_{i=1}^n c = c * n$$

$$4. \sum_{i=1}^n c * s_i = c * \sum_{i=1}^n s_i$$

$$5. \sum_{i=1}^n s_i + t_i = \sum_{i=1}^n s_i + \sum_{i=1}^n t_i$$

Proof:

$$1. \sum_{i=0}^n s_i = \sum_{i \in A} s_i + \sum_{i \in B} s_i = \sum_{i=0}^m s_i + \sum_{i=m+1}^n s_i.$$

We argue by induction on n . Let $P(n)$ be the proposition given by

$$\sum_{i=1}^n s_i = \sum_{i \in A} s_i + \sum_{i \in B} s_i = \sum_{i=1}^m s_i + \sum_{i=m+1}^n s_i$$

The base case $P(0)$ we have that $A = \emptyset$ and $B = \mathbb{N}_0 \setminus A = \{0\}$ as we have by assumption that $m < n$. Hence

$$\sum_{i=0}^0 s_i = s_0$$

Likewise we have

$$\sum_{i \in A} s_i + \sum_{i \in B} s_i = 0 + \sum_{i=0}^0 s_i = s_0$$

So the base case holds. Now suppose that the $P(n)$ that is

$$\sum_{i=1}^n s_i = \sum_{i \in A} s_i + \sum_{i \in B} s_i = \sum_{i=1}^m s_i + \sum_{i=m+1}^n s_i$$

we need to show that

$$\sum_{i=1}^{n+1} s_i = \sum_{i=1}^m s_i + \sum_{i=m+1}^{n+1} s_i$$

also holds. By definition we have that

$$\sum_{i=1}^{n+1} s_i = s_0 + s_1 + s_2 + \cdots + s_n + s_{n+1} = \sum_{i=0}^n s_i + s_{n+1}$$

Now we have that

$$\begin{aligned}
\sum_{i=1}^{n+1} s_i &= \sum_{i=0}^n s_i + s_{n+1} \\
&= \sum_{i=1}^m s_i + \sum_{i=m+1}^n s_i + s_{n+1}, \text{ By the induction hypothesis} \\
&= \sum_{i=1}^m s_i + \sum_{i=m+1}^{n+1} s_i, \text{ By definition}
\end{aligned}$$

Hence $P(n+1)$ holds and the result follows by induction.

$$2. \sum_{i=a}^n s_i = \sum_{i=a}^m s_i + \sum_{i=m+1}^n s_i :$$

This follows by a similar argument as 1. but starting the induction at a .

$$3. \sum_{i=1}^n c = c * n:$$

We argue by induction on n . Let $P(n)$ be the proposition given by

$$\sum_{i=1}^n c = c * n$$

For the base case $P(1)$

$$\sum_{i=1}^1 c = c = c * 1$$

Now suppose that $P(n)$ holds we need to show that $P(n+1)$ holds, that is

$$\sum_{i=1}^{n+1} nc = c * (n+1)$$

We have that

$$\sum_{i=1}^{n+1} c = \sum_{i=1}^n c + c = n * c + c = c * n + c = c * S(n) = c * (n+1)$$

The result follows by induction.

$$4. \sum_{i=1}^n c * s_i = c * \sum_{i=1}^n s_i:$$

We have by definition of summation that

$$\sum_{i=1}^n c * s_i = c * s_1 + c * s_2 + \cdots + c * s_n$$

Now as multiplication distributes over addition we have

$$\sum_{i=1}^n c * s_i = c * s_1 + c * s_2 + \cdots + c * s_n = c(s_1 + s_2 + \cdots + s_n) = c * \sum_{i=1}^n s_i$$

$$5. \sum_{i=1}^n s_i + t_i = \sum_{i=1}^n s_i + \sum_{i=1}^n t_i:$$

We argue by induction. Let $P(n)$ denote be the proposition given by

$$\sum_{i=1}^n s_i + t_i = \sum_{i=1}^n s_i + \sum_{i=1}^n t_i$$

For the base case $P(1)$ we have that

$$\sum_{i=1}^1 s_i + t_i = s_1 + t_1$$

Likewise we have

$$\sum_{i=1}^1 s_i + \sum_{i=1}^n t_i = s_1 + t_1$$

So the base case holds. Now suppose $P(n)$ holds so we need to show $P(n+1)$ holds. By definition we have

$$\sum_{i=1}^{n+1} s_i + \sum_{i=1}^n s_i + t_i + s_{n+1} + t_{n+1} = \sum_{i=1}^n s_i + \sum_{i=1}^n t_i + s_{n+1} + t_{n+1}$$

By the induction hypothesis. Now addition is commutative so we get

$$\sum_{i=1}^{n+1} s_i + t_i = \sum_{i=1}^n s_i + \sum_{i=1}^n t_i + s_{n+1} + t_{n+1} = \sum_{i=1}^n s_i + s_{n+1} + \sum_{i=1}^n t_i + t_{n+1} = \sum_{i=1}^{n+1} s_i + \sum_{i=1}^{n+1} t_i$$

The result follows by induction.

□

The summation notation allows us to deduce an additional property of multiplication .

Proposition 1.3.3. *Product of two naturals being zero implies one of the numbers is zero*

Let $a, b \in \mathbb{N}$. If $ab = 0$ then at least one of a or b is zero.

Proof:

Let $a, b \in \mathbb{N}$ and let $ab = 0$. Using the summation notation we have that

$$ab = \sum_{i=1}^b a = \underbrace{a + a + a + \cdots + a}_{b \text{ times}} = 0$$

From which we can see that this holds for $a = 0$ and only $a = 0$. Suppose that $a \neq 0$ then

$$\sum_{i=1}^b a = \underbrace{a + a + a + \cdots + a}_{b \text{ times}} > 0$$

A contradiction to the hypothesis.

A similar result holds for $ab = \sum_{i=1}^a b$. Finally if both a and b are zero the result is trivial.

The result has been shown. □.

A similar definition can be made for multiplication, called product notation

Definition 1.3.19. *Product notation*

Let $s \in \mathbb{N}^{n+1}$ be an ordered $n + 1$ -tuple of Natural numbers where $s = (s_0, s_1, s_1, s_2, \dots, s_n)$ and define $\mathbb{N}_n = \{0, 1, 2, 3, \dots, n\}$. Let $f : \mathbb{N}_n \rightarrow \mathbb{N}$ be a mapping defined by

$$\begin{aligned} f : \mathbb{N}_n &\rightarrow \mathbb{N} \\ i &\mapsto f(i) = s_i \end{aligned}$$

We define the product notation by

$$\prod_{i=0}^n f(i) = f(0) * f(1) * f(2) * \dots * f(n)$$

This can also be written as

$$\prod_{i=0}^n s_i = s_0 * s_1 * s_2 * \dots * s_n$$

We call i the index of the product and that $i = 0$ as the lower starting point of the product for some $a \in \mathbb{N}$ and that n is the upper point of the product. In the case that $s \in \emptyset$ then we define the product to be 1 and call such a product an empty product.

We can also define the product over a subset of \mathbb{N}_n which allows for starting the product at a starting point other than $i = 0$. Let $T \subseteq \mathbb{N}$. We can define the product over the set T by

$$\prod_{i \in T} s_i$$

If we have a mapping $g : \mathbb{N} \rightarrow \mathbb{N}$ for some mapping g then we can define a product over g by

$$\prod_{i \in T} g(s_i)$$

Finally, we can define a product over a predicate $P(i)$ for $i \in T$ giving

$$\sum_{P(i)} g(s_i)$$

which means to take the product of the $g(s_i)$ where i satisfies the predicate P . If the predicate is not satisfied by any i then the product is also said to be an empty product and given a value of 1. In light of definition a product of a predicate we have that if $a > n$ where a is the lower index of the product and n the upper point of product then the product would be by definition equal to 1. That is to say

$$\sum_{i=a}^n s_i = 1, \text{ If } a > n$$

Example 1.3.12. Let $s = (2, 3, 4, 8) \in \mathbb{N}^4$ then we have that

$$\prod_{i=0}^3 s_i = 2 * 3 * 4 * 8 = 192$$

Example 1.3.13. Let $g(n) = n$ and let $k = 4$ then we have that

$$\prod_{i=0}^{4-1} g(i) = \prod_{i=0}^3 i = 1 * 2 * 3 * 4 = 24$$

Example 1.3.14. Let $s_1 \in \mathbb{N}$ then we have

$$\prod_{i=1}^1 s_1 = s_1$$

Example 1.3.15. Let $g(n) = n * n$ and let $T = \{2, 6, 11\} \subseteq \mathbb{N}^{11}$ then

$$\prod_{i \in T} g(i) = g(2) * g(6) * g(11) = (2 * 2) + (6 * 6) + (11 * 11) = 4 * 36 * 121 = 17424$$

Example 1.3.16. Let $g(n) = n$, let $P(n)$ be the predicate such that

$$P(n) = \begin{cases} 1, & \text{If } n = 2, 4, 6 \\ 0, & \text{Otherwise} \end{cases}$$

Let $T = \{2, 6, 11\} \subseteq \mathbb{N}^{11}$ then we have for the $i \in T$ that satisfies $P(i)$ is given by

$$\sum_{P(i)} i = 2 * 4 = 12$$

There is an some immediate properties of product notation that are clear

Proposition 1.3.4. *Properties of product notation*

Let $n, m \in \mathbb{N}$ such that $m < n$. Let $s, t \in \mathbb{N}^n$ and let $c \in \mathbb{N}$. In addition define $A = \mathbb{N}_m$ and $B = \mathbb{N}_n \setminus A = \{m+1, m+2, \dots, n\}$ so that $A \cup B = \mathbb{N}_n$. Let $a \in \mathbb{N}$ be the lower index summation. We have that the following properties hold.

1. $\prod_{i=0}^n s_i = \prod_{i \in A} s_i * \prod_{i \in B} s_i = \prod_{i=0}^m s_i + \prod_{i=m+1}^n s_i$
2. $\prod_{i=a}^n s_i = \prod_{i=a}^m s_i * \prod_{i=m+1}^n s_i$
3. $\prod_{i=1}^n s_i t_i = \prod_{i=1}^n s_i \prod_{i=1}^n t_i$

Proof:

1. $\prod_{i=0}^n s_i = \prod_{i \in A} s_i * \prod_{i \in B} s_i = \prod_{i=0}^m s_i + \prod_{i=m+1}^n s_i$

We argue by induction on n . Let $P(n)$ be the proposition given by

$$\prod_{i=0}^n s_i = \prod_{i=0}^n s_i = \prod_{i \in A} s_i * \prod_{i \in B} s_i = \prod_{i=0}^m s_i + \prod_{i=m+1}^n s_i$$

The base case $P(0)$ we have that $A = \emptyset$ and $B = \mathbb{N}_0 \setminus A = \{0\}$ as we have by assumption that $m < n$. Hence

$$\prod_{i=0}^0 s_i = s_0$$

Likewise we have

$$\prod_{i \in A} s_i + \prod_{i \in B} s_i = 0 + \prod_{i=0}^0 s_i = s_0$$

So the base case holds. Now suppose that the $P(n)$ that is

$$\prod_{i=1}^n s_i = \prod_{i=1}^m s_i + \prod_{i=m+1}^n s_i$$

we need to show that

$$\prod_{i=1}^{n+1} s_i = \prod_{i=1}^m s_i + \prod_{i=m+1}^{n+1} s_i$$

also holds. By definition we have that

$$\prod_{i=1}^{n+1} s_i = s_0 * s_1 * s_2 * \cdots * s_n * s_{n+1} = \prod_{i=0}^n s_i * s_{n+1}$$

Now we have that

$$\begin{aligned} \prod_{i=1}^{n+1} s_i &= \prod_{i=0}^n s_i * s_{n+1} \\ &= \prod_{i=1}^m s_i * \prod_{i=m+1}^n s_i * s_{n+1}, \text{ By the induction hypothesis} \\ &= \prod_{i=1}^m s_i * \prod_{i=m+1}^{n+1} s_i, \text{ By definition} \end{aligned}$$

Hence $P(n+1)$ holds and the results follows by induction.

$$2. \prod_{i=a}^n s_i = \prod_{i=a}^m s_i * \prod_{i=m+1}^n s_i.$$

A similar argument as in part 1 shows this.

$$3. \prod_{i=1}^n s_i t_i = \prod_{i=1}^n s_i \prod_{i=1}^n t_i:$$

We argue by induction. Let $P(n)$ denote the proposition

$$\prod_{i=1}^n s_i t_i = \prod_{i=1}^n s_i \prod_{i=1}^n t_i$$

In the base case $P(1)$ we have

$$\prod_{i=1}^1 s_i t_i = s_1 t_1$$

Likewise

$$\prod_{i=1}^1 s_i \prod_{i=1}^1 t_i = s_1 * t_1 = s_1 t_1$$

Which shows the base case. Now suppose $P(n)$ is true, we show that $P(n+1)$ is true. We have that

$$\begin{aligned} \prod_{i=1}^{n+1} s_i t_i &= \prod_{i=1}^n s_i t_i * s_{n+1} t_{n+1} \\ &= \prod_{i=1}^n s_i \prod_{i=1}^n t_i * s_{n+1} * t_{n+1} \\ &= \prod_{i=1}^n s_i * s_{n+1} * \prod_{i=1}^n t_i * t_{n+1} \\ &= \prod_{i=1}^{n+1} s_i \prod_{i=1}^{n+1} t_i \end{aligned}$$

The result follows by induction.

□

1.3.2.12 Exponentiation

With the product notation defined we can define another operation called exponentiation

Definition 1.3.20. *Exponentiation of Natural numbers*

Let $(m, n) \in \mathbb{N} \times \mathbb{N}$ and let $\wedge : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$. We define the exponentiation of m by n to be m multiplied by itself $n - 1$ times

$$\wedge : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$$

$$(m, n) \mapsto \wedge(m, n) = \begin{cases} 1, & \text{If } n = 0 \text{ and } m = 0 \\ 1, & \text{If } n = 0 \\ \wedge(m, n) = \prod_{i=1}^n m = 1 * \prod_{i=1}^n m, & n \neq 0 \end{cases}$$

We will write $\wedge(m, n)$ as m^n . We say that m is the base and n is the exponent. We sometimes say that m has been raised to the power of n . In the case that $n = 0$ and $m = 0$ we have a vacuous product and so an empty product which by definition has a value of 1.

With the above definition, we make a quick remark. We know that an empty product has a value of 1 and as multiplication by 1 doesn't change the value we can write exponentiation as

$$\prod_{i=1}^n m = 1 * \prod_{i=1}^n m$$

This makes it clear that exponentiation is multiplication of 1 by n copies of m .

Example 1.3.17. Let $n = 2$ and $m = 2$ then we have that $2 = S(1)$ then

$$\wedge(2, 2) = \prod_{i=1}^2 2 = 2 * 2 = 4$$

Example 1.3.18. Let $m = 4$ and $n = 1$ then we have that

$$\wedge(4, 1) = 4^1 = 4$$

Example 1.3.19. Let $m = 5$ and $n = 0$ then we have that

$$\wedge(5, 0) = 5^0 = 1$$

Example 1.3.20. Let $m = 2$ and $n = 7$ then we have that

$$\wedge(5, 0) = 5^0 = 1$$

As we have defined a new operation we should check that the operation is meaningful

Theorem 1.3.13. *Exponentiation is closed*

For all $n, m \in \mathbb{N}$ we have that

$$\wedge(n, m) \in \mathbb{N}$$

Proof:

There are two cases to consider $m = 0$ and $m \neq 0$. When $m = 0$ the operation is defined such that

$$\wedge(n, 0) = 1$$

which is in \mathbb{N} . When $m \neq 0$ then $\wedge(n, m) \in \mathbb{N}$ as multiplication in \mathbb{N} is closed. \square

We should also verify that the other properties that we have verified for addition and multiplication either hold or do not. For example we can find examples that show that exponentiation is not commutative.

Proposition 1.3.5. *Exponentiation is non-commutative*

There exist $n, m \in \mathbb{N}$ such that

$$\wedge(n, m) \neq \wedge(m, n)$$

Proof:

Let $n = 3$ and $m = 4$ then we have that

$$\wedge(3, 4) = 81$$

$$\wedge(4, 3) = 64$$

from which it is clear that $81 \neq 64$. \square

Proposition 1.3.6. *Exponentiation is non-associative*

There exist $a, b, c \in \mathbb{N}$ such that

$$(a^b)^c \neq a^{(b^c)}$$

Proof:

Let $a = 2, b = 3$ and $c = 4$ then we have that

$$(2^3)^4 = 8^3 = 4096$$

$$2^{(3^4)} = 2^{81} = 2417851639229258349412352$$

Clearly $4096 \neq 2417851639229258349412352$. \square

The non-associativity of exponentiation shows an important point, that the order in which we do exponentiation can give drastically different result and so the order in which the exponentiation should be done will depend on the context. We will bracket for each case as required. There is an interesting property for the case $(a^b)^c$ called the power law of exponentiation.

Proposition 1.3.7. Power law of exponentiation

Let $a, b, c \in \mathbb{N}$. We have that

$$(a^b)^c = a^{bc}$$

Proof:

By definition of exponentiation we have that

$$(a^b)^c = \prod_{i=1}^c a^b = \prod_{i=1}^c \left(\prod_{j=1}^b a \right)$$

That is we are multiplying $\prod_{j=1}^b a$ by itself c times. The product $\prod_{j=1}^b a$ itself is the multiplication of a by itself b times. We can therefore express the above by

$$\begin{aligned} (a^b)^c &= \underbrace{\prod_{j=1}^b a * \prod_{j=1}^b a * \cdots * \prod_{j=1}^b a}_{c \text{ times}} \\ &= \underbrace{(a * a * \cdots * a) * (a * a * \cdots * a) * \cdots * (a * a * \cdots * a)}_{c \text{ times}} \\ &\quad \underbrace{\qquad \qquad \qquad}_{b \text{ times}} \quad \underbrace{\qquad \qquad \qquad}_{b \text{ times}} \quad \underbrace{\qquad \qquad \qquad}_{b \text{ times}} \end{aligned}$$

There are therefore $b * c$ multiplications of a with itself, as we need to perform c iterations of $\prod_{j=1}^b a$. Hence we have that

$$(a^b)^c = \underbrace{a * a * a * \cdots * a}_{b*c \text{ times}} = \prod_{i=1}^{bc} a = a^{bc}$$

As required. \square

There are some additional properties that we can deduce. Consider 2^m for $m \in \mathbb{N}$ we have for $m = 0, 1, 2, 3$ and 4 that $2^0 = 1, 2^1 = 2, 2^2 = 4, 2^3 = 8$ and $2^4 = 16$. Notice that multiplying any 2^m by 2 adds one to the power. In fact multiplying any 2^m by 4 adds to the power. It looks like the powers multiply together. For example $2^m * 2^n = 2^{m+n}$. We can show this is true for bases other than 2.

Proposition 1.3.8. Multiplying exponents of same base adds the powers

Let $a, m, n \in \mathbb{N}$. We have that

$$a^n * a^m = a^{n+m}$$

Proof:

Let $a, n, m \in \mathbb{N}$. If $n = 0$ and $m \geq 0$ then $a^n = 1$ and we have that $a^n * a^m = a^{n+m} = 1 * a^m = a^{0+m} = a^m$. Likewise for the case $m = 0$ and $n \geq 0$. So suppose that $m > 0$ and $m > 0$. We have by definition of exponentiation that

$$a^n * a^m = \prod_{i=1}^n a * \prod_{i=1}^m a = \underbrace{a * a * \cdots * a}_n * \underbrace{a * a * \cdots * a}_m = \underbrace{a * a * \cdots * a}_{n+m} = a^{n+m}$$

as required. \square

We also have the following result that combines multiplying two numbers and raising that result to a power. As an example consider $(2 * 3)^2 = 6^2 = 36$. Now consider $2^2 = 4$ and $3^2 = 9$ and we clearly have $4 * 9 = 36$. The powers can come through to each of the numbers of the multiplication.

Proposition 1.3.9. *Power of product is product of powers*

Let $a, b, n \in \mathbb{N}$. We have that

$$(a * b)^n = a^n * b^n$$

Proof:

If $n = 0$ then $(a * b)^0 = 1$ by definition and $a^0 * b^0 = 1$. So suppose that $n > 0$ then we have that

$$\begin{aligned}
(a * b)^n &= \prod_{i=1}^n ab = \underbrace{ab * ab * ab \cdots * ab}_{n \text{ times}} \\
&= \left(\underbrace{a * a * a \cdots * a}_{n \text{ times}} \right) * \left(\underbrace{b * b * b \cdots * b}_{n \text{ times}} \right), \text{ By commutativity of multiplication} \\
&= a^n * b^n
\end{aligned}$$

The proposition has been shown. \square

1.3.2.13 Subtraction

We can define an operation that will allow us to at least partially undo addition. To define this operation we need to make use of the less than operator.

Definition 1.3.21. *Subtraction of natural number*

Let $n, m \in \mathbb{N}$ such that $m \leq n$. Let $d \in \mathbb{N}$ such that $n = m + d$. We define subtraction by

$$d = n - m$$

We call d the difference between n and m .

There is an immediate result from the definition of subtraction

Proposition 1.3.10. $a + (b - c) = (a + b) - c$

Let $a, b, c \in \mathbb{N}$ with $b \geq c$. We have that

$$a + (b - c) = (a + b) - c$$

Proof:

We argue by induction. Let $P(n)$ denote the proposition

$$a + (n - c) = (a + n) - c$$

For the base case $n = 0$ we have by definition $c = 0$ and so

$$a + (0 - 0) = a = (a + 0) - 0$$

Now suppose that $P(n)$ holds, we show that $P(n + 1)$ is true that is

$$a + ((n + 1) - c) = (a + (n + 1)) - c$$

We have that $n + 1 = (n + 0) + 1 = n + (0 + 1)$ and so

$$\begin{aligned}
a + ((n + 1) - c) &= a + (n + (0 + 1) - c) \\
&= a + (n + (1 - c)) \\
&= (a + n) + 1 - c \\
&= a + (n + 1) - c
\end{aligned}$$

As required. \square

We immediately see that subtraction is not commutative that is $a - b \neq b - a$ in fact it is not even defined for $b - a$ unless $b \geq a$ but then it is not defined for $a - b$ and visa-versa. Likewise it is not associative as for example $(8 - 4) - 2 = 2$ but $8 - (4 - 2) = 6$. We do however retain the fact that multiplication is commutative over subtraction

Proposition 1.3.11. *Multiplication distributes over subtraction*

Let $a, b, c \in \mathbb{N}$ with $b \geq c$ and let $a \in \mathbb{N}$. We have that

1. $a(b - c) = ab - ac$
2. $(b - c)a = ba - ca = ab - ac$

Proof:

1. $a(b - c) = ab - ac$:

Let $a \in \mathbb{N}$ be arbitrary. We argue by induction of the proposition $P(n)$ given by

$$a(n - m) = an - am$$

where by definition $m \leq n$. For the base case we have $P(0)$ we have that $n = m = 0$ and so

$$a(0 - 0) = a * 0 = 0 = a * 0 - a * 0$$

Showing the base case. Now suppose that $P(n)$ holds we show that $P(n + 1)$ is true, that is we show

$$a((n + 1) - m) = a(n + 1) - am$$

where $m \leq (n + 1)$. There are two cases to consider if $m = n + 1$ then we have

$$a((n + 1) - m) = a * 0 = 0 = a(n - 1) - am$$

Now suppose that $m < (n + 1)$ then

$$a((n + 1) - m) = a(n + 1) - am$$

by the induction hypothesis. The result follows by induction.

2. $(b - c)a = ba - ca = ab - ac$:

As multiplication is commutative we have that

$$\begin{aligned} (b - c)a &= a(b - c) \\ &= ab - ac \\ &= ba - ca \end{aligned}$$

The result follows. \square

1.3.2.14 The principle of strong induction

The final property of the natural we shall look at is that of the principle of strong induction, although as we will see, this is actually equivalent to usual induction. There is one more version of induction that is sometimes useful, this is the so-called principle of strong induction, this is instead of assuming $P(n)$ is true and showing that $P(n+1)$. We instead assume that for all $n \leq k$ for some $k \in \mathbb{N}$ we have that $P(n)$ is true for all $n \leq k$ and we show that this implies that $P(k+1)$ is true.

Theorem 1.3.14. *The principle of strong induction*

Let $P(n)$ be a proposition about a natural number $n \in \mathbb{N}$. Moreover, suppose that

1. $P(0)$ is true.
2. $\forall k \in \mathbb{N} : P(0) \wedge P(1) \wedge P(2) \wedge \cdots \wedge P(k)$ all being true implies that $P(k+1)$ is true.

If these two statements are true, we have that $P(n)$ is true for any natural number n , and we say the proposition $P(n)$ holds by the principle of strong mathematical induction.

Proof:

Define $\tilde{P}(n)$ to be the following proposition

$$\tilde{P}(n) = P(0) \wedge P(1) \wedge P(2) \wedge \cdots \wedge P(n)$$

We show that $\tilde{P}(n)$ for all $n \geq 0$. By assumption $\tilde{P}(n)$ is true as $\tilde{P}(n) = P(0)$. Now suppose that $\tilde{P}(n)$ is true for some $n \in \mathbb{N}$, that is

$$\tilde{P}(n) = P(0) \wedge P(1) \wedge P(2) \wedge \cdots \wedge P(n)$$

is true, we show that $\tilde{P}(n+1)$ is true, that is

$$\tilde{P}(n+1) = P(0) \wedge P(1) \wedge P(2) \wedge \cdots \wedge P(n) \wedge P(n+1)$$

By assumption 2. as we have that $\forall n \in \mathbb{N} : P(0) \wedge P(1) \wedge P(2) \wedge \cdots \wedge P(n)$ implies that $P(n+1)$ is true. Hence we have that

$$\tilde{P}(n+1) = \tilde{P}(n) \wedge P(n+1) = \tilde{P}(n+1)$$

is true.

Hence by the principle of mathematical induction we have that $\tilde{P}(n)$ is true for all $n \geq 0$. \square

As mentioned earlier, we said that strong induction and the usual induction are equivalent, we shall prove this. We used induction to prove strong induction so it is left to show that given the assumptions for strong induction, we can deduce the truth $\forall n \in \mathbb{N}$ of the proposition $P(n)$ only using induction.

Theorem 1.3.15. *Strong induction is equivalent to the usual induction*

Suppose that the assumptions of strong induction hold. That is suppose $P(n)$ be a proposition about a natural number $n \in \mathbb{N}$ and moreover suppose that

1. $P(0)$ is true.
2. $\forall k \in \mathbb{N} : P(0) \wedge P(1) \wedge P(2) \wedge \cdots \wedge P(k)$ all being true implies that $P(k+1)$ is true.

We have that the truth of $P(n)$ for all $n \in \mathbb{N}$ can be deduced using only regular induction.

Proof:

Let $\tilde{P}(n)$ be the proposition be given by

$$\forall k \leq n \text{ we have } P(k) \text{ is true}$$

We show by the principle of induction that

1. $\tilde{P}(0)$ is true
2. $\tilde{P}(n)$ being true implies $\tilde{P}(n+1)$ is true for any natural number n .

1. $\tilde{P}(0)$ is true:

To see this, we have that $\tilde{P}(0)$ is given by

$$\forall k \leq 0 \text{ we have } P(0) \text{ is true}$$

This clearly holds as the only natural number that is less than or equal to zero is zero. Hence $P(0)$ is true and so $\tilde{P}(0)$.

2. $\tilde{P}(n)$ being true implies $\tilde{P}(n+1)$ is true for any Natural number n :

Suppose that $\tilde{P}(n)$ is true, that is

$$\forall k \leq n \text{ we have } P(k) \text{ is true}$$

we show that $\tilde{P}(n+1)$ is true, that is

$$\forall k \leq n+1 \text{ we have } P(k) \text{ is true}$$

Let $k \leq n+1$ be a natural number, have two cases to consider.

- (a) If $k < n+1$ then we must have that $k \leq n$. Now, we know that $\tilde{P}(n)$ is true by assumption, moreover by the assumptions of strong induction holding true we can conclude that $P(0) \wedge P(1) \wedge P(2) \wedge \dots \wedge P(n)$ all being true gives us $P(n+1)$ is true. Hence we can conclude that $\tilde{P}(n+1)$ holds.
- (b) Now, the remaining case is $k = n+1$. In this case $\tilde{P}(n+1)$ is the statement

$$\forall k \leq n+1 \text{ we have } P(k) \text{ is true}$$

Now, we have that $\tilde{P}(n+1) = \tilde{P}(n) \wedge P(n+1)$, from which we can assume the truth of $\tilde{P}(n)$ by assumption and by hypothesis this allows us to deduce the truth of $P(n+1)$. This gives us the truth of $\tilde{P}(n+1)$.

Hence, in both cases we conclude the truth of $\tilde{P}(n+1)$.

Hence the proposition follows by mathematical induction. Which is to say, strong induction can be proven using regular induction. \square

We have now shown the equivalence of regular and strong induction.

1.3.2.15 The well-ordering principle

Consider the way we constructed the natural numbers, we started with one element $0 = \emptyset$, and build each element in turn by the successor function. That is

$$\begin{aligned} 1 &= S(0) = 0 \cup \{0\} \\ 2 &= S(1) = 1 \cup \{1\} \\ 3 &= S(2) = 2 \cup \{2\} \end{aligned}$$

This is clearly constructing some form of ordering on \mathbb{N} , in particular we can consider this in two different ways. Firstly we can see the successor map under set inclusion, that is

$$0 \subset 1 \subset 2 \subset 3 \subset 4 \subset 5 \subset \dots$$

likewise we can consider this ordering in the more intuitive sense of the less than or equal to operator.

$$0 \leq 1 \leq 2 \leq 3 \leq 4 \leq 5 \leq \dots$$

This just doesn't hold for the entirety of \mathbb{N} . For example consider the set $S = \{2, 4, 6, 8\}$, we have from the successor mapping that $2 \in 4 \in 6 \in 8$, hence 2 is the smallest element of S , with respect to the inclusion of sets. We phrase this in the following proposition

Theorem 1.3.16. *Well-ordering principle*

Let $S \subseteq \mathbb{N}$ be a subset of \mathbb{N} with the possibility of being the entirety of \mathbb{N} . We have that $\exists x \in S$ such that x is the smallest element of S with respect to set inclusion. This is to say $\exists x \in S$ such that $\forall y \in S$ we have $x \subseteq y$.

Proof:

As 0 is by construction included in every natural number it is enough to show that any subset of $\mathbb{N} \setminus \{0\}$ has no minimal element with respect to set inclusion. For this purpose we will define $M = \mathbb{N} \setminus \{0\}$

Let $S \subseteq M$ that has no smallest element with respect to set inclusion. We argue by strong induction on S

By assumption S has no smallest element with respect to inclusion then $1 \notin S$ otherwise it would be by definition the smallest element with respect to inclusion. Define T to be the complement of S and then we $0 \in T$.

Now suppose that every $n \in M$ such that $k \leq n$ is in T . If $n+1 \in S$ then it would be a minimal element as every element less than $n+1$ is in the complement of S , hence $n+1 \in T$. This implies that every element of M is in T by strong induction.

It follows that $S = \emptyset$. Hence the result. \square

We have shown in some sense that \mathbb{N} is well-ordered. We will see that the idea of well-ordering is an example of a so-called relation.

1.3.2.16 Rules for the inequality operators

Now that we have a firm grasp of the natural numbers we can deduce some properties that relate to inequalities. In the natural numbers, there are a few results which can be deduced.

Proposition 1.3.12. *Properties of inequalities for natural numbers*

Let $a, b, c, d \in \mathbb{N}$. We have the following properties for inequalities

1. $a \leq b$ is the same as $b \geq a$
2. $a < b$ is the same as $b > a$
3. If $a \leq b$ and $b \leq c$ then $a \leq c$
4. If $a < b$ and $b \leq c$ then $a < c$
5. If $a \leq b$ and $b < c$ then $a < c$
6. If $a < b$ and $b < c$ then $a < c$
7. If $a \geq b$ and $b \geq c$ then $a \geq c$
8. If $a > b$ and $b \geq c$ then $a > c$
9. If $a \geq b$ and $b > c$ then $a > c$
10. If $a > b$ and $b > c$ then $a > c$
11. If $a \leq b$ then $a + c \leq b + c$

12. If $a < b$ then $a + c < b + c$
13. If $a \geq b$ then $a + c \geq b + c$
14. If $a > b$ then $a + c > b + c$
15. If $a \leq b$ then $ac \leq bc$
16. If $a < b$ then $ac < bc$
17. If $a \geq b$ then $ac \geq bc$
18. If $a > b$ then $ac > bc$

Proof:

1. $a \leq b$ is the same as $b \geq a$:

Suppose that $a \leq b$ then by definition of $a \leq b$ we have that $a \subseteq b$. We then clearly have that $b \not\subseteq a$ and so either $b > a$ by definition or $b = a$. In other words $b \geq a$.

2. $a < b$ is the same as $b > a$:

Similar to the first part. If $a < b$ then by definition a is a strict subset of b , that is $a \subset b$. If a is a strict subset of b then $b \not\subseteq a$ by definition of a subset. Hence $b > a$ by definition of greater than.

3. If $a \leq b$ and $b \leq c$ then $a \leq c$:

Suppose that $a \leq b$ and $b \leq c$. By definition, we have that $a \subseteq b$ and $b \subseteq c$ and so by proposition 1.2.2 we have $a \subseteq c$ which is to say $a \leq c$.

4. If $a < b$ and $b \leq c$ then $a < c$:

As $a < b$ and $b \leq c$ then $a \subset b$ and $b \subseteq c$. Applying proposition 1.2.3 gives $a \subset c$ and so $a < c$

5. If $a \leq b$ and $b < c$ then $a < c$:

Similar to part 4. As $a \leq b$ then $a \subseteq b$ and likewise as $b < c$ then $b \subset c$. Applying 1.2.3 gives $a \subset c$ and hence $a < c$.

6. If $a < b$ and $b < c$ then $a < c$:

Similar to part 4. and 5. As $a < b$ then $a \subset b$ and likewise as $b < c$ then $b \subset c$. By proposition 1.2.4 we have that $a \subset c$ and hence $a < c$.

7. If $a \geq b$ and $b \geq c$ then $a \geq c$:

By the first part of the proposition we have that $a \geq b$ and $b \geq c$ then $a \geq c$ is the same as $b \leq a$ and $c \leq b$ then $c \leq a$, and so part 3. of the proposition applies.

8. If $a > b$ and $b \geq c$ then $a > c$:

Applying part 2. of this proposition to $a > b$ and $a > c$ and part 1. to $b \geq c$ gives the equivalent statement $b < a$ and $c \leq a$ then $c < a$, and so part 4. of the proposition applies.

9. If $a \geq b$ and $b > c$ then $a > c$:

Applying part 1. of this proposition to $a \geq b$ and part 1. to $b > c$ and $a > c$ gives the equivalent statement $b \leq a$ and $c < b$ then $c < a$, and so part 5. of the proposition applies.

10. If $a > b$ and $b > c$ then $a > c$:

Applying part 2. to $a > b$, $b > c$ and $c > a$ gives the equivalent statement $b < a$ and $c < b$ then $c < a$ and so part 6. applies.

11. If $a < b$ then $a + c < b + c$:

Suppose that $a < b$, then $a \subset b$. We argue by induction on c that $(a + c) \subset (b + c)$.

Let $P(c)$ be the proposition given by

$$(a + c) \subset (b + c)$$

For the base case $c = 0$ and we trivially have $a < b$ by hypothesis. Hence $P(0)$ is true.

So suppose that $P(c)$ is true, that is to say

$$(a + c) \subset (b + c)$$

We need to show that $P(c + 1) = P(S(c))$ is true. That is

$$(a + S(c)) \subset (b + S(c))$$

We know from the definition of addition that $\forall m \in \mathbb{N}$ and $n \neq 0$ that

$$m + n = m + S(n) = S(m + n)$$

Hence we have

$$(a + S(c)) \subset (b + S(c)) \Rightarrow S(a + c) \subset S(b + c)$$

By the induction hypothesis, we know that $a + c \subset b + c$. Let $x, y \in \mathbb{N}$ with $x = a + c$ and $y = b + c$. Then we have to show that

$$S(x) \subset S(y)$$

Now, we have that $x = x + 0$, likewise $y = y + 0$ and so

$$\begin{aligned} S(x) &\subset S(y) \\ S(x + 0) &\subset S(y + 0) \\ x + S(0) &\subset y + S(0) \\ a + c + S(0) &\subset b + c + S(0) \\ a + S(c + 0) &\subset b + S(c + 0) \\ a + S(c) &\subset b + S(c) \end{aligned}$$

Hence $P(S(c)) = P(c + 1)$ holds.

The result follows by induction. Therefore $a + c \subset b + c$ for all $c \in \mathbb{N}$ and therefore $a + c < b + c$.

12. If $a \leq b$ then $a + c \leq b + c$:

Suppose that $a \leq b$. If $a < b$ then by part 11. we have $a + c < b + c$. So suppose that $a = b$ then we must have that $a + c = b + c$ and so by definition $a + c \leq b + c$.

13. If $a > b$ then $a + c > b + c$:

Applying part 2. of the proposition give the equivalent statement of $b < a$ then $b + c < a + c$ and so we can apply part 11.

14. If $a \geq b$ then $a + c \geq b + c$:

Applying part 1. of the proposition give the equivalent statement of $b \leq a$ then $b + c \leq a + c$ and so we can apply part 12.

15. If $a < b$ then $ac < bc$:

Suppose that $a < b$, then $a \subset b$. We argue by induction on c that $ac \subseteq bc$.

Let $P(c)$ be the proposition given by

$$(ac) \subset (bc)$$

For the base case $c = 0$ and we trivially have $a * 0 < b * 0 \Rightarrow 0 < 0$ is vacuously true. Hence $P(0)$ is true.

So suppose that $P(c)$ is true, that is to say

$$(ac) \subset (bc)$$

We need to show that $P(c + 1) = P(S(c))$ is true. That is

$$(aS(c)) \subset (bS(c))$$

We know from the definition of multiplication that $\forall m \in \mathbb{N}$ and $n \neq 0$ that

$$m * n = m * S(n) = m * n + m$$

Hence we have

$$(aS(c)) \subset (bS(c)) \Rightarrow a * c + c \subset b * c + c$$

By the induction hypothesis, we know that $ac < bc$ and so by part 11. we conclude that $ac + c \subset bc + c$ which is to say $aS(c) \subset bS(c)$. Hence $P(S(c)) = P(c + 1)$ is true and the result follows by induction. Hence we conclude that $ac \subset bc$.

16. If $a \leq b$ then $ac \leq bc$:

Suppose $a \leq b$ then if $a < b$ we apply part 15. Otherwise, we have that $a = b$ and so by definition $ac = bc$ which is to say $ac \leq bc$.

17. If $a > b$ then $ac > bc$:

Applying part 2. of the proposition gives the equivalent statement of $b < a$ then $bc < ac$ and so we apply part 15. of the proposition.

18. If $a \geq b$ then $ac \geq bc$:

Applying part 1. of the proposition gives the equivalent statement of $b \leq a$ then $bc \leq ac$ and so we apply part 16. of the proposition.

The result has been shown. \square

1.4 Cardinality, countability, relations

God created infinity, and man, unable to understand infinity, had to invent finite sets.

Gian-Carlo Rota

1.4.1 Cardinality

In the previous chapter when constructing the natural numbers, we made continuous reference to the idea that the number $1 = \{\emptyset\}$ is somehow the set that contains a single element, $3 = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}$ somehow contains three individual elements. We can make this a rigorous definition, to do so we will use the idea we have been using all along. This is to say, the natural number n is a set that has n elements.

Definition 1.4.1. *Cardinality of a natural number*

We define the cardinality of a natural number $n \in \mathbb{N}$, which we will denote $|n|$ to be the same as the identity mapping. That is to say

$$\begin{aligned} |\cdot| : \mathbb{N} &\rightarrow \mathbb{N} \\ n &\mapsto |n| = n \end{aligned}$$

Example 1.4.1. *Consider 1 and 3 from before. We have that*

$$|1| = |\{\emptyset\}| = 1$$

and

$$|3| = |\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}| = 3$$

Indeed, we have captured the essence of that intuitive idea that a natural number n is a set that has n elements.

Now that we have a notion of size for natural numbers, we can extend this idea to sets in general. In particular, how many elements does a given set have? To build this idea we will also be making use of mappings. For an example, suppose we have the set $S = \{2, 4, 6, 8\}$. Intuitively we know that this is a set which has four elements, and by our definition above we know that $|4| = |\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}| = 4$, is a set that contains 4 elements. Now, consider the mapping $f : S \rightarrow 4$, given by

$$\begin{aligned} f(2) &= \emptyset \\ f(4) &= \{\emptyset\} \\ f(6) &= \{\emptyset, \{\emptyset\}\} \\ f(8) &= \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} \end{aligned}$$

With f being defined as it is, we conclude that f is a bijection. Hence we know that each element $x \in S$ maps exactly to one and only one of the elements $y \in 4$. This somehow tells us that S is a set which has 4 elements, as 4 is a set which contains 4 elements. We can make this a definition of the size of a set, and in doing so define the notion of a finite and “Infinite” set

Definition 1.4.2. *Cardinality of a set*

Let S be a set.

1. Suppose that $n \in \mathbb{N}$. We define the cardinality of the set S , denoted by $|S| = n$, to be n if and only if there exists a bijective mapping $f : S \rightarrow n$. We write this

$$|S| = n$$

If such a mapping exist we say that S is a finite set of size n . We recall that each element of n is nothing but a set whose elements are sets.

2. Suppose that $f : S \rightarrow \mathbb{N}$ be a bijective mapping. We say that the cardinality of the set S is infinite. Informally we denote this by ∞ , but formally we say that $|S| = \aleph_0 = |\mathbb{N}|$, where \aleph_0 is pronounced Aleph-Null.
3. Suppose that $f : S \rightarrow T$ is a bijective mapping. We say that the sets S and T have the same cardinality and write $|S| = |T|$.

This definition made reference to the idea of an “infinite” set. We know that the axiom of infinity gives us the existence of one infinite set, this infinite set that is defined by the axiom of infinite includes the natural numbers but it also includes the so called ordinal numbers. The natural numbers are what are called cardinal numbers, they refer to the size of collections of objects or the amount of some quantity, they can also be used to list (enumerate) a collection. For example we can think of a race between 20 drivers. We must have that one driver comes first, another second, another third, and so on. Each driver can be listed using a number from 1 to 20 inclusive, alternatively a number from 0 to 19. When used in this way, the natural numbers order by enumeration the positions the drivers in the race finished. Now, elements in the infinite set that are not the natural numbers also have this property that they can be used to enumerate the finishing positions of race drivers, however to use them we first would have to go through every single natural number first. The first such non-natural number ordinal is usually denoted ω , and so to label something ω -th, infinitely many things would have to come before.

This gets complicated quickly and as such we won’t go into more details for now. Instead the idea that the natural numbers can be used for enumeration turns out to be a useful one, especially later down the line when we start considering sets like \mathbb{R} . For now, we are only interested in sets whose cardinality is either finite or \aleph_0 and we will continue the exploration of cardinality.

To continue this exploration we will need to relate the ideas of subsets to that of cardinality.

Proposition 1.4.1. *Proper subset of a finite set has strictly smaller cardinality*

Let S and T be finite sets such that $S \subset T$, then we have that $|S| < |T|$.

Proof:

Let S and T be finite sets such that $S \subset T$ with say $|T| = n$, we argue by induction on n , the cardinality of the set T .

Let $P(n)$ be the proposition given by

If T is a finite set with $S \subset T$ and $|T| = n$ then S is a finite set and $|S| < |T| = n$

We need to show that

1. $P(0)$ is true.
2. If $P(n)$ is true then $P(n+1)$ is true.

1. $P(0)$ is true:

We have that $|T| = 0$ and so $T = \emptyset$. As $T = \emptyset$ then there are no subsets $S \subset \emptyset$ for if there were then $T \neq \emptyset$. Hence the base case is vacuously true.

2. If $P(n)$ is true then $P(n+1)$ is true:

Suppose that $P(n)$ holds for some $n \in \mathbb{N}$ which is the statement

If T is a finite set with $S \subset T$ and $|T| = n$ then S is a finite set and $|S| < |T| = n$

We need to show that $P(n+1)$ also holds that is we show that

If T is a finite set with $S \subset T$ and $|T| = n+1$ then S is a finite set and $|S| < |T| = n+1$

So suppose that $|T| = n+1$ for some $n \in \mathbb{N}$ such that $S \subset T$. As S is a strict subset of T we know that $\exists t \in T$ with $t \notin S$. Hence we have that $S \subseteq T \setminus \{t\}$. We need to now show that $|T \setminus \{t\}| = n$

Lemma 1.4.1. Set of cardinality $n+1$ minus an element has cardinality n

Let S be a finite set with cardinality $n+1$. Consider the set $S \setminus \{s\}$ where $s \in S$ is an arbitrary element of S . We have that $|S \setminus \{s\}| = n$

Proof:

We need to show that for the set $S \setminus \{s\}$ that there exists a bijective mapping to a set of n elements. We know that S has cardinality $n+1$, hence there exists a bijection $f : S \rightarrow n+1$. We know by construction that $n+1 = n \cup \{n\}$, hence we have that $n = n+1 \setminus \{n\}$.

Consider the mapping given by g defined as follows

$$g : S \setminus \{s\} \rightarrow n = n+1 \setminus \{n\}$$

$$x \mapsto g(x) = \begin{cases} f(x) : \text{If } f(x) \neq \{n\} \\ f(s) : \text{If } f(x) = \{n\} \end{cases}$$

This is to say g is a mapping that takes each $x \in S$ and maps it to $f(x)$ if $f(x) \neq \{n\} \in n+1$, that is if f doesn't map x to the removed element of the set $n+1$, otherwise if f does map an element $x \in S$ to $\{n\}$ then g maps x to whatever f takes the removed element s to.

For example suppose that $S = \{0, 1, 2\}$, i.e we are considering the case $n = 2$, let $f : S \rightarrow 3$ be the identity mapping, this is a bijection. Suppose we now consider $S \setminus \{2\} = \{0, 1\}$ and consider the mapping $g : S \setminus \{2\} \rightarrow 2$ given by

$$g : S \setminus \{2\} \rightarrow 2 = 3 \setminus \{2\} = \{\emptyset, \{\emptyset\}\}$$

$$x \mapsto g(x) = \begin{cases} f(x) : \text{If } f(x) \neq \{2\} \\ f(2) : \text{If } f(x) = \{2\} \end{cases}$$

We have that $g(0) = 0 = \emptyset$ and $g(1) = 1 = \{\emptyset\}$. We could have instead considered $S \setminus \{1\} = \{0, 2\}$ again with f being the identity mapping. We have that in this case g is the mapping given by

$$g : S \setminus \{1\} \rightarrow 2 = 3 \setminus \{2\} = \{\emptyset, \{\emptyset\}\}$$

$$x \mapsto g(x) = \begin{cases} f(x) : \text{If } f(x) \neq \{2\} \\ f(1) : \text{If } f(x) = \{2\} \end{cases}$$

In this case we have that $g(0) = 0 = \emptyset$ but $g(2) = f(1) = 1 = \{\emptyset\}$.

Now, we need to show the general case where g is given by

$$g : S \setminus \{s\} \rightarrow n = n+1 \setminus \{n\}$$

$$x \mapsto g(x) = \begin{cases} f(x) : \text{If } f(x) \neq \{n\} \\ f(s) : \text{If } f(x) = \{n\} \end{cases}$$

is a bijection.

(a) g is an injection:

To see that g is an injection, suppose that $x, y \in S \setminus \{s\}$ and that $x \neq y$. There are three cases to consider.

i. $f(x) \neq \{n\}$ and $f(y) \neq \{n\}$:

We have by definition of the mapping g that $f(x) = g(x)$ and $f(y) = g(y)$. Moreover we know that f is a bijection and in particular an injection, hence as $f(x) \neq f(y)$ we must have that $g(x) \neq g(y)$.

ii. $f(x) = \{n\}$:

By the definition of the mapping g we have that $g(x) = f(s)$. Now, recall that $y \in S \setminus \{s\}$, thus it follows that $y \neq s$. Now, by the injectivity of f we have that $f(y) \neq f(s) = g(x)$. Moreover by the injectivity of f we have that $f(y) \neq \{n\}$. It now follows by definition of g that

$$g(y) = f(y) \neq f(x) = g(x)$$

That is $g(y) \neq g(x)$.

iii. $f(y) = \{n\}$:

This is the same as $f(x) = \{n\}$ except the roles of x and y are swapped, for completeness we give the details.

By the definition of the mapping g we have that $g(y) = f(s)$. Now, as $x \in S \setminus \{s\}$ it follows that $x \neq s$. By the injectivity of f we have that $f(x) \neq f(s) = g(y)$. Moreover by the injectivity of f we have that $f(x) \neq \{n\}$. It now follows by definition of g that

$$g(x) = f(x) \neq f(y) = g(y)$$

That is $g(y) \neq g(x)$.

This shows that g is an injection.

(b) g is a surjection:

We need to show that $\forall y \in n, \exists x \in S$ such that $g(x) = y$. Let $y \in n$. We know that f is a bijection and in particular it is a surjection and so by definition we know we must have

$$\forall y \in n+1, \exists x \in S : f(x) = y$$

Consider the definition of g . We know that $g : S \setminus \{s\} \rightarrow n$, hence to show that g is surjective we need to show that any $y \in n$ has an element $x' \in S$ with $f(x') = y$. Moreover as S doesn't have the element s we can't use $x = s$ in the surjectivity of f to show surjectivity of g .

$$\forall y \in n = n+1 \setminus \{n\}, \exists x' \in S \setminus \{s\} : x' \neq a \text{ and } f(x') = y$$

Finally, we need to consider $f(s)$ and in particular the two cases of $f(s) \neq \{n\}$ and $f(s) = \{n\}$, from the definition of g .

i. $f(s) \neq \{n\}$:

Suppose that $f(s) \neq \{n\}$. As f is a bijection we have that f is invertible, in particular we must have that $f^{-1}(\{n\}) \neq s$. There are two additional cases to consider now, $f(s) = y = f(x)$ and $f(s) \neq y = f(x)$.

A. $f(s) = y = f(x)$:

Suppose that $f(s) = y$, by definition of g we have that

$$g(f^{-1}(\{n\})) = y$$

as $f^{-1}(\{n\}) \neq s$. So let $x' = f^{-1}(\{n\})$.

B. $f(s) \neq y = f(x)$:

Suppose that $f(s) \neq y$, by assumption of surjectivity of f we have that $f(x) = y$. Hence $f(s) \neq f(x)$ and so by injectivity of f we have that $x \neq s$, hence we can simply take $x' = x$,

ii. $f(s) = \{n\}$:

Now suppose that $f(s) = \{n\}$. We know that $\{n\} \notin n$ and so by assumption we have that $f(x) = y \neq \{n\}$. Thus we conclude that $x \neq s$ so we let $x' = x$.

In each case we have found a valid choice for x' and so surjectivity has been shown.

It follows that $g : S \setminus \{s\} \rightarrow n$ is a bijection and by definition of set cardinality we conclude that $S \setminus \{s\}$ has cardinality n . As required. \square

Now, by the lemma we have that $T \setminus \{t\}$ is set of cardinality n . Now if $S = T \setminus \{t\}$ then $|S| = n < n + 1 = S(n)$ and so is finite by definition, otherwise we must have that S is a proper subset of $T \setminus \{t\}$. So the induction hypothesis holds, that is S is a finite set with less than n elements. Moreover as $n < n + 1$ it follows that S has less than $n + 1$ elements.

Hence $P(n + 1)$ holds.

The result now follows by induction. \square

This proposition has an immediate consequence.

Lemma 1.4.2. *Subset of a finite set has at most the same cardinality*

Let S and T be finite sets such that $S \subseteq T$, we have that $|S| \leq |T|$.

Proof:

There are two cases to consider. Firstly if $S = T$ we have by definition that S and T have the same elements and therefore the identity map is a bijection between the two sets. Hence $|S| = |T|$. The finally case is $S \subset T$ which is simply proposition 1.4.1. \square

We defined the cardinality of a set S in terms of a bijective mapping from S to \mathbb{N} , although this doesn't mean we can't deduce things about cardinality for say injective mappings or surjective mappings. We will assume unless stated otherwise that the sets we are dealing with are finite.

Proposition 1.4.2. *Cardinality of finite sets in an injective mapping*

Let S and T be two finite sets, and suppose that $f : S \rightarrow T$ is an injection. We have that

$$|S| \leq |T|$$

Proof:

Suppose that $f : S \rightarrow T$ is an injective mapping between finite sets with $|S| = n$ and $|T| = m$. Now consider the mapping given by $g : S \rightarrow \text{Image}(f)$. We have by proposition 1.2.18 that an injective mapping to the image is a bijection and so by definition $|\text{Image}(f)| = n$. Additionally by definition of the image of f we have that $\text{Image}(f) \subseteq T$. It follows that as $\text{Image}(f) \subseteq T$ then $n = |S| = |\text{Image}(f)| \leq |T| = m$, that is $|S| \leq |T|$. As required. \square

Proposition 1.4.3. *Cardinality of finite sets in a surjective mapping*

Let S, T be two finite sets, and suppose that $f : S \rightarrow T$ is a surjection. We have that

$$|T| \leq |S|$$

Proof:

Suppose that $f : S \rightarrow T$ is a surjective mapping between finite sets with $|S| = n$ and $|T| = m$. For each $t \in T$ define $x_t \in f^{-1}(\{t\})$, x_t exists because f is surjective and so by definition for any $t \in T$ there is some $s \in S$ such that $f(s) = t$.

Define $X = \{x_t \in S : t \in T\}$, that is X is the set of all such elements defined by the pre-image above. Clearly $X \subseteq S$ and so by 1.4.2 we have that $|X| \leq |S|$.

Now, consider the restriction mapping $f|_X$. We have that for all $t \in T$ that $x_t \in X$ and that $f|_X(x_t) = t$, and so $f|_X$ is surjective. Moreover if we have so some $x_t, x_v \in X$ with $f|_X(x_t) = t$ and $f|_X(x_v) = v$ and $t = v$ then by definition we have that $x_t = x_v$ and $f|_X$ is a bijection. Hence by definition $|T| = |X| \leq |S|$. \square

If we have two sets of finite cardinality, what can we say about the Cartesian product? This should also have finite cardinality. If we have a set S of cardinality n and a set T of cardinality m . The Cartesian product $S \times T$ has elements of the form (s, t) for $s \in S$ and $t \in T$. For some element s_0 we can have that every element $t \in T$ is in $S \times T$ for which there are precisely m such elements of this form. We can do this for each element in $s \in S$ for which there are n such elements. Hence we expect the total number of elements in $S \times T$ to be nm .

Proposition 1.4.4. *Cardinality of the Cartesian product of finite sets*

Let S and T be two sets with cardinalities $|S| = n$ and $|T| = m$ then

$$|S \times T| = |S| |T| = nm$$

Proof:

If either one of $|S| = 0$ or $|T| = 0$ then $S = \emptyset$ or $T = \emptyset$ and so $S \times T = \emptyset$. So let S and T be as given then $|S| = n$ and $|T| = m$. Let $s \in S$ and define the following mapping

$$\begin{aligned} f : T &\rightarrow \{s\} \times T \\ t &\mapsto f(t) = (s, t) \end{aligned}$$

We show that f is a bijection. Indeed suppose that $f(a) = f(b)$ where $a, b \in T$ then $(s, a) = (s, b)$ and as s is fixed we conclude that $a = b$ which shows injectivity. Now let $t \in \{s\} \times T$ then $t = (s, t')$ for some $t' \in T$ but then clearly $f(t') = t$ and so $\forall t \in \{s\} \times T, \exists t' \in T$ such that $f(t') = t$. Hence f is surjective and therefore we have that f is a bijection. By proposition 1.4.2 as f is an injective mapping between finite sets then $|T| \leq |\{s\} \times T|$. Likewise by proposition 1.4.3 we conclude that $|\{s\} \times T| \leq |T|$ hence $|T| = |\{s\} \times T| = m$.

Now define the set K by

$$K = \{\{s\} \times T : s \in S\}$$

for any $s \in S$. Define the following mapping

$$\begin{aligned} g : S &\rightarrow K \\ x &\mapsto g(x) = \{x\} \times T \end{aligned}$$

We show that g is a bijection. Clearly if $g(a) = g(b)$ then $\{a\} \times T = \{b\} \times T$ and as T is a fixed set then $a = b$ and injectivity holds. Now let $k \in K$ then $k = \{k'\} \times T$ where $k' \in S$ then clearly $g(k') = k$ so surjectivity holds. Hence g is a bijection and so by a similar argument with the mapping f we conclude that $|S| = |K| = n$

We now need to show that set K partitions $S \times T$. This is to say we need to show that

1. $\forall x, y \in K$ we have that $x \cap y = \emptyset$ whenever $x \neq y$
2. $\forall x \in K$ we have that

$$S \times T = \bigcup_{x \in K} x$$

3. $\forall x \in K$ we have that $x \neq \emptyset$

1. $\forall x, y \in K$ we have that $x \cap y = \emptyset$ whenever $x \neq y$

We can make use of the fact that g is a bijection. If $g(x) = g(y)$ then $x = y$ and so $x \cap y = x = y \neq \emptyset$. Now if $g(x) \neq g(y)$ then $x \neq y$ say $x = \{s_1\} \times T$ and $y = \{s_2\} \times T$ with $s_1 \neq s_2$. It follows that $x \cap y = \emptyset$.

2. $\forall x \in K$ we have that

$$S \times T = \bigcup_{x \in K} x$$

By definition we have that any $x \in K$ has the form $\{s\} \times T$ where $s \in S$. Let $y \in \{s\} \times T$ then $y = (s, t)$ for some $t \in T$ and so $y \in S \times T$ therefore

$$\bigcup_{x \in K} x \subseteq S \times T$$

Likewise suppose that $x \in S \times T$ then $x = (s, t)$ for some $s \in S$ and $t \in T$. This implies that $x \in \{s\} \times T$ and as $\{s\} \times T \in K$ then $x \in K$ so that

$$S \times T \subseteq \bigcup_{x \in K} x$$

It follows that

$$S \times T = \bigcup_{x \in K} x$$

for all $x \in K$.

3. $\forall x \in K$ we have that $x \neq \emptyset$

Let $x \in K$ then $x \neq \emptyset$ as $S \neq \emptyset$ and $T \neq \emptyset$. Hence $\forall x \in K$ $x \neq \emptyset$.

It follows that K partitions $S \times T$. Now as K is a set containing n elements and K partitions $S \times T$ and each element of K is a set containing m elements. We have that the cardinality of $S \times T$ is the sum of the cardinalities of each set $x \in K$ which is $m * n$. That is to say

$$|S \times T| = nm$$

and the result is shown. \square

1.4.2 Countability

Definition 1.4.3. *Countable Set*

Let S be a set. Let $T \subseteq \mathbb{N}$ allowing for the possibility that $T = \mathbb{N}$. We say that S is a countable set if and only if the mapping $f : S \rightarrow T$ is a bijection.

If T is a finite subset of \mathbb{N} we say that S is a finitely countable set and thus countable. If $T = \mathbb{N}$ we say that S is a countably infinite set. If S is not a finitely countable set or a countably infinite set we say that S is an uncountably infinite set.

Informally, a set S is finitely countable or countably infinite if we have some process for which we can enumerate each element of S , that is to say list out each element in some way. We have an immediate result. We can make the notion of an enumeration rigorous

Definition 1.4.4. *Enumeration*

Let S be a finitely countable set with cardinality $|S| = n$ and define $\mathbb{N}_n = \{1, 2, 3, \dots, n\}$ for some $n \in \mathbb{N}$. We define an enumeration of S to be a bijective mapping $f : \mathbb{N}_n \rightarrow S$ or a bijective mapping $g : S \rightarrow \mathbb{N}_n$.

If S is a countably infinite we define an enumeration of S to be the bijection $f : \mathbb{N} \rightarrow S$ or a bijective mapping $g : S \rightarrow \mathbb{N}$.

It is clear that in either case the if f is an enumeration of a countable set S then so is f^{-1} is also an enumeration of S .

Proposition 1.4.5. *Inverse of an enumeration mapping is an enumeration mapping*

1. Let S be a finitely countable set with cardinality $|S| = n$ have enumeration $f : \mathbb{N}_n \rightarrow S$ then $f^{-1} : S \rightarrow \mathbb{N}_n$ is an enumeration of S where f and f^{-1} define the same enumeration of the elements of S
2. Let S be a countable set have enumeration $f : \mathbb{N} \rightarrow S$ then $f^{-1} : S \rightarrow \mathbb{N}$ is an enumeration of S where f and f^{-1} define the same enumeration of the elements of S

Proof:

1. Let S be a finitely countable set with cardinality $|S| = n$ have enumeration $f : \mathbb{N}_n \rightarrow S$ then $f^{-1} : S \rightarrow \mathbb{N}_n$ is an enumeration of S where f and f^{-1} define the same enumeration of the elements of S :

As f is a bijection then it has an inverse $f^{-1} : S \rightarrow \mathbb{N}_n$ which is also a bijection. Hence f^{-1} is an enumeration. To show that f and f^{-1} define the same enumeration of the elements of S we note that $f \circ f^{-1} = \text{id}_{\mathbb{N}_n}$ and $f^{-1} \circ f = \text{id}_S$.

2. Let S be a countable set have enumeration $f : \mathbb{N} \rightarrow S$ then $f^{-1} : S \rightarrow \mathbb{N}$ is an enumeration of S where f and f^{-1} define the same enumeration of the elements of S :

As f is a bijection then it has an inverse $f^{-1} : S \rightarrow \mathbb{N}$ which is also a bijection. Hence f^{-1} is an enumeration. To show that f and f^{-1} define the same enumeration of the elements of S we note that $f \circ f^{-1} = \text{id}_{\mathbb{N}}$ and $f^{-1} \circ f = \text{id}_S$.

The result is shown. \square

Proposition 1.4.6. *The natural numbers are countably infinite*

We have that \mathbb{N} is a countably infinite set.

Proof:

To show that \mathbb{N} is countable we need to find a bijective mapping $f : \mathbb{N} \rightarrow \mathbb{N}$. We can clearly take $\text{id}_{\mathbb{N}}$, that is the identity mapping on \mathbb{N} . That is to say

$$\begin{aligned} \text{id}_{\mathbb{N}} : \mathbb{N} &\rightarrow \mathbb{N} \\ x &\mapsto \text{id}_{\mathbb{N}}(x) = x \end{aligned}$$

As required. \square

We also have the following immediate result.

Proposition 1.4.7. *Any subset of \mathbb{N} is countable*

Let $S \subseteq \mathbb{N}$ then S is countable.

Proof:

Let $S \subseteq \mathbb{N}$ and suppose that S is not finite, for if it is by definition it is countable. As \mathbb{N} is well-ordered we have by theorem 1.3.16 that S is well-ordered and so have a set inclusion minimal element say s_0 . As S is infinite then $S \setminus \{s_0\}$. We will use this as the basis for induction.

Suppose we have $s_n \in S \setminus \{s_0, s_1, s_2, \dots, s_{n-1}\}$ then another application of the well-order principle means there is some set inclusion minimal element s_{n+1} with $s_{n+1} \in S \setminus \{s_0, s_1, s_2, \dots, s_n\}$. This holds for all $n \in \mathbb{N}$ and so we conclude that $S = \{s_0, s_1, s_2, \dots\}$ is countable by defining the bijective mapping mapping

$$\begin{aligned} f : \mathbb{N} &\rightarrow S \\ x &\mapsto f(x) = s_x \end{aligned}$$

The result follows. \square

Proposition 1.4.8. *The empty-set is countable*

We have that \emptyset is a countable set.

Proof:

The empty-set has cardinality 0 which is finite. \square

There are some results that can be deduced which give equivalent conditions for a set to be countable. Two of these results follow by definition of a countable set.

Proposition 1.4.9. *Equivalence definitions of a countable set*

Let S be a set. The following hold.

1. S is countable if and only if there is an injection $f : S \rightarrow T$ for some subset $T \subseteq \mathbb{N}$
2. S is countable if and only if $S = \emptyset$ or there is a surjection $f : T \rightarrow S$ for some subset $T \subseteq \mathbb{N}$

Proof:

1. S is countable if and only if there is an injection $f : S \rightarrow T$ for some subset $T \subseteq \mathbb{N}$:

(\Rightarrow): Suppose that S is countable then by definition there is a bijection $f : S \rightarrow T$ for some $T \subseteq \mathbb{N}$. As f is a bijection then f is an injection and we are done.

