MA136: Introduction to abstract algebra 2021-22

Contents

Lecture 14 - Quotient Groups

Lecture 15 - The Alternating Group

Lecture 16 - Rings

Lecture 17 - The Unit Group of a Ring

Lecture 18 - Congruences Revisited

Many thanks to Samir Siksek for allowing me to use his notes as the basis for these course notes. The vast majority of what you read here is written by Samir with a few additions and alterations by me. Please e-mail r.m.lissaman@warwick.ac.uk with details of any errors you might notice on the notes here. You are advised to download a new copy of these notes at intervals throughout the course as each version is subject to change. This will mainly be correcting errors and typos. Details of any substantial changes will be posted to the Moodle page for MA136.

Richard Lissaman

September 2021

Lecture 1 - Prologue

In this course we shall look at some *algebraic structures.* The main examples we'll see are *groups* and *rings.* Let's not worry about what these are just yet.

First of all, let's explore the title of the module and the words *abstract* and *algebra.* When we think of the word algebra we think of symbols replacing specific instances of numbers and manipulationg them to various ends.

Think about some typical algebra that you may have seen at school/college.

$$\begin{aligned} x(x+y) &= yx \\ \Rightarrow x^2 + xy - yx &= 0 \\ \Rightarrow x^2 &= 0 \\ \Rightarrow x &= 0 \end{aligned}$$

Let's think very precisely (much more precisely than usual) about which properties of numbers we are using here.

First of all we have used the fact that $x(x+y) = x^2 + xy$ for any two numbers.

We've also used that facts that $xy = yx$, that $yx + (-yx) = 0$ and that $x^2 + 0 = x^2$.

Finally we've used the fact that the only number whose square is 0 is 0 itself when we've concluded that $x^2 = 0 \Rightarrow x = 0$.

All of these facts are certainly true if x and y are regular numbers, '+' means regular addition and xy mean $x \times y$ the regular multiplication of two numbers x and y. However some of these facts would not be true in other cases.

For example if x and y are both matrices and xy means matrix multiplication of x and y then we can't say that $xy = yx$.

Here is a case in point

$$\begin{pmatrix} 2 & 1 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 3 & 2 \end{pmatrix} \neq \begin{pmatrix} 2 & 0 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 3 & 2 \end{pmatrix}.$$

We can't say that if the square of a matrix is the zero matrix then that matrix itself must be the zero matrix since

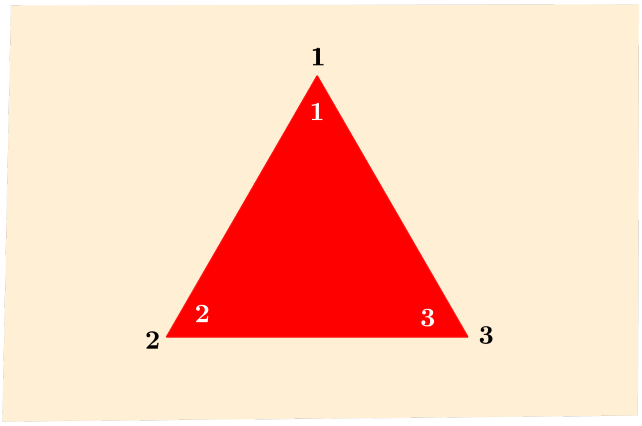
$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

If x and y are vectors and xy means the vector product of x and y then we can't say that $xy = yx$.

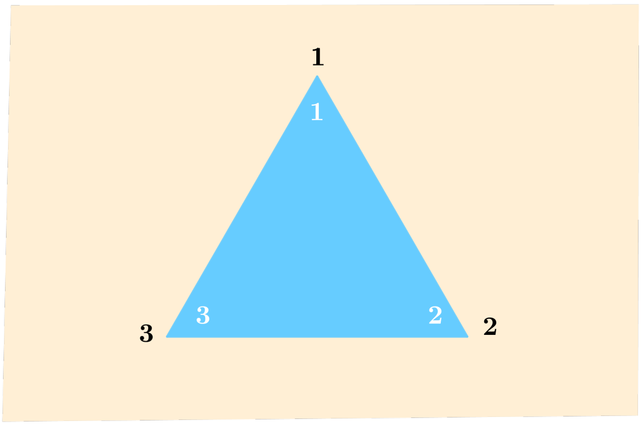
In this module we'll be paying a lot of attention to the exact properties of operations like this. It will be important to be able to decide which objects and operations have which properties because we want to explore what happens when we abstract these from any particular instance of it. Probably this last sentence doesn't mean much to you now but it should mean more to you as the module goes on.

Sometimes, we can also think about the algebraic structures we'll be exploring geometrically. This approach reflects more closely the origins of the abstract algebra as a discipline in its own right. Here is an example, to give a flavour of this.

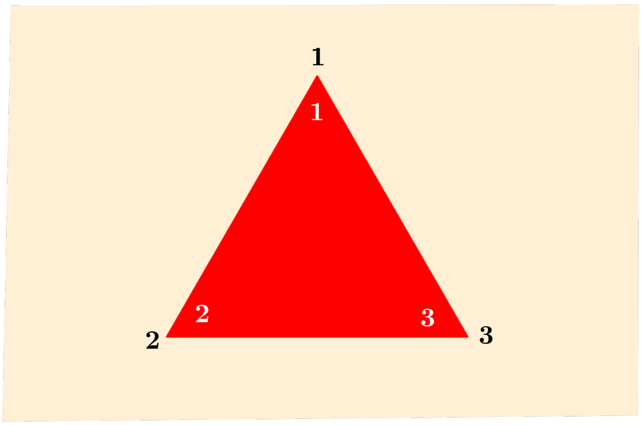
Imagine the following, an equilateral triangle made of red paper fitting exactly into a yellow paper frame.



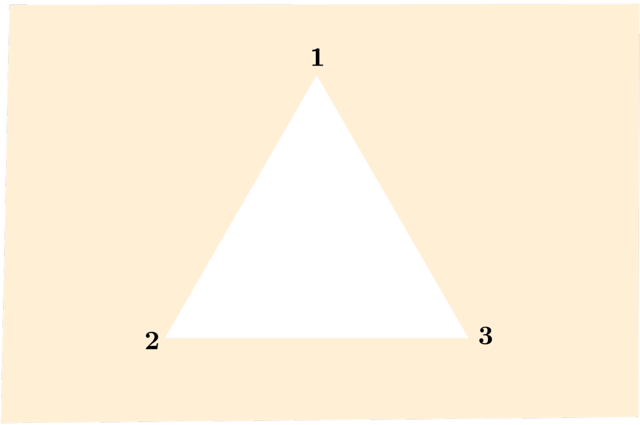
The back of the triangle is blue, from the back it looks like this (notice the 2 and 3 are swapped on this reverse view).



Let's turn it back over again...

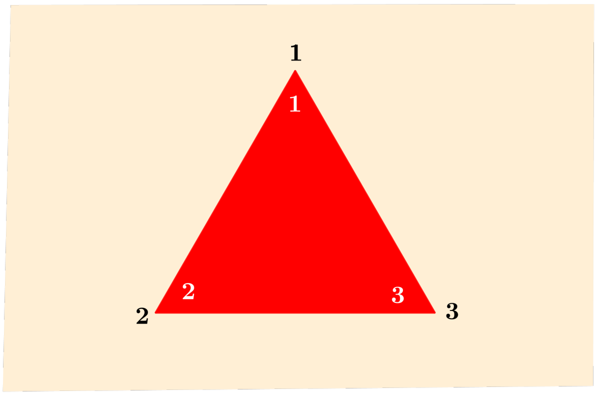


and lift the triangle out of its frame.

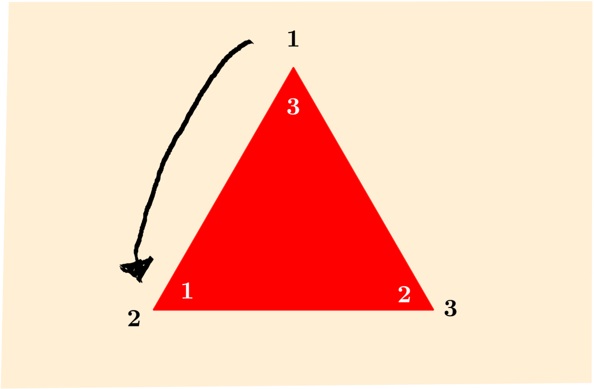


How many ways can we put the triangle back into its frame so that it occupies the frame exactly? Let's count them.

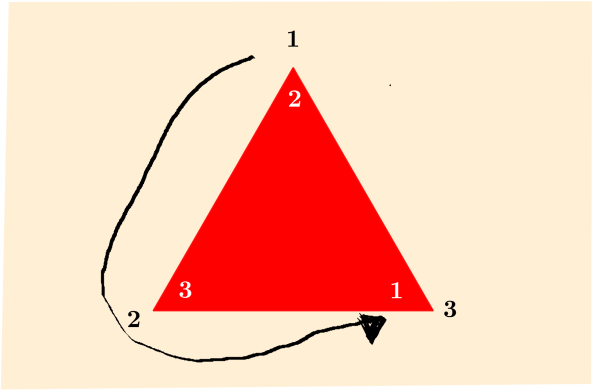
1. We could put it straight back down, exactly in the position it was in when we picked it up:



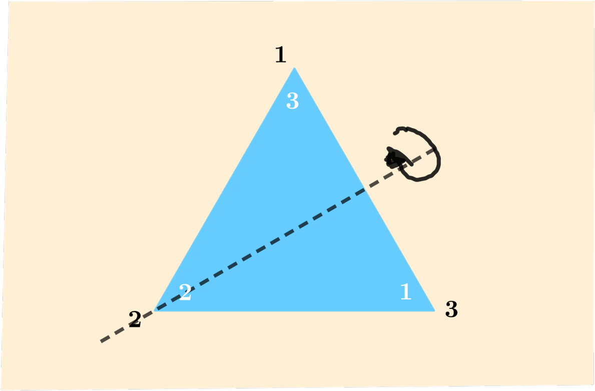
2. We could rotate it by $\frac{1}{3}$ of a turn anticlockwise.



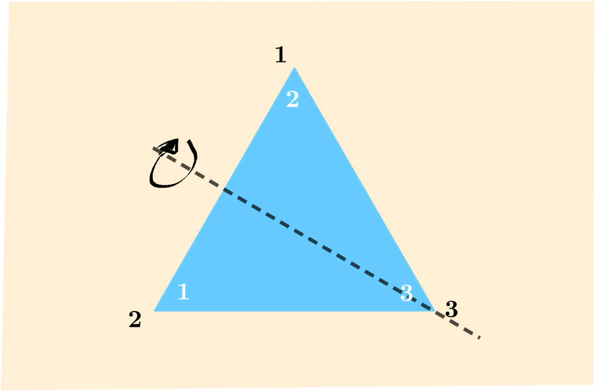
3. We could rotate it by $\frac{2}{3}$ of a turn anti-clockwise.



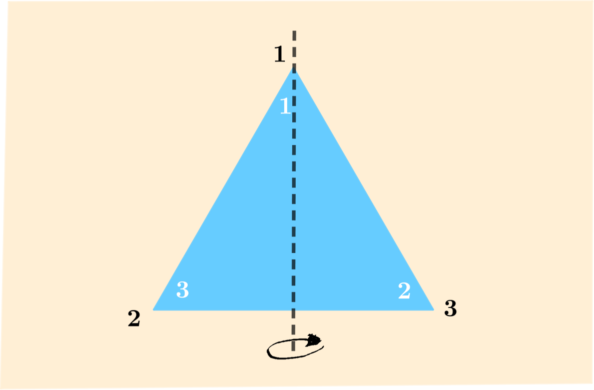
4. We could reflect it in this line (you might prefer to think of this as a 180 degree rotation about this axis given by the line). This means the triangle will be the other way up afterwards and we'll see the blue side.



5. or about this line/axis



6. or about this one



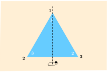
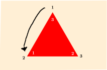
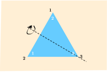
These six are all the possibilities. Here is an argument as to why there can be no more than 6.

For any 'way' the corner labelled 1 can end in one of three 'frame corners'. After that the corner labelled 2 can end in either of the two remaining frame corners. After that the corner labelled 3 will end in the other frame corner.

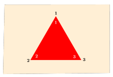
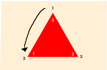
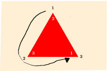
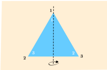
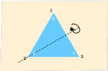
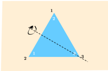
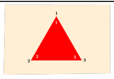
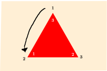
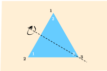
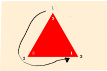
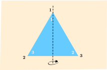
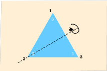
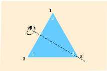
This means there are at most $3 \times 2 \times 1 = 6$ possible 'ways'. Since we have already found six different 'ways' these must be all six of them.

At this point it's probably worth pointing out the the 'ways' are called 'symmetries'. More precisely these are the symmetries of an equilateral triangle.

Finally, and very importantly, think about what would happen if you pick the triangle up from its orginal position, move it according to on of the 'ways', then move the new triangle according to another (possibly the same) 'way'. The triangle is occupying its frame after these two moves so what you have done must be the same as one of the six 'ways'.

Convince yourself that doing this to the triangle  and then doing this to the new triangle  is the same as just doing this to the original triangle .

Considering this for all possible choices of 'do thing one' then 'do thing two' we can start to fill in this table.

		Do this first					
	○						
Do this second							
							
							
							
							
							

The complete table looks like this.

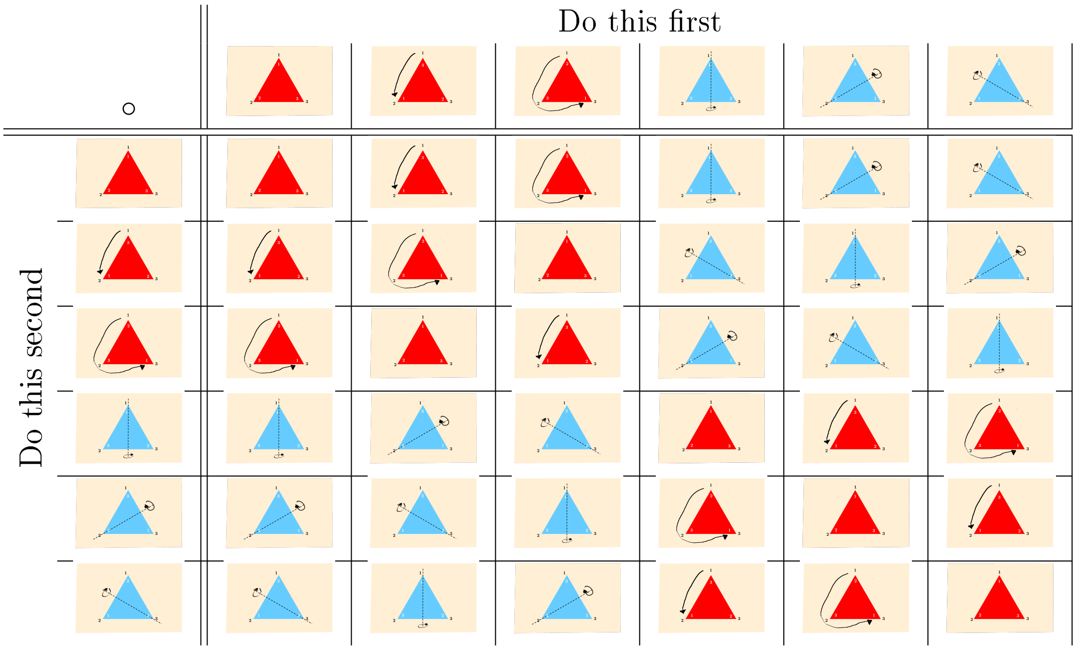


Figure 1: The six symmetries of an equilateral triangle and how any two of them 'combine'.

What did we do?

We took a mathematical object (an equilateral triangle) and we looked at transformations of it that preserve some property (sitting in the frame) and which were 'undo-able' (i.e. we could apply some other transformation to get the object back to its orginal state).

Later we'll see that groups provide a way to study situations like this in an abstract sense.

Before we can get on to groups and rings we need to review some algebraic fundamentals on *sets and functions* and *matrices* in the next two lectures.

Lecture 2 - Sets and Functions

2.1 What is a set?

2.1.1 Definition A *set* is simply a collection of objects.

We use curly brackets to denote sets. For example, if we write

$$A = \{2, 5, 13\},$$

then we're saying that the set A consists of the elements 2, 5, 13. This is one way of specifying a set; we simply list all its elements between curly brackets. The notation $x \in S$ means *x is a member of the set S* and the notation $x \notin S$ means *x is not a member of the set S*. For the set A above, we know $13 \in A$ but $11 \notin A$. We can also specify some infinite sets in this fashion; for example, the set of all integers

$$\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}.$$

This is absolutely standard notation: when you see $\mathbb{Z}$, you're expected to know that it's the set of integers. The set of natural numbers is

$$\mathbb{N} = \{0, 1, 2, 3, 4 \dots\}.$$

Again this is standard notation (but not all mathematicians include 0 in the natural numbers).i

Here is an example of another way of specifying a set:

$$B = \{x \in \mathbb{Z} \mid x^2 = 16\}.$$

This is saying that B is the set of all integers x satisfying the equation $x^2 = 16$. Of course, another way of specifying the same set would be to write $B = \{-4, 4\}$.

If we write

$$C = \{x \in \mathbb{N} \mid x^2 = 16\},$$

then $C = \{4\}$.

To get more practice with this notation, observe that another way of specifying the natural numbers is to write

$$\mathbb{N} = \{x \in \mathbb{Z} \ \mid \ x \geq 0\}.$$

Yet another correct—although admittedly silly way—is to write

$$\mathbb{N} = \{x \in \mathbb{Z} \ \mid \ x \geq -0.5\}.$$

2.2 The empty set

2.2.1 Definition The *empty set* is the set containing no objects. It is denoted by $\emptyset$.

If we write

$$D = \{u \in \mathbb{Z} \mid u^3 = 2\},$$

then D is the set of integers u satisfying $u^3 = 2$. There are no integers satisfying this equation, so D is the empty set. We denote the empty set by $\emptyset$, so we can write $D = \emptyset$.

Here are a couple more examples of empty sets:

$$\{w \in \mathbb{N} \mid w \leq -1\} = \emptyset, \qquad \{v \in \mathbb{Z} \mid 3.01 \leq v \leq 3.99\} = \emptyset.$$

2.3 More sets (and more notation)

Here are some other sets you need to know:

1. $\mathbb{Q}$ is the set of *rational numbers.* We can write this as

$$\mathbb{Q} = \left\{ \frac{a}{b} \mid a, b \in \mathbb{Z}, \quad b \neq 0 \right\}.$$

Examples of elements of $\mathbb{Q}$ are 0, 5, $-7/11$, $3/2$, $6/4$ (the last two being the same element). From *Foundations* or *Sets and Numbers* you should/will know that $\sqrt{2}$ is irrational. You can write this statement in set notation: $\sqrt{2} \notin \mathbb{Q}$. Other examples of irrational numbers are e and π.

2. $\mathbb{R}$ is the set of *real numbers.* It isn't possible to write $\mathbb{R}$ in straightforward way as for the sets above, but you can think of the elements of $\mathbb{R}$ as points on the real line. Examples of elements of $\mathbb{R}$ are -7, $3/5$, 3.85, $\sqrt{7}$, $(\pi+1)/2$, $\sin 5$.

3. $\mathbb{C}$ is the set of *complex numbers.* You have seen complex numbers in your *Further Mathematics* A-Level.Recall that i is a symbol that satisfies $i^2 = -1$. We can write the set of complex numbers as

$$\mathbb{C} = \{ \ a + bi \mid a, b \in \mathbb{R} \ \}.$$

4. We define *Euclidean n-space* as

$$\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) \mid x_1, x_2, \ldots, x_n \in \mathbb{R}\}.$$

Thus $\mathbb{R}^2$ is the set of vectors in the plane, and $\mathbb{R}^3$ is the set of vectors in 3-dimensional space.

Notice that in the above notation vectors are written as rows, for example $(1, 2) \in \mathbb{R}^2$.

You might be more used to writing vectors as columns, so you might write

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

instead of $(x_1, x_2, \ldots, x_n)$. For example,

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

instead of $(1, 2)$.

Just like $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ where we have the four basic arithmetic operations, $\mathbb{R}^n$ has some additional structure defined on it. *Vector addition* is defined by

$$(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) = (x_1 + y_1, x_2, +y_2, \ldots, x_n + y_n)$$

or, in column notation

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}$$

For example, in $\mathbb{R}^3$,

$$(2, 3, -4) + (2, 1, 0) = (4, 4, -4)$$

or, in column notation,

$$\begin{pmatrix} 2 \\ 3 \\ -4 \end{pmatrix} + \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 4 \\ 4 \\ -4 \end{pmatrix}$$

Scalar multiplication is defined as follows. If λ is a scalar (i.e. $\lambda \in \mathbb{R}$) and $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ is a vector, we define

$$\lambda\mathbf{x} = \lambda(x_1, x_2, \ldots, x_n) = (\lambda x_1, \lambda x_2, \ldots, \lambda x_n).$$

For example, in $\mathbb{R}^3$, if $\lambda = 3$ and $\mathbf{x} = (1, 2, 3)$ then

$$3(1, 2, 3) = (3, 6, 9).$$

You learn in *Linear Algebra* about how these two operations give $\mathbb{R}^n$ the structure of a *vector space* over $\mathbb{R}$ ('over $\mathbb{R}$' just means that, in this case, the scalars are real numbers).

2.4 What is a function?

2.4.1 Definitions Given two sets X and Y, a *function*, f, is a rule which associates a unique element $y \in Y$ to each element $x \in X$.

We write $f : X \to Y$. If $y \in Y$ is the unique element of Y associated with $x \in X$ we write $y = f(x)$.

X is called the *domain* of the function and Y is called the *codomain*. The subset $\{f(x) \mid x \in X\}$ of Y is called the *image* of f.

Here are some examples:

1. $f : \mathbb{R} \to \mathbb{R}$ given by the rule $f(x) = 2x + 1$. For this function $f(5) = 11$ and $f(-1) = -1$.

   It's often useful to have a way of visualising a function. In this case, there is the usual the straight line $y = f(x) = 2x + 1$, shown in figure 2.

   Another way to visualise this function, shown in figure 3, is in a diagram where we have the domain on the left, the codomain on the right (both $\mathbb{R}$) in this case and arrows showing the element in the codomain associated with each element in the domain (note: don't be deceived by this picture - the function is defined for all reals, not just the integers).

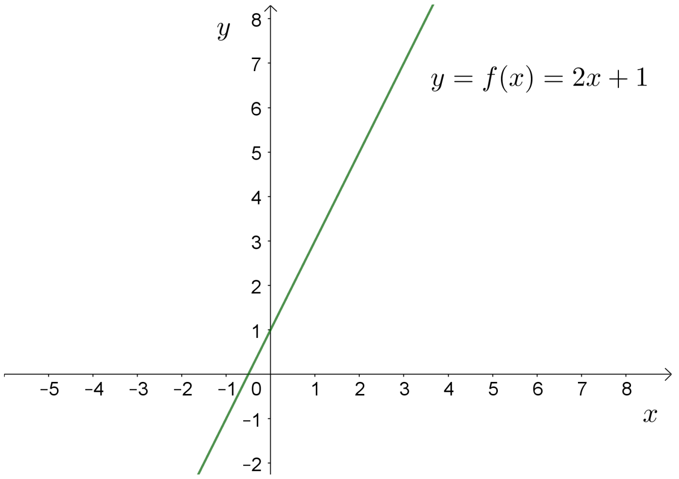


Figure 2: Usual visual depiction of a function as a graph in x-y plane

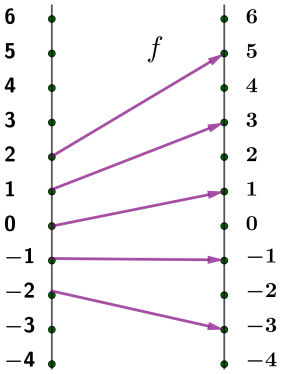


Figure 3: Another depiction of a function where arrows show what is mapped to what.

2. $f : \{1, 2, 3\} \to \{A, B\}$ given by $f(1) = B$, $f(2) = B$, $f(3) = A$

   Figure 4 shows is a way to think of this function as a picture.

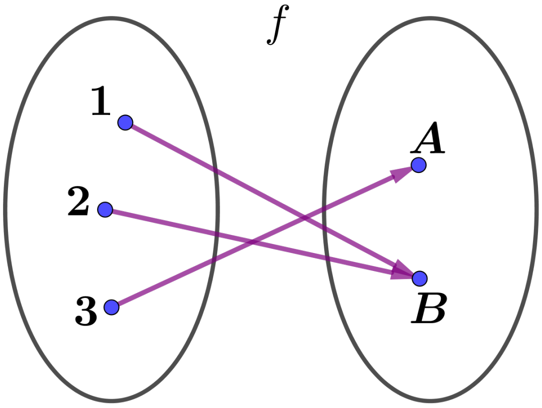


Figure 4: Here arrows show what is mapped to what, 'circles' represent the sets and the 'dots' their elements.

2.5 When are two functions equal?

2.5.1 Definition Let X, Y, A and B be sets and let $f : X \to Y$ and $g : A \to B$ be functions.

We say that $f = g$ if all of the following hold

1. $X = A$

2. $Y = B$

3. $f(x) = g(x)$ for all $x \in X$. ◊

This means that the functions $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ and $g : \mathbb{R} \to \mathbb{R}$ given by $g(x) = |x|^2$ are equal.

If, in addtion, $h : \mathbb{Z} \to \mathbb{R}$ is given by $h(x) = x^2$ then h is not equal to f.

2.6 Composition of functions and inverse functions

2.6.1 Definition Let S_1, S_2 and S_3 be sets and f, g be functions

$$f : S_1 \to S_2, \qquad g : S_2 \to S_3.$$

We can define the *composition* $g \circ f : S_1 \to S_3$ by the rule: $(g \circ f)(x) = g(f(x))$, i.e. $g \circ f$ is the function obtained by substituting f into g.

2.6.2 Example Here is an example of composition of functions.

Let $S_1 = \{1, 2, 3\}$. Let $S_2 = \{A, B, \}$. Let $S_3 = \{\text{apple}, \text{banana}, \text{pear}\}$.

$$f : S_1 \to S_2, \qquad f(1) = B, f(2) = B, f(3) = A$$

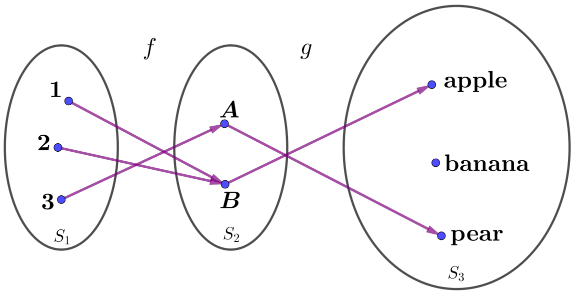
and

$$g : S_2 \to S_3, \qquad g(A) = \text{pear}, g(B) = \text{apple}$$

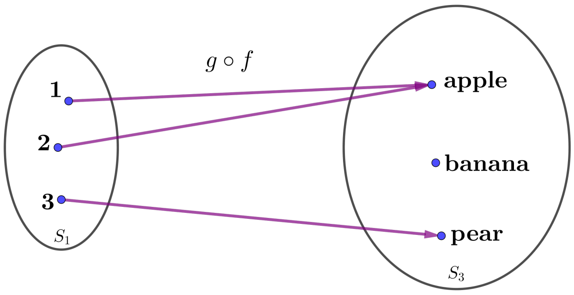
Then

$$\begin{aligned}(g \circ f)(1) = g(f(1)) = g(B) = \text{apple}\\ (g \circ f)(2) = g(f(2)) = g(B) = \text{apple}\\ (g \circ f)(3) = g(f(3)) = g(A) = \text{pear}\end{aligned}$$

You could think of this in terms of pictures.



Following the arrows from the set S_1 on the left to the set S_3 on the right gives the picture for $g \circ f$:



2.6.3 Example Here is another example of composition of functions. Let

$$f : \mathbb{R} \to \mathbb{R}, \qquad f(x) = x^2 - 5$$

and

$$g : \mathbb{R} \to \mathbb{R}, \qquad g(x) = 3x + 2.$$

Notice that, unlike in Example 2.6.2, because the codomain and the domain are both $\mathbb{R}$ for both f and g we can compose these functions both 'ways round':

$$\begin{aligned}(f \circ g)(x) &= f(g(x)) = f(3x+2) = (3x+2)^2 - 5 = 9x^2 + 12x - 1,\\ (g \circ f)(x) &= g(f(x)) = g(x^2 - 5) = 3(x^2 - 5) + 2 = 3x^2 - 13.\end{aligned}$$

The order matters here: $f \circ g$ is the result of substituting g into f, and $g \circ f$ is the result of substituting f into g.

The following lemma might look quite basic, but it one of the most important results we shall meet in this module, and we shall use it again and again.

2.6.4 Lemma Let S_1, S_2, S_3, S_4 be sets and let f, g, h be functions

$$h : S_1 \to S_2, \qquad g : S_2 \to S_3, \qquad f : S_3 \to S_4.$$

Then $f \circ (g \circ h) = (f \circ g) \circ h$.

Proof
Think about what these two functions do to an element $x \in S_1$.

Let's start with $f \circ (g \circ h)$. Here we 'do' $g \circ h$ to x first, this will give us $g(h(x))$. Then we 'do' f to this. So we will end up with $f(g(h(x)))$.

Now think about the effect of $(f \circ g) \circ h)$ on x Here we 'do' h to x first, this will give us $h(x)$. Then we 'do' $f \circ g$ to this, this means doing g to it, then doing f to the result. So we will again end up with $f(g(h(x)))$.

Therefore $[f \circ (g \circ h)](x) = f(g(h(x))) = [(f \circ g) \circ h](x)$ for all $x \in S_1$ and this is what it means to say that $f \circ (g \circ h)$ and $(f \circ g) \circ h$ are equal functions. $\Diamond$

2.6.5 Definition Let X and Y be sets. Let $f : X \to Y$ be a function. Suppose there exists a function $g : Y \to X$ such that $(g \circ f)(x) = x$ for all $x \in X$ and $(f \circ g)(y) = y$ for all $y \in Y$. Then f is said to be *invertible* and g is said to be the *inverse function* to f.

Note that then g is also invertible and f is the inverse function to g by the symmetry in the definition. Also it is usual to then write g as f^{-1}.

2.7 Injective, surjective and bijective functions

2.7.1 Definitions Let X and Y be sets. The function $f : X \to Y$ is said to be *injective* or *one-to-one* if whenever $x_1, x_2 \in X$ with $x_1 \neq x_2$ then $f(x_1) \neq f(x_2)$. Note that this is equivalent to saying that if $x_1, x_2, \in X$ and $f(x_1) = f(x_2)$ then $x_1 = x_2$.

The function $f : X \to Y$ is said to be *surjective* or *onto* if for all $y \in Y$ there exists $x \in X$ such that $f(x) = y$.

The function $f : X \to Y$ is said to be *a bijection* it is both injective and surjective.

2.7.2 Theorem Let X and Y be sets. The function $f : X \to Y$ is bijective if and only if it is invertible.

Proof
Notice that this is an 'if and only if' statement and therefore we need to prove it 'from left to right' and 'from right to left'.

Let's start with 'from right to left'. Here we start by assuming that f is invertible and need to deduce, from that starting point, that f is injective.

Since, in this scenario, f is invertible there exists a function $g : Y \to X$ such that $g \circ f(x) = x$ for all $x \in X$ and $f \circ g(y) = y$ for all $y \in Y$. We use this function to show that f is both injective and surjective.

Let's start by showing that f is injective. Suppose we have $x_1, x_2, \in X$ and $f(x_1) = f(x_2)$. Then, by applying g to both sides we get $g(f(x_1)) = g(f(x_2))$ or $(g \circ f)(x_1) = (g \circ f)(x_2)$. But, since g is the inverse of f, we have that $g(f(x_1)) = x_1$ and that $g(f(x_2)) = x_2$. So we have that $x_1 = x_2$. Therefore f is injective.

Now let's show that f is surjective. Let $y \in Y$. Then $g(y) \in X$ and $f(g(y)) = y$. Therefore we have found an element of X, namely $g(y)$ that f maps to y. We can conclude that f is surjective.

So f is a bijection and this concludes the proof 'from right to left'.

Now let's deal with 'from left to right'. Here we start by assuming that f is a bijection and we need to deduce, from that starting point that f has an inverse.

We need to define what a function $g : Y \to X$ which will be the inverse to f. This means, given a $y \in Y$ we need to say what g does to y, in other words we need to specify $g(y)$. Since f is both injective and surjective, there is a unique element $x \in X$ such that $f(x) = y$. This is the element g is going to send y to. So define $g(y) = x$ (notice that it's important that x is unique in the above to be able to do this).

Let's now show that g is the inverse of f. Let $x \in X$. Then $f(x) \in Y$ and by the way that f is defined $g(f(x)) = x$, i.e. $(g \circ f)(x) = x$.

Now let $y \in Y$. Again by the way that g is construction $f(g(y)) = y$ i.e. $(f \circ g)(y) = y$. These two facts mean that g is the inverse of f and f is invertible. $\Diamond$

Note that you will cover this material again in *Foundations* or *Sets and Numbers*. It's so fundamental in mathematics that this is not a bad thing.

Lecture 3 - Matrices

You almost certainly met matrices during A-Levels, and you'll see them again in *Linear Algebra*. In any case you need to know about matrices for this module. This chapter summarizes what you need to know. We'll not cover this chapter in much detail in lectures because most of it should be familar to you.

Even if you think you know all about matrices I advise you to read this chapter carefully: do you know why matrix multiplication is defined the way it is? Do you know why matrix multiplication is associative?

3.1 What are Matrices?

3.1.1 Definition Let m, n be positive integers. An *$m \times n$ matrix* (or a *matrix of size or order $m \times n$*) is a rectangular array consisting of mn numbers arranged in m rows and n columns:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{pmatrix}.$$

◊

3.1.2 Examples Let

$$A = \begin{pmatrix} 1 & -2 & 0 \\ -1 & 7 & 14 \end{pmatrix}, \qquad B = \begin{pmatrix} 3 & -2 \\ -1 & 8 \\ 2 & 5 \end{pmatrix}, \qquad C = \begin{pmatrix} 3 & 1 & 5 \\ -6 & -8 & 12 \\ 2 & 5 & 0 \end{pmatrix}.$$

A, B, C are matrices. The matrix A has size (or order) 2×3 because it has 2 rows and 3 columns. Likewise B has size 3×2 and C has size 3×3. ◊

Writing

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{pmatrix}.$$

wastes a lot of space. It is convenient to abbreviate this matrix by the notation $A = (a_{ij})_{m \times n}$. This means that A is a matrix of size or order $m \times n$ (i.e. m rows and n columns) and that we shall refer to the element that lies at the intersection of the i-th row and j-th column by a_{ij}.

3.1.3 Example Let $A = (a_{ij})_{2\times 3}$. We can write A out in full as

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}.$$

Notice that A has 2 rows and 3 columns. The element a_{12} belongs to the 1st row and the 2nd column. ♢

3.1.4 Definition $M_{m\times n}(\mathbb{R})$ is the set of $m \times n$ matrices with entries in $\mathbb{R}$. We similarly define $M_{m\times n}(\mathbb{C})$, $M_{m\times n}(\mathbb{Q})$, $M_{m\times n}(\mathbb{Z})$, etc. ♢

3.1.5 Example

$$M_{2\times 2}(\mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \;:\; a,\, b,\, c,\, d \in \mathbb{R} \right\}.$$

♢

3.2 Matrix Operations

3.2.1 Definitions Given matrices $A = (a_{ij})$ and $B = (b_{ij})$ of size $m \times n$, we define the **sum** $A + B$ to be the $m \times n$ matrix whose (i, j)-th element is $a_{ij} + b_{ij}$. We define the **difference** $A - B$ to be the $m \times n$ matrix whose (i, j)-th element is $a_{ij} - b_{ij}$.

Let λ be a scalar. We define λA to be the $m \times n$ matrix whose (i, j)-th element is λa_{ij}.

We let $-A$ be the $m \times n$ matrix whose (i, j)-th element is $-a_{ij}$. Thus $-A = (-1)A$. ♢

Note that the sum $A + B$ is defined only when A and B have the same size. In this case $A + B$ is obtained by adding the corresponding elements.

3.2.2 Examples Let

$$A = \begin{pmatrix} 2 & -5 \\ -2 & 8 \end{pmatrix}, \qquad B = \begin{pmatrix} 4 & 3 \\ 1 & 0 \\ -1 & 2 \end{pmatrix}, \qquad C = \begin{pmatrix} -4 & 2 \\ 0 & 6 \\ 9 & 1 \end{pmatrix}.$$

Then $A+B$ is undefined because A and B have different sizes. Similarly $A+C$ is undefined. However $B+C$ is defined and is easy to calculate:

$$B+C = \begin{pmatrix} 4 & 3 \\ 1 & 0 \\ -1 & 2 \end{pmatrix} + \begin{pmatrix} -4 & 2 \\ 0 & 6 \\ 9 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 5 \\ 1 & 6 \\ 8 & 3 \end{pmatrix}.$$

Likewise $A-B$ and $A-C$ are undefined, but $B-C$ is:

$$B-C = \begin{pmatrix} 4 & 3 \\ 1 & 0 \\ -1 & 2 \end{pmatrix} - \begin{pmatrix} -4 & 2 \\ 0 & 6 \\ 9 & 1 \end{pmatrix} = \begin{pmatrix} 8 & 1 \\ 1 & -6 \\ -10 & 1 \end{pmatrix}.$$

Scalar multiplication is always defined. Thus, for example

$$-A = \begin{pmatrix} -2 & 5 \\ 2 & -8 \end{pmatrix}, \qquad 2\,B = \begin{pmatrix} 8 & 6 \\ 2 & 0 \\ -2 & 4 \end{pmatrix}, \qquad 1.5C = \begin{pmatrix} -6 & 3 \\ 0 & 9 \\ 13.5 & 1.5 \end{pmatrix}.$$

3.2.3 Definition The zero matrix of size $m \times n$ is the unique $m \times n$ matrix whose entries are all 0. This is denoted by $0_{m\times n}$, or simply 0 if there is no possibility of confusion. ♢

3.2.4 Definition Let $A = (a_{ij})_{m\times n}$ and $B = (b_{ij})_{n\times p}$. We define the **product** AB to be the matrix $C = (c_{ij})_{m\times p}$ such that

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} + \cdots + a_{in}b_{nj}.$$

♢

Note the following points:

- For the product AB to be defined we demand that the number of columns of A is equal to the number of rows of B.
- The ij-th element of AB is obtained by taking the dot product of the i-th row of A with the j-th column of B.

3.2.5 Example Let

$$A = \begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix}, \qquad B = \begin{pmatrix} 5 & -3 \\ 0 & -2 \end{pmatrix}.$$

Both A and B are 2×2. From the definition we know that $A \times B$ will be a 2×2 matrix. We see that

$$AB = \begin{pmatrix} 1 \times 5 + 2 \times 0 & 1 \times -3 + 2 \times -2 \\ -1 \times 5 + 3 \times 0 & -1 \times -3 + 3 \times -2 \end{pmatrix} = \begin{pmatrix} 5 & -7 \\ -5 & -3 \end{pmatrix}.$$

Likewise

$$BA = \begin{pmatrix} 5 \times 1 + -3 \times -1 & 5 \times 2 - 3 \times 3 \\ 0 \times 1 - 2 \times -1 & 0 \times 2 + -2 \times 3 \end{pmatrix} = \begin{pmatrix} 8 & 1 \\ 2 & -6 \end{pmatrix}.$$

We make a very important observation: $AB \neq BA$ in this example. So **matrix multiplication is not commutative**. $\Diamond$

3.2.6 Example Let A be as in the previous example, and let

$$C = \begin{pmatrix} 2 & 1 & 3 \\ 3 & -4 & 0 \end{pmatrix}.$$

Then

$$AC = \begin{pmatrix} 8 & -7 & 3 \\ 7 & -13 & -3 \end{pmatrix}.$$

However, CA is not defined because the number of columns of C is not equal to the number of rows of A. $\Diamond$

If $m \neq n$, then we can't multiply two matrices in $M_{m\times n}(\mathbb{R})$. However, matrix multiplication is defined on $M_{n\times n}(\mathbb{R})$ and the result is again in $M_{n\times n}(\mathbb{R})$. In other words, multiplication is a binary operation on $M_{n\times n}(\mathbb{R})$.