(\Leftarrow): Suppose that there is an injection $f : S \rightarrow T$ for some $T \subseteq \mathbb{N}$. Consider the mapping $g : S \rightarrow \text{Image}(f)$. By proposition 1.2.15 we have that g is a surjection. By definition of a surjection we have that $\forall y \in \text{Image}(f)$ there is some $x \in S$ such that $f(x) = y$. It follows that g is a bijection as g is also an injection by definition of the image of a mapping. Therefore $|S| = |\text{Image}(f)|$ and as $\text{Image}(f) \subseteq T \subseteq \mathbb{N}$ we have that S is countable.

2. S is countable if and only if $S = \emptyset$ or there is a surjection $f : T \rightarrow S$ for some subset $T \subseteq \mathbb{N}$:

(\Rightarrow): Suppose that S is countable then there is a bijection $f : T \rightarrow S$ and by definition is therefore a surjection.

(\Leftarrow): Suppose that $f : T \rightarrow S$ is a surjection. If $S = \emptyset$ then $f : T \rightarrow S$ is vacuously injective and surjective and therefore $|S| = |\emptyset| = |T|$ and therefore countable. So suppose that $S \neq \emptyset$. By proposition 1.2.14 we have for any mapping $g : X \rightarrow Y$ that the pre-image of $g^{-1}(Y) = X$, therefore $f^{-1}(S) = T$. By assumption $T \subseteq \mathbb{N}$ and is therefore either finite or some countably infinite subset of \mathbb{N} possibly being \mathbb{N} itself. If T is finite then we have that $|S| \leq |T|$ by definition of f being surjective and therefore $|S|$ is finite and therefore countable. So suppose that $|T| = \aleph_0$ then T is either a countable subset of \mathbb{N} or \mathbb{N} itself.

Let $g : T \rightarrow \mathbb{N}$ be a bijection then $g^{-1} : \mathbb{N} \rightarrow T$ is an bijection by proposition 1.2.35 and we have that $f \circ g^{-1} : \mathbb{N} \rightarrow S$ is a surjection by proposition 1.2.20. It is left to show that $f \circ g^{-1}$ being surjective implies S is countable. Proposition 1.2.28 gives that $f \circ g^{-1}$ being surjective means there exists a right inverse h such that $h : S \rightarrow \mathbb{N}$. By proposition 1.2.30 we have that h is injective. It follows by part 1 that S is countable.

The result is shown. \square

Proposition 1.4.10. *Set is countable if cardinality of set equals cardinality of a countable set*

Let S, T be sets such that $|S| = |T|$ then if S is countable so is T .

Proof:

Suppose that S is countable. We have that as $|S| = |T|$ then there exists a bijection $f : S \rightarrow T$, in particular there exists a bijection $g : T \rightarrow S$. Now as S is countable there exists injection $h : S \rightarrow \mathbb{N}$. Now as g is a bijection we have that g is an injection. The mapping $h \circ g : T \rightarrow \mathbb{N}$ is an injection as h and g are. Hence as $h \circ g$ is an injection it follows that T is countable by proposition 1.4.9. \square

1.4.3 Relations

1.4.3.1 Definition of a relation

So far we have seen a few notations that relate elements of a set to another. An example that relates elements of a set is equality of natural numbers, two natural numbers are equal if and only if there are the same element. Another example that we have seen on the natural numbers is the less than operator $<$. A natural number x is less than y if and only if $x \subseteq y$. A more fundamental example of a relation is that of

a mapping $f : S \rightarrow T$. We can consider a function as relating any $s \in S$ and $t \in T$ to the pair (s, t) where $f(s) = t$.

In a sense, we have that the idea of relations is somehow as fundamental as sets and mappings, in fact we just described a mapping as some form of relation so the idea of relations is more fundamental than that of a mapping. Using the examples of the comparison operators on \mathbb{N} we can motivate a definition for a relation.

Definition 1.4.5. *Relation*

Let S be a set and consider the Cartesian product $S \times S$. A relation is a subset $R \subseteq S \times S$. We write an element $(a, b) \in R$ as aRb or we also write $a \sim b$ and we say that a relates to b . If $(a, b) \notin R$ we write $a \not R b$ or we write $a \not\sim b$.

We can recast the ideas at the start of this section into the language of relations.

Example 1.4.2. Consider equality on \mathbb{N} . We can define equality as a relation $\mathbb{N} \times \mathbb{N}$ where $a \sim b$ if and only if $a \subseteq b$ and $b \subseteq a$. Explicitly we have that R is a subset of $\mathbb{N} \times \mathbb{N}$ given by

$$R = \{(0, 0), (1, 1), (2, 2), \dots\}$$

Example 1.4.3. Consider the less than operator on \mathbb{N} . We have that the less than operator is a relation where $a \sim b$ is given by $a \subset b$. To see this consider $T = \{0, 1, 2\}$. Then the less than relation on T is given by the relation

$$R = \{(0, 1), (0, 2), (1, 2)\}$$

Example 1.4.4. Let $S = \{0, 1\} \subseteq \mathbb{N}$ and define $T = P(S)$ be the power set of S given by

$$T = \{\emptyset, \{0\}, \{1\}, \{0, 1\}, S\}$$

We can define a relation $R \subseteq T \times T$ by

$$R = \{(\emptyset, \emptyset), (\emptyset, \{0\}), (\emptyset, \{1\}), (\emptyset, \{0, 1\}), (\emptyset, S), (\{0\}, \{0\}), (\{0\}, \{0, 1\}), (\{0\}, S), \\ (\{1\}, \{1\}), (\{1\}, \{0, 1\}), (\{1\}, S), (\{0, 1\}, \{0, 1\}), (\{0, 1\}, S), (S, S)\}$$

This relation expresses inclusive subset inclusion, \subseteq , on S .

Example 1.4.5. Let $S = \{0, 1, 2\}$ and $T = S$. Define $T \times T$ by

$$T \times T = \{(0, 0), (0, 1), (0, 2), (1, 0), (1, 1), (1, 2), (2, 0), (2, 1), (2, 2)\}$$

We can use the less than or equal to operator, \leq , to define a relation. We have that

$$R = \{(0, 0), (0, 1), (0, 2), (1, 1), (1, 2), (2, 2)\}$$

1.4.3.2 Reflexive Relation

All of the examples from the previous section, except the strictly less than example, share a common property. Each element is related to itself, that is in each example there is some element $s \in S$ such that $(s, s) \in R \subseteq S \times S$. We formalise this in the following definition.

Definition 1.4.6. *Reflexive relation*

Let S be a set with a relation $R \subseteq S \times S$. We say that the relation R is reflexive if and only if $\forall s \in S$ we have that $(s, s) \in R$. If there is an $s \in S$ such that $(s, s) \notin R$ then we say that the relation is anti-reflexive.

We have given examples of reflexive relations and one example of an anti-reflexive relation. We give an additional example of an anti-reflexive relation.

Example 1.4.6. We have for $a, b \in \mathbb{N}$ that $a = b$ if and only if $a \subseteq b$ and $b \subseteq a$. If this doesn't hold then $a \neq b$ and either one of $a \subseteq b$ or $b \subseteq a$ is true but not both. It follows that the relation $a \sim b$ meaning $a \neq b$ is anti-reflexive. This also implies that if $a \neq b$ then either $a \leq b$ or $b \leq a$.

The examples given so far have allowed us to see some examples of relations and one particular type of relation, a reflexive relation. Unfortunately only considering relations on elements a single set S currently gives us few practical examples to work with. A simple extension to the idea of a relation can fix this.

Definition 1.4.7. *Binary Relation*

Let S and T be sets. We define a binary relation to be a subset $R \subseteq S \times T$. We write an element $(s, t) \in R$ as sRt or write $s \sim t$ and we say that s relates to t . If $(s, t) \notin R$ we write $s \not R t$ or we write $s \not\sim t$.

We can extend this the notion of a relation and binary relation to that of any finite Cartesian product

Definition 1.4.8. *n-ary Relation*

Let $S_1, S_2, S_3, \dots, S_n$ be sets. We define an n -ary relation to be a subset $R \subseteq S_1 \times S_2 \times S_3 \times \dots \times S_n = \mathbb{S}$. An element of R has the form $r = (r_1, r_2, r_3, \dots, r_n)$ and we say that the elements of r relate. We write this as $R(r) = R(r_1, r_2, r_3, \dots, r_n)$

In light of these previous definitions we would like to extend the definition of a reflexive relation to binary and n -ary relations. To see how we could extend a reflexive relation to a binary relation suppose we have two sets S and T . The definition of a reflexive relation of a set Z is that $(z, z) \in R_z \subseteq Z \times Z$ where $z \in Z$ and R_z is the relation defined on Z . A natural way to extend this two S and T is to have either $(s, s) \in R \subseteq S \times T$ or $(t, t) \in R \subseteq S \times T$ where R is a binary relation for S and T . Hence for a reflexive binary relation to makes sense we must have that $s, t \in S \cap T$ and therefore the relation would have to be defined on $S \cap T$.

In the first case $(s, s) \in R \subseteq S \times T$ we have by definition of an ordered tuple that $(s, s) \in R$ if and only if $s \in S$ and $s \in T$. Likewise for $(t, t) \in R \subseteq S \times T$ we must have $s \in S$ and $t \in T$ which is to say $s, t \in S \cap T$. If $S \neq T$ then there will exist at least one element $(s, t) \in R \subseteq S \times T$ where either $s \in S$ and $s \notin T$ or $t \in T$ and $t \notin S$, in this case it is not possible for a reflexive relation to exist.

Definition 1.4.9. *Reflexive binary relation*

Let S and T with relation $R \subseteq S \times T$. We say that the relation R is reflexive if and only if $S = T$.

A similar argument shows there can be no reflexive n -ary relation unless all of the sets that make the relation are the same. For example consider the sets X, Y and Z . The natural way to represent a relation $R \subseteq X \times Y \times Z$ would be to have either $(x, x, x) \in R$, $(y, y, y) \in R$ or $(z, z, z) \in R$ where $x \in X$, $y \in Y$ and $z \in Z$. If $(x, x, x) \in R$ then by definition we must have $x \in Y$ and $x \in Z$, likewise if $(y, y, y) \in R$ then $y \in X$ and $y \in Z$ and finally if $(z, z, z) \in R$ then $z \in X$ and $z \in Y$. Any of the cases implies that $x, y, z \in X \cap Y \cap Z$

Definition 1.4.10. *Reflexive n-ary relation*

Let $S_1, S_2, S_3, \dots, S_n$ be sets with relation $R \subseteq S_1 \times S_2 \times S_3 \times \dots \times S_n$. We say that the relation R is reflexive if and only if $S_i = S_j$ for all $i, j \in \{1, 2, 3, \dots, n\}$

This means when talking about a reflexive relation we only need to consider a single set.

An example of a binary relation is a mapping.

Example 1.4.7. Let $S = T = \mathbb{N}$ and define the mapping $f : S \rightarrow T$ given by $f(s) = s$. We have that f defines a relation as we have that

$$R = \{(0, 0), (1, 1), (2, 2), (3, 3), \dots\} \subseteq \mathbb{N} \times \mathbb{N}$$

Example 1.4.8. Let $S = \{1, 2\}$ and $T = \{3, 4\}$. Define the mapping $f : S \rightarrow T$ by $f(1) = 4$ and $f(2) = 3$. We have f defines a relation as

$$R = \{(1, 4), (2, 3)\} \subseteq S \times T$$

We can consider operators as relations by using the n -ary notion of a relation

Example 1.4.9. Let $X = Y = Z = \mathbb{N}$. We can consider the operator $+$ as a mapping given by

$$\begin{aligned} f : X \times Y &\rightarrow Z \\ (x, y) &\mapsto f(x, y) = x + y \end{aligned}$$

A relation can be defined by f . A sample of this relation R looks as follows

$$R = \{(0, 0, 0), (0, 1, 1), (4, 3, 7), (3, 4, 7), (2, 2, 4), \dots\} \subseteq \mathbb{N} \times \mathbb{N} \times \mathbb{N}$$

In general, R has the following definition

$$R = \{(x, y, x + y) : x, y \in \mathbb{N}\}$$

We note that as $X = Y$ then for any $x \in X$ we have $x \in Y$ and likewise for any $y \in Y$ we have that $y \in X$. We therefore have that $R(x, y, x + y) = R(y, x, y + x)$. This is confirming the fact that addition is commutative.

Example 1.4.10. Let $X = Y = Z = \mathbb{N}$. We can consider the operator $*$ as a mapping given by

$$\begin{aligned} f : X \times Y &\rightarrow Z \\ (x, y) &\mapsto f(x, y) = x * y \end{aligned}$$

The relation defined by f looks as follows

$$R = \{(0, 0, 0), (0, 1, 0), (4, 3, 12), (3, 4, 12), (2, 2, 4), \dots\} \subseteq \mathbb{N} \times \mathbb{N} \times \mathbb{N}$$

In general, R has the following definition

$$R = \{(x, y, x * y) : x, y \in \mathbb{N}\}$$

As before, we have that as $X = Y$ then for any $x \in X$ we have $x \in Y$ and likewise, for any $y \in Y$ we have that $y \in X$. We, therefore, have that $R(x, y, x * y) = R(y, x, y * x)$, again confirming the fact that multiplication is commutative.

Example 1.4.11. Let $X = Y = Z = \mathbb{N}$. We can consider the operator \wedge as a mapping given by

$$\begin{aligned} f : X \times Y &\rightarrow Z \\ (x, y) &\mapsto f(x, y) = \wedge(x, y) = x^y \end{aligned}$$

The relation defined by f looks as follows

$$R = \{(0, 0, 1), (0, 1, 0), (2, 3, 8), (8, 2, 64), (3, 2, 9), \dots\} \subseteq \mathbb{N} \times \mathbb{N} \times \mathbb{N}$$

In general, R has the following definition

$$R = \{(x, y, x^y) : x, y \in \mathbb{N}\}$$

As before, we have that as $X = Y$ then for any $x \in X$ we have $x \in Y$ and likewise, for any $y \in Y$ we have that $y \in X$. We, therefore, have that $R(x, y, x^y) \neq R(y, x, y^x)$, which confirms that in general exponentiation is not commutative.

The last three examples expose another property that relations can have. If two or more elements relate then it doesn't matter which way the relation is written, that is if $x \sim y$ then we can have the case that $y \sim x$. Such a relation is called symmetric.

Definition 1.4.11. *Symmetric relation*

Let S be a set with relation $R \subseteq S \times S$. We say that R is a symmetric relation if and only if $\forall x, y \in S$ we have that xRy implies yRx , equivalently we can write R is symmetric if and only if $x \sim y$ implies $y \sim x$. If R is not symmetric we say that R is an anti-symmetric relation.

As with reflexive relations we can show that trying to extend the idea of a symmetric relation on a single set to multiple sets we have to conclude the sets have to be the same.

Indeed suppose that S and T are sets with a relation $R \subseteq S \times T$. The natural extension for a symmetric relation would be $\forall s \in S$ that $sRt \Rightarrow tRs$ for $t \in T$. This implies that $t \in S$ and $s \in T$ and therefore $s, t \in S \cap T$.

Definition 1.4.12. *Symmetric binary relation*

Let S and T be sets with relation $R \subseteq S \times T$. We say that R is symmetric if and only if $S = T$

Likewise a similar argument holds for n -ary symmetric relations

Definition 1.4.13. *Symmetric n -ary relation*

Let $S_1, S_2, S_3, \dots, S_n$ be sets with relation $R \subseteq S_1 \times S_2 \times S_3 \times \dots \times S_n$. We say that the relation R is symmetric if and only if $S_i = S_j$ for all $i, j \in \{1, 2, 3, \dots, n\}$

The comparison, less than, less than or equal to, greater than, and greater than or equal to operators on the naturals also give insight into another interesting property. The following examples will make it more clear

Example 1.4.12. Let $S = T = \mathbb{N}$ and define $x \sim y$ by $x \leq y$. Consider $a, b, c \in \mathbb{N}$ with $a = 2$, $b = 4$ and $c = 6$. We have that $a \sim b$ as $2 \leq 4$ and we have that $b \sim c$ as $4 \leq 6$, we clearly also have $a \sim c$ as $2 \leq 6$. In general if we have $a, b, c \in \mathbb{N}$ with $a \leq b \leq c$ we have that $a \sim b$ and $b \sim c$ implies $a \sim c$.

Example 1.4.13. Let $S = T = \mathbb{N}$ and define $x \sim y$ by $x \geq y$. Consider $a, b, c \in \mathbb{N}$ with $a = 8$, $b = 3$ and $c = 1$. We have that $a \sim b$ as $8 \geq 3$ and we have that $b \sim c$ as $3 \geq 1$, we also have $a \sim c$ as $8 \geq 1$. More generally if we have $a, b, c \in \mathbb{N}$ with $a \geq b \geq c$ we have that $a \sim b$ and $b \sim c$ implies $a \sim c$.

Example 1.4.14. Let $S = T = \mathbb{N}$ and define $x \sim y$ by $x = y$. Consider $a, b, c \in \mathbb{N}$ with $a = 2$, $b = 2$ and $c = 2$. We have that $a \sim b$ as $2 = 2$ and we have that $b \sim c$ as $2 = 2$, we also have $a \sim c$ as $2 = 2$. More generally if we have $a, b, c \in \mathbb{N}$ with $a = b = c$ we have that $a \sim b$ and $b \sim c$ implies $a \sim c$.

We see that with certain relations that if $a \sim b$ is true and $b \sim c$ is true then we can conclude that $a \sim c$ is true. Such a relation is called a Transitive relation.

Definition 1.4.14. *Transitive relation*

Let S be a set with relation $R \subseteq S \times S$. We say that R is a transitive relation if and only if $\forall a, b, c \in S$ we have that if aRb and bRc then we have that aRc .

We again consider if a transitive relation can be extended to multiple sets. Suppose that we have a binary relation $R \subseteq S \times T$ for some sets S and T . The natural extension to make R a transitive relation is to have $s \sim t$ and $t \sim u$ implies $s \sim u$ for $s, t \in S$ and $t, u \in T$. Hence we must have $s, t \in S$ but need not have $u \in S$. As we aren't assuming anything else about the relation R there is nothing more we can deduce about a binary transitive relation.

Definition 1.4.15. *Transitive binary relation*

Let S and T be sets with relation $R \subseteq S \times T$. We say that R is transitive if and only if the set \tilde{R} given by

$$\tilde{R} = \{(x, z) \in S \times T : \forall x \in S \wedge \forall z \in T : \exists y \in S \cap T : (x, y) \in R \wedge (y, z) \in R\}$$

is non-empty.

A definition can be made for a transitive n -ary relation. I AM NOT SURE HOW TO DEFINE THIS YET, PAIR-WISE RELATION OF EACH SET???????????????? We can make use of a binary relation in order to define

Definition 1.4.16. *Transitive n -ary relation*

Let $S_1, S_2, S_3, \dots, S_n$ be sets with relation $R \subseteq S_1 \times S_2 \times S_3 \times \dots \times S_n$. We say that the relation R is transitive if and only if the set \tilde{R} given by

$$\tilde{R} = \{(x, z) \in\}$$

is non-empty

1.4.3.3 Equivalence Relations

Of all the examples of relations we have seen so far there is one in particular that is special, the equality operator $=$. This relation is reflexive, symmetric and transitive.

Proposition 1.4.11. *The equality relation on the natural numbers is reflexive, symmetric and transitive*

Let $S = T = \mathbb{N}$ and for $x, y \in \mathbb{N}$ define the relation $x \sim y$ by $x = y$. We have that

1. \sim is reflexive, that is $\forall x \in \mathbb{N}$ we have $x \sim x$
2. \sim is symmetric, that is $\forall x, y \in \mathbb{N}$ we have $x \sim y \Rightarrow y \sim x$
3. \sim is transitive, that is $\forall x, y, z \in \mathbb{N}$ we have that if $x \sim y$ and $y \sim z$ then $x \sim z$

Proof:

1. \sim is reflexive, that is $\forall x \in \mathbb{N}$ we have $x \sim x$:

Let $x \in \mathbb{N}$ then by definition of equality we have that for $y, z \in \mathbb{N}$ that $y = z$ if and only if $y \subseteq z$ and $z \subseteq y$. It is clear that $x = x$ and therefore $x \sim x$ proving reflexivity.

2. \sim is symmetric, that is $\forall x, y \in \mathbb{N}$ we have $x \sim y \Rightarrow y \sim x$:

Let $x, y \in \mathbb{N}$ with $x \sim y$. We have that as $x \sim y$ then $x = y$. By definition of equality we also have $y = x$ and so $y \sim x$ showing that \sim is symmetric.

3. \sim is transitive, that is $\forall x, y, z \in \mathbb{N}$ we have that if $x \sim y$ and $y \sim z$ then $x \sim z$:

Let $x, y, z \in \mathbb{N}$ such that $x \sim y$ and $y \sim z$, then $x = y$ and $y = z$. By definition of equality it follows that $x = z$ and so $x \sim z$ showing transitivity.

The result follows. \square

What does it mean for a relation to be reflexive, symmetric and transitive? In the case of equality on the natural numbers we see that reflexivity tells us that an element is equal to itself. Equality being symmetric tells us that if $x = y$ then $y = x$ that is it does not matter which we say the two numbers are equal. Finally transitivity tells us that if $x = y$ and $y = z$ we are able to deduce that $x = z$. In this context, equality being reflexive, symmetric and transitive allows us to quantify which elements are equivalent. In the case of equality it is clear which elements are equivalent, the ones that are equal!

Example 1.4.15. *Consider $X = Y = \mathbb{N}$ and for $x, y \in \mathbb{N}$ define the relation $R = \mathbb{N} \times \mathbb{N}$. We have that R is reflexive as for any $x \in \mathbb{N}$ we have that $(x, x) \in R$. Likewise R symmetric as $\forall x, y \in \mathbb{N}$ we have that $(x, y) \in R \Rightarrow (y, x) \in R$. as $X = Y$. Finally R is transitive as $\forall x, y, z \in \mathbb{N}$ we have that $(x, y) \in R$ and $(y, z) \in R$ and $(x, z) \in R$.*

What does R being reflexive, symmetric and transitive mean? In this case R being reflexive, symmetric and transitive means that every $x \in X$ and $y \in Y$ are related and we can see R as a relation meaning "is an element of \mathbb{N} ". This means that we have shown that X and Y are equivalent, which we already know by the fact we set $X = Y = \mathbb{N}$.

Based on the two examples we motivate the following definition.

Definition 1.4.17. *Equivalence relation*

Let S be a set and $R \subseteq S \times S$ a relation. We say that R is an equivalence relation if and only if

1. R is reflexive
2. R is symmetric
3. R is transitive

Proposition 1.4.11 is equivalent to saying that equality is an equivalence relation on \mathbb{N} . The two examples also show a disparity between the two equivalence relations shown. In the case of the equality the relation R was a strict subset of $\mathbb{N} \times \mathbb{N}$ where as in the second example R was equal to $\mathbb{N} \times \mathbb{N}$. This raises the question what is different? We can answer this by looking at the set of elements that relate to a given element. Such a set is called an equivalence class.

Definition 1.4.18. *Equivalence class*

Let S be a set, let $x \in S$ and let R be an equivalence relation on S . We define an equivalence class, denoted $[x]$ to be the set

$$[x] = \{y \in S : xRy\}$$

If the context doesn't make clear the relation we are referring we explicitly write $[x]_R$ to be the equivalence class of x under the relation R .

We say that an element $y \in [x]$ is a representative of the equivalence class of x

To get a feel for equivalence classes we consider the, non-mathematical, following example.

Example 1.4.16. Consider the set X to be the set of all people currently alive. Define a relation, \sim , on X by

$$\forall (x, y) \in X \times X : x \sim y \iff x \text{ and } y \text{ where born in the same year}$$

We have that \sim is an equivalence relation. Clearly if $x \sim x$ as x was born in some year D . We have that if $x \sim y$ then x and y are born in the same year and clearly $y \sim x$. Now if $x \sim y$ and $y \sim z$ then x and y are born in the same year and y and z are born in the same year. This therefore means x and z are born in the same year so $x \sim z$ showing transitivity.

Now let $x \in X$ and consider the equivalence class $[x]_{\sim}$. By definition of an equivalence class we have that

$$[x]_{\sim} = \{y \in X : x \sim y\}$$

This means that the equivalence class $[x]_{\sim}$ is the set of all people currently alive that were born in the same year. As X was the set of all currently alive people we have found a way to extract a subset of X such that they are all born in the same year. If we now pick another element of X , say a , such that $x \not\sim a$ then by definition a was not born in the same year as x and $[a]_{\sim}$ is another subset of X of currently alive people born in the same year. Moreover we have that $[x]_{\sim} \neq [a]_{\sim}$. We can do this for every element of X and get a collection of sets that correspond to all of the possible different years that anyone currently alive could possibly be in.

The previous example has shown that we are able to construct a partition of a set S which has an equivalence relation \sim . We can prove this more generally, firstly we recall the definition of a set partition.

Let S be a set and define \mathbb{S} to be the set of subsets of S . We say that \mathbb{S} is a partition of S if the following hold.

1. $\forall S_1, S_2 \in \mathbb{S}$ we have $S_1 \cap S_2 = \emptyset$ whenever $S_1 \neq S_2$
2. Taking the union of every $T \in \mathbb{S}$ gives us S that is

$$S = \bigcup_{T \in \mathbb{S}} T$$

3. $\forall T \in \mathbb{S}$ we have that $T \neq \emptyset$.

Before we can show that the equivalence classes partition the set we must first show that there can be no empty equivalence class.

Proposition 1.4.12. *Equivalence class is non-empty*

Let S be a set with an equivalence relation \sim . Let $x \in S$ then we have that $[x]_{\sim} \neq \emptyset$

Proof:

Let S be a set with an equivalence relation \sim . By definition of an equivalence relation we have that $\forall x, y, z \in S$ that

1. \sim is reflexive, that is $x \sim x$
2. \sim is symmetric, that is $x \sim y \Rightarrow y \sim x$
3. \sim is transitive, that is $x \sim y$ and $y \sim x$ implies that $x \sim z$

Consider the equivalence class $[x]_{\sim}$. By definition of an equivalence class we know that

$$[x]_{\sim} = \{y \in S : x \sim y\}$$

As \sim is reflexive we have that $x \sim x$ and so $x \in [x]_{\sim}$ and therefore $[x]_{\sim} \neq \emptyset$. \square

We can prove that an equivalence relation partitions the set it is defined on.

Theorem 1.4.1. *Equivalence classes of a relation partitions the set*

Let S be a set with an equivalence relation \sim . Let \mathbb{S} denote the equivalence classes of \sim for each $s \in S$. We have that \mathbb{S} is a partition of S .

Proof:

Let S be a set with an equivalence relation \sim and let \mathbb{S} be the set of equivalence classes of \sim for each $s \in S$. Let $x \in S$ then x belongs to at least one equivalence class by proposition 1.4.12. We therefore have that

$$S = \bigcup_{x \in S} [x]_{\sim}$$

It is left to show that if $[x]_{\sim} \neq [y]_{\sim}$ for $x, y \in S$ then we have that $[x]_{\sim} \cap [y]_{\sim} = \emptyset$. This is equivalent to saying that if $[x]_{\sim} \cap [y]_{\sim} \neq \emptyset$ then $[x]_{\sim} = [y]_{\sim}$. So suppose that $[x]_{\sim} \cap [y]_{\sim} \neq \emptyset$ then $[x]_{\sim} \cap [y]_{\sim}$ has at least one element z . Suppose that $z \in [x]_{\sim}$ then by definition we have that $x \sim z$. Let $a \in [x]_{\sim}$ be an arbitrary element of the equivalence class of x . We have that $a \sim x$ then by transitivity of \sim we conclude that $a \sim z$. However as $z \in [x]_{\sim} \cap [y]_{\sim}$ then we have that $z \in [y]_{\sim}$ and so $y \sim z$. As \sim is symmetric we have $z \sim y$ and again by transitivity we conclude that $a \sim y$. Hence $a \in [y]_{\sim}$ and so $[x]_{\sim} \subseteq [y]_{\sim}$.

A similar argument shows $[y]_{\sim} \subseteq [x]_{\sim}$ and therefore we have that $[x]_{\sim} = [y]_{\sim}$. Finally we conclude that unequal equivalence classes are disjoint and therefore the set of equivalence classes \mathbb{S} is a partition for S .

The result is shown. \square

1.5 Construction of the Integers

The trouble with integers is that we have examined only the very small ones. Maybe all the exciting stuff happens at really big numbers, ones we can't even begin to think about in any very definite way.

Ronald Graham

We now have enough theory to consider extending the natural numbers \mathbb{N} . One reason to do this is to provide a completion to the idea of subtraction. Recall that $n - m$ is only defined in \mathbb{N} if and only if $m \leq n$. This is a limiting idea. For example, the idea of debt can't be explained using only \mathbb{N} . We know that if the balance on your bank account is negative then you owe money to someone, if your balance is positive you have money to spend⁸. The natural numbers don't have a concept of "negative" or debt, we can only deal with "positive" values. To keep the financial institutions happy we should resolve this issue.

To do this we need to consider exactly what it is we want to achieve. Firstly we want to be able to define $n - m$ for all $n, m \in \mathbb{N}$. Clearly, if $n \geq m$ then such a number already exists in \mathbb{N} . Secondly, such a number $n - m$ could have many different representations, for example, $6 - 2 = 4$ and $5 - 1 = 4$. We need a way to say that any of these different representations actually represents the same thing. Formally if we have $a, b, c, d \in \mathbb{N}$ such that $a - b = c - d$ then $a - b$ and $c - d$ represent the same number, this is equivalent to $a + d = b + c$. Thinking of $-$ as a relation we can use the language of equivalence relations to solve this issue. That is a relation where $(a, b) \sim (c, d)$

1.5.1 Defining the Integers

We start by recasting the defining of subtraction to be defined as an ordered tuple.

Definition 1.5.1. *Subtraction as an ordered tuple*

Let $a, b \in \mathbb{N}$. We define the subtraction as an ordered tuple $(a, b) \in \mathbb{N}^2$ to mean $(a - b)$. We will call an element $x \in \mathbb{N}^2$ a subtraction tuple. We note that if $a \geq b$ we have $(a - b) \in \mathbb{N}$

From this we can define a relation

Definition 1.5.2. *Relation on subtraction*

Let $(a, b), (c, d) \in \mathbb{N}^2$ be subtraction tuples. We define the relation \sim such that $(a, b) \sim (c, d)$ if and only if $a + d = b + c$

We have that this relation is an equivalence relation.

Proposition 1.5.1. *Relation on subtraction ordered tuples is an equivalence relation*

Let $x, y \in \mathbb{N}^2$ be subtraction tuples and define the relation $x \sim y$ as above. We have that \sim is an equivalence relation.

Proof:

Let $x, y, z \in \mathbb{N}^2$ be subtraction tuples such that $x = (a, b)$, $y = (c, d)$ and $z = (e, f)$. We need to show that \sim is an equivalence relation, that is

1. \sim is reflexive
2. \sim is symmetric
3. \sim is transitive

1. \sim is reflexive:

We have that $x = (a, b)$ and by definition of \sim we know that $x \sim x$ if and only if $a + b = a + b$ which is clear by definition of equality on the natural numbers. Hence $x \sim x$ and \sim is reflexive.

⁸Hopefully not all at once!

2. \sim is symmetric:

We have that $x = (a, b)$ and $y = (c, d)$. Suppose that $x \sim y$ then we have that $a + d = b + c$. By commutativity of equality of natural numbers that $a + d = b + c \Rightarrow b + c = a + d$. By commutativity of addition on the natural numbers we have that $b + c = a + d$ is the same as $c + b = d + a$. Hence we have that $(c, d) \sim (a, b)$ by definition of \sim and so $y \sim x$ showing that \sim is symmetric.

3. \sim is transitive:

We know that $x = (a, b)$, $y = (c, d)$ and $z = (e, f)$. Now suppose that $x \sim y$ and $y \sim z$ then by definition we have that $(a, b) \sim (c, d)$ and $(c, d) \sim (e, f)$ and hence by definition of \sim we have $a + d = c + b$ and $c + f = e + d$.

Consider $a + c + f$ we have

$$a + c + f = a + e + d = a + d + e = c + b + e$$

That is to say $a + c + f = c + b + e$. We have by the cancellation laws on the natural numbers that $a + f = b + e$ which implies that $(a, b) \sim (e, f)$. Which is to say $x \sim z$. Hence transitivity has been shown.

It follows that \sim is an equivalence relation. \square

Now that we have shown that \sim is an equivalence relation we can solve the multiple representation problem by considering the equivalence classes of \mathbb{N}^2 with the relation \sim . Let $x \in \mathbb{N}^2$ with $x = (a, b)$ then the equivalence class of x is given by

$$[x]_{\sim} = [(a, b)]_{\sim} = \{(c, d) \in \mathbb{N}^2 : (a, b) \sim (c, d)\}$$

We know by theorem 1.4.1 that for each $x \in \mathbb{N}^2$ there is set of equivalence classes partition \mathbb{N}^2 and that each equivalence class is disjoint. This is to say if $x, y \in \mathbb{N}^2$ then we have that if $[x]_{\sim} \cap [y]_{\sim} \neq \emptyset$ then $[x]_{\sim} = [y]_{\sim}$. This solves the multiple representation issue.

Let us have a look at some equivalence classes

Example 1.5.1. Let $x \in \mathbb{N}^2$ with $x = (1, 3)$ by definition we have that x represents $1 - 3$. Consider the equivalence class of x , $[x] = \{y \in \mathbb{N}^2 : x \sim y\}$ and let $y \in [x]$. We have that $y = (c, d)$ and that $1 + d = 3 + c$, one possible y where this is true is given by $y = (0, 2)$ and y represents $0 - 2$, As we have $y \in [x]$ then we have that $[x] = [y]$ so we shall take y to be the canonical representative of this equivalence class.

Now that we have that the subtraction tuples are in equivalence classes we can consider the following. Suppose that $a, b, c \in \mathbb{N}$ then what is $a - (b - c)$? For example if $a = 10, b = 6$ and $c = 3$ then we have that $10 - (6 - 3) = 10 - 3 = 7$. This is also the same as $10 + 3 - 6 = 13 - 6 = 7$. This holds in general where we have that $(a, b - c) \sim (a + c, b)$

Lemma 1.5.1. $(a, b - c) \sim (a + c, b)$

Let $a, b, c \in \mathbb{N}$ with $a > b \geq c$. We have that

$$(a, b - c) \sim (a + c, b)$$

Proof:

Let $a, b, c \in \mathbb{N}$ be as given. By definition of \sim we have $(x, y) \sim (u, v)$ if and only if $x + v = u + y$. We argue by contradiction, suppose that $(a, b - c) \not\sim (a + c, b)$ then by definition we have that

$$\begin{aligned}
a + b &\neq a + c + (b - c) \\
b &\neq c + (b - c), \text{ By the cancellation law} \\
b &\neq (c + b) - c, \text{ By proposition 1.3.10} \\
b &\neq (b + c) - c, \text{ By commutativity} \\
b &\neq b + (c - c), \text{ By proposition 1.3.10} \\
0 &\neq (c - c), \text{ By the cancellation law} \\
0 &\neq 0
\end{aligned}$$

A contradiction. \square

By this lemma it follows that $a - (b - c) = (a + c) - b$.

We now look at the definition of what the set of equivalence relations looks like. We make the following definition

Definition 1.5.3. *Quotient set*

Let S be a set with an equivalence relation \sim . Let $x \in S$ and consider the equivalence class $[x]_{\sim}$. We define the quotient set of S , denoted by S/\sim by

$$S/\sim = \{[x]_{\sim} : x \in S\}$$

Why have we called the set of the equivalence classes a quotient set? We can see why with a few examples.

Example 1.5.2. We reconsider the example where X is the set of all people currently alive with the relation \sim given by

$$\forall (x, y) \in X \times X : x \sim y \iff x \text{ and } y \text{ where born in the same year}$$

We know that \sim is an equivalence relation and we know that the equivalence classes define a set of all people currently alive born in a certain year. We can identify the quotient set X/\sim as the set of all of the possible years that all people currently alive could live in. As an example suppose that person $x \in X$ was born in 1983. Then by the definition of \sim we have that $x \sim y$ if and only if y is also born in 1983 and that $[x]_{\sim}$ is the equivalence class of all people born in 1983. As $[x]_{\sim} \in X/\sim$ then $[x]_{\sim}$ is the set in X/\sim that represents the year 1983. That is the quotient set has taken the set X of all currently alive people who were born in a certain year and turned it into the set of all possible years.

Example 1.5.3. Let X be the set of all possible cars and define the equivalence relation \sim such that $x \sim y$ if and only if x and y are the same colour. We have that \sim is an equivalence relation. Reflexivity is clear as if x is a certain colour then clearly $x \sim x$ will be true. Now if $x \sim y$ then both x and y are the same colour and so $y \sim x$. Finally if $x \sim y$ and $y \sim z$ then x and y are the same colour and so are y and z so it follows that $x \sim z$.

Suppose now that $x \in X$, then the equivalence class $[x]_{\sim}$ is the set where all cars are the same colours. Hence the quotient set X/\sim will be the set of all possible car colours. The quotient set has taken the set of all possible cars and turned it into the set of all possible car colours.

If we had a different relation R where xRy if and only if x and y have exactly two doors then R is also an equivalence relation and X/R would take all of the possible cars X and turn it into the set of all of the cars that have exactly two doors.

These examples show that the quotient set takes a set of objects S and extracts a given property defined by the equivalence relation \sim defined on S . How can we use the quotient set on the equivalence classes of the subtraction tuples?

We have that the the quotient set of \mathbb{N}^2/\sim is given by

$$\mathbb{N}^2/\sim = \{[x]_{\sim} : x \in \mathbb{N}^2\}$$

What do these elements actually look like? Let $(a, b) = x \in \mathbb{N}^2$ and consider the equivalence class $[x]_{\sim}$. Firstly, in the naturals, we know that $0 = 0 - 0$ and more generally that $0 = a - a$ for any $a \in \mathbb{Z}$. Hence $0 \in [(0, 0)]$.

Now, consider $[(a, 0)]$ then we would have that any $(c, d) = y \in [(a, 0)]$ is such that $(a, 0) \sim (c, d)$ if and only if $a - 0 = c - d$. Hence each a is equivalent to some subtraction tuple. Moreover each $(a, 0) = a \in \mathbb{N}$, therefore we have a canonical representation for each element $a \in \mathbb{N}$. What happens if we have a tuple (a, b) where $a \geq b$? We can see that if $(a, b) \sim (c, d)$ then $a + d = c + b$. For example we have that $(0, 3) \sim (1, 4)$ which gives

$$(8, 11) \sim (0, 3) = 8 - 11 = 0 - 38 + 3 = 11$$

$$0 - 3 = 1 - 4 \Rightarrow 0 + 4 = 1 + 3 \Rightarrow 4 = 4$$

Hence we can define a canonical representation for each $(0, a)$ where $a \in \mathbb{N}$. We will write the element $(0, a)$ by $-a$ for each $a \in \mathbb{N}$. We have define the set of Integers.

Definition 1.5.4. *Integers*

Let \mathbb{N}^2 have the equivalence relation \sim defined by $(a, b) \sim (c, d)$ if and only if $a + d = b + c$. We define the set of Integers, denoted \mathbb{Z} , as the quotient set \mathbb{N}^2 / \sim . The set \mathbb{Z} has the form

$$\mathbb{Z} = \{\dots, -4, -3, -2, -1, 0, 1, 2, 3, 4, \dots\} \quad (12)$$

We make two additional definitions based on the definition of the canonical form the equivalence classes

Definition 1.5.5. *Positive Integer*

Let $a \in \mathbb{Z}$. We say that a is a positive integer if and only if $a \in [(b, 0)]$ for some $b \in \mathbb{N}$ with $b \neq 0$.

Definition 1.5.6. *Negative Integer*

Let $a \in \mathbb{Z}$. We say that a is a negative integer if and only if $a \in [(0, b)]$ for some $b \in \mathbb{N}$ with $b \neq 0$.

We can use these two definitions to define an occasionally useful idea.

Definition 1.5.7. *Sign of an integer*

Let $x \in \mathbb{Z}$. We define the sign of x , denoted by $\text{sgn}(x)$ to be the following function

$$\begin{aligned} \text{sgn} : \mathbb{Z} &\rightarrow \{-1, 0, 1\} \\ x &\mapsto \text{sgn}(x) = \begin{cases} 1, & \text{If } x \text{ is a positive integer} \\ -1, & \text{If } x \text{ is a negative integer} \\ 0, & \text{Otherwise} \end{cases} \end{aligned}$$

We also have the following, clear result

Proposition 1.5.2. *The natural numbers are a subset of the integers*

We have that $\mathbb{N} \subseteq \mathbb{Z}$

Proof:

We have that the elements of the equivalence class $[(x, 0)]$ have the form $x - 0 = x \in \mathbb{N}$. Let $a \in \mathbb{N}$ then we have that $a \in [(a, 0)]$. This holds for every $a \in \mathbb{N}$ and so $\mathbb{N} \subseteq \mathbb{Z}$. \square

We will let $[(a, b)]$ be denoted by $[a, b]$ and extend the operations of addition and multiplication to the integers by defining how they work on the equivalence classes.

1.5.2 Extending equality to the integers

Equality for the integers is easy to define.

Definition 1.5.8. *Equality of integers*

Let $x, y \in \mathbb{Z}$ be two integers numbers. We define that two integers are equal, denoted $x = y$ if and only if $x \sim y$. This is the same as saying both x and y belong to the same equivalence class. In the case where $x \not\sim y$, we say that x is not equal to y and write $x \neq y$.

1.5.3 Extending inequality operators to the integers

Inequality operators extend in a natural way.

Definition 1.5.9. *Less than operator*

Let $x, y \in \mathbb{Z}$ where $x \in [a, b]$ and $y \in [c, d]$ for some $a, b, c, d \in \mathbb{N}$. The less than operator, denoted by $x < y$ is defined by the logical proposition

$$< (x, y) = \begin{cases} 1, & \text{If } a + d < b + c \\ 0, & \text{Otherwise} \end{cases}$$

This can equivalently be express as

$$x < y \iff a + d < b + c$$

Definition 1.5.10. *Less than or equal to operator*

Let $x, y \in \mathbb{Z}$ where $x \in [a, b]$ and $y \in [c, d]$ for some $a, b, c, d \in \mathbb{N}$. The less than or equal operator, denoted by $x \leq y$ is defined by the logical proposition

$$\leq (x, y) = \begin{cases} 1, & \text{If } a + d \leq b + c \\ 0, & \text{Otherwise} \end{cases}$$

This can equivalently be express as

$$x \leq y \iff a + d \leq b + c$$

Definition 1.5.11. *Greater than operator*

Let $x, y \in \mathbb{Z}$ where $x \in [a, b]$ and $y \in [c, d]$ for some $a, b, c, d \in \mathbb{N}$. The greater than operator, denoted by $x > y$ is defined by the logical proposition

$$> (x, y) = \begin{cases} 1, & \text{If } a + d > b + c \\ 0, & \text{Otherwise} \end{cases}$$

This can equivalently be express as

$$x > y \iff a + d > b + c$$

Definition 1.5.12. *Greater than or equal to operator*

Let $x, y \in \mathbb{Z}$ where $x \in [a, b]$ and $y \in [c, d]$ for some $a, b, c, d \in \mathbb{N}$. The greater than or equal to operator, denoted by $x \geq y$ is defined by the logical proposition

$$\geq (x, y) = \begin{cases} 1, & \text{If } a + d \geq b + c \\ 0, & \text{Otherwise} \end{cases}$$

This can equivalently be express as

$$x \geq y \iff a + d \geq b + c$$

1.5.4 Extending addition to the integers

We have an understanding of addition on the natural numbers, mainly the recursive definition given by

$$\begin{aligned} &+ : \mathbb{N}^2 \rightarrow \mathbb{N} \\ (m, n) &\mapsto + (m, n) = \begin{cases} m + 0 = m, & \text{If } n = 0 \\ m + S(n) = S(m + n), & \text{If } n \neq 0 \end{cases} \end{aligned}$$

Now if we take $a, b \in \mathbb{Z}$ with a, b being positive integers then we have that $a \in [(a, 0)]$ and $b \in [(b, 0)]$. We then have that $a + b$ will be in $[(a + b, 0)]$. Now suppose that $a, b \in \mathbb{N}$ with a, b being negative integers then we have that $a \in [(0, a)]$ and $b \in [(0, b)]$. Intuitively we know that $-2 + -3 = -5$ so we want these to add like in the positive integer case. This is to say we have $a + b$ will be in the class $[(0, a + b)]$.

We can combine these two observations to define addition on the integers.

Definition 1.5.13. *Addition on the Integers*

Let $x, y \in \mathbb{Z}$ with $x = (a, b)$ and $y = (c, d)$. We define addition on the integers by

$$[a, b] + [c, d] = [a + c, b + d] \quad (13)$$

To check this definition makes sense consider $x = 4, y = 3$. Both x and y belong to some equivalence class, for example $x \in [(5, 1)]$ and $y \in [(8, 5)]$. Then we have that $x + y = 7$ and

$$(5, 1) + (8, 5) = (5 + 8, 1 + 5) = (13, 6) \Rightarrow 13 - 6 = 7$$

1.5.5 Extending multiplication to the integers

We also extend multiplication to the integers. We have the definition of multiplication on the naturals given by

$$\begin{aligned} * : \mathbb{N} \times \mathbb{N} &\rightarrow \mathbb{N} \\ (m, n) &\mapsto * (m, n) = \begin{cases} m * 0 = 0, & \text{If } n = 0 \\ m * S(n) = m * n + m, & \text{If } n \neq 0 \end{cases} \end{aligned}$$

As before, if we take $x, y \in \mathbb{Z}$ with x, y being positive integers then we have that $x \in [(x, 0)]$ and $b \in [(x, 0)]$ we have that $x * y \in [(x * y, 0)]$.

Suppose that $x, y \in \mathbb{Z}$ with $x = (a, b)$ and $y = (c, d)$. We have that

$$\begin{aligned} (a - b) * (c - d) &= (a - b) c - (a - b) d \\ &= ac - bc - (ad - bd) \\ &= ac - bc + bd - ad \\ &= ac + bd - bc - ad \\ &= ac + bd - (ad + bc) \end{aligned}$$

This is $(a, b) * (c, d) = (ac + bd, ad + bc)$

This will be the definition of multiplication of the integers.

Definition 1.5.14. *Multiplication on the Integers*

Let $x, y \in \mathbb{Z}$ with $x = (a, b)$ and $y = (c, d)$. We define multiplication on the integers by

$$[a, b] * [c, d] = [ac + bd, ad + bc] \quad (14)$$

1.5.6 Closure properties of addition and multiplication

As with the natural numbers we need to show that the operations of addition and multiplication are closed. Additionally we want to prove our claim at the start of this section that the integers allow us to completely perform subtraction.

Theorem 1.5.1. *Addition and multiplication on the integers are well-defined operators and closed*

We have that $\forall x, y \in \mathbb{Z}$ that

1. $x + y \in \mathbb{Z}$

2. $x * y \in \mathbb{Z}$

Proof:

1. $x + y \in \mathbb{Z}$:

We need to show that if $(a, b) \sim (a', b')$ and $(c, d) \sim (c', d')$ then $(a + c, b + d) \sim (a' + c', b' + d')$ as this will show equivalent elements produce the same result when added and therefore integer addition is well-defined.

We have by definition that $(a, b) \sim (a', b')$ that $a + b' = a' + b$, likewise we have $(c, d) \sim (c', d')$ gives $c + d' = c' + d$.

Now, we have that

$$\begin{aligned} a + b' + c + d' &= a' + b + c' + d \\ a + c + b' + d' &= a' + c' + b + d \\ \Rightarrow (a + c, b + d) &\sim (a' + c', b' + d') \end{aligned}$$

Hence $[(a + c, b + d)] = [(a' + c', b' + d')]$ and so addition is well-defined.

It is left to prove closure. Let $x, y \in \mathbb{Z}$ with $x = (a, b)$ and $y = (c, d)$. By definition of integer addition we have that $x + y = (a + c, b + d)$ and moreover we have $a + c \in \mathbb{N}$ and $b + d \in \mathbb{N}$. Hence $(a + c, b + d) \in [a + c, b + d]$ and therefore $x + y \in \mathbb{Z}$ showing closure.

2. $x * y \in \mathbb{Z}$:

As with addition we need to show that if $(a, b) \sim (a', b')$ and $(c, d) \sim (c', d')$ then $(a, b) * (c, d) \sim (a', b') * (c', d')$. As before we have that

We have that

$$(a, b) * (c, d) = (ac + bd, ad + bc) \iff ac + bd - (ad + bc)$$

Now as $(a, b) \sim (a', b')$ then $a + b' = b + a'$ and $(c, d) \sim (c', d')$ then $c + d' = d + c'$. Hence

$$\begin{aligned} ac + bd - (ad + bc) &= (ac - ad) + (bd - bc) \\ &= a(c - d) + b(d - c) \\ &= a(c' - d') + b(d' - c'), \text{ By assumption } ac + d' = d + c' \Rightarrow c - d = c' - d' \\ &= ac' - ad' + bd' - bc' \\ &= (ac' - bc') + (bd' - ad'), \text{ By commutativity of the Naturals} \\ &= c'(a - b) + d'(b - a) \\ &= c'(a' - b') + d'(b' - a'), \text{ By assumption as } a + b' = b + a' \Rightarrow a - b = a' - b' \\ &= (c'a' - c'b') + (d'b' - d'a') \\ &= c'a' - c'b' + d'b' - d'a' \\ &= a'c' - b'c' + b'd' - a'd', \text{ By commutativity of the Naturals} \\ &= (a'c + b'd') - b'c' - a'd' \\ &= (a'c + b'd') - (a'd' + b'c'), \text{ By lemma 1.5.1} \end{aligned}$$

This shows that multiplication is well-defined. It is left to show closure. Let $x, y \in \mathbb{Z}$ with $x = (a, b)$ and $y = (c, d)$. By the definition of multiplication on the integers we have that $x * y = (ac + bd, ad + bc)$ with $ac + bd \in \mathbb{N}$ and $ad + bc \in \mathbb{N}$. Hence we conclude that $(ac + bd, ad + bc) \in [ac + bd, ad + bc]$, and so by definition $x * y \in \mathbb{Z}$.

The result is shown. \square

Now that we have shown closure we can deduce an immediate property.

Proposition 1.5.3. *Multiplication of an integer by -1*

Let $x \in \mathbb{Z}$ where $x \in [a, b]$ for some $a, b \in \mathbb{N}$. We have that

1. $-1 * x = -1 * (a, b) = (b, a)$
2. $x * -1 = (a, b) * -1 = (b, a)$

Proof:

1. $-1 * x = -1 * (a, b) = (b, a)$:

We have that $-1 \in [0, 1]$ and so

$$\begin{aligned} -1 * x &= (0, 1) * (a, b) \\ &= (0 * a + 1 * b, 0 * b + 1 * a) \\ &= (b, a) \end{aligned}$$

2. $x * -1 = (a, b) * -1 = (b, a)$:

Likewise we have

$$\begin{aligned} x * -1 &= (a, b) * (0, 1) \\ &= (a * 0 + b * 1, a * 1 + b * 0) \\ &= (b, a) \end{aligned}$$

As required. \square

Corollary 1.5.1. *Multiplication of a positive integer by -1 makes it a negative integer and multiplication of a negative integer by -1 makes it a positive integer*

1. *If x is a positive integer then $-1 * x$ is a negative integer.*
2. *If x is a negative integer then $-1 * x$ is a positive integer.*

Proof:

*By definition if $x \in \mathbb{Z}$ is positive then $x \in [a, 0]$ for some $a \in \mathbb{N}$. By proposition 1.5.3 we have that $-1 * x = (0, a) = x * -1$, which is by definition a negative integer.*

*Likewise if $x \in \mathbb{Z}$ is negative then $x \in [0, a]$ for some $a \in \mathbb{N}$. By proposition 1.5.3 we have that $-1 * x = (a, 0) = x * -1$, which is by definition a positive integer.*

\square

1.5.7 Associativity of integer addition and multiplication

The associativity of addition and multiplication of the naturals also extends to the integers.

Theorem 1.5.2. *Let $x, y, z \in \mathbb{Z}$. We have that*

1. $x + (y + z) = (x + y) + z$
2. $x(yz) = (xy)z$

Proof:

1. $x + (y + z) = (x + y) + z$:

Let $x, y, z \in \mathbb{Z}$ be such that $x = (a, b)$, $y = (c, d)$ and $z = (e, f)$ where $a, b, c, d, e, f \in \mathbb{N}$ and we have that $(a, b) \in [a, b]$, $(c, d) \in [c, d]$ and $(e, f) \in [e, f]$. We have that

$$\begin{aligned}
 x + (y + z) &= (a, b) + ((c, d) + (e, f)) \\
 &= (a, b) + (c + e, d + f) \\
 &= (a + (c + e), b + (d + f)) \\
 &= ((a + c) + e, (b + d) + f), \text{ By associativity of addition for natural numbers} \\
 &= (a + c, b + d) + (e, f) \\
 &= ((a, b) + (c, d)) + (e, f) \\
 &= (x + y) + z
 \end{aligned}$$

Which shows associativity of addition.

2. $x(yz) = (xy)z$:

As with addition, let $x, y, z \in \mathbb{Z}$ be such that $x = (a, b)$, $y = (c, d)$ and $z = (e, f)$ where $a, b, c, d, e, f \in \mathbb{N}$ and we have that $(a, b) \in [a, b]$, $(c, d) \in [c, d]$ and $(e, f) \in [e, f]$. We then have that

$$\begin{aligned}
 x(yz) &= (a, b) * ((c, d)(e, f)) \\
 &= (a, b)(ce + df, cf + de) \\
 &= (a(ce + df) + b(cf + de), a(cf + de) + b(ce + df)) \\
 &= (ace + adf + bcf + bde, acf + ade + bce + bdf) \\
 &= (ace + bde + adf + bcf, acf + bdf + ade + bce), \text{ By associativity of addition for natural numbers} \\
 &= ((ac + bd)e + (ad + bc)f, (ac + bd)f + (ad + bc)e) \\
 &= (ac + bd, ad + bc)(e, f) \\
 &= ((a, b)(c, d))(e, f) \\
 &= (xy)z
 \end{aligned}$$

Showing associativity of multiplication.

The result follows. \square

1.5.8 Commutativity of integer addition and multiplication

As with the naturals, addition and multiplication in the integers both satisfy commutativity.

Theorem 1.5.3. *Addition and multiplication are commutative*

For all $x, y \in \mathbb{Z}$ we have that

1. $x + y = y + x$

2. $xy = yx$

Proof:

1. $x + y = y + x$:

Let $x, y \in \mathbb{Z}$. By definition we have that $x \in [a, b]$ and $y \in [c, d]$ for some $a, b, c, d \in \mathbb{N}$. Let $x = (a, b)$ and $y = (c, d)$. We then have by definition of addition that

$$\begin{aligned}
x + y &= (a, b) + (c, d) \\
&= (a + c, b + d) \\
&= (c + a, d + b), \text{ By commutativity of addition for natural numbers} \\
&= (c, d) + (a, b) &= y + x
\end{aligned}$$

Showing commutativity holds for addition in the integers.

2. $xy = yx$:

Let $x, y \in \mathbb{Z}$ by definition we have that $x \in [a, b]$ and $y \in [c, d]$ for some $a, b, c, d \in \mathbb{N}$. So let $x = (a, b)$ and $y = (c, d)$. By definition of multiplication we have

$$\begin{aligned}
xy &= (a, b) * (c, d) \\
&= (ac + bd, ad + bc) \\
&= (ca + db, da + bc), \text{ By commutativity of multiplication of the naturals} \\
&= (ca + db, da + bc), \text{ By commutativity of addition of the naturals} \\
&= (c, d) * (a, b) \\
&= yx
\end{aligned}$$

Showing commutativity for integer multiplication.

The result has been shown. \square

1.5.9 Multiplication distributes over addition

Another result that extends from the naturals is that multiplication distributes over addition.

Theorem 1.5.4. *Multiplication distributes over addition*

For all $x, y, z \in \mathbb{Z}$ we have that

1. $x(y + z) = xy + xz$
2. $(y + z)x = yx + zx = xy + xz$

Proof:

Let $x, y, z \in \mathbb{Z}$ then $x \in [a, b]$, $y \in [c, d]$ and $z \in [e, f]$ for some $a, b, c, d, e, f \in \mathbb{N}$.

So let $x = (a, b)$, $y = (c, d)$ and $z = (e, f)$.

1. $x(y + z) = xy + xz$:

We have that

$$\begin{aligned}
x(y + z) &= (a, b)((c, d) + (e, f)) \\
&= (a, b)(c + e, d + f) \\
&= (a(c + e) + b(d + f), a(d + f) + b(c + e)) \\
&= (ac + ae + bd + bf, ad + af + bc + be) \\
&= (ac + bd + ae + bf, ad + bc + af + be) \\
&= (ac + bd, ad + bc) + (ae + bf, af + be) \\
&= (a, b)(c, d) + (a, b)(e, f) \\
&= xy + xz
\end{aligned}$$

2. $(y + z)x = yx + zx = xy + xz$:

Now that we have the previous part the proof of this part is quick. We have

$$\begin{aligned}(y + z)x &= x(y + z), \text{ By commutativity of multiplication} \\ &= xy + xz, \text{ By part 1.} \\ &= yx + zx, \text{ By commutativity of multiplication}\end{aligned}$$

As required. \square

1.5.10 The Zero and Identity laws

The zero and identity laws from the naturals extend to the integers.

Theorem 1.5.5. *The zero and Identity laws*

Let $x \in \mathbb{Z}$. We have that

1. $x + 0 = x = 0 + x$

2. $1 * x = x = x * 1$

Proof:

Let $x \in \mathbb{Z}$ then we have that $x = (a, b)$ for some $a, b \in \mathbb{N}$

1. $x + 0 = x = 0 + x$:

We have that $0 \in [0, 0]$. Hence we have that

$$x + 0 = (a, b) + (0, 0) = (a + 0, b + 0) = (a + b) = (0 + a, 0 + b) = (0, 0) + (a, b) = 0 + x$$

2. $x * 1 = x = 1 * x$:

As $1 \in [1, 0]$ then

$$\begin{aligned}x * 1 &= (a, b) * (1, 0) \\ &= (a * 1 + b * 0, b * 1 + a * 0) \\ &= (a + 0, b + 0) \\ &= (a, b) = x \\ &= (1 * a + 0 * b, 0 * a + 1 * b) \\ &= (1, 0) (a, b) \\ &= 1 * x\end{aligned}$$

The result follows. \square

1.5.11 Extending subtraction to the integers

As we have a notion of subtraction on the naturals, we can ask about extending this to the integers. We defined subtraction on the naturals as follows. Let $n, m \in \mathbb{N}$ such that $n \leq m$. Let $d \in \mathbb{N}$ such that $n = m + d$. We define subtraction by

$$d = n - m$$

Where we called d the difference between n and m . We also have the notion of a positive and negative integer. Recall that $x \in \mathbb{Z}$ is a positive integer if and only if $x \in [0, \infty)$. Let $x \in \mathbb{Z}$. We say that x is a positive integer if and only if $x \in [(b, 0)]$ for some $b \in \mathbb{N}$. Likewise x is a negative integer if and only if $x \in [(0, b)]$ for some $b \in \mathbb{N}$. In order to extend subtraction to the integers we need to consider a few things.

Definition 1.5.15. *Negation of a natural number*

Let $x \in \mathbb{Z}$ so that x is a positive integer, i.e. a natural number. We define the negation of x , denoted $-x$ by

$$-x = -1 * x = (0, 1) * x$$

where $(0, 1) \in [(0, -1)]$. That is $(0, 1)$ is an element of the equivalence class $[(0, 1)]$ which represents all possible elements that are -1 .

We can extend this result to include a general integer.

Proposition 1.5.4. *Negation of an integer*

Let $x \in \mathbb{Z}$ so that $x \in [(a, b)]$ for some $a, b \in \mathbb{N}$. We have that

$$-1 * x = -1 * (a, b) = (b, a)$$

Proof:

Let $x \in \mathbb{Z}$ be as given by the hypothesis. We have that

$$\begin{aligned} -1 * x &= -1 * (a, b) \\ &= (0, 1) * (a, b) \\ &= (0 * a + b * 1, 0 * b + 1 * a) \\ &= (b, a) \end{aligned}$$

As required. \square

In light of this, we can define subtraction for integers.

Definition 1.5.16. *Integer subtraction*

Let $x, y \in \mathbb{Z}$. We define the subtraction of y from x , denoted $x - y$ by

$$x - y = x + (-y) = x + (-1 * y)$$

We immediately get that subtraction is closed, from the fact that both addition and multiplication are closed. We do not have associativity of subtraction in general.

Proposition 1.5.5. *Integer subtraction is not associative*

Let $x, y, z \in \mathbb{Z}$. We have that

$$x - (y - z) \neq (x - y) - z$$

Proof:

Let $x = 2, y = 4$ and $z = 6$, we have $x \in [2, 0], y \in [4, 0]$ and $z \in [0, 6]$ so $x \in (2, 0), y \in (4, 0)$ and $z \in (0, 6)$. We have that

$$\begin{aligned} x - (y - z) &= (2, 0) - ((4, 0) - (6, 0)) \\ &= (2, 0) - ((4, 0) + (-1 * (6, 0))) \\ &= (2, 0) - ((4, 0) + (0, 6)) \\ &= (2, 0) - (4, 6) \\ &= (2, 0) + (-1 * (4, 6)) \\ &= (2, 0) + (6, 4) \\ &= (8, 4) \end{aligned}$$

On the other side we have

$$\begin{aligned}
(x - y) - z &= ((2, 0) - (4, 0)) - (6, 0) \\
&= ((2, 0) + (-1 * (4, 0))) - (6, 0) \\
&= ((2, 0) + (0, 4)) - (6, 0) \\
&= (2, 4) - (6, 0) \\
&= (2, 4) + (-1 * (6, 0)) \\
&= (2, 4) + (0, 6) \\
&= (2, 10)
\end{aligned}$$

Clearly $(8, 4) \neq (2, 10)$. Indeed they are not even equivalent. Suppose that $(8, 4) \sim (2, 10)$ then we have that $8 + 10 = 4 + 2$. However $18 \neq 6$. \square

We can also immediately see the following result, which allows us to formally show that subtraction is an inverse to addition.

Proposition 1.5.6. *Subtracting an integer from itself gives zero*

Let $x \in \mathbb{Z}$. We have that

$$x - x = 0$$

Proof:

Let $x \in \mathbb{Z}$ where $x \in [a, b]$ for some $a, b \in \mathbb{N}$. We have

$$\begin{aligned}
x - x &= (a, b) - (a, b) \\
&= (a, b) + (b, a) \\
&= (a + b, b + a)
\end{aligned}$$

It is left to show that $(a + b, b + a) \sim (0, 0)$. Indeed

$$(a + b) + 0 = (b + a) + 0 \Rightarrow a + b = b + a$$

The result is shown. \square

1.5.12 The cancellation laws

We can now deduce that the cancellation laws also extend to the integers.

Theorem 1.5.6. *The cancellation laws*

Let $x, y, z \in \mathbb{Z}$.

- 1. If $x + y = x + z$ then we have $y = z$.*
- 2. For $x \neq 0$, if $xy = xz$ then we have that $y = z$*

Proof:

- 1. If $x + y = x + z$ then we have $y = z$:*

Let $x, y, z \in \mathbb{Z}$. We have that

$$\begin{aligned}
x + y &= x + z \\
\Rightarrow -x + x + y &= -x + x + z, \text{ Adding the negative of } x \text{ to both sides} \\
\Rightarrow (-x + x) + y &= (-x + x) + z, \text{ Associativity of integers} \\
\Rightarrow 0 + y &= 0 + z, \text{ By proposition 1.5.6} \\
\Rightarrow y &= z
\end{aligned}$$

2. For $x \neq 0$, if $xy = xz$ then we have that $y = z$:

Let $x, y, z \in \mathbb{Z}$ where $x \neq 0$. Suppose that $x \in [a, b]$, $y \in [c, d]$ and $z \in [e, f]$. We have

$$\begin{aligned} xy &= (a, b)(c, d) = (ac + bd, ad + bc) \\ xz &= (a, b)(e, f) = (ae + bf, af + be) \end{aligned}$$

Now assume $xy = xz$ then we have that $(ac + bd, ad + bc) \sim (ae + bf, af + be)$ which is to say

$$ac + bd + af + be = ae + bf + ad + bc$$

Observe that

$$\begin{aligned} ac + bd + af + be &= a(c + f) + b(d + e) \\ ae + bf + ad + bc &= a(e + d) + b(f + c) \end{aligned}$$

Which gives

$$a(c + f) + b(d + e) = a(e + d) + b(f + c)$$

There are now two cases to consider, $a < b$ and $a > b$. Firstly suppose that $a < b$ then we can write that $b = a + h$ for some $h > 0$, this is well-defined as $a, b \in \mathbb{N}$. We then have

$$\begin{aligned} a(c + f) + b(d + e) &= a(e + d) + b(f + c) \\ a(c + f) + (a + h)(d + e) &= a(e + d) + (a + h)(f + c) \\ a(c + f) + a(d + e) + h(d + e) &= a(e + d) + a(f + c) + h(f + c) \\ a(d + e) + h(d + e) &= a(e + d) + h(f + c), \text{ Cancelling } a(c + f) \\ h(d + e) &= h(f + c), \text{ Cancelling } a(d + e) \\ (d + e) &= (f + c), \text{ Cancelling } h \end{aligned}$$

Now as $d + e = f + c$ we have that $c - d = e - f \Rightarrow (c, d) \sim (e, f)$ which is the same as saying $y = z$.

Now if $a > b$ then we write $b = a - h$ for some $h > 0$, again being well-defined as $a, b \in \mathbb{N}$. Thus

$$\begin{aligned} a(c + f) + b(d + e) &= a(e + d) + b(f + c) \\ a(c + f) + (a - h)(d + e) &= a(e + d) + (a - h)(f + c) \\ a(c + f) + a(d + e) - h(d + e) &= a(e + d) + a(f + c) - h(f + c) \\ a(d + e) - h(d + e) &= a(e + d) - h(f + c), \text{ Cancelling } a(c + f) \\ -h(d + e) &= -h(f + c), \text{ Cancelling } a(d + e) \\ (f + c) &= (d + e), \text{ By adding each side to the other and cancelling } h \end{aligned}$$

As $f + c = d + e$ then we have by similar logic to before the $y = z$

The result is shown. \square

1.5.13 Extending the summation and product notations to integers

Summation and product notation has been defined on the naturals. As with the theme of this section the notations extend in a natural way to integers. As before we need to define a few things.

Let $z \in \mathbb{Z}^{n+m+1}$ be an ordered $n+m+1$ tuple of integers where $z = (z_{-m}, z_{-m+1}, \dots, z_{-1}, z_0, z_1, z_2, \dots, z_n)$ and define $\mathbb{Z}_m^n = \{-m, -m+1, -m+2, \dots, -1, 0, 1, \dots, n-1, n\}$. Define $f : \mathbb{Z}_m^n \rightarrow \mathbb{Z}$ by

$$\begin{aligned} f : \mathbb{Z}_m^n &\rightarrow \mathbb{Z} \\ i &\mapsto f(i) = z_i \end{aligned}$$

As before, f simply maps gets the value of z_i from the ordered tuple z .

Definition 1.5.17. *Summation notation for the integers*

Let $z \in \mathbb{Z}^{n+m+1}$ be ordered $n+m+1$ tuple of integers where $z = (z_{-m}, z_{-m+1}, \dots, z_{-1}, z_0, z_1, z_2, \dots, z_n)$. Define \mathbb{Z}_m^n by $\mathbb{Z}_m^n = \{-m, -m+1, -m+2, \dots, -1, 0, 1, \dots, n-1, n\}$. Let $f : \mathbb{Z}_m^n \rightarrow \mathbb{Z}$ defined by

$$\begin{aligned} f : \mathbb{Z}_m^n &\rightarrow \mathbb{Z} \\ i &\mapsto f(i) = z_i \end{aligned}$$

We define the summation notation for integers by

$$\sum_{i=-m}^n f(i) = f(-m) + f(-m+1) + \dots + f(-1) + f(0) + f(1) + \dots + f(n)$$

Alternatively this is written

$$\sum_{i=-m}^n z_i = z_{-m} + z_{-m+1} + \dots + z_{-1} + z_0 + z_1 + \dots + z_n$$

We have that i is called the index of summation and that $i = -m$ is the starting index of the summation, and n the ending index of the summation. If $z = \emptyset$ then we define the summation to be 0 and call a summation an empty sum.

We can also define the summation of some subset of \mathbb{Z}_m^n which allows for starting a summation at some starting point other than $i = -m$. Let $T \subseteq \mathbb{Z}_m^n$. We define the summation over the set T by

$$\sum_{i \in T} z_i$$

If we have a mapping $g : \mathbb{Z} \rightarrow \mathbb{Z}$ we can define a summation over g by

$$\sum_{i \in T} g(z_i)$$

Finally we can define a summation over a predicate $P(i)$ for $i \in T$ by

$$\sum_{P(i)} g(z_i)$$

where we take the sum of the $g(z_i)$ for the i that satisfy the predicate P . We note that if we have $k > n$ for some $k \in \mathbb{N}$ then the sum

$$\sum_{i=k}^n z_i = 0$$

The proprieties shown for summations with natural numbers also extend to the integer version.

Proposition 1.5.7. *Properties of summation notation*

Let $n, m \in \mathbb{Z}$ such that $m < n$. Let $s, t \in \mathbb{N}^{n+m+1}$ and let $c \in \mathbb{Z}$.

Let $a, b \in \mathbb{Z}$ with $m < a < b < n$. Define $A = \mathbb{Z}_a^b$ and define

$$B = \mathbb{Z}_m^n \setminus A = \{-m, -m+1, \dots, a-1, b+1, \dots, n-1, n\}$$

so that $A \cup B = \mathbb{Z}_m^n$. Let $k \in \mathbb{Z}$ be the starting index summation such that $k < n$. We have the following properties hold.

$$1. \sum_{i=-m}^n s_i = \sum_{i \in A} s_i + \sum_{i \in B} s_i = \sum_{i=-m}^{-1} s_i + \sum_{i=0}^n s_i$$

$$2. \sum_{i=k}^n s_i = \sum_{i=k}^d s_i + \sum_{i=d+1}^n s_i$$

$$3. \sum_{i=k}^n c * s_i = c * \sum_{i=k}^n s_i \text{ for some } c \in \mathbb{Z}$$

$$4. \sum_{i=k}^n c = c(n+1-k) \text{ for some } c \in \mathbb{Z}$$

$$5. \sum_{i=k}^n s_i + t_i = \sum_{i=k}^n s_i + \sum_{i=k}^n t_i$$

Proof:

$$1. \sum_{i=-m}^n s_i = \sum_{i \in A} s_i + \sum_{i \in B} s_i = \sum_{i=-m}^{-1} s_i + \sum_{i=0}^n s_i:$$

This follows by applying the definition. We have that

$$\begin{aligned} \sum_{i=-m}^n s_i &= s_{-m} + s_{-m+1} + s_{-m+2} + \dots + s_{-1} + s_0 + s_1 + \dots + s_{n-1} + s_n \\ &= (s_{-m} + s_{-m+1} + s_{-m+2} + \dots + s_{-1}) + (s_0 + s_1 + \dots + s_{n-1} + s_n) \\ &= \sum_{i=-m}^{-1} s_i + \sum_{i=0}^n s_i \end{aligned}$$

Additionally note that

$$\begin{aligned} \sum_{i=-m}^n s_i &= s_{-m} + s_{-m+1} + s_{-m+2} + \dots + s_{-1} + s_0 + s_1 + \dots + s_{n-1} + s_n \\ &= (s_{-m} + s_{-m+1} + s_{-m+2} + \dots + s_{a-2} + s_{a-1}) + (s_a + s_{a+1} + \dots + s_{b-1} + s_b) \\ &\quad + (s_{b+1} + s_{b+2} + \dots + s_{n-1} + s_n) \\ &= (s_{-m} + s_{-m+1} + s_{-m+2} + \dots + s_{a-2} + s_{a-1}) + (s_{b+1} + s_{b+2} + \dots + s_{n-1} + s_n) \\ &\quad + (s_a + s_{a+1} + \dots + s_{b-1} + s_b) \\ &= \sum_{i \in B} s_i + \sum_{i \in A} s_i = \sum_{i \in A} s_i + \sum_{i \in B} s_i \end{aligned}$$

$$2. \sum_{i=k}^n s_i = \sum_{i=k}^d s_i + \sum_{i=d+1}^n s_i.$$

The proof is similar to part 1, replacing $-m$ by k .

$$3. \sum_{i=k}^n c * s_i = c * \sum_{i=k}^n s_i \text{ for some } c \in \mathbb{Z}$$

We have by definition that

$$\sum_{i=k}^n c * s_i = c * s_k + c * s_{k+1} + c * s_{k+2} + \cdots + c * s_n$$

By multiplication distributing over addition we have

$$\sum_{i=1}^n c * s_i = c * s_k + c * s_{k+1} + c * s_{k+2} + \cdots + c * s_n = c(s_k + s_{k+1} + \cdots + s_n) = c * \sum_{i=k}^n s_i$$

$$4. \sum_{i=k}^n c = c(n+1-k) \text{ for some } c \in \mathbb{Z}$$

If $n > 0$ and $k \geq 0$ then the result is the same as for natural numbers. So suppose that $k < 0$. Consider the following set of the indices given by

$$S = \{k, k+1, k+2, \dots, -1, 0, 1, \dots, n-1, n\}$$

We have that the cardinality of S is $n+1-k$. Indeed consider the following mapping

$$\begin{aligned} f : S &\rightarrow \mathbb{N} \\ s &\mapsto f(s) = s - k \end{aligned}$$

Define the mapping $g : S \rightarrow \text{Image}(f)$ then we have that g is a bijection. Suppose that $g(x) = g(y)$ for some $x, y \in S$ then

$$\begin{aligned} g(x) &= g(y) \\ x - k &= y - k \\ x &= y \end{aligned}$$

showing injectivity. Now as g is a mapping from S to the image of f we have by proposition 1.2.15 that g is surjective. Hence we conclude that g is a bijection.

Now we have that

$$\begin{aligned} \text{Image}(f) &= \{f(x) : x \in S\} \\ &= \{k-k, (k+1)-k, (k+2)-k, \dots, -1-k, 0-k, 1-k, \dots, (n-1)-k, n-k\} \\ &= \{0, 1, 2, \dots, k-1, k, k-1, \dots, n-1-k, n-k\} \end{aligned}$$

Hence $|S| = |\text{Image}(f)| = n-k+1$. Hence the sum is adding c to itself $n+1-k$ times. This is to say

$$\sum_{i=k}^n c = c(n+1-k)$$

$$5. \sum_{i=k}^n s_i + t_i = \sum_{i=k}^n s_i + \sum_{i=k}^n t_i:$$

This follows by the definition. We have

$$\begin{aligned} \sum_{i=k}^n s_i + t_i &= (s_k + t_k) + (s_{k+1} + t_{k+1}) + \dots \\ &\quad + (s_{-1} + t_{-1}) + (s_0 + t_0) + (s_1 + t_1) + \dots + (s_{n-1} + t_{n-1}) + (s_n + t_n) \\ &= (s_k + s_{k+1} + \dots + s_{-1} + s_0 + s_1 + \dots + s_{n-1} + s_n) + \\ &\quad + (t_k + t_{k+1} + \dots + t_{-1} + t_0 + t_1 + \dots + t_{n-1} + t_n) \\ &= \sum_{i=k}^n s_i + \sum_{i=k}^n t_i \end{aligned}$$

□

We make a similar definition for product notation.

Definition 1.5.18. *Product notation for the integers*

Let $z \in \mathbb{Z}^{n+m+1}$ be ordered $n+m+1$ tuple of integers where $z = (z_{-m}, z_{-m+1}, \dots, z_{-1}, z_0, z_1, z, \dots, z_n)$. Define \mathbb{Z}_m^n by $\mathbb{Z}_m^n = \{-m, -m+1, -m+2, \dots, -1, 0, 1, \dots, n-1, n\}$. Let $f : \mathbb{Z}^{n+m+1} : \mathbb{Z}$ defined by

$$\begin{aligned} f : \mathbb{Z}^{n+m+1} &\rightarrow \mathbb{Z} \\ i &\mapsto f(i) = z_i \end{aligned}$$

We define the summation notation for integers by

$$\prod_{i=-m}^n f(i) = f(-m) * f(-m+1) * \dots * f(-1) * f(0) * f(1) * \dots + f(n)$$

Alternatively this is written

$$\prod_{i=-m}^n z_i = z_{-m} * z_{-m+1} * \dots * z_{-1} * z_0 * z_1 * \dots * z_n$$

We have that i is called the index of the product and that $i = -m$ is the starting index of the product, and n the ending index of the product. If $z \in \emptyset$ then we define the product to be 1 and call a product an empty sum.

We can also define the product of some subset of \mathbb{Z}_m^n which allows for starting a product at some starting point other than $i = -m$. Let $T \subseteq \mathbb{Z}_m^n$. We define the product over the set T by

$$\prod_{i \in T} z_i$$

If we have a mapping $g : \mathbb{Z} \rightarrow \mathbb{Z}$ we can define a product over g by

$$\prod_{i \in T} g(z_i)$$

Finally we can define a product over a predicate $P(i)$ for $i \in T$ by

$$\prod_{P(i)} g(z_i)$$

where we take the sum of the $g(z_i)$ for the i that satisfy the predicate P . We note that if we have $k > n$ for some $k \in \mathbb{N}$ then the product

$$\prod_{i=k}^n z_i = 1$$

Proposition 1.5.8. *Properties of product notation*

Let $n, m \in \mathbb{Z}$ such that $m < n$. Let $s, t \in \mathbb{Z}^{n+m+1}$ and let $c \in \mathbb{Z}$. Let $a, b \in \mathbb{Z}$ so that $m < a < b < n$. Define $A = \mathbb{Z}_a^b$ and define

$$B = \mathbb{Z}_m^n \setminus A = \{-m, -m+1, \dots, a-1, b+1, \dots, n-1, n\}$$

so that $A \cup B = \mathbb{Z}_m^n$. Let $k \in \mathbb{Z}$ be the lower index of the product.

We have that the following properties hold.

1. $\prod_{i=-m}^n s_i = \prod_{i \in A} s_i * \prod_{i \in B} s_i = \prod_{i=-m}^{-1} s_i * \prod_{i=0}^n s_i$
2. $\prod_{i=k}^n s_i = \prod_{i=k}^m s_i * \prod_{i=m+1}^n s_i$
3. $\prod_{i=k}^n s_i t_i = \prod_{i=k}^n s_i \prod_{i=1}^n t_i$

Proof:

1. $\prod_{i=-m}^n s_i = \prod_{i \in A} s_i * \prod_{i \in B} s_i = \prod_{i=-m}^{-1} s_i * \prod_{i=0}^n s_i.$

This follows by the definition of the product. We have that

$$\begin{aligned} \prod_{i=-m}^n s_i &= s_{-m} * s_{-m+1} * s_{-m+2} * \dots * s_{-1} * s_0 * s_1 * \dots * s_{n-1} * s_n \\ &= (s_{-m} * s_{-m+1} * s_{-m+2} * \dots * s_{-1}) * (s_0 * s_1 * \dots * s_{n-1} * s_n) \\ &= \prod_{i=-m}^{-1} s_i * \prod_{i=0}^n s_i \end{aligned}$$

Likewise we have

$$\begin{aligned} \prod_{i=-m}^n s_i &= s_{-m} * s_{-m+1} * s_{-m+2} * \dots * s_{-1} * s_0 * s_1 * \dots * s_{n-1} * s_n \\ &= (s_{-m} * s_{-m+1} * s_{-m+2} * \dots * s_{a-2} * s_{a-1}) * (s_a * s_{a+1} * \dots * s_{b-1} * s_b) \\ &\quad * (s_{b+1} * s_{b+2} * \dots * s_{n-1} * s_n) \\ &= (s_{-m} * s_{-m+1} * s_{-m+2} * \dots * s_{a-2} * s_{a-1}) * (s_{b+1} * s_{b+2} * \dots * s_{n-1} * s_n) \\ &\quad * (s_a * s_{a+1} * \dots * s_{b-1} * s_b) \\ &= \prod_{i \in B} s_i * \prod_{i \in A} s_i = \prod_{i \in A} s_i * \prod_{i \in B} s_i \end{aligned}$$

$$2. \prod_{i=k}^n s_i = \prod_{i=k}^m s_i * \prod_{i=m+1}^n s_i.$$

The proof is similar to part 1. We replace $-m$ with k .

$$3. \prod_{i=k}^n s_i t_i = \prod_{i=k}^n s_i \prod_{i=1}^n t_i:$$

Observe that

$$\begin{aligned} \prod_{i=k}^n s_i t_i &= s_k t_k * s_{k+1} t_{k+1} * s_{k+2} t_{k+2} * \cdots * s_{-1} t_{-1} * s_0 t_0 * s_1 t_1 * \cdots * s_{n-1} t_{n-1} * s_n t_n \\ &= (s_k * s_{k+1} * s_{k+2} * \cdots * s_{-1} * s_0 * s_1 * \cdots * s_{n-1} * s_n) \\ &\quad * (t_k * t_{k+1} * t_{k+2} * \cdots * t_{-1} * t_0 * t_1 * \cdots * t_{n-1} * t_n) \\ &= \prod_{i=k}^n s_i * \prod_{i=k}^n s_i \end{aligned}$$

□

We can now consider extending the result of proposition 1.3.3. I.e if the product of $ab = 0$ for $a, b \in \mathbb{Z}$ then at least one of a or b is zero.

Proposition 1.5.9. *Product of two integers being zero implies one of the numbers is zero*

Let $x, y \in \mathbb{Z}$. If $xy = 0$ then at least one of x or y is zero.

Proof:

Let $x, y \in \mathbb{Z}$. If $x = y = 0$ then the result is trivial. So suppose that $x = (a, b)$ and $y = (c, d)$, moreover suppose $y \neq 0$. By definition of integer multiplication we have that

$$xy = (a, b) * (c, d) = (ac + bd, ad + bc) = (0, 0)$$

By assumption. We have that

$$\begin{aligned} (ac + bd, ad + bc) = (0, 0) &\iff ac + bd + 0 = ad + bc + 0 \\ &\Rightarrow ac + bd = ad + bc \end{aligned}$$

Now suppose without loss of generality suppose that $c > d$ then we have that $\exists p \in \mathbb{N}$ such that $d + p = c$. We hence have

$$\begin{aligned} ac + bd &= ad + bc \\ a(d + p) + bd &= ad + b(d + p) \\ ad + ap + bd &= ad + bd + bp \\ ap &= bp \\ a &= b, \text{ By the cancellation laws for the natural numbers} \\ a + 0 &= b + 0 \Rightarrow (a, b) = (0, 0) \end{aligned}$$

A similar argument applies for $c < d$.

Hence $x = 0$. A similar argument assuming $x \neq 0$ shows that $y = 0$. The result is shown. □

1.5.14 Extending the rules for inequalities to the integers

For the natural numbers, we were able to derive some rules for how inequalities behave, we can extend those results to the integers. Before we do so we have an additional consideration. As $\mathbb{N} \subset \mathbb{Z}$ then we can view every non-zero $n \in \mathbb{N}$ as a positive integer in \mathbb{Z} . Hence for positive $a, b, c \in \mathbb{Z}$ the results from the proposition 1.3.12 instantly extend to those integers.

To extend the results fully we need to consider negative integers as well. Consider $x = -3$ and $y = 6$, clearly $x < y$. Now consider $-1 * x = 3$ and $-1 * y = -6$, we have that $-1 * x > -1 * y$. This can be shown in general.

Proposition 1.5.10. *Multiplication by -1 changes the inequality sign*

Let $x, y \in \mathbb{Z}$. We have the following

1. *If $x < y$ then $-x > -y$*
2. *If $x \leq y$ then $-x \geq -y$*
3. *If $x > y$ then $-x < -y$*
4. *If $x \geq y$ then $-x \leq -y$*

Proof:

1. *If $x < y$ then $-x > -y$:*

Let $x, y \in \mathbb{Z}$ so that $x < y$. There are three cases to consider

- (a) $x \geq 0$ and $y \geq 0$
- (b) $x < 0$ and $y \geq 0$
- (c) $x < 0$ and $y < 0$

- (a) $x \geq 0$ and $y \geq 0$:

Suppose that $x \geq 0$ and $y \geq 0$ then $x \in [(a, 0)]$ for some $a \in \mathbb{N}$ and $y \in [(b, 0)]$ for some $b \in \mathbb{N}$. As $x < y$ then we must have $a + 0 < b + 0 \Rightarrow a < b$.

We have that

$$\begin{aligned} -x &= -1 * x = -1 * (a, 0) = (0, a) \\ -y &= -1 * y = -1 * (b, 0) = (0, b) \end{aligned}$$

Now, by proposition 1.3.12 part 2. we know that $a < b$ is the same as $b > a$. Now we have $-x > -y$ by definition of greater than for integers as we have

$$-x > -y \iff 0 + b > a + 0$$

- (b) $x < 0$ and $y \geq 0$:

Now suppose that $x < 0$ and $y \geq 0$ then we have that $x \in [(0, a)]$ and $y \in [(b, 0)]$ where $a, b \in \mathbb{N}$.

$$\begin{aligned} -x &= -1 * x = -1 * (0, a) = (a, 0) \\ -y &= -1 * y = -1 * (b, 0) = (0, b) \end{aligned}$$

Now, we have that if $-x > -y$ then we have

$$a + b > 0 + 0$$

However as $a, b \in \mathbb{N}$ and $x < 0 \implies a > 0$. We conclude that $a + b \geq a > 0$ and so $-x > -y$.

(c) $x < 0$ and $y < 0$:

Now suppose that $x < 0$ and $y < 0$ then $x \in [(0, a)]$ for some $a \in \mathbb{N}$ and $y \in [(0, b)]$ for some $b \in \mathbb{N}$. As $x < y$ then we have that $b < a$, which is the same as $a > b$.

We have that

$$\begin{aligned} -x &= -1 * x = -1 * (0, a) = (a, 0) \\ -y &= -1 * y = -1 * (0, b) = (b, 0) \end{aligned}$$

Applying the definition of $>$ to $-x$ and $-y$ gives

$$-x > -y \iff a > b$$

Which we know to be true. Hence $-x > -y$.

This shows part 1.

2. If $x \leq y$ then $-x \geq -y$:

If $x < y$ then we apply part 1. to get $-x > -y$ from which it follows that $-x \geq -y$ by definition. It is left to check when $x = y$. This is clear however as $x = y \implies -x = -y$ and so $-x \geq -y$.

3. If $x > y$ then $-x < -y$:

The proof of this part is similar to part 1. As in part 1. there are three cases to consider

(a) $x \geq 0$ and $y \geq 0$

(b) $x \geq 0$ and $y < 0$

(c) $x < 0$ and $y < 0$

(a) $x \geq 0$ and $y \geq 0$:

Suppose that $x \geq 0$ and $y \geq 0$ then $x \in [(a, 0)]$ for some $a \in \mathbb{N}$ and $y \in [(b, 0)]$ for some $b \in \mathbb{N}$. As $x > y$ then we must have $a + 0 > b + 0 \implies a > b$.

We have that

$$\begin{aligned} -x &= -1 * x = -1 * (a, 0) = (0, a) \\ -y &= -1 * y = -1 * (b, 0) = (0, b) \end{aligned}$$

Now, by proposition 1.3.12 part 2. we know that $a > b$ is the same as $b < a$. Now we have $-x < -y$ by definition of less than for integers as we have

$$-x < -y \iff 0 + b < a + 0$$

(b) $x \geq 0$ and $y < 0$:

Now suppose that $x \geq 0$ and $y < 0$ then we have that $x \in [(a, 0)]$ and $y \in [(0, b)]$ where $a, b \in \mathbb{N}$. We have

$$\begin{aligned} -x &= -1 * x = -1 * (a, 0) = (0, a) \\ -y &= -1 * y = -1 * (0, b) = (b, 0) \end{aligned}$$

Now, we have that if $-x < -y$ then we have

$$0 + 0 < a + b$$

However as $a, b \in \mathbb{N}$ and $y < 0 \implies b > 0$. We conclude that $0 < b \leq a + b$ and so $-x < -y$

(c) $x < 0$ and $y < 0$:

Now suppose that $x < 0$ and $y < 0$ then $x \in [(0, a)]$ for some $a \in \mathbb{N}$ and $y \in [(0, b)]$. A $x > y$ then we have that $0 + b > a + 0 \Rightarrow b > a$ which is the same as $a < b$.

$$-x = -1 * x = -1 * (0, a) = (a, 0)$$

$$-y = -1 * y = -1 * (0, b) = (b, 0)$$

Applying the definition of $<$ to $-x$ and $-y$ gives

$$-x < -y \iff a + 0 < b + 0 \Rightarrow a < b$$

Which we know to be true. Hence $-x < -y$.

4. If $x \geq y$ then $-x \leq -y$:

If $x > y$ we apply part 3. So instead suppose $x = y$ but then $x = y \Rightarrow -x = -y$ and so by definition we have $-x \leq -y$.

The result is shown. \square

This proposition will play a big role in the following proposition that extends the results for the rules of inequalities to the integers.

Proposition 1.5.11. *Properties of inequalities for the integers*

Let $x, y, z, c \in \mathbb{Z}$. We have the following properties for inequalities

1. $x < y$ is the same as $y > x$:
2. $x \leq y$ is the same as $y \geq x$:
3. If $x < y$ and $y < z$ then $x < z$:
4. If $x \leq y$ and $y < z$ then $x < z$:
5. If $x < y$ and $y \leq z$ then $x < z$:
6. If $x \leq y$ and $y \leq z$ then $x \leq z$:
7. If $x > y$ and $y > z$ then $x > z$:
8. If $x \geq y$ and $y > z$ then $x > z$:
9. If $x > y$ and $y \geq z$ then $x > z$:
10. If $x \geq y$ and $y \geq z$ then $x \geq z$:
11. If $x < y$ then $x + z < y + z$:
12. If $x \leq y$ then $x + z \leq y + z$:
13. If $x > y$ then $x + z > y + z$:
14. If $x \geq y$ then $x + z \geq y + z$:
15. If $x < y$ and $z \geq 0$ then $xz < yz$:
16. If $x < y$ and $z < 0$ then $xz > yz$:
17. If $x \leq y$ and $z \geq 0$ then $xz \leq yz$:
18. If $x \leq y$ and $z < 0$ then $xz \geq yz$:

19. If $x > y$ and $z \geq 0$ then $xz > yz$:
20. If $x > y$ and $z < 0$ then $xz < yz$:
21. If $x \geq y$ and $z \geq 0$ then $xz \geq yz$:
22. If $x \geq y$ and $z < 0$ then $xz \leq yz$:

Proof:

1. $x < y$ is the same as $y > x$:

Let $x, y \in \mathbb{Z}$ with $x < y$. Similar reasoning as in proposition 1.5.10 can be used. As in the proposition, there are three cases to consider.

- (a) $x \geq 0$ and $y \geq 0$
- (b) $x < 0$ and $y \geq 0$
- (c) $x < 0$ and $y < 0$

- (a) $x \geq 0$ and $y \geq 0$:

Suppose $x \geq 0$ and $y \geq 0$ then $x \in [(a, 0)]$ and $y \in [(b, 0)]$ for some $a, b \in \mathbb{N}$. We have that $x < y$ only holds if $a < b$, which is equivalent to $b > a$ by proposition 1.3.12. But by definition of $>$ for integers, we have that

$$b > a \iff y > x$$

- (b) $x < 0$ and $y \geq 0$:

Suppose that $x < 0$ and $y \geq 0$, then $x \in [(0, a)]$ and $y \in [(b, 0)]$ for some $a, b \in \mathbb{N}$. By definition of $<$ we have that

$$x < y \iff 0 + 0 < a + b \implies y > x \iff a + b > 0$$

Now, $x < 0 \implies a > 0$ and so we have that $a + b \geq a > 0$ and so $y > x$.

- (c) $x < 0$ and $y < 0$:

Now suppose that $x < 0$ and $y < 0$, it follows that $x \in [(0, a)]$ and $y \in [(0, b)]$ for some $a, b \in \mathbb{N}$. By definition of $<$ we have that

$$x < y \iff b < a \implies y > x \iff a > b$$

Hence, as $b < a$, we have that $a > b$ and so $y > x$.

2. $x \leq y$ is the same as $y \geq x$:

If $x < y$ then we apply part 1. Otherwise, we have that $x = y$ and so clearly $y = x$ and hence $y \geq x$.

3. If $x < y$ and $y < z$ then $x < z$:

Suppose that $x < y$ and $y < z$. There are four cases to consider.

- (a) $x \geq 0, y \geq 0$ and $z \geq 0$
- (b) $x < 0, y \geq 0$ and $z \geq 0$
- (c) $x < 0, y < 0$ and $z \geq 0$
- (d) $x < 0, y < 0$ and $z < 0$

- (a) $x \geq 0, y \geq 0$ and $z \geq 0$:

Suppose that $x \geq 0, y \geq 0$ and $z \geq 0$ then the result follows immediately by proposition 1.3.12 part 6. as $x \geq 0, y \geq 0$ and $z \geq 0$ gives $x \in [(a, 0)], y \in [(b, 0)]$ and $z \in [(c, 0)]$ for some $a, b, c \in \mathbb{N}$ and therefore $x, y, z \in \mathbb{N}$.

(b) $x < 0$, $y \geq 0$ and $z \geq 0$:

Now suppose that $x < 0$, $y \geq 0$ and $z \geq 0$. We have that $x \in [(0, a)]$, $y \in [(b, 0)]$ and $z \in [(c, 0)]$ for some $a, b, c \in \mathbb{N}$. Now we have that

$$x < 0 \iff a > 0$$

$$y \geq 0 \iff b \geq 0$$

$$z \geq 0 \iff c \geq 0$$

By assumption $x < y$ and so we have that $0 < a + b$, moreover by assumption $y < z$ and so we have $b < c$. We hence have that $0 < a + b < a + c$. Now as $0 < a + c$ we have by definition of $<$ that

$$0 < a + c \iff 0 + 0 < a + c \iff (0, a) < (c, 0) \iff x < z$$

(c) $x < 0$, $y < 0$ and $z \geq 0$:

Now suppose that $x < 0$, $y < 0$ and $z \geq 0$. We have that $x \in [(0, a)]$, $y \in [(0, b)]$ and $z \in [(c, 0)]$ for some $a, b, c \in \mathbb{N}$

$$x < 0 \iff a > 0$$

$$y < 0 \iff b > 0$$

$$z \geq 0 \iff c \geq 0$$

By assumption $x < y$ and so we have that $b < a$, moreover by assumption $y < z$ and so we have $0 < b + c$. As $b < a$ then we have $0 < b + c < a + c$, moreover we have by the definition of $<$ that

$$0 < a + c \iff 0 + 0 < a + c \iff (0, a) < (c, 0) \iff x < z$$

(d) $x < 0$, $y < 0$ and $z < 0$:

Suppose that $x < 0$, $y < 0$ and $z < 0$. We have that $x \in [(0, a)]$, $y \in [(0, b)]$ and $z \in [(0, c)]$ for some $a, b, c \in \mathbb{N}$. Observe

$$x < 0 \iff a > 0$$

$$y < 0 \iff b > 0$$

$$z < 0 \iff c > 0$$

As $x < y$ we have that $b < a$, likewise as $y < z$ we have that $c < b$, hence we have that $c < b < a$ and so $c < a$. Hence by definition of $<$ we have

$$c < a \iff 0 + c < a + 0 \iff (0, a) < (0, c) \iff x < z$$

4. If $x \leq y$ and $y < z$ then $x < z$:

Suppose that $x \leq y$ and $y < z$. If $x < y$ then we apply part 3. So suppose that $x = y$, then we must have that $y < z \iff x < z$ and hence the result.

5. If $x < y$ and $y \leq z$ then $x < z$:

As with part 5. Suppose $x < y$ and $y \leq z$, then if $y < z$ we apply part 3. Then we are left with the case $y = z$ and hence we have that $x < y \iff x < z$.

6. If $x \leq y$ and $y \leq z$ then $x \leq z$:

Suppose that $x \leq y$ and $y \leq z$, then if $x < y$ and $y < z$ we apply part 3. If $x \leq y$ and $y < z$ we apply part 4. If $x < y$ and $y \leq z$ we apply part 5. Hence we are left with the case where $x = y$ and $y = z$. The result follows immediately.

7. If $x > y$ and $y > z$ then $x > z$:

By part 1. of the proposition we have that this is equivalent to $y < x$ and $z < y$ then $z < x$ and so part 3. applies.

8. If $x \geq y$ and $y > z$ then $x > z$:

Applying part 2 to $x \geq y$ and part 1. to $y > z$ and $x > z$ gives the equivalent statement of $y \leq x$ and $z < y$ then $z < x$ and so part 4. applies.

9. If $x > y$ and $y \geq z$ then $x > z$:

As with part 8. Applying parts 2. and 1. gives the equivalent statement of $y < x$ and $z \geq y$ then $z < x$ and so part 5. applies

10. If $x \geq y$ and $y \geq z$ then $x \geq z$:

Solely applying part 2 of the proposition gives the statement $y \leq x$ and $z \leq y$ then $z \leq x$, so part 6. applies.

11. If $x < y$ then $x + z < y + z$:

Suppose that $x < y$ where $x \in [(a, b)]$ and $y \in [(c, d)]$ for some $a, b, c, d \in \mathbb{N}$. Let $z \in [(e, f)]$. By assumption we know that

$$x < y \iff a + d < b + c$$

Now, we have that

$$\begin{aligned} x + z &= (a, b) + (e, f) = (a + e, b + f) \\ y + z &= (c, d) + (e, f) = (c + e, d + f) \end{aligned}$$

Now, suppose that $x + z < y + z$. We have that

$$x + z < y + z \iff (a + e) + (d + f) < (b + f) + (c + e)$$

Observe that

$$\begin{aligned} (a + e) + (d + f) &< (b + f) + (c + e) \\ \underbrace{(a + d)}_{=j} + \underbrace{(e + f)}_{=k} &< \underbrace{(b + c)}_{=l} + \underbrace{(f + e)}_{=k} \end{aligned}$$

For some $j, k, l \in \mathbb{N}$. We see that $j < l$ as $a + d < b + c$. Hence by proposition 1.3.12 part 12. that $j < l \Rightarrow j + k < l + k$ and so we have $x + z < y + z$.

12. If $x \leq y$ then $x + z \leq y + z$:

If $x < y$ then the result follows from part 11. Otherwise, we have $x = y$ and then $x + z = y + z$ and so we have $x + z \leq y + z$.

13. If $x > y$ then $x + z > y + z$:

As has been the case so far, applying part 1. gives us the statement $y < x$ then $y + z < x + z$ and so part 11. applies.

14. If $x \geq y$ then $x + z \geq y + z$:

By part 2. we get the equivalent statement of $y \leq x$ then $y + z \leq x + z$ from which we can apply part 12.

15. If $x < y$ and $z \geq 0$ then $xz < yz$:

Suppose that $x < y$ where $x \in [(a, b)]$ and $y \in [(c, d)]$ for some $a, b, c, d \in \mathbb{N}$. Let $z \in [(e, 0)]$ for some $e \in \mathbb{N}$. As $x < y$ we have

$$x < y \iff a + d < b + c$$

Now we have that

$$\begin{aligned} xz &= (a, b) (e, 0) = (ae, be) \\ yz &= (c, d) (e, 0) = (ce, de) \end{aligned}$$

Now, consider $xz < yz$ then

$$xz < yz \iff ae + de < be + ce \iff e \underbrace{(a + d)}_{=m} < e \underbrace{(b + c)}_{=n}$$

The result now follows from proposition 1.3.12 part 16.

16. If $x < y$ and $z < 0$ then $xz > yz$:

Suppose that $x < y$ where $x \in [(a, b)]$ and $y \in [(c, d)]$ for some $a, b, c, d \in \mathbb{N}$. Let $z \in [(0, e)]$ for some $e \in \mathbb{N}$. As $x < y$ we have

$$x < y \iff a + d < b + c$$

Now we have that

$$\begin{aligned} xz &= (a, b) (0, e) = (be, ae) \\ yz &= (c, d) (0, e) = (de, ce) \end{aligned}$$

Now, we want to show that $xz > yz$, by definition we have

$$xz > yz \iff be + ce > ae + de \iff e \underbrace{(b + c)}_{=m} < e \underbrace{(a + d)}_{=n}$$

The result now follows from proposition 1.3.12 part 16.

17. If $x \leq y$ and $z \geq 0$ then $xz \leq yz$:

Suppose that $x \leq y$ then if we have that $x < y$ we apply part 15. Otherwise, $x = y$ and the result is trivial.

18. If $x \leq y$ and $z < 0$ then $xz \geq yz$:

Likewise, if $x < y$ then we apply part 16. So suppose that $x = y$ then $xz = yz$ and we, therefore, have $xz \geq yz$.

19. If $x > y$ and $z \geq 0$ then $xz > yz$:

Let $z \geq 0$ and by applying part 1. we get the equivalent statement of $y < x$ and $z \geq 0$ then $yz < xz$ for which we apply part 15.

20. If $x > y$ and $z < 0$ then $xz < yz$:

Applying part 1. we get the equivalent statement of $y < x$ and $z < 0$ then $yz < xz$ for which we apply part 16.

21. If $x \geq y$ and $z \geq 0$ then $xz \geq yz$:

Part 2 of this proposition gives the equivalent statement of $y \leq x$ and $z \geq 0$ then $yz \leq xz$ and so part 17. applies.

22. If $x \geq y$ and $z < 0$ then $xz \leq yz$:

Now, part 2 gives us the expression $y \leq x$ and $z < 0$ then $yz \geq xz$ and so we apply part 18.

The result has been shown. \square

1.5.15 The absolute value function

After the construction of the natural numbers, we explored the notion of cardinality. That was assigning a notion of size to a natural number. Recall the definition,

$$\begin{aligned} |\cdot| : \mathbb{N} &\rightarrow \mathbb{N} \\ n &\mapsto |n| = n \end{aligned}$$

To extend this we consider the following. We know that $a \in \mathbb{N}$ has a cardinality $|a| = a$ as $a \in \mathbb{N}$ refers to a set containing a elements. Unfortunately, the notion of a set containing a elements doesn't extend in a natural way to the integers. For example, what does it mean for a set to contain -3 elements? Instead, we need to re-think the notion of size.

Armed with subtraction we can re-cast our this understanding of size into a more useful form. Consider for example $6 - 3 = 3$, we can interpret this expression as saying that the number 3 is 3 less than 6, or equivalently the number 6 is 3 bigger than 3. Stated in another way, if we were to get a ruler and measure something to be $6cm$ long and we want to cut it in half we will measure the halfway point at $3cm$ along from where we start measuring. That is to say, the halfway point would be $6cm - 3cm = 3cm$.

What we have done is rather than think about the number of elements, we have thought about things in terms of distances. This turns out to be a very powerful idea, there is an entire subject in mathematics which studies this idea of distances, formally called metrics, which we will see later. We have only considered the positive case so far, what about $3 - 6$?

We know that $3 - 6 = -3$ and using similar logic this is saying that the number -3 is 6 away from 3, equivalently 3 is 6 more than -3 .

We make a definition.

Definition 1.5.19. *Distance function for integers*

Let $x, y \in \mathbb{Z}$. Define the function $d : \mathbb{Z}^2 \rightarrow \mathbb{N}$ by

$$\begin{aligned} d : \mathbb{Z}^2 &\rightarrow \mathbb{N} \\ (x, y) &\mapsto d(x, y) = \begin{cases} x - y, & \text{If } x \geq y \\ -(x - y), & \text{If } x < y \end{cases} \end{aligned}$$

We must verify that this is well defined

Proposition 1.5.12. *The distance function for the integers is well-defined*
Let $x, y \in \mathbb{Z}$. We have that

$$d(x, y) = \begin{cases} x - y, & \text{If } x \geq y \\ -(x - y), & \text{If } x < y \end{cases}$$

is well-defined.

Proof:

Let $x, y \in \mathbb{Z}$. There are two cases to consider $x \geq y$ and $x < y$.

1. $x \geq y$:

Suppose that $x \geq y$, then by proposition 1.5.11 part 14. we have

$$x \geq y \Rightarrow (x + (-y)) \geq (y + (-y)) \Rightarrow x - y \geq 0$$

Hence $x - y \in \mathbb{N}$.

2. $x < y$:

*As $x < y$ we have by definition of d that $d(x, y) = -(x - y)$ where we have that $x - y < 0$. However we have that $-(x - y) = -1 * (x - y)$ and so by part 16 of proposition 1.5.11 we have that $-1 * (x - y) > 0$ which is to say $-(x - y) \in \mathbb{N}$*

The result has been shown. \square

In light of the definition of the distance function, we can define the so-called absolute value function. This will give us a notion of the magnitude of an integer.

Definition 1.5.20. *Absolute value function*

Let $x \in \mathbb{Z}$ we define the absolute value function, denoted by $|x|$ by the function

$$|x| = d(x, 0) = \begin{cases} x, & \text{If } x \geq 0 \\ -x, & \text{If } x < 0 \end{cases}$$

With this definition, we have generalised the idea of “size” to the integers. That is the size of an integer is its distance from 0. We have the basic properties of the absolute value

Proposition 1.5.13. *Properties of the absolute value*

Let $x, y, z \in \mathbb{Z}$. We have that the absolute value function has the following properties

1. $|x| \geq 0$ for all $x \in \mathbb{Z}$
2. $|x| = 0 \iff x = 0$
3. $|x - y| = 0 \iff x = y$
4. $|xy| = |x| |y|$
5. $||x|| = |x|$
6. $|-x| = |x|$
7. $|x| \leq y \iff -y \leq x \leq y$
8. $|x| \geq y \iff x \leq -y \text{ or } x \geq y$
9. $|x + y| \leq |x| + |y|$
10. $|x - y| \leq |x - z| + |z - y|$

11. $|x - y| \geq ||x| - |y||$

12. $|\cdot|$ is not injective

13. $|\cdot|$ is not surjective

Proof:

1. $|x| \geq 0$ for all $x \in \mathbb{Z}$:

This follows by proposition 1.5.12.

2. $|x| = 0 \iff x = 0$:

We have by definition that $|x| = 0$, if and only if $x = 0$.

3. $|x - y| = 0 \iff x = y$:

(\Rightarrow): Suppose that $|x - y| = 0$. There are two cases to consider.

*Firstly if $x \geq y$, then by definition we have that $|x - y| = x - y = 0$ from which we clearly have $x = y$. The other case is $x < y$ from which we get $|x - y| = -(x - y) = 0$. In other words, we have $-1 * (x - y) = 0$. Now by proposition 1.5.9 we know that for integers a, b that if $ab = 0$, at least one of a or b is zero. As $-1 \neq 0$ we conclude that $x - y = 0$ from which we get $x = y$.*

(\Leftarrow): Suppose that $x = y$ then $x - y = 0$ and so $|x - y| = 0$.

4. $|xy| = |x| |y|$:

Let $x, y \in \mathbb{Z}$. There are four cases to consider.

(a) $x \geq 0$ and $y \geq 0$

(b) $x \geq 0$ and $y < 0$

(c) $x < 0$ and $y \geq 0$

(d) $x < 0$ and $y < 0$

(a) $x \geq 0$ and $y \geq 0$:

If $x \geq 0$ and $y \geq 0$ then $xy \geq 0$ and so $|xy| = xy$. Likewise $|x| = x$ and $|y| = y$. Hence $|xy| = |x| |y|$.

(b) $x \geq 0$ and $y < 0$:

If $x \geq 0$ then $|x| = x$ by definition, and if $y < 0$ then $|y| = -y$. Now $|xy| = -xy$ as $y < 0$. Moreover, we have that

$$-xy = (-1)(x)(y) = (x)(-1)(y) = (x)(-y) = |x| |y|$$

Hence we get $|xy| = |x| |y|$

(c) $x < 0$ and $y \geq 0$:

This is similar to the above but swapping the roles of x and y .

(d) $x < 0$ and $y < 0$:

*Suppose that $x < 0$ and $y < 0$, then we have that $|x| = -x$ and $|y| = -y$ by definition. Moreover, we have that $-x * -y = xy$. Hence $|xy| = xy = (-x)(-y) = |x| |y|$*

5. $||x|| = |x|$:

We have that $|x| = x$ if $x \geq 0$ and $-x$ if $x < 0$.

So if $x \geq 0$, we have

$$||x|| = |x| = x = |x|$$

Now if $x < 0$ then

$$||x|| = |-x| = \underbrace{-x}_{As -x > 0} = |x|$$

6. $|-x| = |x|$:

As $-x = -1 * x$ we have by part 4 that

$$|-x| = |-1 * x| = |-1| |x| = 1 * |x| = |x|$$

7. $|x| \leq y \iff -y \leq x \leq y$:

(\Rightarrow): Suppose that $|x| \leq y$. If $x \geq 0$ then we get that $|x| = x \leq y$. From this, it is clear that $-y \leq x \leq y$ as $x \geq 0$ and $x \leq y \Rightarrow y \geq 0$.

Now if $x < 0$, then $|x| = -x \leq y$. Clearly $x \leq -x$ as $x < 0$ hence we conclude that $x \leq -x \leq y$. Now by part 18 of proposition 1.5.11 we have we have

$$(-1) * (-x) \geq (-1)(y) \iff x \geq -y$$

Now $x \geq -y$ is the same as $-y \leq x$ and so we have $-y \leq x \leq -x \leq y$.

Hence $-y \leq x \leq y$.

(\Leftarrow): Suppose that $-y \leq x \leq y$. There are two cases to consider.

(a) $x \geq 0$

(b) $x < 0$

(a) $x \geq 0$:

Suppose $x \geq 0$, then clearly as $x \leq y$ then $|x| \leq |y| = y$. Moreover, we have that $-y \leq x$ is the same $x \geq -y$ and by part 22. of proposition 1.5.11 when applied to $x \geq -y$ gives

$$(-1) * (x) \leq (-1)(-y) \iff -x \leq y$$

We have that $|-x| = |x|$ by part 6. Hence $|-x| = |x| \leq |y| = y$.

(b) $x < 0$:

Suppose $x < 0$. By assumption $x \leq y$ so either $y \geq 0$ or $y < 0$. We can't have $y < 0$ as for example take $x = -4$ and $y = -2$ then we would have $2 \leq -4 \leq -2$ a contradiction.

So suppose that $y \geq 0$ then as $x \leq y$ we have $|x| \leq |y| = y$. Now as $-y \leq x$ by assumption we have that $x \geq -y$ and so part 22. of proposition 1.5.11 gives

$$(-1) * (x) \leq (-1)(-y) \iff -x \leq y$$

Hence part 6. applies and we get that $|x| \leq y$

8. $|x| \geq y \iff x \leq -y$ or $x \geq y$:

(\Rightarrow): Suppose that $|x| \geq y$. If $x \geq 0$ then $|x| = x \geq y$. So suppose that $x < 0$ then by definition we have that $|x| = -x$ and so $-x \geq y$ and the result follows when applying part 22. of proposition 1.5.11.

(\Leftarrow): Suppose that either $x \leq -y$ or $x \geq y$. We have three cases to consider.

(a) $x \leq -y$

(b) $x \geq y$

(c) $x \leq -y$ and $x \geq y$

(a) $x \leq -y$:

Suppose that $x \leq -y$ holds. If $x \geq 0$ then we have that $-y \geq 0$, Hence $y < 0$. Moreover, we have that by part 18. of proposition 1.5.11 that

$$(-1) * (x) \geq (-1)(-y) \iff -x \geq y$$

Now part 6. applies and we see that $|-x| = |x| \geq |y| = y$. This is to say $|x| \geq y$.

Now suppose that $x < 0$. Then as $x \leq -y$ we have that either $-y \geq 0$ or $-y < 0$. In the former case $-y \geq 0$ gives $y < 0$. Hence by part 18. of proposition 1.5.11 we conclude that

$$(-1) * (x) \geq (-1)(y) \iff -x \geq y$$

As $x < 0$ then $-x \geq 0$. The result follows when taking the absolute value.

Now suppose that $-y < 0$ then $y \geq 0$. Following similar logic to the previous case, we see that

$$(-1) * (x) \geq (-1)(y) \iff -x \geq y$$

The result again follows after taking the absolute value.

(b) $x \geq y$:

This case is trivial.

(c) $x \leq -y$ and $x \geq y$:

Suppose that $x \leq -y$ and $x \geq y$ are both true. We know by the first case that $x \leq -y$ gives $|x| \geq y$ and $x \leq y$ also implies $|x| \geq y$ by the second case. Hence both inequalities being true at the same time implies the result $|x| \geq y$.

9. $|x + y| \leq |x| + |y|$:

Let $x, y \in \mathbb{Z}$. There are four cases to consider.

(a) $x \geq 0$ and $y \geq 0$

(b) $x \geq 0$ and $y \leq 0$

(c) $x \leq 0$ and $y \geq 0$

(d) $x \leq 0$ and $y \leq 0$

(a) $x \geq 0$ and $y \geq 0$:

Suppose $x \geq 0$ and $y \geq 0$, then we have that

$$|x + y| = x + y = |x| + |y| \Rightarrow |x + y| \leq |x| + |y|$$

(b) $x \geq 0$ and $y \leq 0$

By assumption we have that $|x| = x$ and $|y| = -y$. We have two cases based on the absolute value, $|x| \leq |y|$ and $|x| \geq |y|$.

So suppose that $|x| \leq |y|$ then by definition $x \leq -y$ and so by part 12. of proposition 1.5.11 we have that

$$x \leq -y \Rightarrow x + y \leq 0$$

Moreover, as $x \geq 0$ then $y \leq x + y \leq 0$. Hence we have by the definition of the absolute value that

$$|x + y| = -(x + y) \leq -y = |y|$$

As $-y > 0$.

In the case $|x| \geq |y|$ we have by definition that $x \geq -y$ and so $x + y \geq 0$. Additionally it is clear that $x \geq x + y$ as $y \leq 0$ and $|x| \geq |y|$. Hence by definition of the absolute value we have that

$$|x + y| = x + y \leq x = |x|$$

Now, it is clear to see that $|x| \leq |x| + |y|$ and likewise $|y| \leq |x| + |y|$.

We have hence shown that $|x + y| \leq |x| + |y|$.

(c) $x \leq 0$ and $y \geq 0$:

This is similar to above, interchanging the roles of x and y .

(d) $x \leq 0$ and $y \leq 0$:

Suppose that $x \leq 0$ and $y \leq 0$ then by definition we have that $|x + y| = -(x + y) = -x - y$. As $x \leq 0$ and $y \leq 0$ then we have that $|y| = -y$ which shows $|x + y| = |x| + |y| \leq |x| + |y|$

10. $|x - y| \leq |x - z| + |z - y|$:

We have that

$$\begin{aligned} |x - y| &= |x - (z - z) - y| \\ &= |x - z + z - y| \\ &\leq |x - z| + |z - y| \end{aligned}$$

11. $|x - y| \geq ||x| - |y||$:

We have that

$$\begin{aligned} |x| &= |(x - y) + y| \leq |x - y| + |y| \Rightarrow |x| - |y| \leq |x - y| \\ |y| &= |(y - x) + x| \leq |x - y| + |x| \Rightarrow |y| - |x| \leq |x - y| \end{aligned}$$

Hence we have

$$\begin{aligned} |x| - |y| &\leq |x - y| \Rightarrow ||x| - |y|| \leq |x - y| \\ |y| - |x| &= (-1)(|x| - |y|) \leq |x - y| \Rightarrow ||x| - |y|| \leq |x - y| \end{aligned}$$

Hence we have the result.

12. $|\cdot|$ is not injective:

To see that the absolute value function is not injective consider $|3| = |-3|$. We have that $|3| = 3$ and $|-3| = 3$ but $3 \neq -3$.

13. $|\cdot|$ is not surjective:

We have that the absolute value function as there are no $x \in \mathbb{Z}$ so that $|x| = -1$ for example.

This ends the proposition. \square

1.5.16 Extending exponentiation to the integers

We can extend the idea of exponentiation to include integers. We are now able to consider negative bases. In other words, expressions of the form x^n for $x \in \mathbb{Z}$ with $x < 0$. This extension is somewhat trivial and extends naturally from the definition of the naturals. We first look at the case where $n \geq 0$

Definition 1.5.21. *Exponentiation of integer numbers*

Let $\mathbb{Z}^+ = \{x \in \mathbb{Z} : x \geq 0\}$. Let $(x, n) \in \mathbb{Z} \times \mathbb{Z}$ with $n \geq 0$ and let $\wedge : \mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{Z}$. We define the exponentiation of x by n to be x multiplied by itself $n - 1$ times

$$\wedge : \mathbb{Z} \times \mathbb{Z}^+ \rightarrow \mathbb{Z}$$

$$(x, n) \mapsto \wedge(x, n) = \begin{cases} 1, & \text{If } x = 0 \text{ and } n = 0 \\ 1, & \text{If } n = 0 \\ \prod_{i=1}^n x, & \text{If } x \neq 0 \text{ and } n \geq 0 \end{cases}$$

We will write $\wedge(x, n)$ as x^n . We say that x is the base and n is the exponent. We sometimes say that x has been raised to the power of n . In the case that $x = 0$ and $m = 0$ we have a vacuous product and so an empty product which by definition has a value of 1.

We will explore this definition by first considering $x = -1$

$$\begin{aligned} x * x &= x^1 = -1 = -1 \\ x * x &= x^2 = -1 * -1 = 1 \\ x * x * x &= x^3 = -1 * -1 * -1 = -1 \\ x * x * x * x &= x^4 = -1 * -1 * -1 * -1 = 1 \end{aligned}$$

This leads to the following proposition.

Proposition 1.5.14. *Negative one to power of $2n$ is 1* Let $n \in \mathbb{N}$. We have that

$$(-1)^{2n} = 1$$

Proof:

We argue by induction on n . The base case is $n = 0$ and by definition, we have that

$$(-1)^{2*0} = (-1)^0 = 1 = 1$$

Now suppose the result holds for some $n = k$, that is

$$(-1)^{2k} = 1$$

We show that

$$(-1)^{2*(k+1)} = 1$$

We have

$$\begin{aligned}
(-1)^{2(k+1)} &= (-1)^{2k+2} \\
&= \prod_{i=1}^{2k+2} (-1) \\
&= \prod_{i=1}^{2k} (-1) * \prod_{i=2k+1}^{2k+2} (-1) \\
&= 1 * ((-1)(-1)) &= 1 * (1) = 1
\end{aligned}$$

Which shows the result. \square

This result generalises for any negative integer.

Proposition 1.5.15. *Negative integer to the power of $2n$ is positive*
Let $x \in \mathbb{Z}$ with $x < 0$. Let $n \in \mathbb{N}$. We have that

$$x^{2n} > 1$$

Proof:

By definition we have

$$\begin{aligned}
x^{2n} &= \prod_{i=1}^{2n} x \\
&= \prod_{i=1}^{2n} (-1 * -x) \\
&= \prod_{i=1}^{2n} (-1) * \prod_{i=1}^{2n} (-x) \\
&= 1 * \underbrace{\prod_{i=1}^{2n} (-x)}_{\geq 1} \geq 1
\end{aligned}$$

As $-x > 0$ because $x < 0$. \square

We also note that exponentiation is neither commutative nor associative as they were not for the naturals. However, the following results do extend.

Proposition 1.5.16. *Power law of exponentiation for positive exponents*
Let $x \in \mathbb{Z}$ and let $n, m \in \mathbb{N}$ with $n \geq 0$ and $m \geq 0$. We have that

$$(x^n)^m = x^{nm}$$

Proof:

By the definition of exponentiation, we have that

$$(x^n)^m = \prod_{i=1}^m x^n = \prod_{i=1}^m \left(\prod_{j=1}^n x \right)$$

Hence we have

$$\begin{aligned}
(x^n)^m &= \underbrace{\prod_{j=1}^n x * \prod_{j=1}^n x * \cdots * \prod_{j=1}^n x}_{n \text{ times}} \\
&= \underbrace{\underbrace{x * x * \cdots * x}_{n \text{ times}} * \underbrace{x * x * \cdots * x}_{n \text{ times}} * \cdots * \underbrace{x * x * \cdots * x}_{n \text{ times}}}_{m \text{ times}}
\end{aligned}$$

Therefore, there are $n * m$ total multiplications of x with itself. Which is to say

$$(x^n)^m = \underbrace{x * x * x * \cdots * x}_{n*m \text{ times}} = \prod_{i=1}^{nm} x = x^{nm}$$

As promised. \square

Proposition 1.5.17. *Multiplying exponents of the same base adds the powers*

Let $x \in \mathbb{Z}$ be a fixed integer and let $n, m \in \mathbb{N}$. We have that

$$x^n * x^m = x^{n+m}$$

Proof:

Let $x \in \mathbb{Z}$ and $n, m \in \mathbb{N}$. If $n = 0$ or $m = 0$ or both then the result is trivial. Likewise if $n = 0$ and $m \geq 0$ or $n \geq 0$ and $m = 0$ again the result is trivial. So suppose that $n > 0$ and $m > 0$. We have by definition of exponentiation that

$$x^n * x^m = \prod_{i=1}^n x * \prod_{i=1}^m x = \underbrace{x * x * \cdots * x}_{n \text{ times}} * \underbrace{x * x * \cdots * x}_{m \text{ times}} = \underbrace{x * x * \cdots * x}_{n+m \text{ times}} = x^{n+m}$$

As expected. \square

Proposition 1.5.18. *Power of product is product of powers*

Let $x, y \in \mathbb{Z}$ and $n \in \mathbb{N}$. Then

$$(x * y)^n = x^n * y^n$$

Proof:

If $n = 0$ then $(x * y)^n = 1$ and clearly $x^0 * y^0 = 1$. So let $n > 0$ then we have

$$\begin{aligned}
(x * y)^n &= \prod_{i=1}^n xy = \underbrace{xy * xy * \cdots * xy}_{n \text{ times}} \\
&= \left(\underbrace{x * x * \cdots * x}_{n \text{ times}} \right) * \left(\underbrace{y * y * \cdots * y}_{n \text{ times}} \right), \quad \text{By commutativity of multiplication} \\
&= x^n * y^n
\end{aligned}$$

Showing the proposition. \square

The awake reader may have noticed how we have only dealt with positive exponents so far in our extension of exponentiation to the integers. What about negative exponents? We can, loosely, justify why we can't yet consider negative exponents by considering proposition 1.5.17. For a second suppose that instead of $n, m \in \mathbb{N}$ we consider $n, m \in \mathbb{Z}$. In particular $n = 1$ and $m = -1$, then we have that

$$x^1 * x^{-1} = x^{1+(-1)} = x^0 = 1$$

Hence we have that when x^1 is multiplied by x^{-1} we get back to 1. Hence in a sense x^{-1} cancels with x . If we let $x = 2$ we have $x^1 = 2$ and so $x^1 * x^{-1} = 1$ gives us the equation $2 * x^{-1} = 1$. We intuitively know that $x^{-1} = \frac{1}{2}$ which we know is not an integer. Hence if 1.5.17 held for all integer powers we have the implied existence of a new type of object. This object has the potential that when an integer is multiplied by the appropriate member of this new type of object, assuming such an object even exists, then integer multiplication is undone.

1.6 Construction of the Rationals

A man is like a fraction whose numerator is what he is and whose denominator is what he thinks of himself. The larger the denominator, the smaller the fraction.

Leo Tolstoy

We have now built a theory of integer numbers. One main reason for doing this was to be able to always undo subtraction. We still have a glaring issue at hand, however. How do we undo multiplication? For example, we are unable to express in mathematical language how many times one quantity goes into another. If we have 6 pints and 3 friends we know that each friend should get 2 pints as $3 * 2 = 6$. In a sense we have that 2 goes into 6 a total of 3 times and 3 goes into 6 a total of 2 times. The integers don't have a concept of how many times one integer can go into another. This is what we call division and we write $\frac{6}{2} = 3$ and $\frac{6}{3} = 2$ for each situation respectively.

Thankfully the method used to construct the integers can be used again on the integers themselves to construct an even richer theory. As with the integers, we should consider what we want to do. We seek a way to undo the multiplication of integers. Consider $a, b, c, d \in \mathbb{Z}$ $a = 6, b = 3, c = 12$ and $d = 6$, with these values we intuitively know that $\frac{6}{3} = 2$ and $\frac{12}{6} = 2$. We also note that $6 * 6 = 36$ and $3 * 12 = 36$. This gives us a clue on how to proceed. We have that $\frac{6}{3}$ and $\frac{12}{6}$ are hence similar. If we temporarily use the language of relations we have that $(a, b) \sim (c, d)$.

1.6.1 Defining the Rationals

We proceed by defining division as an ordered tuple on integers

Definition 1.6.1. *Division as an ordered tuple*

Let $a, b \in \mathbb{Z}$. We define division as an ordered tuple $(a, b) \in \mathbb{Z}^2$ to mean $\frac{a}{b}$. We will call $x \in \mathbb{Z}^2$ a division tuple in this context.

Hence we can define the relation we considered above.

Definition 1.6.2. *Relation for division*

Let $(a, b), (c, d) \in \mathbb{Z}^2$ be division tuples. We define the relation \sim such that $(a, b) \sim (c, d)$ if and only if $ad = bc$

With this definition there is something we need to consider that we have heard since school, you can't divide by zero, that is for any integer a we have $\frac{a}{0}$ is not defined.

Suppose that $(a, 0) \sim (c, d)$ for some $a, c, d \in \mathbb{Z}$. We have by definition of the relation that

$$(a, 0) \sim (c, d) \iff ad = 0 * c = 0$$

By proposition 1.5.9 we have that either $a = 0$ or $d = 0$ or both.

If $a = 0$ then we have $(0, 0) \sim (c, d) \Rightarrow 0 = 0$ for all $c, d \in \mathbb{Z}$. This means that every division tuple in \mathbb{Z}^2 would be equivalent to $(0, 0)$. Likewise if $d = 0$ we get $(a, 0) \sim (c, 0) \Rightarrow 0 = 0$ again meaning for all division tuples in \mathbb{Z}^2 would be equivalent. Finally if both $a = 0$ and $d = 0$ then $(0, 0) \sim (c, 0)$ and so $0 = 0 * c = 0$ and again every division tuple would be equivalent.

This is a problem as this relation would imply that all elements are essentially the same⁹. This is not a useful definition to be using so we will avoid this by not allowing $b = 0$ in $(a, b) \in \mathbb{Z}^2$. We revise the definition

⁹We can think of this as some sort of singularity

Definition 1.6.3. *Division as an ordered tuple*

Let $a, b \in \mathbb{Z}$ with $b \neq 0$. We define division as an ordered tuple $(a, b) \in \mathbb{Z}^2$ to mean $\frac{a}{b}$. We will call $x \in \mathbb{Z}^2$ a division tuple in this context.

Definition 1.6.4. *Relation for division*

Let $(a, b), (c, d) \in \mathbb{Z}^2$ be division tuples where $b \neq 0$ and $d \neq 0$. We define the relation \sim such that $(a, b) \sim (c, d)$ if and only if $ad = bc$

We can show that this revised definition is an equivalence relation.

Proposition 1.6.1. *Relation for division ordered tuple is an equivalence relation*

Let $x, y, z \in \mathbb{Z}^2$ be division tuples and defined the relation $x \sim y$ as above. We have that \sim is an equivalence relation.

Proof:

Let $x, y, z \in \mathbb{Z}^2$ be division tuples such that $x = (a, b), y = (c, d)$ and $z = (e, f)$. We show that \sim is an equivalence relation, in other words.

1. \sim is reflexive

2. \sim is symmetric

3. \sim is transitive

1. \sim is reflexive:

We have that for $x = (a, b)$ that $x \sim x$ as $x \sim x$ if and only if $ab = ab$.

2. \sim is symmetric:

Suppose that $x = (a, b)$ and $y = (c, d)$. Suppose that $x \sim y$ then we have that $ad = bc$. Hence $bc = ad \Rightarrow cb = ad$ and so $(c, d) \sim (a, b)$ and so $y \sim x$.

3. \sim is transitive:

Suppose that $x \sim y$ and $y \sim z$ then by definition we have that $ad = bc$ and $cf = de$. We have that

$$\begin{aligned} ad &= bc \\ adf &= bcf \\ adf &= bde \\ af &= be \end{aligned}$$

Hence $(a, b) \sim (e, f)$ and so $x \sim z$.

It follows that \sim is an equivalence relation. \square

We can now turn our attention to the set $\mathbb{Z}^2 / \sim = \{[x]_{\sim} : x \in \mathbb{Z}^2\}$.

Definition 1.6.5. *Rationals*

Let \mathbb{Z}^2 have the equivalence relation \sim defined by $(a, b) \sim (c, d)$ if and only if $ad = bc$. We define the set of rational numbers, denoted \mathbb{Q} , as the quotient set \mathbb{Z}^2 / \sim . The set has the form

$$\mathbb{Q} = \left\{ \dots, -\frac{2}{3}, -\frac{1}{3}, -\frac{1}{2}, 0, \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \dots \right\} \quad (15)$$

1.6.2 Extending equality to the rationals

As with the integers, it is easy to extend equality.

Definition 1.6.6. *Equality of rationals*

Let $x, y \in \mathbb{Q}$ be two rational numbers. We define that two rationals are equal, denoted $x = y$ if and only if $x \sim y$. That is x and y are in the same equivalence class. If $x \not\sim y$ then we say that x is not equal to y and write $x \neq y$.