So, what can go wrong in terms of commutativity?

- Give a pair of matrices A, B, such that AB is defined but BA isn't.
- Give a pair of matrices A, B, such that both AB and BA are defined but they have different sizes.
- Give a pair of matrices A, B, such that AB and BA are defined and of the same size but are unequal.
- Give a pair of matrices A, B, such that $AB = BA$.

3.3 Matrices and transformations

Given a 2×2 matrix, there is a way to define a function from $\mathbb{R}^2$ to itself which has useful properties with respect to the vector addition and scalar multiplication on $\mathbb{R}^2$. Later on in Linear Algebra you learn that the function you get is more usually referred to as a '*transformation*' or, more precisely a '*linear transformation*'.

Suppose you are given $A \in M_{2\times 2}(\mathbb{R})$. In other words, A is a 2×2 matrix with real entries. Maybe $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$. You could use it to define a function $T_A : \mathbb{R}^2 \to \mathbb{R}^2$ as follows (notice the use of 'column' notation for elements of $\mathbb{R}^2$ below rather than (x, y) notation):

$$T_A \begin{pmatrix} x \\ y \end{pmatrix} = A \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

If you have done any work on matrices already, the chances are you will be familiar with the above idea as a 'matrix as a transformation'. You just put the 2×2 matrix on the left of the point or vector in $\mathbb{R}^2$ that you want to apply it to and multiply out.

Some people might refer to A itself as the function (rather than making the distinction between it and the function T_A it is used to define).

Let us look at some examples of these functions T_A for specific matrices A.

3.3.1 Examples of matrix transformations.

Let

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \qquad C = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Then A defines a function $T_A : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T_A(\mathbf{u}) = A\mathbf{u}$. Let us calculate T_A explicitly:

$$T_A \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y \\ x \end{pmatrix}.$$

We note that, geometrically speaking, T_A represents reflection in the line $y = x$.

Similarly $T_B \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2x \\ y \end{pmatrix}$, which geometrically represents stretching by a factor of 2 in the x-direction.

Also $T_C \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ y \end{pmatrix}$. Thus geometrically, T_C represents projection onto the y-axis.

If we choose not to distinguish between the matrix and the function it defines, we would say that A represents reflection in the line $y = x$, B represents stretching by a factor of 2 in the x-direction, and C represent projection onto the y-axis.

3.4 Compostion of transformations given by matrices

Now suppose you have two matrices $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ and $B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$.This gives rise to two functions $T_A : \mathbb{R}^2 \to \mathbb{R}^2$ and $T_B : \mathbb{R}^2 \to \mathbb{R}^2$.

Notice also that the two matrices can be multiplied to give a further 2×2 matrix $C = BA$ (you could also multiply them the other way round to get AB but that won't be important in this discussion). This matrix C gives rise to a further function $T_C : \mathbb{R}^2 \to \mathbb{R}^2$.

The definition of matrix multiplication is exactly what we need to ensure that T_C is the same map as 'T_A followed by T_B' (or $T_B \circ T_A$). Let's check this.

Let's take a vector $\begin{pmatrix} x \\ y \end{pmatrix}$, apply T_A to it and then apply T_B to whatever we get after that. We calculate this as follows

$$T_B \left(T_A \begin{pmatrix} x \\ y \end{pmatrix} \right) = B \left(A \begin{pmatrix} x \\ y \end{pmatrix} \right) = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \left(\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \right).$$

Doing the matrix multiplication in the order specified by the brackets gives:

$$\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} a_{11}x + a_{12}y \\ a_{21}x + a_{22}y \end{pmatrix} = \begin{pmatrix} b_{11}a_{11}x + b_{11}a_{12}y + b_{12}a_{11}x + b_{12}a_{12}y \\ b_{21}a_{11}x + b_{21}a_{12}y + b_{22}a_{21}x + b_{22}a_{22}y \end{pmatrix}.$$

On the other hand

$$C = BA = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} b_{11}a_{11} + b_{12}a_{21} & b_{11}a_{12} + b_{12}a_{22} \\ b_{21}a_{11} + b_{22}a_{21} & b_{21}a_{12} + b_{22}a_{22} \end{pmatrix}$$

Therefore

$$\begin{aligned} T_C \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} b_{11}a_{11} + b_{12}a_{21} & b_{11}a_{12} + b_{12}a_{22} \\ b_{21}a_{11} + b_{22}a_{21} & b_{21}a_{12} + b_{22}a_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\ &= \begin{pmatrix} b_{11}a_{11}x + b_{12}a_{21}x + b_{11}a_{12}y + b_{12}a_{22}y \\ b_{21}a_{11}x + b_{22}a_{21}x + b_{21}a_{12}y + b_{22}a_{22}y \end{pmatrix}. \end{aligned}$$

So $T_B \circ T_A$ does exactly the same as T_C to any vector and we can say that $T_B \circ T_A = T_C$ whenever $C = BA$ (here A, B and C are all 2×2 matrices but this would also work with $n \times n$ matrices which give functions from $\mathbb{R}^n$ to $\mathbb{R}^n$, or indeed if A was an $m \times l$ matrix and so T_A is a function from $\mathbb{R}^l$ to $\mathbb{R}^m$ and B was an $n \times m$ matrix and so T_B is a function from $\mathbb{R}^m$ to $\mathbb{R}^n$, then $C = BA$ is an $n \times l$ matrix and T_C goes from $\mathbb{R}^l$ to $\mathbb{R}^n$). This is why matrix multiplcation is defined as it is.

Now is a good time to revisit the non-commutativity of matrices. Let us see a geometric example of why matrix multiplication is not commutative.

Consider the matrices AB and BA where A, B are the above matrices. As a reminder Let

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}.$$

We saw earlier how the definition of matrix multiplication ensures that $(AB)\mathbf{u} = A(B\mathbf{u})$, where $\mathbf{u} \in \mathbb{R}^2$. This means stretch $\mathbf{u}$ by a factor of 2 in the x-direction, then reflect it in the line $y = x$.

Also $(BA)\mathbf{u} = B(A\mathbf{u})$, which means reflect $\mathbf{u}$ in the line $y = x$ and then stretch by a factor of 2 in the x-direction.

The two are not the same as you can see from Figure 5. Therefore $AB \neq BA$ (if $AB = BA$ they would have the same effect on any vector and 'A followed by B' would be the same as 'B followed by A'). ◊

Remark. Matrices don't give us all possible functions $\mathbb{R}^2 \to \mathbb{R}^2$. As mentioned earlier, and you will see this in *Linear Algebra*, that they give us what are called the *linear transformations*. For now, think about

$$S : \mathbb{R}^2 \to \mathbb{R}^2, \qquad S \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y+1 \end{pmatrix}.$$

Geometrically, S translates a vector by 1 unit in the y-direction. Can we get S from a matrix A?

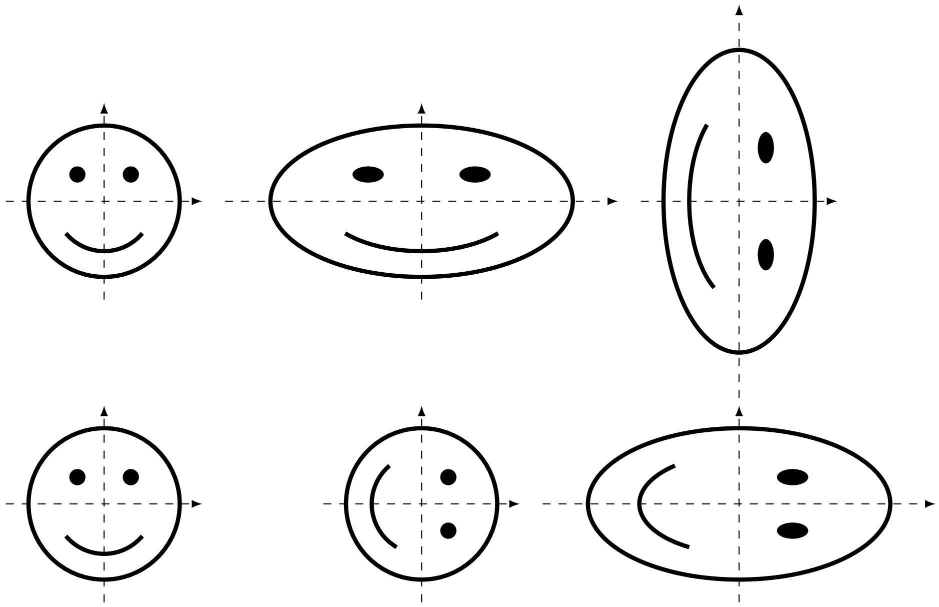


Figure 5: Non-commutativity of matrix multiplication. The matrix A represents reflection in the line $y = x$ and the matrix B represents stretching by a factor of 2 in the x-direction. On the top row we apply B first then A; the combined effect is represented by AB. On the bottom we apply A first then B; the combined effect is represented by BA. It is obvious from comparing the last picture on the top row and the last one on the bottom row that $AB \neq BA$.

Suppose we can, so $S = T_A$ for some matrix A. What this means is that $S\mathbf{u} = T_A\mathbf{u}$ for all $\mathbf{u} \in \mathbb{R}^2$. But $T_A\mathbf{u} = A\mathbf{u}$. So $S\mathbf{u} = A\mathbf{u}$. Now let $\mathbf{u} = \mathbf{0}$. We see that

$$S\mathbf{u} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \qquad A\mathbf{u} = \mathbf{0}$$

which contradicts $S\mathbf{u} = A\mathbf{u}$. So we can't get S from a matrix, and the reason as you'll see in Term 2 is that S is not a linear transformation.

3.5 Why is matrix multiplication associative?

We've seen that if A and B are 2×2 matrices and $\mathbf{u}$ is 2×1 matrix then $(AB)\mathbf{u} = A(B\mathbf{u})$.

This looks like the associative rule for multiplication and is an example of the associated rule for matrix multiplication.

In fact matrix multiplication is associative whenever it is defined.

3.5.1 Theorem Let A be an $m \times n$ matrix, B be an $n \times p$ matrix and C a $p \times q$ matrix, then

$$(AB)C = A(BC). \tag{1}$$

Proof. In terms of functions, (1) is saying

$$(T_A \circ T_B) \circ T_C = T_A \circ (T_B \circ T_C).$$

This holds by Lemma 2.6.4. ♢

Note that in Theorem 3.5.1 T_A, T_B, T_C respectively are functions $\mathbb{R}^n \to \mathbb{R}^m$, $\mathbb{R}^p \to \mathbb{R}^n$, $\mathbb{R}^q \to \mathbb{R}^p$. Also, we have used a 'hidden' result which is that A is the only matrix that gives rise to the function T_A. The following exercise for you to try looks at this.

3.5.2 Exercise Suppose that A and B are matrices and that $T_A = T_B$. Then $A = B$. ♢

When you do *Linear Algebra* in Term 2 you will see a much more computational proof, but the proof above is enlightening too.

3.6 The identity matrix, inverses and determinants

One natural question to ask is what is the multiplicative identity for 2×2 matrices?

You might be wondering what is meant by the *multiplicative identity*? You of course know that $a \cdot 1 = 1 \cdot a = a$ for all real numbers a; we say that 1 is the *multiplicative identity* in $\mathbb{R}$. Likewise the multiplicative identity for 2×2 matrices will be a 2×2 matrix, which we happen to call I_2 (the subscript 2 refers to the , satisfying $AI_2 = I_2A = A$ for all 2×2 matrices A.

Another natural question is given a 2×2 matrix A, what is its *multiplicative inverse* A^{-1}? In other words, if such a thing exists, what is the matrix you can multiply A by to get the result of I_2?

What would be the geometric meaning of the multipicative identity for 2×2 matrices, I_2? In other words what function from $\mathbb{R}^2$ to $\mathbb{R}^2$ should it define (in the sense describes at the start of this section 3.3).

We want I_2 to have the geometric meaning of 'do nothing', as opposed to reflect, stretch, project, etc, or, equivalently, we want the function from $\mathbb{R}^2$ to $\mathbb{R}^2$ it defines to be the one that 'leaves things the same '.

In symbols we want a 2×2 matrix I_2 so that $I_2\mathbf{u} = \mathbf{u}$ for all $\mathbf{u} \in \mathbb{R}^2$. Since I_2 is a 2×2 matrix we can write

$$I_2 = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

where a, b, c, d are numbers. Let us also write

$$\mathbf{u} = \begin{pmatrix} x \\ y \end{pmatrix}.$$

We want

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}.$$

We want this to be true for all values of x, y, because we want the matrix I_2 to mean 'do nothing to all vectors'.

We see that the choices $a = 1$, $b = 0$, $c = 0$, $d = 1$ work. So the matrix

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

has the effect of 'do nothing'. Let's check algebraically that I_2 is a multiplicative identity for 2×2 matrices. What we want to check is that

$$AI_2 = I_2A = A \tag{2}$$

for every 2×2 matrix A. We can write

$$A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}.$$

Now multiplying we find

$$AI_2 = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \alpha \times 1 + \beta \times 0 & \alpha \times 0 + \beta \times 1 \\ \gamma \times 1 + \delta \times 0 & \gamma \times 0 + \delta \times 1 \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = A.$$

In exactly the same way, you can do the calculation to show that $I_2A = A$, so we've established (2).

Before moving on to inverses, it is appropriate to ask in which world does the identity (2) hold? What do I mean by that? Of course A has to be a 2×2 matrix, but are its entries real, complex, rational, integral? If you read the above again, you will notice that we've used properties common to the number systems $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Q}$, $\mathbb{Z}$. So (2) holds for all matrices A in $M_{2\times 2}(\mathbb{R})$, $M_{2\times 2}(\mathbb{C})$, $M_{2\times 2}(\mathbb{Q})$, $M_{2\times 2}(\mathbb{Z})$.

Now what about inverses? Let A be a 2×2 matrix, and let A^{-1} be 'its inverse' whatever that means. If A represents a certain geometric operation then A^{-1} should represent the opposite geometric operation. The matrix A^{-1} should undo the effect of A. The product $A^{-1}A$, which is the result of doing A first then A^{-1}, should now mean 'do nothing'. In other words, we want $A^{-1}A = I_2$ whenever A^{-1} is the inverse of A. Another way of saying the same thing is that if $\mathbf{v} = A\mathbf{u}$ then $\mathbf{u} = A^{-1}\mathbf{v}$.

Should there be such an inverse A^{-1} for every A. No, if $A = 0_{2\times 2}$ then $A^{-1}A = 0_{2\times 2} \neq I_2$. The zero matrix is not invertible, which is hardly surprising. Are there any others? Here it is good to return to the three matrices in Example 3.3.1 and test if they're invertible.

3.6.1 Examples Let

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \qquad C = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Recall that A represents reflection in the line $y = x$. If we repeat a reflection then we end up where we started. So we expect that $A \cdot A = I_2$ (or more economically $A^2 = I_2$). Check this by multiplying. So A is its own inverse.

The matrix B represents stretching by a factor of 2 in the x-direction. It takes a point (or vector if you prefer) an multiplies its x-coordinate (or component if you are thinking in terms of vectors) by 2 whilst leaving its y coordinate/component the same). So its inverse B^{-1} has to represent stretching by a factor of 1/2 (or shrinking by a factor of 2) in the x-direction. We can write

$$B^{-1} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1 \end{pmatrix}.$$

Check for yourself that $B^{-1}B = I_2$. Also note that

$$B^{-1}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x/2 \\ y \end{pmatrix}$$

which does what we want: B^{-1} really is the inverse of B.

Finally recall that C represents projection onto the y-axis. Is there such as thing as unprojecting from the y-axis? Note that

$$C\begin{pmatrix}1\\0\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}, \quad C\begin{pmatrix}2\\0\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}, \quad C\begin{pmatrix}3\\0\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}, \quad C\begin{pmatrix}4\\0\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix},\ldots.$$

Let's assume that C has an inverse and call it C^{-1}. One of the things we want is for $\mathbf{v}=C\mathbf{u}$ to imply $\mathbf{u}=C^{-1}\mathbf{v}$. In other words, C^{-1} is the opposite of C. If there was such an inverse C^{-1} then

$$C^{-1}\begin{pmatrix}0\\0\end{pmatrix}=\begin{pmatrix}1\\0\end{pmatrix}, \quad C^{-1}\begin{pmatrix}0\\0\end{pmatrix}=\begin{pmatrix}2\\0\end{pmatrix}, \quad C^{-1}\begin{pmatrix}0\\0\end{pmatrix}=\begin{pmatrix}3\\0\end{pmatrix}, \quad C^{-1}\begin{pmatrix}0\\0\end{pmatrix}=\begin{pmatrix}4\\0\end{pmatrix},\ldots.$$

This is can't happen! Remember also that a function takes every element in the domain and maps it to a unique element in the codomain so the function T_C can't take $(0,0)$ to all of $(1,0)$, $(2,0)$,$(3,0)$,$(4,0)$.

Therefore, C is not invertible [1]. For a more graphic illustration of this fact, see Figure 6.

The matrix C is non-zero, but it still doesn't have an inverse. This might come as a shock if you haven't seen matrix inverses before. So let's check it in a different way. Write

$$C^{-1}=\begin{pmatrix}a & b\\ c & d\end{pmatrix}.$$

We want $C^{-1}C=I_2$. But

$$C^{-1}C=\begin{pmatrix}a & b\\ c & d\end{pmatrix}\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}=\begin{pmatrix}0 & b\\ 0 & d\end{pmatrix}.$$

We see that no matter what choices of a, b, c, d we make, this will not equal I_2 as the top-left entries don't match. So C is not invertible. $\Diamond$

The next theorem tells us exactly when a 2×2 matrix is invertible.

[1] In *Foundations or Sets and Numbers*, one of the things you'll learn (or have already done) is that a function is invertible if and only if it is bijective. To be bijective a function has to be injective and surjective. We have shown that the 'function' C is not injective, therefore it is not bijective, therefore it is not invertible. If this footnote does not make sense to you yet, return to it at the end of term.

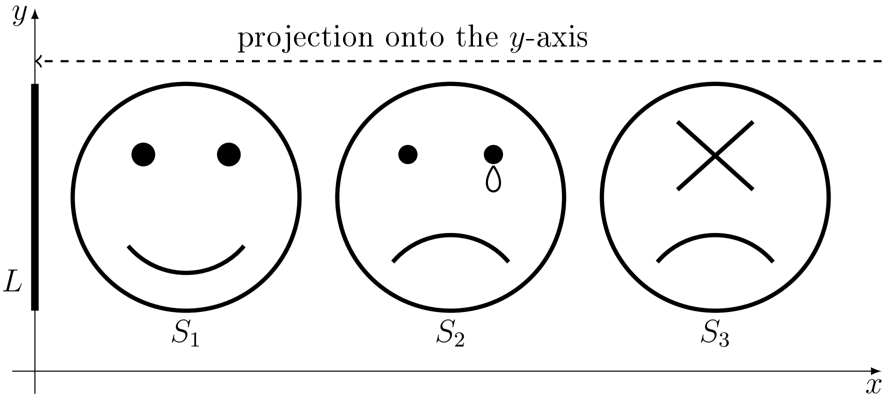


Figure 6: Some non-zero matrices don't have inverses. The matrix C represents projection onto the y-axis. Note that C sends the three 'smileys' S_1, S_2, S_3 to the line segment L. If C had an inverse, would this inverse send L to S_1, S_2 or S_3? We see that C^{-1} does not make any sense!

3.6.2 Theorem A matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible if and only if $ad - bc \neq 0$. Then, the inverse is

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Proof. We'll start by considering the proof from left to right. Suppose that A is invertible. This means there is a 2×2 matrix which we'll call A^{-1} such that $AA^{-1} = I_2 = A^{-1}A$. Suppose the entries in A^{-1} are x, y (top row) and z, w (bottom row) so that

$$A^{-1} = \begin{pmatrix} x & y \\ z & w \end{pmatrix}.$$

We need to show that $ad - bc \neq 0$ and that $A^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$. This means finding x, y, z, w in terms of a, b, c, d. We have

$$\begin{pmatrix} x & y \\ z & w \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Multiplying and equating entries we arrive at four equations:

$$ax + cy = 1 \tag{3}$$
$$bx + dy = 0 \tag{4}$$
$$az + cw = 0 \tag{5}$$
$$bz + dw = 1. \tag{6}$$

We treat the first two equations as simultaneous equations in x and y. Let's eliminate y and solve for x. Multiply the first equation by d, the second by c and subtract. We obtain $(ad-bc)x = d$. By doing similar eliminations you'll find that

$$\begin{cases} (ad-bc)x = d, & (ad-bc)y = -b, \\ (ad-bc)z = -c, & (ad-bc)w = a. \end{cases} \tag{7}$$

Notice that if $ad-bc=0$, we could deduce from the equations (7) that $a=b=c=d=0$ which would mean that $A = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ which is definitely not invertible. This contradiction means that $ad-bc \neq 0$. Then, we have

$$x = \frac{d}{ad-bc}, \quad y = \frac{-b}{ad-bc}, \quad z = \frac{-c}{ad-bc}, \quad w = \frac{a}{ad-bc}.$$

Thus

$$A^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

This proves the theorem from 'left to right'.

Now let's consider the theorem from 'right to left'. Here we start by assuming that $ad-bc \neq = 0$. This means that the matrix $\frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ certainly exists.

Now check by direct multiplication that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and that

$$\frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

This means that

$$\frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

is the inverse of A. Therefore A is invertible and

$$A^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

This completes the proof. $\Diamond$

3.6.3 Definition Let A be a 2×2 matrix and write

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

We define the **determinant** of A, written $\det(A)$ to be

$$\det(A) = ad - bc.$$

Another common notation for the determinant of the matrix A is the following

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

$\Diamond$

From Theorem 3.6.2 we know that a 2×2 matrix A is invertible if and only if $\det(A) \neq 0$.

3.6.4 Theorem (Properties of Determinants) Let A, B be 2×2 matrices.

(a) $\det(I_2) = 1$.

(b) $\det(AB) = \det(A)\det(B)$.

(c) If A is invertible then $\det(A) \neq 0$ and $\det(A^{-1}) = \dfrac{1}{\det(A)}$.

Proof. The proof is mostly left as an exercise. Parts (a), (b) follow from the definition and calculations(make sure you do them). For (c) note that, once (a) and (b) have been proved,

$$\det(A^{-1}A) = \det(I_2) = 1.$$

Now applying (ii) we have $1 = \det(A^{-1})\det(A)$. We see that $\det(A) \neq 0$ and $\det(A^{-1}) = 1/\det(A)$. $\Diamond$

What is the geometric meaning of the determinant? This exercise answers that question.

3.6.5 Exercise Let A be a 2×2 matrix and write

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Let $\mathbf{u} = \begin{pmatrix} a \\ c \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} b \\ d \end{pmatrix}$; in other words, $\mathbf{u}$ and $\mathbf{v}$ are the columns of A. Show that $|\det(A)|$ is the area of the parallelogram with adjacent sides $\mathbf{u}$ and $\mathbf{v}$ (See Figure 7). $\Diamond$

This tells you the meaning of $|\det(A)|$, but what about the sign of $\det(A)$? What does it mean geometrically? Write down and sketch a few examples and see if you can make a guess. Can you prove your guess?

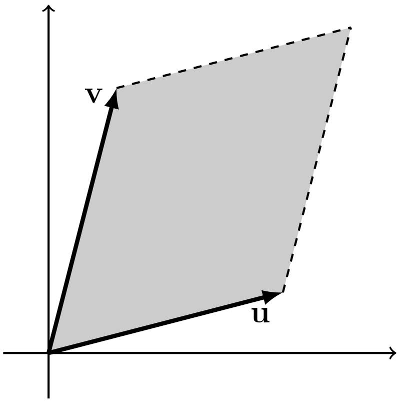


Figure 7: If $\mathbf{u}$ and $\mathbf{v}$ are the columns of A then the shaded area is $|\det(A)|$.

3.6.6 Exercise Suppose $\mathbf{u} = \begin{pmatrix} a \\ c \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} b \\ d \end{pmatrix}$ are non-zero vectors, and let A be the matrix with columns $\mathbf{u}$ and $\mathbf{v}$; i.e. $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Show (algebraically) that $\det(A) = 0$ if and only if $\mathbf{u}$, $\mathbf{v}$ are parallel. Explain this geometrically.

3.6.7 Exercise Let $A = \begin{pmatrix} \alpha & 1 \\ 1 & 1 \end{pmatrix}$. For which values of α is $\det(A^5 - A^4) = -16$? (**Warning: This will be very tricky unless you use the properties of determinants**)

3.7 Matrices and rotations

We saw above some examples of transformations in the plane: reflection, stretching, projection. In this section we take a closer look at rotations about the origin. Let $P = \begin{pmatrix} x \\ y \end{pmatrix}$ be a point in $\mathbb{R}^2$.

Suppose that this point is rotated anticlockwise about the origin through an angle of θ. We want to write down the new point $P' = \begin{pmatrix} x' \\ y' \end{pmatrix}$ in terms of x, y and θ. The easiest way to do this to use polar coordinates.

Let the distance of P from the origin O be r and let the angle $\overrightarrow{OP}$ makes with the positive x-axis be ϕ; in other words the polar coordinates for P are (r, ϕ). Thus

$$x = r\cos\phi, \qquad y = r\sin\phi.$$

Since we rotated P anticlockwise about the origin through an angle θ to obtain P', the polar coordinates for P' are $(r, \phi + \theta)$. Thus

$$x' = r\cos(\phi + \theta), \qquad y' = r\sin(\phi + \theta).$$

We expand $\cos(\phi + \theta)$ to obtain

$$\begin{aligned} x' &= r\cos(\phi + \theta) \\ &= r\cos\phi\cos\theta - r\sin\phi\sin\theta \\ &= x\cos\theta - y\sin\theta. \end{aligned}$$

Similarly

$$y' = x\sin\theta + y\cos\theta.$$

We can rewrite the two relations

$$x' = x\cos\theta - y\sin\theta, \qquad y' = x\sin\theta + y\cos\theta,$$

in matrix notation as follows

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

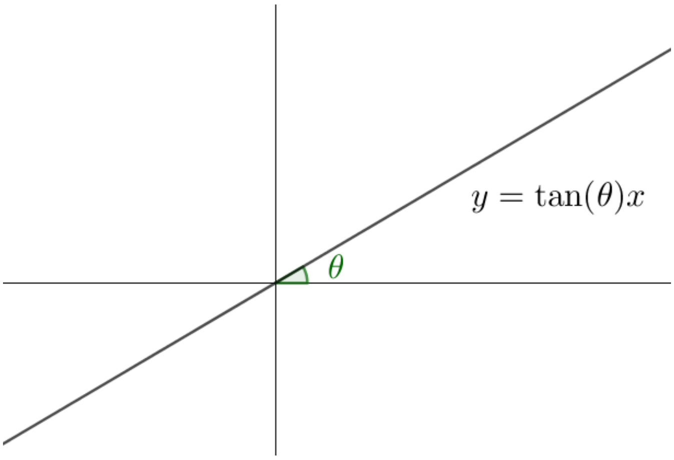
Thus anticlockwise rotation about the origin through an angle θ can be achieved by multiplying by the matrix

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

3.7.1 Exercise You know that R_θ represents anticlockwise rotation about the origin through angle θ. Describe in words the transformation associated to $-R_\theta$. (**Warning: be careful.**) ♢

3.8 Matrices and reflections

The line $y = \tan(\theta)x$ looks like this.



Convince yourself that this is true because the gradient of the line that makes an angle θ with the x-axis like this is $\tan(\theta)$. What is the matrix representing a reflection in this line?

The following exercise will establish this. You will need the following identities on the way

$$\tan(2\theta) = \frac{2\tan(\theta)}{1-\tan^2(\theta)}, \sin(2\theta) = \frac{2\tan(\theta)}{1+\tan^2(\theta)}, \cos(2\theta) = \frac{1-\tan^2(\theta)}{1+\tan^2(\theta)}.$$

3.8.1 Exercise Let $P = \begin{pmatrix} x \\ y \end{pmatrix}$ be a point in $\mathbb{R}^2$. Suppose that $P' = \begin{pmatrix} x' \\ y' \end{pmatrix}$ is the image of P after reflection in the line $y = mx$. In the case $m = \tan(\theta)$, use the fact that the midpoint of PP' is on the line $y = mx$ and the line segment PP' is perpendicular to the line $y = mx$, to show that

$$x' = x\cos(2\theta) + y\sin(2\theta), y' = x\sin(2\theta) - y\cos(2\theta).$$

Conclude that the reflection in the line $y = \tan(\theta)x$ can be achieved by multiplying by the matrix

$$\mathrm{Ref}_\theta = \begin{pmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{pmatrix}.$$

◊

You'll learn much more about matrices in *Linear Algebra.*

Lecture 4 - Binary operations

4.1 What is a binary operation?

4.1.1 Definition Let S be a set. A *binary operation* on S is a rule by which any two elements of S can be combined to give another element of S. $\Diamond$

We are going to use the symbol $\star$ when for binary operations. It's use will mostly be reserved for when we are talking about a general, non-specified, binary operation.

So given $s_1, s_2 \in S$ we have a further element $s_1 \star s_2 \in S$.

4.1.2 Examples

1. Addition is a binary operation on $\mathbb{R}$, because given any two real numbers, their sum is a real number. One way mathematicians like to say this is, "$\mathbb{R}$ *is closed under addition*". All that means is that the sum of two real numbers is a real number. $\Diamond$

2. Addition is also a binary operation on $\mathbb{C}$, $\mathbb{Q}$, $\mathbb{Z}$ and $\mathbb{N}$. Likewise, multiplication is a binary operation on $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$. $\Diamond$

3. Is subtraction a binary operation? This question does not make sense because we haven't specified the set. Subtraction is a binary operation on $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$. Subtraction is not a binary operation on $\mathbb{N}$; for example $1,\ 2 \in \mathbb{N}$ but $1 - 2 = -1 \notin \mathbb{N}$. Thus $\mathbb{N}$ *is not closed under subtraction.* $\Diamond$

4. Is division a binary operation on $\mathbb{R}$? No, because 1, 0 are real numbers but $1/0$ is not defined. Thus $\mathbb{R}$ *is not closed under division.* $\Diamond$

5. Let us define $\mathbb{R}^*$ to be the set of non-zero real numbers:

$$\mathbb{R}^* = \{\, x \in \mathbb{R} \mid \ \neq 0 \,\}.$$

Now division is a binary operation on $\mathbb{R}^*$. But notice that addition is no longer a binary operation on $\mathbb{R}^*$; for example $5,\ -5 \in \mathbb{R}^*$ but $5 + (-5) = 0 \notin \mathbb{R}^*$. $\Diamond$

4.2 Vector operations

Recall that *Euclidean n-space* is defined as

$$\mathbb{R}^n = \{(x_1, x_2, \dots, x_n) \mid x_1, x_2, \dots, x_n \in \mathbb{R}\}.$$

and that in the above notation vectors are written as rows, for example $(1, 2) \in \mathbb{R}^2$ but they can equally be written as columns:

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Vector addition defined by

$$(x_1, x_2, \dots, x_n) + (y_1, y_2, \dots, y_n) = (x_1 + y_1, x_2, +y_2, \dots, x_n + y_n)$$

is then a binary operation on $\mathbb{R}^n$, the result of adding two vectors in $\mathbb{R}^n$ is another vector in $\mathbb{R}^n$.

Vector subtraction, given by

$$(x_1, x_2, \dots, x_n) - (y_1, y_2, \dots, y_n) = (x_1 - y_1, x_2 - y_2, \dots, x_n - y_n),$$

is another binary operation.

What about multiplication by a scalar? Recall that if λ is a scalar (i.e. $\lambda \in \mathbb{R}$) and $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ is a vector, we define

$$\lambda\mathbf{x} = \lambda(x_1, x_2, \dots, x_n) = (\lambda x_1, \lambda x_2, \dots, \lambda x_n).$$

Notice that the result is in $\mathbb{R}^n$, but still multiplication by a scalar is *not* a binary operation on $\mathbb{R}^n$, because we're not 'combining' two elements of $\mathbb{R}^n$, but one element of $\mathbb{R}$ which is λ, and one element of $\mathbb{R}^n$ which is $\mathbf{x}$.

What about the dot product? The dot product is defined on $\mathbb{R}^n$ for all n. If $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$ we define their dot product to be

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

Notice that the result is in $\mathbb{R}$, not $\mathbb{R}^n$, so the dot product is not a binary operation.

What about the cross product (also known as the vector product)? It is defined on $\mathbb{R}^3$ only as follows. Mixing row and column notation (!), if $\mathbf{x} = (x_1, x_2, x_3)$ and $\mathbf{y} = (y_1, y_2, y_3) \in \mathbb{R}^3$ then

$$\mathbf{x} \times \mathbf{y} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \times \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} x_2 y_3 - y_2 x_3 \\ -(x_1 y_3 - y_1 x_3) \\ x_1 y_2 - y1 x_2 \end{pmatrix}$$

So, if $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^3$ then $\mathbf{x} \times \mathbf{y}$ is again in $\mathbb{R}^3$. This means that the cross product is a binary operation on $\mathbb{R}^3$.

4.3 Operations on polynomials as binary operations

4.3.1 Definition $\mathbb{R}[x]$ is the set of polynomials in x with real coefficients. Elements of $\mathbb{R}[x]$ are polynomials and so have the form

$$a_n x^n + a_{n_1} x^{n-1} + a_1 x + a_0$$

where $a_0, a_1, \dots a_n$ are real numbers.

We can vary the set that we take the coefficients coefficients from to get other sets of polynomials.

$\mathbb{C}[x]$ is the set of polynomials in x with complex coefficients, $\mathbb{Q}[x]$ is the set of polynomials in x with rational coefficients, and $\mathbb{Z}[x]$ is the set of polynomials in x with integer coefficients. $\diamondsuit$

Notice that $Z[x] \subseteq \mathbb{Q}[x] \subseteq \mathbb{R}[x] \subseteq \mathbb{C}[x]$.

Addition, subtraction and multiplication of polynomials are defined as you would expect. For example in $\mathbb{Z}[x]$ we have

$$(4x^3 - 2x^2 + 3x + 6) + (x^2 - 10x + 3) = 4x^3 + x^2 - 7x + 9$$

$$(4x^3 - 2x^2 + 3x + 6) - (x^2 - 10x + 3) = 4x^3 - 3x^2 + 13x + 3$$

and

$$(4x^3 - 2x^2 + 3x + 6) \times (x^2 - 10x + 3) = 4\,x^5 - 42\,x^4 + 35\,x^3 - 30\,x^2 - 51\,x + 18$$

This addition, multiplication and subtraction each give a binary operation on each of $Z[z], \mathbb{Q}[x], \mathbb{R}[x], \mathbb{C}[x]$.

You'll be aware that we can create functions by dividing a polynomial by another, so called *rational functions.* For example $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \setminus \{1\} \to \mathbb{R}$ given by

$$f(x) = \frac{x^3+3x+1}{x^2+1}, \; g(x) = \frac{x+2}{x-1}.$$

However $\dfrac{x^3+3x+1}{x^2+1}$ cannot be written as a polynomial. So polynomial division is not a binary operation on any of $Z[z], \mathbb{Q}[x], \mathbb{R}[x], \mathbb{C}[x]$.

4.4 Composition of functions as a binary operation

In example 2.6.3 we started with functions $\mathbb{R} \to \mathbb{R}$ (i.e. with domain and codomain which are both the set of real numbers) and composed them to obtain functions $\mathbb{R} \to \mathbb{R}$. Likewise, in definition 2.6.1, if $S_1 = S_2 = S_3 = S$ say, so that f and g are functions $S \to S$ then $g \circ f$ is a function $S \to S$. In this case (i.e. when the domains and codomains are equal) $\circ$ is a binary operation.

It's easy to get confused about this. Although the set S is involved in this, this is not a binary operation on S, because it doesn't take two elements of S and give us another element. It is a binary operation on the set of functions from S to itself.

4.5 Composition/mutiplication tables

Recall our definition of a binary operation on a set S: it is simply a rule which for any pair of elements of S produces a third "output" element. This binary operation does not have to be 'natural', whatever that means. It does not have to be something we met before, like addition, multiplication, composition of functions, etc. We can simply invent a set S and binary operation on it. If the S is finite, this is easy by means of a *composition/mutiplication table* which tells us for any pair of elements of S what the output element is.

Let $S = \{a, b, c\}$. Let $\star$ be the binary operation on S with the following composition/mutiplication table:

$\star$	a	b	c
a	b	c	a
b	a	c	a
c	b	b	c

The result of the composition $a \star b$, is found at the intersection of the row headed by a with the column headed by b. In other words, for composition tables, the first element determines the row and the second determines the column. Thus for the composition table above,

$$a \star b = c, \qquad b \star a = a, \qquad c \star b = b, \qquad a \star a = b, \dots.$$

You might think that this example is somewhat contrived, and you'd be right. But later on we'll meet more natural composition tables that arise from studying groups, permutations, etc. ♢

4.6 Commutativity and associativity

4.6.1 Definitions Let S be a set and $\circ$ a binary operation.

We say that the binary operation $\star$ is *commutative on S* if $a \star b = b \star a$ for all a, $b \in S$.

We say that the binary operation $\star$ is *associative on S* if $(a \star b) \star c = a \star (b \star c)$ for all a, b, $c \in S$. ♢

Let's consider this for some of the binary operations we've met already:

4.6.2 Examples

1. Addition and multiplication on $\mathbb{R}$ (or $\mathbb{C}$ or $\mathbb{R}[x]$ or …) are both commutative and associative. When operations are commutative and associative, order and bracketing do not matter (though it's suprisingly tricky to give a formal proof of this, we'll not give one here):

   $$e + ((c + b) + (d + a)) = a + b + c + d + e, \qquad e \cdot ((c \cdot b) \cdot (d \cdot a)) = a \cdot b \cdot c \cdot d \cdot e.$$

   Of course subtraction is neither commutative nor associative (write some examples). ♢

2. Addition is commutative and associative on $\mathbb{R}^n$. The cross product is not commutative on $\mathbb{R}^3$. You should know that if $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^3$ then

   $$\mathbf{y} \times \mathbf{x} = -\mathbf{x} \times \mathbf{y}.$$

   We say that the cross product is *anti-commutative.* ♢

3. Let $S = \{a, b, c\}$ and let $\star$ be the binary operation given by the composition table we looked at in section 4.5. Then $\star$ is not commutative; for example

   $$a \star b = c, \qquad b \star a = a.$$

It is also not associative; for example

$$(a \star b) \star c = c \star c = c, \qquad a \star (b \star c) = a \star a = b.$$

$\Diamond$

4. Composition of functions from a set A to itself is associative but not commutative. We know that it is associative from Lemma 2.6.4. We know that it isn't commutative by Example 2.6.3. When a binary operation is associative bracketing doesn't matter. For example,

$$(a \star b) \star ((c \star d) \star e) = (a \star (b \star c)) \star (d \star e).$$

As long as we keep a, b, c, d, e in the same order from left to right, then the order in which we do the compositions does not matter. Thus there would be no ambiguity in writing

$$(a \star b) \star ((c \star d) \star e) = a \star b \star c \star d \star e.$$

This fact that bracketing doesn't matter as long as we keep the same order is called the general associativity theorem. For a proper formulation and proof see

`https://proofwiki.org/wiki/General_Associativity_Theorem/Formulation_2/Proof_1` $\Diamond$

5. Are there binary operations that are commutative but not associative? Yes but it isn't easy to come up with 'natural' examples. However it is easy to invent a finite set and a composition table that is commutative but not associative. Let $S = \{a, b, c\}$. Let $\star$ be the binary operation on S with the following composition table:

$\star$	a	b	c
a	b	c	a
b	c	c	a
c	a	a	c

Note that $\star$ is commutative; you can see this by noting that the table is symmetric about the diagonal from the top left corner to the bottom right corner. But it isn't associative. For example,

$$(b \star c) \star a = a \star a = b, \qquad b \star (c \star a) = b \star a = c.$$

$\Diamond$

4.6.3 Exercise In the following, is $\circ$ a binary operation on A? If so, is it commutative? Is it associative? In each case justify your answer.

(a) $A = \mathbb{R}$ is the set of real numbers and $a \star b = a/b$.

(b) $A = \{1, 2, 3, 4, \dots\}$ is the set of positive integers and $a \star b = a^b$.

(c) $A = \{\dots, 1/8, 1/4, 1/2, 1, 2, 4, 8, \dots\}$ is the set of powers of 2 and $a \circ b = ab$.