1.6.3 Extending inequality operators to the rationals

The inequality operators can be extended to the rationals in a natural way.

Definition 1.6.7. *Less than operator*

Let $x, y \in \mathbb{Q}$ where $x \in [a, b]$ and $y \in [c, d]$ for some $a, b, c, d \in \mathbb{Z}$. The less than operator, denoted by $x < y$ is defined by the logical proposition

$$< (x, y) = \begin{cases} 1, & \text{If } ad < bc \\ 0, & \text{Otherwise} \end{cases}$$

This can equivalently be express as

$$x < y \iff ad < bc$$

Definition 1.6.8. *Less than or equal to operator*

Let $x, y \in \mathbb{Q}$ where $x \in [a, b]$ and $y \in [c, d]$ for some $a, b, c, d \in \mathbb{Z}$. The less than or equal operator, denoted by $x \leq y$ is defined by the logical proposition

$$\leq (x, y) = \begin{cases} 1, & \text{If } ad \leq bc \\ 0, & \text{Otherwise} \end{cases}$$

This can equivalently be express as

$$x \leq y \iff ad \leq bc$$

Definition 1.6.9. *Greater than operator*

Let $x, y \in \mathbb{Q}$ where $x \in [a, b]$ and $y \in [c, d]$ for some $a, b, c, d \in \mathbb{Z}$. The greater than operator, denoted by $x > y$ is defined by the logical proposition

$$> (x, y) = \begin{cases} 1, & \text{If } ad > bc \\ 0, & \text{Otherwise} \end{cases}$$

This can equivalently be express as

$$x > y \iff ad > bc$$

Definition 1.6.10. *Greater than or equal to operator*

Let $x, y \in \mathbb{Q}$ where $x \in [a, b]$ and $y \in [c, d]$ for some $a, b, c, d \in \mathbb{Z}$. The greater than or equal to operator, denoted by $x \geq y$ is defined by the logical proposition

$$\geq (x, y) = \begin{cases} 1, & \text{If } ad \geq bc \\ 0, & \text{Otherwise} \end{cases}$$

This can equivalently be express as

$$x \geq y \iff ad \geq bc$$

1.6.4 Extending addition to the rationals

We can extend addition to the rationals. To do so we need to consider how integers are represented in the rationals. As we know an element $(a, b) \in \mathbb{Q}$ is going to represent $\frac{a}{b}$. So we can start by considering what an integer will look like. We know by the definition of the equivalence relation that for $(a, b), (c, d) \in \mathbb{Z}^2$ that

$$(a, b) \sim (c, d) \iff ad = bc$$

Hence if we have for $b = d = 1$ that

$$(a, 1) \sim (c, 1) \iff a = c$$

Hence an integer can be represented in the rationals by an element of the form $(k, 1)$ for all $k \in \mathbb{Z}$. Therefore if $x, y \in \mathbb{Z}$ they will have the representation $x = (x_1, 1)$ and $y = (y_1, 1)$ for some $x_1, y_1 \in \mathbb{Z}$. Hence by integer addition, we have that

$$x + y = (x_1, 1) + (y_1, 1) = (x_1 + y_1, 1)$$

Now what happens if $a = c = 1$? From the definition of the equivalence relation we have that

$$(1, b) \sim (1, d) \iff d = b$$

So we see that $(1, b) \sim (1, d)$ means that intuitively $\frac{1}{b} = \frac{1}{d}$. The question now becomes what is $\frac{1}{b} + \frac{1}{b}$?

For example consider $\frac{1}{2} + \frac{1}{2} = 1$, or $\frac{1}{3} + \frac{1}{3} = \frac{2}{3}$. It seems the result we need is that $\frac{1}{b} + \frac{1}{b} = \frac{2}{b}$. We hence have that

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

Hence more generally we have that

$$(a, b) + (c, b) = (a + c, b)$$

Now, from intuition, we know that for example $\frac{1}{3} = \frac{2}{6} = \frac{1 * 2}{3 * 2}$. In the language of the relation we have defined, we have that

$$(a, b) \sim (ad, bd)$$

With these facts, we have enough to recover the definition of the addition of rational numbers we were told in school.

We have that

$$\begin{aligned} (a, b) + (c, d) &\sim (ad, bd) + (bc, bd) \\ &\sim (ad + bc, bd) \end{aligned}$$

Indeed, we have for example

$$\frac{1}{2} + \frac{1}{3} = \frac{3 * 1 + 2 * 1}{3 * 2} = \frac{5}{6}$$

We make the required definition.

Definition 1.6.11. *Addition on the Rationals*

Let $x, y \in \mathbb{Q}$ with $x = [a, b]$ and $y = [c, d]$ so that $b \neq 0$ and $d \neq 0$. We define addition on the rationals by

$$x + y = [a, b] + [c, d] = [ad + bc, bd] \tag{16}$$

1.6.5 Extending multiplication to the rationals

We can extend multiplication to the rationals as well. As with extending addition, we should consider how integers are represented in the rationals. As before an integer in the rationals is of the form $(a, 1)$ and given the definition from the integers we know we must have

$$(a, 1) * (b, 1) = (ab, 1)$$

Now we need to answer the question of $(1, b) * (1, d)$. Taking a similar approach as to addition we will consider some examples. We intuitively know that $1 * \frac{1}{2} = \frac{1}{2}$. This is to say that

$$(1, 1) * (1, 2) = (1, 2)$$

We also know that $2 * 2 = 4$ and so we know $\frac{4}{2} = 2$. In other words we must have that

$$(4, 1) * (1, 2) = (2, 1) \sim (4, 2)$$

Now, suppose we have $\frac{3}{2} = 1.5$, what is $\frac{3}{2} * \frac{1}{3}$? Again we know intuitively that $0.5 + 0.5 + 0.5 = 3(0.5) = 1.5$, hence we can write

$$(3, 2) * (1, 3) = (1, 2) \sim (3, 6)$$

We can now see how to handle $(1, b) * (1, d)$ and more generally $(a, b) * (c, d)$. We make the definition.

Definition 1.6.12. *Multiplication on the Rationals*

Let $x, y \in \mathbb{Q}$ with $x \in [a, b]$ and $y = [c, d]$ so that $b \neq 0$ and $d \neq 0$. We define multiplication on the rationals by

$$x * y = [a, b] * [c, d] = [ac, bd] \quad (17)$$

1.6.6 Closure properties of addition and multiplication

As with the natural numbers and integers we need to show that the operations of addition and multiplication on the rationals are closed and well-defined.

Theorem 1.6.1. *Addition and multiplication on the rational are well-defined operators and closed*

We have that $\forall x, y \in \mathbb{Q}$ that

$$1. \ x + y \in \mathbb{Q}$$

$$2. \ x * y \in \mathbb{Q}$$

Proof:

$$1. \ x + y \in \mathbb{Q}:$$

We must show that if $(a, b) \sim (a', b')$ and $(c, d) \sim (c', d')$ then we have

$$(ad + bc, bd) \sim (a'd' + b'c', b'd')$$

By definition we have that $(a, b) \sim (a', b')$ holds if and only if $ab' = ba'$, likewise $(c, d) \sim (c', d')$ holds if and only if $cd' = c'd$. It is left to show $(ad + bc, bd) \sim (a'd' + b'c', b'd')$. By definition of the equivalence relation we have that

$$(ad + bc, bd) \sim (a'd' + b'c', b'd') \iff (ad + bc)b'd' = bd(a'd' + b'c')$$

We have that

$$\begin{aligned} (ad + bc)b'd' &= adb'd' + bcb'd' \text{ As integer multiplication distributes over the addition} \\ &= (ab')(dd') + (cd')(bb') \text{ By commutativity} \\ &= (ba')(dd') + (dc')(bb') \text{ By the equivalence relation} \\ &= (bd)(a'd') + (bd)(b'c') \text{ By commutativity} \\ &= bd(a'd' + b'c') \text{ As integer multiplication distributes over the addition} \end{aligned}$$

Which is what we wished to show. Hence addition is well-defined. It is left to show closure. Let $x, y \in \mathbb{Q}$ with $x = (a, b)$ and $y = (c, d)$ so that $b \neq 0$ and $d \neq 0$. By definition of addition we have that

$$(a, b) + (c, d) = (ad + bc, bd)$$

As $ad + bc \in \mathbb{Z}$ and $bd \in \mathbb{Z}$ then $(ad + bc, bd) \in [ad + bc, bd]$ and so $x + y \in \mathbb{Q}$.

2. $x * y \in \mathbb{Q}$:

As with addition we need to show that if $(a, b) \sim (a', b')$ and $(c, d) \sim (c', d')$ that

$$(ac, bd) \sim (a'c', b'd')$$

As $(a, b) \sim (a', b')$ holds if and only if $ab' = ba'$, likewise $(c, d) \sim (c', d')$ holds if and only if $cd' = c'd$. It is left to show $(ac, bd) \sim (a'c', b'd')$, that is

$$(ac, bd) \sim (a'c', b'd') \iff acb'd' = bda'c'$$

We have

$$\begin{aligned} acb'd' &= (ab')(cd') \text{ By commutativity} \\ &= (ba')(c'd) \text{, By the equivalence relation} \\ &= bda'c' \text{, By commutativity} \end{aligned}$$

Showing that multiplication is well-defined. To show closure let $x, y \in \mathbb{Q}$ with $x = (a, b)$ and $y = (c, d)$ so that $b \neq 0$ and $d \neq 0$ then by definition we have that

$$(a, b) * (c, d) = (ac, bd)$$

From which it is clear that $ac, bd \in \mathbb{Z}$ so $x * y \in \mathbb{Q}$

The result is shown. \square

1.6.7 Associativity of rational addition and multiplication

The associativity of addition and multiplication extends to the rationals.

Theorem 1.6.2. Let $x, y, z \in \mathbb{Q}$. We have that

1. $x + (y + z) = (x + y) + z$
2. $x(yz) = (xy)z$

Proof:

1. $x + (y + z) = (x + y) + z$:

Let $x, y, z \in \mathbb{Q}$ be such that $x = (a, b)$, $y = (c, d)$ and $z = (e, f)$ where $a, b, c, d, e, f \in \mathbb{N}$ and we have that $(a, b) \in [a, b]$, $(c, d) \in [c, d]$ and $(e, f) \in [e, f]$. We have that

$$\begin{aligned}
x + (y + z) &= (a, b) + ((c, d) + (e, f)) \\
&= (a, b) + (cf + de, df) \\
&= (adf + b(cf + de), bdf) \\
&= (adf + bcf + bde, bdf) \\
&= ((ad + bc)f + bde, bdf) \\
&= ((ad + bc)f + bde, bdf), \text{ By associativity of addition for integer numbers} \\
&= (ad + bc, bd) + (e, f) \\
&= ((a, b) + (c, d)) + (e, f) \\
&= (x + y) + z
\end{aligned}$$

Which shows associativity of addition.

2. $x(yz) = (xy)z$:

As with addition, let $x, y, z \in \mathbb{Q}$ be such that $x = (a, b)$, $y = (c, d)$ and $z = (e, f)$ where $a, b, c, d, e, f \in \mathbb{Z}$ and we have that $(a, b) \in [a, b]$, $(c, d) \in [c, d]$ and $(e, f) \in [e, f]$. We then have that

$$\begin{aligned}
x(yz) &= (a, b) * ((c, d)(e, f)) \\
&= (a, b) * (ce, df) \\
&= (ace, bdf) \\
&= (ac, bd) * (e, f) \\
&= ((a, b) * (c, d)) * (e, f)
\end{aligned}$$

Showing associativity of multiplication.

The result follows. \square

1.6.8 Commutativity of rational addition and multiplication

As with the naturals and integers, addition and multiplication in the rationals both satisfy commutativity.

Theorem 1.6.3. *Addition and multiplication are commutative*

For all $x, y \in \mathbb{Q}$ we have that

$$1. \ x + y = y + x$$

$$2. \ xy = yx$$

Proof:

$$1. \ x + y = y + x:$$

Let $x, y \in \mathbb{Q}$. By definition we have that $x \in [a, b]$ and $y \in [c, d]$ for some $a, b, c, d \in \mathbb{Z}$. Let $x = (a, b)$ and $y = (c, d)$. We then have by definition of addition that

$$\begin{aligned}
x + y &= (a, b) + (c, d) \\
&= (ad + bc, bd) \\
&= (bc + ad, bd), \text{ By associativity of addition for the integers} \\
&= (cb + da, db), \text{ By commutativity of addition for the integers} \\
&= (c, d) + (a, b) &= y + x
\end{aligned}$$

Showing commutativity holds for addition in the integers.

2. $xy = yx$:

Let $x, y \in \mathbb{Q}$ by definition we have that $x \in [a, b]$ and $y \in [c, d]$ for some $a, b, c, d \in \mathbb{Z}$. So let $x = (a, b)$ and $y = (c, d)$. By definition of multiplication we have

$$\begin{aligned} xy &= (a, b) * (c, d) \\ &= (ac, bd) \\ &= (ca, db), \text{ By commutativity of multiplication of the integers} \\ &= (c, d) * (a, b) \\ &= yx \end{aligned}$$

Showing commutativity for integer multiplication.

The result has been shown. \square

1.6.9 The Zero and Identity laws

The zero and identity laws from both the naturals and integers extend to the rationals. But first, we show the following result.

Lemma 1.6.1. *Representation of zero in the rationals*

We have that $0 = [0, a]$ for all $a \in \mathbb{Z}$ with $a \neq 0$

Proof:

Let $x, y \in [0, a]$ with $x = (0, a_1)$ and $y = (0, a_2)$. We hence have that $x \sim y$ and

Where the final $0 = 0$ is the zero of the integers, from which the result is clear. \square

We take the natural representation of 0 for the rationals.

Theorem 1.6.4. *The zero and Identity laws*

Let $x \in \mathbb{Q}$. We have that

$$1. \ x + 0 = x = 0 + x$$

$$2. \ 1 * x = x = x * 1$$

Proof:

Let $x \in \mathbb{Q}$ then we have that $x = (a, b)$ for some $a, b \in \mathbb{Z}$

$$1. \ x + 0 = x = 0 + x:$$

We have that $0 \in [0, 1]$. Hence we have that

$$\begin{aligned} x + 0 &= (a, b) + (0, 1) \\ &= (a * 1 + b * 0, b * 1) \\ &= (a, b) = x \\ &= (1 * a + 0 * b, 1 * b) \\ &= (0, 1) * (a, b) \\ &= 0 + x \end{aligned}$$

$$2. \ x * 1 = x = 1 * x:$$

As $1 \in [1, 1]$ then

$$\begin{aligned}
x * 1 &= (a, b) * (1, 1) \\
&= (a * 1, b * 1) \\
&= (a, b) \\
&= (a, b) = x \\
&= (1 * a, 1 * b) \\
&= (1, 0) (a, b) \\
&= 1 * x
\end{aligned}$$

The result follows. \square

1.6.10 Multiplication distributes over addition

Yet another result that extends to the rationals is that multiplication distributes over addition.

Theorem 1.6.5. *Multiplication distributes over addition*

For all $x, y, z \in \mathbb{Q}$ we have that

1. $x(y + z) = xy + xz$
2. $(y + z)x = yx + zx = xy + xz$

Proof:

Let $x, y, z \in \mathbb{Q}$ then $x \in [a, b], y \in [c, d]$ and $z \in [e, f]$ for some $a, b, c, d, e, f \in \mathbb{Z}$.

Let $x = (a, b), y = (c, d)$ and $z = (e, f)$.

1. $x(y + z) = xy + xz$:

We have that

$$\begin{aligned}
x(y + z) &= (a, b)((c, d) + (e, f)) \\
&= (a(cf + ed), bdf) \\
&= (acf + aed, bdf), \text{ By multiplication distributes over addition for the integers} \\
&= (acf + aed, bdf) * (1, 1), \text{ By the identity law for the rationals} \\
&= (acf + aed, bdf) * (b, b), \text{ As } (1, 1) \sim (b, b) \\
&= ((acf + aed)b, bdfb) \\
&= (acfb + aedb, bdfb), \text{ By multiplication distributes over addition for the integers} \\
&= (acbf + aebd, bdbf), \text{ By commutativity of integer multiplication} \\
&= (ac, bd) + (ae, bf) \\
&= (a, b)(c, d) + (a, b)(e, f) \\
&= xy + xz
\end{aligned}$$

2. $(y + z)x = yx + zx = xy + xz$:

Invoking the previous part of the proof we have that

$$\begin{aligned}
(y + z)x &= x(y + z), \text{ By commutativity of multiplication} \\
&= xy + xz, \text{ By part 1.} \\
&= yx + zx, \text{ By commutativity of multiplication}
\end{aligned}$$

As required. \square

1.6.11 Extending subtraction to the rationals

We can extend subtraction from the integers to the rationals. Recall that subtraction was defined for $x, y \in \mathbb{Z}$ by

$$x - y = x + (-y) = x + (-1 * y)$$

That is to say subtraction was defined by adding the negation of y to x . We will use a similar idea to define subtraction on the rationals. Firstly we need to consider what it means to negate a rational number. To do so we need to define what it means for a rational number to be "positive" or "negative".

We know that any integer x can be expressed as a rational by $(x, 1)$ and so in this case $(x, 1)$ is positive if x is positive and $(x, 1)$ is negative if x is negative. Hence a general rational number (a, b) being positive or negative will depend on a and b being positive or negative. There are a few cases to consider.

1. Suppose that a is positive and b is positive. We have that for $(a, b) \sim (c, d)$ for some $c, d \in \mathbb{Z}$ that

$$ad = cb$$

As a and b are positive then we are forced to conclude that c and d are also positive for if not then one side of this equation would have a different sign.

2. Suppose that a is positive and b is negative. Then as before we have that for $(a, b) \sim (c, d)$ to be true that

$$ad = cb$$

As b was negative then we have that cb is either positive or negative depending on c . If c is positive then cb is negative and so d must also be negative. Likewise if c is negative then cb is positive and d must be positive.

The cases for when a is negative and b is either positive or negative are similar. We can use this to make a definition for a positive and negative rational number.

Definition 1.6.13. *Positive and negative rational number*

Let $x \in \mathbb{Q}$ so that $x = (a, b)$ for some $a, b \in \mathbb{Z}$. We say that x is a positive rational number if and only if a is positive and b is positive. That is to say $x \in \mathbb{Q}$ is positive if and only if $\text{sgn}(a) = \text{sgn}(b)$ with $\text{sgn}(a) \neq 0$ and $\text{sgn}(b) \neq 0$ where sgn denotes the sign function of an integer.

If $\text{sgn}(a) \neq \text{sgn}(b)$ and $\text{sgn}(a) \neq 0$ and $\text{sgn}(b) \neq 0$ then we have that x is a negative rational number.

Finally if $\text{sgn}(a) = 0$ and $\text{sgn}(b) \neq 0$ then we say that x is neither positive or negative.

We can summarise this definition using sgn just like we did for the integers.

Definition 1.6.14. *Sign of a rational number*

Let $x \in \mathbb{Q}$ where $x = (a, b)$ with $a, b \in \mathbb{Z}$ and $b \neq 0$. We define the sign of x , denoted by $\text{sgn}(x)$ to be the following function

$$\begin{aligned} \text{sgn} : \mathbb{Q} &\rightarrow \{-1, 0, 1\} \\ x &\mapsto \text{sgn}(x) = \begin{cases} 1, & \text{If } x \text{ is a positive rational number} \\ -1, & \text{If } x \text{ is a negative rational number} \\ 0, & \text{If } \text{sgn}(a) = 0 \end{cases} \end{aligned}$$

Now that we have defined the notion of a positive and negative rational number we can consider what it means to negate a rational number. The definition follows immediately from the representation of -1 in \mathbb{Q} being $(-1, 1)$. Indeed for any $x \in \mathbb{Q}$ with $x = (a, b)$ we have

$$-x = -1 * x = (-1, 1) * (a, b) = (-a, b)$$

We make the formal definition.

Definition 1.6.15. *Negation of a rational number*

Let $x \in \mathbb{Q}$. We define the negation of x , denoted $-x$ by

$$-x = -1 * x = (-1, 1) * x$$

where $(-1, 1) \in [(-1, 1)]$. That is $(-1, 1)$ is an element of the equivalence class $[(-1, 1)]$ which represents all possible elements that are -1 .

We can now make a definition for subtraction for the rational numbers

Definition 1.6.16. *Rational number subtraction*

Let $x, y \in \mathbb{Q}$. We define the subtraction of y from x , denoted $x - y$ by

$$x - y = x + (-y) = x + (-1 * y)$$

We immediately get that subtraction is closed, from the fact that both addition and multiplication is closed. We do not have associativity of subtraction in general.

Proposition 1.6.2. *Rational number subtraction is not associative*

Let $x, y, z \in \mathbb{Q}$. We have that

$$x - (y - z) \neq (x - y) - z$$

Proof:

Let $x = \frac{1}{2}, y = \frac{1}{4}$ and $z = \frac{1}{6}$, we have $x \in [(1, 2)], y \in [(1, 4)]$ and $z \in [(1, 6)]$ so $x = (1, 2), y = (1, 4)$ and $z = (1, 6)$. We have that

$$\begin{aligned} x - (y - z) &= (1, 2) + ((1, 4) - (1, 6)) \\ &= (1, 2) - ((1, 4) + (-1 * (1, 6))) \\ &= (1, 2) - ((1, 4) + (-1, 6)) \\ &= (1, 2) - ((1 * 6 + 4 * -1, 4 * 6)) \\ &= (1, 2) - ((2, 24)) \\ &= (1, 2) + (-1 * ((2, 24))) \\ &= (1, 2) + (-2, 24) \\ &= (1 * 24 + 2 * -1, 2 * 24) \\ &= (22, 48) \end{aligned}$$

On the other hand we have

$$\begin{aligned} (x - y) - z &= (1, 2) - ((1, 4) - (1, 6)) \\ &= ((1, 2) + (-1 * (1, 4))) - (1, 6) \\ &= ((1, 2) + (-1, 4)) - (1, 6) \\ &= (1 * 4 + 2 * -1, 2 * 4) - (1, 6) \\ &= (2, 8) - (1, 6) \\ &= (2, 8) + (-1 * (1, 6)) \\ &= (2, 8) + (-1, 6) \\ &= (2 * 6 + 8 * -1, 8 * 6) \\ &= (4, 48) \end{aligned}$$

It is left to show that $(22, 48) \neq (4, 48)$. Indeed to have $(22, 48) = (4, 48)$ we need $(22, 48) \sim (4, 48)$ which occurs if and only if $22 * 48 = 48 * 8$. However on the left hand side 48 is multiplied by 22 and on the right-hand side 48 is multiplied by 8 so they clearly can not be equal.

The result is shown. \square

As with subtraction with integers, we can now show that formally, subtraction is an inverse to addition.

Proposition 1.6.3. *Subtracting an integer from itself gives zero*

Let $x \in \mathbb{Q}$. We have that

$$x - x = 0$$

Proof:

Let $x \in \mathbb{Q}$ where $x \in [(a, b)]$ for some $a, b \in \mathbb{Z}$ and $b \neq 0$. We have

$$\begin{aligned} x - x &= (a, b) - (a, b) \\ &= (a, b) + (-a, b) \\ &= (ab + b * -a, b * b) \\ &= (ab - ba, b * b) \\ &= (ab - ab, b * b) \\ &= (0, b * b) \end{aligned}$$

It is left to show that $(0, b * b) \sim (0, 1)$. Indeed

$$0 * 1 = b * b * 0 \Rightarrow 0 = 0$$

The result is shown. \square

1.6.12 The cancellation laws

We can now deduce that the cancellation laws extend to the rational numbers.

Theorem 1.6.6. *The cancellation laws*

Let $x, y, z \in \mathbb{Q}$.

1. If $x + y = x + z$ then we have $y = z$.
2. For $x \neq 0$, if $xy = xz$ then we have that $y = z$

Proof:

1. If $x + y = x + z$ then we have $y = z$:

Let $x, y, z \in \mathbb{Q}$. We have that

$$\begin{aligned} x + y &= x + z \\ \Rightarrow -x + x + y &= -x + x + z, \text{ Adding the negative of } x \text{ to both sides} \\ \Rightarrow (-x + x) + y &= (-x + x) + z, \text{ Associativity of the rationals} \\ \Rightarrow 0 + y &= 0 + z, \text{ By proposition 1.6.3} \\ \Rightarrow y &= z \end{aligned}$$

2. For $x \neq 0$, if $xy = xz$ then we have that $y = z$:

Let $x, y, z \in \mathbb{Q}$ where $x \neq 0$. Suppose that $x \in [(a, b)]$, $y \in [(c, d)]$ and $z \in [(e, f)]$. We have

$$\begin{aligned} xy &= (a, b)(c, d) = (ac, bd) \\ xz &= (a, b)(e, f) = (ae, bf) \end{aligned}$$

Now suppose that $xy = xz$ then we have that $(ac, bd) \sim (ae, bf)$ which is to say

$$acbf = aebd$$

Observe that

$$\begin{aligned} acbf &= aebd \\ a(cb f) &= a(ebd) \\ cbf &= ebd, \text{ By the cancellation laws for the integers} \\ bcf &= bed, \text{ By commutativity of the integers} \\ b(cf) &= b(ed) \\ cf &= ed, \text{ By the cancellation laws for the integers} \\ \Rightarrow (c, d) &\sim (e, f), \text{ By definition of the equivalence relation} \end{aligned}$$

It hence follows that as $(c, d) \sim (e, f)$ then $y = z$

The result is shown. \square

1.6.13 Defining multiplicative inverses and division

When we extended the naturals to the integers we were able to extend the notion of subtraction in such a way that we could undo any addition operation. We were not able to do the same for multiplication in general. For example if we have $x * 2 = 1$ where $1, 2, x \in \mathbb{Z}$ then there is no integer x that when multiplied by 2 gives 1.

What happens if we consider instead the situation where we have $1, 2, x \in \mathbb{Q}$? Let $x = (a, b)$ for some $a, b \in \mathbb{Z}$ with $b \neq 0$ and taking the natural representations for 1 and 2 of $1 = (1, 1)$ and $2 = (2, 1)$. We have that

$$\begin{aligned} x * 2 &= 1 \\ (a, b)(2, 1) &= (1, 1) \\ (2a, b) &= (1, 1) \\ \Rightarrow (2a, b) &\sim (1, 1) \iff 2a = b \end{aligned}$$

We don't seem to be in a better position then when we asked this question for \mathbb{Z} . However as a, b were arbitrary, of course with $b \neq 0$, we are free to vary them. For example $a = 1$ gives us $b = 2$, $a = 2$ gives $b = 4$, $a = 3$ yields $b = 6$ and so on. We hence have that there is a family of possible value for x which satisfies $x * 2 = 1$ over the rational numbers, in particular we have $x = (a, 2a)$ for $a \in \mathbb{Z}$ and $a \geq 0$. Moreover we clearly have

$$(a, 2a) \sim (1, 2) \iff 2a = 2a$$

Hence we have that $(a, 2a)$ somehow undoes multiplication by 2. Indeed consider $45 * 2 = 90$. We have that

$$90 * (a, 2a) = (90, 1) * (a, 2a) = (90a, 2a)$$

Where we have $(90a, 2a) \sim (45, 1)$ as $90a * 1 = 45 * 2a \iff 90a = 90a$. We can generalise this to $x * y = 1$ for any $y \in \mathbb{Q}$. Indeed let $x = (a, b)$ and $y = (c, d)$ where $a, b, c, d \in \mathbb{Z}$ and $c \neq 0$ and $d \neq 0$ then we have

$$\begin{aligned} x * y &= (a, b) * (c, d) \\ &= (ac, bd) = (1, 1) \\ \Rightarrow (ac, bd) &\sim (1, 1) \iff bd = ac \end{aligned}$$

This is a somewhat unsatisfactory conclusion as it doesn't tell us what a or b should actually be equal to in order for $x * y = 1$, likewise, it doesn't tell us what c or d should be either.

Perhaps then we should consider a more simple setup. Suppose that $x \in \mathbb{Z}$ then is there $y \in \mathbb{Q}$ where $y = (c, d)$ with $d \neq 0$, such that $x * y = 1$? We have

$$x * y = (x, 1) * (c, d) = (xc, d) = (1, 1)$$

Hence

$$(xc, d) \sim (1, 1) \iff xc = d$$

Hence $y = (c, xc)$ satisfies this relation. However we can see that $(c, cx) \sim (1, x)$. Hence for any integer $x \neq 0$ we have a solution to $x * y = 1$ with $y \in \mathbb{Q}$. We call y a multiplicative inverse of x and x a multiplicative inverse of y .

Definition 1.6.17. *Multiplicative inverse of an integer*

Let $x \in \mathbb{Z}$ be such that $x \neq 0$. Then there is a $y \in \mathbb{Q}$ such that

$$x * y = 1 = y * x$$

where $y = (1, x)$. We can write this as $y = \frac{1}{x}$ or $y = x^{-1}$. We sometimes say that x^{-1} is a reciprocal of x or a multiplicative inverse of x .

In light of this, we have the immediate result

Proposition 1.6.4. *Multiplicative inverse of an integer times its multiplicative inverse is the original number*

Let $x \in \mathbb{Z}$ so that $x^{-1} \in \mathbb{Q}$ where $x^{-1} = \frac{1}{x}$ is the multiplicative inverse to x in the rationals. The following result holds.

$$x * x^{-1} * x = x$$

Proof:

By definition of a multiplicative inverse we have that

$$x * x^{-1} = x * \frac{1}{x} = (x, 1) * (1, x) = (x, x) \sim (1, 1) = 1$$

Hence as x^{-1} is a multiplicative inverse for x it follows that x is a multiplicative inverse for x^{-1} and so

$$x * x^{-1} * x = 1 * x = x$$

As required. \square

Armed with this definition we can answer the original question. In order to find an x so that $x * y = 1$ we have that we need to find a multiplicative inverse for c and a multiplicative inverse for $d^{-1} = \frac{1}{d}$. Clearly we have that $c^{-1} = \frac{1}{c}$ and a multiplicative inverse for d^{-1} is simply d . Hence a candidate for x is given by $x = (d, c)$. Indeed we have that

$$x * y = (d, c) * (c, d) = (cd, cd) \sim (1, 1) = 1$$

We can hence extend the idea of multiplicative inverses to the rationals.

Definition 1.6.18. *Multiplicative inverse of a rational number*

Let $x \in \mathbb{Q}$ such that $x = (a, b)$ with $a, b \in \mathbb{Z}$ and $b \neq 0$. Then there is a $y \in \mathbb{Q}$ such that

$$x * y = 1 = y * x$$

where $y = (b, a)$. Hence we must also have $a \neq 0$. We write this as $y = \frac{b}{a}$ or as $x^{-1} = y = \frac{b}{a}$. We sometimes say that x^{-1} is a reciprocal of x or a multiplicative inverse of x .

A similar result holds as for proposition 1.6.4

Proposition 1.6.5. *Multiplicative inverse of a rational number times its multiplicative inverse is the original number*

Let $x \in \mathbb{Q}$ with $x = (a, b)$ and $a, b \in \mathbb{Z}$ so that $a \neq 0$ and $b \neq 0$. Let x^{-1} denote the multiplicative inverse of x . The following result holds.

$$x * x^{-1} * x = x$$

Proof:

By definition of a multiplicative inverse we have that

$$x * x^{-1} = 1$$

Hence

$$x * x^{-1} * x = 1 * x = x$$

As required. \square

We now have a solid grasp of undoing multiplication in the rational numbers. In fact we are now in a position to define the operation of division. However we are already done due to the work we have just done, and our original motivation for defining the rational numbers in the first place. We use the idea of multiplicative inverses!

Definition 1.6.19. *Division*

Let $a, b \in \mathbb{Z}$ so that $b \neq 0$. We define the division of a by b , denoted $\frac{a}{b}$ by

$$\frac{a}{b} = a * b^{-1} = (a, 1) * (1, b) = (a, b)$$

We can extend the notion of division even further by considering $a, b \in \mathbb{Q}$ rather than \mathbb{Z} . At first it appears we have a problem, we defined the rationals using integers and division in terms of integers, so how could we possibly assign any meaning to an expression like $\frac{1}{\frac{1}{2}}$?

Consider for example the following

$$\frac{1}{\frac{1}{2}} * \frac{1}{2}$$

If we were suppose the rule for multiplication that we defined extends to this situation then we get

$$\frac{1}{\frac{1}{2}} * \frac{1}{2} = \frac{1 * 1}{\frac{1}{2} * 2} = \frac{1}{1} = 1$$

In the context of the work we have just done we have that $\frac{1}{\frac{1}{2}}$ is a multiplicative inverse of $\frac{1}{2}$. However we know that $\frac{1}{2}$ has a multiplicative inverse of 2. Does this mean that $\frac{1}{\frac{1}{2}} = 2$? A deeper analysis of expressions of the form $\frac{1}{\frac{1}{a}}$.

We know from before that $\frac{1}{a} = a^{-1}$ for some non-zero $a \in \mathbb{Z}$. Hence we have that by definition $a^{-1} \in \mathbb{Z}$. Hence we are considering the expression

$$\frac{1}{\frac{1}{a}} = \frac{1}{a^{-1}}$$

Therefore we know from the definition of the multiplicative inverse of a rational number that there is some $y \in \mathbb{Q}$ so that

$$\frac{1}{a^{-1}} * y = 1$$

By the definition we also know what y must be $\frac{a^{-1}}{1} = a^{-1} = \frac{1}{a}$. Hence we can justify our “temporary” assumption of extending the multiplication rule. Hence hence make the following deduction

Proposition 1.6.6. *One divided by multiplicative inverse of an integer is the integer itself*

Let $x \in \mathbb{Q}$ so that $x = \frac{1}{\frac{1}{a}}$ for some $a \in \mathbb{Z}$ with $a \neq 0$. we have that

$$\frac{1}{\frac{1}{a}} = a$$

Proof:

Let $x \in \mathbb{Q}$ be such that $x = \frac{1}{\frac{1}{a}}$ for some non-zero $a \in \mathbb{Z}$. We know by definition that

$$x = \frac{1}{\frac{1}{a}} = a^{-1}$$

where $a^{-1} \in \mathbb{Z}$ and therefore $x = \frac{1}{a^{-1}}$. Moreover this is still a rational number by definition and so there exists some rational y so that

$$x * y = 1$$

*where $y = \frac{a^{-1}}{1} = a^{-1}$. It follows that $y = \frac{1}{a}$. Again by definition there is some $z \in \mathbb{Q}$ so that $y * z = 1$ where $z = \frac{a}{1} = a$ that is to say z is a multiplicative inverse of y .*

We therefore have that

$$x * y = 1 = y * z$$

Hence by theorem 1.6.6 we have that $x = z$ which is to say

$$\frac{1}{\frac{1}{a}} = a$$

As required. \square

We hence get an immediate corollary

Corollary 1.6.1. *One divided by rational number*

Let $x \in \mathbb{Q}$ be such that $x = \frac{a}{b}$. We have that

$$\frac{1}{x} = \frac{1}{\frac{a}{b}} = \frac{b}{a}$$

Proof:

We have

$$\frac{1}{x} = \frac{1}{\frac{a}{b}} = \frac{1}{a \frac{1}{b}} = \frac{1}{a b^{-1}} = \frac{1}{a} * \frac{1}{b^{-1}} = \frac{1}{a} b = \frac{b}{a}$$

As required. \square

1.6.14 Extending the summation and product notations to the rationals

Summation and product notation has been defined on the naturals as well as the integers. We can extend the notation to include the rational numbers.

Let $q \in \mathbb{Q}^{n+m+1}$ be an ordered $n + m + 1$ tuple of rational numbers where

$$q = (q_{-m}, q_{-m+1}, \dots, q_{-1}, q_0, q_1, \dots, q_n)$$

Define $\mathbb{Z}_m^n = \{-m, -m+1, -m+2, \dots, -1, 0, 1, \dots, n-1, n\}$ to be a set of indices and define $f : \mathbb{Z}_m^n \rightarrow \mathbb{Q}$ by

$$\begin{aligned} f : \mathbb{Z}_m^n &\rightarrow \mathbb{Q} \\ i &\mapsto f(i) = q_i \end{aligned}$$

Definition 1.6.20. *Summation notation for rational numbers*

Let $z \in \mathbb{Q}^{n+m+1}$ be ordered $n + m + 1$ tuple of integers where $q = (q_{-m}, q_{-m+1}, \dots, q_{-1}, q_0, q_1, \dots, q_n)$. Define \mathbb{Z}_m^n by $\mathbb{Z}_m^n = \{-m, -m+1, -m+2, \dots, -1, 0, 1, \dots, n-1, n\}$. Let $f : \mathbb{Z}^{n+m+1} : \mathbb{Q}$ defined by

$$\begin{aligned} f : \mathbb{Z}^{n+m+1} &\rightarrow \mathbb{Q} \\ i &\mapsto f(i) = q_i \end{aligned}$$

We define the summation notation for the rational numbers by

$$\sum_{i=-m}^n f(i) = f(-m) + f(-m+1) + \dots + f(-1) + f(0) + f(1) + \dots + f(n)$$

Alternatively this is written

$$\sum_{i=-m}^n q_i = q_{-m} + q_{-m+1} + \dots + q_{-1} + q_0 + q_1 + \dots + q_n$$

We have that i is called the index of summation and that $i = -m$ is the starting index of the summation, and n the ending index of the summation. If $q \in \emptyset$ then we define the summation to be 0 and call the summation an empty sum.

We can also define the summation of some subset of \mathbb{Z}_m^n which allows for starting a summation at some starting point other than $i = -m$. Let $T \subseteq \mathbb{Z}_m^n$. We define the summation over the set T by

$$\sum_{i \in T} z_i$$

If we have a mapping $g : \mathbb{Q} \rightarrow \mathbb{Q}$ we can define a summation over g by

$$\sum_{i \in T} g(z_i)$$

Finally we can define a summation over a predicate $P(i)$ for $i \in T$ by

$$\sum_{P(i)} g(z_i)$$

where we take the sum of the $g(z_i)$ for the i that satisfy the predicate P . We note that if we have $k > n$ for some $k \in \mathbb{N}$ then the sum

$$\sum_{i=k}^n z_i = 0$$

The usual proprieties shown for summations with integer numbers also extend to the rational number version.

Proposition 1.6.7. *Properties of summation notation*

Let $n, m \in \mathbb{Z}$ such that $m < n$. Let $s, t \in \mathbb{Q}^{n+m+1}$ and let $c \in \mathbb{Q}$.

Let $a, b \in \mathbb{Z}$ with $m < a < b < n$. Define $A = \mathbb{Z}_a^b$ and define

$$B = \mathbb{Z}_m^n \setminus A = \{-m, -m+1, \dots, a-1, b+1, \dots, n-1, n\}$$

so that $A \cup B = \mathbb{Z}_m^n$. Let $k \in \mathbb{Z}$ be the starting index summation such that $k < n$. We have that the following properties hold.

1. $\sum_{i=-m}^n s_i = \sum_{i \in A} s_i + \sum_{i \in B} s_i = \sum_{i=-m}^{-1} s_i + \sum_{i=0}^n s_i$
2. $\sum_{i=k}^n s_i = \sum_{i=k}^d s_i + \sum_{i=d+1}^n s_i$
3. $\sum_{i=k}^n c * s_i = c * \sum_{i=k}^n s_i$ for some $c \in \mathbb{Q}$
4. $\sum_{i=k}^n s_i + t_i = \sum_{i=k}^n s_i + \sum_{i=k}^n t_i$

Proof:

1. $\sum_{i=-m}^n s_i = \sum_{i \in A} s_i + \sum_{i \in B} s_i = \sum_{i=-m}^{-1} s_i + \sum_{i=0}^n s_i$

The proof is the same as for the integer case. We give it again for completeness

We have that

$$\begin{aligned} \sum_{i=-m}^n s_i &= s_{-m} + s_{-m+1} + s_{-m+2} + \dots + s_{-1} + s_0 + s_1 + \dots + s_{n-1} + s_n \\ &= (s_{-m} + s_{-m+1} + s_{-m+2} + \dots + s_{-1}) + (s_0 + s_1 + \dots + s_{n-1} + s_n) \\ &= \sum_{i=-m}^{-1} s_i + \sum_{i=0}^n s_i \end{aligned}$$

Additionally note that

$$\begin{aligned} \sum_{i=-m}^n s_i &= s_{-m} + s_{-m+1} + s_{-m+2} + \dots + s_{-1} + s_0 + s_1 + \dots + s_{n-1} + s_n \\ &= (s_{-m} + s_{-m+1} + s_{-m+2} + \dots + s_{a-2} + s_{a-1}) + (s_a + s_{a+1} + \dots + s_{b-1} + s_b) \\ &\quad + (s_{b+1} + s_{b+2} + \dots + s_{n-1} + s_n) \\ &= (s_{-m} + s_{-m+1} + s_{-m+2} + \dots + s_{a-2} + s_{a-1}) + (s_{b+1} + s_{b+2} + \dots + s_{n-1} + s_n) \\ &\quad + (s_a + s_{a+1} + \dots + s_{b-1} + s_b) \\ &= \sum_{i \in B} s_i + \sum_{i \in A} s_i = \sum_{i \in A} s_i + \sum_{i \in B} s_i \end{aligned}$$

$$2. \sum_{i=k}^n s_i = \sum_{i=k}^d s_i + \sum_{i=d+1}^n s_i:$$

The proof is similar to part 1, replacing $-m$ by k .

$$3. \sum_{i=k}^n c * s_i = c * \sum_{i=k}^n s_i \text{ for some } c \in \mathbb{Q}$$

We have by definition that

$$\sum_{i=k}^n c * s_i = c * s_k + c * s_{k+1} + c * s_{k+3} + \cdots + c * s_n$$

By multiplication distributing over addition we have

$$\sum_{i=1}^n c * s_i = c * s_k + c * s_{k+1} + c * s_{k+3} + \cdots + c * s_n = c (s_k + s_{k+1} + \cdots + s_n) = c * \sum_{i=k}^n s_i$$

$$4. \sum_{i=k}^n s_i + t_i = \sum_{i=k}^n s_i + \sum_{i=k}^n t_i:$$

This follows by the definition. We have

$$\begin{aligned} \sum_{i=k}^n s_i + t_i &= (s_k + t_k) + (s_{k+1} + t_{k+1}) + \cdots \\ &\quad + (s_{-1} + t_{-1}) + (s_0 + t_0) + (s_1 + t_1) + \cdots + (s_{n-1} + t_{n-1}) + (s_n + t_n) \\ &= (s_k + s_{k+1} + \cdots + s_{-1} + s_0 + s_1 + \cdots + s_{n-1} + s_n) + \\ &\quad + (t_k + t_{k+1} + \cdots + t_{-1} + t_0 + t_1 + \cdots + t_{n-1} + t_n) \\ &= \sum_{i=k}^n s_i + \sum_{i=k}^n t_i \end{aligned}$$

□

We make a similar definition for product notation.

Definition 1.6.21. *Product notation for the rational numbers*

Let $z \in \mathbb{Q}^{n+m+1}$ be ordered $n+m+1$ tuple of integers where $q = (q_{-m}, q_{-m+1}, \dots, q_{-1}, q_0, q_1, \dots, q_n)$. Define \mathbb{Z}_m^n by $\mathbb{Z}_m^n = \{-m, -m+1, -m+2, \dots, -1, 0, 1, \dots, n-1, n\}$. Let $f : \mathbb{Z}^{n+m+1} \rightarrow \mathbb{Z}$ defined by

$$\begin{aligned} f : \mathbb{Z}^{n+m+1} &\rightarrow \mathbb{Q} \\ i &\mapsto f(i) = z_i \end{aligned}$$

We define the summation notation for integers by

$$\prod_{i=-m}^n f(i) = f(-m) * f(-m+1) * \cdots * f(-1) * f(0) * f(1) * \cdots * f(n)$$

Alternatively this is written

$$\prod_{i=-m}^n q_i = q_{-m} * q_{-m+1} * \cdots * q_{-1} * q_0 * q_1 * \cdots * q_n$$

We have that i is called the index of the product and that $i = -m$ is the starting index of the product, and n the ending index of the product. If $z \in \emptyset$ then we define the product to be 1 and call a product an empty product.

We can also define the product of some subset of \mathbb{Z}_m^n which allows for starting a product at some starting point other than $i = -m$. Let $T \subseteq \mathbb{Z}_m^n$. We define the product over the set T by

$$\prod_{i \in T} z_i$$

If we have a mapping $g : \mathbb{Z} \rightarrow \mathbb{Z}$ we can define a product over g by

$$\prod_{i \in T} g(z_i)$$

Finally we can define a product over a predicate $P(i)$ for $i \in T$ by

$$\prod_{P(i)} g(z_i)$$

where we take the sum of the $g(z_i)$ for the i that satisfy the predicate P . We note that if we have $k > n$ for some $k \in \mathbb{N}$ then the product

$$\prod_{i=k}^n z_i = 1$$

Proposition 1.6.8. *Properties of product notation*

Let $n, m \in \mathbb{Z}$ such that $m < n$. Let $s, t \in \mathbb{Q}^{n+m+1}$ and let $c \in \mathbb{Z}$. Let $a, b \in \mathbb{Z}$ so that $m < a < b < n$. Define $A = \mathbb{Z}_a^b$ and define

$$B = \mathbb{Z}_m^n \setminus A = \{-m, -m+1, \dots, a-1, b+1, \dots, n-1, n\}$$

so that $A \cup B = \mathbb{Z}_m^n$. Let $k \in \mathbb{Z}$ be the lower index of the product.

We have that the following properties hold.

1. $\prod_{i=-m}^n s_i = \prod_{i \in A} s_i * \prod_{i \in B} s_i = \prod_{i=-m}^{-1} s_i * \prod_{i=0}^n s_i$
2. $\prod_{i=k}^n s_i = \prod_{i=k}^m s_i * \prod_{i=m+1}^n s_i$
3. $\prod_{i=k}^n s_i t_i = \prod_{i=k}^n s_i \prod_{i=1}^n t_i$

Proof:

1. $\prod_{i=-m}^n s_i = \prod_{i \in A} s_i * \prod_{i \in B} s_i = \prod_{i=-m}^{-1} s_i * \prod_{i=0}^n s_i$

The proof is the same for the integer case.

We have that

$$\begin{aligned} \prod_{i=-m}^n s_i &= s_{-m} * s_{-m+1} * s_{-m+2} * \dots * s_{-1} * s_0 * s_1 * \dots * s_{n-1} * s_n \\ &= (s_{-m} * s_{-m+1} * s_{-m+2} * \dots * s_{-1}) * (s_0 * s_1 * \dots * s_{n-1} * s_n) \\ &= \prod_{i=-m}^{-1} s_i * \prod_{i=0}^n s_i \end{aligned}$$

Likewise we have

$$\begin{aligned}
\prod_{i=-m}^n s_i &= s_{-m} * s_{-m+1} * s_{-m+2} * \cdots * s_{-1} * s_0 * s_1 * \cdots * s_{n-1} * s_n \\
&= (s_{-m} * s_{-m+1} * s_{-m+2} * \cdots * s_{a-2} * s_{a-1}) * (s_a * s_{a+1} * \cdots * s_{b-1} * s_b) \\
&\quad * (s_{b+1} * s_{b+2} * \cdots * s_{n-1} * s_n) \\
&= (s_{-m} * s_{-m+1} * s_{-m+2} * \cdots * s_{a-2} * s_{a-1}) * (s_{b+1} * s_{b+2} * \cdots * s_{n-1} * s_n) \\
&\quad * (s_a * s_{a+1} * \cdots * s_{b-1} * s_b) \\
&= \prod_{i \in B} s_i * \prod_{i \in A} s_i = \prod_{i \in A} s_i * \prod_{i \in B} s_i
\end{aligned}$$

$$2. \prod_{i=k}^n s_i = \prod_{i=k}^m s_i * \prod_{i=m+1}^n s_i.$$

The proof is similar to part 1. We replace $-m$ with k .

$$3. \prod_{i=k}^n s_i t_i = \prod_{i=k}^n s_i \prod_{i=1}^n t_i.$$

Observe that

$$\begin{aligned}
\prod_{i=k}^n s_i t_i &= s_k t_k * s_{k+1} t_{k+1} * s_{k+2} t_{k+2} * \cdots * s_{-1} t_{-1} * s_0 t_0 * s_1 t_1 * \cdots * s_{n-1} t_{n-1} * s_n t_n \\
&= (s_k * s_{k+1} * s_{k+2} * \cdots * s_{-1} * s_0 * s_1 * \cdots * s_{n-1} * s_n) \\
&\quad * (t_k * t_{k+1} * t_{k+2} * \cdots * t_{-1} * t_0 * t_1 * \cdots * t_{n-1} * t_n) \\
&= \prod_{i=k}^n s_i * \prod_{i=k}^n s_i
\end{aligned}$$

□

We can now extend the result of proposition 1.3.3 and proposition 1.5.9. I.e if the product of $ab = 0$ for $a, b \in \mathbb{Q}$ then at least one of a or b is zero.

Proposition 1.6.9. *Product of two rational numbers being zero implies one of the numbers is zero*

Let $x, y \in \mathbb{Q}$. If $xy = 0$ then at least one of x or y is zero.

Proof:

Let $x, y \in \mathbb{Q}$. If $x = y = 0$ then the result is trivial. So suppose that $x = (a, b)$ and $y = (c, d)$, moreover suppose $y \neq 0$. By definition of rational number multiplication we have that

$$xy = (a, b) * (c, d) = (ac, bd) = (0, 1)$$

Hence we must have $ac = 0$. Therefore by proposition 1.5.9 we must have that either $a = 0$ or $c = 0$ or both. As we have assumed $y \neq 0$ then $a = 0$ and so $x = 0$. A similar argument assuming $x \neq 0$ shows that $y = 0$. The result is shown. □

1.6.15 Extending the rules for inequalities to the integers

For the natural numbers and the integers, we have a theory of inequalities. These results extend to the rationals. Additionally, as rational numbers represent the division of integers there are some additional properties that now hold.

we were able to derive some rules for how inequalities behave, we can extend those results to the integers. Before we do so we have an additional consideration.

To extend the results fully we need to consider negative rational numbers as well. We follow a similar layout to the section on integer inequalities.

Proposition 1.6.10. *Multiplication by -1 changes the inequality sign*

Let $x, y \in \mathbb{Q}$. We have the following

1. *If $x < y$ then $-x > -y$*
2. *If $x \leq y$ then $-x \geq -y$*
3. *If $x > y$ then $-x < -y$*
4. *If $x \geq y$ then $-x \leq -y$*

Proof:

Let $x, y \in \mathbb{Q}$ so that $x = \frac{a}{b}$ and $y = \frac{c}{d}$ where $b \neq 0$ and $d \neq 0$.

1. *If $x < y$ then $-x > -y$:*

Let $x, y \in \mathbb{Q}$ so that $x < y$. By definition of $<$ for the rationals we have that

$$x < y \iff ad < bc$$

Applying proposition 1.5.10 we have

$$ad < bc \Rightarrow -ad > -bc$$

Hence $-x > -y$.

2. *If $x \leq y$ then $-x \geq -y$:*

Let $x, y \in \mathbb{Q}$ so that $x \leq y$. Applying the definition of \leq for the rationals gives

$$x \leq y \iff ad \leq bc$$

Proposition 1.5.10 gives

$$ad \leq bc \Rightarrow -ad \geq -bc$$

Hence $-x \geq -y$.

3. *If $x > y$ then $-x < -y$:*

Let $x, y \in \mathbb{Q}$ so that $x > y$. By definition of $>$ for the rationals we have that

$$x > y \iff ad > bc$$

Proposition 1.5.10 shows us that

$$ad > bc \Rightarrow -ad < -bc$$

Hence $-x < -y$.

4. If $x \geq y$ then $-x \leq -y$:

Let $x, y \in \mathbb{Q}$ so that $x > y$. By definition of \geq for the rationals, we have that

$$x \geq y \iff ad \geq bc$$

Proposition 1.5.10 we have

$$ad \geq bc \Rightarrow -ad \leq -bc$$

Hence $-x \leq -y$.

The result is shown. \square

There is another useful lemma that will be useful for extending the rules of inequalities to the rationals.

Lemma 1.6.2. *Strictly larger rational minus a smaller is positive*

Let $x, y \in \mathbb{Q}$. We have that $x < y \iff y - x > 0$

Proof:

(\Rightarrow) : Let $x, y \in \mathbb{Q}$, then $x = \frac{a}{b}$ and $y = \frac{c}{d}$ for some $a, b, c, d \in \mathbb{Z}$ and $b \neq 0$ and $d \neq 0$. As $x < y$ then $ad < bc$. Now $y - x$ is given by

$$\begin{aligned} y - x &= y + (-1 * x) \\ &= (c, d) + (-a, b) \\ &= (cb - ad, bd) \end{aligned}$$

Now, saying $y - x$ is positive is the same as $y - x > 0$. By definition of greater than, and the fact that $0 \in [(0, 1)]$ we would have that

$$(cb - ad) * 1 > 0 * (bd) \Rightarrow bc - ad > 0$$

Which is true as $ad < bc$. Hence $y - x > 0$

(\Leftarrow) : Suppose that $y - x > 0$ where $x, y \in \mathbb{Q}$, with $x = \frac{a}{b}$ and $y = \frac{c}{d}$ for some $a, b, c, d \in \mathbb{Z}$ and $b \neq 0$ and $d \neq 0$. We have that

$$y - x = (cb - ad, bd)$$

Moreover, $y - x > 0$ implies that

$$(cb - ad) * 1 > 0 * (bd) \Rightarrow bc - ad > 0$$

This is to say $bc > ad$, which by part 1 of proposition 1.5.11 is the same as $ad < bc$ which is equivalent to saying that $x < y$.

As required. \square

Corollary 1.6.2. *Larger or equal rational minus a smaller is positive*

Let $x, y \in \mathbb{Q}$. We have that $x \leq y \iff y - x > 0$ or $y = x$.

Proof:

(\Rightarrow) : Suppose $x \leq y$. If $x < y$ then lemma 1.6.2 applies. Otherwise $x = y$.

(\Leftarrow) : Suppose that one of $y - x > 0$ or $y = x$ holds. In the first case, $y - x > 0$ implies $x < y$ by lemma 1.6.2 and clearly we will have $x \leq y$. If $x = y$ then we clearly also have $x \leq y$ by definition. \square

We can now extend the properties of inequalities to the rationals.

Proposition 1.6.11. *Properties of inequalities for the rationals*

Let $x, y, z, c \in \mathbb{Q}$. We have the following properties for inequalities

1. $x < y$ is the same as $y > x$
2. $x \leq y$ is the same as $y \geq x$
3. If $x < y$ and $y < z$ then $x < z$
4. If $x \leq y$ and $y < z$ then $x < z$
5. If $x < y$ and $y \leq z$ then $x < z$
6. If $x \leq y$ and $y \leq z$ then $x \leq z$
7. If $x > y$ and $y > z$ then $x > z$
8. If $x \geq y$ and $y > z$ then $x > z$
9. If $x > y$ and $y \geq z$ then $x > z$
10. If $x \geq y$ and $y \geq z$ then $x \geq z$
11. If $x < y$ then $x + z < y + z$
12. If $x \leq y$ then $x + z \leq y + z$
13. If $x > y$ then $x + z > y + z$
14. If $x \geq y$ then $x + z \geq y + z$
15. If $x < y$ and $z \geq 0$ then $xz < yz$
16. If $x < y$ and $z < 0$ then $xz > yz$
17. If $x \leq y$ and $z \geq 0$ then $xz \leq yz$
18. If $x \leq y$ and $z < 0$ then $xz \geq yz$
19. If $x > y$ and $z \geq 0$ then $xz > yz$
20. If $x > y$ and $z < 0$ then $xz < yz$
21. If $x \geq y$ and $z \geq 0$ then $xz \geq yz$
22. If $x \geq y$ and $z < 0$ then $xz \leq yz$
23. If $x < y$ and $z > 0$ then $\frac{x}{z} < \frac{y}{z}$
24. If $x \leq y$ and $z > 0$ then $\frac{x}{z} \leq \frac{y}{z}$
25. If $x > y$ and $z > 0$ then $\frac{x}{z} > \frac{y}{z}$
26. If $x \geq y$ and $z > 0$ then $\frac{x}{z} \geq \frac{y}{z}$
27. If $x < y$ and $z < 0$ then $\frac{x}{z} > \frac{y}{z}$
28. If $x \leq y$ and $z < 0$ then $\frac{x}{z} \geq \frac{y}{z}$
29. If $x > y$ and $z < 0$ then $\frac{x}{z} < \frac{y}{z}$
30. If $x \geq y$ and $z < 0$ then $\frac{x}{z} \leq \frac{y}{z}$

31. If $x < y$ and $x > 0$ and $y > 0$ then $\frac{1}{x} > \frac{1}{y}$

32. If $x < y$ and $x < 0$ and $y < 0$ then $\frac{1}{x} > \frac{1}{y}$

33. If $x \leq y$ and $x > 0$ and $y > 0$ then $\frac{1}{x} \geq \frac{1}{y}$

34. If $x \leq y$ and $x < 0$ and $y < 0$ then $\frac{1}{x} \geq \frac{1}{y}$

35. If $x > y$ and $x > 0$ and $y > 0$ then $\frac{1}{x} < \frac{1}{y}$

36. If $x > y$ and $x < 0$ and $y < 0$ then $\frac{1}{x} < \frac{1}{y}$

37. If $x \geq y$ and $x > 0$ and $y > 0$ then $\frac{1}{x} \leq \frac{1}{y}$

38. If $x \geq y$ and $x < 0$ and $y < 0$ then $\frac{1}{x} \leq \frac{1}{y}$

Proof:

Let $x, y, z \in \mathbb{Q}$. Let $x = \frac{a}{b}$, $y = \frac{c}{d}$, $z = \frac{e}{f}$ for $a, b, c, d, e, f, g, h \in \mathbb{Z}$ and $b \neq 0$, $d \neq 0$, $f \neq 0$.

1. $x < y$ is the same as $y > x$:

Suppose that $x < y$ then by definition we have $ad < bc$. Applying part 1. of proposition 1.5.11 gives $be > af$ and so $y > x$.

2. $x \leq y$ is the same as $y \geq x$:

If $x < y$ then part 1 applies, Otherwise we have $x = y$ and so $y = x$ and clearly $y \geq x$.

3. If $x < y$ and $y < z$ then $x < z$:

Suppose that $x < y$ and $y < z$ then $y - x > 0$ and $z - y > 0$ by lemma 1.6.2. Now we have that

$$(y - x) + (z - y) = z - x > 0$$

As $y - x$ and $z - y$ are both greater than 0. Hence as $z - x > 0$ then $x < z$.

4. If $x \leq y$ and $y < z$ then $x < z$:

Suppose that $x \leq y$ and $y < z$. If $x < y$ then the previous part applies, so suppose not. Then $x = y$ and so $x < z$.

5. If $x < y$ and $y \leq z$ then $x < z$:

Suppose that $x < y$ and $y \leq z$. By lemma 1.6.2 we have that $y - x > 0$, likewise by corollary 1.6.2 we have that $y \leq z$ means either $z - y > 0$ or $y = z$.

If $z - y > 0$ then the result is the same as part 3. So suppose $y = z$ then clearly $x < z$.

6. If $x \leq y$ and $y \leq z$ then $x \leq z$:

If $x \leq y$ and $y \leq z$ then either $x < y$ and $y < z$ in which case part 3. applies, or $x < y$ and $y \leq z$ so part 5. applies, or $x \leq y$ and $y < z$ so part 4 applies. Finally, we have the case $x = y$ and $y = z$ so clearly $x = z$ so that $x \leq z$.

7. If $x > y$ and $y > z$ then $x > z$:

By part 1. this is equivalent to $y < x$ and $z < y$ then $z < x$ so part 3. applies.

8. If $x \geq y$ and $y > z$ then $x > z$:

Using parts 1. and 2. gives us the equivalent expression $y \leq x$ and $z < y$ then $z < x$ and so part 4 applies.

9. If $x > y$ and $y \geq z$ then $x > z$:

As with the previous part, applying parts 1. and 2. gives the statement $y < x$ and $z \leq y$ then $z < x$ so part 5. applies.

10. If $x \geq y$ and $y \geq z$ then $x \geq z$:

Using part 2. gives us $y \leq x$ and $z \leq y$ then $z \leq x$ so part 6. applies.

11. If $x < y$ then $x + z < y + z$:

Suppose that $x < y$ then $y - x > 0$ by lemma 1.6.2. Observe that

$$\begin{aligned} y - x &= y - (z - z) - x \\ &= (y - z) + (z - x) \\ &= (y - z) - (x - z) > 0 \end{aligned}$$

So $(y - z) - (x - z) > 0$ and so by the same lemma we conclude that $x + z < y + z$.

12. If $x \leq y$ then $x + z \leq y + z$:

If $x < y$ then the previous part applies. Otherwise $x = y$ and clearly $x + z = y + z$ and so $x + z \leq y + z$.

13. If $x > y$ then $x + z > y + z$:

Applying part 1. and then part 11. gives the equivalent result $y < x$ then $y + z < x + z$.

14. If $x \geq y$ then $x + z \geq y + z$:

Applying part 2. and then part 12. gives the equivalent result $y \leq x$ then $y + z \leq x + z$.

15. If $x < y$ and $z \geq 0$ then $xz < yz$:

Suppose $x < y$ then $y - x > 0$ by lemma 1.6.2. Hence, by distributivity, we have $z(y - x) > 0$ as $z \geq 0$. Hence

$$z(y - x) = zy - zx = yz - xz \Rightarrow xz < yz$$

16. If $x < y$ and $z < 0$ then $xz > yz$:

Suppose $x < y$, as $z < 0 \Rightarrow -z > 0$, then applying part 15. with $-z$ gives $-xz < -yz$. Finally by proposition 1.6.10 part 1 yields $xz > yz$.

17. If $x \leq y$ and $z \geq 0$ then $xz \leq yz$:

If $x \leq y$ there are two cases to consider. If $x < y$ then part 15. applies. Otherwise $x = y$ and clearly $xz = yz$ giving $xz \leq yz$.

18. If $x \leq y$ and $z < 0$ then $xz \geq yz$:

Likewise, if $x \leq y$ there are two cases. The case $x < y$ is covered by part 16. Otherwise $x = y$ gives $xz = yz$ and again we have $xz \geq yz$.

19. If $x > y$ and $z \geq 0$ then $xz > yz$:

We have $x > y$ is the same as $y < x$ and so $x - y > 0$. By distributivity, we have that $z(x - y) > 0$. Therefore we have $zx - zy = xz - yz > 0$ and so $yz < xz$ which is the same as $xz > yz$ by part 1.

20. If $x > y$ and $z < 0$ then $xz < yz$:

We have $x > y$. Additionally, $z < 0 \Rightarrow -z > 0$ so applying part 19. gives $-xz > -yz$ and so by part 1. we conclude $xz < yz$.

21. If $x \geq y$ and $z \geq 0$ then $xz \geq yz$:

There are two cases to consider. If $x > y$ then we apply part 19. Otherwise $x = y$ and $xz = yz$ so that $xz \geq yz$.

22. If $x \geq y$ and $z < 0$ then $xz \leq yz$:

Again there are two cases to consider. If $x > y$ then the result holds by part 20. Otherwise $x = y$ and so $xz = yz$ to give the result $xz \leq yz$.

23. If $x < y$ and $z > 0$ then $\frac{x}{z} < \frac{y}{z}$:

This follows by part 15.

24. If $x \leq y$ and $z > 0$ then $\frac{x}{z} \leq \frac{y}{z}$:

This follows by part 17.

25. If $x > y$ and $z > 0$ then $\frac{x}{z} > \frac{y}{z}$:

This follows by part 19.

26. If $x \geq y$ and $z > 0$ then $\frac{x}{z} \geq \frac{y}{z}$:

This follows by part 21.

27. If $x < y$ and $z < 0$ then $\frac{x}{z} > \frac{y}{z}$:

This follows by part 16.

28. If $x \leq y$ and $z < 0$ then $\frac{x}{z} \geq \frac{y}{z}$:

This follows by part 18.

29. If $x > y$ and $z < 0$ then $\frac{x}{z} < \frac{y}{z}$:

This follows by part 20.

30. If $x \geq y$ and $z < 0$ then $\frac{x}{z} \leq \frac{y}{z}$:

This follows by part 22.

31. If $x < y$ and $x > 0$ and $y > 0$ then $\frac{1}{x} > \frac{1}{y}$:

Suppose that $x < y$ then $ad < bc$. Moreover as $x > 0$ that either $a > 0$ and $b > 0$ or $a < 0$ and $b < 0$. Likewise as $y > 0$ then either $c > 0$ and $d > 0$ or $c < 0$ and $d < 0$. Hence there are four cases to consider.