(d) $A = \mathbb{C}$ is the set of complex numbers and $a \star b = |a - b|$.

Lecture 5 - Groups

5.1 The Definition of a Group

A *group* is a pair $(G, \star)$ where G is a set and $\star$ is a binary operation on G, such that the following four properties hold:

(i) (closure) for all a, $b \in G$, $a \star b \in G$;

(ii) (associativity) for all a, b, $c \in G$,

$$a \star (b \star c) = (a \star b) \star c;$$

(iii) (existence of the identity element) there is an element $e \in G$ such that for all $a \in G$,

$$a \star e = e \star a = a;$$

(iv) (existence of inverses) for every $a \in G$, there is an element $b \in G$ (called the inverse of a) such that

$$a \star b = b \star a = e.$$

$\Diamond$

If $\star$ is a binary operation then (i) automatically holds. So why did is it listed in the definition? It's there for good measure! When you suspect an operation gives you a group the first thing you should check is that the operation is really a binary operation.

99% of mathematicians call (i)–(iv) the "group axioms"even through they are the "defining properties of a group" . This could be considered to be a bit odd, the word axiom is usually reserved for statements of 'universal truth'.

5.2 First Examples (and Non-Examples)

5.2.1 Example $(\mathbb{R}, +)$ is a group. We know already that addition is a binary operation on $\mathbb{R}$, so 'closure' holds. We know addition of real numbers is associative. What is the identity element? We want an element $e \in \mathbb{R}$ so that $a + e = e + a = a$ for all $a \in \mathbb{R}$. It is clear that $e = 0$ works and is the only possible choice. Moreover, the (additive) inverse of a is $-a$: $a + (-a) = (-a) + a = 0$. $\Diamond$

5.2.2 Example Recall our definition of the natural numbers:

$$\mathbb{N} = \{0, 1, 2, \dots\}.$$

Is $(\mathbb{N}, +)$ a group? Conditions (i), (ii) are satisfied. For condition (iii) we can take the identity element to be 0 (again the only possible choice). But (iv) does not hold. For example, if we take $a = 1$, there is no $b \in \mathbb{N}$ such that $a + b = b + a = 0$. Thus $(\mathbb{N}, +)$ is not a group. ◊

5.2.3 Example $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$ and $(\mathbb{C}, +)$ are groups. ◊

5.2.4 Example Recall we defined

$$\mathbb{R}^* = \{\alpha \in \mathbb{R} : \alpha \neq 0\}.$$

Then $(\mathbb{R}^*, \cdot)$ is a group, where $\cdot$ means multiplication. Again closure and associativity are obvious. If e is the identity element then it has to satisfy $\alpha \cdot e = e \cdot \alpha = \alpha$ for all $\alpha \in \mathbb{R}$. Thus $e = 1$ and this is the only choice possible. Then the inverse of α is α^{-1}.

We can define $\mathbb{C}^*$ and $\mathbb{Q}^*$ in the same way and obtain groups $(\mathbb{C}^*, \cdot)$ and $(\mathbb{Q}^*, \cdot)$.

Can we obtain from $\mathbb{Z}$ a group with respect to multiplication? In view of the above, the obvious candidate is

$$U = \{\alpha \in \mathbb{Z} : \alpha \neq 0\}.$$

But $(U, \cdot)$ is not a group. It is true that (i), (ii) and (iii) hold with 1 being the identity element. But, for example, $2 \in U$ does not have an inverse: there is no $b \in U$ such that $b \cdot 2 = 2 \cdot b = 1$. So $(U, \cdot)$ is not a group. But the answer is not no; all we've done is shown that the obvious choice for a group $(\mathbb{Z}^*, \cdot)$ made up of integers does not work. We'll return to this question and answer it fully later. ◊

5.2.5 Example $(\mathbb{R}^2, +)$ is a group. Let's prove this. We're allowed to assume the usual properties of the real numbers (see Section 4.6.2). The elements of $\mathbb{R}^2$ are pairs (a_1, a_2) where a_1, a_2 are real numbers. Addition is defined by

$$(a_1, a_2) + (b_1, b_2) = (a_1 + b_1, a_2 + b_2).$$

Note that the entries $a_1 + b_1$ and $a_2 + b_2$ are real numbers, and so $(a_1 + b_1, a_2 + b_2)$ is a pair of real numbers. Hence $(a_1 + b_1, a_2 + b_2)$ is in $\mathbb{R}^2$. In other words, $\mathbb{R}^2$ is closed under addition, which shows that $(\mathbb{R}^2, +)$ satisfies condition (i). Next we want to prove associativity of addition. Consider $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ in $\mathbb{R}^2$. We can write

$$\mathbf{a} = (a_1, a_2), \qquad \mathbf{b} = (b_1, b_2), \qquad \mathbf{c} = (c_1, c_2).$$

Here a_1, a_2, b_1, b_2 and c_1, c_2 are real numbers. Note that

$$(\mathbf{a} + \mathbf{b}) + \mathbf{c} = ((a_1 + b_1) + c_1, (a_2 + b_2) + c_2) \, .$$

Likewise,

$$\mathbf{a} + (\mathbf{b} + \mathbf{c}) = (a_1 + (b_1 + c_1), a_2 + (b_2 + c_2)) \, .$$

Because addition of real numbers is associative, we know that

$$(a_1 + b_1) + c_1 = a_1 + (b_1 + c_1), \qquad (a_2 + b_2) + c_2 = a_2 + (b_2 + c_2).$$

Hence

$$(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c}).$$

This shows that $(\mathbb{R}^2, +)$ satisfies (ii).

Next we need an identity element, and the obvious candidate is $\mathbf{0} = (0, 0)$. Then

$$(a_1, a_2) + (0, 0) = (a_1 + 0, a_2 + 0) = (a_1, a_2),$$

and

$$(0, 0) + (a_1, a_2) = (0 + a_1, 0 + a_2) = (a_1, a_2).$$

Thus (iii) is satisfied.

Finally we want an inverse. If $\mathbf{a} = (a_1, a_2)$ is in $\mathbb{R}^2$ then the inverse we choose (there's no other choice) is $\mathbf{b} = (-a_1, -a_2)$. This is in $\mathbb{R}^2$ and satisfies

$$\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a} = (0, 0).$$

Hence (iv) is satisfied and so $(\mathbb{R}^2, +)$ is a group. $\Diamond$

What matters is that you realize that the properties of addition in $\mathbb{R}^2$ simply follow from the definition of addition in $\mathbb{R}$ and corresponding properties of the real numbers.

The proofs for the following examples 5.2.6, 5.2.7, and 5.2.8 are similar.

5.2.6 $(\mathbb{R}^n, +)$ is a group for any $n \geq 2$. $\Diamond$

5.2.7 $(\mathbb{R}[x], +)$ is a group. $\diamondsuit$

5.2.8 $(M_{m\times n}(K), +)$ are groups for $K = \mathbb{C}$, $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{Z}$, with $0_{m\times n}$ the identity element. $\diamondsuit$

5.2.9 All the groups we have met so far are infinite. Here is an example of a finite group. Let $A = \{+1, -1\}$. Then $(A, \cdot)$ is a group (where of course $\cdot$ is multiplication). $\diamondsuit$

5.2.10 Let $B = \{1, i, -1, -i\}$, where $i = \sqrt{-1}$. Then $(B, \cdot)$ is another example of a finite group. $\diamondsuit$

5.2.11 Let $C = \{1, i\}$. Then $(C, \cdot)$ is not a group since it isn't closed; for example $i \cdot i = -1 \notin C$. $\diamondsuit$

5.3 Abelian Groups

5.3.1 Definition We say that a group $(G, \star)$ is *abelian* if (in addition to properties (i)–(iv) in definition 5.1) it also satisfies

(v) (commutativity) for all a, $b \in G$,

$$a \star b = b \star a.$$

$\diamondsuit$

All the groups we have seen above are actually abelian: $(\mathbb{R}, +)$, $(\mathbb{C}, +)$, $(\mathbb{Q}, +)$, $(\mathbb{R}[x], +)$, $(\mathbb{R}^n, +)$, $(\mathbb{R}^*, \cdot)$, $(\mathbb{C}^*, \cdot)$, $(M_{m\times n}(\mathbb{R}), +)$, ...

Are there any non-abelian groups? There are many, but perhaps not ones that you're used to thinking about. In the next section we give an example of a non-abelian group.

5.4 D4 - the symmetries of a square

In many ways the examples above are misleading for three reasons:

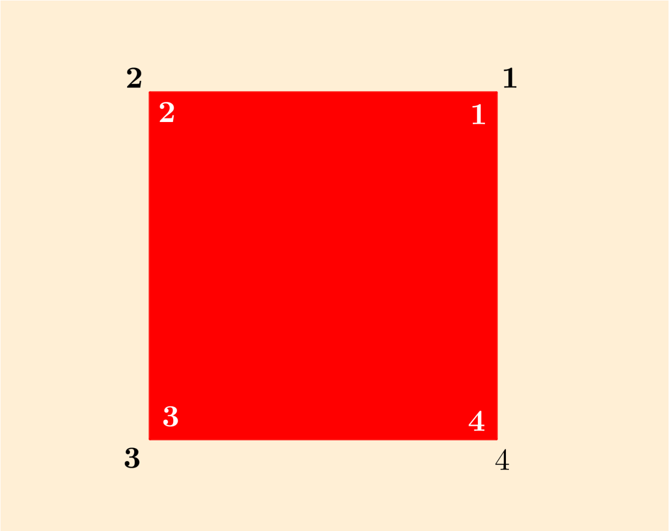
- Most of the examples of groups we have met above have additional structure. For example, in $\mathbb{R}$ we can add, but we can also multiply and we can divide by non-zero numbers.

In fact $\mathbb{R}$ is an example of a *field*. Like in $\mathbb{R}^2$ we have addition and scalar multiplication, so $\mathbb{R}^2$ is an example of a *vector space*. This doesn't stop $(\mathbb{R}, +)$ and $(\mathbb{R}^2, +)$ from being groups, but if you want to test your own ideas in group theory, it is best to also look at examples where there aren't any of these additional structures.

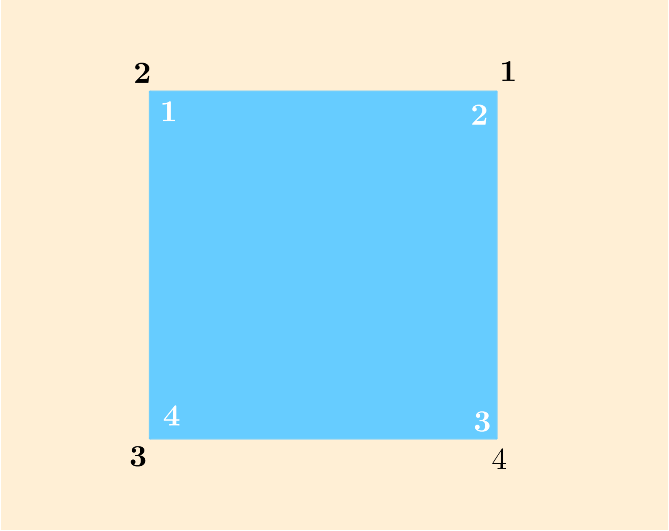
- The groups above are abelian. The theory of abelian groups is rather close in flavour to linear algebra. Many of the most interesting groups that you'll come across during your degree will be non-abelian.
- All the groups above, except for Example 5.2.9, are infinite. Although infinite groups are important and interesting, most theorems we will do in this course will apply only to finite groups. Thus it is essential to become familiar with examples of finite groups.

Here is a great example of a group!

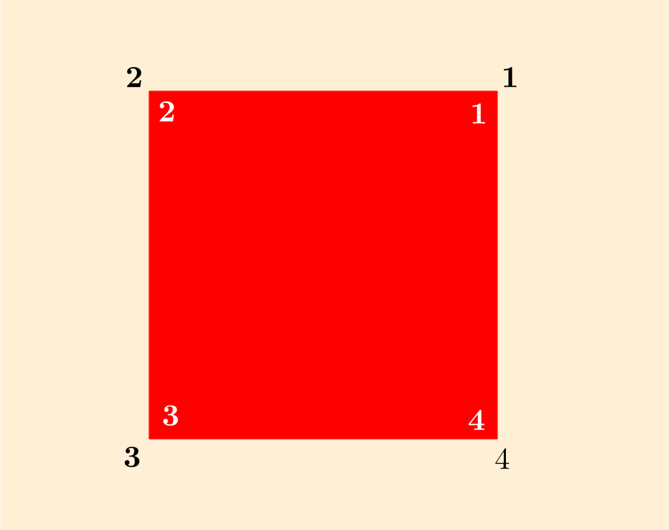
Imagine the following (this might ring a bell); a square made of red paper fitting exactly into a yellow paper frame.



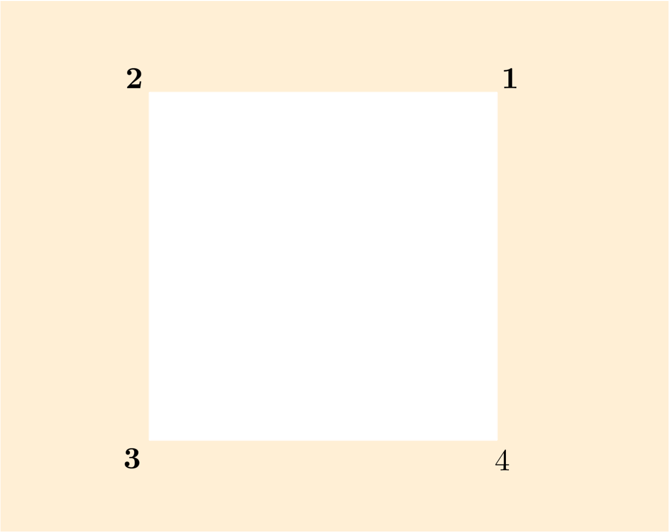
On the back it looks like this.



Let’s turn it back over again.

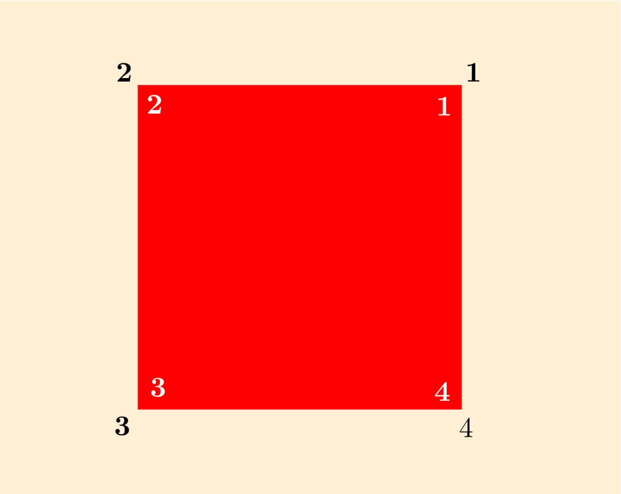


and lift the square out of its frame.

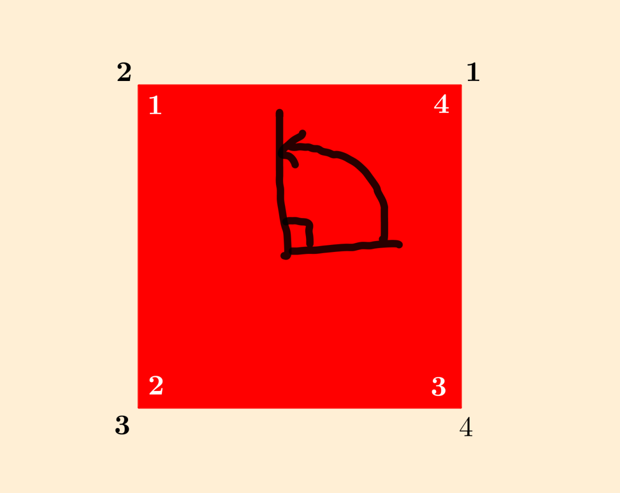


How many ways can we put the square back into its frame so that it occupies the frame exactly? Let's count them.

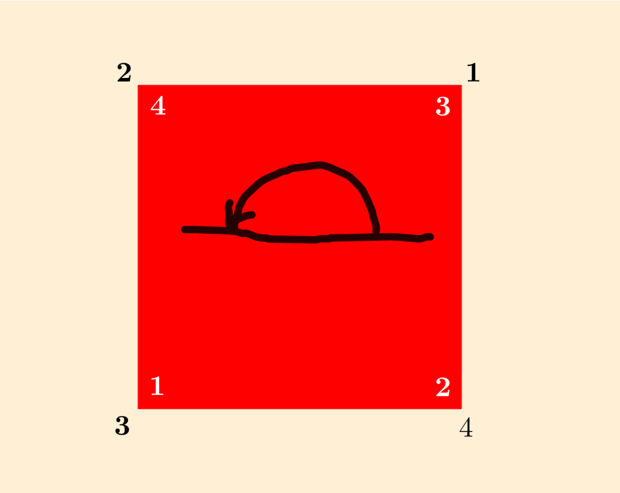
1. We could put it straight back down, exactly in the position it was in when we picked it up:



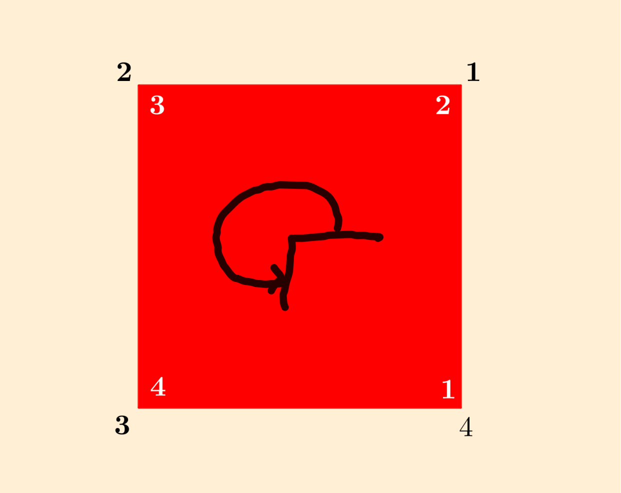
2. We could rotate it by $\frac{1}{4}$ of a turn anticlockwise.



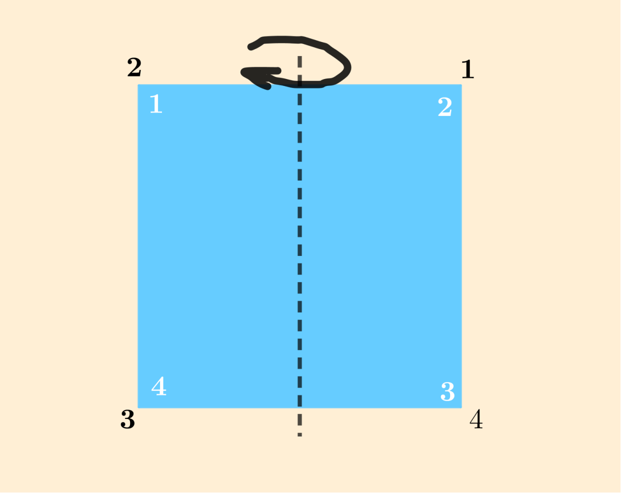
3. We could rotate it by $\frac{1}{2}$ of a turn anitclockwise.



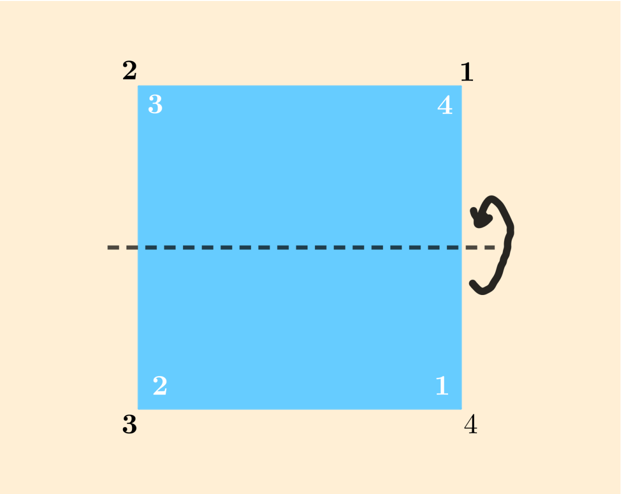
4. We could rotate it by $\frac{3}{4}$ of a turn anitclockwise.



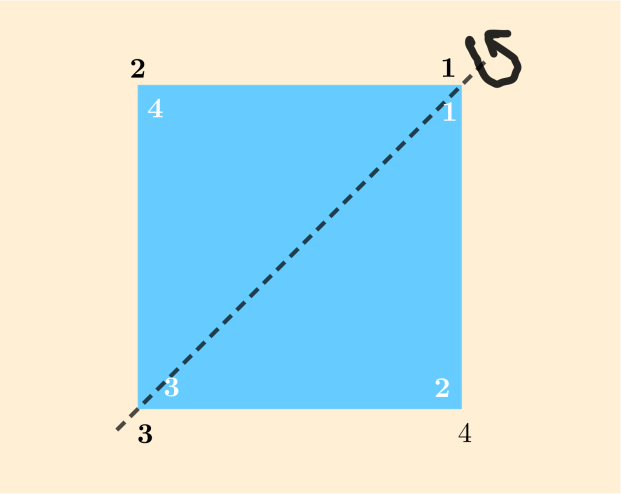
5. We could reflect it in this line (you might prefer to think of this as a 180 degree rotation about this axis given by the line). This means the square will be the other way up afterwards and we'll see the blue side.



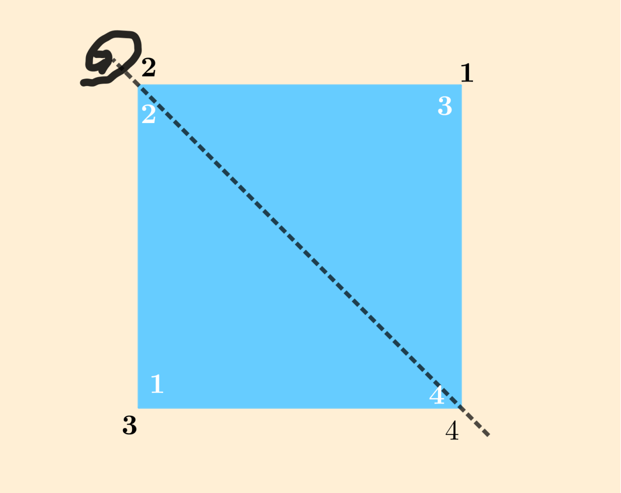
6. or about this line/axis



7. or about this one



8. or this one

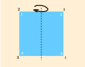
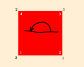
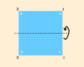


These eight are all the possibilities.

Here is an argument as to why there can be no more than 8 'ways'. For any 'way' the corner labelled 1 can end in one of four 'frame corners'. After that the corner labelled 2 can end in either of the two corners adjacent to the one the the corner labelled 1 landed in.After that the new positions of corner 3 and corner 4 are completely determined.

This means there are at most $4 \times 2 = 8$ possible 'ways'. Since we have already found eight different 'ways' these must be all eight of them.

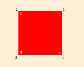
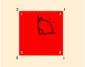
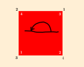
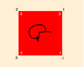
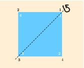
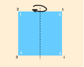
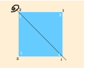
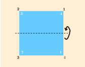
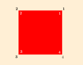
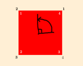
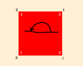
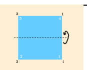
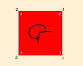
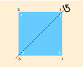
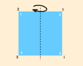
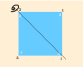
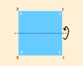
Finally, and very importantly, think about what would happen if you pick the square up from its orginal position, move it according to on of the 'ways', then move the newly positioned square according to another (possibly the same) 'way'. The square is occupying its frame after these two moves so what you have done must be the same as one of the eight 'ways'.

Convince yourself that doing this to the triangle  and then doing this to the new triangle  is the same as just doing this to the original triangle .

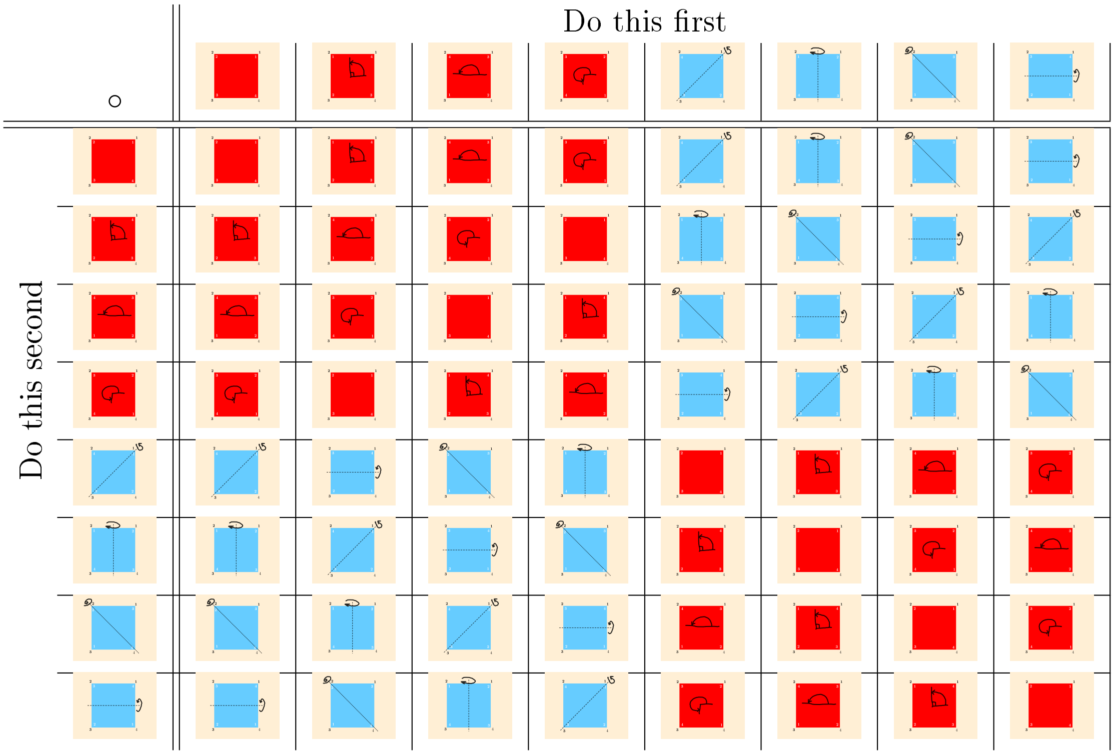
Considering this for all possible choices of 'do thing one' then 'do thing two' we can start to fill in this table.

Do this first

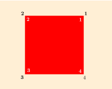
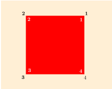
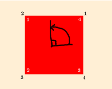
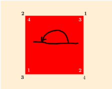
Do this second

∘								
								
								
								
								
								
								
								
								

The complete table looks like this.


Do this first
Do this second
∘

You may be thinking that at this point we really need some notation to save drawing the pictures representing the 'moves' every time!

Let ρ_0, ρ_1, ρ_2, ρ_3 be anticlockwise rotations of the square about O by 0°, 90°, 180° and 270°. So these are , ,  and  respectively.

Let's give the reflections names too. As in Figure 8:

- σ_0 the reflection about the diagonal joining the top-right vertex to the bottom-left vertex;
- σ_1 the reflection about the line joining the midpoint of top side and the midpoint of bottom side;
- σ_2 the reflection about the diagonal joining top-left vertex and the bottom-right vertex;
- σ_3 the reflection about the line joining the midpoint of the left side and the midpoint of the right side.

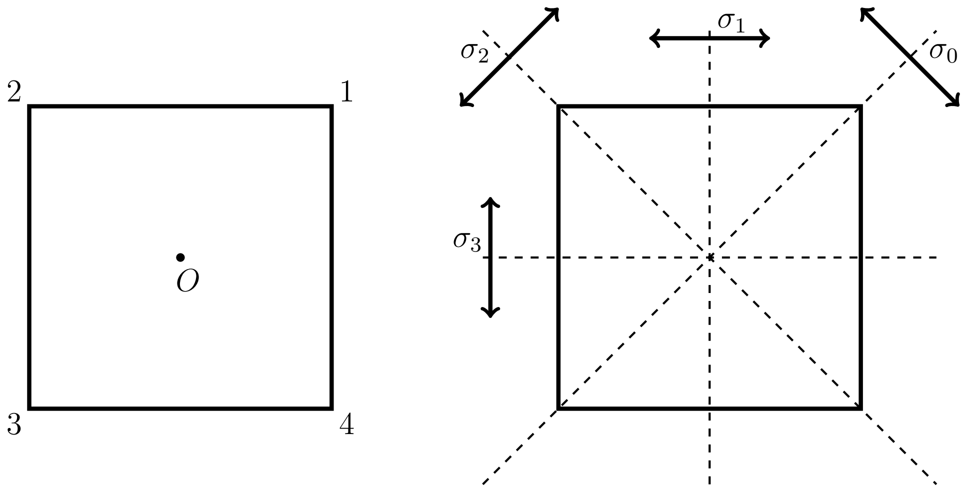


Figure 8: Left: the square with vertices labelled 1, 2, 3, 4. Right: the reflections σ_0, σ_1, σ_2, σ_3.

These are the 'symmetries of a square' and put into a set they look like this.

$$D_4 = \{\rho_0, \rho_1, \rho_2, \rho_3, \sigma_0, \sigma_1, \sigma_2, \sigma_3\}.$$

We talked about a group of symmetries, so it is not enough to just list the symmetries, but we have to specify a binary operation.

We've discussed that the compostion/multiplication is 'do the first move, then do the second move'. Notice that if we think of the moves as functions then this is just compostion of functions. (In fact you could put the square in the x, y plane with its centre at $(0, 0)$ and write down the matrix which corresponds to each of the moves.)

We need to be clear about how the notation works here. If $\alpha, \beta \in D_4$ then $\alpha \circ \beta$ means the symmetry which is "apply β first then α" (not the other way round). This might feel a bit strange. The reason for this choice is that we want to sometimes think of the elements of D_4 as functions, and when we do that we want composition in D_4 to agree with the usual composition of functions. Recall that $f \circ g$ means apply g first then f.

Now we can write out a composition/multiplication table (remember: $\alpha \circ \beta$ means you take α from the left column of 'row headings' and β from the top row of 'column headings'):

$\circ$	ρ_0	ρ_1	ρ_2	ρ_3	σ_0	σ_1	σ_2	σ_3
ρ_0	ρ_0	ρ_1	ρ_2	ρ_3	σ_0	σ_1	σ_2	σ_3
ρ_1	ρ_1	ρ_2	ρ_3	ρ_0	σ_1	σ_2	σ_3	σ_0
ρ_2	ρ_2	ρ_3	ρ_0	ρ_1	σ_2	σ_3	σ_0	σ_1
ρ_3	ρ_3	ρ_0	ρ_1	ρ_2	σ_3	σ_0	σ_1	σ_2
σ_0	σ_0	σ_3	σ_2	σ_1	ρ_0	ρ_3	ρ_2	ρ_1
σ_1	σ_1	σ_0	σ_3	σ_2	ρ_1	ρ_0	ρ_3	ρ_2
σ_2	σ_2	σ_1	σ_0	σ_3	ρ_2	ρ_1	ρ_0	ρ_3
σ_3	σ_3	σ_2	σ_1	σ_0	ρ_3	ρ_2	ρ_1	ρ_0

It is not worth your while to check every entry in the table, but make sure you check four or five entries at random to get an idea of how to compose symmetries, and let me know if there are any mistakes!

Let's convince ourselves that $(D_4, \circ)$ is a group.

The first thing we should ask about is closure. This is clear from the table (either the picture one or the symbol one); when you compose two elements of D_4 you get an element of D_4.

It is clear that ρ_0 (=do nothing) is an identity element.

It is also (geometrically) clear that every element has an inverse which does belong to D_4. If you reflect twice in the same line you end up where you started, so $\sigma_i \circ \sigma_i = \rho_0$; in other words, σ_i is its own inverse for $i = 0, 1, 2, 3$. The inverse of an anticlockwise rotation around O by 90° is an anticlockwise rotation around O by 270°. We find that the inverses of ρ_0, ρ_1, ρ_2 and ρ_3 respectively are ρ_0, ρ_3, ρ_2 and ρ_1.

What's left is to prove associativity. Composing symmetries is associative for exactly the same reason as composing functions is associative.

Doing "symmetry A followed by symmetry B" to the square and then doing symmetry C to what you got is the same as doing symmetry A to the square and then doing "symmetry B followed by symmetry C" to what you got (they both are doing A then B then C to the square).

If you think the square as being centred at the origin and the matrices for each symmetry then you really can think of the elements as functions.

D_4 non-abelian group is our first example of a non-abelian group. To check that it isn't abelian all we have to do is give a pair of symmetries that don't commute.

For example,

$$\sigma_0 \circ \rho_1 = \sigma_3, \qquad \rho_1 \circ \sigma_0 = \sigma_1.$$

5.4.1 Subgroups of D_4

The set D_4 contains rotations and reflections. Let us now look at the rotations on their own and the reflections on their own:

$$R = \{\rho_0, \rho_1, \rho_2, \rho_3\}, \qquad S = \{\sigma_0, \sigma_1, \sigma_2, \sigma_3\}.$$

For now let us look at the part of the composition table that involves only rotations:

$\circ$	ρ_0	ρ_1	ρ_2	ρ_3
ρ_0	ρ_0	ρ_1	ρ_2	ρ_3
ρ_1	ρ_1	ρ_2	ρ_3	ρ_0
ρ_2	ρ_2	ρ_3	ρ_0	ρ_1
ρ_3	ρ_3	ρ_0	ρ_1	ρ_2

Notice from the table that if we compose two rotations we obtain a rotation. We didn't really need the table for this; we can see it from the geometry. Thus $\circ$ is a binary operation on R (as well as being a binary operation on D_4).

We can ask whether $(R, \circ)$ is a group, and it is easy to see that the answer is yes (with the same reasoning as before). We have an interesting phenomenon, which is a group $(R, \circ)$ contained in another group $(D_4, \circ)$. We say that $(R, \circ)$ is a *subgroup* of $(D_4, \circ)$.

We will discuss subgroups at length later. It is also interesting to note that $(R, \circ)$ is abelian. An algebraic way of seeing the $(R, \circ)$ is abelian is to note that its composition table is symmetric about the leading diagonal. But you should also see geometrically that if you compose two rotations (centred at the same point) then the order does not matter. So $(R, \circ)$ is an abelian subgroup of the non-abelian group $(D_4, \circ)$.

What about $(S, \circ)$? Do the reflections of the square form a group? By looking at the composition table the first thing we notice is that S is **not** closed under composition. So $(S, \circ)$ is not a group. Are there any other subgroups inside $(D_4, \circ)$ besides $(R, \circ)$? Yes. See Figure 9 for a complete list.

Again, check that a couple of these are subgroups. Don't waste time checking there aren't other subgroups of $(D_4, \circ)$; when you know a lot more about groups and subgroups you can come back to this question.

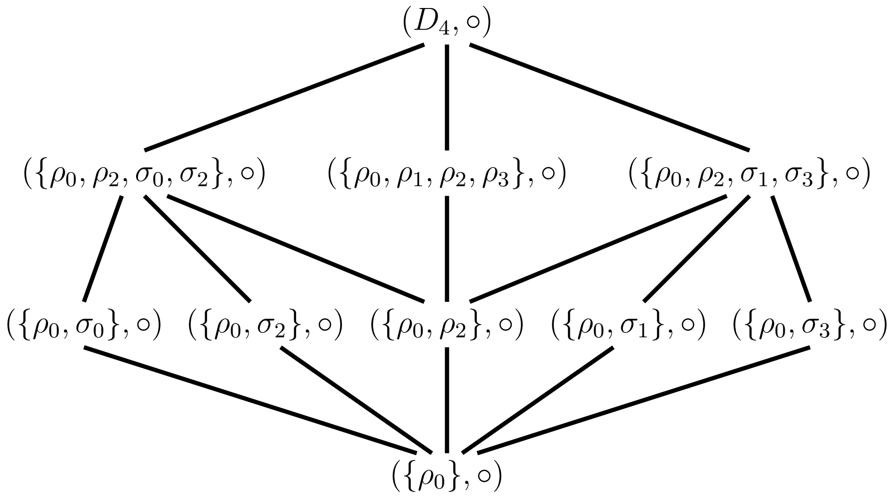

$(D_4, \circ)$
$(\{\rho_0, \rho_2, \sigma_0, \sigma_2\}, \circ)$
$(\{\rho_0, \rho_1, \rho_2, \rho_3\}, \circ)$
$(\{\rho_0, \rho_2, \sigma_1, \sigma_3\}, \circ)$
$(\{\rho_0, \sigma_0\}, \circ)$
$(\{\rho_0, \sigma_2\}, \circ)$
$(\{\rho_0, \rho_2\}, \circ)$
$(\{\rho_0, \sigma_1\}, \circ)$
$(\{\rho_0, \sigma_3\}, \circ)$
$(\{\rho_0\}, \circ)$

Figure 9: The figure shows the subgroups of $(D_4, \circ)$ and how they fit inside each other.

5.4.2 Exercise In this exercise you will write out the composition table for the group D_3 which is the group of symmetries of an equilateral triangle. Sketch an equilateral triangle and label the vertices 1, 2, 3 in anticlockwise order (see Chapter for a reminder about this). Label the centre of the triangle with O. Let ρ_0, ρ_1, ρ_2 denote anticlockwise rotations about O through angles 0, $2\pi/3$ and $4\pi/3$. Let σ_1, σ_2, σ_3 denote reflections about the lines respectively joining vertices 1, 2, 3 to O. Let

$$D_3 = \{\rho_0, \rho_1, \rho_2, \sigma_1, \sigma_2, \sigma_3\}.$$

Write down a composition table for D_3 and explain why it is a group [2]. Is it abelian? It has six subgroups; write them down.

5.4.3 Exercise Write down the symmetries of a triangle that is isoceles but not equilateral and a composition table for them. Do they form a group?

[2] More generally, D_n denotes the group of symmetries of a regular polygon with n sides. These are called the *dihedral groups*. Some mathematicians denote D_n by D_{2n} because it has $2n$ elements. Mysteriously, they don't denote S_n by $S_{n!}$.

Lecture 6 - First theorems and notation

Our first two theorems deal with subconscious assumptions. One of the defining properties of a group is the 'existence of the identity element' (property (iii)). The word 'the' contains a hidden assumption; how do we know there is only one identity element? Shouldn't we be talking about the 'existence of an identity element'?

6.1 Uniqueness of the identity element

6.1.1 Theorem Let $(G, \star)$ be a group. Then $(G, \star)$ has a unique identity element.

Proof. Suppose that e and e' are identity elements. Thus, for all $a \in G$ we have

$$a \star e = e \star a = a, \tag{8}$$

and

$$a \star e' = e' \star a = a. \tag{9}$$

Now let us try evaluating $e \star e'$. If we let $a = e$ and use (9) we find

$$e \star e' = e.$$

But if we let $a = e'$ and use (8) we find

$$e \star e' = e'.$$

Thus $e = e'$. In other words, the identity element is unique. ◊

6.1.2 Theorem Let $(G, \star)$ be a group and let a be an element of G. Then a has a unique inverse.

Proof. Our proof follows the same pattern as the proof of Theorem 6.1.1, and you'll see this pattern again and again during your undergraduate career. Almost all uniqueness proofs follow the same pattern: suppose that there are two of the thing that we want to prove unique; show that these two must be equal; therefore it is unique.

For our proof we suppose that b and c are both inverses of a. We want to show that $b = c$. By definition of inverse (property (iv) in the definition of a group) we know that

$$a \star b = b \star a = e, \qquad a \star c = c \star a = e,$$

where e is of course the identity element of the group. Thus

$$\begin{aligned} b &= b \star e && \text{by (iii) in the definition of a group} \\ &= b \star (a \star c) && \text{from the above } a \star c = e \\ &= (b \star a) \star c && \text{by (ii) in the definition of a group} \\ &= e \star c && \text{from the above } b \star a = e \\ &= c && \text{by (iii) again.} \end{aligned}$$

Thus $b = c$. Since any two inverses of a must be equal, we see that the inverse of a is unique. $\Diamond$

6.2 Getting Relaxed about Notation

It is quite tedious to keep writing $\star$ for the group operation. If $(G, \star)$ is a group and a, $b \in G$, we shall write ab for $a \star b$, unless there is reason for possible confusion.