(a) $a > 0$ and $b > 0$ and $c > 0$ and $d > 0$

(b) $a > 0$ and $b > 0$ and $c < 0$ and $d < 0$

(c) $a < 0$ and $b < 0$ and $c > 0$ and $d > 0$

(d) $a < 0$ and $b < 0$ and $c < 0$ and $d < 0$

(a) $a > 0$ and $b > 0$ and $c > 0$ and $d > 0$:

Observe that

$$\begin{aligned}
 ad &< bc \\
 a^{-1}ad &< a^{-1}bc, \text{ By part 15. } asa^{-1} > 0 \\
 d &< a^{-1}bc, \text{ As multiplication of an element by its inverse is 1} \\
 dc^{-1} &< a^{-1}bcc^{-1}, \text{ By part 15. as } c^{-1} > 0 \\
 dc^{-1} &< a^{-1}b, \text{ As multiplication of an element by its inverse is 1} \\
 \frac{d}{c} &< \frac{b}{a}, \text{ By the definition of an inverse element}
 \end{aligned}$$

Hence $\frac{d}{c} < \frac{b}{a}$ which is equivalent to $\frac{b}{a} > \frac{d}{c}$, which is to say $\frac{1}{x} > \frac{1}{y}$.

(b) $a > 0$ and $b > 0$ and $c < 0$ and $d < 0$:

We have that as $c < 0$ and $d < 0$ then $ad < 0$ and $bc < 0$ and $ad < bc$. Hence observe that

$$\begin{aligned}
 ad &< bc \\
 a^{-1}ad &> a^{-1}bc, \text{ By part 16. as } a^{-1} < 0 \\
 d &> a^{-1}bc \\
 dc^{-1} &< a^{-1}bcc^{-1}, \text{ By part 20. as } c^{-1} < 0 \\
 dc^{-1} &< a^{-1}b \\
 \frac{d}{c} &< \frac{b}{a}
 \end{aligned}$$

Hence we again conclude that $\frac{1}{x} > \frac{1}{y}$.

(c) $a < 0$ and $b < 0$ and $c > 0$ and $d > 0$:

The argument is similar to the previous one, swapping the roles of a, b, c and d .

(d) $a < 0$ and $b < 0$ and $c < 0$ and $d < 0$:

This is similar to the first part. We give the full argument. As $a < 0$, $b < 0$, $c < 0$ and $d < 0$ then $ad > 0$ and $bc > 0$ and $ad < bc$. Hence we can see that

$$\begin{aligned}
 ad &< bc \\
 a^{-1}ad &> a^{-1}bc, \text{ By part 16. as } a^{-1} < 0 \\
 d &> a^{-1}bc \\
 dc^{-1} &< a^{-1}bcc^{-1}, \text{ By part 20. as } c^{-1} < 0 \\
 dc^{-1} &< a^{-1}b \\
 \frac{d}{c} &< \frac{b}{a}
 \end{aligned}$$

Giving the result.

32. If $x < y$ and $x < 0$ and $y < 0$ then $\frac{1}{x} > \frac{1}{y}$:

This is similar to the previous part. Suppose that $x < y$ then $ad < bc$. Moreover as $x < 0$ that either $a > 0$ and $b < 0$ or $a < 0$ and $b > 0$. Likewise as $y < 0$ then either $c > 0$ and $d < 0$ or $c < 0$ and $d > 0$. Hence there are four cases to consider.

- (a) $a > 0$ and $b < 0$ and $c > 0$ and $d < 0$
(b) $a > 0$ and $b < 0$ and $c < 0$ and $d > 0$
(c) $a < 0$ and $b > 0$ and $c > 0$ and $d < 0$
(d) $a < 0$ and $b > 0$ and $c < 0$ and $d > 0$

- (a) $a > 0$ and $b < 0$ and $c > 0$ and $d < 0$:

As $a > 0$ and $b < 0$ and $c > 0$ and $d < 0$ then we have that $ad < 0$ and $bc < 0$ and $ad < bc$. We have that

$$\begin{aligned} ad &< bc \\ a^{-1}ad &< a^{-1}bc \\ d &< a^{-1}bc \\ dc^{-1} &< a^{-1}bcc^{-1} \\ dc^{-1} &< a^{-1}b \\ \frac{d}{c} &< \frac{b}{a} \end{aligned}$$

Giving $\frac{1}{x} > \frac{1}{y}$

- (b) $a > 0$ and $b < 0$ and $c < 0$ and $d > 0$:

We have $a > 0$ and $b < 0$ and $c < 0$ and $d > 0$ then we have that $ad > 0$ and $bc > 0$ and $ad < bc$. We have that

$$\begin{aligned} ad &< bc \\ a^{-1}ad &< a^{-1}bc \\ d &< a^{-1}bc \\ dc^{-1} &< a^{-1}bcc^{-1} \\ dc^{-1} &< a^{-1}b \\ \frac{d}{c} &< \frac{b}{a} \end{aligned}$$

Giving $\frac{1}{x} > \frac{1}{y}$

- (c) $a < 0$ and $b > 0$ and $c > 0$ and $d < 0$:

This time we have $a < 0$ and $b > 0$ and $c > 0$ and $d < 0$ then we have that $ad > 0$ and $bc > 0$ and $ad < bc$

$$\begin{aligned} ad &< bc \\ -ad &> bc \\ a^{-1}(-ad) &> a^{-1}(-bc) \\ -d &> a^{-1}(-bc) \\ -dc^{-1} &> a^{-1}(-bc)c^{-1} \\ -dc^{-1} &> -a^{-1}b \\ \frac{d}{c} &< \frac{b}{a} \end{aligned}$$

Giving the result.

(d) $a < 0$ and $b > 0$ and $c < 0$ and $d > 0$:

Finally, $a < 0$ and $b > 0$ and $c < 0$ and $d > 0$ which gives $ad < 0$ and $bc < 0$ with $ad < bc$. Once again we have that

$$\begin{aligned} ad &< bc \\ a^{-1}ad &> a^{-1}bc \\ d &> a^{-1}bc \\ dc^{-1} &< a^{-1}bcc^{-1} \\ dc^{-1} &< a^{-1}b \\ \frac{d}{c} &< \frac{b}{a} \end{aligned}$$

Which concludes this part of the proposition

33. If $x \leq y$ and $x > 0$ and $y > 0$ then $\frac{1}{x} \geq \frac{1}{y}$:

If $x < y$ then we apply part 31. Otherwise $x = y$ and so $\frac{1}{x} = \frac{1}{y}$ hence the result.

34. If $x \leq y$ and $x < 0$ and $y < 0$ then $\frac{1}{x} \geq \frac{1}{y}$

Likewise if $x < y$ we apply part 32. Otherwise $x = y$ and $\frac{1}{x} = \frac{1}{y}$ so the result is clear.

35. If $x > y$ and $x > 0$ and $y > 0$ then $\frac{1}{x} < \frac{1}{y}$:

Applying part 1. the equivalent statement is $y < x$ and $x < 0$ and $y < 0$ then $\frac{1}{y} > \frac{1}{x}$ so part 32. applies.

36. If $x > y$ and $x < 0$ and $y < 0$ then $\frac{1}{x} < \frac{1}{y}$:

Likewise by part 1. this is the same as $y < x$ and $x > 0$ and $y > 0$ then $\frac{1}{y} > \frac{1}{x}$ so part 31. applies.

37. If $x \geq y$ and $x > 0$ and $y > 0$ then $\frac{1}{x} \leq \frac{1}{y}$:

If $x > y$ then part 35 applies. Otherwise, $x = y$ and the result is clear.

38. If $x \geq y$ and $x < 0$ and $y < 0$ then $\frac{1}{x} \leq \frac{1}{y}$:

Finally, if $x > y$ then we apply part 36. Otherwise $x = y$ and we get the result.

The result has been shown.¹⁰ \square

¹⁰Pheh!

1.6.16 Extending exponentiation to the rational numbers

Recall the definition of exponentiation from the integers.

$$\wedge : \mathbb{Z} \times \mathbb{Z}^+ \rightarrow \mathbb{Z}$$

$$(x, n) \mapsto \wedge(x, n) = \begin{cases} 1, & \text{If } x = 0 \text{ and } n = 0 \\ 1, & \text{If } n = 0 \\ \prod_{i=1}^n x, & \text{If } x \neq 0 \text{ and } n \geq 0 \end{cases}$$

where $\mathbb{Z}^+ = \{x \in \mathbb{Z} : x \geq 0\}$. We noted in the section on extending exponentiation to the integers that we were unable to consider the case of negative exponents. By assuming that they did we deduced that a new type of object exists that undoes integer multiplication. As we have seen in this section, that object type is actually a rational number. Indeed we showed that in proposition 1.6.4 that if $x \in \mathbb{Z}$ then there is some $x^{-1} \in \mathbb{Q}$ so that $x * x^{-1} = 1 = x^0$. This would generalise proposition 1.5.17 to all integers rather than positive exponents. We hence generalise the definition of exponentiation and prove the results to all integer exponents rather than the positive.

Definition 1.6.22. *Exponentiation of integer numbers*

Let $(x, y) \in \mathbb{Z} \times \mathbb{Z}$ and let $\wedge : \mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{Q}$. We define the exponentiation of x by y by

$$\wedge : \mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{Q}$$

$$(x, y) \mapsto \wedge(x, y) = \begin{cases} 1, & \text{If } x = 0 \text{ and } y = 0 \\ 1, & \text{If } x = 0 \\ \prod_{i=1}^y x, & \text{If } x \neq 0 \text{ and } y \geq 0 \\ \prod_{i=1}^{|y|} \frac{1}{x}, & \text{If } x \neq 0 \text{ and } y < 0 \end{cases}$$

We can now extend the results shown in the section on integer exponentiation extension.

Proposition 1.6.12. *Power law of exponentiation for positive exponents*

Let $x \in \mathbb{Z}$ and let $n, m \in \mathbb{Z}$. We have that

$$(x^n)^m = x^{nm}$$

Proof:

If $n, m \geq 0$ the result is the same as proposition 1.5.16. So we must consider the following cases

1. $n \geq 0$ and $m < 0$

2. $n < 0$ and $m \geq 0$

3. $n < 0$ and $m < 0$

1. $n \geq 0$ and $m < 0$:

By definition of integer exponentiation, we have that $x^n = \prod_{i=1}^n x$. Now applying the general definition of integer exponentiation we see that

$$\begin{aligned}
(x^n)^m &= \prod_{i=1}^{|m|} \frac{1}{x^n} \\
&= \underbrace{\left(\frac{1}{x^n}\right) * \left(\frac{1}{x^n}\right) * \left(\frac{1}{x^n}\right) * \cdots * \left(\frac{1}{x^n}\right)}_{|m| \text{ times}}
\end{aligned}$$

Now, we know by definition of multiplication for rationals that $\frac{1}{a} * \frac{1}{b} = \frac{1}{ab}$ and so.

$$\begin{aligned}
(x^n)^m &= \underbrace{\left(\frac{1}{x^n}\right) * \left(\frac{1}{x^n}\right) * \left(\frac{1}{x^n}\right) * \cdots * \left(\frac{1}{x^n}\right)}_{|m| \text{ times}} \\
&= \frac{1}{x^{n|m|}} \\
&= \prod_{i=1}^{n|m|} \frac{1}{x} = x^{nm}
\end{aligned}$$

By definition.

2. $n < 0$ and $m \geq 0$:

As $n < 0$ then we have that

$$x^n = \prod_{i=1}^{|n|} \frac{1}{x} = \frac{1}{x^n}$$

We can now apply similar logic to the first part to conclude the result.

3. $n < 0$ and $m < 0$:

Using similar logic to the two previous parts deduces the result.

As promised. \square

Proposition 1.6.13. *Multiplying exponents of the same base adds the powers*

Let $x \in \mathbb{Z}$ be a fixed integer and let $n, m \in \mathbb{Z}$. We have that

$$x^n * x^m = x^{n+m}$$

Proof:

If $n, m \geq 0$ the result is the same as proposition 1.5.17, so we have to consider the following three cases

1. $n \geq 0$ and $m < 0$

2. $n < 0$ and $m \geq 0$

3. $n < 0$ and $m < 0$

1. $n \geq 0$ and $m < 0$:

Let $m = -k$ for some $k \in \mathbb{Z}$ with $k > 0$. We know that $x^m = x^{-k} = \prod_{i=1}^{-k} \frac{1}{x} = x^{-k}$. Now we have

$$x^n * x^m = x^n x^{-k} = x^{n-k}$$

Which is equivalent to x^{n+m} .

2. $n < 0$ and $m \geq 0$:

Like the previous part let $n = -k$ for some $k \in \mathbb{Z}$ with $k > 0$ then we get

$$x^n * x^m = x^{-k} * x^m = x^{-k+m} = x^{n+m}$$

3. $n < 0$ and $m < 0$:

Let $n = -k$ and $m = -j$ for $k, j \in \mathbb{Z}$ with $k > 0$ and $j > 0$. Then

$$x^n * x^m = x^{-k} * x^{-j} = x^{-k-j} = x^{n+m}$$

As required. \square

Proposition 1.6.14. *Power of product is product of powers*

Let $x, y \in \mathbb{Z}$ and $n \in \mathbb{Z}$. Then

$$(x * y)^n = x^n * y^n$$

Proof:

If $n = 0$ then $(x * y)^n = 1$ and clearly $x^0 * y^0 = 1$. So suppose $n > 0$ then we have

$$\begin{aligned} (x * y)^n &= \prod_{i=1}^n xy = \underbrace{xy * xy * \cdots * xy}_{n \text{ times}} \\ &= \left(\underbrace{x * x * \cdots * x}_{n \text{ times}} \right) * \left(\underbrace{y * y * \cdots * y}_{n \text{ times}} \right), \quad \text{By commutativity of multiplication} \\ &= x^n * y^n \end{aligned}$$

Finally, let $n < 0$ then a similar argument shows that

$$(x * y)^n = \frac{1}{x^n * y^n}$$

Showing the proposition. \square

We have extended integer exponentiation. What can we say about rational exponentiation? We can clearly extend the base of exponentiation to an arbitrary rational number. We have already used special cases of this when we considered denominators and numerators separately in the previous proofs. We formalise this to a fully general rational number. Firstly, we know that if $n < 0$ then $x^n = \frac{1}{x^{-n}}$. Additionally if $x \in \mathbb{Z}$ then a multiplicative inverse of x in the rationals is given by $x^{-1} = \frac{1}{x}$. We combine the two into a general definition.

Definition 1.6.23. *Exponentiation for negative indices*

Let $x \in \mathbb{Z}$ with $x \neq 0$. We extend exponentiation to negative $n \in \mathbb{Z}$ by

$$x^{-n} = (x^{-1})^n$$

Clearly we have in general that $x^{-n} \in \mathbb{Q}$

Now we can consider the more general case of $\left(\frac{a}{b}\right)^n$ for $a, b, n \in \mathbb{Z}$ and $b \neq 0$. We have the following proposition

Proposition 1.6.15. *Rational number raised to an integer exponent*

Let $x \in \mathbb{Q}$ with $x = \frac{a}{b}$ and $b \neq 0$. Let $n \in \mathbb{Z}$. We have that

$$\left(\frac{a}{b}\right)^n = \frac{a^n}{b^n}$$

Proof:

We have that

$$\begin{aligned} \left(\frac{a}{b}\right)^n &= (a * b^{-1})^n \\ &= \underbrace{(ab^{-1}) (ab^{-1}) \dots (ab^{-1})}_{n \text{ times}} \\ &= \underbrace{a * a * a * \dots * a}_{n \text{ times}} * \underbrace{b^{-1} * b^{-1} * b^{-1} * \dots * b^{-1}}_{n \text{ times}} \\ &= a^n (b^{-1})^n \\ &= a^n * b^{-n} \\ &= \frac{a^n}{b^n} \end{aligned}$$

As required. \square

The rules of integer exponentiation extend when the base is rational.

Proposition 1.6.16. *Power law of exponentiation for positive exponents*

Let $x \in \mathbb{Q}$ and let $n, m \in \mathbb{Z}$. We have that

$$(x^n)^m = x^{nm}$$

Proof:

Let $x = \frac{a}{b}$ with $a, b \in \mathbb{Z}$ and $b \neq 0$. We have that

$$\begin{aligned} (x^n)^m &= \left(\left(\frac{a}{b}\right)^n\right)^m \\ &= \left(\frac{a^n}{b^n}\right)^m \\ &= (a^n * b^{-n})^m \\ &= a^{nm} * b^{-nm} \\ &= \frac{a^{nm}}{b^{nm}} \\ &= x^{nm} \end{aligned}$$

\square

Proposition 1.6.17. *Multiplying exponents of the same base adds the powers*

Let $x \in \mathbb{Q}$ be a fixed integer and let $n, m \in \mathbb{Z}$. We have that

$$x^n * x^m = x^{n+m}$$

Proof:

Let $x = \frac{a}{b}$ with $a, b \in \mathbb{Z}$ and $b \neq 0$. Observe that

$$\begin{aligned}
x^n * x^m &= \left(\frac{a}{b}\right)^n * \left(\frac{a}{b}\right)^m \\
&= \frac{a^n}{b^n} * \frac{a^m}{b^m} \\
&= \frac{a^n * a^m}{b^n * b^m} \\
&= \frac{a^{n+m}}{b^{n+m}} \\
&= \left(\frac{a}{b}\right)^{n+m} \\
&= x^{n+m}
\end{aligned}$$

As required. \square

Proposition 1.6.18. *Power of product is product of powers*
Let $x, y \in \mathbb{Q}$ and $n \in \mathbb{Z}$. Then

$$(x * y)^n = x^n * y^n$$

Proof:

Let $x = \frac{a}{b}$ with $a, b \in \mathbb{Z}$ and $b \neq 0$ and let $y = \frac{c}{d}$ with $c, d \in \mathbb{Z}$ and $d \neq 0$. We have

$$\begin{aligned}
(x * y)^n &= \left(\frac{a}{b} * \frac{c}{d}\right)^n \\
&= \left(\frac{ac}{bd}\right)^n \\
&= \frac{(ac)^n}{(bd)^n} \\
&= \frac{a^n c^n}{b^n d^n} \\
&= \frac{a^n}{b^n} * \frac{c^n}{d^n} \\
&= x^n * y^n
\end{aligned}$$

What about rational exponents? Can we assign meaning to expressions of the form $\wedge\left(\frac{a}{b}, \frac{c}{d}\right)$? Using a similar argument to when we considered extending integer exponentiation. Suppose that proposition 1.6.17 holds for rational exponents. In particular we have for some $x \in \mathbb{Q}$ that

$$x^{\frac{1}{2}} * x^{\frac{1}{2}} = x^1$$

Now, suppose that $x = 2$. We are hence saying that

$$2^{\frac{1}{2}} * 2^{\frac{1}{2}} = 2$$

If we suppose that $2^{\frac{1}{2}} \in \mathbb{Q}$ with say $y = 2^{\frac{1}{2}}$ we are saying that $y^2 = 2$. Unfortunately, there is no such rational y that satisfies this. Moreover, we lack the theory required to prove this at this time. This will be corrected in part 2.

1.6.17 Extending the absolute value function

When we constructed the integers we recast the notion of size into that of distance. This was achieved using the so-called absolute value function given by

$$|x| = d(x, 0) = \begin{cases} x, & \text{If } x \geq 0 \\ -x, & \text{If } x < 0 \end{cases}$$

where

$$d : \mathbb{Z}^2 \rightarrow \mathbb{N}$$

$$(x, y) \mapsto d(x, y) = \begin{cases} x - y, & \text{If } x \geq y \\ -(x - y), & \text{If } x < y \end{cases}$$

Now that we have constructed the rational numbers we can consider how this idea extends. One thing that is clear from the definition of d for integers is that the smallest possible non-zero distance that can be achieved is 1, for example, $d(2, 1)$. However, consider

$$1 - \frac{1}{2} = \frac{1}{2}$$

If this idea of distance is to extend to the rationals we will clearly have that distances smaller than 1 are now possible. In other words, the mapping for d when used with rational numbers can no longer map into \mathbb{N} . This is easily remedied by defining the following set.

Definition 1.6.24. *Positive rationals*

We define the set of positive rationals by

$$\mathbb{Q}^+ = \{x \in \mathbb{Q} : x > 0\}$$

It is clear from the definitions for the integers how to extend the distance function and the absolute value function to the rationals.

Definition 1.6.25. *Distance function for the rationals*

Let $x, y \in \mathbb{Q}$. Define the function $d : \mathbb{Q}^2 \rightarrow \mathbb{Q}^+$ by

$$d : \mathbb{Q}^2 \rightarrow \mathbb{Q}^+$$

$$(x, y) \mapsto d(x, y) = \begin{cases} x - y, & \text{If } x \geq y \\ -(x - y), & \text{If } x < y \end{cases}$$

As before we prove that this distance function is well-defined.

Proposition 1.6.19. *The distance function for the rationals is well-defined*

Let $x, y \in \mathbb{Q}$. We have that

$$d(x, y) = \begin{cases} x - y, & \text{If } x \geq y \\ -(x - y), & \text{If } x < y \end{cases}$$

is well-defined.

Proof:

Let $x, y \in \mathbb{Q}$. There are two cases to consider $x \geq y$ and $x < y$.

1. $x \geq y$:

Suppose that $x \geq y$, then by proposition 1.6.11 part 14. we have

$$x \geq y \Rightarrow (x + (-y)) \geq (y + (-y)) \Rightarrow x - y \geq 0$$

Hence $x - y \in \mathbb{Q}^+$.

2. $x < y$:

As $x < y$ we have by definition of d that $d(x, y) = -(x - y)$ where we have that $x - y < 0$. However we have that $-(x - y) = -1 * (x - y)$ and so by part 16 of proposition 1.6.11 we have that $-1 * (x - y) > 0$ which is to say $-(x - y) \in \mathbb{Q}^+$

The result has been shown. \square

We can now generalise the absolute value function.

Definition 1.6.26. *Absolute value function*

Let $x \in \mathbb{Q}$ we define the absolute value function, denoted by $|x|$ by the function

$$|x| = d(x, 0) = \begin{cases} x, & \text{If } x \geq 0 \\ -x, & \text{If } x < 0 \end{cases}$$

We have generalised the idea of “size” to the rationals. We can now also generalise the properties of the absolute value function explored in the construction of the integers.

Proposition 1.6.20. *Properties of the absolute value*

Let $x, y, z \in \mathbb{Q}$. We have that the absolute value function has the following properties

1. $|x| \geq 0$ for all $x \in \mathbb{Q}$
2. $|x| = 0 \iff x = 0$
3. $|x - y| = 0 \iff x = y$
4. $|xy| = |x||y|$
5. $\left|\frac{x}{y}\right| = \frac{|x|}{|y|}$ with $y \neq 0$
6. $||x|| = |x|$
7. $|-x| = |x|$
8. $|x| \leq y \iff -y \leq x \leq y$
9. $|x| \geq y \iff x \leq -y \text{ or } x \geq y$
10. $|x + y| \leq |x| + |y|$
11. $|x - y| \leq |x - z| + |z - y|$
12. $|x - y| \geq ||x| - |y||$
13. $|\cdot|$ is not injective
14. $|\cdot|$ is not surjective

Proof:

1. $|x| \geq 0$ for all $x \in \mathbb{Q}$:

This follows by proposition 1.6.19.

2. $|x| = 0 \iff x = 0$:

We have by definition that $|x| = 0$, if and only if $x = 0$.

3. $|x - y| = 0 \iff x = y$:

(\Rightarrow): Suppose that $|x - y| = 0$. There are two cases to consider.

Firstly if $x \geq y$, then by definition we have that $|x - y| = x - y = 0$ from which we clearly have $x = y$. The other case is $x < y$ from which we get $|x - y| = -(x - y) = 0$. In other words, we have $-1 * (x - y) = 0$. Now by proposition 1.6.9 we know that for rationals a, b that if $ab = 0$, at least one of a or b is zero. As $-1 \neq 0$ we conclude that $x - y = 0$ from which we get $x = y$.

(\Leftarrow): Suppose that $x = y$ then $x - y = 0$ and so $|x - y| = 0$.

4. $|xy| = |x| |y|$:

Let $x, y \in \mathbb{Q}$. There are four cases to consider.

(a) $x \geq 0$ and $y \geq 0$

(b) $x \geq 0$ and $y < 0$

(c) $x < 0$ and $y \geq 0$

(d) $x < 0$ and $y < 0$

(a) $x \geq 0$ and $y \geq 0$:

If $x \geq 0$ and $y \geq 0$ then $xy \geq 0$ and so $|xy| = xy$. Likewise $|x| = x$ and $|y| = y$. Hence $|xy| = |x| |y|$.

(b) $x \geq 0$ and $y < 0$:

If $x \geq 0$ then $|x| = x$ by definition, and if $y < 0$ then $|y| = -y$. Now $|xy| = -xy$ as $y < 0$. Moreover, we have that

$$-xy = (-1)(x)(y) = (x)(-1)(y) = (x)(-y) = |x| |y|$$

Hence we get $|xy| = |x| |y|$

(c) $x < 0$ and $y \geq 0$:

This is similar to the above but swapping the roles of x and y .

(d) $x < 0$ and $y < 0$:

Suppose that $x < 0$ and $y < 0$, then we have that $|x| = -x$ and $|y| = -y$ by definition. Moreover, we have that $-x * -y = xy$. Hence $|xy| = xy = (-x)(-y) = |x| |y|$

5. $\left| \frac{x}{y} \right| = \frac{|x|}{|y|}$ with $y \neq 0$:

This follows by part 4.

6. $||x|| = |x|$:

We have that $|x| = x$ if $x \geq 0$ and $-x$ if $x < 0$.

So if $x \geq 0$, we have

$$||x|| = |x| = x = |x|$$

Now if $x < 0$ then

$$||x|| = |-x| = \underbrace{-x}_{\text{As } -x > 0} = |x|$$

7. $|-x| = |x|$:

As $-x = -1 * x$ we have by part 4 that

$$|-x| = |-1 * x| = |-1| |x| = 1 * |x| = |x|$$

8. $|x| \leq y \iff -y \leq x \leq y$:

(\Rightarrow): Suppose that $|x| \leq y$. If $x \geq 0$ then we get that $|x| = x \leq y$. From this, it is clear that $-y \leq x \leq y$ as $x \geq 0$ and $x \leq y \Rightarrow y \geq 0$.

Now if $x < 0$, then $|x| = -x \leq y$. Clearly $x \leq -x$ as $x < 0$ hence we conclude that $x \leq -x \leq y$. Now by part 18 of proposition 1.6.11 we have we have

$$(-1) * (-x) \geq (-1)(y) \iff x \geq -y$$

Now $x \geq -y$ is the same as $-y \leq x$ and so we have $-y \leq x \leq -x \leq y$.

Hence $-y \leq x \leq y$.

(\Leftarrow): Suppose that $-y \leq x \leq y$. There are two cases to consider.

(a) $x \geq 0$

(b) $x < 0$

(a) $x \geq 0$:

Suppose $x \geq 0$, then clearly as $x \leq y$ then $|x| \leq |y| = y$. Moreover, we have that $-y \leq x$ is the same $x \geq -y$ and by part 22. of proposition 1.5.11 when applied to $x \geq -y$ gives

$$(-1) * (x) \leq (-1)(-y) \iff -x \leq y$$

We have that $|-x| = |x|$ by part 6. Hence $|-x| = |x| \leq |y| = y$.

(b) $x < 0$:

Suppose $x < 0$. By assumption $x \leq y$ so either $y \geq 0$ or $y < 0$. We can't have $y < 0$ as for example take $x = -4$ and $y = -2$ then we would have $2 \leq -4 \leq -2$ a contradiction.

So suppose that $y \geq 0$ then as $x \leq y$ we have $|x| \leq |y| = y$. Now as $-y \leq x$ by assumption we have that $x \geq -y$ and so part 22. of proposition 1.6.11 gives

$$(-1) * (x) \leq (-1)(-y) \iff -x \leq y$$

Hence part 6. applies and we get that $|x| \leq y$

9. $|x| \geq y \iff x \leq -y$ or $x \geq y$:

(\Rightarrow): Suppose that $|x| \geq y$. If $x \geq 0$ then $|x| = x \geq y$. So suppose that $x < 0$ then by definition we have that $|x| = -x$ and so $-x \geq y$ and the result follows when applying part 22. of proposition 1.6.11.

(\Leftarrow): Suppose that either $x \leq -y$ or $x \geq y$. We have three cases to consider.

(a) $x \leq -y$

(b) $x \geq y$

(c) $x \leq -y$ and $x \geq y$

(a) $x \leq -y$:

Suppose that $x \leq -y$ holds. If $x \geq 0$ then we have that $-y \geq 0$, Hence $y < 0$. Moreover, we have that by part 18. of proposition 1.6.11 that

$$(-1) * (x) \geq (-1)(-y) \iff -x \geq y$$

Now part 6. applies and we see that $|-x| = |x| \geq |y| = y$. This is to say $|x| \geq y$.

Now suppose that $x < 0$. Then as $x \leq -y$ we have that either $-y \geq 0$ or $-y < 0$. In the former case $-y \geq 0$ gives $y < 0$. Hence by part 18. of proposition 1.6.11 we conclude that

$$(-1) * (x) \geq (-1)(y) \iff -x \geq y$$

As $x < 0$ then $-x \geq 0$. The result follows when taking the absolute value.

Now suppose that $-y < 0$ then $y \geq 0$. Following similar logic to the previous case, we see that

$$(-1) * (x) \geq (-1) (y) \iff -x \geq y$$

The result again follows after taking the absolute value.

(b) $x \geq y$:

This case is trivial.

(c) $x \leq -y$ and $x \geq y$:

Suppose that $x \leq -y$ and $x \geq y$ are both true. We know by the first case that $x \leq -y$ gives $|x| \geq y$ and $x \leq y$ also implies $|x| \geq y$ by the second case. Hence both inequalities being true at the same time implies the result $|x| \geq y$.

10. $|x + y| \leq |x| + |y|$:

Let $x, y \in \mathbb{Q}$. There are four cases to consider.

(a) $x \geq 0$ and $y \geq 0$

(b) $x \geq 0$ and $y \leq 0$

(c) $x \leq 0$ and $y \geq 0$

(d) $x \leq 0$ and $y \leq 0$

(a) $x \geq 0$ and $y \geq 0$:

Suppose $x \geq 0$ and $y \geq 0$, then we have that

$$|x + y| = x + y = |x| + |y| \Rightarrow |x + y| \leq |x| + |y|$$

(b) $x \geq 0$ and $y \leq 0$

By assumption we have that $|x| = x$ and $|y| = -y$. We have two cases based on the absolute value, $|x| \leq |y|$ and $|x| \geq |y|$.

So suppose that $|x| \leq |y|$ then by definition $x \leq -y$ and so by part 12. of proposition 1.6.11 we have that

$$x \leq -y \Rightarrow x + y \leq 0$$

Moreover, as $x \geq 0$ then $y \leq x + y \leq 0$. Hence we have by the definition of the absolute value that

$$|x + y| = -(x + y) \leq -y = |y|$$

As $-y > 0$.

In the case $|x| \geq |y|$ we have by definition that $x \geq -y$ and so $x + y \geq 0$. Additionally it is clear that $x \geq x + y$ as $y \leq 0$ and $|x| \geq |y|$. Hence by definition of the absolute value we have that

$$|x + y| = x + y \leq x = |x|$$

Now, it is clear to see that $|x| \leq |x| + |y|$ and likewise $|y| \leq |x| + |y|$.

We have hence shown that $|x + y| \leq |x| + |y|$.

(c) $x \leq 0$ and $y \geq 0$:

This is similar to above, interchanging the roles of x and y .

(d) $x \leq 0$ and $y \leq 0$:

Suppose that $x \leq 0$ and $y \leq 0$ then by definition we have that $|x + y| = -(x + y) = -x - y$. As $x \leq 0$ and $y \leq 0$ then we have that $|y| = -y$ which shows $|x + y| = |x| + |y| \leq |x| + |y|$

11. $|x - y| \leq |x - z| + |z - y|:$

We have that

$$\begin{aligned} |x - y| &= |x - (z - z) - y| \\ &= |x - z + z - y| \\ &\leq |x - z| + |z - y| \end{aligned}$$

12. $|x - y| \geq ||x| - |y||:$

We have that

$$\begin{aligned} |x| &= |(x - y) + y| \leq |x - y| + |y| \Rightarrow |x| - |y| \leq |x - y| \\ |y| &= |(y - x) + x| \leq |x - y| + |x| \Rightarrow |y| - |x| \leq |x - y| \end{aligned}$$

Hence we have

$$\begin{aligned} |x| - |y| &\leq |x - y| \Rightarrow ||x| - |y|| \leq |x - y| \\ |y| - |x| &= (-1)(|x| - |y|) \leq |x - y| \Rightarrow ||x| - |y|| \leq |x - y| \end{aligned}$$

Hence we have the result.

13. $|\cdot|$ is not injective:

This follows as the absolute value function was not injective for the integers

14. $|\cdot|$ is not surjective:

This follows as the absolute value function was not surjective for the integers

As required. \square

Part 2

Elementary Number Theory

2.1 Introduction

Mathematics is the queen of the sciences and
Number Theory is the queen of mathematics.

Carl Friedrich Gauss

In the previous part, we have gone from only having the axioms of ZFC, the rules of logic and knowledge of mappings and have built two types of numbers, the naturals and the integers. Unfortunately, we need to make a detour from constructing new objects. We need to start using the objects we have constructed to provide a guide on how to proceed with building more mathematical objects.

We will start with Number Theory. Number Theory primarily deals with the properties of the integers \mathbb{Z} as well as mappings defined on \mathbb{Z} . This includes properties about the operations on the integers, properties about the compositions and ways of expressing relationships between certain “types” of integers, solving equations involving the integers and more.

The applications of Number Theory to the modern world are numerous. One main example of the usage of Number Theory is encryption, the art of obfuscating information so that it can only be read by trusted individuals¹¹. We will later consider an example of encryption called RSA.

Additionally, the ideas that we will develop when studying Number Theory are key to providing crucial insights into other branches of mathematics. We will come to see that many of the key properties of the integers are also enjoyed by many other types of mathematical objects, especially in an abstract setting.

2.2 Divisibility

Now where there are no parts, neither
extension, shape, nor divisibility is possible.
And these monads are the true atoms of
nature and, in a word, the elements of things.

Gottfried Leibniz

2.2.1 Definition of divisibility of integers

Although we have a concrete construction of the integers, we haven’t even discussed some of their most basic properties! We know how to add, subtract and multiply them, but we don’t know how to divide them without the rational numbers \mathbb{Q} . It is with \mathbb{Q} that we can hope to find a rule that says that $\frac{a}{b} \in \mathbb{Z}$ for some $a, b \in \mathbb{Z}$.

Recall that in \mathbb{Q} we defined an equivalence relation \sim so that for $(a, b), (c, d) \in \mathbb{Z}^2$ we have that

$$(a, b) \sim (c, d) \iff ad = bc$$

where we had $b \neq 0$ and $d \neq 0$. We also saw that $(x, 1) \in [(x, 1)]$ represented an integer. Hence the question we are resolving is when does $(a, b) \sim (x, 1)$. We have that

$$(a, b) \sim (x, 1) \iff a = bx$$

That is b divides a and gives an integer if and only if $a = bx$. We make this our first formal definition in the field of Number Theory.

¹¹Until someone manages to find a way to get past the elegant mathematics of the encryption scheme!

Definition 2.2.1. *Integer divisibility*

Let $a, b \in \mathbb{Z}$ with $b \neq 0$. We say that a is divisible by b , or b divides a , written as $b \mid a$ if and only if $\exists c \in \mathbb{Z}$ so that $a = bc$. We say that b is a divisor of a .

If b does not divide a we write $b \nmid a$.

Example 2.2.1.

We have that $3 \mid 6$ as $6 = 3 * 2$.

Obverse that $2 \nmid 3$. Indeed there is no integer x so $3 = 2x$.

We make a definition based on the definition of divisibility. Namely based on if a number can be divided into two equal parts.

Definition 2.2.2. *Even number*

Let $x \in \mathbb{Z}$. We say that x is even if we have that $2 \mid x$.

This immediately gives another definition.

Definition 2.2.3. *Odd number*

Let $x \in \mathbb{Z}$. We say that x is odd if we have that $2 \nmid x$.

We can make another definition, based on divisibility.

Definition 2.2.4. *Integer multiple*

Let $a, b \in \mathbb{Z}$ so that $b \mid a$. We say that b is a multiple of a .

There are two results that we can derive based on an even number, an odd number and integer multiples.

Proposition 2.2.1. *Integer is even if it is a multiple of 2*

Let $x \in \mathbb{Z}$. We have that x is even if and only if x is a multiple of 2.

Proof:

(\Rightarrow) : Suppose that x is even, then by definition we have that $2 \mid x$ and so by the definition of divisibility we have that $x = 2c$ for some $c \in \mathbb{X}$. By the definition of being an integer multiple we have that x is a multiple of 2.

(\Leftarrow) : Suppose that x is a multiple of 2. By definition of being an integer multiple, we have that $x = 2r$ for some $r \in \mathbb{Z}$. Hence by the definition of divisibility, we have that $2 \mid x$ and so by definition of an even number we have that x is even. \square

We can find a similar proposition for odd numbers. Observe that by the previous proposition that x being even means that $x = 2n$ for some integer n . Also, we have that $2n + 2 = 2(n + 1)$ is even, so what can we say about $2n + 1$?

Proposition 2.2.2. *Integer is odd if and only if it is not a multiple of 2*

Let $x \in \mathbb{Z}$. We have that x is odd if and only if x is not a multiple of 2.

Proof:

The proof follows by the contra-positive, that is x is a multiple of 2 if and only if x is even, which is the previous proposition. \square

Hence we need to determine if $2n + 1$ is even or odd. We need to develop the theory of divisibility.

The definition of divisibility gives an immediate result. Namely that when considering the divisibility of integers we need only concern ourselves with positive integers, as negative integers will also be divisors. That is if $b \mid a$ then so does $-b$.

Proposition 2.2.3. *Integer dividing another implies negative integer also divides*

Let $a, b \in \mathbb{Z}$ with $b \mid a$. We also have that $-b \mid a$.

Proof:

Let $a, b \in \mathbb{Z}$ with $b \mid a$. By definition of divisibility, we have that $\exists c \in \mathbb{Z}$ so that $a = bc$. We know that $-1 * 1 = 1$ and so we have that

$$a = bc = (-1 * -1)bc = -b * -c$$

As $-c \in \mathbb{Z}$ then it follows by definition that $-b \mid a$. \square

Hence by proposition 2.2.3 we will restrict our view to positive divisors only, knowing that any results about a positive divisor will extend to negative divisors.

One clear divisor of any integer a is itself, that is $a \mid a$ as $a = a * 1$. We will find it interesting to consider the more non-trivial divisors of some integers. Hence we make the following definition

Definition 2.2.5. *Proper divisor*

Let $a, b \in \mathbb{Z}$ with $b \mid a$. If we have that $0 < b < a$ then we say that b is a proper divisor of a .

There are some clear results about divisibility.

Proposition 2.2.4. *Properties of divisibility*

Let $a, b, c \in \mathbb{Z}$. We have the following properties for divisibility

1. $a \mid b \Rightarrow a \mid bc$ for any $c \in \mathbb{Z}$
2. $a \mid b$ and $b \mid c$ implies that $a \mid c$
3. $a \mid b$ and $a \mid c$ implies that $a \mid (bx + cy)$ for any $x, y \in \mathbb{Z}$
4. $a \mid b$ and $b \mid a$ implies $a = \pm b$, that is either $a = b$ or $a = -b$.
5. $a \mid b$ and $a > 0$ and $b > 0$ implies that $a \leq b$.
6. If $m \in \mathbb{Z}$ is such that $m \neq 0$ then $a \mid b$ is true if and only if $ma \mid mb$.
7. For all $a \in \mathbb{Z}$ with $a \neq 0$ we have $a \mid 0$

Proof:

1. $a \mid b \Rightarrow a \mid bc$ for any $c \in \mathbb{Z}$:
2. $a \mid b$ and $b \mid c$ implies that $a \mid c$:

Suppose that $a \mid b$, then by definition there exists $d \in \mathbb{Z}$ so that $b = ad$. Hence we have that

$$bc = adc \Rightarrow a \mid bc$$

as $dc \in \mathbb{Z}$.

3. $a \mid b$ and $a \mid c$ implies that $a \mid (bx + cy)$ for any $x, y \in \mathbb{Z}$:

Suppose that $a \mid b$ and $a \mid c$, then by the definition of divisibility, and by part 1., we have that $b = ax$ and $c = ay$ for all $x, y \in \mathbb{Z}$. We hence see that

$$c = a y$$

Hence as $xy \in \mathbb{Z}$ then we conclude that $a \mid c$.

4. $a \mid b$ and $b \mid a$ implies $a = \pm b$, that is either $a = b$ or $a = -b$:

Let $a \mid b$ and $a \mid c$, then there are $d, e \in \mathbb{Z}$ such that $b = ad$ and $c = ae$. Now, let $x, y \in \mathbb{Z}$ then we have that $bx = adx$ and $cy = aey$ and $bx + cy = adx + aey = a(dx + ey)$. Hence $a \mid (bx + cy)$.

5. $a \mid b$ and $a > 0$ and $b > 0$ implies that $a \leq b$:

If $a \mid b$ then $\exists x \in \mathbb{Z}$ so that $b = ax$, likewise if $b \mid a$ then $\exists y \in \mathbb{Z}$ so that $a = by$. It follows that $b = byx$. We have that $b = byx$ is true if and only if $yx = 1$. Therefore either $x = y = 1$ or $x = y = -1$.

The result is clear after substituting y into $a = by$.

6. If $m \in \mathbb{Z}$ is such that $m \neq 0$ then $a \mid b$ is true if and only if $ma \mid mb$:

(\Rightarrow): Let $m \in \mathbb{Z}$ be non-zero and let $a \mid b$. By definition, there is some $c \in \mathbb{Z}$ so that $b = ac$. Multiplying both sides by m gives

$$bm = acm = amc$$

and so $am \mid bm$.

(\Leftarrow): Suppose that $am \mid bm$, then again by the definition of divisibility we have that there is some $c \in \mathbb{Z}$ so that $bm = amc$. By the cancellation law, we can cancel the m to get $b = ac$ and the result follows.

7. For all $a \in \mathbb{Z}$ with $a \neq 0$ we have $a \mid 0$:

Let $a \in \mathbb{Z}$, where $a \neq 0$. We have that $0 = ka$ has the solution $k = 0$ by part I proposition 1.5.9. Hence $a \mid 0$.

As required. \square

Part 3. of the previous proposition can be generalised. We will work with an example to see how this can be achieved.

Example 2.2.2. Let $a = 2$, $b = 16$ and $c = 32$. Clearly we have that $a \mid b$ as $16 = 4 * 2$ and likewise $a \mid c$ as $32 = 5 * 2$.

Now part 3. states that if $a \mid b$ and $a \mid c$ then we must have that $a \mid (bx + cy)$ for any $x, y \in \mathbb{Z}$.

Indeed, for example, we can see that $2 \mid (-5(16) + 7(32))$. As $-5(16) + 7(32) = -80 + 224 = 144$. Now suppose that $d = 64$ and say $z = 5$. We can see that

$$-5(16) + 7(32) + 5(64) = 144 + 320 = 464 \Rightarrow 2 \mid (-5(16) + 7(32) + 5(64))$$

We prove the general statement now.

Proposition 2.2.5. Divisor that divides a set of integers divides a combination of the set

Let $a \in \mathbb{Z}$ and let $S = \{b_1, b_2, b_3, \dots, b_n\}$ be a set of n integers where $b_i \in \mathbb{Z}$ for each b_i . Moreover suppose that $a \mid b_i$ for each $b_i \in S$. We have that

$$a \mid \sum_{i=1}^n b_i x_i$$

for any $x_i \in \mathbb{Z}$.

Proof:

We argue by induction on n . The base case is $n = 2$ which is shown in proposition 2.2.4. So suppose that the result holds for some $k \geq 1$, which is to say that if $S = \{b_1, b_2, \dots, b_k\}$ and we have that $a \mid b_i$ for each $b_i \in S$ then

$$a \mid \sum_{i=1}^k b_i x_i$$

We need to show that the result holds for $k + 1$. That is if $\tilde{S} = S \cup \{b_{k+1}\}$ so that $a \mid b_i$ for each $b_i \in \tilde{S}$ then

$$a \mid \sum_{i=1}^{k+1} b_i x_i$$

So take $\tilde{S} = S \cup \{b_{k+1}\}$ so that $a \mid b_i$ for each $b_i \in \tilde{S}$. By applying part 1. of proposition 2.2.4 to each $a \mid b_i$ we know that for all $x_i \in \mathbb{Z}$ that $a \mid b_i x_i$.

Now, by the induction hypothesis we know that $\forall b_i \in S$ that $a \mid b_i$ and moreover we have that

$$a \mid \sum_{i=1}^k b_i x_i$$

Let $d = \sum_{i=1}^k b_i x_i$. Again by part 1 of proposition 2.2.4 we have that $a \mid ad$. Additionally we know that $a \mid b_{k+1}$ and so by part 3. of 2.2.4, As $d \in \mathbb{Z}$, we have that

$$\begin{aligned} a &\mid (1 * d + b_{k+1} x_{k+1}) \\ a &\mid \left(\sum_{i=1}^k b_i x_i + b_{k+1} x_{k+1} \right) \\ a &\mid \left(\sum_{i=1}^{k+1} b_i x_i \right) \end{aligned}$$

Which implies the result holds for $k+1$ and hence for any $n \in \mathbb{N}$ by induction. \square

2.2.2 The greatest common divisor and the least common multiple

Now that we have a solid grasp of the basics of integer divisibility, we can start looking towards some applications. One immediate question is given a set of integers say

$$S = \{a_1, a_2, a_3, \dots, a_n\}$$

What is the largest integer which divides each $a_i \in S$. and what is the largest integer m so that m has each $a_i \in S$ as a proper divisor? An immediate use of these two ideas is very useful when doing arithmetic with rational numbers. For example, consider trying to simplify the fraction $\frac{525}{2925}$. To simplify this we need to find the integers that multiply to make 525 and those that multiply to make 2925. If there are any in common then we know from the construction of the rationals that $\frac{x}{x} = 1$ and in particular we have that $\frac{xy}{xz} = \frac{y}{z} * \frac{x}{x} = 1$.

Likewise suppose we wanted to add $\frac{1}{4}$ and $\frac{1}{7}$. It is true that by definition of addition, we would have

$$\frac{1}{4} + \frac{1}{7} = \frac{1 * 7 + 1 * 4}{7 * 4} = \frac{7 + 4}{7 * 4} = \frac{11}{28}$$

The key stage was $\frac{1 * 7 + 1 * 4}{7 * 4}$, breaking this down we see that

$$\frac{1 * 7 + 1 * 4}{7 * 4} = \frac{1 * 7}{7 * 4} + \frac{1 * 4}{7 * 4}$$

In other words, we are finding a multiple in common with 7 and 4 to turn the denominator into. It is therefore worthwhile to work out the theory of working out common divisors and common multiples.

We will start by working out common divisors, by first making a definition.

Definition 2.2.6. *Common divisor*

Let $a, b, c \in \mathbb{Z}$ be non-zero integers. We say that c is a common divisor of a and b if $c \mid a$ and $c \mid b$.

Example 2.2.3. Consider the integers 35 and 25. The divisors of 35 are 1, 5 and 7 and 35, likewise the divisors of 25 are 1 and 5 and 25. The largest common divisor is therefore 5.

Example 2.2.4. Consider the integers 24 and 54. Doing the same as before, we can see that the divisors of 24 are 1, 2, 3, 4, 6, 8, 12 and 24. Looking at the divisors of 54 we see that they are 1, 2, 3, 6, 9, 18, 27 and 54.

The common divisors of 24 and 54 are therefore 1, 2, 3 and 6,

Example 2.2.5. Consider the common divisors of 3 and 5. The divisors of 3 are simply 1 and 3, likewise the divisors of 5 are 1 and 5. The only common divisor is 1.

We can see from the previous examples that there was a largest, or greatest common divisor between the pairs of integers in each case. We can show that for any two integers, there is always a greatest common divisor.

Theorem 2.2.1. The greatest common divisor of two integers exists

Let $a, b \in \mathbb{Z}$ so that $a \neq 0$ or $b \neq 0$. Then there exists $d \in \mathbb{Z}$ so that d is the largest possible common divisor, that is there is no $g \in \mathbb{Z}$ with $g > d$ so that $g \mid a$ and $g \mid b$.

Proof:

Firstly, we note that as $1 \mid a$ and $1 \mid b$, the largest possible common divisor is at least 1, proving existence. To show that there is the largest possible common divisor we must show that this divisor can't exceed some integer, say M , where M depends on a and b . Moreover by proposition 2.2.3 we only need to consider the case where $a \geq 0$ and $b \geq 0$.

So, suppose that $c \mid a$ and $c \mid b$ for some $c \geq 1$. By part 5. of proposition 2.2.4 we have that as $c \mid a$ then $c \leq a$, likewise as $c \mid b$ then $c \leq b$. There are three possibilities to consider

1. $a = b$
2. Without any loss of generality we have $a < b$
3. One of $a = 0$ or $b = 0$ but not both at the same time.

1. $a = b$:

In this case we easily take M to be the largest divisor of a , or equivalently b , then $c \leq M$

2. Without any loss of generality we have $a < b$:

Without loss of generality, we take $a < b$, if this is not the case we simply swap the roles of a and b . In this case, we take M to be the largest divisor so that $M \leq a$. For if we took a M so that $M \leq b$ then by the fact $a < b$ we could have the case that $M > a$ a contradiction to the fact that $c \leq a$ as $c \mid a$.

3. One of $a = 0$ or $b = 0$ but not both at the same time:

Suppose that $a = 0$ and $b \neq 0$, then we have that for all $M \in \mathbb{Z}$ that $M \mid a$, but as $c \mid b$ then $c \leq b$ and so we take $M = b$ as $b \mid b$. Likewise if we assume $b = 0$ and $a \neq 0$.

In each case we found a M so that if we take $c \leq M$ then $c \mid a$ and $c \mid b$.

We have shown that for any two integers a greatest common divisor always exists. We can make a formal definition.

Definition 2.2.7. Greatest common divisor

Let $a, b \in \mathbb{Z}$ so that $a \neq 0$ and $b \neq 0$. Let $d \in \mathbb{Z}$ be such that $d \mid a$ and $d \mid b$. We say the largest value of d where $d \mid a$ and $d \mid b$ is the greatest common divisor of a and b , denoted $d = \text{GCD}(a, b)$, sometimes written $\text{gcd}(a, b)$ and in some texts simply by (a, b) .

As $a \mid 0$ for any integer a . We define $\text{GCD}(a, 0) = a$, similarly $\text{GCD}(0, b) = b$.

We will use the notation GCD in this text and we will usually abbreviate saying the greatest common divisor to GCD. Although we have proved that the greatest common divisor exists, we do not yet actually have a method of calculating what it is other than trying through trial and error. To see how we can attempt to construct a method of finding GCD we should look to cases where integer division does not fail and to cases where it does fail.

Example 2.2.6. It is clear that $2 \nmid 3$ as there is no integer x so that $3 = 2x$. If we take $x = 1$ we get the false equality of $3 = 2$, if we take $x = 2$ we get another false equality of $3 = 4$. We observe however that $3 = 2 * 1 + 1$.

Example 2.2.7. Let $a = 25$ and $b = 7$. It is clear that $7 \nmid 25$. The first couple multiples of 7 are $7 = 7 * 1$, $14 = 7 * 2$, $21 = 7 * 3$, $28 = 7 * 4$ and so on. However, we can see that $25 = 7 * 3 + 4$.

Example 2.2.8. Let $a = 36$ and $b = 12$. Clearly that $12 \mid 36$ as $36 = 12 * 3$. The first couple multiples of 12 are $12 = 12 * 1$, $24 = 12 * 2$, $36 = 12 * 3$, $48 = 12 * 4$ and so on.

Example 2.2.9. This time, let $a = 8$ and $b = 2$. Then we have that $2 \mid 8$ as $8 = 2 * 4$. In a similar way to the previous examples we see that $8 = 2 * 4 + 0$

If we let $a, b \in \mathbb{Z}$ so that $b \nmid a$ then, in the previous examples it seems that we can always find a multiple of b so that $bx \leq a$ for some $x \in \mathbb{Z}$ and in particular we have that

$$a = bx + (a - bx)$$

In the case that $b \mid a$ then $a - bx = 0$. Interpreting what $a - bx$ means, when $b \nmid a$ then $a - bx \neq 0$ and when $b \mid a$ we had that $a - bx = 0$. Hence $a - bx \neq 0$ is a measure of how far off we are from having $b \mid a$. This is to say that if $a - bx > 0$ then we are a little short of making a multiple of a from b and if $a - bx < 0$ we are a little over of making a multiple of a from b .

In general, we can see that any integer division can be viewed in this way, that is if $a, b \in \mathbb{Z}$ we can see the result of a divided by b in the form $a = qb + r$ for some $q, r \in \mathbb{Z}$.

Theorem 2.2.2. The division algorithm

Let $a, b \in \mathbb{Z}$ so that $b > 0$, then there exist $q, r \in \mathbb{Z}$ with q, r being unique so that

$$a = bq + r$$

where $0 \leq r < b$

Proof:

There are three cases to consider

1. $a = b$

2. $a < b$

3. $a > b$

1. $a = b$:

If $a = b$ then $b \mid a$ holds trivially and we see that $a = 1 * b + 0$ where $q = 1$ and $r = 0$.

2. $a < b$:

If $a < b$ then we also see that trivially we have that $a = 0 * b + a$ where $q = 0$ and $r = a$.

3. $a > b$:

This case is the meat of the theorem. To prove the division theorem we will argue by induction on a . The base case is $a = 1$ where we either have $a = b$ or $a < b$ which have been dealt with. So now suppose that the result holds for some $k > 1$. Likewise in the base case, we only need to consider the case of $k + 1 > b$, or equivalently $b < k + 1$

As $b < k + 1$ we have that $1 \leq (k + 1) - b$ and so by the induction hypothesis we have that there are integers $q, r \in \mathbb{Z}$ so that

$$(k + 1) - b = bq + r$$

where $0 \leq r < b$. From this, we clearly get $k + 1 = bq' + r$ where $q' = 1 + q$ which shows the induction step. The result now follows by induction.

Now that the existence has been shown, it is left to show the uniqueness of q and r . So suppose that q_1, r_1 and q_2, r_2 are two such pairs that satisfy the conditions of the theorem. Firstly suppose that $r_1 \neq r_2$ then we have that, without loss of generality that $r_1 < r_2$ so that $0 < r_2 - r_1 < b$ and then by the theorem we have that

$$r_2 - r_1 = b(q_2 - q_1)$$

which implies that $b \mid (r_2 - r_1)$. This is a contradiction to theorem 2.2.4 part 5. as this part implies that $b \leq r_2 - r_1$. Therefore $r_1 = r_2$ and from $r_1 = r_2$ we have that $0 = b(q_2 - q_1)$ and by part 1 proposition 1.5.9 as $b > 0$ then $q_2 - q_1 = 0$ giving $q_2 = q_1$. \square

Based on this theorem we make a definition.

Definition 2.2.8. *Quotient and remainder*

Let $a, b \in \mathbb{Z}$ so that $b > 0$. We have by the division algorithm that

$$a = qb + r$$

where $q, r \in \mathbb{Z}$ and $0 \leq r < b$. We say that q is the quotient of the division and that r is the remainder.

In the theorem, we assumed that $b > 0$. However by proposition 2.2.3 we know that negative divisors are also valid. To resolve this we reformulate theorem 2.2.2 so that $0 \leq r < |b|$.

Theorem 2.2.3. *The division algorithm (Extended)*

Let $a, b \in \mathbb{Z}$ so that $b \neq 0$, then there exist $q, r \in \mathbb{Z}$ with q, r being unique so that

$$a = bq + r$$

where $0 \leq r < |b|$

Proof:

By the division algorithm, theorem 2.2.2 we have for $|a|$ and $|b|$ that there exist unique $q, r \in \mathbb{Z}$ so that

$$|a| = q|b| + r$$

where $0 \leq r < |b|$. There are a few cases to consider.

1. $r = 0$
2. $r > 0$ and $a \geq 0$
3. $r > 0$ and $a < 0$

1. $r = 0$:

If $r = 0$, then $|a| = q|b|$ and so by the properties of the absolute value we have that $a = \pm qb$, hence $a = b(\pm q)$ and we have the result.

2. $r > 0$ and $a \geq 0$:

Now suppose $r > 0$ and $a \geq 0$. We hence have that $a = q|b| + r$ which gives

$$\begin{aligned} a &= bq + r, \text{ If } b > 0 \\ a &= (-b)q + r, \text{ If } b < 0 \end{aligned}$$

The first is simply the first version of the division algorithm and the second can be written as $a = b(-q) + r$ which gives the result.

3. $r > 0$ and $a < 0$:

Finally if $r > 0$ and $a < 0$ then we have

$$-a = |b|q + r \Rightarrow a = -|b|q - r$$

This is a problem as it would give a negative remainder. We can employ a trick that doesn't change the value of a but allows us to express $a = -|b|q - r$ in a more suitable form.

$$\begin{aligned} a &= -|b|q - r \\ a &= -|b|q + (|b| - |b|) - r \\ a &= -|b|q + |b| + (|b|r) \\ a &= |b|(-1 - q) + (|b|r) \end{aligned}$$

By assumption we have that $0 < r < |b|$ implies that $0 < |b| - r < |b|$, so we re-write the above as

$$a = bq' + r'$$

where $r' = |b| - r$ and $q' = -1 - q$, if $b > 0$ and for $b < 0$ we write $q' = 1 + q$.

This completes the proof. \square

We can now go back to a problem from the first section, namely showing that $2n + 1$ must be odd

Proposition 2.2.6. *Integer is odd if and only if it is a multiple of $2n + 1$*

Let $x \in \mathbb{Z}$. We have that x is odd if and only if it is a multiple of $2n + 1$ where $x = 2n + 1$ for $n \in \mathbb{Z}$. Then n is odd.

Proof:

Suppose $x \in \mathbb{Z}$, then by the division algorithm we have that

$$x = 2q + r$$

where $0 \leq r < |2|$. Hence the only remainders possible are $r = 0$ or $r = 1$. Hence either $x = 2q$ or $x = 2q + 1$. In the first case we have $x = 2q$ is even by definition. In the case $x = 2q + 1$ we have that $2 \nmid 2n + 1$ and so x can't be even by definition. It follows that x is odd. \square

With this proposition and proposition 2.2.1 we can derive the evenness or oddness when adding or multiplying even or odd integers.

Proposition 2.2.7. *Even and oddness for addition and multiplication*

Let $x, y \in \mathbb{Z}$. We have that

1. If x is even and y is even then $x + y$ is even and xy is even.
2. If x is even and y is odd then $x + y$ is odd and xy is even.
3. If x is odd and y is even then $x + y$ is odd and xy is even.
4. If x is odd and y is odd then $x + y$ is even and xy is odd.

Proof:

1. If x is even and y is even then $x + y$ is even and xy is even:

Suppose that x and y are even, then by proposition 2.2.1 we have $x = 2n$ for some $n \in \mathbb{Z}$ and $y = 2m$ for some $m \in \mathbb{Z}$. We have that $x + y = 2n + 2m = 2(n + m)$ hence $x + y$ is even by proposition 2.2.1. Likewise, we have that $xy = 2n * 2m = 2(n * m)$ and therefore even.

2. If x is even and y is odd then $x + y$ is odd and xy is odd:

Suppose that x is even and y is odd. By we have that $x = 2n$ for some $n \in \mathbb{Z}$ by 2.2.1 and by proposition 2.2.6 we have that $y = 2m + 1$ for some $m \in \mathbb{Z}$.

We have $x + y = 2n + 2m + 1 = 2(n + m) + 1$ and so $x + y$ is odd by proposition 2.2.6. Additionally, $xy = 2n(2m + 1) = 2(2mn + n)$ and so by proposition 2.2.1 we have that xy is even.

3. If x is odd and y is even then $x + y$ is odd and xy is even:

Similar to above, swapping the roles of x and y .

4. If x is odd and y is odd then $x + y$ is even and xy is odd:

By proposition 2.2.6 we have that $x = 2n + 1$ for some $n \in \mathbb{Z}$ and $y = 2m + 1$ for some $m \in \mathbb{Z}$.

Now, $x + y = (2n + 1) + (2m + 1) = 2(n + m) + 2 = 2((n + m) + 1)$. So by proposition 2.2.1 we have $x + y$ is even.

Finally, $xy = (2n + 1)(2m + 1) = 4nm + 2n + 2m + 1 = 2(2nm + (n + m)) + 1$ and so by proposition 2.2.6 is odd.

As required. \square

Continuing with our quest to find a method to compute the greatest common divisor. At first, it might seem that we haven't made much progress in finding a way to calculate the GCD. However, consider the following examples.

Example 2.2.10. Consider $a = 56$ and $b = 24$. By the division algorithm, we have that $56 = 2 * 24 + 8$. Now what about $a = 24$ and $b = 8$? Again, by the division algorithm, we have that $24 = 3 * 8 + 0$.

Now, the divisors of 56 are 1, 2, 4, 7, 8, 14, 28 and 56, the divisors of 24 are 1, 2, 3, 4, 6, 8, 12 and 24. The largest common divisor was 8, which was the remainder after the first use of the division algorithm. Likewise, it was the quotient in the second application of the division algorithm.

Example 2.2.11. Consider $a = 4947$ and $b = 1552$. By the division algorithm, we have that $4947 = 3 * 1552 + 291$. Applying the division algorithm to $a = 1552$ and $b = 291$ gives $1552 = 5 * 291 + 97$. A third application of the division algorithm to $a = 291$ and $b = 97$ gives $291 = 3 * 97 + 0$.

Unlike with the previous example, there may be potentially too many divisors for 4947 to list them out by trying each integer $0 < x \leq 4947$. The same is true for 1552. However, if we follow the same logic as the previous example we might suspect that 97 is the greatest common divisor, as by the division algorithm for $a = 4947$ and $b = 97$ we get $4947 = 51 * 97 + 0$. Applying the division algorithm to $a = 1552$ and $b = 97$ gives $1552 = 16 * 97 + 0$.

Based on these two examples we might be tempted to make a conjecture on how we can potentially calculate the GCD. A further example is needed.

Example 2.2.12. Let $a = 574$ and $b = 34$. By the division algorithm, we have that $574 = 16 * 34 + 30$. Applying the algorithm again to $a = 34$ and $b = 30$ gives $34 = 1 * 30 + 4$. Another application gives $30 = 7 * 4 + 2$ and finally a last application gives $4 = 2 * 2$.

Now, applying the division algorithm to 574 and 2 gives $574 = 287 * 2 + 0$ and applying it to 34 and 2 gives $34 = 17 * 2 + 0$. So we suspect that $\text{GCD}(574, 34) = 2$.

If what we suspect is true, then repeated applications of the division algorithm might provide a way to compute the greatest common divisor of any two integers. We can provide more evidence that this must be the case by considering the examples in reverse.

Example 2.2.13. Consider $a = 56$ and $b = 24$. We saw that applying the division algorithm twice gave us that

$$56 = 2 * 24 + 8$$

$$24 = 3 * 8$$

By substituting $24 = 3 * 8$ into $56 = 2 * 24 + 8$ we get

$$\begin{aligned}56 &= 2 * 24 + 8 \\56 &= 2 * (3 * 8) + 8 \\56 &= 6 * 8 + 8 \\56 &= 7 * 8\end{aligned}$$

And hence by the definition of divisibility $8 \mid 56$, likewise by $24 = 3 * 8$ we have that $8 \mid 24$.

Now, suppose that d is a common divisor of 56 and 24. We have that as $d \mid 56$ and $d \mid 24$, in particular we must have that $d \mid (2 * 24 + 8)$ as $d \mid 56$. Hence $d \mid 8$ as $d \mid 24$, and clearly the largest such $d \mid 8$ is 8 itself.

Example 2.2.14. In the example where $a = 4947$ and $b = 1552$. We saw that applying the division algorithm three times gave us

$$\begin{aligned}4947 &= 3 * 1552 + 291 \\1552 &= 5 * 291 + 97 \\291 &= 3 * 97 + 0\end{aligned}$$

By substituting $291 = 3 * 97$ into $1552 = 5 * 291 + 97$ we get

$$\begin{aligned}1552 &= 5 * 291 + 97 \\1552 &= 5 (3 * 97) + 97 \\1552 &= 15 * 97 + 97 \\1552 &= 16 * 97\end{aligned}$$

Which gives us that $97 \mid 1552$. Now substituting $1552 = 16 * 97$ and $291 = 3 * 97$ into $4947 = 3 * 1552 + 291$ yields.

$$\begin{aligned}4947 &= 3 * 1552 + 291 \\4947 &= 3 * (16 * 97) + 3 * 97 \\4947 &= 48 * 97 + 3 * 97 \\4947 &= 51 * 97\end{aligned}$$

Showing $97 \mid 4947$. Now as in the previous example, suppose that d is a common divisor of 4947 and 1552. As $d \mid 4947$ and $d \mid 1552$ then $d \mid (3 * 1552 + 291)$ which gives $d \mid 291$. Applying similar logic, we see that as $d \mid 1552$ and $d \mid 97$ then $d \mid (5 * 291 + 97)$ and so $d \mid 97$. The largest such d satisfying this is $d = 97$.