For example if $(G, \star) = (\mathbb{R}, +)$ then it is silly to write ab for $a + b$ because the usual meaning for ab is "$a \times b$". But it is OK most of the time, and when it is OK we will do it. Moreover, we shall often say "let G be a group", without giving an explicit name to the binary operation. When we talk of the groups $\mathbb{R}$, $\mathbb{R}^2$, $\mathbb{R}[x]$, $\mathbb{R}^*$, etc. we shall mean the groups $(\mathbb{R}, +)$, $(\mathbb{R}^2, +)$, $(\mathbb{R}[x], +)$, $(\mathbb{R}^*, \cdot)$, etc. This may feel a bit odd at first but it's something you will get used to.

If G is a group, and we're writing ab for $a \star b$, then it makes sense to use 1 to denote the identity element instead of e. We write a^{-1} for the (unique) inverse of a. So

$$a \star b = b \star a = e,$$

where b is the inverse of a, becomes

$$aa^{-1} = a^{-1}a = 1,$$

which looks familiar.

Here are a couple of crucial results that you should get used to.

6.2.1 Theorem Let G be a group and $a \in G$. Then

$$(a^{-1})^{-1} = a.$$

Proof. We're being asked to prove that a is the inverse of a^{-1}. Thinking carefully about what this would mean, we want to show that

$$a^{-1}a = 1 = aa^{-1}$$

But this is clearly true because a^{-1} is the inverse of a. ◊

The above proof is a good exercise in getting your head around definitions. Make sure you understand how we are proving that a is acting as the inverse element to a^{-1} in the above (usually we think of this the other way round).

6.2.2 Theorem Let G be a group and a, $b \in G$. Then

$$(ab)^{-1} = b^{-1}a^{-1}.$$

Notice that we reverse the order when taking inverse. You have probably seen this before when you did matrices at school/college

Proof. We're being asked to prove that $b^{-1}a^{-1}$ is the inverse of ab. So we want to show that

$$(b^{-1}a^{-1})(ab) = 1 = (ab)(b^{-1}a^{-1}).$$

Now

$$\begin{aligned}(b^{-1}a^{-1})(ab) &= b^{-1}(a^{-1}a)b \qquad \text{by associativity}\\ &= b^{-1}1b\\ &= 1,\end{aligned}$$

and similarly $(ab)(b^{-1}a^{-1}) = 1$. ◊

Note that you shouldn't write $\dfrac{a}{b}$ unless the group is abelian. This notation is ambiguous; does $\dfrac{a}{b}$ mean $b^{-1}a$ or ab^{-1}? The two aren't the same in a non-abelian group.

6.2.3 Exercise Use D_4 to give counterexamples to the following:

- $b^{-1}a = ab^{-1}$,
- $(ab)^{-1} = a^{-1}b^{-1}$,
- $a^{-1}ba = b$.

6.2.4 Exercise Let G be a group satisfying $a^2 = 1$ for all a in G. Show that G is abelian.
Going back to our discussion of notation, if n is a positive integer we shall define

$$a^n = \underbrace{aa\cdots a}_{n \text{ times}}.$$

We define $a^0 = 1$. If n is a negative integer we define $a^n = (a^{-n})^{-1}$. Again we should reflect a little to make sure we're not being reckless. Does a^3 mean $(a \star a) \star a$ or $a \star (a \star a)$? It doesn't matter because of the associativity property of a group.

6.2.5 Example Let $\star$ be the binary operation on $S = \{a, b, c\}$ in Example 4.6.2 3. Note that $(S, \star)$ is definitely not a group, as $\star$ is not associative. Now you can check that

$$(a \star a) \star a = a, \qquad a \star (a \star a) = c.$$

Thus writing a^3 in this context does not make any sense. $\diamondsuit$

The following theorem deals with some consequences of this notation, which should look reasonably familiar to you.

6.2.6 Theorem Let G be a group, and let $a \in G$. Then

1. $a^n \in G$ for all $n \in \mathbb{Z}$.

2. If $n \in \mathbb{Z}$ then $(a^{-1})^n = (a^n)^{-1} = a^{-n}$.

3. Moreover, if m, n are integers then

$$(a^m)^n = a^{mn}, \qquad a^m a^n = a^{m+n}.$$

4. Further, if the group G is abelian, a, $b \in G$ and n an integer then

$$(ab)^n = a^n b^n.$$

Proof.

1. Let's deal with the case of $n = 0$ separately. $a^0 = 1$ and certainly $1 \in G$ so the statement is true then.

Now let's prove it for positive integers by induction on n. If $n = 1$ then $a^n = a^1 = a$ since $a \in G$ so the statement is true when $n = 1$.

Now suppose we know that $a^k \in G$ for some positive integer k. Then $a^{k+1} = a^k a \in G$ since both a^k and a are in G which is closed under the binary operation. It follows by induction that $a^n \in G$ for any positive integer n.

Now let's prove that $a^n \in G$ when n is a negative integer.

Then $a^n = (a^{-n})^{-1}$. But we know that $a^{-n} \in G$ since $-n$ is positive by earlier in the proof. So $(a^{-n})^{-1}$ is also in G by the definition of a group. Therefore $a^n \in G$.

2. Clearly this is true if $n = 0$ since all three of the expressions are equal to 1.

   If n is a positive integer then $a^n(a^{-1})^n = \underbrace{aa\cdots a}_{n \text{ times}}\underbrace{a^{-1}a^{-1}\cdots a^{-1}}_{n \text{ times}}$.

   Each aa^{-1} in the centre collapses to a 1 and eventually we see that this expression is 1. Therefore it's true to say that $(a^{-1})^n = (a^n)^{-1}$. Also $(a^n)^{-1} = a^{-n}$ just by notational definition.

   If n is a negative integer then

$$a^n(a^{-1})^n = (a^{-n})^{-1}((a^{-1})^{-n})^{-1} = (a^{-1})^{-n}((a^{-1})^{-1})^{-n} = (a^{-1})^{-n}a^{-n} = 1$$

   where the expressions either side of the second and third equals sign are the same by what we have already proved for positive integers. The last equals sign follows in a similar way to above (remembering that $-n$ is a positive integer).

   Again we can conclude that $(a^{-1})^n = (a^n)^{-1}$. Also $(a^n)^{-1} = ((a^{-n})^{-1})^{-1} = a^{-n}$.

3. Here we need to first prove that for any integers m and n,

$$(a^m)^n = a^{mn}, \qquad a^m a^n = a^{m+n}.$$

If m and n are positive integers then

$$(a^m)^n = \underbrace{\underbrace{aa\cdots a}_{m \text{ times}}\underbrace{aa\cdots a}_{m \text{ times}}\cdots\underbrace{aa\cdots a}_{m \text{ times}}}_{n \text{ times}}.$$

From this it can be seen that $(a^m)^n = a^{mn}$.

If m is positive and n is negative then we have

$$(a^m)^n = ((a^m)^{-n})^{-1} = (a^{-mn})^{-1} = ((a^{mn})^{-1})^{-1} = a^{mn}$$

using what we have just proved for positive integers and part ii).

The other cases (m negative and n positive and m and n both negative are similar).

We also need to show that $a^m a^n = a^{m+n}$.

If m and n are positive integers then

$$a^m a^n = \underbrace{aa\cdots a}_{m \text{ times}}\underbrace{aa\cdots a}_{n \text{ times}} = a^{m+n}.$$

If m is positive and n is negative then, since $n = -k$ where $k = -n$ is a postive integer we have

$$a^m a^n = a^m a^{-k} = a^m(a^{-1})^k \underbrace{aa\cdots a}_{m \text{ times}}\underbrace{a^{-1}a^{-1}\cdots a^{-1}}_{k \text{ times}} = a^{m-k} = a^{m+n}.$$

Again, the other cases are similar.

4. Suppose n is positive and G is abelian.

   Then
   $$(ab)^n = \underbrace{abab\cdots ab}_{n \text{ times}} = \underbrace{aa\cdots a}_{n \text{ times}}\underbrace{bb\cdots b}_{n \text{ times}} = a^n b^n.$$

   Note that the expressions either side of the second equals sign are the same only because G is abelian.

   Then, if n is negative, we have

   $$(ab)^n = ((ab)^{-n})^{-1} = ((ab)^{-1})^{-n} = (b^{-1}a^{-1})^{-n} = (b^{-1})^{-n}(a^{-1})^{-n} = b^n a^n = a^n b^n.$$

The equalities in the above rely on what we have already proved in this theorem and results from earlier in this chapter (it's a useful exercise to make sure you can see which ones). ♢

The next example shows that we must have an ablelian group for 4. in the theorem 6.2.6 above to hold.

6.2.7 Example In D_4 you can check that

$$\rho_1^2\sigma_0^2 = \rho_2, \qquad (\rho_1\sigma_0)^2 = \rho_0,$$

and so $\rho_1^2\sigma_0^2 \neq (\rho_1\sigma_0)^2$. ♢

6.3 Additive Notation

For some groups the binary operation is 'addition' (whatever that means). These include $(\mathbb{R}, +)$, $(\mathbb{Z}, +)$, $(\mathbb{R}[x], +)$, $(\mathbb{R}^2, +)$ etc. An important convention is that additive notation is only ever used for abelian groups. A multiplicative group can be abelian, such as $(\mathbb{R}^*, \cdot)$, and can be non-abelian, such as $(D_4, \circ)$.

You need to rephrase statements appropriately when using additive notation. For example, instead of speaking of

$$a^n = \underbrace{aa \cdots a}_{n \text{ times}},$$

you need to talk about

$$na = \underbrace{a + a + \cdots + a}_{n \text{ times}}.$$

Instead of b^{-1} write $-b$. We will mostly state and prove theorems in multiplicative notation, but it's up to you to translate these into additive notation for groups where the binary operation is addition. Let's do this for Theorem 6.2.6. Here is the translation.

6.3.1 Theorem Let G be an (abelian) group with addition as the binary operation, and let $a \in G$. Then

1. $na \in G$ for all $n \in \mathbb{Z}$.
2. $n(-a) = -(na) = (-n)a$ for any $n \in \mathbb{Z}$
3. Moreover, if m, n are integers then

$$m(na) = (mn)a, \qquad ma + na = (m + n)a.$$

4. Further, if a, $b \in G$ and n an integer then

$$n(a+b) = na + nb.$$

Lecture 7 - More examples of groups

Examples are an integral part of abstract algebra. They are as important as the definitions and theorems.

7.1 Matrix Groups I

We saw that $(M_{2\times 2}(\mathbb{R}), +)$ is a group. This in fact is **not** a particularly interesting group because addition of matrices is not a very interesting operation. Multiplication of matrices is a far more interesting and natural operation; as we saw, if A, B represent certain geometric operations (e.g. scaling, reflection, rotation, etc.) then BA is the operation that one obtains from doing A first then B; if this doesn't sound familiar look again at Section 3.4 and at Example 3.3.1. Can we obtain a group out of (say) 2×2 matrices under multiplication?

To answer, let's look back to Example 5.2.4. There we obtained a multiplicative group from the real numbers by removing 0. We had to remove 0 because it doesn't have a multiplicative inverse.

It will not be enough for us to exclude the zero matrix, because there are non-zero matrices that do not have an inverse—see for example 3.6.1. What if we exclude all non-invertible matrices; do we get a group under multiplication?

7.1.1 Definition Define

$$\mathrm{GL}_2(\mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a,\, b,\, c,\, d \in \mathbb{R} \text{ and } ad - bc \neq 0 \right\}.$$

◊

Recall that $ad - bc$ is the determinant of the 2×2-matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, and the matrix is invertible if and only if this determinant is non-zero (Theorem 3.6.2). So $\mathrm{GL}_2(\mathbb{R})$ contains all the invertible 2×2 matrices (with real entries) and none of the non-invertible ones.

7.1.2 Theorem $\mathrm{GL}_2(\mathbb{R})$ is group under multiplication of matrices. We call $\mathrm{GL}_2(\mathbb{R})$ the *general linear group.*

Proof. The first thing to check is that $\mathrm{GL}_2(\mathbb{R})$ is closed under multiplication. If A and B are in $\mathrm{GL}_2(\mathbb{R})$ then AB is a 2×2 matrix with real entries. Also, we know that $\det(AB) = \det(A)\det(B)$ (by Theorem 3.6.4). Because A and B have non-zero determinants, so does AB. So AB is in $\mathrm{GL}_2(\mathbb{R})$.

Next we want to show associativity. But we already know that matrix multiplication is associative thanks to Theorem 3.5.1.

The identity matrix

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

is in $\mathrm{GL}_2(\mathbb{R})$ (because it has non-zero determinant) and is the multiplicative identity element; it satisfies $AI_2 = I_2A = A$ for any 2×2 matrix A.

Finally, we should ask if every matrix in $\mathrm{GL}_2(\mathbb{R})$ has an inverse. We defined $\mathrm{GL}_2(\mathbb{R})$ so every element is invertible, but we need to make sure that the inverse is also in $\mathrm{GL}_2(\mathbb{R})$. If $A \in \mathrm{GL}_2(\mathbb{R})$ then $\det(A) \neq 0$. We know by Theorem 3.6.4 that $\det(A^{-1}) \neq 0$ and indeed

$$\det(A^{-1}) = \frac{1}{\det(A)}.$$

Moreover, A^{-1} is a 2×2 matrix with real entries. Hence $A^{-1} \in \mathrm{GL}_2(\mathbb{R})$. $\diamondsuit$

We can define $\mathrm{GL}_2(\mathbb{Q})$ and $\mathrm{GL}_2(\mathbb{C})$ in a similar way and show that they are groups. However, as this very important exercise shows we can't do this with the integers.

7.1.3 Exercise Show that

$$\left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} | a,\, b,\, c,\, d \in \mathbb{Z} \text{ and } ad - bc \neq 0 \right\}$$

is **not** a group with respect to multiplication.

It turns out that there is a natural definition for a group $\mathrm{GL}_2(\mathbb{Z})$. We'll return to this in Example 17.1.10.

7.2 Congruence Classes

Let $m \geq 2$ be an integer. By $\mathbb{Z}/m\mathbb{Z}$ we mean the set of congruence classes modulo m. In *Foundations* this is denoted by $\mathbb{Z}/m$ and in most algebra textbooks by $\mathbb{Z}_m$. Our notation is the least economical, but also the least arbitrary. There is an excellent reason for taking an approach based on congreunce classes and using this notation $\mathbb{Z}/m\mathbb{Z}$ instead of $\mathbb{Z}/m$ and $\mathbb{Z}_m$. You'll later see in Lecture 14 that this idea is a specific example of a general idea in group theory and so taking this approach will prepare you for that.

If a is an integer, we shall write $\overline{a}$ for the congruence class of a modulo m. Thus

$$\overline{a} = \{\dots, a-3m, a-2m, a-m, a, a+m, a+2m, a+3m, \dots\}.$$

In other words, $\overline{a}$ consists of all integers congruent to a modulo m. From *Foundations/Sets and Numbers* you know that

$$\mathbb{Z}/m\mathbb{Z} = \{\overline{0}, \overline{1}, \overline{2}, \dots, \overline{m-1}\}$$

and that the classes $\overline{0}, \overline{1}, \dots, \overline{m-1}$ are distinct, so $\mathbb{Z}/m\mathbb{Z}$ consists of exactly m classes. You know how addition and multiplication is defined on $\mathbb{Z}/m\mathbb{Z}$:

$$\overline{a} + \overline{b} = \overline{a+b}, \qquad \overline{a} \cdot \overline{b} = \overline{ab}.$$

It's important to understand that the two binary operations above are *well-defined.*

You know that there is more than one way to write a particular congruence class. For example if $m = 3$ then $[2] = [5] = [8]$, they are all the set $\{\dots, -4, -1, 2, 5, 8, \dots\}$.

This means we need to check that, in general, if $[a] = [c]$ and $[b] = [d]$ then both $[a+b] = [c+d]$ and $[a \cdot b] = [c \cdot d]$. This was covered in *Foundations/Sets and Numbers* so it won't be covered here again (but I will go through it in the live lecture). However, when we generalise this idea later in the module we will encounter this again and consider it there.

7.2.1 Example The addition and multiplication tables for $\mathbb{Z}/6\mathbb{Z}$ are in Table 1. ♢

7.2.2 Exercise Write down the addition and multiplication tables for $\mathbb{Z}/4\mathbb{Z}$ and $\mathbb{Z}/5\mathbb{Z}$.

$+$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$	$\overline{5}$
$\overline{0}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$	$\overline{5}$
$\overline{1}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$	$\overline{5}$	$\overline{0}$
$\overline{2}$	$\overline{2}$	$\overline{3}$	$\overline{4}$	$\overline{5}$	$\overline{0}$	$\overline{1}$
$\overline{3}$	$\overline{3}$	$\overline{4}$	$\overline{5}$	$\overline{0}$	$\overline{1}$	$\overline{2}$
$\overline{4}$	$\overline{4}$	$\overline{5}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$
$\overline{5}$	$\overline{5}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$

$\times$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$	$\overline{5}$
$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$
$\overline{1}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$	$\overline{5}$
$\overline{2}$	$\overline{0}$	$\overline{2}$	$\overline{4}$	$\overline{0}$	$\overline{2}$	$\overline{4}$
$\overline{3}$	$\overline{0}$	$\overline{3}$	$\overline{0}$	$\overline{3}$	$\overline{0}$	$\overline{3}$
$\overline{4}$	$\overline{0}$	$\overline{4}$	$\overline{2}$	$\overline{0}$	$\overline{4}$	$\overline{2}$
$\overline{5}$	$\overline{0}$	$\overline{5}$	$\overline{4}$	$\overline{3}$	$\overline{2}$	$\overline{1}$

Table 1: The addition and multiplication tables for $\mathbb{Z}/6\mathbb{Z}$.

7.2.3 Theorem Let m be an integer satisfying $m \geq 2$. Then $(\mathbb{Z}/m\mathbb{Z}, +)$ is an abelian group.

Proof. To show that $\mathbb{Z}/m\mathbb{Z}$ a group, we want to check that $\mathbb{Z}/m\mathbb{Z}$ is closed under addition, that addition is associative, that there is an identity element, and that every element has an additive inverse.

We defined $\mathbb{Z}/m\mathbb{Z}$ to be the set of congruence classes modulo m. We defined the sum of classes $\overline{a}$ and $\overline{b}$ to be $\overline{a+b}$ which is a congruence class modulo m. So $\mathbb{Z}/m\mathbb{Z}$ is closed under addition. Let's prove associativity. Note

$$\begin{aligned}(\overline{a}+\overline{b})+\overline{c} &= \overline{a+b}+\overline{c}\\ &= \overline{(a+b)+c}\\ &= \overline{a+(b+c)} \qquad \text{addition in } \mathbb{Z} \text{ is associative}\\ &= \overline{a}+\overline{b+c}\\ &= \overline{a}+(\overline{b}+\overline{c}).\end{aligned}$$

Thus addition in $\mathbb{Z}/m\mathbb{Z}$ is associative. Obviously $\overline{0}$ is the additive identity. What about the additive inverse? Note that $\overline{a} + \overline{-a} = \overline{0}$ so every class has an additive inverse [3].

Thus $(\mathbb{Z}/m\mathbb{Z}, +)$ is a group. We leave the proof that it is abelian as an easy exercise.

[3] Perhaps you prefer the inverse of $\overline{a}$ where $0 \leq a < m$ to be of the form $\overline{b}$ where b also satisfies $0 \leq b < m$. In this case, if $0 < a < m$, then observe that $0 < m - a < m$, and $\overline{a} + \overline{m-a} = \overline{0}$, since $a + (m-a) \equiv 0 \pmod{m}$. Moreover $-0 \equiv 0 \pmod{m}$, thus $-\overline{0} = \overline{0}$.

Lecture 8 - The order of an element

We return to using multiplicative notation. In Theorem 6.2.6 we observed that if G is a group containing an element a, then a^n is also in G for all integers n. It seems at first sight that this makes every group infinite: just pick an element a and you have an infinite list of elements

$$\ldots, a^{-4}, a^{-3}, a^{-2}, a^{-1}, 1, a, a^2, a^3, a^4, a^5, \ldots.$$

The group D_4 is finite, so what goes wrong? Take $a = \rho_1 \in D_4$ which represents anti-clockwise rotation by 90°. Then $a^4 = 1$. Thus the seemingly infinite list above simply becomes

$$\ldots, 1, a, a^2, a^3, 1, a, a^2, a^3, 1, \ldots.$$

In reality the list consists of exactly four elements $1, a, a^2, a^3$.

8.1 The Order of an Element

The above discussion leads us to the following definition.

8.1.1 Definition The **order** of an element a in a group G is the smallest positive integer n such that $a^n = 1$. If there is no such positive integer n, we say a has **infinite order**.

8.1.2 Example The order of ρ_1 is D_4 is 4. The order of ρ_2 is 2. The order of ρ_0 is 1. What are the orders of the other elements? $\Diamond$

8.1.3 Example In $(\mathbb{R}^*, \cdot)$, the element 1 has order 1 and the element -1 has order 2. What is the order of 7? Is there a *positive integer* n such that $7^n = 1$? No. Thus 7 has infinite order.

What are the elements of finite order in $\mathbb{R}^*$. These are the non-zero real numbers a such that $a^n = 1$ for some positive integer n. You should know that the only such real numbers are 1 and -1. So the only elements of finite order in $\mathbb{R}^*$ are 1 and -1 and all the other elements have infinite order. $\Diamond$

8.1.4 Example When you saw the equation $a^n = 1$ in the above example, you may have thought of the n-th roots of unity. The n-th roots of unity don't all live in $\mathbb{R}$; they live in $\mathbb{C}$. In fact, they live in $\mathbb{C}^*$.

For concreteness we take $n = 3$. You will know from *Foundations/Sets and Numbers* that there are three cube roots of unity. These are $1, \zeta, \zeta^2$, where $\zeta =$

$e^{2\pi i/3}$. See Figure 10. Let us think of these inside the group $\mathbb{C}^*$. Then ζ and ζ^2 have order 3. Let's check this for ζ^2. We note

$$(\zeta^2)^1 = \zeta^2, \qquad (\zeta^2)^2 = \zeta^4 = \zeta \cdot \zeta^3 = \zeta, \qquad (\zeta^2)^3 = (\zeta^3)^2 = 1^2 = 1.$$

So the least positive integer n such that $(\zeta^2)^n = 1$ is $n = 3$, so ζ^2 has order 3. Don't forget that 1 has order 1. So there are three cube roots of unity. Two have order 3 and one has order 1.

Now let us think briefly about the fourth roots of unity. These are $1, i, i^2, i^3$. Again see Figure 10. Note that $i^2 = -1$ and $i^3 = -i$. Of the four, only two have order 4 and these are i and i^3 (check). Of course, -1 has order 2 and 1 has order 1. $\Diamond$

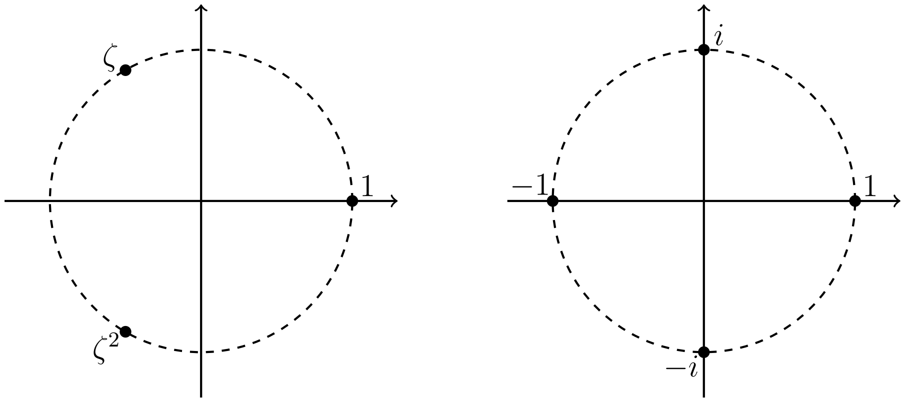


Figure 10: On the left, the three cube roots of unity: here $\zeta = e^{2\pi i/3}$. On the right, the four fourth roots of unity. Note that $e^{2\pi i/4} = e^{\pi i/2} = i$, so the fourth roots of unity are 1, i, $i^2 = -1$, and $i^3 = -i$.

8.1.5 Exercise Write down and sketch the sixth roots of unity. What are their orders? Repeat with the eighth roots of unity.

8.1.6 Exercise $\mathbb{C}^*$ has lots of elements of infinite order. Find a few.

8.1.7 Exercise Let $G = \mathrm{GL}_2(\mathbb{R})$. Show that

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

belong to G. Determine their orders.

Whilst reading the above examples and working out your own, you may have noticed the following:

8.1.8 Lemma Let G be a group and g be an element of G.

(i) g has order 1 if and only if g is the identity element.

(ii) Let m be a **non-zero** integer. Then $g^m = 1$ if and only if g has finite order d with $d \mid m$.

Proof. Let G be a group. Suppose g has order 1. By definition of order, $g^1 = 1$. Thus $g = 1$ which is the identity element of G. Conversely, the identity element clearly has order 1. This proves (i).

Part (ii) is an 'if and only if' statement. Suppose that g has order d and $d \mid m$. Then $g^d = 1$ and $m = qd$ where q is an integer. So $g^m = (g^d)^q = 1$. Let us prove the converse. Suppose $g^m = 1$ where m is a non-zero integer. Then $g^{|m|} = 1$, and $|m|$ is a positive integer. Thus g has finite order, which we denote by d. By the *division algorithm* which you met in *Foundations/Sets and Numbers* we may write

$$m = qd + r, \qquad q,\, r \in \mathbb{Z} \text{ and } 0 \le r < d.$$

Now $g^d = 1$ by definition of order, so $1 = g^m = (g^d)^q \cdot g^r = g^r$. But $0 \le r < d$. As d is the order, it is the **least positive** integer such that $g^d = 1$. So $g^r = 1$ is possible with $0 \le r < d$ if and only if $r = 0$. This happens if and only if $m = qd$ which is the same as $d \mid m$. ♢

8.1.9 Exercise Let G be an abelian group. Suppose a, b are elements of orders m and n. Let $d = \operatorname{lcm}(m, n)$. Show that $(ab)^d = 1$, ensuring that you point out where you have used the fact the G is abelian. Give a counterexample to show that this does not have to be true if G is non-abelian. **Hint:** Look at D_3.♢

Now we return to our examples. We've looked at various multiplicative groups, but what about additive groups? If $(G, +)$ is a group where the binary operation is addition, what is the order of an element a? Of course, it is the smallest positive integer n such that $na = 0$. If there is no such positive integer that a has infinite order.

8.1.10 Example In $(\mathbb{R}, +)$, $(\mathbb{Z}, +)$, $(\mathbb{R}[x], +)$, $(\mathbb{C}, +)$, the only element of finite order is 0, which has order 1. All other elements have infinite order.

How do we know this. Look at the equation $na = 0$ with a in the group and n a positive integer. We can divide both sides by n and obtain $a = 0$. $\Diamond$

You're may be wondering if in every additive group, the identity element 0 is the only one of finite order. The following example shows that this isn't true.

8.1.11 Example Observe that in $(\mathbb{Z}/m\mathbb{Z}, +)$, every element a has finite order. Indeed, $ma \equiv 0 \pmod{m}$ and so $m\overline{a} = \overline{0}$. This does not mean that every element has order m, since the order of a is defined to be the *least* positive integer n such that $n\overline{a} = \overline{0}$. However, we do know by Lemma 8.1.8 that the order n is a divisor of m.

Let us look at the elements of $(\mathbb{Z}/6\mathbb{Z}, +)$ and determine their orders. We quickly find that $\overline{0}$ has order 1 (as usual); $\overline{1}$ and $\overline{5}$ have order 6; $\overline{2}$ and $\overline{4}$ have order 3; and $\overline{3}$ has order 2. $\Diamond$

8.1.12 Exercise Find the orders of the elements of $(\mathbb{Z}/4\mathbb{Z}, +)$ and $(\mathbb{Z}/5\mathbb{Z}, +)$.

8.2 First preview of Lagrange's Theorem

Lagrange's theorem is a beautiful result which says something about how only the number of elements in a finite group determines some of its internal structure. You'll see the full result later on in the module but for now, without stating Lagrange's theorem itself(!), we'll look at a corollary.

8.2.1 Definition Let G be a group. The *order* of G is the number of elements that G has. We denote the order of G by $|G|$ or $\#G$.

8.2.2 Corollary (of Lagrange's Theorem which is to follow in Lecture 13.) Let G be a finite group, and let g be an element of G. The order of g divides the order of G. $\Diamond$

Here's a corollary of the corollary..

8.2.3 Corollary Let G be a finite group of order n, and let g be an element of G. Then $g^n = 1$.

Proof. Let d be the order of g. By definition of the order of an element, $g^d = 1$. By Lagrange's Theorem, d divides n. Thus $n = kd$ for some integer k. Now

$$g^n = (g^d)^k = 1^k = 1,$$

which is what we set out to prove. $\Diamond$

8.2.4 Example Lagrange's Theorem applies to finite groups of which you haven't seen many examples yet. One example of a finite group is D_4 which has order 8. So every element of D_4 must have order dividing 8. In fact the elements of D_4 have orders 1, 2 and 4. $\Diamond$

8.2.5 Example The set $\{1, i, -1, -i\}$ forms a group of order 4 under multiplication (convince yourself that this is true). Then 1 has order 1; -1 has order 2; i and $-i$ have order 4. This is all consistent with the Corollary 8.2.3. $\Diamond$

Lecture 9 - Subgroups

We met subgroups in the last chapter when we discussed the group D_4. Let us write down the formal definition and give some examples.

9.1 Definition and examples

9.1.1 Definition Let $(G, \star)$ be a group. Let H be a subset of G and suppose that $(H, \star)$ is also a group. Then we say that H is a subgroup of G (or more formally $(H, \star)$ is a subgroup of $(G, \star)$). $\Diamond$

For H to be a subgroup of G, we want H to a group with respect to *the same binary operation* that makes G a group.

9.1.2 Example $\mathbb{R}^*$ is a subset of $\mathbb{R}$ and both are groups. But $\mathbb{R}^*$ is **not** a subgroup of $\mathbb{R}$, since the operation that makes $\mathbb{R}^*$ a group is multiplication and the operation that makes $\mathbb{R}$ a group is addition. $\Diamond$

9.1.3 Example $\mathbb{Z}$ is a subgroup of $\mathbb{R}$ (or more formally, $(\mathbb{Z}, +)$ is a subgroup of $(\mathbb{R}, +)$); because $\mathbb{Z}$ is a subset of $\mathbb{R}$ and both are groups with respect to the same binary operation which is addition. $\Diamond$

9.1.4 Example $\mathbb{R}$ is a subgroup of $\mathbb{R}[x]$ since any real number can be viewed as a polynomial of degree 0. $\Diamond$

9.1.5 Example $(\emptyset, +)$ is **not** a subgroup of $(\mathbb{R}, +)$, simply because $(\emptyset, +)$ is not a group; a group has to be non-empty since it has to contain an identity element. $\Diamond$

9.2 Criterion for a Subgroup

9.2.1 Theorem Let G be a group. A subset H of G is a subgroup if and only if it satisfies the following three conditions

(a) $1 \in H$,

(b) if a, $b \in H$ then $ab \in H$,

(c) if $a \in H$ then $a^{-1} \in H$.

Proof. The theorem has an "if and only if" statement.

It's usual when proving an "if and only if" statement to break it up into an "if" part, and an "only if" part, and prove each part separately. This is what we will do here. The "if" part says: "if H is a subset of G that satisfies (a),(b),(c) then it is a subgroup of G". The "only if" part says: "if H is a subgroup of G then H satisfies (a), (b), (c)".

Let us do the "if" part of the proof first. We have a group G and a subset H of G. All we have been told is that H satisfies conditions (a), (b), (c) in the statement of the theorem. We want to show that H is a group, where the binary operation on H is the same as the binary operation on G. This means that we have to show that H satisfies properties (i), (ii), (iii), (iv) in the definition of a group.

Property (i) is 'closure': we want that if a, $b \in H$ then $ab \in H$. But this is what (b) is saying. So (i) is satisfied.

Property (ii) is associativity. We want to show that for all a, b, $c \in H$, we have $(ab)c = a(bc)$. But if a, b, c are elements of H then they are also elements of G. We know that associativity holds in G: $(ab)c = a(bc)$. So (ii) holds [4].

Property (iii) is the existence of the identity element in H. But (a) tells us that $1 \in H$. This 1 is the identity element of G and so satisfies $a1 = 1a = a$ for all a in G. Since every a in H is also in G we have that $a1 = 1a = a$ for all a in H so 1 is the identity element of H, and so (iii) holds.

Finally, property (iv) asserts the existence of an inverse for every $a \in H$. This follows from (c). Hence H is a group contained in G and so a subgroup. We have now finished the proof of the "if" part.

Next we do the "only if"part of the proof. Here we assume that H is a subgroup of $|G|$ as in Definition 9.1.1 and need to show that it then satisfies conditions (a), (b) and (c) in the statement of this theorem.

In some ways proving (a) is the most tricky. Remember the 1 in (a) is the identity from G. Let's call that element 1_G in this part of the proof to remind us of that. All we know is that H is a group and so it will contain an identity element. Let's call this 1_H. So $1_H \in G$ and if we can show that $1_H = 1_G$ then it will follow that $1 \in H$.

[4] There is a subtle point here that is camouflaged by our notation, and that is that the binary operation we're using on H is precisely the same one as the binary operation we're using on G. If it was different we would have no right to say: because associativity holds in G it holds in H.

We have $1_H = 1_H 1_H$. Now $1_H \in G$ and so it has an inverse in G, 1_H^{-1} with the property that $1_H 1_H^{-1} = 1_G$ (think carefully about this and make sure you are happy with why 1_G is on the right hand side as opposed to 1_H).

Multiplying both sides of $1_H = 1_H 1_H$ on the left by 1_H^{-1} gives $1_H^{-1} 1_H = 1_H^{-1} 1_H 1_H$ or $1_G = 1_G 1_H = 1_H$.

So the identity element of G equals the identity element of H and in particular (returning to the usual notation $1_G = 1$) $1 \in H$.

(b) follows because, as a group in its own right under the same binary operation as G, H is closed under that binary operation.

To show (c) we have to show that if $a \in H$ then the inverse of a in G is in H. Because H is a subgroup it will have an inverse in H, call this a^{-1}. Then $1_H = aa^{-1} = a^{-1}a$. But, since $1 = 1_H$, this means that $1 = aa^{-1} = a^{-1}a$ and a^{-1} is the inverse of a in G (by the uniqueness of inverses, Theorem 6.1.2). $\diamondsuit$

Now let's try out the theorem.

9.2.2 Example Let's take $G = \mathbb{R}^*$ and H the subset of positive real numbers:

$$H = \{a \in \mathbb{R}^* : a > 0\}.$$

Let's show that H is a subgroup of G. First, 1 is positive, so $1 \in H$. Hence condition (a) is satisfied.

To check (b), suppose that a, b are in H. Thus a and b are positive, and so their product ab is also positive. Hence $ab \in H$ and we know that (b) is satisfied.

Finally, we want to check condition (c). Suppose a is an element of H. Then a is positive, and so a^{-1} is positive. Hence a^{-1} is also an element of H. It follows that condition (c) is satisfied.

By Theorem 9.2.1, H is a subgroup of $\mathbb{R}^*$. $\diamondsuit$

9.2.3 Example Let

$$2\mathbb{Z} = \{2a : a \in \mathbb{Z}\} = \{\ldots, -6, -4, -2, 0, 2, 4, 6, \ldots\}.$$

In other words, $2\mathbb{Z}$ is the set of even integers. Now $2\mathbb{Z}$ is a subset of $\mathbb{Z}$, but is it a subgroup of $\mathbb{Z}$? We should check the three conditions in the theorem, where $G = \mathbb{Z}$ and $H = 2\mathbb{Z}$. Condition (a) is "$1 \in H$". What does that mean in our context? 1 is not the number 1. The 1 in the theorem is the identity element for the group operation on $\mathbb{Z}$. The group operation on $\mathbb{Z}$ is addition. The identity element is 0. As 0 is an even number (after all $0 = 2 \times 0$) we have $0 \in 2\mathbb{Z}$. Thus condition (a) is satisfied.

Let's move on to condition (b). This says "if a, $b \in H$ then $ab \in H$". Again ab doesn't always mean the product of a and b; it is shorthand for $a \star b$ where $\star$ is the binary operation on G. Here $G = \mathbb{Z}$ and the binary operation on $\mathbb{Z}$ is $+$. So to check (b) what we must check is the following "if a, $b \in 2\mathbb{Z}$ then $a + b \in 2\mathbb{Z}$". In words this just says "the sum of two even integers is even", which is true so (b) holds.

Finally we have to interpret (c) in our context. Here a^{-1} is the inverse of a with respect to addition, so it just means $-a$. Thus to check (c) we want to check that "if a is an even integer then $-a$ is also even". Again this is true, so (c) holds.

It follows from Theorem 9.2.1 that $2\mathbb{Z}$ is a subgroup of $\mathbb{Z}$.

By contrast, the set of odd integers

$$\{\dots, -5, -3, -1, 1, 3, 5, \dots\}$$

is not a subgroup of $\mathbb{Z}$. For example, it does not contain the identity element 0, so does not satisfy (a). ♢

9.2.4 Example In Subsection 5.4.1, we listed the ten subgroups of D_4. Go back to that list, and use Theorem 9.2.1 to verify that a couple of them are indeed subgroups. ♢

9.2.5 Example Let

$$V = \{(a, a) : a \in \mathbb{R}\}.$$

In other words V is the subset of $\mathbb{R}^2$ where the x-coordinate equals the y-coordinate. Thus V is the line $y = x$ in $\mathbb{R}^2$. It is geometrically obvious that V contains the origin, which is the identity element of $\mathbb{R}^2$; that if we add two vectors belonging to it the result also belongs to it; and that if we multiply any vector belonging to this diagonal by -1 the result also belongs to V. Figure 11 will help you visualise this. But you also need to be able to write a proof in symbols. Let us do that:

First note that $\mathbf{0} = (0,0) \in V$. Secondly, suppose $\mathbf{u} \in V$ and $\mathbf{v} \in V$. By definition of V, $\mathbf{u} = (a,a)$ and $\mathbf{v} = (b,b)$ for some a, $b \in \mathbb{R}$. Thus $\mathbf{u} + \mathbf{v} = (a+b, a+b)$ which again belongs to V. Finally, suppose that $\mathbf{v} \in V$. By definition of V, $\mathbf{v} = (a,a)$ for some $a \in \mathbb{R}$. So $-\mathbf{v} = (-a,-a)$ which is in V. This shows that V is a subgroup of $\mathbb{R}^2$. ♢

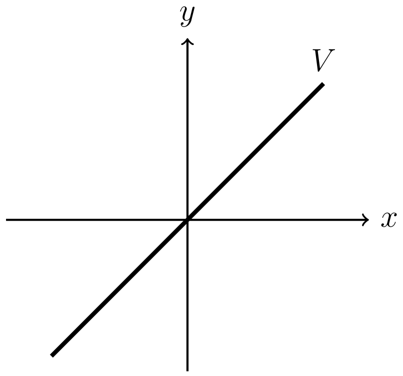


Figure 11: The set $V = \{(a,a) : a \in \mathbb{R}\}$ is the line $y = x$. It contains the identity element $(0,0)$, is closed under addition and negation. Therefore it is a subgroup of $\mathbb{R}^2$.

9.2.6 Example This time we take $W = \{(a,a) : a \in \mathbb{R},\ a \geq 0\}$. The set W is not all the line $y = x$ but a 'ray' as in Figure 12. Note that W does satisfy the first two conditions (a), (b) for being a subgroup. However, it does not satisfy

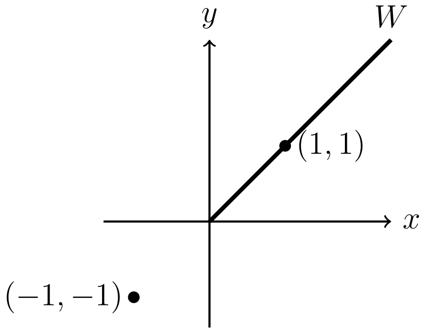


Figure 12: The ray $W = \{(a,a) : a \in \mathbb{R},\ a \geq 0\}$ is not a subgroup of $\mathbb{R}^2$. It contains the identity element $(0,0)$ and is closed under addition. The problem is with the existence of additive inverses; e.g. $(1,1)$ is in W but its inverse $(-1,-1)$ isn't in W.

condition (c); for example, $\mathbf{v} = (1, 1)$ belongs to W but $-\mathbf{v} = (-1, -1)$ does not. Hence W is not subgroup of $\mathbb{R}^2$.