It is therefore clear that for integers a and b repeated applications of the division algorithm on b and the remainder r give a candidate for the greatest common divisor. When this candidate is used candidate through the equations generated by each use of the division algorithm proves that it is the largest such common divisor of a and b . Hence, informally, we have found the method for computing the GCD! It is left to formalise this discovery.

From working the examples in reverse we have an important proposition that will be crucial for proving the result. Namely that the greatest common divisor of a and b is also equal to the greatest common divisor of b and r where r is the remainder from the division algorithm.

Proposition 2.2.8. $\text{GCD}(a, b) = \text{GCD}(b, r)$

Let $a, b \in \mathbb{Z}$ so that $b \neq 0$. By the division algorithm we have that $a = qb + r$ where $q, r \in \mathbb{Z}$ and $0 \leq r < |b|$.

We have that

$$\text{GCD}(a, b) = \text{GCD}(b, r)$$

Proof:

Let $d = \text{GCD}(a, b)$. By definition of the greatest common divisor, we have that $d \mid a$ and $d \mid b$. By the division algorithm we have that $a = qb + r$ where $q, r \in \mathbb{Z}$ and $0 \leq r < |b|$.

Hence as $d \mid a$ then $d \mid (qb + r)$. Now, as $r = a - qb$ then $d \mid r$. Hence by definition of the greatest common divisor, we must have that $d \leq \text{GCD}(b, r)$ as d is a common divisor of b and r .

Now suppose that $g = \text{GCD}(b, r)$ then $g \mid b$ and $g \mid r$. However, by proposition 2.2.4 part 3. as $g \mid b$ and $g \mid r$ then $\forall x, y \in \mathbb{Z}$ we have that $g \mid (bx + yr)$. In particular, we have that $g \mid (qb + r)$. But if $g \mid (qb + r)$ then as $a = qb + r$ we have that $g \mid a$.

Therefore we have that $g \leq \text{GCD}(a, b)$. Combining the two directions gives us that

$$\begin{aligned} d &= \text{GCD}(a, b) \leq \text{GCD}(b, r) \\ g &= \text{GCD}(b, r) \leq \text{GCD}(a, b) \end{aligned}$$

That is, $d \leq g$ and $g \leq d$ which is true if and only if $d = g$. Which is to say $\text{GCD}(a, b) = \text{GCD}(b, r)$. As required. \square

We are almost ready to formalise the process of computing the greatest common divisor. The last step to show is that repeatedly applying the division algorithm doesn't result in a process that never ends. We have for integers a and b that the division algorithm gives $a = qb + r$ where $0 \leq r < |b|$. Another application applied to b and r would give $b = q'r + \tilde{r}$ where we have $\leq \tilde{r} < |r| < |b|$.

Clearly then, applying multiple stages of the division algorithm will always cause the remainder at each stage to decrease, and by the condition that $0 \leq r < |b|$ this process ultimately will give a remainder of 0. For if not then there would be some integer x so that $0 \leq x < 1$ is a contradiction. We formally prove this result.

Proposition 2.2.9. *Remainders from multiple applications of division algorithm decrease to 0*

Let $a, b \in \mathbb{Z}$ with $b \neq 0$. Consider the result of the division algorithm on a, b , i.e

$$a = qb + r, \quad 0 \leq r < |b|$$

Likewise consider applying the division algorithm to b and r to get

$$b = \tilde{q}r + \tilde{r}, \quad 0 \leq \tilde{r} < r$$

If we continually apply this process we have that the remainder is eventually zero.

Proof:

By proposition 2.2.8, we know that $\text{GCD}(a, b) = \text{GCD}(b, r)$ where r is the remainder from the division algorithm and $0 \leq r < |b|$.

Applying the division algorithm to b and r gives us again, by proposition 2.2.8 that $\text{GCD}(b, r) = \text{GCD}(r, r_1)$ where $0 \leq r_1 < |r|$.

Continuing in this fashion for n applications we get the chain of inequalities

$$0 \leq r_n < |r_{n-1}| < |r_{n-2}| < \cdots < |r_2| < |r_1| < |r|$$

Now, for any integers $x, y \in \mathbb{Z}$, where $x \geq 0$ and $y \geq 0$, we have that the largest value of x so that $x < y$ is given by $x = y - 1$. Hence, in the chain of inequalities for the remainder, the smallest decrease from one remainder to the next is 1 and hence there can only be at most r such decreases. If there were more than r decreases, then at the n -th application we would have $r_n < 0$ a contradiction to the division algorithm.

This bounds the length of the chain of inequalities to be at most r and therefore we eventually get to 0 as required. \square

We can now formalise the process for computing the greatest common divisor using repeated applications of the division algorithm.

Theorem 2.2.4. *The Euclidean algorithm*

Let $a, b \in \mathbb{Z}$ so that $b \neq 0$, and suppose that $|a| \geq |b|$. Let $x, y \in \mathbb{Z}$ so that $x = a$ and $y = b$. We have that following these steps computes the greatest common divisor of a and b .

1. Let $d = \text{GCD}(x, y)$. If $b = 0$ then $d = a$ and there is nothing more to do.
2. Otherwise, $b \neq 0$ so use the division algorithm to write $a = qb + r$ where $0 \leq r < |b|$.
3. Let $x = b$ and $y = r$, then by the division algorithm we have that $|b| \geq |r|$.
4. Go back to step 1. and repeat until $y = 0$.

Following these steps gives us that $d = \text{GCD}(a, b)$ is the value of x after these steps have been performed. This is to say we have that $d = \text{GCD}(a, b) = x$

Proof:

Let $a, b \in \mathbb{Z}$ be as stated in the theorem. Let $x = a$ and $y = b$. By the division algorithm we know that $a = qb + r$ for some $q, b \in \mathbb{Z}$ where $0 \leq r < |b|$. Moreover by proposition 2.2.8 we have that $\text{GCD}(a, b) = \text{GCD}(b, r)$.

By this proposition, we are therefore looking for the value of $\text{GCD}(b, r)$. By proposition 2.2.9 we know that the chain of remainders that are generated by repeatedly using the division algorithm must eventually be 0. Hence at some point, we are computing $\text{GCD}(r_n, 0)$ after some step n . The value of $\text{GCD}(r_n, 0) = r_n$. Which is the required greatest common divisor. \square

Theorem 2.2.4 has shown that we can calculate the greatest common divisor for any integers $a, b \in \mathbb{Z}$ where $b \neq 0$. With this theorem, we can now assume that whenever $d = \text{GCD}(a, b)$ is stated we know the value of d by applying this algorithm. We can now consider properties of the GCD. One such example is $\text{GCD}(ma, mb)$ for some $m \in \mathbb{Z}$. Clearly if $d = \text{GCD}(a, b)$ then $d \mid ma$ and $d \mid mb$ so $d \mid \text{GCD}(ma, mb)$. As we will see it turns out that we must have in fact, that $d = \text{GCD}(ma, mb)$. Another property is a particular application of proposition 2.2.4 part 3.

We know from part 3. that if $a \mid b$ and $a \mid c$ then for all integers $x, y \in \mathbb{Z}$ that $a \mid (bx + cy)$. Now suppose that $d = \text{GCD}(a, b)$, then by definition we have that $d \mid a$ and $d \mid b$ then $d \mid (ax + by)$ for any $x, y \in \mathbb{Z}$. By the definition of divisibility, we have that $ax + by = cd$ for some $c \in \mathbb{Z}$. The question now is, is it possible to have $c = 1$?

As it turns out the answer is yes.

Theorem 2.2.5. *Bézout's Identity*

Let $a, b \in \mathbb{Z}$ so that $b \neq 0$ and consider $d = \text{GCD}(a, b)$. Then, there exists $x, y \in \mathbb{Z}$ so that

$$d = ax + by$$

Proof:

Let $a, b \in \mathbb{Z}$ be as given and let $d = \text{GCD}(a, b)$. By proposition 2.2.4 part 3. we have that as $d \mid a$ and $d \mid b$ then we have that for all $x, y \in \mathbb{Z}$ that $d \mid (ax + by)$.

Let S denote the set of all such $ax + by$, that is

$$S = \{ax + by : x, y \in \mathbb{Z}\}$$

Now, it is clear that there are $s \in S$ where $s < 0$ and $s \in S$ where $s > 0$. Moreover, we clearly have $0 \in S$ as we can take $x = 0$ and $y = 0$.

Now consider the set \tilde{S} given by

$$\tilde{S} = \{s \in S : s > 0\}$$

We have by definition of \tilde{S} that $\forall s \in \tilde{S}$ that $s > 0$ and so $\tilde{S} \subset \mathbb{N}$. Hence by the well-ordering principle, theorem 1.3.16, there is a smallest element, say \bar{s} . By definition of being an element of \tilde{S} we have that $\bar{s} = ax_0 + by_0$ for some $x_0, y_0 \in \mathbb{Z}$, where x_0, y_0 each have a fixed value.

We show that $\bar{s} \mid a$ and $\bar{s} \mid b$. Suppose instead that $\bar{s} \nmid a$, then by the division algorithm we have that $a = q\bar{s} + r$ where $0 < r < |\bar{s}|$. It hence follows that

$$\begin{aligned} a &= q\bar{s} + r \\ r &= a - q\bar{s} \\ r &= a - q(ax_0 + by_0) \\ r &= a - qax_0 - qby_0 \\ r &= a(1 - qx_0) + b(-qy_0) \end{aligned}$$

This gives us at $r \in \tilde{S}$. We know that by the division algorithm that $0 < r < |\bar{s}|$ hence $r < \bar{s}$ which gives a contradiction to the well-ordering principle. Meaning that $\bar{s} \nmid a$ is false so it must be the case that $\bar{s} \mid a$. A similar argument shows that $\bar{s} \mid b$.

Now, we have that $d = \text{GCD}(a, b)$ and so $a = md$ and $b = nd$ for some $n, m \in \mathbb{Z}$. Moreover, we have that $\bar{s} = ax_0 + by_0$. So we have that

$$\begin{aligned} \bar{s} &= ax_0 + by_0 \\ \bar{s} &= (md)x_0 + (nd)y_0 \\ \bar{s} &= d(mx_0 + ny_0) \end{aligned}$$

Hence by the definition of divisibility, we conclude that $d \mid \bar{s}$. Applying part 5. of proposition 2.2.4 we have that $d \leq \bar{s}$. But as d is the greatest common divisor of a and b we can't have $d < \bar{s}$, so it follows $d = \bar{s}$ as required. \square

We now note the more standard properties of the greatest common divisor.

Proposition 2.2.10. *Properties of the greatest common divisor*

Let $a, b \in \mathbb{Z}$ with $b \neq 0$. We have the following properties of the GCD hold.

1. $\text{GCD}(a, a) = a$
2. $\text{GCD}(a, b) = \text{GCD}(b, a)$
3. Let D be the set of all common divisors of a and b . then $\forall d \in D$ we have that $d \mid \text{GCD}(a, b)$
4. We have that $\text{GCD}(a, b)$ is the smallest such $ax + by$ where $x, y \in \mathbb{Z}$ so that $\text{GCD}(a, b) = ax + by$
5. Let $m \in \mathbb{Z}$ with $m > 0$, then $\text{GCD}(am, bm) = m * \text{GCD}(a, b)$
6. If $d \mid a$ and $d \mid b$ where $d \in \mathbb{Z}$ and $d > 0$ then $\text{GCD}\left(\frac{a}{d}, \frac{b}{d}\right) = \frac{1}{d} \text{GCD}(a, b)$
7. If $\text{GCD}(a, b) = d$ then $\text{GCD}\left(\frac{a}{d}, \frac{b}{d}\right) = 1$

Proof:

1. $\text{GCD}(a, a) = a$:

Clearly, we have that $a \mid a$. Now by proposition 2.2.4 part 5. We have that if $a \mid a$ with $a > 0$ then $a \leq a$. Hence a is the largest such divisor so $\text{GCD}(a, a) = a$

2. $\text{GCD}(a, b) = \text{GCD}(b, a)$:

This is trivial. If $d = \text{GCD}(a, b)$ then d is the largest common divisor of a and b .

3. Let D be the set of all common divisors of a and b . then $\forall d \in D$ we have that $d \mid \text{GCD}(a, b)$:

Let D be defined as above, then

$$D = \{x \in \mathbb{Z} : x > 0 \text{ and } x \mid a \text{ and } x \mid b\}$$

Then by definition of D we have that $\forall d \in D$ that d is a common divisor of a and d is a common divisor of b . Clearly then $d \mid \text{GCD}(a, b)$ as $\text{GCD}(a, b)$ is the largest such common divisor of a and b and therefore $\text{GCD}(a, b) \in D$.

4. We have that $\text{GCD}(a, b)$ is the smallest such $ax + by$ where $x, y \in \mathbb{Z}$ so that $\text{GCD}(a, b) = ax + by$:

This follows from the proof of theorem 2.2.5. For it it were not we would have a contradiction.

5. Let $m \in \mathbb{Z}$ with $m > 0$, then $\text{GCD}(a, b) = m \text{GCD}(a, b)$:

By the previous part we have that $\text{GCD}(a, b)$ is the smallest such element of the set

$$S = \{ax + by : x, y \in \mathbb{Z}\}$$

Let $s \in S$ denote the smallest such $ax + by$, that is $s = ax + by$ and $s = \text{GCD}(a, b)$.

As $s = \text{GCD}(a, b)$ then $s \mid a$ and $s \mid b$. As $s \mid a$ then $a = ks$ for some $k \in \mathbb{Z}$ and so $am = k(ms)$ which is to say $ms \mid am$. Likewise as $s \mid b$ then $b = ls$ for some $l \in \mathbb{Z}$ and hence $bm = l(ms)$ giving $ms \mid bm$.

Now as $s = ax + by$ then we have that $ms = m(ax + by) = a(mx) + b(my)$. Moreover, as $s \in S$ is the smallest such $ax + by$ then $m(ax + by)$ will be the smallest such element of the set

$$\tilde{S} = \{amx + bmy : x, y \in \mathbb{Z}\}$$

Hence we have that $amx + bmy = \text{GCD}(am, bm) = ms = m * \text{GCD}(a, b)$.

6. If $d \mid a$ and $d \mid b$ where $d \in \mathbb{Z}$ and $d > 0$ then $\text{GCD}\left(\frac{a}{d}, \frac{b}{d}\right) = \frac{1}{d} \text{GCD}(a, b)$

Let $a, b, d \in \mathbb{Z}$ so that $d \mid a$ and $d \mid b$. As $d \mid a$ then we have that $\frac{a}{d} \in \mathbb{Z}$, likewise as $d \mid b$ then $\frac{b}{d} \in \mathbb{Z}$. The result now follows by applying the previous part.

7. If $\text{GCD}(a, b) = d$ then $\text{GCD}\left(\frac{a}{d}, \frac{b}{d}\right) = 1$:

This follows by the previous part.

Concluding the proof. \square

We have talked a lot about the greatest common divisor but nothing about the least common multiple. As with common divisors, we start by making a definition of a common multiple.

Definition 2.2.9. *Common multiple*

Let $a, b, c \in \mathbb{Z}$ so that $a \mid m$ and $b \mid m$. We say that m is a common multiple of a and b .

Example 2.2.15. Let $a = 2$, $b = 4$ and $c = 8$. We have that $2 \mid 8$ and $4 \mid 8$ and so 8 is a common multiple of 2 and 4. In fact, 4 is a common multiple of 2 and 4.

Example 2.2.16. Let $a = 4$ and $b = 14$. Listing multiples of 2 we have 4, 8, 12, 16, 20, 24, 28, 32 and so on. Doing a similar procedure for 14 we see we have 14, 28, 42 and so on. We see that 28 is a common multiple of 4 and 14.

Example 2.2.17. Consider $a = 24$ and $b = 54$. Listing the first ten multiples of a and b we have

24, 48, 72, 96, 120, 144, 168, 192, 216, 240, ...
 54, 108, 162, 216, 270, 324, 378, 432, 486, 540, ...

The first common multiple is 216. Interestingly, we saw that $\text{GCD}(a, b)$ was 6. We have that $216 * 6 = 1296$ and $24 * 54 = 1296$.

Example 2.2.18. We observe for any integer a that $a \mid 0$ as $0 = am$ for some $m \in F$ and by proposition 1.5.9 we must have either $a = 0$ or $m = 0$. Hence 0 can be argued to be a common multiple of any integers a and b . This result is not particularly useful.

These examples indicate that a common multiple always exists. In fact, there is always a smallest common multiple

Theorem 2.2.6. The least common multiple of two integers exists

Let $a, b \in \mathbb{Z}$ where $a > 0$ and $b > 0$. We have that $\exists m \in \mathbb{Z}$ with $m > 0$ so that m is the smallest common multiple of a and b . That is m is the smallest such integer so that $a \mid m$ and $b \mid m$.

Proof:

We first prove that a non-trivial common multiple of a and b exists. That is some $m \neq 0$ as 0 can be viewed as a common divisor of any two integers a, b . Clearly ab is a common multiple of a and b as $a \mid ab$ and $b \mid ab$. Hence a non-trivial common multiple exists.

It is left to show that there is a minimal common multiple. Let S be the set of all positive common multiples of a and b . By the well-ordering principle, S has a smallest element as $S \subset \mathbb{N}$. The result follows. \square

We can now make a formal definition. However, first, we can note that the restriction of $a > 0$ and $b > 0$ is not needed.

Corollary 2.2.1. Let $a, b \in \mathbb{Z}$, where $a \neq 0$ and $b \neq 0$. We have that $\exists m \in \mathbb{Z}$ with $m > 0$ so that m is the smallest common multiple of a and b . This is, m is the smallest such integer so that $a \mid m$ and $b \mid m$.

Proof:

The proof is similar to theorem 2.2.6. We have that ab is a common multiple of a and b as is $-ab$. Hence we have that one of $ab > 0$ or $-ab > 0$. Let S be the set of all positive common multiples of a and b . Then the well-ordering principle gives us that S has the smallest such element. \square .

Definition 2.2.10. Least common multiple

Let $a, b \in \mathbb{Z}$ so that $a \neq 0$ and $b \neq 0$. We say that the smallest positive value m so that $a \mid m$ and $b \mid m$ is the least common multiple of a and b , denoted $m = \text{LCM}(a, b)$, sometimes written $\text{lcm}(a, b)$.

It is important to note why we say that the least common multiple is positive. If we allowed a negative least common multiple, say $-m$, then for all $n \in \mathbb{Z}$ with $n > 0$ we have that $-nm$ is a smaller common multiple than $-m$ and so we could always find a smaller such multiple.

As with the greatest common divisor, we need a way to compute the least common multiple. We should look again at the example where $a = 24$ and $b = 54$. We saw that the first, smallest, common multiple was 216, and that the greatest common divisor was 6. We also noted that the product $ab = 1296$ which is also the product $216 * 6$. We should look to more examples to see if this holds in other cases.

Example 2.2.19. Let $a = 14$ and $b = 21$. Using the method of writing multiples out we have

14, 28, 42, 56, ...
 21, 42, 63, 84, ...

So the smallest positive common multiple is 42. Now, $\text{GCD}(14, 21) = 7$. Finally, $14 * 21 = 294$ and $7 * 42 = 294$.

Hence we have that $\text{LCM}(14, 21) = \frac{14 * 21}{\text{GCD}(14, 21)}$.

In general we might expect that $\text{LCM}(a, b) = \frac{a * b}{\text{GCD}(a, b)}$

Example 2.2.20. Let $a = 6$ and $b = 36$. Using our expected result, we have that $\text{LCM}(a, b) = \frac{a * b}{\text{GCD}(a, b)}$.

So computing $\text{GCD}(a, b)$ we see that $\text{GCD}(a, b) = 6$ and so we suspect that $\text{LCM}(6, 36) = \frac{6 * 36}{6} = 36$.

Writing out the multiples of both 6 and 36

6, 12, 18, 24, 30, 36, 42, ...
36, 72, 108, ...

So the smallest common multiple is indeed 36.

We have enough evidence to postulate and prove the following theorem.

Theorem 2.2.7. Least common multiple by greatest common divisor equals product

Let $a, b \in \mathbb{Z}$ so that $a > 0$ and $b > 0$. We have that

$$\text{GCD}(a, b) * \text{LCM}(a, b) = ab$$

Proof:

Let $d = \text{GCD}(a, b)$, then by definition we have that $d \mid a$ so by proposition 2.2.4 part 1. implies that $d \mid ac$ for any $c \in \mathbb{Z}$ and in particular $d \mid ab$. Hence by the definition of divisibility, there exists $n \in \mathbb{Z}$ so that $ab = dn$.

Now as $d \mid a$ then there is an integer u so that $a = du$, likewise as $d \mid b$ then there is an integer v so that $b = dv$. Hence we have that

$$dn = dub \Rightarrow n = ub, \text{ By the cancellation law for the integers}$$

$$dn = adv \Rightarrow n = av, \text{ By the cancellation law for the integers}$$

Hence as $n = ub$ we have that $b \mid n$ and likewise as $n = av$ we have that $a \mid n$. Hence it follows that n is a common multiple of a and b . We need to show that n is the smallest such multiple so then $\text{LCM}(a, b) = n$.

So, let S denote the set of positive common multiples of a and b and let $s \in S$ be a common multiple of a and b . By definition of a common multiple, we have that there exists some $k_1, k_2 \in \mathbb{Z}$ so that $s = ak_1$ and $s = bk_2$.

Now, we have by Bézout's identity we have that $\exists x, y \in \mathbb{Z}$ so that

$$\text{GCD}(a, b) = d = ax + by$$

Now, consider sd , we have that

$$\begin{aligned} sd &= s(ax + by) \\ &= sax + sby \\ &= (bk_2)ax + (ak_1)by \\ &= abk_2x + abk_1y \\ &= ab(k_2x + k_1y) \\ &= dn(k_2x + k_1y) \\ s &= n(k_2x + k_1y), \text{ By the cancellation law for the integers} \end{aligned}$$

Now $(k_2x + k_1y) \in \mathbb{Z}$ and so we have that $n \mid s$. Now by proposition 2.2.4 part 5. we have that $n \leq s$. As $s \in S$ was arbitrary we have that n divides the smallest element of S by the well-ordering principle, i.e n is the smallest common divisor and so by definition $\text{LCM}(a, b) = n$.

Hence we have that $ab = dn = \text{GCD}(a, b) \text{LCM}(a, b)$. As required. \square .

We can now justify the following corollary to compute the least common multiple.

Corollary 2.2.2. *Least common multiple is product divided by greatest common divisor*

Let $a, b \in \mathbb{Z}$ so that $a > 0$ and $b > 0$. We have that

$$\text{LCM}(a, b) = \frac{ab}{\text{GCD}(a, b)}$$

Proof:

By theorem 2.2.7 we have that

$$\text{GCD}(a, b) * \text{LCM}(a, b) = ab$$

Let $d = \text{GCD}(a, b)$ then by definition we have that $d \mid a$ and $d \mid b$ so that $d \mid ab$. Hence $\frac{ab}{d} \in \mathbb{Z}$. Hence $\text{LCM}(a, b) \in \mathbb{Z}$. \square

We can now show some similar results to proposition 2.2.10

Proposition 2.2.11. *Properties of the least common multiple*

Let $a, b \in \mathbb{Z}$ with $a > 0$ $b > 0$. We have the following properties of the LCM hold.

1. $\text{LCM}(a, a) = a$
2. $\text{LCM}(a, b) = \text{LCM}(b, a)$
3. *Let M be the set of all positive common multiples of a and b . then $\forall m \in M$ we have that $\text{LCM}(a, b) \mid m$*
4. *We have that $\text{LCM}(a, b)$ is the greatest $\frac{ab}{ax + by}$ where $\text{GCD}(a, b) = ax + by$.*

Proof:

1. $\text{LCM}(a, a) = a$:

*As $\text{GCD}(a, a) = a$ and $a * a = a^2$, we have by corollary 2.2.2 that*

$$\text{LCM}(a, a) = \frac{a * a}{\text{GCD}(a, a)} = \frac{a^2}{a} = a$$

2. $\text{LCM}(a, b) = \text{LCM}(b, a)$:

This follows as $\text{GCD}(a, b) = \text{GCD}(b, a)$ and integer multiplication is commutative, this is to say

$$\text{LCM}(a, b) = \frac{a * b}{\text{GCD}(a, b)} = \frac{b * a}{\text{GCD}(b, a)} = \text{LCM}(b, a)$$

3. *Let M be the set of all positive common multiples of a and b . then $\forall m \in M$ we have that $\text{LCM}(a, b) \mid m$:*

Let M be the set of all positive common multiples. By the well-ordering principle, there is a smallest element \tilde{m} . By the definition of the least common multiple we have that $\text{LCM}(a, b)$ divides any other common multiple, so $\text{LCM}(a, b) \mid \tilde{m}$. For every $m \in M$, we have that $m \geq \tilde{m}$ and so $\text{LCM}(a, b) \mid m$ for every $m \in M$.

4. We have that $\text{LCM}(a, b)$ is the greatest $\frac{ab}{ax + by}$ where $\text{GCD}(a, b) = ax + by$:

By proposition 2.2.10 part 4. we have that $\text{GCD}(a, b) = ax + by$ for some $x, y \in \mathbb{Z}$ is the smallest such $ax + by$. Hence

$$\text{LCM}(a, b) = \frac{ab}{\text{GCD}(a, b)}$$

Will be the greatest such fraction. For if not then there is either $x_0, y_0 \in \mathbb{Z}$ so that $ax_0 + by_0 < ax + by$ a contradiction to part 4. of proposition 2.2.10, or we have that there is $x_1, y_1 \in \mathbb{Z}$ with $ax_1 + by_1 > ax + by$ then by part 35. of proposition 1.6.11 we have that

$$\frac{ab}{ax_1 + by_1} < \frac{ab}{ax + by}$$

Concluding the proof. \square

2.3 Prime and co-prime numbers

God may not play dice with the universe, but something strange is going on with the prime numbers.

Paul Erdos

So far we have been building a theory of divisibility. This theory has allowed us to define what it means to be an odd or an even integer. To know when one integer divides another, and computing the largest divisor of two integers. Where do we go from here? One question we could ask is how many divisors does a given integer have?

2.3.1 The divisor function

We start with the following definition.

Definition 2.3.1. *The Divisor function*

Let $x \in \mathbb{Z}$. We define $\sigma : \mathbb{Z} \rightarrow \mathbb{Z}$ by

$$\sigma : \mathbb{Z} \rightarrow \mathbb{Z}$$

$$x \mapsto \sigma(x) = \sum_{d|x} 1$$

here we are summing over all of the divisors d of x , where if $d \mid x$ then we add one to the sum total.

Rather than work with explicit examples we will provide a table of the first 20 integers.

x	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$\sigma(x)$	1	2	2	3	2	4	2	4	3	4	2	6	2	4	4	5	2	6	2	6

Table 7: The divisor function for the integers $1 \leq x \leq 20$

There are a few things to note from this table. Firstly the only integer with a single divisor is 1. Secondly, there are many examples of integers having only 2 divisors. These are 2, 3, 5, 7, 11, 13, 17 and 19. As 1 is a divisor of every integer we can conclude the other divisors in the case of $\sigma(x) = 2$ must be x itself.

What about the case when $\sigma(x) > 2$. Looking at 6 we see the divisors are 1, 2, 3 and 6 itself, and from the table $\sigma(2) = \sigma(3) = 2$. Moreover, we have that $6 = 2 * 3$.

Similarly with 12 we have that the divisors are 1, 2, 3, 4, 6 and 12. Again, we have that $\sigma(2) = \sigma(3) = 2$. Now, as $12 = 2 * 6$ and $6 = 2 * 3$ then we have that $12 = 2 * 2 * 3$. In both cases, we have seen that a number x with $\sigma(x) > 2$ can be written into a product of integers with exactly 2 divisors. We can ask does this hold in general? To do so we need to make some definitions.

2.3.2 Prime numbers

With the remarks of the previous section, we give a special name to any integer x where $\sigma(x) = 2$.

Definition 2.3.2. *Prime number*

Let $x \in \mathbb{Z}$ with $x \geq 2$. We say that x is a prime number, or simply that x is prime, if and only if $\sigma(x) = 2$. In other words, we say that x is prime, if and only if the only two distinct positive divisors of x are 1 and itself. If x is not prime we say that x is composite.

We noted that there were many $x \in \mathbb{Z}$ with $\sigma(x) = 2$. A natural question that arises is are there infinitely many such x , or are there only finitely so many? To answer this we need to see how primes and divisibility interact. We first have to make another definition based on the greatest common divisor of two integers. We show some examples to motivate this new definition.

Example 2.3.1. Let $a = 6$ and $b = 35$. By the Euclidean algorithm, we see that

$$\begin{aligned} 35 &= 5(6) + 5 \\ 6 &= 5 + 1 \\ 5 &= 5(1) \end{aligned}$$

Hence $\text{GCD}(a, b) = 1$.

Example 2.3.2. Let $a = 2$ and $b = 3$. By the Euclidean algorithm, we see that

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

Hence $\text{GCD}(a, b) = 1$. We note that a and b are prime.

Example 2.3.3. Let $a = 4$ and $b = 9$. By the Euclidean algorithm, we see that

$$\begin{aligned} 9 &= 2(4) + 1 \\ 4 &= 4(1) \end{aligned}$$

Hence $\text{GCD}(a, b) = 1$.

We see that there are integers $a, b \in \mathbb{Z}$ so that $\text{GCD}(a, b) = 1$. Meaning that they have no common divisors other than 1. This situation turns out to happen enough in Number Theory to warrant a definition.

Definition 2.3.3. *Co-prime Integers*

Let $a, b \in \mathbb{Z}$. We say that a is co-prime to b , or a and b are co-prime, or a and b are relatively prime, if and only if $\text{GCD}(a, b) = 1$.

We have some immediate results.

Proposition 2.3.1. *Bézout's Identity for co-prime integers*

Let $a, b \in \mathbb{Z}$ so that $\text{GCD}(a, b) = 1$. We have that $\exists x, y \in \mathbb{Z}$ so that

$$1 = ax + by$$

Proof:

This immediately follows from theorem 2.2.5. \square

Proposition 2.3.2. *Distinct prime numbers are co-prime*

Let $p, q \in \mathbb{Z}$ so that p and q are prime. We have that $\text{GCD}(p, q) = 1$.

Proof:

Let $p, q \in \mathbb{Z}$ so that p and q are prime and $p \neq q$. As p is prime then the only positive divisors are p and 1, likewise for q . Hence the largest divisor of both p and q is 1 so that $\text{GCD}(p, q) = 1$ by definition. \square

Corollary 2.3.1. *Prime not dividing integer implies co-prime*

Let $a, p \in \mathbb{Z}$ where p is prime. If $p \nmid a$ then $\text{GCD}(a, p) = 1$

Proof:

Let $a, p \in \mathbb{Z}$ where p is prime and where $p \nmid a$. Suppose that $\text{GCD}(a, p) = d$ for some $d \in \mathbb{Z}$. By definition of the greatest common divisor, we have that $d \mid p$ and by definition of p prime, we have that either $d = 1$ or $d = p$. But if $d = p$ then $p \mid a$ by definition of the greatest common divisor, contradicting the assumption that $p \nmid a$. Hence $d = 1$. \square

Proposition 2.3.3. *Product of co-prime integers is equal to their least common multiple*

Let $a, b \in \mathbb{Z}$ so that $\text{GCD}(a, b) = 1$. We have that $ab = \text{LCM}(a, b)$.

Proof:

Let $a, b \in \mathbb{Z}$ be as given in the proposition. We have by corollary 2.2.2 that

$$\text{LCM}(a, b) = \frac{ab}{\text{GCD}(a, b)}$$

As a and b are co-prime, we have $\text{GCD}(a, b) = 1$, hence the result. \square

Proposition 2.3.4. *Coefficients in Bézout's identity are co-prime*

Let $a, b \in \mathbb{Z}$ with $d = \text{GCD}(a, b)$ so that by Bézout's identity we have $\exists x, y \in \mathbb{Z}$ so that

$$d = ax + by$$

We have that $\text{GCD}(x, y) = 1$

Proof:

Let $a, b \in \mathbb{Z}$ with $d = \text{GCD}(a, b)$. By Bézout's identity we have that there exists $x, y \in \mathbb{Z}$ so that

$$d = ax + by$$

Now, dividing by d gives

$$1 = \frac{a}{d}x + \frac{b}{d}y$$

As $d \mid a$ and $d \mid b$. Hence we have that $1 = k_1x + k_2y$ where $k_1 = \frac{a}{d}$ and $k_2 = \frac{b}{d}$. Hence $\text{GCD}(x, y) = 1$ and so by definition x and y are co-prime. \square

With some basic results out of the way, we can start seeing more meaningful consequences of defining prime and co-prime numbers. One of the first things we should do is see how primes divide other integers.

Example 2.3.4. *Let $n = 10$, we have that $2 \mid 10$ and $\sigma(2) = 2$, hence 2 is prime. Moreover $10 = 2 * 5$ and clearly $2 \mid 2$.*

Example 2.3.5. *let $n = 4$, clearly $4 = 2 * 2$ and so $2 \mid 4$. Moreover, $2 \mid 2$.*

Example 2.3.6. *Let $n = 14 = 2 * 7$. Both 2 and 7 are prime and so $2 \mid 14$ and $7 \mid 14$.*

Then, if a prime p divides $n = ab$ we seem to have that either $p \mid a$ or $p \mid b$.

Lemma 2.3.1. *Euclid's Lemma*

Let $a, b \in \mathbb{Z}$ and let $p \in \mathbb{Z}$ be prime. Suppose that $p \mid ab$. We have that either $p \mid a$ or $p \mid b$.

Proof:

Let $p \mid ab$. Suppose that $p \nmid b$. As the only divisors of p are 1 and itself then we have that $\text{GCD}(p, b) = 1$ by corollary 2.3.1. Now by proposition 2.3.1 we have that $\exists x, y \in \mathbb{Z}$ so that

$$1 = px + by$$

Multiplying by a gives $a = apx + aby$ and as $p \mid apx$ and $p \mid ab$ we have that $p \mid a$. Likewise if $p \nmid a$. \square

This result generalises to products of more than two integers.

Lemma 2.3.2. *Generalised Euclid's lemma*

Let $p \in \mathbb{Z}$ be prime. Let $n \in \mathbb{Z}$ be such that

$$n = \prod_{i=1}^m a_i$$

where $a_i \in \mathbb{Z}$ for each i . Suppose that $p \mid n$, then there exists an $i \in \mathbb{N}$ so that $p \mid a_i$.

Proof:

We argue by induction on m . The base case is $m = 2$ which follows by Euclid's lemma. So suppose the result holds for some $k > 2$ that is if n is such that

$$n = \prod_{i=1}^k a_i$$

then there is some $i \in \mathbb{N}$ so that $p \mid a_i$. We show that if n is such that

$$n = \prod_{i=1}^{k+1} a_i$$

then there is some $i \in \mathbb{N}$ so that $p \mid a_i$. So suppose that $p \mid n$, then

$$p \mid \prod_{i=1}^{k+1} a_i$$

We have that

$$\begin{aligned} p &\mid \prod_{i=1}^{k+1} a_i \\ p &\mid \left(\prod_{i=1}^k a_i * a_{k+1} \right) \end{aligned}$$

By the induction hypothesis we have that as $p \mid \prod_{i=1}^k a_i$ then there is some $i \in \mathbb{N}$ so that $p \mid a_i$ where $1 \leq i \leq k$. Hence we have that either $p \mid a_i$ or $p \mid a_{k+1}$. The result now follows by induction. \square

With Euclid's lemma, we can provide a very famous theorem. Namely, there is no $x \in \mathbb{Q}$ so that $x^2 = 2$. We first need a definition, based on co-prime integers.

Definition 2.3.4. *Reduced fraction*

Let $x \in \mathbb{Q}$ where $x = \frac{a}{b}$ and $b \neq 0$. We say that x is a reduced fraction, or a fraction in its lowest terms if $\text{GCD}(a, b) = 1$.

We give some examples.

Example 2.3.7. Let $x = \frac{1}{2}$. As $\text{GCD}(1, 2) = 1$ we have that x is a reduced fraction.

Example 2.3.8. Let $x = \frac{3}{6}$. We can compute that $\text{GCD}(3, 6) = 3$, hence we have that $3 \mid 3$ and $3 \mid 6$. We hence can write

$$x = \frac{3}{6} = \frac{3 * 1}{3 * 2} = \frac{1}{2}$$

And as $\text{GCD}(1, 2) = 1$ we can conclude x is now in its lowest terms.

We can now show the theorem.

Theorem 2.3.1. *No rational exists whose square is 2*

We have that $\nexists x \in \mathbb{Q}$ with $x^2 = 2$.

Proof:

Suppose instead that $x \in \mathbb{Q}$ where $x = \frac{a}{b}$ with $b \neq 0$. Moreover assume that x is a reduced fraction, i.e. $\text{GCD}(a, b) = 1$. We can make this assumption as otherwise we can reduce x until it is reduced without affecting the proof.

We have that

$$\begin{aligned}x^2 &= 2 \\ \frac{a^2}{b^2} &= 2 \\ a^2 &= 2b^2\end{aligned}$$

Hence by the definition of divisibility, we have $2 \mid a^2$ and so by Euclid's lemma we have that $2 \mid a$ as 2 is prime. So write $a = 2k$ for some $k \in \mathbb{Z}$. Then we have that

$$\begin{aligned}a^2 &= 2b^2 \\ (2k)^2 &= 2b^2 \\ 4k^2 &= 2b^2 \\ 2k^2 &= b^2\end{aligned}$$

Hence $2 \mid b^2$ and again by Euclid's lemma we have that $2 \mid b$. We have a contradiction as $2 \mid a$ and $2 \mid b$ implies that $\text{GCD}(a, b) \geq 2$ and so x can't have been a reduced fraction. But then if x was not a reduced fraction and a and b can't be co-prime then we can conclude that there is no rational x so that $x^2 = 2$. \square

This raises the question if there is no rational x whose square is 2 then what exactly is x ? Unfortunately, we are not quite ready to properly answer this question in a satisfying way, all we can is that we have seen a hint of a new type of number. One that we can define but not study in more detail at the moment.

Definition 2.3.5. *Irrational number*

If we have $x \notin \mathbb{Q}$, then we say that x is irrational. In other words, x is irrational if and only if $x = \frac{a}{b}$ where $a, b \in \mathbb{Z}$ and $b \neq 0$.

Clearly, if S denotes the set of irrational numbers then by theorem 2.3.1 that $S \neq \emptyset$. Perhaps then it makes sense, for now, to consider which elements of $x \in \mathbb{Q}$ so that $x^2 = y$ where $y \in \mathbb{Z}$, or more restrictively, which $x \in \mathbb{Z}$ are such that we have $x^2 = y$ where $y \in \mathbb{Z}$.

Before we start answering this question, we note one useful result by generalising Euclid's lemma from the prime case to the co-prime case.

Lemma 2.3.3. *Euclid's lemma for co-primes*

Let $a, b, c \in \mathbb{Z}$ and suppose that $c \mid ab$ and $\text{GCD}(b, c) = 1$. We have that $c \mid a$.

Proof:

Let $a, b, c \in \mathbb{Z}$ be such that $c \mid ab$ and $\text{GCD}(b, c) = 1$. As $\text{GCD}(b, c) = 1$, we have by proposition 2.3.1 that there exists integers $x, y \in \mathbb{Z}$ so that

$$bx + cy = 1$$

On multiplication by a we have that $abx + acy = a$. Clearly $c \mid abx$ and $c \mid acy$ and so $c \mid a$ as required. \square

There is a useful application of this lemma.

Example 2.3.9. Let $a, b \in \mathbb{Z}$ and let $d = \text{GCD}(a, b)$. We know by Bézout's identity that $\exists x, y \in \mathbb{Z}$ so that

$$ax + by = d$$

The theorem for Bézout's identity, theorem 2.2.5, doesn't state anything about there not being another pair x', y' so that

$$ax' + by' = d$$

For example, consider $a = 30$ and $b = 105$, then $\text{GCD}(a, b) = 15$ and we have that $15 = -3 * 30 + 1 * 105$, i.e $x = -3$ and $y = 1$ in this case. We could have also have $x = -10$ and $y = 3$ as $-10 * 30 + 3 * 105 = -300 + 315 = 15$.

So supposing that $a, b \in \mathbb{Z}$ and $d = \text{GCD}(a, b)$ we know that $\exists x, x', y, y' \in \mathbb{Z}$ with

$$\begin{aligned} ax + by &= d \\ ax' + by' &= d \end{aligned}$$

Can we find a relation between the pair x and y and the pair x' and y' ? As $d \mid a$ then there exists $a' \in \mathbb{Z}$ so that $a = a'd$ and likewise as $d \mid b$ then there exists $b' \in \mathbb{Z}$ so that $b = b'd$. Hence we see that

$$\begin{aligned} ax + by &= d \\ a'dx + b'dy &= d \\ a'x + b'y &= 1 \end{aligned}$$

Now, we have that x and y are co-prime so we can deduce that a' and b' are also co-prime. Now, we have that

$$ax + by = d = ax' + by'$$

So, re-arranging we see that

$$\begin{aligned} ax - ax' &= by' - by \\ a(x - x') &= b(y' - y) \end{aligned}$$

Dividing by d gives

$$a'(x - x') = b'(y' - y)$$

Now, as a' and b' are co-prime, we have by Euclid's lemma for co-primes that $a' \mid (y' - y)$, We, therefore have that $\exists k \in \mathbb{Z}$ so that

$$y' - y = a'k \Rightarrow y' = y + a'k$$

But as $y' - y = a'k$ we have that

$$\begin{aligned} a'(x - x') &= b'(a'k) \\ x - x' &= b'k \\ x' &= x - b'k \end{aligned}$$

Therefore, we can conclude that

$$\begin{aligned} x' &= x - \frac{b}{d}k \\ y' &= y + \frac{a}{d}k \end{aligned}$$

where $k \in \mathbb{Z}$. To check this is the case we return to the example of $a = 30$ and $b = 105$ where we had that $\text{GCD}(a, b) = 15$. We saw that $x = -3$ and $y = 1$. Using these values in the equations above we get

$$\begin{aligned}x' &= -3 - \frac{105}{15}k \Rightarrow x' = -3 - 7k \\y' &= 1 + \frac{30}{15}k \Rightarrow y' = 1 + 2k\end{aligned}$$

Using $k = 1$ gives us the alternative solution we saw of $x' = -10$ and $y' = 3$.

From Euclid's lemma for co-primes we have deduced the full set of values where $d = \text{GCD}(a, b)$ and $d = ax + by$.

We now return to the problem at hand. We wish to consider the elements of $x \in \mathbb{Z}$ so that $x^2 = y$ where $y \in \mathbb{Z}$. As is the theme of this section we will do some exploratory examples.

Example 2.3.10. Let $x \in \mathbb{Q}$ be such that $x = \frac{2}{1}$, then $x^2 = \frac{4}{1} = 4 \in \mathbb{Z}$. In particular, we have that $4 = 2 * 2 = 2^2$.

Example 2.3.11. Consider $x = \frac{10}{1} = 10$. Clearly $x^2 = 100 \in \mathbb{Z}$. We have that

$$100 = 2 * 50 = 2 * 2 * 25 = 2 * 2 * 5 * 5 = 2^2 * 5^2$$

Example 2.3.12. We generalise the example of there being no $x \in \mathbb{Q}$ so that $x^2 = 2$. We will show that for a prime $p \in \mathbb{Z}$, there is no $x \in \mathbb{Q}$ so that $x^2 = p$. So suppose there is such an x , that is $x = \frac{a}{b}$ so that $a, b \in \mathbb{Z}$ and $b \neq 0$ and moreover suppose that x is a reduced fraction, which is to say $\text{GCD}(a, b) = 1$. We then have that

$$x^2 = \frac{a^2}{b^2} = p \Rightarrow a^2 = pb^2$$

Hence $p \mid a^2$. Hence by Euclid's lemma, we have that $p \mid a$. Hence let $a = pk$ for some $k \in \mathbb{Z}$. We then have that

$$\begin{aligned}a^2 &= pb^2 \\(pk)^2 &= pb^2 \\p^2k^2 &= pb^2 \\pk &= b^2\end{aligned}$$

Therefore $p \mid b^2$ and so by Euclid's lemma we have that $p \mid b$, a contradiction to the assumption that x was a fraction in reduced form.

This last example shows that for any prime p there is no rational number x with $x^2 = p$. We also saw an example of when $x^2 = p^2$, namely when $p = 2$. Also an example of a product of primes satisfying $x^2 = p^2 * q^2$ for some primes p and q . It seems therefore that the question of what $x \in \mathbb{Z}$ so that $x^2 = y$ for some integer y is deeply connected to primes. In particular, we have seen that the powers of the primes must be even. We need more examples before we can make a claim.

Example 2.3.13. Consider $x = 4$, we have that $x^2 = 16$ and 16 is not prime as $\sigma(16) = 31$, with divisors 1, 2, 4 and 8. However, we have that $16 = 2^4$ and we know that 2 is prime.

Example 2.3.14. Let $y = 3^2 * 5^4 = 5625$, a product of primes. We can see that we can take $x = 3 * 5^2 = 75$.

With these examples, we can see that to answer the question of what $x \in \mathbb{Z}$ are such that $x^2 = y$ for some $y \in \mathbb{Z}$, it is enough to consider the structure of the primes that make y . This leads us to, perhaps, the most important theorem of elementary Number Theory¹².

¹²If there is only one theorem you learn when studying Number Theory, it has to be this one!

Theorem 2.3.2. *The fundamental theorem of arithmetic*

Let $n \in \mathbb{Z}$ be such that $n \geq 2$. We have that n can be expressed as a product of one or more primes. This product is uniquely up to the order of the primes. This is to say we have that

$$n = p_1^{e_1} * p_2^{e_2} * p_3^{e_3} * \cdots * p_k^{e_k}$$

where p_i are the primes and e_i are the powers for the prime p_i . Here uniquely up to the order of the primes means that, for example, $6 = 2 * 3 = 3 * 2$ are considered the same product.

Proof:

There are two parts to this theorem, firstly we must show that every integer $n \geq 2$ is expressible as a product of primes. Secondly that this product is unique up to the ordering of the primes.

As a result, we will break this theorem down into two sub-theorems.

Theorem 2.3.3. *Every integer greater than one is expressible as a product of primes*

Let $n \in \mathbb{Z}$ be such that $n > 1$. We have that

$$n = p_1 * p_2 * p_3 * \cdots * p_k$$

where p_i are the primes.

Proof:

We argue by induction on n . The base case is $n = 2$ for which we have $n = 2$ which is a prime. So the base case is immediate. So suppose the result holds for some $k > 2$, that is $n = k$ can be written as a product of primes. We show that $n = k + 1$ can be written as a product of primes.

If $k + 1$ is itself prime we are done, so suppose not, then $\sigma(k + 1) > 2$ and so there are some factors, say a and b so that $k + 1 = ab$, where $2 \leq a < k + 1$ and $2 \leq b < k + 1$. However, this means that we have $2 \leq a \leq k$ and $2 \leq b \leq k$ and so by the induction hypothesis we can write a and b as a product of primes. But then ab will be a product of primes and so $k + 1$ is a product of primes.

The result follows by induction. \square

Theorem 2.3.4. *The product of primes expression for an integer is unique*

Let $n \in \mathbb{Z}$ be such that $n \geq 2$. We have that the expression for n as a product of primes is unique.

Proof:

Let $n \in \mathbb{Z}$ be as given. Suppose that n has two different representations into a product of primes, that is

$$n = p_1 p_2 p_3 \dots p_r$$

$$n = q_1 q_2 q_3 \dots q_s$$

where without loss of generality we suppose that $r \leq s$. Moreover, Without loss of generality suppose that we have the primes in ascending order, that is, $p_1 \leq p_2 \leq p_3 \leq \cdots \leq p_r$ and that $q_1 \leq q_2 \leq q_3 \leq \cdots \leq q_s$.

Now as $p_1 \mid q_1 q_2 q_3 \dots q_s$ we have by Euclid's lemma that $p_1 \mid q_i$ for some $1 \leq i \leq s$. Therefore $p_1 \geq q_1$ as the primes are in ascending order. Likewise, as $q_1 \mid p_1 p_2 p_3 \dots p_r$, then $q_1 \mid p_j$ for some $1 \leq j \leq r$. Hence $q_1 \geq p_1$. As $p_1 \geq q_1$ and $q_1 \geq p_1$ we must have that $p_1 = q_1$. Hence we have

$$p_1 p_2 p_3 \dots p_r = q_1 q_2 q_3 \dots q_s$$

$$p_1 p_2 p_3 \dots p_r = p_1 q_2 q_3 \dots q_s$$

$$p_2 p_3 \dots p_r = q_2 q_3 \dots q_s$$

This process can be repeated for each prime p_j for the remaining $2 \leq j \leq r$. Now if $r < s$ we will eventually get to

$$1 = q_{r+1} q_{r+2} q_{r+3} \dots q_s$$

However the only divisors of 1 are 1 and -1 , hence none of the q_i for $r + 1 \leq i \leq s$ can't be prime, a contradiction. So $r = s$. If $r = s$ then we must have that $p_i = q_i$ for $1 \leq i \leq s$. Hence the two expressions of for n are equal giving us uniqueness. \square

The fundamental theorem of arithmetic now follows from theorem 2.3.3 and theorem 2.3.4. The final result involving the powers of primes is trivial to see. Suppose that n is a product of primes given by

$$n = p_1 p_2 p_3 \dots p_k$$

We will have that some of the p_i will be the same and others will not, if we combine the primes that are equal then we will get that, after re-labelling so that k is once again the largest index that appears,

$$n = p_1^{e_1} * p_2^{e_2} * p_3^{e_3} * \dots * p_k^{e_k}$$

The result is shown. \square

This theorem is of great importance. It ultimately allows us to deal with problems of divisibility by recasting them into statements about the primes that make the integer. We make a quick definition.

Definition 2.3.6. Prime factorisation of an integer

Let $n \in \mathbb{Z}$ where $n = p_1^{e_1} * p_2^{e_2} * p_3^{e_3} * \dots * p_k^{e_k}$. We say that the expression for n is the prime factorisation of n , or simply the factorisation of n

We have shown that any integer can be factored into a product of primes, a natural question we can now ask, and answer, is how many primes are there. Could it be the case that the set of primes finite, if very large? We can see that the number of primes is infinite.

Theorem 2.3.5. Number of primes is infinite

We have that the number of primes is infinite.

Proof:

We will argue by contradiction. Suppose that there are only a finite number of primes, say

$$P = \{p_1, p_2, p_3, \dots, p_n\}$$

where we have that $p_i < p_j$ for $i < j$ and $1 \leq i, j \leq n$, i.e $p_1 = 2$, $p_2 = 3$ etc. Let N be the integer

$$N = (p_1 p_2 p_3 \dots p_n) + 1$$

Clearly, N is not prime as otherwise we would have $N \in P$ but $N > p_n$, which would be a contradiction. So N is composite and by the fundamental theorem of arithmetic, we have that N has a factorisation into primes. Clearly, none of the p_i divide N , but then none of the p_i divide the prime factorisation of N from the fundamental theorem of arithmetic, a contradiction. Hence P can't be a finite set and the number of primes must be infinite. \square

The fundamental theorem of arithmetic can be used to recast some previous results for the greatest common divisor. We start with a result for integers being co-primes.

Proposition 2.3.5. Greatest common divisor is 1 if and only if no-common prime in factorisation

Let $a, b \in \mathbb{Z}$ with $b \neq 0$. We have that $\text{GCD}(a, b) = 1$ if and only if a and b share no common primes in their factorisations.

Proof:

We have that $\text{GCD}(a, b) = 1$ if and only if the largest divisor of both a and b is 1, which occurs if and only if there are no primes in the factorisation of a and in the factorisation of b in common. \square

We can compute the greatest common divisor by considering the prime factorisations of a and b . To do so we need a helpful result.

Proposition 2.3.6. Expression for integers as powers of same primes

Let $a, b \in \mathbb{Z}$ with $a \geq 2$ and $b \geq 2$. Consider the prime factorisations of a and b given by

$$\begin{aligned}
a &= p_1^{e_1} p_2^{e_2} p_3^{e_3} \dots p_n^{e_n} \\
&= \prod_{\substack{p_i | a \\ p_i \text{ is prime}}} p_i^{e_i} \\
b &= q_1^{f_1} q_2^{f_2} q_3^{f_3} \dots q_m^{f_m} \\
&= \prod_{\substack{q_i | b \\ q_i \text{ is prime}}} q_i^{f_i}
\end{aligned}$$

where n need not be equal to m . We have that there exist prime numbers

$$t_1 < t_2 < t_3 \dots < t_v$$

So that

$$\begin{aligned}
a &= t_1^{g_1} t_2^{g_2} t_3^{g_3} \dots t_v^{g_v} \\
b &= t_1^{h_1} t_2^{h_2} t_3^{h_3} \dots t_v^{h_v}
\end{aligned}$$

Proof:

Let $a, b \in \mathbb{Z}$ be as given. We have that

$$\begin{aligned}
a &= p_1^{e_1} p_2^{e_2} p_3^{e_3} \dots p_n^{e_n} \\
b &= q_1^{f_1} q_2^{f_2} q_3^{f_3} \dots q_m^{f_m}
\end{aligned}$$

In particular. Let $A = \{p_1, p_2, p_3, \dots, p_n\}$ and let $B = \{q_1, q_2, q_3, \dots, q_m\}$. We can therefore define the set $T = A \cup B$. Where we have

$$T = \{t_1, t_2, \dots, t_v\}$$

where clearly $v \leq (n + m)$. We now need a way to pick the primes in the factorisation of a , and b , from the set T .

Define ι_A and ι_B as follows

$$\begin{aligned}
\iota_A : A &\rightarrow T \\
x &\mapsto \iota_A(x) = x
\end{aligned}$$

$$\begin{aligned}
\iota_B : B &\rightarrow T \\
x &\mapsto \iota_B(x) = x
\end{aligned}$$

That is, ι_A and ι_B simply map elements of either A or B to the same element in T . Using these mappings we can see that

$$\begin{aligned}
a &= \prod_{i=1}^n p_i^{e_i} \\
&= \prod_{i=1}^n \iota_A(p_i)^{e_i} \\
&= \prod_{p_i \in A} p_i^{e_i} \\
&= \prod_{p_i \in A} p_i^{e_i} * \prod_{t_i \in T \setminus A} t_i^0 \\
&= \prod_{t_i \in T} t_i^{g_i} \text{ where } g_i = \begin{cases} e_i, & \text{If } t_i = p_i \\ 0, & \text{If } t_i \notin A \end{cases} \\
&= t_1^{g_1} t_2^{g_2} t_3^{g_3} \dots t_v^{g_v}
\end{aligned}$$

Likewise, for b we have

$$\begin{aligned}
b &= \prod_{j=1}^m q_j^{f_j} \\
&= \prod_{j=1}^m \iota_B(q_j)^{f_j} \\
&= \prod_{q_j \in B} q_j^{f_j} \\
&= \prod_{q_j \in B} q_j^{f_j} * \prod_{t_j \in T \setminus B} t_j^0 \\
&= \prod_{t_j \in T} t_j^{h_j} \text{ where } h_j = \begin{cases} f_j, & \text{If } t_j = q_j \\ 0, & \text{If } t_j \notin B \end{cases} \\
&= t_1^{h_1} t_2^{h_2} t_3^{h_3} \dots t_v^{h_v}
\end{aligned}$$

Hence, a and b have been expressed as a product of the same set of primes, where possibly one or more of the powers in either of the products could be zero. As required. \square

In other words, proposition 2.3.6 is saying that given the prime factorisations of a and b , we can always construct a new set containing the common primes of a and b and use this new set to express the factorisations of a and b . Why is this useful? It is useful because it will allow us to find the greatest common divisor of two integers by simply looking at the primes, and the powers of those primes, in common of those integers. We can use some examples to express this idea. The reader is encouraged to also try these examples using the Euclidean algorithm to verify.

Example 2.3.15. Let $a = 2 * 3^2 * 5 = 90$ and $b = 3 * 5^2 * 7 = 525$. We have that the $\text{GCD}(a, b) = 15$. Additionally, we know that $15 = 3 * 5$. The common primes of a and b are 3 and 5.

Example 2.3.16. Let $a = 5 * 11 = 55$ and $b = 2 * 7 = 14$. We have that the $\text{GCD}(a, b) = 1$. Moreover, a and b have no common primes.

Example 2.3.17. Let $a = 7 * 11 * 13 = 1001$ and $b = 7 * 11 * 17 = 1309$. We have that the $\text{GCD}(a, b) = 77$, as the primes in common are 7 and 11.

Example 2.3.18. Let $a = 2 * 3^4 = 162$ and $b = 3^3 * 5 = 135$. We have that the $\text{GCD}(a, b) = 27$, as the primes, with powers, in common is only 3^3 .

We show that looking at the primes in common is sufficient to get the greatest common divisor. To aid in the notation we make a definition

Definition 2.3.7. *The minimum function for integers*

Let $a, b \in \mathbb{Z}$. We define the minimum function, denoted $\min(a, b)$ by

$$\min : \mathbb{Z}^2 \rightarrow \mathbb{Z}$$

$$(a, b) \mapsto \min(a, b) = \begin{cases} a, & \text{If } a \leq b \\ b, & \text{If } b \leq a \end{cases}$$

Proposition 2.3.7. *Greatest common divisor from prime factorisation*

Let $a, b \in \mathbb{Z}$ with $b \neq 0$. By proposition 2.3.6 we know that there exists a set of primes

$$T = \{t_1, t_2, t_3, \dots, t_v\}$$

so that the prime factorisations of a and b are given by

$$a = \prod_{i=1}^v t_i^{e_i}$$

$$b = \prod_{i=1}^v t_i^{f_i}$$

We have that the greatest common divisor $\text{GCD}(a, b)$ is given by

$$\text{GCD}(a, b) = t_1^{\min(e_1, f_1)} t_2^{\min(e_2, f_2)} t_3^{\min(e_3, f_3)} \dots t_v^{\min(e_v, f_v)}$$

Proof:

Let $a, b \in \mathbb{Z}$ be as given as suppose that we have expressed a and b in accordance with proposition 2.3.6. This is to say we have a set T of primes so that

$$T = \{t_1, t_2, t_3, \dots, t_v\}$$

and the prime factorisations of a and b are given by

$$a = \prod_{i=1}^v t_i^{e_i}$$

$$b = \prod_{i=1}^v t_i^{f_i}$$

let $d = \text{GCD}(a, b)$ and let $D = t_1^{\min(e_1, f_1)} t_2^{\min(e_2, f_2)} t_3^{\min(e_3, f_3)} \dots t_v^{\min(e_v, f_v)}$. We need to show that $d = D$. To do so we show

1. $D \leq d$
2. $d \leq D$

Then the result follows from the fact that for $n, m \in \mathbb{Z}$ we have $n \leq m$ and $m \leq n$ we have $n = m$, for ease of notation, let $\sigma_i = \min(e_i, f_i)$ for $1 \leq i \leq v$.

1. $D \leq d$:

We have by definition of the minimum that $\sigma_i \leq e_i$ and $\sigma_i \leq f_i$. Hence $\exists k_i, l_i \in \mathbb{Z}$ so that

$$\begin{aligned} e_i &= \sigma_i + k_i \\ f_i &= \sigma_i + l_i \end{aligned}$$

Hence, we can express a as

$$\begin{aligned} a &= \prod_{i=1}^v t_i^{e_i} \\ &= \prod_{i=1}^v t_i^{\sigma_i + k_i} \\ &= \prod_{i=1}^v t_i^{\sigma_i} t_i^{k_i} \\ &= \prod_{i=1}^v t_i^{\sigma_i} \prod_{i=1}^v t_i^{k_i} \\ &= D * \prod_{i=1}^v t_i^{k_i} \end{aligned}$$

Therefore, as $\prod_{i=1}^v t_i^{k_i} \in \mathbb{Z}$ we have that $D \mid a$. A similar argument for b shows that $D \mid b$. Hence D is a common divisor of a and b and so by definition we have that $D \leq d$.

2. $d \leq D$:

To show that $d \leq D$ we will show that $d \mid D$ then by proposition 2.2.4 part 5. we will have that $d \leq D$. So suppose that $d \mid D$ then by the definition of divisibility we have that there is some $k \in \mathbb{Z}$ so that

$$d = Dk$$

As $k \in \mathbb{Z}$, it has a factorisation into primes by the fundamental theorem of arithmetic. Now, k could have primes in common with D , hence we can take those primes that are in common with D and k and place them into the factorisation of D , this is to say we have that

$$\begin{aligned} d &= Dk \\ d &= t_1^{\sigma_1} t_2^{\sigma_2} t_3^{\sigma_3} \dots t_v^{\sigma_v} k \\ d &= t_1^{\lambda_1} t_2^{\lambda_2} t_3^{\lambda_3} \dots t_v^{\lambda_v} k' \end{aligned}$$

Where we have that λ_i is the new value for each prime after we extract the primes that were in common with D and k , additionally, k' are the primes that were not in common.

To get the result we want we need to show two things.

$$(a) \ k' = 1$$

$$(b) \ \lambda_i \leq \sigma_i \text{ for all } 1 \leq i \leq v$$

(a) $k' = 1$:

Suppose for a contradiction that $k' \neq 1$. As $d > 0$ and $D > 0$ then we must have that $k > 0$ which means that $k' > 0$. Hence as $k' > 0$ we have by the fundamental theorem of arithmetic that k' has a factorisation into primes, say

$$k' = q_1^{r_1} q_2^{r_2} q_3^{r_3} \dots q_c^{r_c}$$

Moreover, no $q_j = t_i$ as k' has no common primes with $t_1^{\lambda_1} t_2^{\lambda_2} t_3^{\lambda_3} \dots t_v^{\lambda_v}$. Pick one of the primes in k' say $q = q_j$ then we have that $q \mid d$. Moreover we have that $d \mid a$ as $d = \text{GCD}(a, b)$ hence we must have that $q \mid a$. Hence we have that q is one of the primes t_i a contradiction. Therefore $k' = 1$.

(b) $\lambda_i \leq \sigma_i$ for all $1 \leq i \leq v$:

Suppose for a contradiction that $\lambda_i > \sigma_i$ for all $1 \leq i \leq v$. Without loss of generality take $i = 1$, for if this is not the case re-label the primes. Now by definition of σ_1 we have that $\sigma_1 = \min(e, f_1)$ and so we must have that either $\sigma_1 = e_1$ or $\sigma_1 = f_1$. Without loss of generality let $\sigma_1 = e_1$ as the case where $\sigma_1 = f_1$ is similar.

We, therefore, have that $\lambda_1 > e_1$. Now, as d is the greatest common divisor of a there is a $s \in \mathbb{Z}$ so that $ds = a$ where $s > 0$ as both a and d are. Now, comparing the prime factorisations of ds and a we have that

$$s * t_1^{\lambda_1} t_2^{\lambda_2} t_3^{\lambda_3} \dots t_v^{\lambda_v} = t_1^{e_1} t_2^{e_2} t_3^{e_3} \dots t_v^{e_v}$$

Dividing by $t_1^{e_1}$ we get that

$$s * t_1^{\lambda_1 - e_1} t_2^{\lambda_2} t_3^{\lambda_3} \dots t_v^{\lambda_v} = t_2^{e_2} t_3^{e_3} \dots t_v^{e_v}$$

Where clearly $t_1^{e_1 - e_1} = 1$. So this can be re-written as

$$s * t_1^{\lambda_1 - e_1} t_2^{\lambda_2} t_3^{\lambda_3} \dots t_v^{\lambda_v} = t_2^{e_2} t_3^{e_3} \dots t_v^{e_v}$$

As $\lambda_1 > e_1$, we have $\lambda_1 - e_1 > 0$. and so t_1 divides the left-hand side of the equation. But by the fundamental theorem of arithmetic if t_1 divides the left-hand side it must also divide the right-hand side and so would appear in the factorisation, but it is not in the factorisation of the right-hand side a contradiction. So $\lambda_i \leq \sigma_i$ for all $1 \leq i \leq v$.