To show that W is not a subgroup, we gave a **counterexample**. This means that we gave an example to show that at least one of the requirements in Theorem 9.2.1 is not satisfied. $\Diamond$

9.2.7 Example Let

$$V = \{(a, a) : a \in \mathbb{R}\}, \qquad V' = \{(-a, a) : a \in \mathbb{R}\}.$$

You know from Example 9.2.5 that V is a subgroup of $\mathbb{R}^2$ (and is the line $y = x$). You can show, in a similar way, that V' (which happens to be the line $y = -x$) is also a subgroup of $\mathbb{R}^2$. What about their union $U = V \cup V'$? You can check that U satisfies conditions (a) and (c) of Theorem 9.2.1. However, $(1, 1)$ and $(-1, 1)$ are in U but their sum $(0, 2)$ is not in U. So U does not satisfy (b), and is therefore not a subgroup of $\mathbb{R}^2$. See Figure 13.

On the other hand, the intersection $V \cap V' = \{(0, 0)\}$ is a subgroup of $\mathbb{R}^2$. $\Diamond$

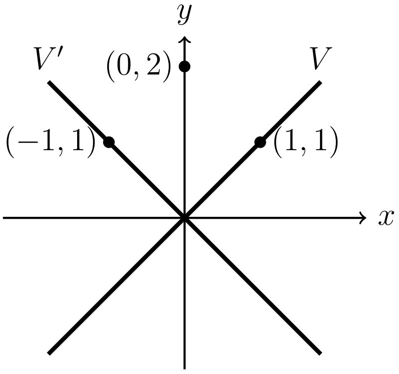


Figure 13: The lines $y = x$ and $y = -x$ are subgroups of $\mathbb{R}^2$. Their union is not.

9.2.8 Exercise Let G be a group and let H_1, H_2 be subgroups. Show that $H_1 \cap H_2$ is also a subgroup of G.

9.2.9 Example Let's take

$$C = \{(a, a^3) : a \in \mathbb{R}\}.$$

Clearly C is a subset of $\mathbb{R}^2$; in fact it is the graph $y = x^3$ (see Figure 14). But is it a subgroup? It contains the identity element $(0, 0)$. Moreover, $-(a, a^3) = (-a, (-a)^3)$. So C satisfies condition (c) for subgroups. But it doesn't satisfy condition (b). To show this we give a counterexample. Note that $(1, 1)$ is in C but $(1, 1) + (1, 1) = (2, 2)$ is not in C. ♢

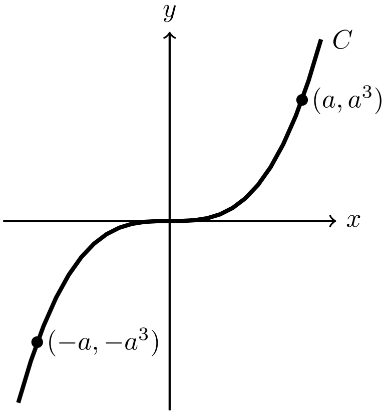


Figure 14: The set $C = \{(a, a^3) : a \in \mathbb{R}\}$ is the graph $y = x^3$. It satisfies conditions (a) and (c) for subgroups but not condition (b).

9.2.10 Example $\mathbb{Z}^2$ is a subgroup of $\mathbb{R}^2$. ♢

9.2.11 Example In Example 9.2.5 we saw that the line $y = x$ in $\mathbb{R}^2$ gives us a subgroup. In this example we would like to think about planes in $\mathbb{R}^3$ and whether they give us subgroups of $\mathbb{R}^3$. One way to specify a plane in $\mathbb{R}^3$ is via the point-normal equation which you probably met at school/college, but which we revise now.

Let Π be a plane in $\mathbb{R}^3$. Let $\mathbf{n}$ be a vector normal to Π (by normal to Π we simply mean perpendicular to Π) as in Figure 15. Choose and fix a point Q on the plane Π and let $\mathbf{u} = \overrightarrow{OQ}$ be the position vector of Q. Suppose now that P is any point on Π and let $\mathbf{x} = \overrightarrow{OP}$ be its position vector. Note that the vector $\overrightarrow{QP} = \mathbf{x} - \mathbf{u}$ is parallel to the plane and so perpendicular to $\mathbf{n}$. Hence $\mathbf{n} \cdot (\mathbf{x} - \mathbf{u}) = 0$. This is the *point-normal equation* for the plane:

$$\Pi \ : \ \mathbf{n} \cdot (\mathbf{x} - \mathbf{u}) = 0. \tag{10}$$

Here $\mathbf{n}$ is any (non-zero) vector normal to the plane, and $\mathbf{u}$ is the position vector of any point on the plane.

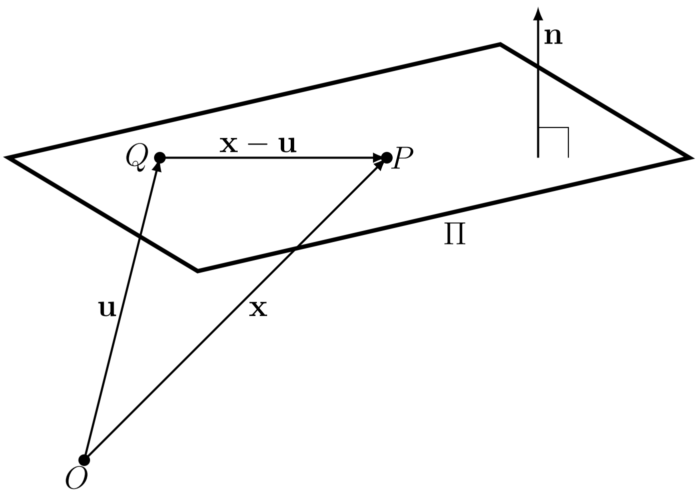


Figure 15: The *point-normal equation* of a plane. Here $\mathbf{n}$ is normal to the plane Π, Q is a fixed point on Π and $\mathbf{u}$ is its position vector. If P is any point on Π with position vector $\mathbf{x}$, then $\mathbf{x} - \mathbf{u}$ is parallel to the plane, and so $\mathbf{n} \cdot (\mathbf{x} - \mathbf{u}) = 0$.

The plane Π in (10) defines a set

$$V_\Pi = \{\mathbf{x} \in \mathbb{R}^3 : \mathbf{n} \cdot (\mathbf{x} - \mathbf{u}) = 0\}.$$

This is the set of points on the plane. It is a subset of the group $\mathbb{R}^3$. Is V_Π a subgroup? Of course to be a subgroup it has to contain the identity element of $\mathbb{R}^3$ which is $\mathbf{0}$. So we can choose $\mathbf{u} = \mathbf{0}$. This doesn't mean that our original Q was the origin. We're free to choose Q anywhere we like on Π, and if Π goes through the origin then we choose it to be the origin, and so take $\mathbf{u} = \mathbf{0}$. With this choice, we can simplify V_Π to obtain

$$V_\Pi = \{\mathbf{x} \in \mathbb{R}^3 : \mathbf{n} \cdot \mathbf{x} = 0\}.$$

Let's check that this is indeed a subgroup of V_Π. If $\mathbf{x}_1$, $\mathbf{x}_2 \in V_\Pi$ then $\mathbf{n} \cdot \mathbf{x}_i = 0$ so

$$\mathbf{n} \cdot (\mathbf{x}_1 + \mathbf{x}_2) = \mathbf{n} \cdot \mathbf{x}_1 + \mathbf{n} \cdot \mathbf{x}_2 = 0 + 0 = 0.$$

Thus $\mathbf{x}_1 + \mathbf{x}_2 \in V_\Pi$. Also

$$\mathbf{n} \cdot (-\mathbf{x}_1) = -\mathbf{n} \cdot \mathbf{x}_1 = -0 = 0.$$

Thus $-\mathbf{x}_1 \in V_\Pi$. Hence V_Π is a subgroup of $\mathbb{R}^3$.

Conclusion: *a plane defines a subgroup of* $\mathbb{R}^3$ *if and only if it passes through the origin.* $\Diamond$

9.2.12 Exercise Which lines in $\mathbb{R}^2$ define a subgroup? Justify your answer.

9.2.13 Example Recall that

$$\mathbb{C}^* = \{\alpha \in \mathbb{C} : \alpha \neq 0\}.$$

Geometrically, $\mathbb{C}^*$ is the whole complex plane minus the origin. We have observed before that $\mathbb{C}^*$ is a group (where the binary operation is multiplication of complex numbers). Let

$$\mathbb{S} = \{\alpha \in \mathbb{C} : |\alpha| = 1\}.$$

The set $\mathbb{S}$ is the set of all points in the complex plane with distance 1 from the origin. Of course this is just the unit circle (the circle centred at the origin with radius 1) as in Figure 16. Let us check that $\mathbb{S}$ is a subgroup of $\mathbb{C}^*$; it is clearly a subset. Of course the identity element of $\mathbb{C}^*$ is 1 and $|1| = 1$ so $1 \in \mathbb{S}$, which proves (a). Suppose α, $\beta \in \mathbb{S}$. Then $|\alpha| = 1$ and $|\beta| = 1$. From the properties of the absolute value [5] we have

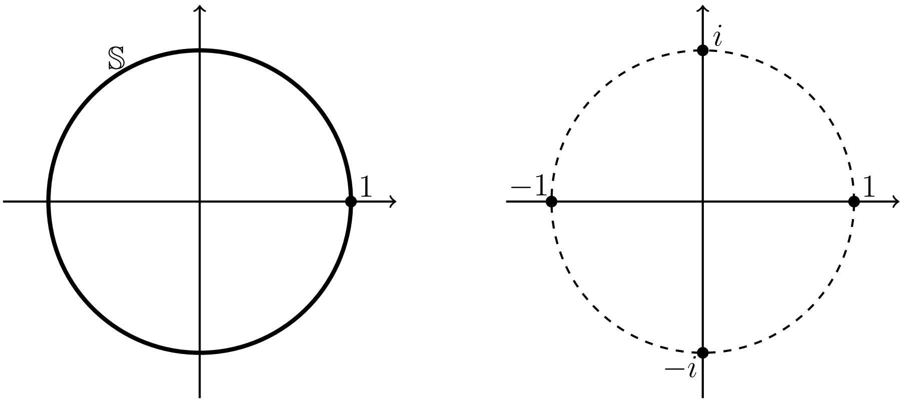


Figure 16: On the left, the group $\mathbb{S}$ which is just the unit circle. On the right, the subgroup of the fourth roots of unity.

$$|\alpha\beta| = |\alpha||\beta| = 1.$$

Thus $\alpha\beta \in \mathbb{S}$. This proves (b).

To check (c), suppose $\alpha \in \mathbb{S}$, so that $|\alpha| = 1$. Then, again from the properties of the absolute value,

$$|\alpha^{-1}| = \frac{1}{|\alpha|} = 1,$$

[5] At school/college you might have called $|\alpha|$ the *modulus* of α. Most mathematicians call $|\alpha|$ the *absolute value* of α.

so $\alpha^{-1} \in \mathbb{S}$. By Theorem 9.2.1, $\mathbb{S}$ is indeed a subgroup of $\mathbb{C}^*$.

We shall call $\mathbb{S}$ the *circle group*. Notice that $\mathbb{S}$ is an infinite subgroup of $\mathbb{C}^*$. But $\mathbb{C}^*$ has plenty of finite subgroups too. An example is $\{1, i, -1, -i\}$. This is the set of solutions to the equation $x^4 = 1$ (check). The solutions to $x^4 = 1$ are called the fourth roots of unity. Check for yourself that $\{1, i, -1, -i\}$ is a subgroup of $\mathbb{C}^*$ (and in fact a subgroup of $\mathbb{S}$). Can you find a finite subgroup of $\mathbb{C}^*$ that isn't a subgroup of $\mathbb{S}$? We'll return to roots of unity later. ♢

9.2.14 Exercise In the following, is H a subgroup of the group G? Give full justification.

Before you start answering: You might be wondering why the binary operation on G isn't specified. Mathematicians generally don't; you're expected to figure it out from the context [6].

(i) $G = \mathbb{R}$, $H = \mathbb{R}^*$.

(ii) $G = \mathbb{R}^*$, $H = \{1, -1\}$.

(iii) $G = \mathbb{C}$, $H = 2\mathbb{Z}$.

(iv) $G = \mathbb{C}$, $H = \{a + ai : a \in \mathbb{R}\}$.

(v) $G = \mathbb{C}^*$, $H = \{\alpha \in \mathbb{C}^* : \alpha^3 = 1\}$.

(vi) $G = \mathbb{Z}$, $H = \mathbb{Z}/2\mathbb{Z}$.

(vii) $G = \mathbb{R}[x]$, $H = \mathbb{Z}[x]$.

(viii) $G = \mathbb{R}[x]$, $H = \{f \in \mathbb{R}[x] : f(0) = 0\}$.

(ix) $G = \mathbb{R}[x]$, $H = \{f \in \mathbb{R}[x] : f(0) = 1\}$.

(x) $G = \mathbb{Z}/10\mathbb{Z}$, $H = \{\overline{0}, \overline{5}\}$.

9.2.15 Exercise Show that [7] the subgroups of $\mathbb{Z}/4\mathbb{Z}$ are $\{\overline{0}\}$, $\{\overline{0}, \overline{2}\}$ and $\mathbb{Z}/4\mathbb{Z}$.

[6] We know that addition makes $\mathbb{R}$ into a group, and multiplication doesn't. But are there really no other binary operations on $\mathbb{R}$ that make it into a group?

Yes, there are binary operations other than addition that make the set of real numbers into a group. But if this was anything other than the usual or obvious operation you'd have been told so.

[7] When answering a maths question, you should always be careful about what is being asked. Here you're being asked to show two things. The first is that the three listed sets are indeed subgroups. The second is that there aren't any other subgroups.

9.2.16 Exercise Show that the only subgroups of $\mathbb{Z}/3\mathbb{Z}$ are $\{\bar{0}\}$ and $\mathbb{Z}/3\mathbb{Z}$.

9.2.17 Exercise Let

$$D = \{\alpha \in \mathbb{C}^* : |\alpha| \le 1\}.$$

Sketch D. Show that D is not a subgroup of $\mathbb{C}^*$.

9.2.18 Exercise Let r be a positive real number. Let

$$\mathbb{S}_r = \{\alpha \in \mathbb{C}^* : |\alpha| = r\}.$$

What does $\mathbb{S}_r$ represent geometrically? For what values of r will $\mathbb{S}_r$ be a subgroup of $\mathbb{C}^*$?

9.3 Roots of Unity

Let n be a positive integer. Let $\zeta = e^{2\pi i/n}$. The n-th roots of unity are the solutions in $\mathbb{C}$ to the equation $x^n = 1$. Recall that there are exactly n of them:

$$1, \zeta, \zeta^2, \ldots, \zeta^{n-1}.$$

See Figure 10 for the roots of unity when $n = 3$ and $n = 4$ and note how they're distributed on the unit circle. Write

$$U_n = \{1, \zeta, \zeta^2, \ldots, \zeta^{n-1}\}.$$

That is, U_n is the set of n-th roots of unity.

9.3.1 Lemma U_n is a subgroup of $\mathbb{C}^*$ of order n.

Proof. Clearly U_n is a subset of $\mathbb{C}^*$ containing 1. Suppose a, $b \in U_n$. We want to check that $ab \in U_n$. But since $a^n = b^n = 1$ we know that $(ab)^n = a^n b^n = 1$. So ab is also an n-th root of unity and so $ab \in U_n$. Likewise, $(a^{-1})^n = (a^n)^{-1} = 1$. So a^{-1} is an n-th root of unity and so $a^{-1} \in U_n$. Thus U_n is indeed a subgroup of $\mathbb{C}^*$. Since it has n elements, it has order n.

Notation Warning. The notation U_n is not standard. Why do I point this out? You must always be careful with notation: do other people understand you? If you write $\mathbb{C}^*$ then this is standard notation and every mathematician will know what you mean. If you write U_n, others (e.g. your tutor and supervisor) will not know what you mean. They will of course know that the n-th roots of unity are a subgroup of $\mathbb{C}^*$, but they will not know that you're denoting this subgroup by U_n. If you write U_n, even in your homework, then you have to say what it is.

9.3.2 Exercise Is $U_2 \cup U_3$ a subgroup of $\mathbb{C}^*$?

9.4 Matrix Groups II

In Section 7.1 you met the general linear group

$$\mathrm{GL}_2(\mathbb{R}) = \{A \in M_{2\times 2}(\mathbb{R}) : \det(A) \neq 0\}.$$

This is group where the operation is multiplication of matrices. In this section we'll meet some subgroups of it.

9.4.1 Exercise Let

$$\mathrm{SL}_2(\mathbb{R}) = \{A \in M_{2\times 2}(\mathbb{R}) \mid \det(A) = 1\}.$$

Show that $\mathrm{SL}_2(\mathbb{R})$ is a group [8] (with respect to multiplication). This is known as the *special linear group* [9].

9.4.2 Exercise Show that

$$\{A \in M_{2\times 2}(\mathbb{Z}) : \det(A) \neq 0\}$$

is not a group under multiplication. Let

$$\mathrm{SL}_2(\mathbb{Z}) = \{A \in M_{2\times 2}(\mathbb{Z}) : \det(A) = 1\}.$$

Show that $\mathrm{SL}_2(\mathbb{Z})$ is a group. This is known as the *modular group* [10].

Now is a good time to revise Section 3.8 on rotation matrices. Recall that the matrix

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

represents anticlockwise rotation about the origin through an angle θ. It is geometrically clear that if compose two rotations about the origin we obtain a rotation about the origin. So it is natural to expect that rotations form a subgroup of $\mathrm{GL}_2(\mathbb{R})$, and indeed this is the case. We define

$$\mathrm{SO}_2(\mathbb{R}) = \{R_\theta \mid \theta \in \mathbb{R}\}.$$

This is called the *special orthogonal group.*

[8] Recall, the easiest way to show that a set is a group is to show that it is a subgroup of a something you already know to be a group.

[9] If you've done Exercise 3.6.5 then you'll see that $\mathrm{SL}_2(\mathbb{R})$ consists of the matrices that preserve area and orientation.

[10] The modular group is a very interesting group in mathematics. Google it!

9.4.3 Theorem $\mathrm{SO}_2(\mathbb{R})$ is a subgroup of $\mathrm{GL}_2(\mathbb{R})$.

Proof. First we have to check that $\mathrm{SO}_2(\mathbb{R})$ is a subset of $\mathrm{GL}_2(\mathbb{R})$. In other words, we want to check that every matrix R_θ has non-zero determinant. Note $\det(R_\theta) = \cos^2\theta + \sin^2\theta = 1$. Hence $\mathrm{SO}_2(\mathbb{R})$ is contained in $\mathrm{GL}_2(\mathbb{R})$. Also [11] $I_2 = R_0$, so $\mathrm{SO}_2(\mathbb{R})$ contains the identity element of $\mathrm{GL}_2(\mathbb{R})$.

Next we have to show that $\mathrm{SO}_2(\mathbb{R})$ is closed under multiplication. Consider two elements of $\mathrm{SO}_2(\mathbb{R})$ and call them R_θ and R_ϕ. Now R_θ and R_ϕ represent anticlockwise rotation about the origin through angles θ and ϕ. Thus $R_\theta R_\phi$ represents the combined effect of rotations through angles ϕ then θ. Clearly, from this geometric reasoning $R_\theta R_\phi = R_{\theta+\phi}$, but let's check this algebraically:

$$\begin{aligned}
R_\theta R_\phi &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \\
&= \begin{pmatrix} \cos\theta\cos\phi - \sin\theta\sin\phi & -\cos\theta\sin\phi - \cos\phi\sin\theta \\ \cos\theta\sin\phi + \cos\phi\sin\theta & \cos\theta\cos\phi - \sin\theta\sin\phi \end{pmatrix} \\
&= \begin{pmatrix} \cos(\theta+\phi) & -\sin(\theta+\phi) \\ \sin(\theta+\phi) & \cos(\theta+\phi) \end{pmatrix} \\
&= R_{\theta+\phi}.
\end{aligned}$$

Thus $\mathrm{SO}_2(\mathbb{R})$ is closed under multiplication.

Finally we must check that the inverse of every matrix in $\mathrm{SO}_2(\mathbb{R})$ is again in $\mathrm{SO}_2(\mathbb{R})$. Geometrically, it's easy to see that the inverse of R_θ is $R_{-\theta}$; I'll leave it to you to check this algebraically. This completes the proof.

Remark. It is clear (at least geometrically) that $R_\theta R_\phi = R_\phi R_\theta$. Thus $\mathrm{SO}_2(\mathbb{R})$ is an abelian subgroup of the non-abelian group $\mathrm{GL}_2(\mathbb{R})$. We saw this phenomenon before: the group D_4 is non-abelian, but its subgroup of rotations is abelian.

9.5 Differential Equations

Let $\mathcal{C}$ be the set of infinitely differentiable real functions. This probably sounds scary, but to reassure you I'll just point out that $\mathcal{C}$ contains all polynomials, as well as $\sin t$, $\cos t$, e^t, e^{-t}. It is a fact that $\mathcal{C}$ is an additive group. Don't worry about the proof; it depends on properties of differentiability that you'll see eventually in term 2 of *Analysis or Mathematical Analysis (term 2).* Addition in $\mathcal{C}$ is done in a common sense way. For example, if $f(t) = t^2 + \sin(t)$ and $g(t) = 2t^2 - e^t$ then

[11] In geometric terms, both I_2 and R_0 mean 'do nothing', so they must be equal.

$f(t)+g(t)=3t^2+\sin(t)-e^t$. The identity element is 0.

Let's dive straight into an example. We define the following subset

$$H=\left\{x(t)\in\mathcal{C}\ :\ t\frac{\mathrm{d}\,x}{\mathrm{d}\,t}-2x=0\right\}.$$

In other words, H is the set of infinitely differentiable functions $x(t)$ that satisfy the differential equation

$$t\frac{\mathrm{d}\,x}{\mathrm{d}\,t}-2x=0. \tag{11}$$

The function $x(t)=0$ (which is the identity element of $\mathcal{C}$) clearly satisfies (11) and so belongs to H. Suppose $x_1(t)$ and $x_2(t)$ are in H. Thus

$$t\frac{\mathrm{d}\,x_1}{\mathrm{d}\,t}-2x_1=0, \qquad t\frac{\mathrm{d}\,x_2}{\mathrm{d}\,t}-2x_2=0.$$

Let $x(t)=x_1(t)+x_2(t)$. By the properties of differentiation,

$$\frac{\mathrm{d}\,x}{\mathrm{d}\,t}=\frac{\mathrm{d}\,x_1}{\mathrm{d}\,t}+\frac{\mathrm{d}\,x_2}{dt}.$$

Thus

$$\begin{aligned} t\frac{\mathrm{d}\,x}{\mathrm{d}\,t}-2x &= t\left(\frac{\mathrm{d}\,x_1}{\mathrm{d}\,t}+\frac{\mathrm{d}\,x_2}{\mathrm{d}\,t}\right)-2(x_1+x_2)\\ &= t\frac{\mathrm{d}\,x_1}{\mathrm{d}\,t}-2x_1+t\frac{\mathrm{d}\,x_2}{\mathrm{d}\,t}-2x_2\\ &= 0. \end{aligned}$$

Therefore $x(t)\in H$. Similarly, using the properties of differentiation, you can show that if $x_1(t)\in H$ then $-x_1(t)\in H$. So H is a subgroup of $\mathcal{C}$.

Note that we didn't have to solve the differential equation to know that its set of solutions is a group; we merely used the properties of differentiation. But in fact it is easy to solve this particular equation using separation of variables. If you do that (try it) you'll find that

$$H=\{at^2:a\in\mathbb{R}\}.$$

Now check again that H forms an additive group.

9.5.1 Exercise Which of the following differential equations define subgroups of $\mathcal{C}$?

(i) $t\dfrac{\mathrm{d}\,x}{\mathrm{d}\,t} - 2x = t^3.$

(ii) $\dfrac{\mathrm{d}^2\,x}{\mathrm{d}\,t^2} - 5\dfrac{\mathrm{d}\,x}{\mathrm{d}\,t} + 6x = 0.$

(iii) $\dfrac{\mathrm{d}\,x}{\mathrm{d}\,t} - x^2 = 0.$

9.6 Non-Trivial and Proper Subgroups

It's very easy for you to prove the following proposition.

9.6.1 Proposition Let G be a group. Then G and $\{1\}$ are subgroups.

Here, of course, $\{1\}$ is the subset containing the identity element of G. We call $\{1\}$ the *trivial* subgroup of G; any other subgroup is called *non-trivial.* A subgroup of G that is not equal to G is called *proper.* The subgroups $\{1\}$ and G aren't particularly interesting, since they're always there. The interesting subgroups are the proper non-trivial subgroups.

9.6.2 Example The trivial subgroup of $\mathbb{Z}$ is $\{0\}$. Examples of a non-trivial subgroups are $\mathbb{Z}$ and $2\mathbb{Z}$. The subgroup $2\mathbb{Z}$ is proper and non-trivial. ◊

9.6.3 Example Consider the group U_4 which is the group of fourth roots of unity. Thus $U_4 = \{1, i, -1, -i\}$; of course the binary operation is multiplication. The trivial subgroup is $\{1\}$. We note that $U_2 = \{1, -1\}$ is a non-trivial proper subgroup. Are there any others?

Suppose H is *another* non-trivial proper subgroup of U_4. Then $1 \in H$, as subgroups always contain the identity element. Since H is non-trivial, and $H \neq \{1, -1\}$, it must contain either i or $-i$. Suppose H contains i. Then H contains $i^2 = -1$ and $i^3 = -i$. Therefore $H = U_4$, which contradicts the assumption that H is proper. Similarly if H contains $-i$ then $H = U_4$ (check). Therefore the only non-trivial proper subgroup of U_4 is $U_2 = \{1, -1\}$. ◊

9.6.4 Exercise For what values of m does $\mathbb{Z}/m\mathbb{Z}$ have non-trivial proper subgroups? Try out a few examples and see if you can make a conjecture. Can you prove your conjecture?

9.7 Second preview of Lagrange's Theorem

Here is another consequence of Lagrange's Theorem. The relationship between this version and the earlier one (Theorem 8.2.2) will be explained once we have studied cyclic groups in the next chapter.

9.7.1 Corollary (of Lagrange's Theorem which is to follow in Lecture 13.) Let G be a finite group, and H a subgroup of G. Then the order of H divides the order of G.

9.7.2 Example We saw in Example 9.6.3 that U_4, the group of 4-th roots of unity, contains U_2, the group of square-roots of unity. Now U_2 has order 2, U_4 has order 4. Lagrange's Theorem tells that the order of U_2 must divide the order of U_4 which is correct. $\Diamond$

9.7.3 Example Recall that D_4 has order 8. In Figure 9 we listed the ten subgroups of D_4. These have orders 1, 2, 4 and 8. This is consistent with Lagrange's Theorem. $\Diamond$

9.7.4 Exercise Let G be a group, and suppose the order of G is p where p is a prime. Show that the only subgroups of G are $\{1\}$ and G.

Lecture 10 - Cyclic groups and cyclic subgroups

Cyclic groups are the simplest groups to understand. These are those groups where the elements are powers of one particular element. We say they are *generated* by that element.

10.1 The cyclic subgroup generated by an element

10.1.1 Theorem Let G be a group, and let g be an element of G. Write $\langle g\rangle$ for the set

$$\langle g\rangle = \{g^n \mid n \in \mathbb{Z}\} = \{\dots, g^{-2}, g^{-1}, 1, g, g^2, g^3, \dots\}.$$

Then $\langle g\rangle$ is a subgroup of G.

10.1.2 Proof We'll prove this using Theorem 9.2.1. So we need to show that the subset $\langle g\rangle$ of G satisfies conditions (a), (b) and (c) in that theorem.

Let's start with (a). We have that $g^0 = 1$ and, since $0 \in \mathbb{Z}$ this means that the identity from G, $1 \in \langle g\rangle$.

Now for (b). We need to show that if $a, b \in \langle g\rangle$ then $ab \in \langle g\rangle$. We do this as follows.

Let $a, b \in \langle g\rangle$. Then $a = g^m$ and $b = g^n$ for some $m, n \in \mathbb{Z}$. Then $ab = g^m g^n = g^{m+n}$ by Theorem 6.2.6. Since $m + n \in \mathbb{Z}$ this means that $ab \in \langle g\rangle$.

Finally (c). We need to show that if $a \in \langle g\rangle$ then $a^{-1} \in \langle g\rangle$:

Let $a \in \langle g\rangle$. The $a = g^m$ for some $m \in Z$. Then $a^{-1} = (g^m)^{-1} = g^{-m}$ by Theorem 6.2.6. Since $-m \in \mathbb{Z}$ this means that $a^{-1} \in \langle g\rangle$. $\Diamond$

10.1.3 Definition We call $\langle g\rangle$ the *cyclic subgroup* generated by g. If $G = \langle g\rangle$ then we call G a *cyclic group*, and we say that g is a *generator* of G.

10.1.4 Example As roots of unity are fresh in your mind, let's start with them. The group of n-th roots of unity U_n is cyclic, since every element is a power of $\zeta = e^{2\pi i/n}$; indeed the elements of U_n are precisely

$$\zeta^0 = 1, \zeta, \zeta^2, \dots, \zeta^{n-1}.$$

Thus $U_n = \langle\zeta\rangle$ and ζ is a generator.

Let's consider U_6, and calculate the cyclic subgroup generated by each element. Write $\zeta = e^{2\pi i/6}$. Note that $\zeta^6 = 1$. Consider for example $h = \zeta^2$. The powers of h are $1, h, h^2$. Indeed, note that $h^3 = \zeta^6 = 1$. Thus

$$h^4 = h, \; h^5 = h^2, \; h^6 = 1, \; h^7 = h, \ldots .$$

What about h^{-1}. We know that $h^3 = 1$; multiplying both sides by h^{-1} we deduce that $h^{-1} = h^2$. Thus

$$h^{-2} = h, \; h^{-3} = 1, \; h^{-4} = h^2, \; h^{-5} = h, \ldots .$$

Thus the distinct powers of h are $1, h, h^2$, which are $1, \zeta^2, \zeta^4$. We can't write all the elements of U_6 as powers of h; therefore h is not a generator of U_6.

However, let us consider $g = \zeta^5$. We can write the powers of g and simplify them using the fact that $\zeta^6 = 1$. For example,

$$g^2 = \zeta^{10} = \zeta^6\zeta^4 = \zeta^4.$$

We find that $1, g, g^2, g^3, g^4, g^5$ are respectively, $1, \zeta^5, \zeta^4, \zeta^3, \zeta^2, \zeta$. Since every element of U_6 is a power of $g = \zeta^5$, we see that g is also a generator of U_6. Table 2 lists the elements of U_6 and the subgroups they generate.

g	$\langle g \rangle$
1	$\{1\}$
ζ	$\{1, \zeta, \zeta^2, \zeta^3, \zeta^4, \zeta^5\}$
ζ^2	$\{1, \zeta^2, \zeta^4\}$
ζ^3	$\{1, \zeta^3\}$
ζ^4	$\{1, \zeta^2, \zeta^4\}$
ζ^5	$\{1, \zeta, \zeta^2, \zeta^3, \zeta^4, \zeta^5\}$

Table 2: The six elements of U_6 and the cyclic subgroups they generate.

◊

10.1.5 Example For each element of the group $\mathbb{Z}/m\mathbb{Z}$, we write down the cyclic group it generates. Note that since $\mathbb{Z}/m\mathbb{Z}$ is an additive group, the subgroup generated by g is $\langle g \rangle = \{ng : n \in \mathbb{Z}\}$. That is, it is the set of multiples of g rather than the set of powers of g. See Table 3. ◊

$\overline{a}$	$\langle\overline{a}\rangle$
$\overline{0}$	$\{\overline{0}\}$
$\overline{1}$	$\{\overline{0},\overline{1},\overline{2},\overline{3},\overline{4},\overline{5}\}$
$\overline{2}$	$\{\overline{0},\overline{2},\overline{4}\}$
$\overline{3}$	$\{\overline{0},\overline{3}\}$
$\overline{4}$	$\{\overline{0},\overline{2},\overline{4}\}$
$\overline{5}$	$\{\overline{0},\overline{1},\overline{2},\overline{3},\overline{4},\overline{5}\}$

Table 3: The six elements of $\mathbb{Z}/6\mathbb{Z}$ and the cyclic subgroups that they generate.

10.1.6 Example Recall the group D_4 of the symmetries of the square. It has 8 elements. It's easy to write down the subgroup generated by each element (see Section 5.4 to remind yourself of the notation):

g	$\langle g\rangle$
1	$\{1\}$
ρ_1	$\{1,\rho_1,\rho_2,\rho_3\}$
ρ_2	$\{1,\rho_2\}$
ρ_3	$\{1,\rho_1,\rho_2,\rho_3\}$
σ_0	$\{1,\sigma_0\}$
σ_1	$\{1,\sigma_1\}$
σ_2	$\{1,\sigma_2\}$
σ_3	$\{1,\sigma_3\}$

None of the elements of D_4 generates it. We see that D_4 is not a cyclic group. $\Diamond$

10.1.7 Theorem Cyclic groups are abelian.

Proof. Let G be a cyclic group generated by g. Let a, b be elements of G. We want to show that $ab = ba$. Now, $a = g^m$ and $b = g^n$ for some integers m and n. So, $ab = g^m g^n = g^{m+n}$ and $ba = g^n g^m = g^{n+m}$. But $m + n = n + m$ (addition of integers is commutative). So $ab = ba$.

Whilst working through the above examples, you will have noticed a pattern about $\langle g\rangle$, which we state in the following theorem.

10.1.8 Theorem Let G be a group and let g be an element of finite order n. Then

$$\langle g\rangle = \{1, g, g^2, \ldots, g^{n-1}\}.$$

In particular, the order of the subgroup $\langle g\rangle$ is equal to the order of g.

Proof. Observe that $\langle g\rangle$ is a set, and $\{1, g, \dots, g^{n-1}\}$ is a set. We want to show that these sets are the same.

Whenever you have two sets, A and B, and you want to prove that they're equal, one way to do this is to show that every element of A belongs to B and every element of B belongs to A. You will see this principle again and again throughout your undergraduate career.

Let's apply this principle in our situation. By definition,

$$\langle g\rangle = \{g^n : n \in \mathbb{Z}\} = \{\dots, g^{-2}, g^{-1}, 1, g, g^2, g^3, \dots\}.$$

That is $\langle g\rangle$ is the set of all powers of g. It is obvious that every element of $\{1, g, \dots, g^{n-1}\}$ belongs to $\langle g\rangle$. What about the other way round. Suppose that h is an element of $\langle g\rangle$. We want to show that h is an element of $\{1, g, \dots, g^{n-1}\}$. We can write $h = g^m$ where m is an integer (positive or negative). We want to show that $h = g^r$ where r is one of $0, 1, 2, \dots, n-1$. For this we will use the *division algorithm* which you met in *Foundations/Sets and Numbers*. We can write

$$m = qn + r, \qquad q, r \in \mathbb{Z}, \quad 0 \le r < n.$$

Here we simply divided m by n; the integers q, r are respectively the quotient and the remainder. Thus

$$h = g^m = g^{qn+r} = (g^n)^q \cdot g^r.$$

However, $g^n = 1$ since g has order n. So $h = g^r$. Since $0 \le r < n$, we see that r is one of $0, 1, \dots, n-1$. Therefore h is in $\{1, g, \dots, g^{n-1}\}$. By our principle, we see that $\langle g\rangle = \{1, g, \dots, g^{n-1}\}$.

10.1.9 Exercise In each of the following groups G, write down the cyclic subgroup generated by g.

(a) $G = \mathbb{S}$, $g = \exp(2\pi i/7)$.

(b) $G = \mathbb{Z}/12\mathbb{Z}$, $g = \overline{8}$.

(c) $G = \mathrm{GL}_2(\mathbb{R})$, $g = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$.

10.1.10 Exercise Which of the following groups G are cyclic? Justify your answer for each, and if G is cyclic then write down a generator.

(a) $G = k\mathbb{Z}$ (where k is a non-zero integer).

(b) $G = \mathbb{Z}/m\mathbb{Z}$ (where m is a positive integer).

(c) D_3.

10.1.11 Exercise In this exercise, you will show using contradiction that $\mathbb{R}^*$ is not cyclic. Suppose that it is cyclic and let $g \in \mathbb{R}^*$ be a generator. Then $\mathbb{R}^* = \langle g \rangle$. In particular, $|g|^{1/2} \in \mathbb{R}^*$ and so $|g|^{1/2} = g^m$ for some integer m. Show that the only solutions to this equation are $g = \pm 1$. Where's the contradiction?

10.1.12 Exercise In this exercise you'll show that $\mathbb{Q}$ is not cyclic. Let a, b be integers with $b \neq 0$. Let p be a prime that does not divide b. Show that $1/p$ cannot be written in the form na/b with n an integer. Deduce that $\mathbb{Q}$ is not cyclic.

10.1.13 Exercise Show that $\mathbb{S}$ is not cyclic.

10.2 Lagrange previews revisited

You have seen two corollarys of Lagrange's Theorem.

This was Corollary 8.2.2:

Let G be a finite group, and let g be an element of G. The order g divides the order of G.

And this was Corollary 9.7.1:

Let G be a finite group, and H a subgroup of G. Then the order of H divides the order of G.

In fact, assuming Corollary 9.7.1 we can deduce Corollary 8.2.2 as follows:

Let G be a finite group and g an element of G. Suppose g has order n. By Theorem 10.1.8, the cyclic subgroup generated by g, denoted $\langle g \rangle$, also has order n. By Version 2, the order of the subgroup $\langle g \rangle$ divides the order of G. Hence n divides the order of G, which is what we wanted to prove. $\Diamond$

This doesn't mean that we've proved Corollary 8.2.2 of Lagrange's Theorem. It does mean that once we prove Corollary 9.7.1, then we will have also proved Version 1.

10.2.1 Exercise Let G be a group of order p, where p is a prime number. Let H be a subgroup. Show that H must either equal G or the trivial subgroup $\{1\}$. Deduce that if $g \in G$ is not the identity element, then $G = \langle g \rangle$.

10.3 Subgroups of $\mathbb{Z}$

The first thing to note about $\mathbb{Z}$ is that it is cyclic. Remember that $\mathbb{Z}$ is an additive group so we need to change notation for this. If G is an additive group, and g is an element of G then

$$\langle g \rangle = \{ng : n \in \mathbb{Z}\} = \{\dots, -2g, -g, 0, g, 2g, 3g, \dots\}.$$

Thus $\mathbb{Z} = \langle 1 \rangle$ and so it is cyclic. In fact, it is infinite and cyclic, unlike for example, U_n.

10.3.1 Lemma Let k be an integer. Write

$$k\mathbb{Z} = \{ka : a \in \mathbb{Z}\}.$$

Then $k\mathbb{Z}$ is a subgroup of $\mathbb{Z}$.

Proof. You can prove this in a similar way to Example 9.2.3. However, it is quicker to note that $k\mathbb{Z} = \langle k \rangle$, and so is a subgroup by Theorem 10.1.1.

Note that $0\mathbb{Z} = \{0\}$ has only the identity element. Also

$$\begin{aligned}(-k)\mathbb{Z} &= \{\dots, -2(-k), -(-k), 0, -k, 2(-k), \dots\}\\ &= \{\dots, 2k, k, 0, -k, -2k, \dots\}\\ &= \{\dots, -2k, -k, 0, k, 2k, \dots\}\\ &= k\mathbb{Z}\end{aligned}$$

because the order of elements in a set does not matter. In other words,

$$-\mathbb{Z} = \mathbb{Z}, \qquad -2\mathbb{Z} = 2\mathbb{Z}, \qquad -3\mathbb{Z} = 3\mathbb{Z}, \dots.$$

So we have an infinite list of subgroups

$$\{0\}, \quad \mathbb{Z}, \quad 2\mathbb{Z}, \quad 3\mathbb{Z}, \quad 4\mathbb{Z}, \dots$$

and we want to know if they're all the subgroups of $\mathbb{Z}$. The following theorem tells us that they are.

10.3.2 Theorem Any subgroup of $\mathbb{Z}$ has the form $k\mathbb{Z}$ for some non-negative [12] integer k. In particular, all subgroups of $\mathbb{Z}$ are cyclic.

Proof. Let H be a subgroup of $\mathbb{Z}$. We want to show that there is a non-negative integer k such that $H = k\mathbb{Z}$. We divide the proof into two cases. The first case is when H is the subgroup $\{0\}$. Then $H = 0\mathbb{Z}$ and we've done what we wanted.

So let's look at the second case where H has non-zero elements. If a is a non-zero element of H then because H is a(n additive) group, $-a$ is also a non-zero element of H but it has a different sign. So we know for sure that H has some positive elements. Let k be the *smallest positive element of* H. We will prove that $H = k\mathbb{Z}$.

Whenever you have two sets, A and B, and you want to prove that they're equal, one way to do it is to show that every element of A belongs to B and every element of B belongs to A.

As k is in H, we know by Theorem 6.3.1 that all the multiples of k belong to H. Thus every element of $k\mathbb{Z}$ belongs to H. We must show the converse: every element of H is a multiple of k.

Let a be an element of H. By the **division algorithm** which you have met in *Foundations/Sets and Numbers*, we can write

$$a = qk + r, \qquad \text{where } q,\, r \text{ are integers and } 0 \le r < k.$$

Now a is in H; qk is in H because it is a multiple of $k \in H$. So $r = a - qk$ is also in H. But $0 \le r < k$, and k is the *smallest positive* element of H. If $r > 0$ then it would be an even smaller positive element of H giving us a contradiction. So $r = 0$. Hence $a = qk$ is a multiple of k.

Thus we've also shown that every element of H is a multiple of k and so belongs to $k\mathbb{Z}$. Hence $H = k\mathbb{Z}$, as required.

10.3.3 Exercise The subgroups of $\mathbb{Z}^2$ are harder to describe. Write down a few.

[12] The *non-negative integers* are $0, 1, 2, 3, \ldots$.

Lecture 11 - Symmetric Groups

Probably the most interesting groups have elements that are functions. Matrix groups are examples, and the groups we are about to meet in this lecture, the symmetric groups, are other examples. It turns out that every finite group can be thought of as a subgroup of one of the symmetric groups. This is called Cayley's Theorem, where this statement is made precise. You won't meet this until later courses in group theory.

11.1 Motivation

Let A be a set, and let f, g be functions from A to itself. We know that we can compose f, g to obtain $f \circ g$ which is also a function from A to itself. We shall write $\text{Map}(A)$ for the set of functions from A to itself. Then $\circ$ is a binary operation on $\text{Map}(A)$. And it's natural to ask if this makes $\text{Map}(A)$ into a group. After all, we know by Lemma 2.6.4 that composition of functions is associative. The following example will help clarify these ideas.

11.1.1 Example Let $A = \{1, 2\}$. You will quickly convince yourself that there are only four functions from A to itself, which are given in Figure 17.

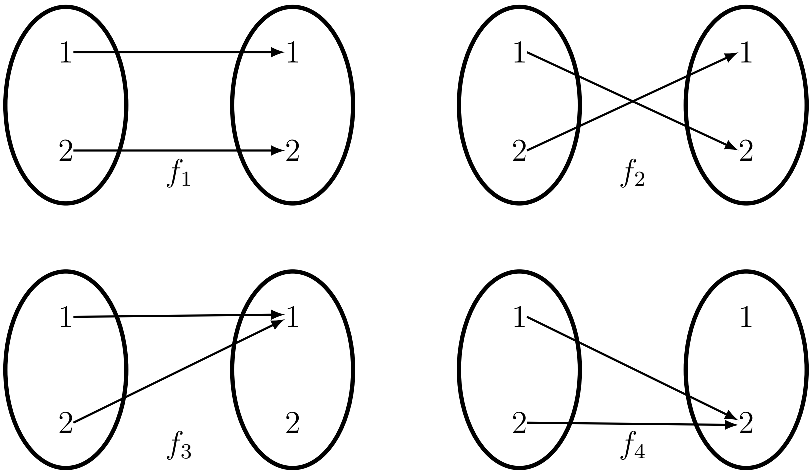


Figure 17: f_1, f_2, f_3 and f_4 are the four functions from $\{1, 2\}$ to itself.

Thus $\text{Map}(A) = \{f_1, f_2, f_3, f_4\}$. Is $\text{Map}(A)$ a group with respect to composition of functions? Here is the composition table for $\text{Map}(A)$:

$\circ$	f_1	f_2	f_3	f_4
f_1	f_1	f_2	f_3	f_4
f_2	f_2	f_1	f_4	f_3
f_3	f_3	f_3	f_3	f_3
f_4	f_4	f_4	f_4	f_4

Make sure you understand the table. The entry for $f_i \circ f_j$ is at the intersection of the i-th row and j-th column. As always, $f_i \circ f_j$ means apply f_j first then f_i. We know that composition of functions is associative by Lemma 2.6.4. Moreover, it is clear from the table that f_1 is the identity element. But f_3 and f_4 don't have inverses; we can't combine either of them with any of the four functions to obtain the identity f_1.

But if you look carefully at the table, you will see a group with respect to composition. It is the subset: $\{f_1, f_2\}$. We already know why f_1, f_2 have inverses (which in this case happen to be f_1 and f_2 respectively), and f_3, f_4 don't: the functions f_1, f_2 are bijections and f_3 and f_4 are not. This was Theorem 2.7.2 which says that a function f from a set X to a set Y has an inverse if and only if it is a bijection. In the example we have been looking at $X = Y = \{1, 2\}$.

$\Diamond$

11.2 The Symmetric Group on a general set A

Let A be a set. We shall denote the set of bijections from A to itself by $\mathrm{Sym}(A)$.

11.2.1 Example In Example 11.1.1 we wrote down all the functions from $A = \{1, 2\}$ to itself and found that exactly two of these are bijections. These were called f_1 and f_2 in Figure 17. Hence $\mathrm{Sym}(A) = \{f_1, f_2\}$. Note that $f_1 = \mathrm{id}_A$, the function which sends every element of A to itself. In that example, we noted that $\{f_1, f_2\}$ is a group under composition with f_1 being the identity element. Check this again, and note that the group is abelian. $\Diamond$

11.2.2 Theorem Let A be a set. Then $(\mathrm{Sym}(A), \circ)$ is a group with id_A as the identity element.

We call $\mathrm{Sym}(A)$ the *symmetric group* on A.

Proof. By Assignment 1, Question 3, $\mathrm{Sym}(A)$ is closed under composition. Moreover, composition of functions is associative by Lemma 2.6.4.

Clearly id_A is a bijection and so is in $\mathrm{Sym}(A)$. We want to check that id_A is the identity for composition, which means that for any $f \in \mathrm{Sym}(A)$ we want $f \circ \mathrm{id}_A = \mathrm{id}_A \circ f = f$. Note

$$(f \circ \mathrm{id}_A)(x) = f(\mathrm{id}_A(x)) = f(x), \qquad (\mathrm{id}_A \circ f)(x) = \mathrm{id}_A(f(x)) = f(x).$$

Thus $f \circ \mathrm{id}_A = \mathrm{id}_A \circ f = f$ holds.

Finally we want every element of $\mathrm{Sym}(A)$ to have an inverse in $\mathrm{Sym}(A)$. This is true by Theorem 2.7.2.

11.2.3 Exercise Let $f : \mathbb{Z} \to \mathbb{Z}$ and $g : \mathbb{R} \to \mathbb{R}$ be given by $x \mapsto 2x$. Show that $f \notin \mathrm{Sym}(\mathbb{Z})$ but $g \in \mathrm{Sym}(\mathbb{R})$. Write down g^n for integers n.

11.2.4 Exercise Let $f : \mathbb{C} \to \mathbb{C}$, $g : \mathbb{C} \to \mathbb{C}$, $h : \mathbb{C} \to \mathbb{C}$ be given by $f(z) = z + 1$, $g(z) = z + i$, $h(z) = iz$. Describe f, g, h geometrically. Show that f, g, h are in $\mathrm{Sym}(\mathbb{C})$. Show that f and g commute. What about f and h or g and h? What are the orders of f, g and h?

11.3 S_n

We define S_n to be the group $\mathrm{Sym}(\{1, 2, \ldots, n\})$. We call S_n the *n-th symmetric group*. In Example 11.2.1 we found that S_2 is a group of order 2.

11.3.1 Theorem S_n has order $n!$.

Proof. S_n is the set of bijections from $\{1, 2, \ldots, n\}$ to itself. So we want to count these bijections. It's clear that any injective function from $\{1, 2, \ldots, n\}$ to itself will be surjective (because if distinct elements get sent to distinct elements then the number of elements that get 'hit' must be n, i.e. all elements of $\{1, 2, \ldots, n\}$ are 'hit' and the function is surjective).

So let's count the injections. Let f be an injection from $\{1, 2, \ldots, n\}$ to itself. Then $f(1)$ can be any of $1, 2, \ldots, n$; that is, there are n choices for $f(1)$. If we fix $f(1)$ then $f(2) \neq f(1)$. So there are $n - 1$ choices for $f(2)$ once we've chosen $f(1)$. Likewise there are $n - 2$ choices for $f(3)$ once we've chosen $f(1)$ and $f(2)$. It is now clear that the number of injections is

$$n \times (n - 1) \times \cdots \times 1 = n!.$$

The elements of S_n are called *permutations*. One way of representing permutations is to use diagrams such as those for f_1, $f_2 \in S_2$ in Figure 17. The following is a more economical way. Let $a_1, a_2, \dots, a_n$ be the numbers $1, 2, \dots, n$ in some order. Then

$$\begin{pmatrix} 1 & 2 & \cdots & n \\ a_1 & a_2 & \cdots & a_n \end{pmatrix}$$

represents the unique permutation in S_n that sends 1 to a_1, 2 to a_2, …, and n to a_n.

11.3.2 Example S_2 has two elements:

$$\begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}.$$

These are respectively the same as f_1, f_2 in Figure 17. The first of these is the identity element. We noted in Example 11.2.1 that $S_2 = \mathrm{Sym}(\{1, 2\})$ is abelian. $\diamondsuit$

11.3.3 Example We know from Theorem 11.3.1 that S_3 has 6 elements. These are

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$$
$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}.$$

Again, the first of these is the identity element. It is important that you know what the notation means and how to multiply two permutations written in this notation, so let's have some practice. Let

$$\rho = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}, \qquad \mu = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}.$$

Never forget that these are bijections from $\{1, 2, 3\}$ to itself. To find out what ρ does, look at the columns. ρ is the function that sends 1 to 3, 2 to 1 and 3 to 2. Thus

$$\rho(1) = 3, \qquad \rho(2) = 1, \qquad \rho(3) = 2. \tag{12}$$

Likewise,

$$\mu(1) = 1, \qquad \mu(2) = 3, \qquad \mu(3) = 2.$$

Now let us compute $\rho\mu$. As always, this means apply μ first then ρ. So

$$\begin{aligned} (\rho\mu)(1) &= \rho(\mu(1)) = \rho(1) = 3; \\ (\rho\mu)(2) &= \rho(\mu(2)) = \rho(3) = 2; \\ (\rho\mu)(3) &= \rho(\mu(3)) = \rho(2) = 1. \end{aligned}$$

Thus

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

Similarly,

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

Note that $\mu\rho \neq \rho\mu$, so S_3 is non-abelian. How do we compute ρ^{-1}? From (12) we find

$$1 = \rho^{-1}(3), \qquad 2 = \rho^{-1}(1), \qquad 3 = \rho^{-1}(2).$$

We rearrange this:

$$\rho^{-1}(1) = 2, \qquad \rho^{-1}(2) = 3, \qquad \rho^{-1}(3) = 1.$$

Hence

$$\rho^{-1} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}.$$

◊

11.3.4 Exercise Write down a multiplication table for S_3 and determine the orders of all six elements checking that your answers are consistent with what we saw in Corollary 8.2.2, namely that, because S_3 is finite, the order of each element divides the order of the group.

11.3.5 Exercise Let ρ and τ be the following permutations:

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

Compute ρ^{-1}, $\rho\tau$, τ^2.

11.3.6 Exercise Show that S_n is non-abelian for $n \geq 3$.

11.4 Cycle Notation

Let $a_1, a_2, \ldots, a_m$ be distinct elements of the set $\{1, 2, \ldots, n\}$. By the notation

$$(a_1, a_2, \ldots, a_m) \tag{13}$$

we mean the element of S_n that takes a_1 to a_2, a_2 to a_3, $\ldots$, a_{m-1} to a_m and a_m back to a_1, and fixes all other elements of $\{1, 2, \ldots, n\}$. The permutation (13) is called a *cycle of length m*. A cycle of length 2 is called a *transposition*.

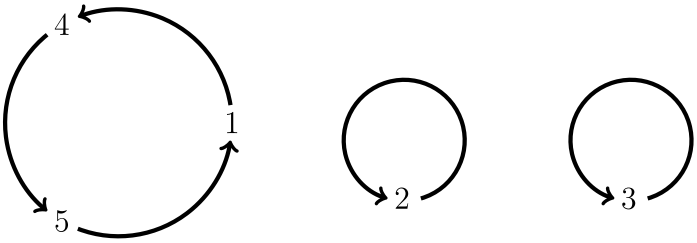


Figure 18: The cycle $(1,4,5) \in S_5$.

11.4.1 Example Let $\mu = (1,4,5) \in S_5$. The cycle μ is of length 3 and is illustrated in Figure 18.
We can write $(1,4,5)$ using our old notation:

$$(1,4,5) = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 2 & 3 & 5 & 1 \end{pmatrix}.$$

Notice that $(1,4,5) = (4,5,1) = (5,1,4)$. However, $(1,4,5) \neq (1,5,4)$.

The transposition $(1,5) \in S_5$ is given in Figure 19.

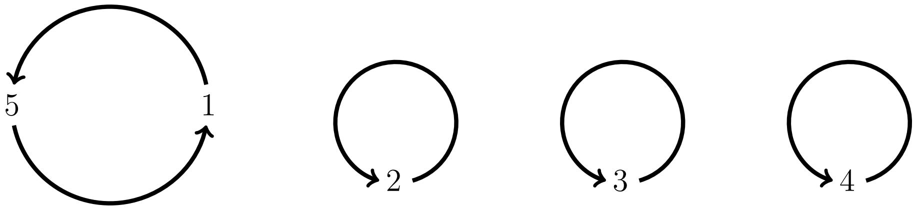


Figure 19: The transposition $(1,5) \in S_5$. This merely swaps 1 and 5, and fixes all other elements.

In our old notation, the transposition $(1,5)$ is written as follows:

$$(1,5) = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 2 & 3 & 4 & 1 \end{pmatrix}.$$

Note that $(1,5) = (5,1)$.

Finally (1) is the cycle that takes 1 to itself and fixes all the other elements. Clearly $(1) = (2) = (3) = (4) = (5) = \mathrm{id}$ is nothing other than the identity permutation. $\Diamond$

I hope that the above example has convinced you that cycle notation is simultaneously more concise and and more transparent than the old notation. If so, the following theorem, where we show that every permutation can be written as a product of disjoint cycles will come as a pleasant surprise!

Before we state it, what does *disjoint* mean? Two cycles $(a_1, a_2, \dots, a_n)$ and $(b_1, b_2, \dots, b_m)$ are said to be disjoint if $a_i \neq b_j$ for all integers i, j with $1 \leq i \leq n$ and $1 \leq j \leq m$. What does *product* mean? The product of two permutations is their composition as functions.

11.4.2 Theorem Every permutation can be written as a product of disjoint cycles.

Proof. Let ρ be an element of S_n. Consider the sequence

$$1, \quad \rho 1, \quad \rho^2 1, \quad \rho^3 1, \dots$$

Every term in this infinite sequence is contained in the finite set $\{1, 2, \dots, n\}$. Thus the sequence must contain repetition. Let $\rho^u 1$ be the first term in the sequence that has already appeared. Thus $\rho^u 1 = \rho^v 1$ for some $0 \leq v < u$. Apply ρ^{-v} to both sides. We obtain $\rho^{u-v} 1 = 1$. Note that $0 < u - v \leq u$. If $u - v < u$, then $\rho^{u-v} 1$ is in fact the first term in the sequence that has already appeared, which contradicts our assumption. Therefore, $u - v = u$ and so $v = 0$. Hence $\rho^u 1 = 1$, and $1, \rho 1, \dots, \rho^{u-1} 1$ are distinct.

Let μ_1 be the cycle of length u

$$\mu_1 = \left(1, \rho 1, \rho^2 1, \dots, \rho^{u-1} 1\right).$$

It is clear that μ_1 has the same effect as ρ on the elements $1, \rho 1, \dots, \rho^{u-1} 1$.

Now let a be the first element of the set $\{1, 2, \dots, n\}$ not appearing in the list $1, \rho 1, \dots, \rho^{u-1} 1$. Repeat the above argument with the sequence

$$a, \rho a, \rho^2 a, \rho^3 a, \dots.$$

We deduce the existence of a cycle

$$\mu_2 = \left(a, \rho a, \dots, \rho^{v-1} a\right)$$

such that μ_2 and ρ have the same effect on the elements $a, \rho a, \dots, \rho^{v-1} a$. Let us show that μ_1 and μ_2 are disjoint. Suppose otherwise. Then $\rho^i 1 = \rho^j a$ for some

$0 \leq i < u$ and $0 \leq j < v$. Now apply ρ^{v-j} to both sides to obtain $\rho^k 1 = a$ where $k = i + v - j$. This contradicts our assumption that a does not appear in the list $1, \rho 1, \ldots, \rho^{u-1} 1$. Hence the cycles μ_1 and μ_2 are disjoint. Now the product $\mu_1 \mu_2$ has the same effect as ρ on the elements $1, \rho 1, \ldots, \rho^{u-1} 1, a, \rho a, \ldots, \rho^{v-1} a$.

We repeat the process, starting with the first element of $\{1, 2, \ldots, n\}$ not appearing in either cycle μ_1, μ_2 to construct a μ_3 that is disjoint from both μ_1 and μ_2, etc. As the set $\{1, 2, \ldots, n\}$ is finite, this process must terminate eventually with some μ_r. The product of disjoint cycles $\mu_1 \mu_2 \ldots \mu_r$ has the same effect on $\{1, \ldots, n\}$ as ρ. Therefore

$$\rho = \mu_1 \mu_2 \cdots \mu_r.$$

$\lozenge$

Let's see an example where we write down a permutation as a product of cycles.

11.4.3 Example Let

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

Write ρ as a product of disjoint cycles.

Answer: We start with 1 are repeatedly apply ρ to it:

$$1 \mapsto 5 \mapsto 8 \mapsto 3 \mapsto 1.$$

Therefore ρ contains the cycle $(1, 5, 8, 3)$. Now we start with an element of the set $\{1, 2, \ldots, 8\}$ that is not contained in the cycle $(1, 5, 8, 3)$. For example start with 2 and repeatedly apply ρ to it:

$$2 \mapsto 7 \mapsto 6 \mapsto 2.$$

So ρ also contains the cycle $(2, 7, 6)$. Note that the cycles $(1, 5, 8, 3)$ and $(2, 7, 6)$ are disjoint, and ρ contains the product (or composition) $(1, 5, 8, 3)(2, 7, 6)$. There still remains one element of the set $\{1, 2, \ldots, 8\}$ that does not appear as either of the two cycles $(1, 5, 8, 3)$ and $(2, 7, 6)$ and this is 4. Applying ρ to 4 we find:

$$4 \mapsto 4.$$

So

$$\rho = (1, 5, 8, 3)(2, 7, 6)(4)$$

as a product of disjoint cycles. Recall that (4) is just the identity, so it is usual to omit it and write,

$$\rho = (1, 5, 8, 3)(2, 7, 6).$$

You might be wondering why we wrote ρ as above and not $\rho = (2, 7, 6)(1, 5, 8, 3)$. This does not matter since **disjoint cycles commute**; more on this below. $\lozenge$

11.4.4 Example Let

$$\sigma = (1,3,10,9)(2,5,6), \qquad \tau = (4,3,10)(1,5,8).$$

Express $\sigma\tau$ and σ^{-1} as a product of disjoint cycles.

Answer: We start with 1 and follow the same procedure as the above example. Note that $\sigma\tau 1$ means apply τ first to 1 and then apply σ to the result. Now $\tau 1 = 5$ and $\sigma 5 = 6$. So $\sigma\tau 1 = 6$. Next we apply $\sigma\tau$ to 6. The permutation τ does not have 6 in its cycle decomposition, so $\tau 6 = 6$. So $\sigma\tau 6 = \sigma 6 = 2$. We keep applying $\sigma\tau$ until we return to 1:

$$1 \mapsto 6 \mapsto 2 \mapsto 5 \mapsto 8 \mapsto 3 \mapsto 9 \mapsto 1.$$

Thus $\sigma\tau$ has the cycle $(1,6,2,5,8,3,9)$ in its decomposition as a product of disjoint cycles. We note that this cycle has no 4 in it. So we apply $\sigma\tau$ repeatedly starting with 4:

$$4 \mapsto 10 \mapsto 4.$$

Hence $\sigma\tau$ has the product $(1,6,2,5,8,3,9)(4,10)$ in its decomposition as a product of disjoint cycles. Finally, note that of the elements of the set $\{1,2,\dots,10\}$, the only one not appearing in the product $(1,6,2,5,8,3,9)(4,10)$ is 7. However $\sigma\tau 7 = 7$. So

$$\sigma\tau = (1,6,2,5,8,3,9)(4,10)$$

as a product of disjoint cycles.

You may have noticed that we were tacitly assuming that σ and τ are elements of S_{10} and computed the product under that assumption. In fact, we would have obtained the same result had σ and τ been elements of $S_{11}, S_{12}, \dots$. Indeed viewed as elements of S_{11}, the permutations σ and τ, and the cycles $(1,6,2,5,8,3,9)$ and $(4,10)$ all fix 11.

To compute σ^{-1} we start with $\sigma = (1,3,10,9)(2,5,6)$ and reverse the arrows: Therefore $\sigma^{-1} = (1,9,10,3)(2,6,5)$. Check for yourself that $\sigma\sigma^{-1}$ is indeed the identity permutation. $\diamondsuit$

11.4.5 Exercise Let ρ and τ be as given in Exercise 11.3.5. Write ρ and τ as products of disjoint cycles.

11.4.6 Exercise Which of the following pairs of permutations are equal elements of S_6?

(i) $(1,2,3)(4,6)$ and $(6,4)(2,3,1)(5)$.

(ii) $(4,5,6)(1,2,3)$ and $(3,1,2)(5,4,6)$.

11.4.7 Exercise Let $\rho = (1,2,3)(4,5)$ and $\tau = (1,2,3,4)$. Write the following in cycle notation (i.e. as a product of disjoint cycles): ρ^{-1}, τ^{-1}, $\rho\tau$, $\tau\rho^2$.

11.4.8 Lemma Disjoint cycles commute.

Proof. Let σ and τ be disjoint cycle in S_n and write

$$\sigma = (a_1, a_2, \dots, a_k), \qquad \tau = (b_1, b_2, \dots, b_\ell).$$

Since σ and τ are disjoint $a_i \neq b_j$ for $i = 1, \dots, k$ and $j = 1, \dots, \ell$.

We want to show that $\sigma\tau = \tau\sigma$. This means that $\sigma\tau x = \tau\sigma x$ for all $x \in \{1, 2, \dots, n\}$. We subdivide into three cases:

Case 1: x does not equal any of the a_i or b_j. Then $\tau x = x$ and $\sigma x = x$. Therefore

$$\sigma\tau x = \sigma x = x = \tau x = \tau\sigma x.$$

Case 2: $x = a_i$ for some $i = 1, \dots, k$. Thus x does not equal any of the b_j, and so $\tau x = x$. Hence $\sigma\tau x = \sigma x = \sigma a_i = a_{i+1}$; here a_{k+1} is interpreted as being a_1. Let's compute $\tau\sigma x$. This is $\tau\sigma a_i = \tau a_{i+1} = a_{i+1}$ since a_{i+1} does not equal any of the b_j. Hence $\sigma\tau x = \tau\sigma x$.

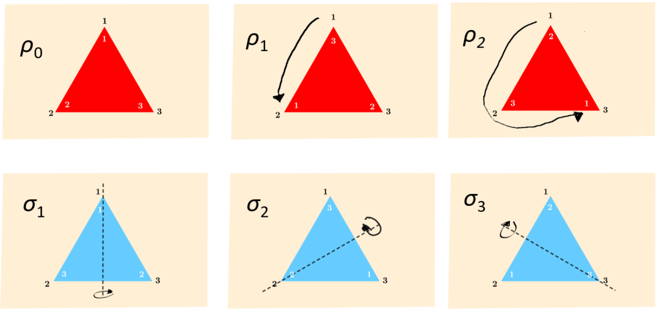
Case 3: $x = b_j$ for some $j = 1, \dots, \ell$. This is similar to Case 2.

We conclude that $\sigma\tau = \tau\sigma$ as required.

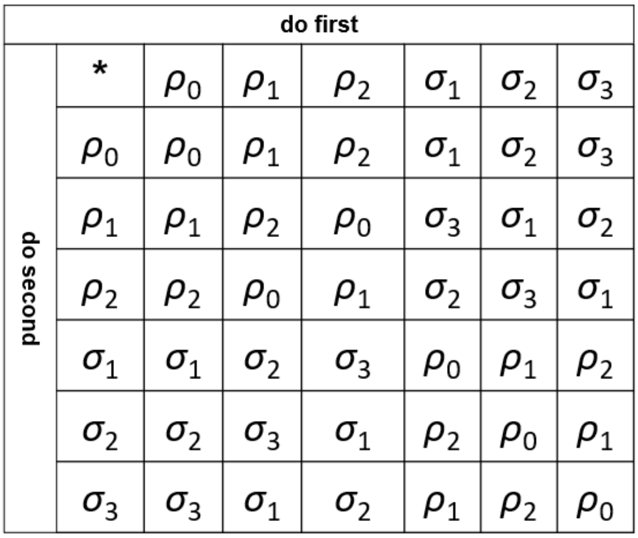
11.5 D_3 and S_3

Cast your mind back to the prologue when we met the six symmetries of an equilateral triangle.

We gave each of the symmetries a name:



And this is the the composition/table using the binary operation ‘followed by’(essentially composition of symmetries):



	do first						
do second	*	ρ_0	ρ_1	ρ_2	σ_1	σ_2	σ_3
	ρ_0	ρ_0	ρ_1	ρ_2	σ_1	σ_2	σ_3
	ρ_1	ρ_1	ρ_2	ρ_0	σ_3	σ_1	σ_2
	ρ_2	ρ_2	ρ_0	ρ_1	σ_2	σ_3	σ_1
	σ_1	σ_1	σ_2	σ_3	ρ_0	ρ_1	ρ_2
	σ_2	σ_2	σ_3	σ_1	ρ_2	ρ_0	ρ_1
	σ_3	σ_3	σ_1	σ_2	ρ_1	ρ_2	ρ_0

This is the group D_3.

Remember that we labelled the vertices of the equilateral triangle with the numbers 1, 2, 3. Notice how there is a natural correspondence between the elements of D_3 and elements of S_3. For example σ_2 swaps vertices 1 and 3 and so it corresponds to $(1,3) \in S_3$.

The elements of S_3 which correspond to

$$\rho_0, \rho_1, \rho_2, \rho_3, \sigma_1, \sigma_2, \sigma_3 \in D_3$$

written in disjoint cycle notation are, respectively:

$$(1); (1,2,3); (1,3,2); (2,3); (1,3); (1,2).$$

Notice that the list above is the whole of S_3. Here is the composition/multiplication table for S_3.

$\circ$	(1)	$(1,2,3)$	$(1,3,2)$	$(2,3)$	$(1,3)$	$(1,2)$
(1)	(1)	$(1,2,3)$	$(1,3,2)$	$(2,3)$	$(1,3$	$(1,2)$
$(1,2,3)$	$(1,2,3$	$(1,3,2)$	(1)	$(1,2)$	$(2,3)$	$(1,3)$
$(1,3,2)$	$(1,3,2)$	(1)	$(1,2,3)$	$(1,3)$	$(1,2)$	$(2,3)$
$(2,3)$	$(2,3)$	$(1,3)$	$(1,2))$	(1)	$(1,2,3)$	$(1,3,2)$
$(1,3)$	$(1,3)$	$(1,2)$	$(2,3)$	$(1,3,2)$	(1)	$(1,2,3)$
$(1,2)$	$(1,2)$	$(2,3)$	$(1,3)$	$(1,2,3)$	$(1,3,2)$	(1)

This is identical to the D_3 table if you carefully swap all the entries according to the correspondence above! From an asbtract algebra point of view, when we are really only interested in the elements and how they combine, is there really any distinction between D_3 and S_3? The next lecture contains more on this.

11.6 D_4 and S_4

Now cast your mind back to Lecture 5 when we met D_4 and labelled the vertices of the square with 1, 2, 3, 4 as in Figure 8. Go back there and remind yourself of the notation for the elements of D_4.

In the notation established there, the elements of D_4 are

$$\rho_0, \rho_1, \rho_2, \rho_3, \sigma_0, \sigma_1, \sigma_2, \sigma_3.$$

The elements of S_4 which correspond to these by considering where the four vertices end up, written in disjoint cycle notation are, respectively:

$$(1); (1,2,3,4); (1,3)(2,4); (1,4,3,2); (2,4); (12)(34); (13); (14)(23).$$

Notice that this is not the whole of S_4 because S_4 has 24 elements. Why is it that the permutation $(1,2,3) \in S_4$, for example, does not correspond to a symmetry of the square?

Here is the composition/multiplication table for D_4

$\circ$	ρ_0	ρ_1	ρ_2	ρ_3	σ_0	σ_1	σ_2	σ_3
ρ_0	ρ_0	ρ_1	ρ_2	ρ_3	σ_0	σ_1	σ_2	σ_3
ρ_1	ρ_1	ρ_2	ρ_3	ρ_0	σ_1	σ_2	σ_3	σ_0
ρ_2	ρ_2	ρ_3	ρ_0	ρ_1	σ_2	σ_3	σ_0	σ_1
ρ_3	ρ_3	ρ_0	ρ_1	ρ_2	σ_3	σ_0	σ_1	σ_2
σ_0	σ_0	σ_3	σ_2	σ_1	ρ_0	ρ_3	ρ_2	ρ_1
σ_1	σ_1	σ_0	σ_3	σ_2	ρ_1	ρ_0	ρ_3	ρ_2
σ_2	σ_2	σ_1	σ_0	σ_3	ρ_2	ρ_1	ρ_0	ρ_3
σ_3	σ_3	σ_2	σ_1	σ_0	ρ_3	ρ_2	ρ_1	ρ_0

Here is the composition/multiplication table for the corresponding 8 elements of S_4.

$\circ$	(1)	$(1,2,3,4)$	$(1,3)(2,4)$	$(1,4,3,2)$	$(2,4)$	$(1,2)(3,4)$	$(1,3)$	$(1,4)(2,3)$
(1)	(1)	$(1,2,3,4)$	$(1,3)(2,4)$	$(1,4,3,2)$	$(1,3)$	$(1,2)(3,4)$	$(1,3)$	$(1,4)(2,3)$
$(1,2,3,4)$	$(1,2,3,4)$	$(1,3)(2,4)$	$(1,4,3,2)$	(1)	$(1,2)(3,4)$	$(1,3)$	$(1,4)(2,3)$	$(1,3)$
$(1,3)(2,4)$	$(1,3)(2,4)$	$(1,4,3,2)$	(1)	$(1,2,3,4)$	$(1,3)$	$(1,4)(2,3)$	$(2,4)$	$(1,2)(3,4)$
$(1,4,3,2)$	$(1,4,3,2)$	(1)	$(1,2,3,4)$	$(1,3)(2,4)$	$(1,4)(2,3)$	$(2,4)$	$(1,2)(3,4)$	$(1,3)$
$(2,4)$	$(2,4)$	$(1,4)(2,3)$	$(1,3)$	$(1,2)(3,4)$	(1)	$(1,4,3,2)$	$(1,3)(2,4)$	$(1,2,3,4)$
$(1,2)(3,4)$	$(1,2)(3,4)$	$(2,4)$	$(1,4)(2,3)$	$(1,3)$	$(1,2,3,4)$	(1)	$(1,4,3,2)$	$(1,3)(2,4)$
$(1,3)$	$(1,3)$	$(1,2)(3,4)$	$(2,4)$	$(1,4)(2,3)$	$(1,3)(2,4)$	$(1,2,3,4)$	(1)	$(1,4,3,2)$
$(1,4)(2,3)$	$(1,4)(2,3)$	$(1,3)$	$(1,2)(3,4)$	$(2,4)$	$(1,4,3,2)$	$(1,3)(2,4)$	$(1,2,3,4)$	(1)

Notice that the above table shows that

$$H = \{(1); (1,2,3,4); (1,3)(2,4); (1,4,3,2); (2,4); (1,2)(3,4); (1,3); (1,4)(2,3)\}$$

is a subgroup of S_4. Again, is this subgroup really any different to D_4 itself?

This means there is a subgroup which is a essentially a 'copy 'of D_4 inside S_4. You will eventually be comfortable enough to say that there is no distinction between D_4 and this copy of D_4 and say that D_4 is a subgroup of S_4. Again, the next lecture contains more on this.

11.6.1 Exercise Think about the group D_5, the symmetries of a regular pentagon. How many elements does it have? How many elements does S_5 have? How is there a subgroups of S_5 which is a 'copy 'of D_5? Think about how this generalises to any integer $n > 2$.

Lecture 12 - Isomorphisms

At the end of the last chapter we saw two examples of two groups (first D_3 and S_3, then D_4 and a particular subgroup of S_4) that have 'the same' multiplication tables' but which have arisen in two diffent ways. In one of the earlier lectures we saw that D_4 has the same multiplicaton table as a subgroup consisting of 8 matrices in $\mathrm{GL}_2(\mathbb{R})$.

Casting your mind back to chapter 10 you probably noticed there's a lot in common between the group of m-th roots of unity U_m, and the group $\mathbb{Z}/m\mathbb{Z}$. If not, take another look Tables 2 and 3.

We'll see here that these are all examples of *isomorphic* groups. What does this mean?

12.1 What is an isomorphism?

12.1.1 Definition Let $(G, \circ)$ and $(H, *)$ be groups. We say that the function $\phi : G \to H$ is an *isomorphism* if it is a bijection and it satisfies

$$\phi(g_1 \circ g_2) = \phi(g_1) * \phi(g_2)$$

for all g_1, g_2 in G. In this case we say that $(G, \circ)$ and $(H, *)$ are *isomorphic.*

Isomorphic groups may look different, but in essence they are the same. An isomorphism is a way of relabeling the elements of one group to obtain another group, as the following examples will make clear.

12.1.2 Example Define $\phi : \mathbb{Z}/m\mathbb{Z} \to U_m$ by the rule

$$\phi(\overline{a}) = \zeta^a, \qquad a = 0, 1, \ldots, m-1.$$

Then ϕ is a bijection and satisfies the property

$$\phi(\overline{a} + \overline{b}) = \phi(\overline{a+b}) = \zeta^{a+b} = \zeta^a \cdot \zeta^b = \phi(\overline{a})\phi(\overline{b}).$$

So ϕ is an isomorphism. $\Diamond$

12.1.3 Example Recall that the matrix

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

represents anticlockwise rotation around the origin through an angle θ. The identity

$$R_{\theta+\phi} = R_\theta R_\phi.$$

turns addition into multiplication, and so it should remind you of the identity $e^{\theta+\phi} = e^\theta e^\phi$. In fact, a more accurate analogy is identity

$$e^{i(\theta+\phi)} = e^{i\theta} e^{i\phi}.$$

The reason is because multiplying a complex number by $e^{i\theta}$ rotates it about the origin anticlockwise through the angle θ (prove this using the exponential form for complex numbers).

Now that you know that R_θ and $e^{i\theta}$ are analogues, you will expect that the groups $\mathrm{SO}_2(\mathbb{R})$ and $\mathbb{S}$ are isomorphic. Recall that $\mathrm{SO}_2(\mathbb{R})$ is the special orthogonal group (Theorem 9.4.3) defined by

$$\mathrm{SO}_2(\mathbb{R}) = \{R_\theta : \theta \in \mathbb{R}\},$$

and $\mathbb{S}$ is the circle group (page 91) given by

$$\mathbb{S} = \{\alpha \in \mathbb{C} : |\alpha| = 1\} = \{e^{i\theta} : \theta \in \mathbb{R}\}.$$

You can satisfy yourself that the map

$$\phi : \mathrm{SO}_2(\mathbb{R}) \to \mathbb{S}, \qquad \phi(R_\theta) = e^{i\theta}$$

is an isomorphism.

By this correspondence, we see that there are matrix analogues of the n-th roots of unity are. If we let

$$\mathcal{Z} = R_{2\pi/n} = \begin{pmatrix} \cos(2\pi/n) & -\sin(2\pi/n) \\ \sin(2\pi/n) & \cos(2\pi/n) \end{pmatrix},$$

then $I_2, \mathcal{Z}, \dots, \mathcal{Z}^{n-1}$ all satisfy the relationship $A^n = I_2$. ♢

12.1.4 Exercise Let $\mathcal{Z} = R_{2\pi/6}$. Show that $\{1, \mathcal{Z}, \dots, \mathcal{Z}^5\}$ is a subgroup of $\mathrm{SO}_2(\mathbb{R})$. Write down the orders of its elements and check that they are consistent with Lagrange's Theorem.

12.1.5 Exercise Suppose groups G and H are isomorphic. Show that G and H have the same order. Show that G is abelian if and only if H is abelian. Show that G is cylic if and only if H is cyclic.

12.2 What's special About S_n?

We started lecture 10 by looking at symmetry groups of arbitrary sets A. Then we restricted ourself to $S_n = \mathrm{Sym}(\{1, 2, \ldots, n\})$. This is not as big a restriction as it looks. Suppose the set A is finite, and let $n = |A|$, the number of elements of A. Then $\mathrm{Sym}(A)$ is isomorphic to S_n. One way of seeing this is convince ourselves that every permutation of $\{1, 2, \ldots, n\}$ gives us a permutation of A. For example, suppose $A = \{a_1, a_2, a_3\}$. Then the permutation $\{1, 2, 3\}$ given by

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$$

corresponds to the permutation of A given by

$$\begin{pmatrix} a_1 & a_2 & a_3 \\ a_2 & a_3 & a_1 \end{pmatrix}.$$

Understanding $\mathrm{Sym}(A)$ with $|A| = n$ is the same as understanding S_n.

That's it on isomorphisms for this module but this is something that you'll hear much more about in future modules covering abstract algebra (and you'll meet *homomorphisms* too).

Lecture 13 - Cosets

Cosets are what we get when we 'shift' a subgroup by the elements of the group. Sometimes there is a clear geometrical interpretation to this, sometimes there is not.

There are many excting things about cosets! One is that they will enable us to prove Lagrange's theorem. Another is that, sometimes, we can make a new group from an existing group where the elements are cosets. Understanding this (potentially simpler) new group can then help us to understand the original group. But let's not get ahead of ourselves, here's the definition of a coset.

13.1 What is a coset?

13.1.1 Definition Let G be a group and H a subgroup. Let g be an element of G. We call the set

$$gH = \{gh \mid h \in H\}$$

a *left coset of* H in G and the set

$$Hg = \{hg \mid h \in H\}$$

a *right coset of* H in G.

13.1.2 Example Let's take G to be D_4 and R the subgroup made up of rotations:

$$R = \{1, \rho_1, \rho_2, \rho_3\}.$$

Revisit Section 5.4 to remind yourself of the notation. Let's compute $\sigma_1 R$. By definition,

$$\begin{aligned}\sigma_1 R &= \{\sigma_1 \cdot 1, \sigma_1\rho_1, \sigma_1\rho_2, \sigma_1\rho_3\}\\ &= \{\sigma_1, \sigma_0, \sigma_3, \sigma_2\}\\ &= \{\sigma_0, \sigma_1, \sigma_2, \sigma_3\}.\end{aligned}$$

Let's try another coset.

$$\begin{aligned}\rho_2 R &= \{\rho_2 \cdot 1, \rho_2\rho_1, \rho_2\rho_2, \rho_2\rho_3\}\\ &= \{\rho_2, \rho_3, 1, \rho_1\}\\ &= \{1, \rho_1, \rho_2, \rho_3\}.\end{aligned}$$

We see that $\rho_2 R$ is equal to R, and $\sigma_1 R$ isn't equal to R. In fact, $\sigma_1 R$ isn't even a subgroup of D_4; why? You can carry on computing all eight left cosets, and you'll find

$$1 \cdot R = \rho_1 R = \rho_2 R = \rho_3 R = \{1, \rho_1, \rho_2, \rho_3\}$$

and

$$\sigma_0 R = \sigma_1 R = \sigma_2 R = \sigma_3 R = \{\sigma_0, \sigma_1, \sigma_2, \sigma_3\}.$$

◊

13.1.3 Exercise Recall that $H = \{1, \sigma_2\}$ is also a subgroup of D_4. Compute its left cosets. Check that $\sigma_1 H \neq H\sigma_1$.