Therefore $d \leq D$

As $D \leq d$ and $d \leq D$ we must have that $d = D$. As required. \square

Proposition 2.3.7 allows us to compute the greatest common divisor by considering the prime factorisations, rather than using the Euclidean algorithm. Unfortunately, we now have a new problem, how do we compute the prime factorisation of an integer? Thankfully to answer this question we have to answer the original question posed, what $x \in \mathbb{Z}$ are such that $x^2 = y$ for some $y \in \mathbb{Z}$? Clearly if $x \in \mathbb{Z}$ then x has some prime factorisation, say

$$x = p_1^{e_1} p_2^{e_2} p_3^{e_3} \dots p_k^{e_k}$$

So that

$$\begin{aligned} x^2 &= (p_1^{e_1} p_2^{e_2} p_3^{e_3} \dots p_k^{e_k}) (p_1^{e_1} p_2^{e_2} p_3^{e_3} \dots p_k^{e_k}) \\ &= (p_1^{e_1} p_1^{e_1}) (p_2^{e_2} p_2^{e_2}) (p_3^{e_3} p_3^{e_3}) \dots (p_k^{e_k} p_k^{e_k}) \\ &= p_1^{2e_1} p_2^{2e_2} p_3^{2e_3} \dots p_k^{2e_k} = y \end{aligned}$$

For each prime p_i , the power of that prime is now of the form $2e_i$ and therefore the power is even. We make this fact a definition.

Definition 2.3.8. *Square number*

Let $y \in \mathbb{Z}$ where $y > 0$, if there exists an $x \in \mathbb{Z}$ so that

$$x^2 = y$$

Then we say that y is a square number.

In light of the above discussion, we have the following result.

Proposition 2.3.8. *Square number if and only if prime factorisation has even powers*

Let $x \in \mathbb{Z}$. We have that x is a square number if and only if the prime factorisation of x only contains even prime powers. This is to say that each prime p_i in the factorisation of x has an exponent of the form $2e_i$.

Proof:

(\Rightarrow): Suppose that x is a square number, by definition there exists $y \in \mathbb{Z}$ so that $y^2 = x$. Let the prime factorisation of y be

$$y = q_1^{f_1} q_2^{f_2} q_3^{f_3} \dots q_k^{f_k}$$

We have that then

$$x = y^2 = q_1^{2f_1} q_2^{2f_2} q_3^{2f_3} \dots q_k^{2f_k}$$

Hence all the prime factors of x have an exponent of the form $2f_i$ making them even.

(\Leftarrow): Suppose that the prime factorisation of x has prime factors which only have even powers, that is

$$x = p_1^{e_1} p_2^{e_2} p_3^{e_3} \dots p_k^{e_k}$$

As each e_i is even we have that $\frac{e_i}{2} \in \mathbb{Z}$. Define y to be

$$y = p_1^{e_1/2} p_2^{e_2/2} p_3^{e_3/2} \dots p_k^{e_k/2}$$

Where clearly $y \in \mathbb{Z}$. We then have that

$$\begin{aligned} y^2 &= \left(p_1^{e_1/2} p_2^{e_2/2} p_3^{e_3/2} \dots p_k^{e_k/2} \right) \left(p_1^{e_1/2} p_2^{e_2/2} p_3^{e_3/2} \dots p_k^{e_k/2} \right) \\ &= \left(p_1^{e_1/2} p_1^{e_1/2} \right) \left(p_2^{e_2/2} p_2^{e_2/2} \right) \left(p_3^{e_3/2} p_3^{e_3/2} \right) \dots \left(p_k^{e_k/2} p_k^{e_k/2} \right) \\ &= p_1^{e_1} p_2^{e_2} p_3^{e_3} \dots p_k^{e_k} = x \end{aligned}$$

Hence as $x = y^2$ for some $y \in \mathbb{Z}$ we conclude that x is a square number. \square

We also have an immediate proposition.

Proposition 2.3.9. *Product of two square numbers is a square number*

Let $x, y \in \mathbb{Z}$ be square numbers. We have that xy is a square number.

Proof:

Let $x, y \in \mathbb{Z}$ be square numbers. We have by definition that $\exists a, b \in \mathbb{Z}$ so that

$$a^2 = x$$

$$b^2 = y$$

Now, consider the product xy , we have

$$xy = a^2 * b^2 = (ab)^2$$

Hence by definition, xy is a square number. \square

With proposition 2.3.8 we can finally answer the question of what $x \in \mathbb{Z}$ are such that $x^2 = y$ for some $y \in \mathbb{Z}$. It is those $x \in \mathbb{Z}$ so that x^2 is a square number! At first, this doesn't seem too useful as we can clearly take any $n \in \mathbb{Z}$ and see that $n^2 \in \mathbb{Z}$. However, the real meaning of this result is actually the converse, given some $n \in \mathbb{Z}$ we can see if there is an $x \in \mathbb{Z}$ so that $x^2 = n$. With this, we make a definition

Definition 2.3.9. *Square root function*

Let $x \in \mathbb{Z}$ be a positive square number. We define the square root function, denoted by $\sqrt{\cdot}$ as follows

$$\begin{aligned} \sqrt{\cdot} : \mathbb{Z} &\rightarrow \mathbb{Z} \\ x &\mapsto \sqrt{x} = \begin{cases} n, & \text{If } n^2 = x \\ \text{Undefined} & \text{otherwise} \end{cases} \end{aligned}$$

That is, we define the square root of an integer x to be the integer n that when squared gives x .

In light of this definition, we have the following result.

Proposition 2.3.10. *Square root of product is product of square roots*

Let $x, y \in \mathbb{Z}$ be square numbers. We have that

$$\sqrt{xy} = \sqrt{x}\sqrt{y}$$

Proof:

Let x, y be as given. By proposition 2.3.9 we have that xy is a square number and so \sqrt{xy} is well-defined. We need to show that

$$\sqrt{xy} = \sqrt{x}\sqrt{y}$$

By definition, we suppose that $\sqrt{xy} = n$, where $n^2 = xy$. Additionally, we can suppose that $\sqrt{x} = a$ where $a^2 = x$ and $\sqrt{y} = b$ where $b^2 = y$. Now, we have that

$$(\sqrt{x}\sqrt{y})^2 = (ab)^2 = a^2b^2 = xy = n^2 = (\sqrt{xy})^2$$

As $n^2 = a^2b^2$ we have that $n = ab$. Hence we have that $\sqrt{xy} = \sqrt{x}\sqrt{y}$ as required. \square

The idea of a square number actually generalises, meaning the question of what $x \in \mathbb{Z}$ are such that $x^2 = y$ for some $y \in \mathbb{Z}$ can be generalised to the question what $x \in \mathbb{Z}$ are such that $x^n = y$ for some $y \in \mathbb{Z}$ and every $n \in \mathbb{N}$.

The generalisation works very similarly to how we got to square numbers. As before let $x \in \mathbb{Z}$ which has a factorisation

$$x = p_1^{e_1} p_2^{e_2} p_3^{e_3} \cdots p_k^{e_k}$$

Now, consider x^n , the factorisation is given by

$$x^n = p_1^{ne_1} p_2^{ne_2} p_3^{ne_3} \cdots p_k^{ne_k}$$

Hence the power of each prime p_i is of the form ne_i . This is the defining characteristic for the next definition.

Definition 2.3.10. *n-th power number*

Let $y \in \mathbb{Z}$ with $y > 0$ and let $n \in \mathbb{N}$, if there exists an $x \in \mathbb{Z}$ so that

$$x^n = y$$

We say that y is the n -th power of x . We have already seen the case of $n = 2$ where y is called a square number. For $n = 3$ we call y a cube number. For $n > 4$, there is no formal term hence the definition using the terminology of n -th power.

The next step is to prove an equivalent proposition to 2.3.8.

Proposition 2.3.11. *n-th power number if and only if prime factorisation has multiples of n powers*

Let $x \in \mathbb{Z}$. We have that x is a n -th power number if and only if the prime factorisation of x only contains prime powers that are a multiple of n . this is to say that each prime p_i in the factorisation of x has an exponent of the form ne_i .

Proof:

(\Rightarrow): Suppose that x is a n -th power number, by definition there exists $y \in \mathbb{Z}$ so that $y^n = x$. Let the prime factorisation of y be

$$y = q_1^{f_1} q_2^{f_2} q_3^{f_3} \dots q_k^{f_k}$$

We have that then

$$x = y^n = q_1^{nf_1} q_2^{nf_2} q_3^{nf_3} \dots q_k^{nf_k}$$

Hence all the prime factors of x have an exponent of the form nf_i , meaning each prime power is a multiple of n .

(\Leftarrow): Suppose that the prime factorisation of x has prime factors which only have multiples of n , that is

$$x = p_1^{ne_1} p_2^{ne_2} p_3^{ne_3} \dots p_k^{ne_k}$$

As each e_i is a multiple of n we have that $\frac{e_i}{n} \in \mathbb{Z}$. Define y to be

$$y = p_1^{e_1/n} p_2^{e_2/n} p_3^{e_3/n} \dots p_k^{e_k/n}$$

Where clearly $y \in \mathbb{Z}$. We then have that

$$\begin{aligned} y^n &= \prod_{i=1}^n \left(p_1^{e_1/n} p_2^{e_2/n} p_3^{e_3/n} \dots p_k^{e_k/n} \right) \\ &= \prod_{j=1}^k \left(\prod_{i=1}^n \left(p_j^{e_j/n} \right) \right) \\ &= \prod_{j=1}^k \left(p_j^{e_j} \right) \\ &= p_1^{e_1} p_2^{e_2} p_3^{e_3} \dots p_k^{e_k} = x \end{aligned}$$

Hence as $x = y^n$ for some $y \in \mathbb{Z}$ we conclude that x is an n -th power number. \square

As before, there is an immediate proposition.

Proposition 2.3.12. *Product of two n – th power numbers is an n-th power number*

Let $x, y \in \mathbb{Z}$ be n -th power numbers. We have that xy is an n -th power number.

Proof:

Let $x, y \in \mathbb{Z}$ be n -th power numbers. By definition, we have that $\exists a, b \in \mathbb{Z}$ so that

$$\begin{aligned} a^n &= x \\ b^n &= y \end{aligned}$$

We have

$$xy = a^n * b^n = (ab)^n$$

Giving the result. \square

We can now now generalise the square root function.

Definition 2.3.11. *n-th root function*

Let $x \in \mathbb{Z}$ be a positive n -th power number. We define the n -th root function, denoted by $\sqrt[n]{}$ is given by

$$\begin{aligned} \sqrt[n]{} : \mathbb{Z} &\rightarrow \mathbb{Z} \\ x \mapsto \sqrt[n]{x} &= \begin{cases} m, & \text{If } m^n = x \\ \text{Undefined} & \text{otherwise} \end{cases} \end{aligned}$$

That is, we define the n -th root of an integer x to be the integer m that when raised to the power of n gives x .

2.4 The integers modulo n

Mathematicians call it “the arithmetic of congruences.” You can think of it as clock arithmetic

John Derbyshire

So far in the study of the divisibility of integers, we have considered what it means for an integer a to divide another b , namely we have that $a \mid b$ if there is some $c \in \mathbb{Z}$ such that $ac = b$. We now explore the implications of the case where $a \nmid b$, in particular, we look at the remainders from the division algorithm.

2.4.1 Remainders after division

Recall that for $a, b \in \mathbb{Z}$ we have that $a \mid b$ if $\exists c \in \mathbb{Z}$ so that $b = ac$. When this is not the case we have that $a \nmid b$. By the division algorithm, when $a \nmid b$ we have that $0 < r < |a|$, that is we have that

$$b = qa + r$$

The question is, what are the possible values for r ? The division algorithm gives us lower and upper bounds on the valid values of r but does not say anything about whether it can take all values in this range. Is only a small subset of this range valid? What happens if we allow b to be an arbitrary integer rather than some fixed integer? Some exploratory examples will be helpful.

Example 2.4.1. Let $a = 2$ and consider some $b > a$. We have by the division algorithm that

$$b = 2q + r$$

Hence r can only be one of 0 or 1. Now if $r = 0$ then we must have that b is an even number and if $r = 1$ we must have that b is an odd number. Then as b is an arbitrary integer we must have that dividing any integer by 2 will give us all of the possible remainders as an integer $x \in \mathbb{Z}$ is either even or odd.

Example 2.4.2. Let $a = 3$ and consider some $b > a$. By the division algorithm we have that r is either 0, 1 or 2. Like before, if $r = 0$ then b is a multiple of 3 so that $b = 3q$.

Now, suppose b is a multiple of 3. We have that $b + 3$ is also a multiple of 3 as

$$b + 3 = 3q + 3 \Rightarrow 3(q + 1)$$

So, as b is a multiple of 3 and $b + 3$ is a multiple of 3 then these will give a remainder $r = 0$ by the division algorithm. What can we say about $b + 1$ and $b + 2$? Using $b + 1$ in the division algorithm with 3 gives

$$b + 1 = 3q + 1$$

as $b = 3q$. Hence the remainder is 1. Likewise using $b + 2$ in the division algorithm with 3 gives a remainder of 2. As b was an arbitrary integer, we can conclude that the possible remainders when dividing an arbitrary integer by 3 are 0, 1 and 2. All of the possibilities are realised for the division of an arbitrary integer.

Example 2.4.3. Let $a = 4$ and consider some $b > a$. The division algorithm gives the possible range of remainders of 0, 1, 2 and 3. Like the previous example, we see that if the remainder is 0 then b is a multiple of 4, so similarly $b + 4$ is a multiple of 4. Looking at $b + 1$, $b + 2$ and $b + 3$ we see by the division algorithm that

$$b + 1 = 4q + 1$$

$$b + 2 = 4q + 2$$

$$b + 3 = 4q + 3$$

So dividing an arbitrary integer by 4 will give a remainder in the range 0 to 3 inclusive.

These examples suggest that when dividing an arbitrary integer b by some $a \in \mathbb{Z}$ with $a > 0$ will always give one value r with $0 < r < |a|$. This is true not just for the examples above but for every $a > 0$. We can prove this but first make an important observation.

Corollary 2.4.1. *Let $a, b \in \mathbb{Z}$ with $b > a$. Consider the division algorithm for b divided by a , that is we have*

$$b = qa + r$$

for some $q, r \in \mathbb{Z}$ and $0 < r < |a|$. We have that $a \mid (b - r)$.

Proof:

Let $a, b \in \mathbb{Z}$ be as given by the hypothesis. The division algorithm applied to a dividing b gives

$$b = qa + r$$

This gives $(b - r) = qa$ and so by the definition of divisibility we have that $a \mid (b - r)$. As required. \square

Corollary 2.4.1 is slightly misleading, this corollary provides the bedrock for the rest of this section.

2.4.2 Congruences and residues (Modular arithmetic)

We are now able to define the main topic behind this section. Corollary 2.4.1 tells us that when we divide b by a , then the difference between b and the remainder r is always divisible by a . This is to say $a \mid (b - r)$. Now suppose we have another integer c so that when c is divided by a the remainder is also r . We then also have $a \mid (c - r)$, in a sense b and c are similar when divided by a . That is they give the same remainder after division. We use this to make the definitions.

Definition 2.4.1. *Congruences, congruent number and residue number*

Let $a, b, n \in \mathbb{Z}$ so that $n > 0$. If we have that a and b have the same remainder when divided by n we say that a and b are congruent modulo n . This is denoted by

$$a \equiv b \pmod{n}$$

We call b a residue of a modulo n . We usually say that a is congruent to b modulo n . We define a congruence to capture the notion of congruent numbers and residue numbers. We call the number n the modulus of the congruence.

If a is not congruent to b , equivalently if b is not a residue of a we write $a \not\equiv b \pmod{n}$.

We can make use of corollary 2.4.1 to connect division to congruences.

Proposition 2.4.1. *Congruent if and only if the difference is divisible by modulus*

Let $a, b, n \in \mathbb{Z}$ and fix $n \geq 1$. We have that $a \equiv b \pmod{n}$ if and only if $n \mid (a - b)$.

Proof:

By the division algorithm, we have that

$$a = qn + r$$

$$b = q'n + r'$$

for some $q, q', r, r' \in \mathbb{Z}$ where $0 < r < |n|$ and $0 < r' < |n|$. Hence we have that

$$a - b = (q - q')n + (r - r')$$

where $-n < r - r' < n$.

(\Rightarrow): Suppose that $a \equiv b \pmod{n}$. By definition of congruences, we have that a and b share the same remainder when divided by n . Hence $r = r'$ and so $r - r' = 0$ so that

$$a - b = (q - q')n$$

which implies that $n \mid (a - b)$.

(\Leftarrow): Now suppose that $n \mid (a - b)$. We then have that

$$(a - b) - (q - q')n = (r - r')$$

Where $-n < r - r' < n$. The only integer strictly between $-n$ and n which is divisible by n is 0. Indeed if there were such a number between $-n$ and n which was divisible by n it would be a multiple of n so the inequality wouldn't be strict. Hence $r - r' = 0$ which implies that $r = r'$. So by the definition of a congruence, we have that $a \equiv b \pmod{n}$. \square

Proposition 2.4.1 gives us a bridge, allowing us to translate statements about divisibility into statements about congruences and visa versa. To get used to working with congruences we will use some examples.

Example 2.4.4. Suppose you were asked given that it is Monday, what would be the day in 100 days times? We can make use of the fact that days repeat in a 7-day cycle. By the division algorithm, we have that

$$100 = 14(7) + 2$$

That is to say, $100 \equiv 2 \pmod{7}$. So we know that if the current day is a Monday, then in 100 days times the day would be a Wednesday.

Example 2.4.5. The previous example can also be used to calculate some time in the future. Suppose we are using a 24-hour clock and we know that it is 1 PM, what would be the time in 164 hours?

To find this we make use of the fact that a 24-hour clock repeats every 24 hours. So we need to compute the remainder of 164 when divided by 24. The division algorithm gives

$$164 = 6(24) + 20$$

Hence $164 \equiv 20 \pmod{24}$. Now on a 24-hour clock, 1 PM is equal to 13. So the time 164 hours later will be given by $13 + 20 = 33$. We have a problem, 33 is not on a 24-hour clock! To find what time this is we need to find the remainder when 33 is divided by 24. We can quickly see that

$$33 = 1 * 24 + 9$$

So $33 \equiv 9 \pmod{24}$. Hence, 164 hours after 1 PM

Example 2.4.6. Let $a, n \in \mathbb{Z}$. We have that $n \mid (a - a)$ and so by proposition 2.4.1 that $a \equiv a \pmod{n}$.

Example 2.4.7. Let $a = 8$, $b = 11$ and $n = 5$. Using the division algorithm we can see that

$$8 \equiv 3 \pmod{5}$$

$$11 \equiv 1 \pmod{5}$$

Now, consider $a + b = 19$. By the division algorithm, we see that

$$19 = 3 * 5 + 4$$

So that $19 \equiv 4 \pmod{5}$. It would seem that we can add congruences together and the result makes sense.

The last two examples hint at some properties of congruences which should be investigated. In particular, $a \equiv a \pmod{n}$ is one criterion of being an equivalence relation. Do the other properties for being an equivalence relation hold? This is to say if $a \equiv b \pmod{n}$ do we have that $b \equiv a \pmod{n}$ and if $a \equiv b \pmod{n}$ and $b \equiv c \pmod{n}$ do we have $a \equiv c \pmod{n}$? As it turns out the answer is yes

Proposition 2.4.2. *Congruences are an equivalence relation*

Let $a, b, n \in \mathbb{Z}$ so that n is fixed and $n \geq 1$. Consider the relation \sim_n where

$$a \sim_n b \iff a \equiv b \pmod{n}$$

We have that \sim_n is an equivalence relation. That is

1. $a \equiv a \pmod{n}$
2. If $a \equiv b \pmod{n}$ then $b \equiv a \pmod{n}$
3. If $a \equiv b \pmod{n}$ and $b \equiv c \pmod{n}$ then $a \equiv c \pmod{n}$

Proof:

1. $a \equiv a \pmod{n}$:

As $n \mid (a - a)$ then by proposition 2.4.1 that $a \equiv a \pmod{n}$.

2. If $a \equiv b \pmod{n}$ then $b \equiv a \pmod{n}$:

Suppose that $a \equiv b \pmod{n}$ then $n \mid (a - b)$. By the definition of divisibility, we have that $\exists k \in \mathbb{Z}$ so that $a - b = kn$. Multiplying both sides by -1 gives $b - a = (-k)n$. So again by the definition of divisibility, we have that $n \mid (b - a)$ and so $b \equiv a \pmod{n}$.

3. If $a \equiv b \pmod{n}$ and $b \equiv c \pmod{n}$ then $a \equiv c \pmod{n}$:

Suppose that $a \equiv b \pmod{n}$ and $b \equiv c \pmod{n}$, then $n \mid (a - b)$ and $n \mid (b - c)$. Property 3. of proposition 2.2.4 then gives us that $n \mid ((a - b) + (b - c))$.

Clearly $(a - b) + (b - c) = a - c$ and so $n \mid (a - c)$ which is to say $a \equiv c \pmod{n}$.

As required. \square

We now know that congruences form an equivalence relation, one for each $n \geq 1$. So we can consider the equivalence classes that are formed by congruences. Let $a \in \mathbb{Z}$, what does the equivalence class $[a]$ look like?

Recall that $[a]_n$ is given by

$$[a]_n = \{x \in \mathbb{Z} : a \sim_n x\} = \{x \in \mathbb{Z} : a \equiv x \pmod{n}\}$$

That is, the equivalence class $[a]$ is a set of integers that are congruent to a modulo n . Equivalently, as \sim_n is an equivalence relation, we have that $a \equiv x \pmod{n}$ is the same as $x \equiv a \pmod{n}$. Hence we can view $[a]$ as the set of integers that x so that x gives a remainder of a when divided by n ¹³. This is to say

$$[a]_n = \{x \in \mathbb{Z} : x \equiv a \pmod{n}\}$$

For example, suppose that $a = 0$, then we have that

$$\begin{aligned} [0]_n &= \{x \in \mathbb{Z} : x \equiv 0 \pmod{n}\} \\ &= \{\dots, -2n, n, 0, n, 2n, \dots\} \end{aligned}$$

Likewise, when $a = 1$ we have that

¹³I prefer this way of thinking.

$$\begin{aligned}[1]_n &= \{x \in \mathbb{Z} : x \equiv 1 \pmod{n}\} \\ &= \{\dots, 1 - 2n, 1 - n, 1, 1 + n, 1 + 2n, \dots\}\end{aligned}$$

That is, the equivalence class of 0 modulo n is simply the multiples of n and the equivalence class of 1 modulo n are one more than a multiple of n .

How many congruence classes are there for a given n ? Clearly, by the definition algorithm, there are at most n such classes, We can show that there are exactly n classes.

Proposition 2.4.3. *Number of equivalence classes for congruence equivalence relation*

Let $a, b, n \in \mathbb{Z}$ so that $n \geq 1$ and consider the relation \sim_n given by

$$a \sim_n b \iff a \equiv b \pmod{n}$$

We have that there are n equivalence classes for the relation \sim_n , one for each possible remainder.

Proof:

By the division algorithm applied to n dividing a we have for $a, b, n \in \mathbb{Z}$ with $n \geq 1$ that

$$a = qn + r$$

for some unique $q, r \in \mathbb{Z}$ and $0 \leq r < |n|$. Hence the possible remainders are in the set

$$R = \{0, 1, 2, \dots, n - 1\}$$

as $n \geq 1$. Firstly we show that no two $i, j \in R$ are congruent modulo n . So suppose, WLOG, that $0 \leq i \leq j < n$, then $j - i > 0$ and $j - i < n$. Then we have that $n \nmid (j - i)$ and so $j \not\equiv i \pmod{n}$. As the choice of i, j was arbitrary we conclude that no two elements of R are congruent. It was shown in the proof of theorem 1.4.1 that unequal equivalence classes are disjoint, which is to say unique. Hence $[i]_n \neq [j]_n$ for all $i, j \in R$.

Now, it is left to show that any given $r \in R$ belongs to exactly one equivalence class. This is clear upon rewriting the result from the division algorithm as

$$a - r = qn$$

which gives that $a \equiv r \pmod{n}$. \square

It follows immediately by theorem 1.4.1 that the equivalence classes modulo n partition \mathbb{Z} . Additionally we have that $[a]_n = [b]_n$ if and only if $a \equiv b \pmod{n}$.

As we have shown that \sim_n is an equivalence relation we can define the quotient set, as in definition 1.5.3

Definition 2.4.2. *The integers modulo n*

Let $a \in \mathbb{Z}$. We define the quotient set \mathbb{Z}_n to be

$$\mathbb{Z}_n = \{[a]_n : a \in \mathbb{Z}\}$$

By proposition 2.4.3 we know there are n such sets which correspond to all the possible remainders when an integer a is divided by n . Hence we can explicitly write

$$\mathbb{Z}_n = \{[0]_n, [1]_n, [2]_n, \dots, [n - 1]_n\}$$

If we take the canonical representative of each class, for example, if the class is $[0]_n$ we take the canonical representative to be 0. We can write \mathbb{Z}_n more cleanly as

$$\mathbb{Z}_n = \{0, 1, 2, \dots, n - 1\}$$

As hinted by an example and because we have defined arithmetic on \mathbb{Z} , the next natural question is how does arithmetic work in \mathbb{Z}_n ? Recall the example, we had that $a = 8$, $b = 11$ and $n = 5$ and

$$\begin{aligned} 8 &\equiv 3 \pmod{5} \\ 11 &\equiv 1 \pmod{5} \end{aligned}$$

When we computed $a + b = 19$ and found that

$$19 \equiv 4 \pmod{5}$$

What about multiplication? We know that $8 * 8 = 64$ and we can see that $64 \equiv 4 \pmod{5}$. Multiplying the residue of 3 with itself we get $3 * 3 = 9$ from which we see $9 \equiv 4 \pmod{5}$. Similarly, we can see that subtraction makes sense in \mathbb{Z}_n . We know that $11 - 8 = 3$ so clearly $3 \equiv 3 \pmod{n}$. Subtracting the residues gives $1 - 3 = -2$. At first, this seems to be a problem, we seem to be saying that $3 \equiv -2 \pmod{5}$. However, a quick review of the definition of congruences tells us that is correct. We know that $a \equiv b \pmod{5}$ if and only if $n \mid (a - b)$, in our case we indeed have that $5 \mid (3 - (-2))$ as $3 - (-2) = 5$.

We can make the idea of addition, subtraction and multiplication rigorous.

Proposition 2.4.4. *Addition, subtraction and multiplication of congruences*

Let $a, b, c, d, n \in \mathbb{Z}$ so that $a \equiv b \pmod{n}$ and $c \equiv d \pmod{n}$. We have that

1. $(a + c) \equiv (b + d) \pmod{n}$
2. $(a - c) \equiv (b - d) \pmod{n}$
3. $(ac) \equiv (bd) \pmod{n}$

Proof:

Let $a, b, c, d, n \in \mathbb{Z}$ be as given by the hypothesis. As $a \equiv b \pmod{n}$ we have by proposition 2.4.1 that $n \mid (a - b)$ and so by the definition of divisibility we have that $a - b = kn$. Likewise, we have that as $c \equiv d \pmod{n}$ then by proposition 2.4.1 that $n \mid (c - d)$ and so by the definition of divisibility we have that $c - d = ln$.

In particular, we have that $a = b + kn$ and $c = d + ln$. It follows that

$$\begin{aligned} a + c &= (b + kn) + (d + ln) \\ &= (b + d) + (kn + ln) \\ &= (b + d) + n(k + l) \\ &\Rightarrow n \mid ((a + c) - (b + d)) \\ &\Rightarrow (a + c) \equiv (b + d) \pmod{n} \end{aligned}$$

Likewise, for subtraction we have

$$\begin{aligned} a - c &= (b + kn) - (d + ln) \\ &= (b - d) + (kn - ln) \\ &= (b - d) + n(k - l) \\ &\Rightarrow n \mid ((a - c) - (b - d)) \\ &\Rightarrow (a - c) \equiv (b - d) \pmod{n} \end{aligned}$$

Finally, for multiplication, we see that

$$\begin{aligned}
ac &= (b + kn)(d + ln) \\
&= bd + bln + dkn + kln^2 \\
&= bd + n(bl + dk + kln) \\
&\Rightarrow n \mid ((ac) - (bd)) \\
&\Rightarrow (ac) \equiv (bd) \pmod{n}
\end{aligned}$$

As required. \square

This proposition provides the backbone of showing that the operations of addition, subtraction and multiplication are well-defined on \mathbb{Z}_n .

Definition 2.4.3. *Addition, subtraction and multiplication on \mathbb{Z}_n*

Let $a, b, n \in \mathbb{Z}$ with $n \geq 1$. We define addition, subtraction and multiplication on \mathbb{Z}_n by

1. $[a]_n + [b]_n = [a + b]_n$
2. $[a]_n - [b]_n = [a - b]_n$
3. $[a]_n [b]_n = [ab]_n$

We prove these are well-defined.

Proposition 2.4.5. *Addition, subtraction and multiplication on \mathbb{Z}_n is well-defined and closed*

Let $n \in \mathbb{Z}$ so that $n \geq 1$. We have that addition, subtraction and multiplication of equivalence classes are well-defined and closed. This is to say $\forall x, y \in \mathbb{Z}_n$ we have that

1. $[x]_n + [y]_n = [x + y]_n \in \mathbb{Z}_n$
2. $[x]_n - [y]_n = [x - y]_n \in \mathbb{Z}_n$
3. $[x]_n [y]_n = [xy]_n \in \mathbb{Z}_n$

Proof:

Suppose that $a \in [x]_n$ and $b \in [y]_n$. By definition, we have that $a \equiv x \pmod{n}$ and $b \equiv y \pmod{n}$.

By proposition 2.4.4 we have that

1. $a + b \equiv x + y \pmod{n}$
2. $a - b \equiv x - y \pmod{n}$
3. $ab \equiv xy \pmod{n}$

So that $a + b \in [x + y]_n$, $a - b \in [x - y]_n$ and $ab \in [xy]_n$, showing the operations are well-defined. Closure is immediate in each case. \square

We now have a well-defined idea of arithmetic on \mathbb{Z}_n . A poor student or a particularly clever dog will realise immediately that we have missed out on some operations that were defined on \mathbb{Z} . In this section for example we defined what integer division means. What about exponentiation?

We will first look at exponentiation. Thankfully there isn't much work to do as we can make use of the definition of multiplication for \mathbb{Z}_n . We can see that

$$\begin{aligned}
([a])^2 &= [a] [a] = [a * a] = [a^2] \\
([a])^3 &= [a]^2 [a] = [a^2 * a] = [a^3] \\
([a])^4 &= [a]^3 [a] = [a^3 * a] = [a^4] \\
&\dots
\end{aligned}$$

So clearly exponentiation is well-defined. Now, what about division? We expect that to get a well-defined definition for division in \mathbb{Z}_n it should respect the definition of divisibility for the integers. Here in lies the problem, division over \mathbb{Z} is not well-defined, for example, $3 \nmid 2$, so it is clear there is no equivalence class for this case. What about the cases where division over \mathbb{Z} is well-defined? This is our definition of being congruent so we can't extend to division of congruences this way either.

However, recall that we have defined the idea of a multiplicative inverse. In particular, we had that for $x \in \mathbb{Z}$ such that $x \neq 0$, then $y \in \mathbb{Q}$ was said to be a multiplicative inverse of x so that

$$x * y = 1 = y * x$$

Perhaps then, we might hope to recover some notion of division modulo n by using multiplicative inverses. Such a definition, of course, would have to respect congruences. So for $x \in \mathbb{Z}_n$ with $x \not\equiv 0 \pmod{n}$, we are looking for $y \in \mathbb{Z}_n$ so that $x * y \equiv 1 \pmod{n}$. To start, it would be wise to look at multiplication for a few small values of $n \geq 2$, to get a feel for what we are looking for.

*	0	1
0	0	0
1	0	1

Table 8: The multiplication table for $n = 2$

*	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

Table 9: The multiplication table for $n = 3$

*	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

Table 10: The multiplication table for $n = 4$

*	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Table 11: The multiplication table for $n = 5$

*	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

Table 12: The multiplication table for $n = 6$

*	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6
2	0	2	4	6	1	3	5
3	0	3	6	2	5	1	4
4	0	4	1	5	2	6	3
5	0	5	3	1	6	4	2
6	0	6	5	4	3	2	1

Table 13: The multiplication table for $n = 7$

What do these tables tell us? Starting with the case $n = 2$, we see that only $1 \equiv 1 \pmod{2}$ has a multiplicative inverse, namely $1 \equiv 1 \pmod{2}$. For $n = 3$, we see that if $x \equiv 1 \pmod{3}$ then we can take $y \equiv 1 \pmod{3}$ and likewise if $x \equiv 2 \pmod{3}$ then we can take $y \equiv 2 \pmod{3}$.

Things get a little more complicated for $n = 4$. We see that if $x \equiv 1 \pmod{4}$ then we take $y \equiv 1 \pmod{4}$ and if $x \equiv 3 \pmod{4}$ we take $y \equiv 3 \pmod{4}$. What about $x \equiv 2 \pmod{4}$? Looking at the table we see that $x * 1 \equiv 2 \pmod{4}$ and $x * 3 \equiv 2 \pmod{4}$, finally $x * 2 \equiv 0 \pmod{4}$. A disaster! We have that 2 does not have a multiplicative inverse modulo 4. Hence not all elements of \mathbb{Z}_4 have a multiplicative inverse. A similar situation occurs for the case $n = 6$, for example, the row for $x \equiv 3 \pmod{6}$ shows only 3 and 0 can be results.

So our quest of being able to define some notion of division for \mathbb{Z}_n in general appears to be at an end.

That being said, the situation looks more promising in the cases of $n = 5$ and $n = 7$. For \mathbb{Z}_5 we have that the following multiplicative inverses, for those $x \not\equiv 0 \pmod{5}$

x	x^{-1}
1	1
2	3
3	2
4	4

Table 14: The elements $x \not\equiv 0 \pmod{5}$ and their respective multiplicative inverses

Likewise for the elements $x \not\equiv 0 \pmod{7}$ for \mathbb{Z}_7 we have

x	x^{-1}
1	1
2	4
3	5
4	2
5	3
6	6

Table 15: The elements $x \not\equiv 0 \pmod{7}$ and their respective multiplicative inverses

We saw similar situations for \mathbb{Z}_2 and \mathbb{Z}_3 , so what do $\mathbb{Z}_2, \mathbb{Z}_3, \mathbb{Z}_5, \mathbb{Z}_7$ have in common? The thing they have in common is that the modulus is a prime! Does this result hold for all primes? If so, why? If not, why not and what primes does it fail for?

We also saw cases in \mathbb{Z}_4 and \mathbb{Z}_6 where certain elements did have a multiplicative inverse. For example in \mathbb{Z}_4 we saw $x \equiv 1 \pmod{4}$ had the multiplicative inverse of 1, similarly we saw $x \equiv 3 \pmod{4}$ had a multiplication inverse of 3. In \mathbb{Z}_6 we can see that 1 has the inverse of 1, and 5 has an inverse of 5. So what is special in the case where n is not prime that allows some elements to have an inverse?

In the case of \mathbb{Z}_4 the elements which had an inverse, 1 and 3 are co-prime to 4. Likewise in \mathbb{Z}_6 the elements that had inverse were 1 and 5 which are again co-prime to 6. When n was prime, we make the trivial observation that all non-zero elements of \mathbb{Z}_n are co-prime to n , for if not then they share a common prime factor and hence the greatest common divisor would be larger than 1. It seems we have recovered our original goal, that is to say, it looks like it is the case that an element of $x \in \mathbb{Z}_n$ for $n \geq 2$ has a multiplicative inverse if $\text{GCD}(x, n) = 1$. Clearly, this is an if and only-if statement.

Proposition 2.4.6. *Existence of inverse element in \mathbb{Z}_n*

Let $n \in \mathbb{Z}$ with $n \geq 2$. Let $x \in \mathbb{Z}_n$. The multiplicative inverse of x in \mathbb{Z}_n exist if and only if $\text{GCD}(x, n) = 1$

Proof:

(\Rightarrow): Let $x \in \mathbb{Z}_n$ have an inverse $y \in \mathbb{Z}_n$. We therefore have that

$$xy \equiv 1 \pmod{n}$$

By the definition of congruences, we therefore have that $xy = 1 + kn$ for some $k \in \mathbb{Z}$. Let $d = \text{GCD}(x, n)$. As d is the greatest common divisor of x and n we have that $d \mid x$ and $d \mid n$ so $d \mid xy - kn$. But $xy - kn = 1$ so $d \mid 1$. Clearly $\text{GCD}(x, n) \geq 1$ and so we conclude that $d = 1$.

(\Leftarrow): Suppose that $\text{GCD}(x, n) = 1$. We have by Bézout's Identity (theorem 2.2.5) that $\exists a, b \in \mathbb{Z}$ so that

$$ax + bn = 1$$

Modulo n , we get that $ax \equiv 1 \pmod{n}$ and so a is the inverse element of x in \mathbb{Z}_n .

As required. \square

Corollary 2.4.2. *All non-zero elements of \mathbb{Z}_n exist if n is prime*

Let p be prime. We have that all the non-zero elements of \mathbb{Z}_p have a multiplicative inverse.

Proof:

By corollary 2.3.1, if p is a prime and $p \nmid a$ for some $a \in \mathbb{Z}$ then $\text{GCD}(a, p) = 1$. Now suppose that $x \in \mathbb{Z}_p$, clearly $x \leq p$. In particular $p \nmid x$. As this is true for every non-zero $x \in \mathbb{Z}_p$ then $\text{GCD}(x, p) = 1$ and so each $x \in \mathbb{Z}_p$ has a multiplicative inverse by proposition 2.4.6. \square

We have now recovered a definition of division of the congruence classes of \mathbb{Z}_n . Now that modular arithmetic is on a solid footing, what can we use it for? One immediate use case is solving problems about divisibility.

Example 2.4.8. *We will show that $6 \mid a(a+1)(a+2)$ for every integer a . We observe that the possible residues of a modulo 6 are 0, 1, 2, 3, 4 and 5. It is enough to check that each is congruent to zero modulo 6.*

When $a \equiv 0 \pmod{6}$ we see that $a+1 \equiv 1 \pmod{6}$ and $a+2 \equiv 2 \pmod{6}$. So that

$$a(a+1)(a+2) \equiv 0 * 1 * 1 \equiv 0 \pmod{6}$$

Now, when $a \equiv 1 \pmod{6}$ we see that $a+1 \equiv 2 \pmod{6}$ and $a+2 \equiv 3 \pmod{6}$, giving

$$a(a+1)(a+2) \equiv 1 * 2 * 3 \equiv 0 \pmod{6}$$

*As $1 * 2 * 3 = 6$ which is congruent to zero modulo 6. We see that, with an abuse of notation for brevity, that*

a	$a \pmod{6}$	$(a+1) \pmod{6}$	$(a+2) \pmod{6}$	$a(a+1)(a+2) \pmod{6}$
2	2	3	4	$24 \equiv 0$
3	3	4	5	$60 \equiv 0$
4	4	5	0	$0 \equiv 0$
5	5	0	1	$0 \equiv 0$

Table 16: The residues of $a \pmod{6}$ for $a \geq 2$, the values of each term and their resultant multiplication modulo 6

We can see that the product is always zero modulo 6. As each product is always congruent to zero modulo 6 then $a(a+1)(a+2) \equiv 0 \pmod{6}$ which implies $6 \mid a(a+1)(a+2)$.

The astute reader may notice that this feels longer than a proof that uses only the definition of divisibility. The astute reader would be correct. In fact we have that $6 \mid m$ for some $m \in \mathbb{Z}$ if and only if $2 \mid m$ and $3 \mid m$.

Indeed, suppose that $6 \mid m$ then $m = 6n$ for some $n \in \mathbb{Z}$, moreover $6 = 2 * 3$ so $m = 2 * 3 * n$ which implies that $2 \mid m$ and $3 \mid m$. Conversely, if $2 \mid m$ and $3 \mid m$ then 6 clearly divides m as 2 and 3 will appear at least once in the prime factorisation of m . So why did we bother with congruences? By first doing the longer calculations and then the shorter proof, we have seen a hint at a possible generalisation to the theory!

That is, if $a \equiv b \pmod{n}$ for some $n \in \mathbb{Z}$ with $n > 0$ with a prime factorisation

$$n = p_1^{e_1} p_2^{e_2} p_3^{e_3} \dots p_k^{e_k}$$

then we might expect that $a \equiv b \pmod{n}$ if and only if $a \equiv b \pmod{p_i^{e_i}}$ where $i = 1, 2, \dots, k$. We can prove this.

Proposition 2.4.7. *Congruent if and only if congruent to each prime in factorisation*

Let $n \in \mathbb{Z}$ so that $n > 0$ and n has a prime factorisation given by

$$n = p_1^{e_1} p_2^{e_2} p_3^{e_3} \dots p_k^{e_k}$$

Let $a, b \in \mathbb{Z}$. We have that $a \equiv b \pmod{n}$ if and only if $a \equiv b \pmod{p_i^{e_i}}$ for each i where $i = 1, 2, \dots, k$

Proof:

Let $n \in \mathbb{Z}$ be as given in the hypothesis. We have that

$$\begin{aligned}
a \equiv b \pmod{n} &\iff n \mid (a - b) \\
&\iff p_1^{e_1} p_2^{e_2} p_3^{e_3} \dots p_k^{e_k} \mid (a - b) \\
&\iff p_i^{e_i} \mid (a - b), \text{ For each } i = 1, 2, \dots, k \\
&\iff a \equiv b \pmod{p_i^{e_i}}, \text{ For each } i = 1, 2, \dots, k
\end{aligned}$$

As required. \square

Another use of congruences is in cryptography, which is a field of study of taking messages and encoding (obfuscating) them in such a way that only the person the message was intended for can read it. This is especially true for the RSA¹⁴ encryption method. We already have some of the mathematical machinery required to explore how this method of cryptography works, namely prime numbers and congruences. On the other hand, we still lack some important theory. If cryptography is the field of encoding messages so that only the person the message was intended for can read it, then there is some method that encodes the message and a method that decodes the message using some information known to both the sender and recipient. This means that using this information the recipient will have some method of finding out the original message! We look at this idea in more detail.

¹⁴RSA stands for Rivest–Shamir–Adleman

2.5 Diophantine equations and Polynomials

I had a Polynomial once. My Doctor removed it.

Micheal Grant

We start with a definition that we have seen numerous times so far but have not formally defined. That of an equation.

Definition 2.5.1. *Equation*

An equation is a mathematical statement that states that two expressions are equal.

This seems simple enough, but what does it mean? Unfortunately, this depends on the situation, different situations will have a different meaning of what a statement is. Thankfully, we have seen equations already throughout the text so this abstract definition is familiar to us. For example

$$1 + 1 = 2$$

is an equation. So is $\text{GCD}(a, b) = d$, in a similar vain we have from Bézout's Identity that $d = ax + by$ is also an equation. So why define something if it is really this simple? Simply put, we can use the idea of an equation in a more complex way. For example,

$$1 + x = 2$$

says that 1 plus x is equal to 2 but we don't know what x is. However, we can see that

$$\begin{aligned} 1 + x &= 2 \\ x &= 1 \end{aligned}$$

That is, we see that $x = 1$, this is an equation! This is where the power of an equation starts to show its worth. If we have a problem where we don't know the value of some quantity of interest, we might be able to work out what that quantity is. We have seen more complex examples of equations, for example $x^2 = 2$ which we have shown has no value of $x \in \mathbb{Q}$ where it is true.

Hence, equations that contain a value, or maybe multiple values that we don't know but want to know, are important. This section is focused on looking at such equations. We make another couple of definitions for when an equation contains a value we don't know.

Definition 2.5.2. *Variable*

A variable is a value that is allowed to be changed either freely or restricted by some constraint or equation. A variable can be taken to be any meaningful value, either inside or outside of some set S . The context of the statement under study usually makes it clear where the variable belongs.

Definition 2.5.3. *Indeterminate variable*

An indeterminate variable is a variable value which has not been specified. As with a variable, it could be inside or outside of some set S .

Definition 2.5.4. *An unknown variable*

An unknown variable, or simply an unknown, is a variable whose value is unknown but we wish to find its value. As before, this unknown variable is to be taken as a member of a set S . If a value for the unknown variable can be found, we call it a solution to the equation.

For example, the equation $5x + 1 = 2$ would have x as the indeterminate variable, if we were solving for x then x would be the unknown variable as well. The equation $2x + 5y = 6$ has two indeterminate variables, x and y . We can potentially have many indeterminate variables in an equation. Moreover, in many problems, we will have a certain type of variable whose value can vary but is not the unknown that we are looking to solve for. We define this type of variable as well.

Definition 2.5.5. *Coefficient*

A variable which can vary but is not the variable that is being solved for is called a coefficient, or a parameter of the equation.

So, let's start simply and consider the simplest equation possible with one unknown variable and two coefficients.

$$x + a = b$$

This is simple to solve for the unknown x , simply take a from both sides to give $x = b - a$. So for example if we let $a, b \in \mathbb{Z}$ say with $a = 5$ and $b = 3$, then we see that $x \in \mathbb{Z}$ with $x = 3 - 5 = -2$. This is also true if we take $a, b \in \mathbb{Q}$. A more complex form of the above equation is

$$ax + b = c$$

Now we hit a problem we are looking for a solution $x \in \mathbb{Z}$. Firstly, we have that $ax = c - b$, but then a solution $x \in \mathbb{Z}$ can occur if and only if $a \mid (c - b)$. If we look for a solution where $x \in \mathbb{Q}$ then no such problem occurs. Therefore, the set that we are looking for solutions in is crucial in solving equations. With our current theory, the situation gets more hopeless the more complicated the equation becomes. For example, if we consider the equation

$$4x^2 + 2x + 3 = 0$$

Does this equation have solutions in \mathbb{Z} ? How about \mathbb{Q} ?. Additionally, what happens if we have more than one equation or unknowns? For example, consider the two equations given by

$$\begin{aligned} 4x + 2y &= 6 \\ -2x + 5y &= 7 \end{aligned}$$

How do we solve equations like this? This section aims to answer questions like these. We make a final definition, a special case for when we only seek integer solutions.

Definition 2.5.6. *Diophantine equation*

An equation for which the solutions have to be integers is called a Diophantine equation¹⁵.

2.5.1 Linear Diophantine equations**2.5.1.1 Linear equations with two variables**

We start where the previous section left off, by looking at the simplest type of equation that can be solved.

Definition 2.5.7. *Linear equation of a single indeterminate variable*

Let S be a set. We say an equation is a linear equation in a single variable x if it has the form

$$ax + b = c$$

for some coefficients $a, b, c \in S$ and an indeterminate variable x . In particular as this equation only has one indeterminate variable we say it is a single-variable linear equation.

We have already seen that solutions to this equation exist in \mathbb{Z} if and only if $a \mid (c - b)$, and a solution always exists if we want $x \in \mathbb{Q}$. Things are a bit more interesting if we introduce a second variable.

Definition 2.5.8. *Linear equation of two indeterminate variables*

Let S be a set. We say an equation is a linear equation in two variables x, y if it has the form

$$ax + by = c$$

for some coefficients $a, b, c \in S$ and indeterminate variables x and y .

¹⁵Named after the 3rd-century mathematician Diophantus of Alexandria

We have seen this type of equation before, in Bézout's Identity (Theorem 2.2.5). In Bézout's Identity, we have that the greatest common divisor, d , of two integers a, b can be expressed as

$$ax + by = d$$

for some $x, y \in \mathbb{Z}$. This gives us examples of already solved equations, but what about the other way? Given an equation of the form

$$ax + by = c$$

with $a, b, c \in \mathbb{Z}$ given, can we find integer values for x and y ? That is, we are considering $ax + by = c$ to be a Diophantine equation. If the reader is sufficiently alert, they will notice that by mentioning Bézout's Identity we are hinting that it will be crucial to finding the solutions.

We know of one solution, namely if $\text{GCD}(a, b) = d$ and $c = d$ then the solution is found by the Euclidean algorithm. Now if c were a multiple of d can we find solutions? Recall proposition 2.2.10 part 4. We have that $\text{GCD}(a, b) = d$ is the smallest such so that $ax + by = d$, given that this is the smallest such then we can show that there exist others, namely these solutions are multiples of d .

Proposition 2.5.1. *Integer has form $ax + by$ if it is a multiple of the greatest common divisor of a and b . Let $a, b \in \mathbb{Z}$ and $d = \text{GCD}(a, b)$. Let $c \in \mathbb{Z}$. We have that*

$$c = ax + by$$

if and only if $d \mid c$. Which is to say c is a multiple of d

Proof:

(\Rightarrow): Clearly if $c = ax + by$ then as $d = \text{GCD}(a, b)$ we have by proposition 2.2.4 part 3 that $d \mid c$.

(\Leftarrow): Suppose that $c = de$ for some $e \in \mathbb{Z}$. By Bézout's Identity, we have that $\exists u, v \in \mathbb{Z}$ so that

$$d = au + bv$$

where $d = \text{GCD}(a, b)$. Multiplying both sides by e we get

$$c = aue + bve = ax + by$$

Hence $x = ue$ and $y = ve$.

As required. \square

Armed with this proposition we can find the solutions to the Diophantine equation $ax + by = c$.

Proposition 2.5.2. *Solutions to the Diophantine equation $ax + by = c$*

Let $a, b, c \in \mathbb{Z}$ be such that

$$ax + by = c$$

for the indeterminate variables x, y and let $d = \text{GCD}(a, b)$. We have that there are solutions so that $x, y \in \mathbb{Z}$ if and only if $d \mid c$.

Moreover, there are infinitely many solutions where the solutions are given by

$$\begin{aligned} x &= x_0 + \frac{bn}{d} \\ y &= y_0 - \frac{an}{d} \end{aligned}$$

where $x_0, y_0 \in \mathbb{Z}$ is one solution.

Proof:

The existence of a solution is given by proposition 2.5.1. It is left to show that the suggested solutions x, y are solutions and that there are infinitely many solutions. This follows the argument in example 2.3.9. We give the argument again to refresh the reader's memory.

Let $x_0, y_0 \in \mathbb{Z}$ be a solution, then we have that

$$ax_0 + by_0 = c$$

For any $n \in \mathbb{Z}$ let

$$\begin{aligned} x &= x_0 + \frac{bn}{d} \\ y &= y_0 - \frac{an}{d} \end{aligned}$$

We then have that $\frac{bn}{d} \in \mathbb{Z}$ as $d \mid b$ by definition of the greatest common divisor, likewise for $\frac{an}{d}$. Hence, we have that

$$\begin{aligned} ax + by &= a \left(x_0 + \frac{bn}{d} \right) + b \left(y_0 - \frac{an}{d} \right) \\ &= ax_0 + a \frac{bn}{d} + by_0 - b \frac{an}{d} \\ &= ax_0 + \frac{abn}{d} + by_0 - \frac{abn}{d} \\ &= ax_0 + by_0 = c \end{aligned}$$

Hence x, y is a solution. Moreover, as $n \in \mathbb{Z}$ is any integer we have shown that there are infinitely many solutions. It is left to show that these are the only solutions.

Let $x, y \in \mathbb{Z}$ be any solution to $ax + by = c$, and let $x_0, y_0 \in \mathbb{Z}$ be a particular solution. Hence

$$ax + by = ax_0 + by_0$$

Subtracting $ax_0 + by_0$ from the right-hand side gives

$$\begin{aligned} ax + by - ax_0 - by_0 &= 0 \\ a(x - x_0) + b(y - y_0) &= 0 \end{aligned}$$

Now, as $d = \text{GCD}(a, b)$ then we have that $d \mid a$ and $d \mid b$ so that

$$\begin{aligned} \frac{a}{d}(x - x_0) + \frac{b}{d}(y - y_0) &= 0 \\ \frac{a}{d}(x - x_0) &= -\frac{b}{d}(y - y_0) \end{aligned}$$

If $a = b = 0$, we are done so suppose not. Then one of a or b is non-zero. Without loss of generality, suppose that $a \neq 0$. We have that by proposition 2.2.10 that if $\text{GCD}(a, b) = d$ then $\text{GCD}\left(\frac{a}{d}, \frac{b}{d}\right) = 1$, moreover by definition of co-prime integers we have that $\frac{a}{d}$ and $\frac{b}{d}$ are co-prime.

By Euclid's lemma for co-primes (lemma 2.3.3) we have that $\frac{a}{d} \mid -(y - y_0)$. Hence there is some $n \in \mathbb{Z}$ so that

$$-(y - y_0) = n \frac{a}{d}$$

Which is to say

$$y = y_0 - \frac{an}{d}$$

Similarly, we have that

$$x = x_0 + \frac{bn}{d}$$

As required. \square

2.5.1.2 Linear equations with more than two variables

A natural question to ask now is what happens when we have more than two indeterminate variables? For example $ax + by + cz = e$? We can take some inspiration from the two variable case.

Recall that for $ax + by = c$ with $d = \text{GCD}(a, b)$ that there are solutions with $x, y \in \mathbb{Z}$ if and only if $d \mid c$. More importantly, we have that if $d = \text{GCD}(a, b)$ then we can express d by $d = ax + by$ for some $x, y \in \mathbb{Z}$ by Bézout's Identity. Moreover by proposition 2.2.5 we have that for a set of n integers $S = \{b_1, b_2, b_3, \dots, b_n\}$ and additionally we have that that $a \mid b_i$ for each $b_i \in S$ then

$$a \mid \sum_{i=1}^n b_i x_i$$

This hints at an extension to Bézout's Identity, given a suitable extension to the definition of the greatest common divisor for more than two inputs. Hence, our goal is to build this suitable extension to the greatest common divisor. We will start by looking at some exploratory examples before moving on with the generalisation.

Example 2.5.1. Let $a = 2, b = 4$ and $c = 6$. What is $\text{GCD}(a, b, c)$? Clearly, by inspection, we have that 2 is the largest divisor of a, b and c . In particular we have that $\text{GCD}(2, 4) = 2$ and $\text{GCD}(2, 6) = 2$. In other words, we have that

$$\text{GCD}(2, 4, 6) = \text{GCD}(\text{GCD}(2, 4), 6)$$

Equivalently, we could have first considered $\text{GCD}(4, 6) = 2$ and then $\text{GCD}(2, 2) = 2$ so we have

$$\text{GCD}(2, 4, 6) = \text{GCD}(2, \text{GCD}(4, 6))$$

Example 2.5.2. Let $a = 3, b = 6$ and $c = 30$. What is $\text{GCD}(a, b, c)$? Breaking this problem down we have that $\text{GCD}(3, 6) = 3$, $\text{GCD}(3, 30) = 3$ and $\text{GCD}(6, 30) = 6$. As the greatest common divisor must divide all of the numbers we must conclude that $\text{GCD}(3, 6, 30) = 3$.

Example 2.5.3. Let $a = 3, b = 5$ and $c = 7$. As a, b and c are all prime we clearly see that $\text{GCD}(a, b, c) = 1$

Example 2.5.4. Let $a = 14, b = 35, c = 7$ and $d = 5$. We again break this down. We see that

$$\text{GCD}(14, 33) = 7$$

$$\text{GCD}(14, 7) = 7$$

$$\text{GCD}(14, 5) = 1$$

$$\text{GCD}(35, 7) = 7$$

$$\text{GCD}(35, 5) = 7$$

$$\text{GCD}(7, 5) = 1$$

Again the greatest common divisor is the smallest value that divides all of the inputs a, b, c and d . The smallest such number here is 1 so $\text{GCD}(14, 35, 7, 5) = 1$.

In these examples, we made use of the fact that the greatest common divisor of two numbers is the smallest number that divides both of the input numbers. We then looked at all of the possible combinations of the inputs and took the smallest value that occurred. This is to be consistent with two variable version of the GCD that we have already developed. This was shown explicitly in the first example with

$$\text{GCD}(2, 4, 6) = \text{GCD}(\text{GCD}(2, 4), 6) = \text{GCD}(2, \text{GCD}(4, 6))$$

Hence an immediate property that we can deduce is that the GCD is associative, in the sense that computing the GCD of three numbers is equivalent to computing the GCD of two of the inputs with the remaining input.

Proposition 2.5.3. *GCD is associative*

Let $a, b, c \in \mathbb{Z}$. We have that

$$\text{GCD}(a, \text{GCD}(b, c)) = \text{GCD}(\text{GCD}(a, b), c)$$

Proof:

Let $x = \text{GCD}(a, \text{GCD}(b, c))$ and $y = \text{GCD}(\text{GCD}(a, b), c)$. We need to show that $x \mid y$ and $y \mid x$ then we can conclude that $x = y$.

As $x = \text{GCD}(a, \text{GCD}(b, c))$ then by definition of the greatest common divisor, we have that $x \mid a$ and $x \mid \text{GCD}(b, c)$. Moreover as $x \mid \text{GCD}(b, c)$ then again by definition of the greatest common divisor we have that $x \mid b$ and $x \mid c$.

As $x \mid a$ and $x \mid b$ then $x \mid \text{GCD}(a, b)$ and likewise $x \mid c$ so $x \mid \text{GCD}(\text{GCD}(a, b), c)$ by definition and so $x \mid y$. The proof that $y \mid x$ is similar.

As $x \mid y$ and $y \mid x$ and $x > 0$ and $y > 0$ we conclude that $x = y$ as required. \square

To extend our definition of the greatest common divisor to more than two inputs, we will use the definition of the GCD given by the decomposition of primes. That is to say, given $a, b \in \mathbb{Z}$, we know that there exists a set of primes

$$T = \{t_1, t_2, \dots, t_v\}$$

So that a and b can be represented by a prime factorisation of primes $t_i \in T$. That is

$$a = \prod_{i=1}^v t_i^{e_i}$$

$$b = \prod_{i=1}^v t_i^{f_i}$$

We then have that the greatest common divisor is given by

$$\text{GCD}(a, b) = t_1^{\min(e_1, f_1)} t_2^{\min(e_2, f_2)} t_3^{\min(e_3, f_3)} \dots t_v^{\min(e_v, f_v)}$$

Firstly, we will extend the result of proposition 2.3.6 to the case of n integers, the proof is similar to proposition 2.3.6.

Proposition 2.5.4. *Expression of set of integers as powers of same primes*

Let $S = \{a_1, a_2, a_3, \dots, a_n\}$ be such that $a_i \in \mathbb{Z}$ and $a_i > 2$ for $1 \leq i \leq n$. For each a_i let its prime factorisation be denoted by

$$a_i = \prod_{\substack{p_{(i,k)} \mid a_i \\ p_{(i,k)} \text{ is prime}}} p_{(i,k)}^{e_{(i,k)}}$$

where (i, k) is a index tuple with i denoting one of the primes and k denoting the k -th element of a_i 's prime factorisation. Then there exists a set of primes

$$T = \{t_1, t_2, t_3, \dots, t_v\}$$

with $t_1 < t_2 < t_3 < \dots < t_v$ so that

$$a_i = \prod_{j=1}^v t_j^{f(i,j)}$$

for each $1 \leq i \leq n$.

Proof:

Let each a_i be as given. That is,

$$a_i = \prod_{\substack{p(i,k) | a_i \\ p(i,k) \text{ is prime}}} p_{(i,k)}^{e(i,k)}$$

Let $A_i = \{p(i,k) : p(i,k) \text{ appears in the prime factorisation of } a_i\}$, that is each A_i denotes the set of the prime factors that appear in a_i . We can therefore take T to be

$$T = \bigcup_{i=1}^n A_i$$

so that

$$T = \{t_1, t_2, t_3, \dots, t_v\}$$

where $v \leq \sum_{i=1}^n |A_i|$. It is now left to show that we can pick the primes in the factorisations of the a_i from T . Define the mapping ι_{A_i} by

$$\begin{aligned} \iota_{A_i} : A_i &\rightarrow T \\ x &\mapsto \iota_{A_i}(x) = x \end{aligned}$$

We have that ι_{A_i} maps the elements of A_i to the same element in T . Therefore, we have for some a_i that

$$\begin{aligned} a_i &= \prod_{j=1}^k p_{(i,j)}^{e(i,j)} \\ &= \prod_{j=1}^k \iota_{A_i}(p_{(i,j)})^{e(i,j)} \\ &= \prod_{p(i,j) \in A_i} p_{(i,j)}^{e(i,j)} \\ &= \prod_{p(i,j) \in A_i} p_{(i,j)}^{e(i,j)} * \prod_{t_i \in T \setminus A_i} t_i^0 \\ &= \prod_{t_i \in T} t_i^{g_i} \text{ where } g_i = \begin{cases} e(i,j), & \text{If } t_i = p_{(i,j)} \\ 0, & \text{If } t_i \notin A_i \end{cases} \\ &= t_1^{g_1} t_2^{g_2} t_3^{g_3} \dots t_v^{g_v} \end{aligned}$$

Which expresses a_i in terms of the primes in T as required. \square

The final ingredient required before we can extend the GCD is to extend the minimum function to multiple inputs. This is a straightforward extension.

Definition 2.5.9. *General minimum function for integers*

Let $S = (a_1, a_2, a_3, \dots, a_n) \in \mathbb{Z}^n$ be a n -tuple of integers. We define the minimum function on S by

$$\min : \mathbb{Z}^n \rightarrow \mathbb{Z}$$

$$S \mapsto \min(S) = \begin{cases} a_1, & \text{If } n = 1 \\ \min(a_1, a_2), & \text{If } n = 2 \\ \min(\min(a_1, a_2, a_3, \dots, a_{n-1}), a_n), & \text{If } n \geq 3 \end{cases}$$

We need to show that this is well-defined.

Proposition 2.5.5. *General minimum function for the integers is well-defined*

Let $S = (a_1, a_2, a_3, \dots, a_n) \in \mathbb{Z}^n$ be a n -tuple of integers. We have that $\min(S)$ is well-defined.

Proof:

We argue by induction on n . The base case is $n = 1$ for which the result is trivial, likewise the case $n = 2$ is trivial. So suppose the result holds for some $k > 2$, then we have that

$$\min(\min(a_1, a_2, a_3, \dots, a_{k-1}), a_k)$$

is well-defined. We show that

$$\min(\min(a_1, a_2, a_3, \dots, a_k), a_{k+1})$$

is well-defined. Evaluating the inner $\min(a_1, a_2, a_3, \dots, a_k)$ we have by definition that

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

Which by hypothesis is well-defined. Hence $\min(a_1, a_2, a_3, \dots, a_k) = m$ for some $m \in \mathbb{Z}$. Hence we have that

$$\min(\min(a_1, a_2, a_3, \dots, a_k), a_{k+1}) = \min(m, a_{k+1})$$

Which is well-defined. Hence by induction, we have that the general minimum function on the integers is well-defined. \square

We also have the following proposition.

Proposition 2.5.6. *The general minimum function is associative*

Let $S = (a_1, a_2, a_3, \dots, a_n) \in \mathbb{Z}^n$ be a n -tuple of integers. We have that

$$\min(\min(a_1, a_2, a_3, \dots, a_{n-1}), a_n) = \min(a_1, \min(a_2, a_3, \dots, a_{n-1}, a_n))$$

Proof:

We argue by induction on n . The case $n = 1$ has nothing to prove. Likewise for $n = 2$, so we shall show it holds for $n = 3$. That is

$$\min(\min(a_1, a_2), a_3) = \min(a_1, \min(a_2, a_3))$$

There are 6 cases to consider.

1. $a_1 \leq a_2 \leq a_3$
2. $a_1 \leq a_3 \leq a_2$
3. $a_2 \leq a_1 \leq a_3$
4. $a_2 \leq a_3 \leq a_1$

$$5. \ a_3 \leq a_1 \leq a_2$$

$$6. \ a_3 \leq a_2 \leq a_1$$

$$1. \ a_1 \leq a_2 \leq a_3:$$

We have that

$$\begin{aligned} \min(\min(a_1, a_2), a_3) &= \min(a_1, a_3) = a_1 \\ \min(a_1, \min(a_2, a_3)) &= \min(a_1, a_2) = a_1 \end{aligned}$$

$$2. \ a_1 \leq a_3 \leq a_2:$$

$$\begin{aligned} \min(\min(a_1, a_2), a_3) &= \min(a_1, a_3) = a_1 \\ \min(a_1, \min(a_2, a_3)) &= \min(a_1, a_3) = a_1 \end{aligned}$$

$$3. \ a_2 \leq a_1 \leq a_3:$$

$$\begin{aligned} \min(\min(a_1, a_2), a_3) &= \min(a_2, a_3) = a_2 \\ \min(a_1, \min(a_2, a_3)) &= \min(a_1, a_2) = a_2 \end{aligned}$$

$$4. \ a_2 \leq a_3 \leq a_1:$$

$$\begin{aligned} \min(\min(a_1, a_2), a_3) &= \min(a_2, a_3) = a_2 \\ \min(a_1, \min(a_2, a_3)) &= \min(a_1, a_2) = a_2 \end{aligned}$$

$$5. \ a_3 \leq a_1 \leq a_2:$$

$$\begin{aligned} \min(\min(a_1, a_2), a_3) &= \min(a_1, a_3) = a_3 \\ \min(a_1, \min(a_2, a_3)) &= \min(a_1, a_3) = a_3 \end{aligned}$$

$$6. \ a_3 \leq a_2 \leq a_1:$$

$$\begin{aligned} \min(\min(a_1, a_2), a_3) &= \min(a_2, a_3) = a_3 \\ \min(a_1, \min(a_2, a_3)) &= \min(a_2, a_3) = a_3 \end{aligned}$$

Hence the base case is shown. Now suppose that the proposition holds for some $k > 3$, that is

$$\min(\min(a_1, a_2, a_3, \dots, a_{k-1}), k_n) = \min(a_1, \min(a_2, a_3, \dots, a_{k-1}, a_k))$$

we show that it holds for $k + 1$, i.e.

$$\min(\min(a_1, a_2, a_3, \dots, a_k), a_{k+1}) = \min(a_1, \min(a_2, a_3, \dots, a_k, a_{k+1}))$$

We have by evaluating the inner minimum of the left-hand side we get

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

And so by the induction hypothesis, we have that

$$\begin{aligned} \min(\min(a_1, a_2, a_3, \dots, a_k), a_{k+1}) &= \min(\min(\min(a_1, a_2, a_3, \dots, a_{k-1}), a_k), a_{k+1}) \\ &= \min(\min(a_1, \min(a_2, a_3, \dots, a_{k-1}, a_k)), a_{k+1}), \text{ Induction hypothesis} \end{aligned}$$

As $\min(a_2, a_3, \dots, a_{k-1}, a_k)$ is well-defined by proposition 2.5.5 then $\min(a_2, a_3, \dots, a_{k-1}, a_k) = M$ say where $M \in \mathbb{Z}$. Therefore, on substituting $\min(a_2, a_3, \dots, a_{k-1}, a_k)$ for M for ease of reading we have

$$\begin{aligned} \min(\min(a_1, \min(a_2, a_3, \dots, a_{k-1}, a_k)), a_{k+1}) &= \min(\min(a_1, M), a_{k+1}) \\ &= \min(a_1, \min(M, a_{k+1})) \\ &= \min(a_1, \min(\min(a_2, a_3, \dots, a_{k-1}, a_k), a_{k+1})) \\ &= \min(a_1, \min(a_2, a_3, \dots, a_k, a_{k+1})) \end{aligned}$$

The result now follows by induction. \square

Proposition 2.5.6 is a useful proposition, it allows us to discard the cumbersome notation of the definition of the general minimum function on the Integers. That is to say, we can now simply, and more easily write

$$\min(a_1, a_2, a_3, \dots, a_n)$$

For convenience, we also define the minimum function for a subset of n integers.

Definition 2.5.10. General minimum function for a subset of integers

Let $A = \{a_1, a_2, a_3, \dots, a_n\} \subset \mathbb{Z}$ be a subset of n integers. Let $S = (a_1, a_2, a_3, \dots, a_n) \in A^n$. We define the minimum of the set of integers A by

$$\min(A) = \min(S) = \min(a_1, a_2, a_3, \dots, a_n)$$

That is, we simply take the element of A^n which corresponds to the set.

Example 2.5.5. Let $A = \{2, 3\}$. We have that

$$A^2 = \{(2, 2), (2, 3), (3, 2), (3, 3)\}$$

We have that $S = (2, 3) \in A^2$ and

$$\min(A) = \min(S) = \min(2, 3) = 2$$

We have all the ingredients required to extend the GCD function. We use a method similar to how we extended the minimum function.

Definition 2.5.11. Generalised greatest common divisor

Let $S = (a_1, a_2, a_3, \dots, a_n) \in \mathbb{Z}^n$ be a n -tuple of integers. We define the greatest common divisor function on S by

$$\text{GCD} : \mathbb{Z}^n \rightarrow \mathbb{Z}$$

$$S \mapsto \text{GCD}(S) = \begin{cases} a_1, & \text{If } n = 1 \\ \text{GCD}(a_1, a_2), & \text{If } n = 2 \\ \text{GCD}(\text{GCD}(a_1, a_2, a_3, \dots, a_{n-1}), a_n), & \text{If } n \geq 3 \end{cases}$$

We show that this is well-defined.

Proposition 2.5.7. *Generalised greatest common divisor function for the integers is well-defined*

Let $S = (a_1, a_2, a_3, \dots, a_n) \in \mathbb{Z}^n$ be a n -tuple of integers. We have that $\gcd(S)$ is well-defined.

Proof:

The argument is by induction on n . The base case is $n = 2$ which is well-defined by theorem 2.2.1. Now suppose the result is true for some $k > 2$, that is

$$\text{GCD}(\text{GCD}(a_1, a_2, a_3, \dots, a_{k-1}), a_k)$$

is well-defined. We show that

$$\text{GCD}(\text{GCD}(a_1, a_2, a_3, \dots, a_k), a_{k+1})$$

is well-defined. Evaluating the inner $\text{GCD}(a_1, a_2, a_3, \dots, a_k)$ we have by definition that

$$\text{GCD}(a_1, a_2, a_3, \dots, a_k) = \text{GCD}(\text{GCD}(a_1, a_2, a_3, \dots, a_{k-1}), a_k)$$

Which by hypothesis is well-defined. Hence $\text{GCD}(a_1, a_2, a_3, \dots, a_k) = d$ for some $d \in \mathbb{Z}$. Hence we have that

$$\text{GCD}(\text{GCD}(a_1, a_2, a_3, \dots, a_k), a_{k+1}) = \text{GCD}(d, a_{k+1})$$

Which is well-defined. The result now follows by induction. \square

As with the minimum function, to avoid cumbersome notation we can show that the generalised greatest common divisor is associative.

Proposition 2.5.8. *Generalised GCD is associative*

Let $S = (a_1, a_2, a_3, \dots, a_n) \in \mathbb{Z}^n$ be a n -tuple of integers. We have that

$$\text{GCD}(\text{GCD}(a_1, a_2, a_3, \dots, a_{n-1}), a_n) = \text{GCD}(a_1, \text{GCD}(a_2, a_3, \dots, a_{n-1}, a_n))$$

Proof:

We argue by induction on n . The cases of $n = 1$ and $n = 2$ are trivial, so we show it holds for $n = 3$.

Let $x = \text{GCD}(a_1, \text{GCD}(a_2, a_3))$ and $y = \text{GCD}(\text{GCD}(a_1, a_2), a_3)$. We need to show that $x \mid y$ and $y \mid x$ then we can conclude that $x = y$.

As $x = \text{GCD}(a_1, \text{GCD}(a_2, a_3))$ then by definition of the greatest common divisor, we have that $x \mid a_1$ and $x \mid \text{GCD}(a_2, a_3)$. Moreover, as $x \mid \text{GCD}(a_2, a_3)$ then again by definition of the greatest common divisor we have that $x \mid a_2$ and $x \mid a_3$.

As $x \mid a_1$ and $x \mid a_2$ then $x \mid \text{GCD}(a_1, a_2)$ and likewise $x \mid a_3$ so $x \mid \text{GCD}(\text{GCD}(a_1, a_2), a_3)$ by definition and so $x \mid y$. The proof that $y \mid x$ is similar.

As $x \mid y$ and $y \mid x$ and $x > 0$ and $y > 0$ we conclude that $x = y$ as required.

Now suppose the result is true for some $k > 2$. That is

$$\text{GCD}(\text{GCD}(a_1, a_2, a_3, \dots, a_{k-1}), a_k) = \text{GCD}(a_1, \text{GCD}(a_2, a_3, \dots, a_{k-1}, a_k))$$

we show that

$$\text{GCD}(\text{GCD}(a_1, a_2, a_3, \dots, a_k), a_{k+1}) = \text{GCD}(a_1, \text{GCD}(a_2, a_3, \dots, a_k, a_{k+1}))$$

Evaluation of the inner GCD of the left-hand side yields

$$\text{GCD}(a_1, a_2, a_3, \dots, a_k) = \text{GCD}(\text{GCD}(a_1, a_2, a_3, \dots, a_{k-1}), a_k)$$

So by the induction hypothesis, we have that

$$\begin{aligned}\text{GCD}(\text{GCD}(a_1, a_2, a_3, \dots, a_k), a_{k+1}) &= \text{GCD}(\text{GCD}(\text{GCD}(a_1, a_2, a_3, \dots, a_{k-1}) a_k), a_{k+1}) \\ &= \text{GCD}(\text{GCD}(a_1, \text{GCD}(a_2, a_3, \dots, a_{k-1}, a_k)), a_{k+1}), \text{ By hypothesis}\end{aligned}$$

As $\text{GCD}(a_2, a_3, \dots, a_{k-1}, a_k)$ is well-defined by proposition 2.5.7, we have $\text{GCD}(a_2, a_3, \dots, a_{k-1}, a_k) = d$ with $d \in \mathbb{Z}$. Hence we have

$$\begin{aligned}\text{GCD}(\text{GCD}(a_1, \text{GCD}(a_2, a_3, \dots, a_{k-1}, a_k)), a_{k+1}) &= \text{GCD}(\text{GCD}(a_1, d), a_{k+1}) \\ &= \text{GCD}(a_1, \text{GCD}(d, a_{k+1})) \\ &= \text{GCD}(a_1, \text{GCD}(\text{GCD}(a_2, a_3, \dots, a_{k-1}, a_k), a_{k+1}))\end{aligned}$$

As required. \square

As with the minimum function, we can now simply write

$$\text{GCD}(a_1, a_2, a_3, \dots, a_{n-1}, a_n)$$

Likewise for convenience, we define the GCD function for a subset of n integers.

Definition 2.5.12. General greatest common divisor function for a subset of integers

Let $A = \{a_1, a_2, a_3, \dots, a_n\} \subset \mathbb{Z}$ be a subset of n integers. Let $S = (a_1, a_2, a_3, \dots, a_n) \in A^n$. We define the GCD of the set of integers A by

$$\text{GCD}(A) = \text{GCD}(S) = \text{GCD}(a_1, a_2, a_3, \dots, a_n)$$

That is, we simply take the element of A^n which corresponds to the set.

Example 2.5.6. Let $A = \{2, 3\}$. We have that

$$A^2 = \{(2, 2), (2, 3), (3, 2), (3, 3)\}$$

We have that $S = (2, 3) \in A^2$ and

$$\text{GCD}(A) = \text{GCD}(S) = \text{GCD}(2, 3) = 1$$

We can now finally generalise the computation of the greatest common divisor from the prime factorisation of the inputs.

Proposition 2.5.9. Generalised version of the greatest common divisor from prime factorisation

Let $S = \{a_1, a_2, a_3, \dots, a_n\} \subset \mathbb{Z}$ be a set of integers so that at least one $a_i \neq 0$ for $1 \leq i \leq n$. By proposition 2.5.4, we know that there exists a set of primes

$$T = \{t_1, t_2, t_3, \dots, t_v\}$$

so that for each a_i we have prime factorisations given by

$$a_i = \prod_{j=1}^v t_j^{f(i,j)}$$

For $1 \leq i \leq n$. Define the family of sets for each $1 \leq j \leq v$

$$P_j = \{f(i,j) : 1 \leq i \leq n\}$$

We have that the greatest common divisor $\text{GCD}(a_1, a_2, a_3, \dots, a_n)$ is given by

$$\text{GCD}(a_1, a_2, a_3, \dots, a_n) = t_1^{\min(P_1)} t_2^{\min(P_2)} t_3^{\min(P_3)} \dots t_v^{\min(P_v)}$$

Proof:

The proof is similar to that of proposition 2.3.7. Let $S = \{a_1, a_2, a_3, \dots, a_n\} \subset \mathbb{Z}$ be as given so by proposition 2.5.4 we have a set of primes

$$T = \{t_1, t_2, t_3, \dots, t_v\}$$

so that for each a_i we have prime factorisations given by

$$a_i = \prod_{j=1}^v t_j^{f_{(i,j)}}$$

Now, let $d = \text{GCD}(a_1, a_2, a_3, \dots, a_n)$ and let $D = t_1^{\min(P_1)} t_2^{\min(P_2)} t_3^{\min(P_3)} \dots t_v^{\min(P_v)}$, we show that $d \leq D$ and $D \leq d$. Define $\sigma_j = \min(\{f_{(i,j)} : 1 \leq i \leq n\})$ for $1 \leq j \leq v$.

1. $D \leq d$:

By the definition of the minimum, we have that $\sigma_j \leq f_{(i,j)}$ for each $1 \leq i \leq n$. Hence, for each i and j there exists $k_{(i,j)} \in \mathbb{Z}$ so that

$$f_{(i,j)} = \sigma_j + k_{(i,j)}$$

So that a_i can be expressed as

$$\begin{aligned} a_i &= \prod_{j=1}^v t_j^{f_{(i,j)}} \\ &= \prod_{j=1}^v t_j^{\sigma_j + k_{(i,j)}} \\ &= \prod_{j=1}^v t_j^{\sigma_j} t_j^{k_{(i,j)}} \\ &= \prod_{j=1}^v t_j^{\sigma_j} \prod_{j=1}^v t_j^{k_{(i,j)}} \\ &= D * \prod_{j=1}^v t_j^{k_{(i,j)}} \end{aligned}$$

As a_i was arbitrary this argument holds for each $1 \leq i \leq n$. Hence, we have that $D \mid a_i$ for each i , so D is a common divisor of each a_i . We conclude that $D \leq d$.

2. $d \leq D$:

Suppose that $d \mid D$ then $\exists k \in \mathbb{Z}$ so that

$$d = DK$$

Now, k has a factorisation into primes by the fundamental theorem of arithmetic. Moreover, k could have primes in common with D , so we can take those primes that are in common with D and k and place them into the factorisation of D . That is

$$\begin{aligned}
d &= Dk \\
d &= t_1^{\sigma_1} t_2^{\sigma_2} t_3^{\sigma_3} \dots t_v^{\sigma_v} k \\
d &= t_1^{\lambda_1} t_2^{\lambda_2} t_3^{\lambda_3} \dots t_v^{\lambda_v} k'
\end{aligned}$$

Where λ_j are the new values for each prime after extracting the primes in common with D and k into D . k' are the primes that are not in common. We need to show that

(a) $k' = 1$

(b) $\lambda_j \leq \sigma_j$ for all $1 \leq j \leq v$

(a) $k' = 1$:

Suppose for a contradiction that $k' \neq 1$. As $d > 0$ and $D > 0$ then $k > 0$ and so $k' > 0$. Now as $k' \neq 1$ we have $k' > 1$ and so by the fundamental theorem of arithmetic we have that k' has a factorisation into primes, say

$$k' = q_1^{r_1} q_2^{r_2} q_3^{r_3} \dots q_c^{r_c}$$

Now, no $q_i = t_j$ as k' has no primes in common with $t_1^{\lambda_1} t_2^{\lambda_2} t_3^{\lambda_3} \dots t_v^{\lambda_v}$. Pick one of the primes in k' , say $q = q_i$ then $q \mid d$. Now as $d = \gcd(a_1, a_2, a_3, \dots, a_n)$ then we have $q \mid a_i$ for at least one a_i . This is a contradiction as then q is one of the primes t_j . We conclude that $k' = 1$

(b) $\lambda_j \leq \sigma_j$ for all $1 \leq j \leq v$:

Suppose for contradiction that $\lambda_j > \sigma_j$ for all $1 \leq j \leq v$. Without loss of generality, take $j = 1$, for if not re-label the primes.

By definition of σ_1 , we have that $\sigma_1 = \min(\{f_{(i,1)} : 1 \leq i \leq n\})$, without loss of generality take $i = 1$ as the case for the other values of i are similar. We have that $\sigma_1 = f_{(1,1)}$ and so $\lambda_1 > f_{(1,1)}$. As d is the greatest common divisor of a_1 then there is an $s \in \mathbb{Z}$ so that $ds = a$ where $s > 0$ as both a and d are.

Comparing the prime factorisations, we get that

$$s * t_1^{\lambda_1} t_2^{\lambda_2} t_3^{\lambda_3} \dots t_v^{\lambda_v} = t_1^{f_{(1,1)}} t_2^{f_{(1,2)}} t_3^{f_{(1,3)}} \dots t_v^{f_{(1,v)}}$$

Dividing by $t_1^{f_{(1,1)}}$ we get that

$$s * t_1^{\lambda_1 - f_{(1,1)}} t_2^{\lambda_2} t_3^{\lambda_3} \dots t_v^{\lambda_v} = t_1^{f_{(1,1)} - f_{(1,1)}} t_2^{f_{(1,2)}} t_3^{f_{(1,3)}} \dots t_v^{f_{(1,v)}}$$

Where clearly $t_1^{f_{(1,1)} - f_{(1,1)}} = 1$. So this can be re-written as

$$s * t_1^{\lambda_1 - f_{(1,1)}} t_2^{\lambda_2} t_3^{\lambda_3} \dots t_v^{\lambda_v} = t_2^{f_{(1,2)}} t_3^{f_{(1,3)}} \dots t_v^{f_{(1,v)}}$$

As $\lambda_1 > f_{(1,1)}$ then $\lambda_1 - f_{(1,1)} > 0$ and so t_1 divides the left-hand side of the equation. By the fundamental theorem of arithmetic, t_1 divides the left-hand side it must also divide the right-hand side and therefore be in the factorisation. It is not in the factorisation on the right-hand side which is a contradiction. It follows $\lambda_j \leq \sigma_j$ for all $1 \leq j \leq v$

Therefore we conclude that $d \leq D$.

As $d \leq D$ and $D \leq d$ we have that $d = D$ and the result is shown. \square

These last few results were somewhat technical. To show that our new generalised GCD works we give an example.

Example 2.5.7. We compute $\text{GCD}(54, 78, 35, 144, 50)$. By inspection of each of the numbers we have that

$$\begin{aligned} 54 &= 2 * 3^3 \\ 78 &= 2 * 3 * 13 \\ 35 &= 5 * 7 \\ 144 &= 2^4 * 3^2 \\ 50 &= 2 * 5^2 \end{aligned}$$

Hence, the set of primes T is given by

$$T = \{2, 3, 5, 7, 13\}$$

Now, by the proposition, we know that

$$\text{GCD}(54, 78, 35, 144, 50) = t_1^{\min(P_1)} t_2^{\min(P_2)} t_3^{\min(P_3)} t_4^{\min(P_4)} t_5^{\min(P_5)}$$

Where P_j will be the powers of the prime t_j that appear in the factorisation of each of the inputs. Taking $t_1 = 2, t_2 = 3, t_3 = 5, t_4 = 7$ and $t_5 = 13$ we have

$$\begin{aligned} P_1 &= \{1, 1, 0, 4, 1\} = \{0, 1, 4\} \\ P_2 &= \{3, 1, 0, 2, 0\} = \{0, 1, 2, 3\} \\ P_3 &= \{0, 0, 1, 0, 2\} = \{0, 1, 2\} \\ P_4 &= \{0, 0, 1, 0, 0\} = \{0, 1\} \\ P_5 &= \{0, 1, 0, 0, 0\} = \{0, 1\} \end{aligned}$$

From which it is clear that the minimum of every P_j is 0. So that

$$\text{GCD}(54, 78, 35, 144, 50) = 1$$

With a generalised GCD function, we can extend Bézout's Identity.

Theorem 2.5.1. Generalised Bézout's Identity

Let $S = \{a_1, a_2, a_3, \dots, a_n\} \subset \mathbb{Z}$ be a set of integers so that at least one $a_i \neq 0$ for $1 \leq i \leq n$. Consider $d = \text{GCD}(a_1, a_2, a_3, \dots, a_n)$. Then, for $i \leq 1 \leq n$ we have $\exists x_i \in \mathbb{Z}$ so that

$$d = a_1 x_1 + a_2 x_2 + a_2 x_2 + \dots + a_n x_n = \sum_{i=1}^n a_i x_n$$

Proof:

Let S be as given by the hypothesis and let $d = \text{GCD}(a_1, a_2, a_3, \dots, a_n)$. By definition, we have that as $d \mid a_i$ for each $1 \leq i \leq n$ then by proposition 2.2.5 we have that

$$d \mid \sum_{i=1}^n a_i x_n$$

for any $x_i \in \mathbb{Z}$. Define the set A by

$$G = \left\{ \sum_{i=1}^n a_i x_n : x_i \in \mathbb{Z} \right\}$$

Clearly, there are both positive and negative elements in G , additionally $0 \in G$ by taking each $x_i = 0$. Define \tilde{G} by

$$\tilde{G} = \{g \in G : g > 0\}$$

It follows that $\tilde{G} \subset \mathbb{Z}$ and so by the well-ordering principle it has a smallest element \tilde{g} of the form

$$\tilde{g} = \sum_{i=1}^n a_i x_n$$

We must show that $\tilde{g} \mid a_i$ for each i . Suppose for contradiction and without loss of generality that $\tilde{g} \nmid a_1$. By the division algorithm, we have that

$$a_1 = q\tilde{g} + r$$

with $0 < r < |\tilde{g}|$. Therefore

$$\begin{aligned} a_1 &= q\tilde{g} + r \\ r &= a_1 - q\tilde{g} \\ r &= a_1 - q \sum_{i=1}^n a_i x_n \\ r &= a_1 - \left(qa_1 x_1 + q \sum_{i=2}^n a_i x_n \right) \\ r &= a_1 - qa_1 x_1 - q \sum_{i=2}^n a_i x_n \\ r &= a_1 (1 - qx_1) - q \sum_{i=2}^n a_i x_n \\ r &= a_1 (1 - qx_1) + \sum_{i=2}^n a_i (-qx_n) \end{aligned}$$

Which shows that $r \in \tilde{G}$. Moreover as $0 < r < |\tilde{g}|$ we have $r < \tilde{g}$ a contradiction, so $\tilde{g} \mid a_1$. It follows that $\tilde{g} \mid a_i$ for each i .

As $d = \text{GCD}(a_1, a_2, a_3, \dots, a_n)$ we have that $a_1 = m_1 d$ for each $m_1 \in \mathbb{Z}$. Combining this with the expression for \tilde{g} shows that

$$\begin{aligned} \tilde{g} &= \sum_{i=1}^n a_i x_n \\ \tilde{g} &= \sum_{i=1}^n (m_i d) x_n \\ \tilde{g} &= d \sum_{i=1}^n m_i x_n \end{aligned}$$

So $d \mid \tilde{g}$ and we have that $d \leq \tilde{g}$. As d is the greatest common divisor we have that $d = \tilde{g}$ as required. \square

We are now at the end of a long road. We can now, partially, generalise proposition 2.5.2 to the n variable case, namely we state the requirement for solutions to exist.

Definition 2.5.13. *Linear equation of n indeterminate variables*

Let S be a set. We say an equation is a linear equation in n -variables if it has the form

$$a_1x_1 + a_2x_2 + a_3x_3 + \cdots + a_nx_n = c$$

for some coefficients $a_i \in S$ and $c \in S$ and n indeterminate variables x_n .

Proposition 2.5.10. *Existence of solutions to n variable linear Diophantine equation*

Let $S = \{a_1, a_2, a_3, \dots, a_n\} \subset \mathbb{Z}$ be such that

$$a_1x_1 + a_2x_2 + a_3x_3 + \cdots + a_nx_n = c$$

for the indeterminate variable x_i with $1 \leq i \leq n$. Let $d = \text{GCD}(a_1, a_2, a_3, \dots, a_n)$. We have that there are solutions so that each $x_i \in \mathbb{Z}$ if and only if $d \mid c$.

Proof:

(\Rightarrow): If $c = \sum_{i=1}^n a_ix_n$ then by proposition 2.2.5 we have that $d \mid c$.

(\Leftarrow): Suppose that $d \mid c$ then $\exists e \in \mathbb{Z}$ so that $c = de$. By the generalised Bézout's Identity for each i that $\exists y_i \in \mathbb{Z}$ so that

$$d = \sum_{i=1}^n a_iy_n$$

where $d = \text{GCD}(a_1, a_2, a_3, \dots, a_n)$. Multiplying both sides by e we see that

$$c = e \sum_{i=1}^n a_iy_i = \sum_{i=1}^n a_i(ey_i)$$

Hence each $x_i = ey_i$.

The result is shown. \square

Unlike proposition 2.5.2, we did not show that there are infinitely many solutions and what form they take. Recall that we used example 2.3.9 to find the general form of the solutions for the two-variable case. We shall see if we can do the same for multiple variables.

Example 2.5.8. *Consider the three-variable Diophantine equation*

$$15x + 9y + 27z = 9$$

Clearly $\text{GCD}(15, 9, 27) = 3$ and $3 \mid 9$ so integer solutions exist. How can we find one?

One idea that might give a solution is to try and reduce this to a two-variable equation. How can this be done? We can see that

$$15x + 9y + 27z = 3(5x + 3y) + 27z = 9$$

As $5x + 3y \in \mathbb{Z}$ for $x, y \in \mathbb{Z}$ we will denote this by $v \in \mathbb{Z}$, that is $v = 5x + 3y$. As $v \in \mathbb{Z}$ we can set it to any integer value as $\text{GCD}(5, 3) = 1$ and 1 divides every integer. So let $v = 1$ to give

$$5x + 3y = 1$$

The Euclidean algorithm shows that $x = 2$ and $y = -3$ is a particular solution. Hence the general solutions will be given by

$$\begin{aligned} x &= 2 + \frac{3n}{1} = 2 + 3n \\ y &= -3 - \frac{5n}{1} = -3 - 5n \end{aligned}$$

In particular, the general solutions satisfy $5x + 3y = 1$ and so the general solutions to $5x + 3y = v$ will be given by

$$\begin{aligned}x &= v(2 + 3n) \\y &= v(-3 - 5n)\end{aligned}$$

Now, consider the remaining equation given by

$$3v + 27z = 9$$

By the Euclidean algorithm, we see that $v = 3$ and $z = 0$ is a particular solution, with the general solutions given by

$$\begin{aligned}v &= 3 + \frac{27n}{3} = 3 + 9k \\z &= -0 - \frac{3k}{3} = -k\end{aligned}$$

Hence we have that

$$\begin{aligned}x &= v(2 + 3n) \\y &= v(-3 - 5n) \\v &= 3 + 9k \\z &= -k\end{aligned}$$

As we have an expression for v we can substitute it into the expressions for x and y to give

$$\begin{aligned}x &= (3 + 9k)(2 + 3n) \\y &= (3 + 9k)(-3 - 5n) \\z &= -k\end{aligned}$$

where $n, k \in \mathbb{Z}$. This is a general solution to $15x + 9y + 27z = 9$. Indeed, if we substitute these back into the original equation we get

$$\begin{aligned}15x + 9y + 27z &= 15((3 + 9k)(2 + 3n)) + 9((3 + 9k)(-3 - 5n)) + 27(-k) \\&= 15(3 + 9k)(2 + 3n) + 9(3 + 9k)(-3 - 5n) - 27k \\&= (3 + 9k)(15(2 + 3n) + 9(-3 - 5n)) - 27k \\&= (3 + 9k)(30 + 45n + (-27 - 45n)) - 27k \\&= (3 + 9k)(3) - 27k \\&= 9 + 27k - 27k = 9\end{aligned}$$

Example 2.5.9. We consider the same example again, but we will find a different solution method. So consider the three-variable equation given by

$$15x + 9y + 27z = 9$$

We will express x in terms of y and z . We have

$$x = \frac{9 - 9y - 27z}{15}$$

Observe that we can express 9 and 27 in terms of 15. We have

$$9 = 15 - 6$$

$$27 = 2 * 15 - 3$$

Hence, we can split the expression for x up into the parts that we know are divisible by 15 and the parts that may or may not be divisible by 15

$$\begin{aligned} x &= \frac{9 - 9y - 27z}{15} \\ &= \frac{(15 - 6) - (15 - 6)y - (2(15) - 3)z}{15} \\ &= 1 - y - 2z + \frac{-6 + 6y + 3z}{15} \end{aligned}$$

As we seek $x \in \mathbb{Z}$ then as $1 - y - 2z \in \mathbb{Z}$ we will also require that $\frac{-6 + 6y + 3z}{15} \in \mathbb{Z}$. Let $\frac{-6 + 6y + 3z}{15} = s$ where $s \in \mathbb{Z}$. Then we have

$$x = 1 - y - 2z + s$$

Now, We have that

$$15s = -6 + 6y - 3z$$

We repeat the above process to get y in terms of z and s . Doing so gives

$$15s = -6 + 6y - 3z$$

$$6y = 15s + 6 + 3z$$

As before, we can express 15 and 3 in terms of 6 to get

$$15 = 2 * 6 + 3$$

$$3 = 6 - 3$$

So that,

$$\begin{aligned} 6y &= 15s + 6 + 3z \\ 6y &= (2(6) + 3)s + 6 + (6 - 3)z \\ y &= \frac{(2(6) + 3)s + 6 + (6 - 3)z}{6} \\ y &= 2s + 1 - z + \frac{3s + 3z}{6} \end{aligned}$$

As we need $y \in \mathbb{Z}$ then we require $\frac{3s+3z}{6} \in \mathbb{Z}$ say $\frac{3s+3z}{6} = t$. Then we have

$$y = 2s + 1 - z + t$$

Finally, we have that

$$6t = 3s + 3z$$

Which can be solved directly for z to give $z = 2t - s$. Substituting the value of z in y gives

$$\begin{aligned} y &= 2s + 1 - z + t \\ y &= 2s + 1 - (2t - s) + t \\ y &= 3s + 1 - t \end{aligned}$$

And on substitution of this y value and z into x we get

$$\begin{aligned} x &= 1 - y - 2z + s \\ x &= 1 - (3s + 1 - t) - 2(2t - s) + s \\ x &= 1 - 3s - 1 + t - 4t + 2s + s \\ x &= -3t \end{aligned}$$

Hence, a general solution is given by,

$$\begin{aligned} x &= -3t \\ y &= 3s + 1 - t \\ z &= 2t - s \end{aligned}$$

for $s, t \in \mathbb{Z}$. This is indeed a general solution as

$$\begin{aligned} 15x + 9y + 27z &= 15(-3t) + 9(3s + 1 - t) + 27(2t - s) \\ &= -45t + 27s + 9 - 9t + 54t - 27s \\ &= 9 \end{aligned}$$

Of course, there was nothing special about using only three variables.

Example 2.5.10. Consider the four-variable Diophantine equation

$$55a + 35b - 77c + 144d = 1$$

As $\text{GCD}(55, 35, 77, 144) = 1$ then integer solutions exist. We will use a similar method to example 2.5.9, with some details omitted for brevity.

Expressing, a in terms of b, c and d we get that

$$a = \frac{1 - 35b + 77c - 144d}{55}$$

Noting that

$$\begin{aligned}
35 &= 55 - 20 \\
77 &= 55 + 22 \\
144 &= 2 * 55 + 34
\end{aligned}$$

We can express a as

$$\begin{aligned}
a &= \frac{1 - 35b + 77c - 144d}{55} \\
&= \frac{1 - (55 - 20)b + (55 + 22)c - (2(55) + 34)d}{55} \\
&= -b + c - 2d + \frac{1 + 20b + 22c - 34d}{55}
\end{aligned}$$

We require that $\frac{1 + 20b + 22c - 34d}{55} \in \mathbb{Z}$ say with $\frac{1 + 20b + 22c - 34d}{55} = u$ for $u \in \mathbb{Z}$. We therefore have that

$$55u = 1 + 20b + 22c - 34d$$

Expressing, b in terms of c, d and u we get

$$b = \frac{55u - 1 - 22c + 34d}{20}$$

We see that

$$\begin{aligned}
55 &= 2 * 20 + 15 \\
22 &= 20 + 2 \\
34 &= 20 + 14
\end{aligned}$$

Hence,

$$b = 2u - c + d + \frac{15u - 1 - 2c + 14d}{20}$$

Set $\frac{15u - 1 - 2c + 14d}{20} = v$ where $v \in \mathbb{Z}$. Then

$$20v = 15u - 1 - 2c + 14d$$

Solving c in terms of d, u and v gives

$$c = \frac{15u - 1 - 20v + 14d}{2}$$

where we get

$$c = 7u - 10v + 7d + \frac{u - 1}{2} = 7u - 10v + 7d + x$$

where $x = \frac{u - 1}{2} \in \mathbb{Z}$. We seem to have hit a problem, we still don't have an expression for d . However, suppose $d \in \mathbb{Z}$ is arbitrary, can we recover a general solution with this assumption?

Firstly, we will express a, b and c in terms of u, v, x and d . We get that

$$\begin{aligned}d &\in \mathbb{Z} \\c &= 7u - 10v + 7d + x \\b &= -6d - 5u + 11v - x \\a &= 11d + 13u - 21v + 2x\end{aligned}$$

Observe that,

$$\begin{aligned}55a &= 55(11d + 13u - 21v + 2x) = 605d + 715u - 1155v + 110x \\35b &= 35(-6d - 5u + 11v - x) = -210d - 175u + 385v - 35x \\77c &= 77(7u - 10v + 7d + x) = 539d + 539u - 770v + 77x\end{aligned}$$

Hence

$$\begin{aligned}55a + 35b - 77c + 144d &= 605d + 715u - 1155v + 110x \\&\quad - 210d - 175u + 385v - 35x \\&\quad - (539d + 539u - 770v + 77x) \\&\quad + 144d \\&= 0d + u + 0v - 2x = u - 2x\end{aligned}$$

However, we know that $x = \frac{u-1}{2}$ so that $2x = u - 1$ so

$$u - 2x = u - (u - 1) = 1$$

Hence

$$\begin{aligned}a &= 11d + 13u - 21v + 2x \\b &= -6d - 5u + 11v - x \\c &= 7u - 10v + 7d + x \\u, v, x, d &\in \mathbb{Z}\end{aligned}$$

Gives a general solution for $u, v, x, d \in \mathbb{Z}$.

It is interesting to note that we have four arbitrary integer variables in the previous example. In the case of three variables, we were able to find solutions requiring only two arbitrary integer variables. Does the other method also give a general solution requiring four arbitrary integer variables?

Example 2.5.11. Consider again

$$55a + 35b - 77c + 144d = 1$$

We have that $\text{GCD}(55, 35) = 5$ so that

$$55a + 35b - 77c + 144d = 5(11a + 7b) - 77c + 144d = 1$$

We have that as $11a + 7b \in \mathbb{Z}$ we can replace this with a variable, say u so that we get the equation

$$11a + 7b = 1$$

By the Euclidean algorithm, we see for $u = 5$ that $a = 2$ and $b = -3$ is a general solution, with the general solutions being given by

$$\begin{aligned} a &= 2 + \frac{7n}{1} = 2 + 7x \\ b &= -3 - \frac{11n}{1} = -3 - 11x \end{aligned}$$

Hence the general solution to $11a + 7b = u$ is given by

$$\begin{aligned} a &= u(2 + 7x) \\ b &= u(-3 - 11x) \end{aligned}$$

Now, the original four-variable equation is the three-variable equation

$$5u - 77c + 144d = 1$$

We have that $\text{GCD}(-77, 144) = 1$. So replace $-77c + 144d$ with a variable, say v so that

$$-77c + 144d = 1$$

By the Euclidean algorithm, a particular solution is $c = 43$ and $d = 23$ and the general solution is

$$\begin{aligned} c &= 43 + \frac{144y}{1} = 43 + 144y \\ d &= 23 - \frac{-77y}{1} = 23 + 77y \end{aligned}$$

So that solution to $-77c + 144d = v$ is given by

$$\begin{aligned} c &= v(43 + 144y) \\ d &= v(23 + 77y) \end{aligned}$$

This turns the three-variable equation into a two-variable equation given by

$$5u + v = 1$$

which clearly has a particular solution of $u = 0$ and $v = 1$ to give general solutions given by

$$\begin{aligned} u &= 0 + \frac{z}{1} = z \\ v &= -1 - \frac{5z}{1} = -1 - 5z \end{aligned}$$

Therefore, we have

$$\begin{aligned}
a &= u(2 + 7x) \\
b &= u(-3 - 11x) \\
c &= v(43 + 144y) \\
d &= v(23 + 77y) \\
u &= z \\
v &= 1 - 5z
\end{aligned}$$

So, substituting u and v where required yields

$$\begin{aligned}
a &= z(2 + 7x) \\
b &= z(-3 - 11x) \\
c &= (1 - 5z)(43 + 144y) \\
d &= (1 - 5z)(23 + 77y)
\end{aligned}$$

Where $x, y, z \in \mathbb{Z}$. We verify that this is a general solution. We have

$$\begin{aligned}
55a + 35b - 77c + 144d &= 55z(2 + 7x) + 35z(-3 - 11x) - 77(1 - 5z)(43 + 144y) + 144(1 - 5z)(23 + 77y) \\
&= z(55(2 + 7x) + 35(-3 - 11x)) + (1 - 5z)(-77(43 + 144y) + 144(23 + 77y)) \\
&= z(110 + 385x - 105 - 385x) + (1 - 5z)(-77(43 + 144y) + 144(23 + 77y)) \\
&= 5z + (1 - 5z)(-3311 - 11088y + 3312 + 11088y) \\
&= 5z + (1 - 5z) = 1
\end{aligned}$$

Hence

$$\begin{aligned}
a &= z(2 + 7x) \\
b &= z(-3 - 11x) \\
c &= (1 - 5z)(43 + 144y) \\
d &= (1 - 5z)(23 + 77y)
\end{aligned}$$

where $x, y, z \in \mathbb{Z}$ is a general solution.

Hence we expressed the 4-variable linear Diophantine equation in terms of three arbitrary variables and found expressions for 3-variable linear Diophantine equations in terms of two arbitrary parameters. Is this always the case, and if so does it hold for any number of variables? The answer to this question can be found by considering the method of replacing parts of the n -variable linear Diophantine equation with variables.

For example, for the 3-variable case we have

$$ax + by + cz = d$$

Suppose that $\text{GCD}(a, b, c) \mid d$. After a potential factoring of $ax + by = g_1(a'x + b'y)$ where $g_1 = \text{GCD}(a, b)$, we can replace $a'x + b'y$ with a variable, say u so that

$$a'x + b'y = u$$

As we have factored out the greatest common divisor we will have $\text{GCD}(a', b') = 1$ by proposition 2.2.10 part 7. Hence we can solve $a'x + b'y = 1$ by the Euclidean algorithm and get a general solution

$$\begin{aligned}x &= u(x_0 + b'n) \\ y &= u(y_0 - a'n)\end{aligned}$$

for some $n \in \mathbb{Z}$. As we have seen, this turns the 3-variable equation into a 2-variable equation given by

$$g_1u + cz = d$$

Which is solvable as $\text{GCD}(g_1, c) \mid d$.
This will have a general solution of

$$\begin{aligned}u &= u_0 + \frac{cm}{\text{GCD}(g_1, c)} \\ z &= z_0 - \frac{g_1m}{\text{GCD}(g_1, c)}\end{aligned}$$

Hence a general form of the general solution to the 3-variable case is given by

$$\begin{aligned}x &= \left(u_0 + \frac{cm}{\text{GCD}(g_1, c)}\right)(x_0 + b'n) \\ y &= \left(u_0 + \frac{cm}{\text{GCD}(g_1, c)}\right)(y_0 - a'n) \\ z &= z_0 - \frac{g_1m}{\text{GCD}(g_1, c)}\end{aligned}$$

Here the arbitrary variables are n and m .
Now, in the 4-variable case we have

$$aw + bx + cy + dz = e$$

Suppose that $\text{GCD}(a, b, c, d) \mid e$. As before, after a potential factoring of $aw + bx = g_1(a'w + b'x)$ where $g_1 = \text{GCD}(a, b)$, we can replace $a'w + b'x$ with a variable, say u so that

$$a'w + b'x = u$$

We can solve $a'w + b'x = 1$ by the Euclidean algorithm and get a general solution for any u

$$\begin{aligned}w &= u(w_0 + b'n) \\ x &= u(x_0 - a'n)\end{aligned}$$

for some $n \in \mathbb{Z}$. This turns the 4-variable equation into a 3-variable equation given by

$$g_1u + cy + dz = e$$

Which is solvable as $\text{GCD}(g_1, c, d) \mid e$. Two choices can be made here: replacing $g_1u + cy$ or $cy + dz$.

Taking the first choice, we have after a potential factoring $g_1u + cy = g_2(g'_1u + c'y)$ where $g_2 = \text{GCD}(g_1, c)$, we set $g'_1u + c'y = v$, solving $g'_1u + c'y = 1$ by the Euclidean algorithm, we get a general solution for any v

$$\begin{aligned}u &= v(u_0 + c'm) \\ y &= v(y_0 - g'_1 m)\end{aligned}$$

We are now left with a 2-variable equation given by

$$g_2 v + dz = e$$

Again, solvable because $\text{GCD}(g_2, d) \mid e$. This has a general solution given by

$$\begin{aligned}v &= v_0 + \frac{dk}{\text{GCD}(g_2, d)} \\ z &= z_0 - \frac{g_2 k}{\text{GCD}(g_2, d)}\end{aligned}$$

Hence, we have a general solution given by.

$$\begin{aligned}w &= \left(v_0 + \frac{dk}{\text{GCD}(g_2, d)}\right)(u_0 + c'm)(w_0 + b'n) \\ x &= \left(v_0 + \frac{dk}{\text{GCD}(g_2, d)}\right)(u_0 + c'm)(x_0 - a'n) \\ y &= \left(v_0 + \frac{dk}{\text{GCD}(g_2, d)}\right)(y_0 - g'_1 m) \\ z &= z_0 - \frac{g_2 k}{\text{GCD}(g_2, d)}\end{aligned}$$

Alternatively, suppose we took the choice of $cy + dz$, after a potential factoring $cy + dz = g_2(c'y + d'z)$ where $g_2 = \text{GCD}(c, d)$. Setting $c'y + d'z = v$ and solving $c'y + d'e = 1$ by the Euclidean algorithm, we get a general solution for any v given by

$$\begin{aligned}y &= v(y_0 + d'm) \\ z &= v(z_0 - c'm)\end{aligned}$$

Which now gives the 2-variable equation

$$g_1 u + g_2 v = e$$

Solutions exists as $\text{GCD}(g_1, g_2) \mid e$. This has a general solution

$$\begin{aligned}u &= u_0 + \frac{g_2 k}{\text{GCD}(g_1, g_2)} \\ v &= v_0 + \frac{g_1 k}{\text{GCD}(g_1, g_2)}\end{aligned}$$

Hence, a different general solution to the original 4-variable equation is given by

$$\begin{aligned}
w &= \left(u_0 + \frac{g_2 k}{\text{GCD}(g_1, g_2)} \right) (w_0 + b'n) \\
x &= \left(u_0 + \frac{g_2 k}{\text{GCD}(g_1, g_2)} \right) (x_0 - a'n) \\
y &= \left(v_0 + \frac{g_1 k}{\text{GCD}(g_1, g_2)} \right) (y_0 + d'm) \\
z &= \left(v_0 + \frac{g_1 k}{\text{GCD}(g_1, g_2)} \right) (z_0 - c'm)
\end{aligned}$$

It seems there are a few things to show. Firstly, is it possible to show that the solution to an n -variable linear Diophantine equation can be expressed in terms of $n - 1$ variables? Secondly, do the general solutions have the form of being a product of 2 variable solutions?

Given an n -variable linear Diophantine equation, replacing two variables with a single variable turns any n -variable equation with $n - 1$ variables. Eventually, this process terminates when, after some number of replacements, we get to a 2-variable equation.

We have therefore a strong understanding of solving Linear Diophantine equations in any number of variables.

2.5.2 Polynomials

Previously, we have seen how to handle linear equations in multiple variables. That is equations of the form

$$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = b, \quad a_i, x_i, b \in \mathbb{Z}$$

A natural question to ask is can we extend this to non-linear equations? For example, we have defined what we mean by a square number 2.3.8. That is, a number $y \in \mathbb{Z}$ is square if $\exists x \in \mathbb{Z}$ so that $x^2 = y$. If we consider x as a variable for a moment, then we have seen many examples of solving this type of equation. We studied this when finding what integer numbers were squares. We can ask the question, what happens if we combine this x^2 variable with just the variable x , for example, what values for $x \in \mathbb{Z}$ or \mathbb{Q} would satisfy

$$x^2 + x = 2$$

We can go further than simply x^2 . For example, we can consider

$$x^n = \prod_{i=1}^n x$$

and combine variables of this form however we wish and multiply them by constants, for example.

$$x^8 + 15x^7 - 8x^3 + 2x^2 + x + 5 = 0$$

We will want to study equations of this form. We will want to

Definition 2.5.14. Monomial

Let X be a variable and let $a \in S$ for some set $S \neq \emptyset$. We define a monomial to be an expression of the form

$$aX^n$$

where $n \in \mathbb{Z}$ with $n \geq 0$

From a monomial, we define a so-called polynomial

Definition 2.5.15. *Polynomial*

Let S be a set and let $n \in \mathbb{Z}$ with $n \geq 0$. Let X be a variable. We define a polynomial to be an expression of the form

$$P(X) = a_n X^n + a_{n-1} X^{n-1} + a_{n-2} X^{n-2} + \cdots + a_1 X + a_0$$

Where $a_0, a_1, \dots, a_n \in S$ are called the coefficients of the polynomial. We say that X is an indeterminate variable and we say that $P(X)$ is a polynomial in X with coefficients in S .

Here we are formally using the $+$ operation associated with S between the terms in the polynomial.

We can, of course, replace X with a particular value to evaluate the polynomial.

Definition 2.5.16. *Evaluation of a polynomial*

Let $S \neq \emptyset$ be a set and let $P(X)$ be a polynomial with coefficients in S . Let $s \in S$. We define the evaluation of the polynomial P at s by

$$P(s) = a_n s^n + a_{n-1} s^{n-1} + a_{n-2} s^{n-2} + \cdots + a_1 s + a_0$$

Example 2.5.12. Let $S = \mathbb{Z}$ and define $P(X)$ by

$$P(X) = 2X^2 - 3X + 5$$

What is $P(1)$?

On substituting $X = 1$ we see

$$P(1) = 2(1)^2 - 3(1) + 5 = 2 - 3 + 5 = 4$$

It will be useful to describe the set of all polynomials whose coefficients lie in some set S .

Definition 2.5.17. *Set of all polynomials with coefficients in a set S*

Let $S \neq \emptyset$. We define the set of all polynomials whose coefficients are in S by the set

$$S[X] = \left\{ \sum_{i=0}^n s_i X^i : n \in \mathbb{N} \text{ and } s_i \in S \right\}$$

We define the polynomials to be the elements of this set and write $P \in S[X]$, with the understanding that P actually means $P(X)$.

From the definition of a polynomial, we have many choices that we can make that allow us to create a polynomial. We can modify the coefficients a_i for $0 \leq i \leq n$ however we wish, so long as they are all in the set S . In particular, the choices we can make are clearly dependent on the value of n we can pick. The value of n is an important property of polynomials.

Definition 2.5.18. *Degree of a polynomial*

Let $P \in S[X]$. We define the degree of the polynomial P to be the largest $n \in \mathbb{Z}$ so that the coefficient of X^n is not equal to zero.

We write $\deg(P) = n$ to mean the degree of the polynomial P is n .

Example 2.5.13. Let $S = \mathbb{Z}$ and define $P(X)$ by

$$P(X) = 2X^2 - 3X + 5$$

We see that the largest n where $X^n \neq 0$ is 2 so $\deg(P) = 2$.

The astute reader might ask the following. Suppose that P is given by

$$P = 0 + 0 * X + 0 * X^2 + 0 * X^3 + \cdots + 0 * X^n$$

what is the degree of P ? On one hand, by our definition, we can't assign it a degree! There are no non-zero coefficients in the polynomial! On the other hand, we intuitively know that the above polynomial represents a meaningful polynomial, especially for the theory we are attempting to develop. It is not clear how to resolve this problem for now. Perhaps, developing the theory as much as we can without it will make it clear how to resolve this issue.

2.5.2.1 Defining addition between two polynomials

We can define how to add two polynomials together. To do so we need to recast how we see a polynomial, and to do so recall the definition of the Cartesian product of n sets 1.2.18.

Let S_1, S_2, \dots, S_n be sets. We define the Cartesian product of S_1, S_2, \dots, S_n , denoted $S_1 \times S_2 \times \dots \times S_n$ to be the set of all ordered pairs of the form (s_1, s_2, \dots, s_n) where $s_1 \in S_1, s_2 \in S_2, \dots, s_n \in S_n$. This is to say that

$$S_1 \times S_2 \times \dots \times S_n = \{(s_1, s_2, \dots, s_n) : s_1 \in S_1, s_2 \in S_2, \dots, s_n \in S_n\}$$

In particular, if all of the sets are the same we denote this by S^n . We can use this idea to define a polynomial of degree n as a tuple.

Firstly, observe that we can write a polynomial $P(X)$ as

$$\begin{aligned} P(X) &= a_n X^n + a_{n-1} X^{n-1} + a_{n-2} X^{n-2} + \dots + a_1 X + a_0 \\ &= a_0 + a_1 X + a_2 X^2 + a_3 X^3 + \dots + a_{n-1} X^{n-1} + a_n X^n \\ &= \sum_{i=0}^n a_i X^i \end{aligned}$$

That is we can express P as the sum of products of coefficients in S and the corresponding power of the indeterminate variable X .

Now, we have that $\deg(P) = n$ so in order to have the correct sized tuple we must consider the Cartesian product of S with itself $n+1$ times, that is

$$S^{n+1} = \prod_{i=0}^n S$$

As each $a_i \in S$ for $0 \leq i \leq n$ we have that the tuple $a = (a_0, a_1, a_2, \dots, a_{n-1}, a_n) \in S^{n+1}$. This is the correspondence we need.

Definition 2.5.19. *Polynomial as an $n+1$ -tuple*

Let $S \neq \emptyset$ and let $n \in \mathbb{Z}$ with $n \geq 0$ so that $\deg(P) = n$ where

$$P(X) = a_n X^n + a_{n-1} X^{n-1} + a_{n-2} X^{n-2} + \dots + a_1 X + a_0$$

We can view a polynomial as an element of the set S^{n+1} , say a with the form

$$a = (a_0, a_1, a_2, \dots, a_{n-1}, a_n)$$

More simply, we can write

$$P = (a_0, a_1, a_2, \dots, a_{n-1}, a_n)$$

where we have the powers of X^n being implicit.

This definition has an immediate consequence, it enables us to have a representation for each X^n for any $n \geq 0$. For example, we see that

$$\begin{aligned} P(X) = 1 = X^0 &\iff a = (1) \\ P(X) = X &\iff a = (0, 1) \\ P(X) = X^2 &\iff a = (0, 0, 1) \\ P(X) = X^3 &\iff a = (0, 0, 0, 1) \\ &\dots \end{aligned}$$

This allows us to build an understanding of how to properly define addition of two polynomials. Suppose we have

$$\begin{aligned}P(X) &= 1 + X + X^2 \\Q(X) &= 4 - 3X + X^2 + X^3\end{aligned}$$

Where the coefficients of P and Q are elements of \mathbb{Z} . We see that $P = (1, 1, 1)$ and $Q = (4, -3, 1, 1)$. Firstly, we have that P has less entries in its tuple than Q . We can account for this by noting that $P(X) = 1 + X + X^2 = 1 + X + X^2 + 0X^3$ and so an alternative representation of P is given by $P = (1, 1, 1, 0)$.

Now, considers the terms in both P and Q which are associated with X^0 i.e. $P_0(X) = 1$ and $Q_0(X) = 4$. As these are simply elements of \mathbb{Z} we would expect that $P_0(X) + Q_0(X) = 1 + 4 = 5$ and so the sum to have the tuple form (5).

Considering the terms in both P and Q which are associated with X^1 , $P_1(X) = X$ and $Q_1(X) = -3X$, we would then expect that $P_1(X) + Q_1(X) = X - 3X = -2X$.

We can continue this process for the other terms X^2 and X^3 to get

$$\begin{aligned}P_0(X) + Q_0(X) &= 1 + 4 = 5 \\P_1(X) + Q_1(X) &= X - 3X = -2X \\P_2(X) + Q_2(X) &= X^2 + X^2 = 2X^2 \\P_3(X) + Q_3(X) &= 0X^3 + X^3 = X^3\end{aligned}$$

This would then suggest that $P(X) + Q(X) = 5 - 2X + 2X^2 + X^3$. Or, expressing this in tuple form, we have

$$(1, 1, 1, 0) + (4, -3, 1, 1) = (5, -2, 2, 1)$$

That is, the addition of two tuples representing polynomials is done by doing an "element-wise" addition of the tuples. There are a few things that would need to be considered for this to become the foundation for defining addition for polynomials.

Firstly, we observed that $P_0(X) + Q_0(X)$ made sense as this represents integer addition. If we picked our coefficients from say \mathbb{N} , we would not be able to consider $P_0(X) - Q_0(X)$; in fact, this holds for each $P_i - Q_i$. It would therefore be useful to have closure of addition, and additionally a notion of subtraction of the elements of S . This puts a restriction on what the set S can be, for example, it is clear that $S \neq \mathbb{N}$ as subtraction is not closed in \mathbb{N} .

This also means our definition of polynomials is going to depend on the underlying set that the coefficients come from. It is therefore a wise idea to, at least temporarily, distinguish between when we are talking about polynomial addition and when we are talking about the addition of the elements of the set S .

We will use $+_S$ when talking about addition between the elements of the set S , and we will use \oplus_S for the polynomial addition¹⁶.

Furthermore, we had that P was of a lesser degree than Q , $\deg(P) = 2$ and $\deg(Q) = 3$. This poses no real issue as we can always extend an m -tuple to an n -tuple, for $m < n$. Indeed, suppose we have an element $s \in S^m$ and we want to extend it to an element of S^n where $m < n$, then we can use the following map

$$\begin{aligned}f : S^m &\rightarrow S^n \\(s_1, s_2, \dots, s_m) &\mapsto f((s_1, s_2, \dots, s_m)) = (s_1, s_2, \dots, s_m, \underbrace{0, 0, \dots, 0}_{n-m \text{ times}})\end{aligned}$$

Or more simply, append $n - m$ 0s to the element of S^m . We provide a general definition.

¹⁶When we have fully defined polynomial addition, we will go with the usual convention of just using $+$ to denote addition

Definition 2.5.20. *Polynomial tuple extension map*

Let $n, m \in \mathbb{Z}$ so that $m \leq n$. We define the polynomial tuple extension map by

$$E_m^n : S^m \rightarrow S^n$$

$$(s_1, s_2, \dots, s_m) \mapsto E_m^n((s_1, s_2, \dots, s_m)) = (s_1, s_2, \dots, s_m, \underbrace{0, 0, \dots, 0}_{n-m \text{ times}})$$

Here, we are using the notation E_m^n to indicate this extends an m -tuple to an n -tuple.

This means that given two polynomials expressed in their tuple forms, we can always extend the one with the fewer elements so that they share the same number of elements. From this, we can define polynomial addition.

Definition 2.5.21. *Polynomial addition*

Let S be a set and let $P, Q \in S[X]$ so that $\deg(P) = n$ and $\deg(Q) = m$ so that without loss of generality we have that $m \leq n$ and

$$P = (p_0, p_1, p_2, \dots, p_{n-1}, p_n)$$

$$Q = (q_0, q_1, q_2, \dots, q_{m-1}, q_m)$$

Furthermore, suppose that S is endowed with an addition $+_S$ such that $+_S$ is well-defined and closed. We define the addition of P and Q , by

$$\oplus_S : S^n \times S^n \rightarrow S^n$$

$$(P, E_m^n(Q)) \mapsto \oplus_S(P, E_m^n(Q)) = (p_0 +_S q_0, p_1 +_S q_1, p_2 +_S q_2, \dots, p_{n-1} +_S 0, p_n +_S 0)$$

Proposition 2.5.11. *Polynomial addition is well-defined and closed*

Let S be a set and let $P, Q \in S[X]$ so that $\deg(P) = n$ and $\deg(Q) = m$ so that without loss of generality we have that $m \leq n$ and

$$P = (p_0, p_1, p_2, \dots, p_{n-1}, p_n)$$

$$Q = (q_0, q_1, q_2, \dots, q_{m-1}, q_m)$$

Furthermore, suppose that S is endowed with an addition $+_S$ such that $+_S$ is well-defined and closed. We have that the polynomial addition of P and Q , denoted $P \oplus_S Q$ is well-defined and closed.

Proof:

This is immediate. By the definition of the polynomial addition, we have that

$$P \oplus_S Q = (p_0 +_S q_0, p_1 +_S q_1, p_2 +_S q_2, \dots, p_{n-1} +_S 0, p_n +_S 0)$$

As $+_S$ is well defined and closed, then we have that $p_i +_S q_i \in S$ for $0 \leq i \leq m$. Moreover, for $m < i \leq n$ we have that $p_i +_S 0 = p_i \in S$. Hence, all the entries in the tuple given by $P \oplus_S Q$ are elements of S so that $P \oplus_S Q \in S^n$. \square

Lemma 2.5.1. *Degree of polynomial from polynomial addition*

Let $P, Q \in S[X]$. Then

$$\deg(P \oplus_S Q) \leq \max(\deg(P), \deg(Q))$$

Proof:

The result is instant if $\deg(P) = \deg(Q)$ so suppose not and without loss of generality suppose that $\deg(P) > \deg(Q)$ where $\deg(P) = n$ and $\deg(Q) = m$. Then as tuples we have that

$$P = (p_0, p_1, p_2, \dots, p_{n-1}, p_n)$$

$$Q = (q_0, q_1, q_2, \dots, q_{m-1}, q_m)$$

As $\deg(Q) < \deg(P)$ we use the tuple extension mapping E_m^n on Q and we have that $\deg(E_m^n(Q)) \leq n$. Hence

$$\deg(P \oplus_S Q) \leq \deg(P \oplus_S E(Q)) \leq n = \max(\deg(P), \deg(Q))$$

□

We are getting an idea for our problem with the polynomial given by

$$P = 0 + 0 * X + 0 * X^2 + 0 * X^3 + \dots + 0 * X^n$$

If we want lemma 2.5.1 to be consistent, we should define the degree of P to be such that it is no larger than the degree of any other polynomial. In particular, for $c \in S$ we have that $Q = c$ with $Q \in S[X]$ has degree 0, we must have that $\deg(P) < \deg(Q) = 0$. This still doesn't fully answer the question, which negative integer should we take for the degree of P ? Maybe, once we have a definition for the multiplication of polynomials, it will provide further insight.

Now, given a potential candidate for defining the addition of two polynomials, we can also consider a potential candidate for defining the subtraction of two polynomials. As before, we take inspiration from \mathbb{Z} .

As we have shown that the addition of integers is closed and well-defined, additionally, for every $x \in \mathbb{Z}$ we have that $\exists y$ so that $x + y = 0$. In particular, we take $y = -x$ so that the expression becomes $x - x = 0$. A sensible definition for polynomial subtraction should also respect these properties; subtracting two polynomials should give another polynomial. This raises a question; suppose $P \in S[X]$, what is $P - P$?

We know that in \mathbb{N} , \mathbb{Z} and \mathbb{Q} , that for an element x that $x - x$ should be 0, but what does it mean for 0 to be an element of S and by extension $S[X]$? In particular is it the same 0 as for \mathbb{N} , \mathbb{Z} and \mathbb{Q} ?

On the other hand, we know that for any x in \mathbb{N} , \mathbb{Z} and \mathbb{Q} that $x + 0 = x = 0 + x$, a similar sort of element of S would be useful and clearly plays an important role for defining a similar element for $S[X]$. This idea is general enough, assuming we have a well-behaved $+_S$, that we can apply it to a set S .

Definition 2.5.22. *Additive Identity of a set S*

Let S be a set so that there is an operation $+_S : S^2 \rightarrow S$ such that $+_S$ is closed and well-defined. Let $e \in S$. If we have that $\forall s \in S$ that $s +_S e = s$, then we say that e is a right additive identity element of S .

Similarly, if $\forall s \in S$ we have that $e +_S s = s$, then we say that e is a left additive identity element of S .

If we have that $\forall s \in S$ that $e +_S s = s = s +_S e$, we simply call e an additive identity element.

If we need to be clear which set the additive inverse belongs to, we will write e_S

It is an immediate consequence of $+_S$ that the identity element is unique.

Proposition 2.5.12. *The additive identity element of a set S is unique*

Let S be a set so that there is an operation $+_S : S^2 \rightarrow S$ such that $+_S$ is closed and well-defined. Let $e, f \in S$ be additive identity elements of S .

We have that $e = f$.

Proof:

Let S and $+_S : S^2 \rightarrow S$ be as given, and let $e, f \in S$ be additive identity elements of S .

By definition, we have that

$$e = e +_S f = f$$

As $+_S$ is well-defined and closed, we have that $e = f$ as required. □

From this, we can immediately identify that 0 in \mathbb{N} , \mathbb{Z} and \mathbb{Q} is unique.

We have resolved one part of this problem, that in \mathbb{N} , \mathbb{Z} and \mathbb{Q} , for an element x that $x - x = 0$. We have answered what it means for "0" to be in S , but what does it mean for $-x \in S$ given $x \in S$? Noting that $x - x = x +_S (-x)$, this is precisely what it means for x to be invertible in S at least with respect to $+_S$. As with the additive identity of S , this idea is also general enough to apply to a more general set S .

Definition 2.5.23. *Additive Inverse of a set S*

Let S be a set so that there is an operation $+_S : S^2 \rightarrow S$ such that $+_S$ is closed and well-defined. Let $s \in S$.

If we have that $\exists x \in S$ such that $s +_S x = e$, then we say that x is a right additive inverse element of s in S .

Similarly, if $\exists x \in S$ such that $x +_S s = e$, then we say that x is a left additive inverse element of s in S .

If we have that $\exists x \in S$ that $x +_S s = s = s +_S x$, we simply call x an additive inverse element of s in S .

As with the additive identity element, we have an immediate consequence that the inverse of an element $s \in S$ is unique.

Proposition 2.5.13. *The additive inverse element of an element of S is unique*

Let S be a set so that there is an operation $+_S : S^2 \rightarrow S$ such that $+_S$ is closed and well-defined. Let $s \in S$ be an arbitrary element of S .

We have that the additive inverse of s is unique.

Proof:

Let S and $+_S : S^2 \rightarrow S$ be as given, and let $s \in S$ be an arbitrary element of S and suppose that s has two inverses x and y .

By definition, we have that

$$\begin{aligned} x &= x +_S e \\ &= x +_S (s +_S y) \\ &= \end{aligned}$$

As $+_S$ is well-defined and closed, we have that $e = f$ as required. \square

It would also be useful to undo the addition of polynomials via polynomial subtraction. The only requirement is that we need $+_S$ to be invertible. In particular, as we are using a well-defined and closed operation on S , that is $+_S$, we have gained a definition of subtraction for free! Using $-_S$ to denote subtraction in S , we have

$$\begin{aligned} \ominus_S : s^n \times s^n &\rightarrow s^n \\ (P, E(Q)) &\mapsto \ominus_S(P, E(Q)) = (p_0 -_S q_0, p_1 -_S q_1, p_2 -_S q_2, \dots, p_{n-1} -_S q_{n-1}, p_n -_S q_n) \end{aligned}$$

Given a notion of subtraction, we can also define what it means for two polynomials to be equal. Firstly, recall what it means for

Definition 2.5.24. *Equality of Polynomials*

Let $P, Q \in S[X]$ where $\deg(P) = n$ and $\deg(Q) = m$ where without loss of generality $m \leq n$.

We say that P and Q are equal as polynomials, written $P = Q$, if and only if

$$P \ominus_S Q = 0 = \left(\underbrace{0, 0, 0, \dots, 0}_{n+1 \text{ times}}, 0 \right)$$

That is, if the difference between the two is the zero polynomial.

We can therefore define the following relation.

Definition 2.5.25.

It is immediate that a polynomial therefore has a unique representation as an $n + 1$ -tuple.

2.5.2.2 Defining multiplication between two polynomials

We can use the same idea of the $n+1$ -tuples to define multiplication of polynomials. Recall that we observed that we can express the intermediate X , and powers of it, as follows

$$\begin{aligned} P(X) = 1 = X^0 &\iff a = (1) \\ P(X) = X &\iff a = (0, 1) \\ P(X) = X^2 &\iff a = (0, 0, 1) \\ P(X) = X^3 &\iff a = (0, 0, 0, 1) \\ &\dots \end{aligned}$$

Intuitively, we want $X^2 = X * X$, $X^3 = X^2 * X$ and so on. That is

$$\begin{aligned} X * X &= (0, 1) * (0, 1) = (0, 0, 1) = X^2 \\ X^2 * X &= (0, 0, 1) * (0, 1) = (0, 0, 0, 1) = X^3 \\ &\dots \end{aligned}$$

What about more complex expressions? Say $X * (X + X^2)$. The answer to this would depend on if multiplication is distributive over addition with respect to the indeterminate, and additionally on multiplication is commutative!. For now, let us assume that this is the case,

It seems therefore that multiplication by X has the effect of “shifting” to the right