For an additive group G, a subgroup H, a left coset would be of the form

$$g + H = \{g + h : h \in H\}$$

for some g in G.

13.1.4 Example $\mathbb{Z}$ is an additive group. The set of even integers $2\mathbb{Z}$ is a subgroup. What are its cosets? Let's compute a few:

$$\begin{aligned}
0 + 2\mathbb{Z} &= \{\dots, 0 + (-4), 0 + (-2), 0 + 0, 0 + 2, 0 + 4, \dots\} = \{\dots, -4, -2, 0, 2, 4, \dots\};\\
1 + 2\mathbb{Z} &= \{\dots, 1 + (-4), 1 + (-2), 1 + 0, 1 + 2, 1 + 4, \dots\} = \{\dots, -3, -1, 1, 3, 5, \dots\};\\
2 + 2\mathbb{Z} &= \{\dots, 2 + (-4), 2 + (-2), 2 + 0, 2 + 2, 2 + 4, \dots\} = \{\dots, -4, -2, 0, 2, 4, \dots\};\\
3 + 2\mathbb{Z} &= \{\dots, 3 + (-4), 3 + (-2), 3 + 0, 3 + 2, 3 + 4, \dots\} = \{\dots, -3, -1, 1, 3, 5, \dots\}.
\end{aligned}$$

You'll quickly discover that

$$\dots = -4 + 2\mathbb{Z} = -2 + 2\mathbb{Z} = 2\mathbb{Z} = 2 + 2\mathbb{Z} = 4 + 2\mathbb{Z} = \dots$$

and

$$\dots = -3 + 2\mathbb{Z} = -1 + 2\mathbb{Z} = 1 + 2\mathbb{Z} = 3 + 2\mathbb{Z} = \dots.$$

So $2\mathbb{Z}$ has two cosets in $\mathbb{Z}$, which happen to be $2\mathbb{Z}$ itself, and $1 + 2\mathbb{Z}$ which is the set of odd integers. ◊

13.1.5 Exercise You know that $\mathbb{Z}^2$ is a group. Let

$$2\mathbb{Z}^2 = \{(2a, 2b) : a, b \in \mathbb{Z}\}.$$

In otherwords, $2\mathbb{Z}^2$ is the set of vectors in $\mathbb{Z}^2$ with both coordinates even. Check that $2\mathbb{Z}^2$ is a subgroup of $\mathbb{Z}^2$, having four cosets. What are they?

13.1.6 Exercise Let $\mathbb{R}^+$ be the subset of $\mathbb{R}^*$ consisting of the positive numbers. Show that $\mathbb{R}^+$ is a subgroup and that it has exactly two cosets in $\mathbb{R}^*$.

13.2 Geometric Examples

As you'll be beginning to appreciate, some of group theory, particularly the proofs, is about manipulating symbols (i.e doing algebra) but very often groups can be interpreted geometrically (i.e via pictures) and this geometrical meaning is the way to build mathematical insight into the underlying concepts. Here we'll focus on the geometrical interpretation of cosets.

13.2.1 Example You'll recall the circle group $\mathbb{S}$ which is the subgroup of $\mathbb{C}^*$ consisting of all elements of absolute value 1; see Example 9.2.13 if you need to refresh your memory.

Let's study the cosets of $\mathbb{S}$ in $\mathbb{C}^*$. Of course $\mathbb{C}^*$ is abelian, and so there is no distinction between left and right cosets; they're the same. A coset of $\mathbb{S}$ in $\mathbb{C}^*$ has the form $\alpha\mathbb{S}$ where α is in $\mathbb{C}^*$ (i.e. α is a non-zero complex number). As such, we can write $\alpha = re^{i\theta}$, where r is positive (it is the absolute value of α), and θ is the argument of α. Consider $e^{i\theta}\mathbb{S}$. Multiplying any complex number by $e^{i\theta}$ simply rotates anticlockwise through angle θ about the origin. So $e^{i\theta}\mathbb{S} = \mathbb{S}$. Now $\alpha\mathbb{S} = r\mathbb{S}$.

What does multiplying by r do? It scales the circle $\mathbb{S}$ by a factor of r. Two different positive real numbers $r_1 \neq r_2$ will give different cosets $r_1\mathbb{S} \neq r_2\mathbb{S}$, since the first has radius r_1 and the second has radius r_2. See Figure 20.

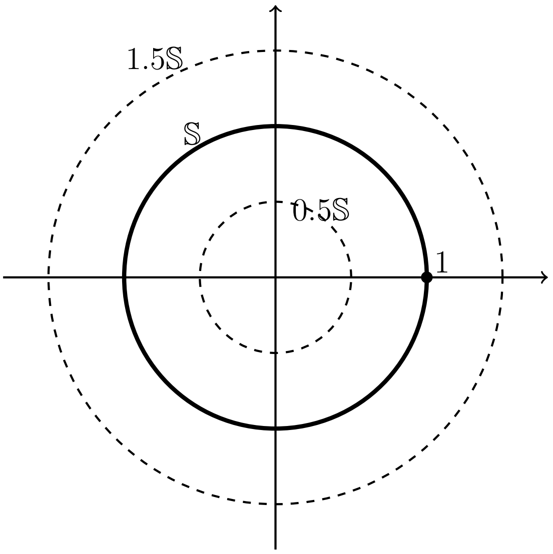


Figure 20: $\mathbb{S}$ and its cosets $0.5\mathbb{S}$ and $1.5\mathbb{S}$ in $\mathbb{C}^*$.

So $\mathbb{S}$ has as many cosets in $\mathbb{C}^*$ as there are positive real numbers.

Summary: $\mathbb{S}$ is the circle centred at the origin of radius 1, and its cosets in $\mathbb{C}^*$ are the circles centred at the origin (of positive radius). ◊

13.2.2 Example In Exercise 9.2.12 you were asked the following question: which lines in $\mathbb{R}^2$ define a subgroup? Let's go back to this question and answer it again, and this time for lines that define a subgroup we want to determine the cosets too.

One convenient way of a specifying a line L in $\mathbb{R}^2$ is as follows. Let Q be a point on L, with position vector $\mathbf{w}$. Let $\mathbf{v}$ be a vector parallel to L. Then L has the parametric form

$$L: \ \mathbf{x} = \mathbf{w} + t\mathbf{v}.$$

This is a slightly clumsy school/college way of saying things. What it means is that the points with position vector $\mathbf{w} + t\mathbf{v}$ are on the line, where t is any 'scalar' (i.e. real number). A much better way is to just write L in set notation:

$$L = \{\mathbf{w} + t\mathbf{v} : t \in \mathbb{R}\}.$$

Now L is a subset of $\mathbb{R}^2$ and we want to know if it defines a subgroup. Of course, if L does not pass through the origin, then it does not contain the identity element, and so cannot be a subgroup. So, let's suppose L passes through the origin. The point Q was any point on the line; we will choose Q to be the origin, and so its position vector is $\mathbf{w} = (0, 0)$. Now we have

$$L = \{t\mathbf{v} : t \in \mathbb{R}\}.$$

Is L a subgroup of $\mathbb{R}^2$? It is straightforward to 'see' geometrically that if we add two vectors in L then the sum is in L. Let's check that algebraically. If $\mathbf{v}_1$ and $\mathbf{v}_2$ are in L then they have the form $\mathbf{v}_1 = t_1\mathbf{v}$ and $\mathbf{v}_2 = t_2\mathbf{v}$. So

$$\mathbf{v}_1 + \mathbf{v}_2 = (t_1 + t_2)\mathbf{v}$$

which in L. Also, $-\mathbf{v}_1 = (-t_1)\mathbf{v}_1$ is in L. Hence L is a subgroup of $\mathbb{R}^2$.

What are the cosets of L in $\mathbb{R}^2$? They have the form

$$\mathbf{w} + L = \{\mathbf{w} + t\mathbf{v} : t \in \mathbb{R}\}$$

where $\mathbf{w}$ is a vector in $\mathbb{R}^2$. This is the line with parametric form $\mathbf{w} + t\mathbf{v}$. Note that both L and its coset $\mathbf{w} + L$ are parallel to $\mathbf{v}$. See Figure 21

Conclusion: *A line in $\mathbb{R}^2$ is a subgroup if and only if it passes through the origin. If it does, then its cosets are the lines parallel to it.* ◊

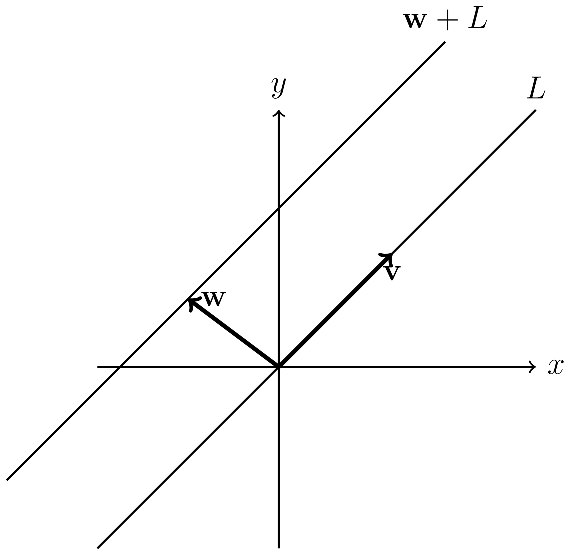


Figure 21: A line L defines a subgroup of $\mathbb{R}^2$ if and only if it passes through the origin. In that case, its cosets are the lines parallel to it.

13.3 Solving Equations

Cosets arise naturally when solving certain types of equations. It's difficult to make this precise at present. Instead let's look at some examples to understand what this means.

13.3.1 Example If you did matrices at school/college, then you will probably know that a system of m linear equations in n variables can be written as a single matrix equation

$$A\mathbf{x} = \mathbf{b} \tag{14}$$

where A is an $m \times n$ matrix, $\mathbf{b}$ is a vector in $\mathbb{R}^m$ and $\mathbf{x}$ is an unknown vector in $\mathbb{R}^n$.

Let

$$K = \{\mathbf{x} \in \mathbb{R}^n \ : \ A\mathbf{x} = \mathbf{0}\}.$$

That is, K is the set of solutions $\mathbf{x}$ of the equation $A\mathbf{x} = \mathbf{0}$. It is easy to show that K is a subgroup of $\mathbb{R}^n$ (exercise!).

We call K the *kernel* of A. What is the relation between K and the solutions of (14)? If (14) has no solutions then there is no relation. So let's suppose it has some solutions, and let's take $\mathbf{x}_0$ to be one of them. Let $\mathbf{x}$ be any other solution.

Then

$$A\mathbf{x} = \mathbf{b}, \qquad A\mathbf{x}_0 = \mathbf{b}.$$

Subtracting we find

$$A(\mathbf{x} - \mathbf{x}_0) = \mathbf{0}.$$

So the difference $\mathbf{x} - \mathbf{x}_0$ belongs to the subgroup K. Thus $\mathbf{x}$ belongs to the coset $\mathbf{x}_0 + K$. In fact, the set of solutions to (14) is precisely the coset $\mathbf{x}_0 + K$. ◊

13.3.2 Example In the *Differential Equations* module, one of things you'll look at are linear second order differential equations. For example, you'll see equations of the form

$$a\frac{d^2x}{dt^2} + b\frac{dx}{dt} + cx = f(t), \tag{15}$$

with a, b, c constants (again, it is likely that you've seen these at school). To solve this you look at *the corresponding homogeneous equation*

$$a\frac{d^2x}{dt^2} + b\frac{dx}{dt} + cx = 0. \tag{16}$$

Convince yourself that the solutions to the homogeneous equation (16) form a group K with respect to addition (revise Section 9.5 if you need to). In some textbooks on differential equations (and some old A-Level maths textbooks), K is called the kernel. Now we ask the same question as in the previous example: what is the relation between the solutions to (15) and K?

Again, if (15) does not have a solution then there is no relation. Suppose it has solutions, and let $x_0(t)$ be one of them. In your Differential Equations module, $x_0(t)$ is called 'a particular integral'. If $x(t)$ is any other solution to (15), then you can check that $x(t)-x_0(t)$ is a solution to the homogeneous equation (16), and so is an element of K. It follows that the set of solutions to (15) is the coset $x_0(t)+K$. ◊

Are the similarities between the above two examples a coincidence? No, they are instances of a recurrent theme in mathematics. This theme is formalized in the **First Isomorphism Theorem**, which you'll meet in *Algebra II*.

So it turns out that you've been using the first isomorphism theorem for a while when solving linear equations (and linear differential equations). After you meet the First Isomorphism Theorem, come back and review these two examples again.

13.4 Index

13.4.1 Definition Let G be a group and H be a subgroup. We shall define the *index* of H in G, denoted by $[G : H]$, to be the number of left cosets of H in G.

13.4.2 Example In Example 13.1.2, we computed the left cosets of $R = \{1, \rho_1, \rho_2, \rho_3\}$ in D_4 and found exactly two of them: namely

$$\{1, \rho_1, \rho_2, \rho_3\} \qquad \text{and} \qquad \{\sigma_0, \sigma_1, \sigma_2, \sigma_3\}.$$

So the index $[D_4 : R] = 2$. ◊

13.4.3 Example In Example 13.1.4 we found that the cosets of $2\mathbb{Z}$ in $\mathbb{Z}$ are $2\mathbb{Z}$ itself, and $1 + 2\mathbb{Z}$, so the index $[\mathbb{Z} : 2\mathbb{Z}] = 2$. If you've done Exercise 13.1.5 then you'll know that $[\mathbb{Z}^2 : 2\mathbb{Z}^2] = 4$. ◊

13.4.4 Example In Example 13.2.1, we found that the cosets of the circle group $\mathbb{S}$ in $\mathbb{C}^*$ are the circles centred at the origin. So the index $[\mathbb{C}^* : \mathbb{S}] = \infty$.

13.4.5 Example Now let's look at the index of the trivial group $\{0\}$ as a subgroup of $\mathbb{Z}$. Note that

$$a + \{0\} = \{a\}.$$

So the cosets of $\{0\}$ in $\mathbb{Z}$ are

$$\ldots, \{-2\}, \ \{-1\}, \ \{0\}, \ \{1\}, \ \{2\}, \ldots$$

Clearly $[\mathbb{Z} : \{0\}] = \infty$. ◊

13.4.6 Exercise Let G be a finite group. Let $\{1\}$ be the trivial subgroup consisting only of the identity element. Explain why $[G : \{1\}] = |G|$.

13.5 The First Innermost Secret of Cosets

Apart from the definition, you need to know two facts about cosets. The first is that a coset of a subgroup has the same size as the subgroup.

13.5.1 Lemma Let G be a group and H a finite subgroup. If $g \in G$ then gH and Hg have the same number of elements as H.

Proof. We'll just prove the lemma for left cosets. The proof for right cosets is nearly the same. Let g be an element of G. We want to show that H and gH has the same number of elements. The sets H and gH are finite. The best way to show that two finite sets have the same number of elements is to set up a bijection between them (thus demonstrating that a bijection exists between them).
Let

$$\phi : H \to gH, \qquad h \mapsto gh.$$

From the definition of gH it is clear that $\phi(h)$ is in the coset gH whenever h is in the subgroup H. So the map ϕ makes sense. To check that it is a bijection we need to show that it is injective and surjective.

Injectiveness: Suppose two elements h_1, h_2 map to the same element in gH. In otherwords, $\phi(h_1) = \phi(h_2)$. We want to show that $h_1 = h_2$. But $\phi(h_1) = \phi(h_2)$ means

$$gh_1 = gh_2.$$

We want to 'cancel' the gs from both sides. Remember we can do this in a group by multiplying by both sides on the left by g^{-1}, to obtain

$$g^{-1}(gh_1) = g^{-1}(gh_2).$$

Thus $h_1 = h_2$.

Surjectiveness: Suppose k is an element of the coset gH. We want to show that k is of the form $\phi(h)$ for some element h of the subgroup H. But by definition, $gH = \{gh \mid h \in H\}$, so $k = gh = \phi(h)$ for some h in H.

Remark. Note that the proof that $\phi : H \to gH$ is a bijection did not assume the finiteness of H; it is true for any subgroup H whether finite or infinite. The finiteness is used to conclude that the number of elements of H and the number of elements of gH are the same.

What happens if H is infinite? Mathematicians still think of H and gH as having the same number of elements, even though they are infinite, simply because there is a bijection between them. Thus $|2\mathbb{Z}| = |1 + 2\mathbb{Z}|$, and $|\mathbb{S}| = |2\mathbb{S}|$. However, $|2\mathbb{Z}| \neq |\mathbb{S}|$, because $2\mathbb{Z}$ is countable and $\mathbb{S}$ is uncountable. See *cardinalities* in *Foundations/Sets and Numbers.*

13.5.2 Example Now is a good time to revisit the examples at the beginning of the chapter and make sure that Lemma 13.5.1 holds for them.

13.6 The Second Innermost Secret of Cosets

In this section we'll find out that, given a subgroup H of a group G any two left/right cosets of H are equal or disjoint. This means that any two left/right cosets are either exactly the same set or they have no elements in common at all. You've seen this in the examples of cosets you've met so far.

13.6.1 Example Look again at Example 13.2.1 and in particular Figure 20. There we looked the cosets of the circle subgroup $\mathbb{S}$ inside $\mathbb{C}^*$. We found that the cosets are the circles centred at the origin of positive radius. It is obvious that two such circles are either equal or disjoint. $\Diamond$

13.6.2 Example In Example 13.2.2, we saw that a line L in $\mathbb{R}^2$ passing through the origin defines a subgroup. The cosets of L are the lines parallel to it. Again it is clear that two lines parallel to L are either equal or disjoint. $\Diamond$

Here comes the main result in this section.

13.6.3 Lemma Let G be a group and H be a subgroup. Let g_1, $g_2 \in G$ so that g_1H and g_2H are left cosets.

Then

i. $g_1H = g_2H$ if and only if $g_2^{-1}g_1 \in H$

ii. $g_1H \cap g_2H = \emptyset$ if and only $g_2^{-1}g_1 \notin H$

This means that any two left cosets g_1H, g_2H are either equal or disjoint. In other words either $g_1H = g_2H$ or $g_1H \cap g_2H = \emptyset$.

Proof.

i. Note that this is and if and only if statement and therefore requires proof in both directions.

   Starting with 'left implies right', suppose $g_1H = g_2H$. Then, since $g_1 = g_1 \cdot 1$ and $1 \in H$, $g_1 \in g_1H = g_2H = \{g_2h \mid h \in H\}$. Therefore there exist $h \in H$ such that $g_1 = g_2h$. But then $g_2^{-1}g_1 = g_2^{-1}g_2h = 1 \cdot h = h$. This means that $g_2^{-1}g_1 \in H$.

   Now for 'right implies left'. Now suppose that $g_2^{-1}g_1 = h^* \in H$. Let g_1h where $h \in H$ be an arbitrary element of g_1H. Then $g_1h = g_2h^*h \in g_2H$ since $hh^* \in H$. Therefore $g_1H \subseteq g_2H$.

   Continuing, note that $(h^*)^{-1} = g_1^{-1}g_2 \in H$, by 6.2.2. By multpilying that by g_1 on the left, $g_2 = g_1(h^*)^{-1}$. Let g_2k where $k \in H$ be an arbitrary element of g_2H. Then $g_2k = g_1(h^*)^{-1}k \in g_1H$. Therefore $g_2H \subseteq g_1H$.

i. Now suppose that $g_1H \cap g_2H = \emptyset$. Suppose $g_2^{-1}g_1 = h \in H$. Then $g_1 \cdot 1 = g_2h$ is in both g_1H and g_2H. This is a condradiction so $g_2^{-1}g_1 \notin H$.

Finally suppose $g_2^{-1}g_1 \notin H$. If $g_1H \cap g_2H \neq \emptyset$ then there must be elements $h_1, h_2 \in H$ such that let $g_1h_1 = g_2h_2 \in g_1H \cap g_2H$. Then, multiplying on the left by g_2^{-1} and on the right by h_1^{-1}, $g_2^{-1}g_1 = h_2h_1^{-1} \in H$, which isn't true. Therefore $g_1H \cap g_2H = \emptyset$.

It follows that, given any two cosets g_1H, g_2H either $g_1H = g_2H$ (in the case $g_2^{-1}g_1 \in H$) or $g_1H \cap g_2H = \emptyset$ (in the case $g_2^{-1}g_1 \notin H$).

♢

13.6.4 Remark In a non-abelian group Lemma 13.6.3 looks like this with right cosets:

Let G be a group and H be a subgroup. Let g_1, $g_2 \in G$ so that Hg_1 and Hg_2 are right cosets.

Then

i. $Hg_1 = Hg_2$ if and only if $g_1g_2^{-1} \in H$

ii. $Hg_1 \cap Hg_2 = \emptyset$ if and only $g_1g_2^{-1} \notin H$

This means that any two right cosets Hg_1, Hg_2 are either equal or disjoint. In other words either $Hg_1 = Hg_2$ or $Hg_1 \cap Hg_2 = \emptyset$.

13.6.5 Remark In an additive abelian group Lemma 13.6.3 looks like this:

Let $(G, +)$ be an abelian group and let H be a subgroup. Let g_1, $g_2 \in G$ so that $g_1 + H$ and $g_2 + H$ are cosets (there is no distinction between left and right cosets here because $g + H = \{g + h \mid h \in H\} = \{h + g \mid h \in H\} = H + g$ for any $g \in G$).

Then

i. $g_1 + H = g_2 + H$ if and only if $g_1 - g_2 \in H$

ii. $(g_1 + H) \cap (g_2 + H) = \emptyset$ if and only $g_1 - g_2 \notin H$

This means that any two cosets $g_1 + H$, $g_2 + H$ are either equal or disjoint. In other words either $g_1 + H = g_2 + H$ or $(g_1 + H) \cap (g_2 + H) = \emptyset$.

For additive abelian groups, one way to think about the cosets in G with respect to H is that they take an element of a group, g, and they bring into one coset, $g + H$, all elements that differ from that element by an element of H. The next two examples illustrate this.

13.6.6 Example Let $X = \{(a, 0) \mid a \in \mathbb{R}\}$. Then X is a subgroup of $\mathbb{R}^2$, it is the x-axis.

Given $(a_1, b_1) \in \mathbb{R}^2$, the coset $(a_1, b_1) + X = \{(a_1, b_1) + (a, 0) \mid a \in \mathbb{R}\} = \{(a_1 + a, b_1) \mid a \in \mathbb{R}\}$. Because, by choosing a appropriately, $a_1 + a$ can be made to equal any real number, this is just all the points with y-coordinate b_1.

Furthermore, given (a_1, b_1) and (a_2, b_2) in $\mathbb{R}^2$, what does it mean for the coset $(a_1, b_1) + X$ to be the same as the coset $(a_2, b_2) + X$?

By Remark 13.6.5, $(a_1, b_1) + X = (a_2, b_2) + X$ if and only if $(a_1 - a_2, b_1 - b_2)$ belongs to X. This happens if and only if $b_1 - b_2 = 0$. So the two cosets $(a_1, b_1) + X$ and $(a_2, b_2) + X$ are equal if and only (a_1, b_1) and (a_2, b_2) have have the same y-coordinate.

$\diamondsuit$

13.6.7 Example Let $G = \mathbb{R}[x]$. Let $H = \{f \in \mathbb{R}[x] : f(0) = 0\}$. This is the set of all polynomials with real coefficients and with no constant term (or whose constant term is 0). H is a subgroup of $\mathbb{R}[x]$.

Given $g \in R[x]$ the coset $g + H = \{g + f \mid f(0) = 0\}$. Because the constant term of f is zero this will be all the polynomials which have the same constant term as g.

Now suppose g, $h \in \mathbb{R}[x]$. What does it mean for the cosets $g + H$ and $h + H$ to be equal?

By Remark 13.6.5, $g + H = h + H$ if and only if $g - h \in H$. Suppose [13]

$$g = a_0 + a_1x + \cdots + a_nx^n, \qquad h = b_0 + b_1x + \cdots + b_nx^n,$$

where $a_0, \ldots, a_n$ and $b_0, \ldots, b_n$ are real numbers. Then $g - h \in H$ if and only if $a_0 - b_0 = 0$ if and only if $a_0 = b_0$ (i.e. g and h have the same constant term). Therefore $g + H$ and $h + H$ are the same cosets if and only the constant term of g equals the constant term of h. $\Diamond$

13.7 Lagrange's Theorem

We stated some corollaries of Lagrange's Theorem a very long time ago. Now, finally, we'll prove it!

13.7.1 Theorem (Lagrange's Theorem) Let G be a finite group and H a subgroup. Then

$$|G| = [G : H] \cdot |H|.$$

Proof. Let $g_1H, g_2H, \ldots, g_mH$ be the distinct left cosets of H. As they are distinct, we know by Lemma 13.6.3 that they are disjoint. Suppose now that g is an element of G. Then gH must equal one of the g_iH. But $g \in gH$, since $1 \in H$. Hence the cosets $g_1H, g_2H, \ldots, g_mH$ are not only disjoint, but every element of G belongs to exactly one of them. Hence

$$|G| = |g_1H| + |g_2H| + \cdots + |g_mH|.$$

Now by Lemma 13.5.1,

$$|g_1H| = |g_2H| = \cdots = |g_mH| = |H|.$$

Hence

$$|G| = m \cdot |H|.$$

What is m? It is the number of left cosets of H in G. We defined this to be the index of H in G, so $m = [G : H]$. This completes the proof.

Now let's revisit the corollaries we met before in lectures 8 and 9. Here is Corollary 9.7.1 from chapter 9.

[13] It seems that we're writing f and g both as polynomials of the same degree n; this looks wrong as there is no reason to suppose that g and h have the same degree. But looks can be misleading. Here we're in fact writing g and h as polynomials of degree *at most* n. For example, if $g = 2 + 7x$ and $h = 4 - 3x + 2x^3$ then we can take $n = 3$ and let $a_0 = 2$, $a_1 = 7$, $a_2 = a_3 = 0$, and $b_0 = 4$, $b_1 = -3$, $b_2 = 0$, $b_3 = 2$

9.7.1 Corollary Let G be a finite group, and H a subgroup of G. Then the order of H divides the order of G.

Proof. By Theorem 13.7.1, since G is finite, $|G| = [G : H] \cdot |H|$. The result follows because $[G : H]$ is the number of cosets of H in G is an integer. $\Diamond$

And here is Corollary 8.2.2 from chapter 8.

8.2.2 Corollary Let G be a finite group, and let g be an element of G. The order of g divides the order of G.

Proof. Since G is finite g must have finite order (otherwise the group G would contain the infinite set $\{g^n \mid n \in \mathbb{Z}\}$). By Theorem 10.1.8 the order of the subgroup $\langle g \rangle$ is equal to the order of g. By Corollary 9.7.1 the order of g divides the order of G.$\Diamond$

Lecture 14 - Quotient Groups

Taking the quotient group with respect to a subgroup is one of the most powerful concepts in group theory. Often it provides us with a simpler group through which we can understand the orginal group.

At this point, it's also worth noting you can do something analogous to this in other branches of abstract algebra, such as with rings, which we'll meet later in this module.

In this chapter we'll be working with additive abelian groups only. Here it's possible to do this for any subgroup. In futher modules covering group theory you'll see that it's possible to take quotients in non-abelian groups too but then only for subgroups which have particular properties. These 'particular properties' always hold in abelian groups. But we are getting ahead of ourselves, let's start with the definition of a quotient group in the case of an additive abelian group.

14.1 Quotient groups in additive abelian groups.

14.1.1 Definition Let $(G, +)$ be an additive abelian group and H a subgroup. We define the quotient group $(G/H, +)$ to be the set of cosets

$$G/H = \{a + H \mid a \in G\}$$

with addition being defined by

$$(a + H) + (b + H) = (a + b) + H. \tag{17}$$

◊

Note carefully that the elements in $(G/H, +)$ are themselves cosets and so the addition defined tell us how to add two cosets to get another coset! This is why quotient groups can take a bit of getting used to.

We will now go on to prove that $(G/H, +)$ is a group (in fact it is abelian). There is a more serious point which is that we need to show that the operation (17) is *well-defined.*

What does this mean? We know that cosets can have more than one 'name'. For example in $\mathbb{Z}/5\mathbb{Z}$ we have $1 + 5\mathbb{Z} = 6 + 5\mathbb{Z}$ so this coset goes both by the name '$1 + 5\mathbb{Z}$' and by the name '$6 + 5\mathbb{Z}$'. The definition above uses this name. So we'd better make sure that the definition is independent of this choice of name. Precisely, we need to check that this is true:

If

$$a + H = a' + H \text{ and } b + H = b' + H,$$

then

$$(a + b) + H = (a' + b') + H.$$

The following theorem checks this.

14.1.2 Theorem - the addition in $(G/H, +)$ is well defined. Let $(G, +)$ be an additive abelian group and H a subgroup. Let a, a', b, b' be elements of G such that in G/H we have

$$a + H = a' + H \text{ and } b + H = b' + H,$$

then

$$(a + b) + H = (a' + b') + H.$$

Proof. Since $a+H = a'+H$ and $b+H = b'+H$ in G/H by Remark 13.6.5 $a-a' = h_1$ and $b - b' = h_2$ where h_1, $h_2 \in H$. Thus (and note that the commutativity of the addtion is used in the step below):

$$(a + b) - (a' + b') = (a - a') + (b - b') = h_1 + h_2.$$

As H is a subgroup containing h_1 and h_2, we know that the sum $h_1 + h_2$ belongs to H. Thus, by Remark 13.6.5 again, the cosets $(a+b)+H$ and $(a'+b')+H$ are equal.

We need to check one more thing: that G/H is indeed a group!

14.1.3 Theorem Let $(G, +)$ be an additive abelian group and H a subgroup. Then $(G/H, +)$ is an abelian group.

Proof. We have to check the defining properties for abelian groups.

First of all the addition in $(G/H, +)$ is closed because if $a + H$ and $b + H$ are two cosets then their sum $(a + H) + (b + H) = (a + b) + H$ is another coset.

The identity element in $(G/H, +)$ is the coset $0 + H$ (note that as a set this coset is equal to H). This is because, if $a \in G$,

$$(a + H) + (0 + H) = (a + 0) + H = a + H = (0 + a) + H = (0 + H) + (a + H).$$

Given $a \in G$, the inverse of the coset $a+H$ is the coset $(-a)+H$. This is because

$$(a+H)+((-a)+H) = (a+(-a)+H = 0+H = ((-a)+a)+H = ((-a)+H)+(a+H)$$

and $0 + H$ is the identity element in $(G/H, +)$.

Finally, let's show that $(G/H, +)$ is an abelian group. Suppose a, $b \in G$. Then

$$\begin{aligned}(b + H) + (a + H) &= (b + a) + H && \text{from the definition of addition in } G/H \\ &= (a + b) + H && b + a = a + b \text{ as } G \text{ is abelian} \\ &= (a + H) + (b + H) && \text{from the definition of addition in } G/H.\end{aligned} \tag{18}$$

14.1.4 Example Let $m \geq 2$ be an integer. We know that $m\mathbb{Z}$ is the subgroup of $\mathbb{Z}$ consisting of the multiples of m. Lets think about the quotient group $(\mathbb{Z}/m\mathbb{Z}, +)$. We're about to find out that we've seen this before.

Remember back in Section 7.2 we defined the congruence class modulo m of an integer a. This was $\overline{a} = \{b \in \mathbb{Z} \mid b \equiv a \pmod{m}\}$. It turns out that the congruence class $\overline{a}$ is equal to the coset $a + m\mathbb{Z}$ as follows:

Let $c \in \overline{a}$. Then $c \equiv a \pmod{m}$ which means that $c - a$ is a multiple of m. So $c - a = mk$ for some integer k. Then $c = a + mk$ and $c \in (a + m\mathbb{Z})$. Therefore $\overline{a} \subseteq a + m\mathbb{Z}$.

Conversely, if $d \in (a + m\mathbb{Z})$ then $d = a + ml$ for some $l \in \mathbb{Z}$ and $d - a$ is a multiple of m. This means that $d \equiv a \pmod{m}$ and $d \in \overline{a}$. Therefore $a+m\mathbb{Z} \subseteq \overline{a}$.

Since both $\overline{a} \subseteq a + m\mathbb{Z}$ and $a + m\mathbb{Z} \subseteq \overline{a}$ we have that $a + m\mathbb{Z} = \overline{a}$.

In Section 7.2 we defined an addition on the congruence classes as

$$\overline{a} + \overline{b} = \overline{a + b}.$$

Notice that we could now write this in coset notation as

$$(a + m\mathbb{Z}) + (b + m\mathbb{Z}) = (a + b) + m\mathbb{Z}$$

and that this exactly how addition in the quotient group $(\mathbb{Z}/m\mathbb{Z}, +)$ was defined in Definition 14.1.1.

So the congruence classes modulo n under addition is *exactly the same group* as the quotient group $(\mathbb{Z}/m\mathbb{Z}, +)$!

For this particular group, we'll stick to the notation $\overline{a}$ rather than the coset notation, $a + m\mathbb{Z}$.

Importantly, this means that taking the quotient group $(G/H, +)$ in the case when $G = \mathbb{Z}$ and $H = n\mathbb{Z}$ corresponds exactly to the group of conjugacy classes modulo n under addition. Thus, taking a quotient group can be seen as a generalisation of defining an addition on conjugacy classes modulo n in the sense that they coincide in this particular instance but that taking quotient groups can be applied more widely. ♢

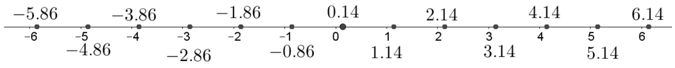
14.2 $\mathbb{R}/\mathbb{Z}$

14.2.1 Example Remember that $\mathbb{Z}$ is a subgroup of $\mathbb{R}$. What is the quotient group $\mathbb{R}/\mathbb{Z}$? Let's start by examining the cosets. These have form $a + \mathbb{Z} = \{a + k \mid k \in \mathbb{Z}\}$ where $a \in \mathbb{R}$.

For example

$$0.14 + \mathbb{Z} = \{0.14 + k \mid k \in \mathbb{Z}\} = \{\dots, -2.86, -1.86, -0.86, 0.14, 1.14, 2.14, \dots\}$$

is the set of all positive real numbers that have '14' after the decimal point together with the negative real numbers which have '86' after the decimal point. Here is what $0.14 + \mathbb{Z}$ looks like on a numberline.



Notice also that $0.14 + \mathbb{Z} = 1.14 + \mathbb{Z} = 11.14 + \mathbb{Z} = -3.86 + \mathbb{Z}$ in accordance with Remark 13.6.5.

You should think of the cosets as 'bringing together' or *identifying* all the real numbers which differ by a whole number. It may seem that this isn't particularly useful. After all, we're concentrating on the small fractional part of number and ignoring the bigger integer part. However in some situations, the fractional part is the important one. Let's see one of those situations. In Example 9.2.13 we defined the circle group

$$\mathbb{S} = \{\alpha \in \mathbb{C} \mid |\alpha| = 1\}.$$

Let

$$f : \mathbb{R} \to \mathbb{S}, \qquad f(\theta) = e^{2\pi i\theta}.$$

What happens to $f(\theta)$ as θ changes? If we start with $\theta = 0 \in \mathbb{R}$ and increase the value of θ, then $f(\theta)$ starts at $1 \in \mathbb{S}$ and moves anticlockwise. When θ reaches $1 \in \mathbb{R}$ then $f(\theta)$ will have done a complete circle and returned to $1 \in \mathbb{S}$. By the time θ reaches $2 \in \mathbb{R}$, $f(\theta)$ will have done another complete circle and returned again to $1 \in \mathbb{S}$.

In fact given any $\theta \in \mathbb{R}$ we have $f(\theta) = f(\theta + 1) = f(\theta - 1) = f(\theta + 2) = f(\theta - 2)$... This certainly means that f, whilst being surjective, is not injective. But notice that that every element in the coset $\theta + \mathbb{Z}$ is mapped to $e^{2\pi i\theta}$ and that this is the only coset with that property.

If we go back to the map

$$f : \mathbb{R} \to \mathbb{S}, \qquad f(\theta) = e^{2\pi i\theta},$$

we can define a similar map,

$$\hat{f} : \mathbb{R}/\mathbb{Z} \to \mathbb{S}, \qquad \hat{f}(\theta + \mathbb{Z}) = e^{2\pi i\theta}.$$

We'll now show that, although f *is not* a bijection, $\hat{f}$ *is* a bijection.

Notice that $\hat{f}$ is defined in terms of cosets so we have to make sure it's well-defined. In fact, we can show that $\hat{f}$ is both well-defined and injective at the same time:

Suppose $\hat{f}(\theta + \mathbb{Z}) = \hat{f}(\phi + \mathbb{Z})$, i.e. $e^{2\pi i\theta} = e^{2\pi i\phi}$. This happens if and only if $2\pi\theta = 2\pi\phi + 2n\pi$ for some $n \in Z$. This happens if and only if $\theta - \phi = n \in \mathbb{Z}$ which implies that $\theta + \mathbb{Z} = \phi + \mathbb{Z}$.

So we have that

$$\hat{f}(\theta + \mathbb{Z}) = \hat{f}(\phi + \mathbb{Z}) \text{ if and only if } \theta + \mathbb{Z} = \phi + \mathbb{Z}.$$

Notice that the 'left to right' implication in the above shows that $\hat{f}$ is injective and the 'right to left' implication in the above shows that $\hat{f}$ is well-defined.

Next, $\hat{f}$ is surjective. If $z \in \mathbb{S}$ then $z = e^{2\pi i\theta}$ for some $\theta \in [0, 1)$ and then $f(\theta + \mathbb{Z}) = z$.

Therefore $\hat{f}$ is a bijection.

It's worth thinking again about what has happened here; f was not injective but $\hat{f}$ is injective. This is because the cosets with respect to the subgroup $\mathbb{Z}$ of $\mathbb{R}$ gather together all the real numbers that get mapped to the same point in $\mathbb{S}$. So where there are many real numbers that get mapped to a particular point in $\mathbb{S}$ under f (preventing injectivity), there is only one coset that gets mapped to that point under $\hat{f}$ (giving injectivity).

We now want to understand the group $\mathbb{R}/\mathbb{Z}$, i.e. to understand it's addtion. Note that every coset in $\mathbb{R}/\mathbb{Z}$ can be written as $x + \mathbb{Z}$ where x is a unique number in the half-open interval

$$[0, 1) = \{x \in \mathbb{R} : 0 \leq x < 1\}.$$

For example

$$5.239 + \mathbb{Z} = 0.239 + \mathbb{Z} \text{ and } 34.1 + \mathbb{Z} = 0.1 + \mathbb{Z}.$$

In other words

$$\mathbb{R}/\mathbb{Z} = \{a + \mathbb{Z} \mid a \in [0, 1)\}.$$

So when we add $(a + \mathbb{Z}) + (b + \mathbb{Z})$, we can take the result of $a + b$, and simplify by subtracting 1 if necessary to obtain $c \in [0, 1)$, and then $(a + \mathbb{Z}) + (b + \mathbb{Z}) = c + \mathbb{Z}$. For example,

$$(0.7 + \mathbb{Z}) + (0.2 + \mathbb{Z}) = 0.9 + \mathbb{Z}, \qquad (0.7 + \mathbb{Z}) + (0.4 + \mathbb{Z}) = 1.1 + \mathbb{Z} = 0.1 + \mathbb{Z}.$$

Returning to our bijection $\hat{f}$ from $\mathbb{R}/\mathbb{Z}$ to $\mathbb{S}$, we'll now show that it's actually an isomorphism. For a reminder about isomorphisms look again at Lecture 12.

Since we already know that $\hat{f}$ is a bijection, we just need to show that $\hat{f}$ 'preserves the operations' between the two groups. $\mathbb{R}/\mathbb{Z}$ is an additive group and $\mathbb{S}$ is a multiplicative group so this means showing that

$$\hat{f}((\theta + \mathbb{Z}) + (\phi + \mathbb{Z})) = \hat{f}(\theta + \mathbb{Z}) \cdot \hat{f}(\phi + \mathbb{Z}).$$

We see this as follows

$$\hat{f}((\theta + \mathbb{Z}) + (\phi + \mathbb{Z})) = \hat{f}((\theta + \phi) + \mathbb{Z}) = e^{2\pi i(\theta+\phi)} = e^{2\pi i\theta}e^{2\pi i\phi} = \hat{f}(\theta + \mathbb{Z}) \cdot \hat{f}(\phi + \mathbb{Z}).$$

So the two groups $(\mathbb{R}/\mathbb{Z}, +)$ and $(\mathbb{S}, \cdot)$ are isomorphic, i.e. they are 'essentially the same'. $\Diamond$

Here is one further thought on how to think about $\mathbb{R}/\mathbb{Z}$. The fact that every coset in $\mathbb{R}/\mathbb{Z}$ can be written as $x + \mathbb{Z}$ where $x \in [0, 1)$ effectively identifies $\mathbb{R}/\mathbb{Z}$ with the interval $[0, 1)$. Think of starting at $0.97 + \mathbb{Z}$ and moving up in small steps of $0.01 + \mathbb{Z}$:

$$0.97 + \mathbb{Z}, 0.98 + \mathbb{Z}, 0.99 + \mathbb{Z}, 1.00 + \mathbb{Z} = 0.00 + \mathbb{Z}, 0.01 + \mathbb{Z}, 0.02 + \mathbb{Z}, 0.03 + \mathbb{Z}, \ldots$$

So we should really think of $\mathbb{R}/\mathbb{Z}$ as the interval $[0, 1)$ with one end joined to the other. If you take a string and join one end to the other you obtain a loop, or a 'circle'. This is what $\hat{f}$ is doing. It is showing that $\mathbb{R}/\mathbb{Z}$ is isomorphic to the unit circle $\mathbb{S}$. Indeed, the 2π in the formula for $\hat{f}$ is a 'stretching factor', since the interval $[0, 1)$ of length 1 has to be 'stretched' around the unit circle of perimeter 2π.

14.2.2 Exercise $\mathbb{R}/\mathbb{Z}$ has four elements of order 5. Find them.

14.2.3 Exercise Let $\alpha \in [0, 1)$. Show that the coset $\alpha + \mathbb{Z} \in \mathbb{R}/\mathbb{Z}$ has finite order in $\mathbb{R}/\mathbb{Z}$ if and only if α is rational.

14.3 $\mathbb{R}^2/\mathbb{Z}^2$

Having thought through $\mathbb{R}/\mathbb{Z}$ should help with understanding $\mathbb{R}^2/\mathbb{Z}^2$ (or $(\mathbb{R}^2/\mathbb{Z}^2, +)$ to give it its full name).

Firstly it's clear that $\mathbb{Z}^2$ is a subgroup of $\mathbb{R}^2$. A coset in $\mathbb{R}^2/\mathbb{Z}^2$ has the form $(a, b) + \mathbb{Z}^2$ where $a, b \in \mathbb{R}$. But we also have that

$$(a, b) + \mathbb{Z}^2 = (a + n, b + k) + \mathbb{Z}^2$$

for any integers n and k. This means that every coset can be written as $(x, y) + \mathbb{Z}^2$ where $x \in [0, 1)$ and $y \in [0, 1)$. For example

$$(3.45, 2.871) + \mathbb{Z}^2 = (0.45, 0.871) + \mathbb{Z}^2.$$

Effectively this identifies $\mathbb{R}^2/\mathbb{Z}^2$ with the unit square

$$\{(x, y) : 0 \le x < 1,\ 0 \le y < 1\}. \tag{19}$$

But you should think of this square as having the top side glued to the bottom side, and the right side glued to the left side! See Figure 22.

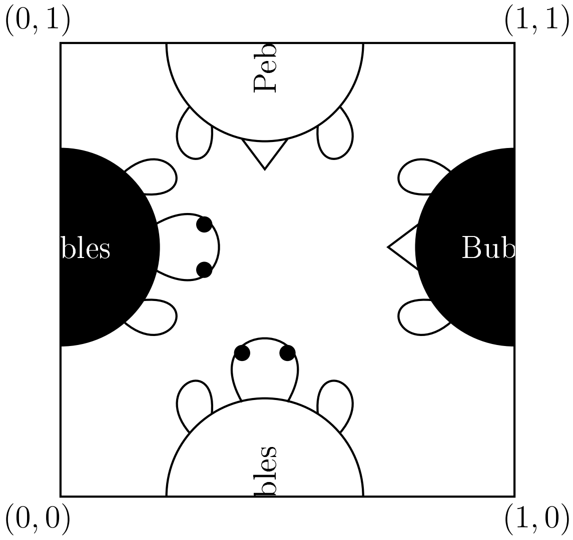


Figure 22: $\mathbb{R}^2/\mathbb{Z}^2$ is really just the unit square with the top side glued to the bottom side, and the right side glued to the left side.

14.3.1 Example In this example, we'll find the elements of order 2 in $\mathbb{R}^2/\mathbb{Z}^2$. Suppose $(x, y)+\mathbb{Z}^2$ is such an element, where x, $y \in \mathbb{R}$ belong to the interval $[0, 1)$. Then

$$((x, y) + \mathbb{Z}^2) + ((x, y) + \mathbb{Z}^2) = (0, 0) + \mathbb{Z}^2 \text{ if and only if } (2x, 2y) + \mathbb{Z}^2 = (0, 0) + \mathbb{Z}^2.$$

This happens if and only if $2x$, $2y$ are integers.

Hence

$$2x = \dots, -1, 0, 1, 2, 3, \dots, \qquad 2y = \dots, -1, 0, 1, 2, 3, \dots.$$

Therefore,

$$x = \cdots, \ -\frac{1}{2}, \ 0, \ \frac{1}{2}, \ 1, \ \frac{3}{2}, \cdots, \qquad y = \cdots, \ -\frac{1}{2}, \ 0, \ \frac{1}{2}, \ 1, \ \frac{3}{2}, \cdots.$$

As x and y belong to the interval $[0, 1)$, we see that $x = 0$ or $1/2$ and $y = 0$ or $1/2$. Hence

$$(x, y) + \mathbb{Z}^2 = (0, 0) + \mathbb{Z}^2, \quad (1/2, 0) + \mathbb{Z}^2, \quad (0, 1/2) + \mathbb{Z}^2, \quad (1/2, 1/2) + \mathbb{Z}^2.$$

However, the first of these has order 1. So the elements of order 2 in $\mathbb{R}^2/\mathbb{Z}^2$ are

$$(x, y) + \mathbb{Z}^2 = (1/2, 0) + \mathbb{Z}^2, \quad (0, 1/2) + \mathbb{Z}^2, \quad (1/2, 1/2) + \mathbb{Z}^2.$$

◊

14.3.2 Exercise Find all elements of order 3 in $\mathbb{R}^2/\mathbb{Z}^2$ (there are eight of them).

14.3.3 Exercise In $\mathbb{Z}^2$ consider $2\mathbb{Z}^2 = \{(2a, 2b) \mid a, b \in \mathbb{Z}\}$. Convince yourself that $2\mathbb{Z}^2$ is a subgroup of $\mathbb{Z}^2$ of index 4 and that

$$\mathbb{Z}^2/2\mathbb{Z}^2 = \{(0,0) + 2\mathbb{Z}^2, (1,0) + 2\mathbb{Z}^2, (0,1) + 2\mathbb{Z}^2, (1,1) + 2\mathbb{Z}^2\}.$$

Write down an addition table for $\mathbb{Z}^2/2\mathbb{Z}^2$.

14.3.4 Exercise How would you describe $\mathbb{C}/\mathbb{Z}[i]$? Is it really different from $\mathbb{R}^2/\mathbb{Z}^2$?

14.3.5 Exercise How would you describe $\mathbb{C}/\mathbb{Z}$? Find all elements of order 2.

14.4 $\mathbb{R}/\mathbb{Q}$

In this section, we shall briefly think about $\mathbb{R}/\mathbb{Q}$. In $\mathbb{R}/\mathbb{Z}$, the zero element is the coset $0 + \mathbb{Q} = \mathbb{Q}$. We treat the rationals as 'zero' (adding the coset $q + Q (= 0 + Q = Q)$ where $q \in Q$ to any coset $x + Q$ where $x \in \mathbb{R}$ leaves it unchanged as the coset $x + Q$).

This is a much trickier quotient group. The trickiness does not come from the definition; there is no difficulty there. We can add in $\mathbb{R}/\mathbb{Q}$ using the definition (17). The problem is with 'simplifying' the result. Let's try some numerical examples so that you see what I mean. If we take $a = 1 + \sqrt{2}$ and $b = 2/3 - \sqrt{2}$, then

$$(a + \mathbb{Q}) + (b + \mathbb{Q}) = (a + b) + \mathbb{Q} = 5/3 + Q = 0 + Q,$$

because $5/3 - 0 = 5/3 \in \mathbb{Q}$. However, if we take $a = \pi$ and $b = e$ (where these have their usual values) then

$$(\pi + \mathbb{Q}) + (e + \mathbb{Q}) = (\pi + e) + \mathbb{Q}.$$

Can we simplify this? For example, is this equal to $0 + \mathbb{Q}$? It is if and only if $\pi + e$ is a rational number. No one knows if the number $\pi + e$ is rational or not (but we know that both π and e are irrational). So we don't know if the result of the above calculation is equal to 'zero' in $\mathbb{R}/\mathbb{Q}$ or not. $\Diamond$

We've only looked at quotients of abelian additive groups. For general groups, things are more tricky. At the heart of the trickiness is that in the non-abelian setting the binary operation on cosets might not be well-defined. For now, if you've got to grips with $\mathbb{Z}/m\mathbb{Z}$, $\mathbb{R}/\mathbb{Z}$ and $\mathbb{R}^2/\mathbb{Z}^2$ then you've made an excellent start with quotients.

14.5 One more thought

We know that in $\mathbb{Z}/m\mathbb{Z}$, not only does addition make sense, but also multiplication makes sense. In the conjugacy class notation we use in $\mathbb{Z}/m\mathbb{Z}$ we defined multiplication by

$$\overline{a} \cdot \overline{b} = \overline{ab}. \tag{20}$$

Remembering that $\overline{a} = a + n\mathbb{Z}$, in coset notation this is

$$a + n\mathbb{Z} \cdot b + n\mathbb{Z} = ab + n\mathbb{Z}. \tag{21}$$

Now, you might ask why we don't define multiplication on $\mathbb{R}/\mathbb{Z}$ in the same way? OK, let's try using the same definition for multiplication on $\mathbb{R}/\mathbb{Z}$ and see what happens:

$$(0.5 + \mathbb{Z}) \times (0.5 + \mathbb{Z}) = 0.25 + \mathbb{Z}, \qquad (1.5 + \mathbb{Z}) \times (0.5 + \mathbb{Z}) = 0.75 + \mathbb{Z}.$$

There is a problem: in $\mathbb{R}/\mathbb{Z}$, the cosets $1.5 + \mathbb{Z}$ and $0.5 + \mathbb{Z}$ are equal, but the cosets $0.75 + \mathbb{Z}$ and $0.25 + \mathbb{Z}$ aren't. Multiplication doesn't make sense on $\mathbb{R}/\mathbb{Z}$.

The problem comes from the 'definition' for multiplication in (21). We're trying to define the product of $a + \mathbb{Z}$ and $b + \mathbb{Z}$ in terms of the representatives a, b of these cosets. But each coset has many representatives. For a definition such as this to make sense, the result must be independent of the choice of representatives. Now you might be wondering why multiplication in $\mathbb{Z}/m\mathbb{Z}$ makes sense. This was actually done in *Foundations/Sets and Numbers* but it is worth looking at the proof again. We return to the conjugacy class notation we use in $\mathbb{Z}/m\mathbb{Z}$ in the below.

14.5.1 Lemma Let $m \geq 2$ be an integer. Let a, a', b, b' satisfying

$$\overline{a} = \overline{a'}, \qquad \overline{b} = \overline{b'},$$

in $\mathbb{Z}/m\mathbb{Z}$. Then

$$\overline{ab} = \overline{a'b'}.$$

As before, we say that multiplication is *well-defined* on $\mathbb{Z}/m\mathbb{Z}$. This means that the result of a product does not depend on the choice of representatitives, even though it defined in terms of those representatives.

Proof. As $\overline{a} = \overline{a'}$ and $\overline{b} = \overline{b'}$ we know that

$$a' = a + km, \qquad b' = b + \ell m,$$

where k and ℓ are integers. So

$$a'b' = ab + m(kb + \ell a + mk\ell).$$

But $kb + \ell a + mk\ell$ is an integer as it is a sum of products of integers. So

$$a'b' \equiv ab \pmod{m},$$

which means

$$\overline{ab} = \overline{a'b'}.$$

Lecture 15 - The Alternating Group

Here we'll meet the *alternating group* A_n which is a subgroup of S_n. This group has had many applications in mathematics. Notably the fact that there is no formula for the roots of a quintic from its coefficients based on the four arithmetic operations and taking n^{th} roots can be proved as a consquence of the internal subgroup structure of this group.

15.1 Permutations and Transpositions

15.1.1 Lemma Every permutation can be written as a product of transpositions.

Note the absence of the word 'disjoint'.

Proof. We know that every permutation can be written a product of cycles. So it is enough to show that a cycle can be written as a product of transpositions. Check for yourself that

$$(a_1, a_2, \ldots, a_m) = (a_1, a_m) \cdots (a_1, a_3)(a_1, a_2). \tag{22}$$

15.1.2 Example Equation (22) gives a recipe for writing any cycle as a product of transpositions. For example,

$$(1, 5, 9) = (1, 9)(1, 5).$$

Note that these transpositions are not disjoint and so they don't have to commute. Check that

$$(1, 9)(1, 5) \neq (1, 5)(1, 9).$$

One thing to be careful about is that decomposition of a permutation as a product of transpositions is not in any way unique. For example, using (22) we have

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

However, you can also check that

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

So we can write $(1, 2, 3, 4)$ as a product of 3 transpositions and as a product of 5 transpositions. Can we write it as a product of 4 transpositions? Spend no more and no less than five minutes thinking about this. ♢

15.2 Even and Odd Permutations

Let $n \geq 2$ be an integer. Let $x_1, x_2, \ldots, x_n$ be variables, and let P_n be the polynomial

$$P_n = \prod_{1 \leq i < j \leq n} (x_i - x_j).$$

The polynomial P_n is called the *n-th alternating polynomial.* It will help us to discover an important subgroup of S_n called the *alternating group* and denoted by A_n. Let us write down the first three alternating polynomials:

$$P_2 = x_1 - x_2, \qquad P_3 = (x_1 - x_2)(x_1 - x_3)(x_2 - x_3),$$
$$P_4 = (x_1 - x_2)(x_1 - x_3)(x_1 - x_4)(x_2 - x_3)(x_2 - x_4)(x_3 - x_4).$$

If $\sigma \in S_n$ then define

$$\sigma(P_n) = \prod_{1 \leq i < j \leq n} (x_{\sigma(i)} - x_{\sigma(j)}).$$

15.2.1 Example Let $\sigma = (1, 2) \in S_3$. Then

$$\begin{aligned} \sigma(P_3) &= (x_{\sigma 1} - x_{\sigma 2})(x_{\sigma 1} - x_{\sigma 3})(x_{\sigma 2} - x_{\sigma 3}) \\ &= (x_2 - x_1)(x_2 - x_3)(x_1 - x_3) \\ &= -P_3. \end{aligned}$$

We obtain the equality in the final step of the calculation by comparing the factors of P_3 with the factors of $\sigma(P_3)$, and **not** by expanding! Note that the first factor of P_3 changed sign and the last two factors are swapped. So $\sigma(P_3) = -P_3$.

Now let $\tau = (1, 2, 3) \in S_3$. Then

$$\begin{aligned} \tau(P_3) &= (x_{\tau(1)} - x_{\tau(2)})(x_{\tau(1)} - x_{\tau(3)})(x_{\tau(2)} - x_{\tau(3)}) \\ &= (x_2 - x_3)(x_2 - x_1)(x_3 - x_1) \\ &= P_3. \end{aligned}$$

Again we obtain equality by comparing factors. Write down $\rho(P_3)$ for the other four elements $\rho \in S_3$. ♢

This example involving P_3 for an element of S_3 is quite straightforward. Let's look at a more difficult example.

15.2.2 Example Let $\tau = (2,4) \in S_5$. We want to check that $\tau(P_5) = -P_5$. Some factors of P_5 are unaffected. For example, $\tau(x_1 - x_3) = x_{\tau(1)} - x_{\tau(3)} = x_1 - x_3$. The ones that aren't affected are the ones that don't contain either of x_2 or x_4. These are,

$$x_1 - x_3, \qquad x_1 - x_5, \qquad x_3 - x_5.$$

We will split the other factors of P_5 into four groups [14]:

$$\begin{array}{lll}
\text{(I)} & x_1 - x_2, & x_1 - x_4, \\
\text{(II)} & x_2 - x_3, & x_3 - x_4, \\
\text{(III)} & x_2 - x_5, & x_4 - x_5, \\
\text{(IV)} & x_2 - x_4. &
\end{array}$$

Let's study what τ does to each group. Note that

$$\tau(x_1 - x_2) = x_1 - x_4, \qquad \tau(x_1 - x_4) = x_1 - x_2.$$

Thus τ swaps the factors in group (I) whilst *keeping their signs the same.* But

$$\tau(x_2 - x_3) = x_4 - x_3 = -(x_3 - x_4), \qquad \tau(x_3 - x_4) = x_3 - x_2 = -(x_2 - x_3).$$

Thus τ swaps the factors in group (II) and *changes the sign of each.* Moreover,

$$\tau(x_2 - x_5) = x_4 - x_5, \qquad \tau(x_4 - x_5) = x_2 - x_5.$$

So τ swaps the factors in group (III) whilst *keeping their signs the same.* Finally,

$$\tau(x_2 - x_4) = x_{\tau 2} - x_{\tau 4} = x_4 - x_2 = -(x_2 - x_4).$$

So the one factor in group (IV) simply *changes sign.* We see that $\tau(P_5)$ has the same factors as P_5 with three sign changes: $\tau(P_5) = (-1)^3 P_5 = -P_5$. $\diamondsuit$

15.2.3 Lemma Let $\tau \in S_n$ be a transposition. Then $\tau(P_n) = -P_n$.

Proof. Let $\tau = (\ell, m)$. The transposition (ℓ, m) swaps ℓ and m, and keeps everything else fixed. In particular $(\ell, m) = (m, \ell)$. So we can suppose that $\ell < m$. Any factor $x_i - x_j$ where neither i nor j is equal to ℓ nor m, is unaffected by τ.

[14] The word "groups" here is used in its English language sense, not in its mathematical sense.

We pair off the other factors as follows:

$$\text{(I)}\quad \begin{cases} x_1 - x_\ell, & x_1 - x_m, \\ x_2 - x_\ell, & x_2 - x_m, \\ \vdots & \vdots \\ x_{\ell-1} - x_\ell, & x_{\ell-1} - x_m, \end{cases}$$

$$\text{(II)}\quad \begin{cases} x_\ell - x_{\ell+1}, & x_{\ell+1} - x_m, \\ x_\ell - x_{\ell+2}, & x_{\ell+2} - x_m, \\ \vdots & \vdots \\ x_\ell - x_{m-1}, & x_{m-1} - x_m, \end{cases}$$

$$\text{(III)}\quad \begin{cases} x_\ell - x_{m+1}, & x_m - x_{m+1}, \\ x_\ell - x_{m+2}, & x_m - x_{m+2}, \\ \vdots & \vdots \\ x_\ell - x_n, & x_m - x_n, \end{cases}$$

$$\text{(IV)}\quad \begin{cases} x_\ell - x_m. \end{cases}$$

Now τ swaps each pair in (I), keeping the signs the same; it swaps each pair in (II) and changes the sign of each; it swaps each pair in (III), keeping the signs the same; it changes the sign of $x_\ell - x_m$. So $\tau(P_n)$ has exactly the same factors as P_n, up to a certain number of sign changes. How many sign changes? The number of sign changes is:

$$2(m - \ell - 1) + 1.$$

The 1 is for changing the sign of $x_\ell - x_m$. There are 2 sign changes coming from each pair in (II). The number of such pairs is $m - \ell - 1$. Since the number of sign changes is odd, we see that $\tau(P_n) = -P_n$.

15.2.4 Lemma If $\sigma \in S_n$ then $\sigma(P_n) = \pm P_n$. More precisely, if σ is a product of an even number of transpositions then $\sigma(P_n) = P_n$ and if σ is a product of an odd number of transpositions then $\sigma(P_n) = -P_n$.

Proof. Recall, by Lemma 15.1.1, that we can write every permutation as a product of transpositions. Every transposition changes the sign of P_n. The lemma follows.

15.2.5 Example We have noted in Example 15.1.2 that the way we express a permutation as a product of transpositions is not unique. Indeed we saw that

$$(1, 2, 3, 4) = (1, 4)(1, 3)(1, 2), \qquad (1, 2, 3, 4) = (2, 3)(1, 3)(3, 5)(3, 4)(4, 5).$$

So we can write $(1,2,3,4)$ as a product of 3 transpositions and as a product of 5 transpositions. We asked the question of whether $(1,2,3,4)$ can be written as a product of 4 transpositions? Write $\sigma = (1,2,3,4)$. From the above lemma, we see that $\sigma(P_n) = -P_n$. If we're able to write σ as a product of an even number of transpositions then $\sigma(P_n) = P_n$. We would then have $P_n = -P_n$ which is a contradiction. Therefore we cannot write σ as a product of 4 transpositions. ♢

This example tells us how to prove the following theorem.

15.2.6 Theorem Every permutation in S_n can be written as a product of either an even number of transpositions, or an odd number of transpositions but ***not both***.

Proof. Let $\sigma \in S_n$. Then, by Lemma 15.1.1, S_n can be written as a product of transpositions. Suppose σ can be written both as a product of an even number of transpositions and a a product of an odd number of transpositions. Then by Lemma 15.2.4 we have $\sigma(P_n) = P_n$ and $\sigma(P_n) = -P_n$. This implies that $P_n = -P_n$, a contradiction. The conclusion follows. ♢

15.2.7 Definition We shall call a permutation *even* if we can write it as a product of an even number of transpositions, and we shall call it *odd* if we can write it as a product of an odd number of transpositions.

15.2.8 Example $(1,2,3,4)$ is an odd permutation because we can write it as the product of 3 transpositions:

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

Indeed, a cycle of length n can be written as product of $n-1$ transpositions by (22). So we need to be careful because a cycle of length n is even if n is odd, and it is odd if n is even!

The permutation $(1,2,3)(4,5)$ is the product of an even permutation which is $(1,2,3)$ and an odd permutation which is the transposition $(4,5)$. Thus $(1,2,3)(4,5)$ is an odd permutation.

What about the identity element id? Note that $\mathrm{id}(P_n) = P_n$, so id must be even. We must be able to write it as a product of an even number of transpositions. A mathematician would say that the identity element is the product of zero transpositions, so it is even. If you don't feel comfortable with that kind of reasoning, instead, note that

$$\mathrm{id} = (1,2)(1,2),$$

which does allow us to check that id is indeed even.

We now come to define a very important group. Let $n \geq 2$. We define the *n-th alternating group* to be

$$A_n = \{\sigma \in S_n \mid \sigma \text{ is even}\}.$$

As usual, all we've done is specify a subset of S_n which we've denoted by A_n and we must indeed show that A_n is a group.

15.2.9 Theorem A_n is a subgroup of S_n.

Proof. We've already seen that the identity element id is even, so $\mathrm{id} \in A_n$. If σ, $\rho \in A_n$ then we can write each as an even number of transpositions. Therefore the product $\sigma\rho$ can be written as an even number of transpositions (even+even=even). Hence $\sigma\rho \in A_n$.

Finally we must show that the inverse of an even permutation is even. Suppose σ is even. We can write

$$\sigma = \tau_1\tau_2 \dots \tau_m$$

where the τ_i are transpositions, and m is even. Now

$$\begin{aligned}\sigma^{-1} &= (\tau_1\tau_2\cdots\tau_m)^{-1}\\ &= \tau_m^{-1}\tau_{m-1}^{-1}\cdots\tau_1^{-1}\\ &= \tau_m\tau_{m-1}\cdots\tau_1.\end{aligned}$$

Here you should convince yourself that $\tau^{-1} = \tau$ for any transposition τ. Since m is even, we find that σ^{-1} is even and so $\sigma^{-1} \in A_n$.

Hence A_n is a subgroup of S_n. ◊

15.2.10 Example Recall that $S_2 = \{\mathrm{id}, (1,2)\}$. We see that $A_2 = \{\mathrm{id}\}$ is the trivial subgroup. ◊

15.2.11 Example Recall that S_3 has $3! = 6$ elements:

$$S_3 = \{\mathrm{id}, (1,2), (1,3), (2,3), (1,2,3), (1,3,2)\}.$$

Then

$$A_3 = \{\mathrm{id}, (1,2,3), (1,3,2)\}.$$

Note that S_3 is non-abelian, but you can check that A_3 is abelian. $\Diamond$

In the above examples we saw that A_n has half the number of elements of S_n for $n = 2$, 3. In fact, this pattern continues.

15.2.12 Theorem Let $n \geq 2$. Then A_n has order

$$|A_n| = \frac{1}{2}|S_n| = \frac{n!}{2}.$$

Proof. We know by Lagrange's Theorem that

$$|S_n| = [S_n : A_n]|A_n|.$$

To prove the theorem it is sufficient to show that the index $[S_n : A_n] = 2$. Fix a transposition τ (e.g. $\tau = (1,2)$). We shall show that the distinct cosets of A_n in S_n are A_n and τA_n. It will then follow that the index $[S_n : A_n] = 2$, completing the proof.

We know that A_n is the subset (indeed subgroup) of S_n consisting of all the even permutations. Thus τA_n consists only of odd permutations. Does τA_n contain all the odd permutations? Suppose σ is odd. Then $\tau\sigma$ is even and is hence in A_n. Therefore $\tau(\tau\sigma)$ is in the coset τA_n. But

$$\tau(\tau\sigma) = \tau^2\sigma = \sigma,$$

since transpositions have order 2, and so $\sigma \in \tau A_n$.

We have now shown that τA_n is the set of all odd permutations, and we know that A_n is the set of all even permutations. Are there any other cosets? If there were any they would have to overlap with either A_n or τA_n, and we know that cosets are either disjoint or equal (Lemma 13.6.3). So there aren't any other cosets and the proof is complete. $\Diamond$

15.2.13 Exercise Let ρ and τ be as given in Exercise 11.3.5. Write ρ and τ as products of transpositions and state if they're even or odd.

15.2.14 Exercise Write down the elements of A_3 and check that it is cyclic (and hence abelian). Show that A_n is non-abelian for $n \geq 4$.

15.2.15 Exercise Let f be a polynomial in variables $x_1, x_2, \dots, x_n$. Let σ be a permutation in S_n. We define $\sigma(f)$ to be the polynomial $f(x_{\sigma 1}, x_{\sigma 2}, \dots, x_{\sigma n})$. For example, if $f = x_1 + x_2^2 + x_3 x_4$ and $\sigma = (1,4)(2,3)$ then σ swaps x_1 and x_4, and swaps x_2 and x_3; thus $\sigma(f) = x_4 + x_3^2 + x_2 x_1$. Compute $\sigma(f)$ for the following pairs f, σ:

(i) $f = x_1^2 - x_2 x_3$, $\sigma = (1,2,3)$.

(ii) $f = x_1 x_2 + x_3 x_4$, $\sigma = (1,3)(2,4)$.

15.2.16 Exercise Let f be a polynomial in variables $x_1, \dots, x_n$.

(a) Let H be a subgroup of S_n. We say that f is *H-invariant* if it satisfies the property that $\sigma(f) = f$ for all $\sigma \in H$. We say that f is *symmetric* if it is S_n-invariant. Find a polynomial in x_1, x_2, x_3, x_4 that is D_4-invariant but not symmetric.

(b) Define $\mathrm{Fix}(f) = \{\sigma \in S_n : \sigma(f) = f\}$. Show that $\mathrm{Fix}(f)$ is a subgroup of S_n. Write down $\mathrm{Fix}(f)$ for the following polynomials in $x_1, \dots, x_4$:

(i) $x_4^2 + x_1 x_2 x_3$.

(ii) $x_1 x_2 + x_3 x_4$.

15.2.17 Exercise Let ρ and $\tau \in S_n$. Show that τ is even if and only if $\rho^{-1}\tau\rho$ is even. (**Hint: It will help to show that if $\rho = c_1 c_2 \cdots c_m$ as a product of transpositions, then $\rho^{-1} = c_m c_{m-1} \dots c_1$**).

Lecture 16 - Rings

16.1 Definition

A *ring* is a triple $(R, +, \cdot)$, where R is a set and $+$, $\cdot$ are binary operations on R such that the following properties hold

(i) (closure) for all a, $b \in R$, $a + b \in R$ and $a \cdot b \in R$;

(ii) (associativity of addition) for all a, b, $c \in R$

$$(a + b) + c = a + (b + c);$$

(iii) (existence of an additive identity element) there is an element $0 \in R$ such that for all $a \in R$,

$$a + 0 = 0 + a = a.$$

(iv) (existence of additive inverses) for all $a \in R$, there an element, denoted by $-a$, such that

$$a + (-a) = (-a) + a = 0;$$

(v) (commutativity of addition) for all a, $b \in R$,

$$a + b = b + a;$$

(vi) (associativity of multiplication) for all a, b, $c \in R$,

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

(vii) (distributivity) for all a, b, $c \in R$,

$$a \cdot (b + c) = a \cdot b + a \cdot c; \qquad (b + c) \cdot a = b \cdot a + c \cdot a;$$

(viii) (existence of a multiplicative identity) there is an element $1 \in R$ so that for all $a \in R$,

$$1 \cdot a = a \cdot 1 = a.$$

Moreover, a ring $(R, +, \cdot)$ is said to be *commutative*, if it satisfies the following additional property:

(ix) (commutativity of multiplication) for all a, $b \in R$,

$$a \cdot b = b \cdot a.$$

Note that the word ‘commutative’ in the phrase ‘commutative ring’ refers to multiplication. Commutativity of addition is part of the definition of ring. Some textbooks omit property (viii) from the definition of a ring. Those textbooks call a ring satisfying (viii) a *ring with unity*. We shall always assume that our rings satisfy (viii).

Observe, from properties (i)–(v), if $(R, +, \cdot)$ is a ring, then $(R, +)$ is an abelian group.

16.2 Examples

16.2.1 Example You know lots of examples of rings: $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\mathbb{R}[x]$, etc. All these examples are commutative rings.

16.2.2 Example Let

$$M_{2\times 2}(\mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, b, c, d \in \mathbb{R} \right\}.$$

This is the set of 2×2 matrices with real entries. From the properties of matrices it is easy to see that $M_{2\times 2}(\mathbb{R})$ is a ring with the usual addition and multiplication of matrices. The additive identity is the zero matrix, and the multiplicative identity is I_2. The ring $M_{2\times 2}(\mathbb{R})$ is an example of a non-commutative ring, as matrix multiplication is non-commutative.

Similarly we define $M_{2\times 2}(\mathbb{C})$, $M_{2\times 2}(\mathbb{Z})$, $M_{2\times 2}(\mathbb{Q})$. These are all non-commutative rings. $\Diamond$

16.2.3 Theorem Let m be an integer satisfying $m \geq 2$. Then $\mathbb{Z}/m\mathbb{Z}$ is a ring.

Proof. We really mean that $(\mathbb{Z}/m\mathbb{Z}, +, \cdot)$ is a commutative ring. We’ve already seen that $\mathbb{Z}/m\mathbb{Z}$ is closed under addition and multiplication, and that $(\mathbb{Z}/m\mathbb{Z}, +)$ is an abelian group. That leaves associativity of multiplication, distributivity and the existence of a multiplicative indentity to check. These all follow from the corresponding properties in $\mathbb{Z}$ as follows. Given $a, b, c \in \mathbb{Z}$ we have:

associativity of multiplication: $\overline{a}.(\overline{b}.\overline{c}) = \overline{a}.\overline{bc} = \overline{a(bc)} = \overline{(ab)c} = \overline{ab}.\overline{c} = (\overline{a}.\overline{b}).\overline{c}$.

distributivity: $\overline{a}.(\overline{b} + \overline{c}) = \overline{a}.\overline{b + c} = \overline{a(b + c)} = \overline{ab + ac} = \overline{ab} + \overline{ac} = \overline{a}.\overline{b} + \overline{a}.\overline{c}$.

existence of a multiplicative identity: $\overline{1}.\overline{a} = \overline{1.a} = \overline{a} = \overline{a.1} = \overline{a}\overline{1}$. $\Diamond$

16.2.4 Example You're familiar with the following two binary operations on $\mathbb{R}^3$: addition and the cross product (also known as the vector product). Is $(\mathbb{R}^3, +, \times)$ a ring? No. First the cross product is not associative. For example,

$$\mathbf{i} \times (\mathbf{j} \times \mathbf{j}) = 0, \qquad (\mathbf{i} \times \mathbf{j}) \times \mathbf{j} = -\mathbf{i}.$$

We only need one of the properties (i)–(viii) to fail for us to conclude that $(\mathbb{R}^3, +, \times)$ is not a ring. We know that (vi) fails. It is interesting to note that (viii) fails too, as we now show. Indeed,

$$\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}. \tag{23}$$

Suppose $\mathbf{1}$ is a vector in $\mathbb{R}^3$ that satisfies

$$\mathbf{a} \times \mathbf{1} = \mathbf{1} \times \mathbf{a} = \mathbf{a}$$

for all $\mathbf{a} \in \mathbb{R}^3$. From (23) we see that $\mathbf{a} = -\mathbf{a}$ for all $\mathbf{a} \in \mathbb{R}^3$. This gives a contradiction. Therefore (viii) fails too. ♢

16.2.5 Example Consider $(\mathbb{R}[x], +, \circ)$, where $\circ$ is composition of polynomials. Is this a ring? No. It is easy to see that all the required properties hold except for distributivity (the "multiplicative identity" is the polynomial $f(x) = x$). Let us give a counterexample to show that distributivity fails. Let

$$f(x) = x^2, \qquad g(x) = x, \qquad h(x) = x.$$

Then

$$f \circ (g + h) = f(2x) = 4x^2; \qquad f \circ g + f \circ h = x^2 + x^2 = 2x^2.$$

♢

16.2.6 Example The **zero ring** is the ring with just one element $\{0\}$. In this ring $1 = 0$, and there is only one possible definition of addition and multiplication: $0 + 0 = 0$, $0 \cdot 0 = 0$. The zero ring is not interesting.

Let R be a ring in which $1 = 0$. Then $a = a \cdot 1 = a \cdot 0 = 0$ for all $a \in R$ and so R is the zero ring. To summarise a ring is the zero ring if and only if $1 = 0$.

16.2.7 Example Let's step back a little and think about $\mathbb{R}^2$. We know that $(\mathbb{R}^2, +)$ is an abelian group. Is there a way of defining multiplication on $\mathbb{R}^2$ so that we obtain a ring? We will define two different multiplications that make $\mathbb{R}^2$ into a ring. The first is more obvious: we define

$$(a_1, a_2) \times (b_1, b_2) = (a_1 b_1, a_2 b_2).$$

With this definition, you can check that $(\mathbb{R}^2, +, \times)$ is a ring, where the multiplicative identity is $\mathbf{1} = (1, 1)$.

The other way is more subtle: we define

$$(a_1, a_2) \times (b_1, b_2) = (a_1b_1 - a_2b_2, a_1b_2 + a_2b_1). \tag{24}$$

Where does this definition come from? Recall that $\mathbb{R}^2$ is represented geometrically by the plane, and $\mathbb{C}$ is represented geometrically by the plane. If we're thinking of points in the plane as elements of $\mathbb{R}^2$ then we write them as ordered pairs of real numbers: (a, b). If we're thinking of points in the plane as elements of $\mathbb{C}$ then we write them in the form $a + ib$ where again a, b are real numbers. We multiply in $\mathbb{C}$ using the rule

$$(a_1 + ia_2) \times (b_1 + ib_2) = (a_1b_1 - a_2b_2) + i(a_1b_2 + a_2b_1). \tag{25}$$

Notice that definitions (24), (25) are exactly the same at the level of points on the plane. We've used the multiplicative structure of $\mathbb{C}$ to define multiplication on $\mathbb{R}^2$. With this definition, $(\mathbb{R}^2, +, \times)$ is a ring. What is the multiplicative identity? It's not $(1, 1)$. For example $(1, 1) \times (1, 1) = (0, 2)$. Think about the multiplicative identity in $\mathbb{C}$. This is simply $1 = 1+0i$. So the multiplicative identity in $(\mathbb{R}^2, +, \times)$ (with multiplication defined as in (24)) is $(1, 0)$. Check for yourself that

$$(a_1, a_2) \times (1, 0) = (1, 0) \times (a_1, a_2) = (a_1, a_2).$$

$\Diamond$

Here a couple of quick lemmas about rings.

16.2.8 Lemma Let R be a ring and $a \in R$. Then $0.a = 0 = a.0$.

Proof. We have (make sure you can see why each of the two equalities in the below is true):

$$0.a = (0 + 0).a = 0.a + 0.a.$$

Adding the additive inverse of $0.a$, namely $-(0.a)$ to both sides of this equation gives $0 = 0.a - 0.a = 0.a + 0.a - 0.a = 0.a$.

That $0 = a.0$ holds similarly. $\Diamond$

16.2.9 Lemma Let R be a ring and $a, b \in R$. Then $-(a.b) = (-a).b = a.(-b)$.

Proof. We have that $ab + (-a).b = (a + (-a)).b = 0.b = 0$ by Lemma 16.2.8. By the uniqueness of additive inverses (see 6.1.2, noting that this is written in mutiplicative notation) it follows that $(-a).b = -(ab)$. To get the other result consider $ab + a.(-b)$ in a similar way. $\Diamond$

16.3 Subrings

Just as we have subgroups, so we have subrings.

16.3.1 Definition Let $(R, +, \cdot)$ be a ring. Let S be a subset of R and suppose that $(S, +, \cdot)$ is also a ring with the same multiplicative identity. Then we say that S is a subring of R (or more formally $(S, +, \cdot)$ is a subring of $(R, +, \cdot)$). $\Diamond$

For S to be a subring of R, we want S to a ring with respect to *the same two binary operations* that makes R a ring, and $1_R \in S$ where 1_R is the multiplicative identity of R.

16.3.2 Example $\mathbb{Z}$ is a subring of $\mathbb{R}$; $\mathbb{Q}$ is a subring of $\mathbb{R}$; $\mathbb{Z}$ is a subring of $\mathbb{Q}$; $\mathbb{R}$ is a subring of $\mathbb{R}[x]$. $\Diamond$

Theorem 9.2.1 gave a criterion for a subset of a group to be a subgroup. As you'd expect we have a similar criterion for a subset of a ring to be a subring.

16.3.3 Theorem Let R be a ring. A subset S of R is a subring if and only if it satisfies the following conditions

(a) $0, 1 \in S$ (that is S contains the additive and multiplicative identity elements of R);

(b) if $a, b \in S$ then $a + b \in S$;

(c) if $a \in S$ then $-a \in S$;

(d) if $a, b \in S$ then $ab \in S$.

Proof.

Let's proof this from 'left to right' first. So suppose that the subset S is a subring of R.

By theorem 9.2.1, since $(S, +)$ is a subgroup of $(R, +)$, $0 \in S$ and both (b) and (c) above are true.

We know from the definition of a subring that the multiplicative identity from R is in S, so we now know that (a) above is true. Also from the definition of a subring, if $a, b \in S$ then $ab \in S$, so (d) above is true.

Now for 'right to left' . Suppose S is a subset of R and $(a),(b),(c),(d)$ above are true.

Since $0 \in S$ and since (b) and (c) are true, by 9.2.1 $(S,+)$ is a subgroup of $(R,+)$.

(a) above ensures the existence of a multiplicative identity in S and (d) tells us that S is closed under multiplication.

All that remains to check is the commutativity of the addition in S, the associativity of the multiplication in S and the distributive rules in S. But these all follow immediately because they hold in R and any elements of S are also elements of R. $\Diamond$

16.3.4 Example In Example 9.2.3, we saw that the set of even integers $2\mathbb{Z}$ is a subgroup of $\mathbb{Z}$. Strictly speaking, $(2\mathbb{Z},+)$ is a subgroup of $(\mathbb{Z},+)$. Now we know that $(\mathbb{Z},+,\cdot)$ is a ring. Is $(2\mathbb{Z},+,\cdot)$ a subring? From Theorem 16.3.3 we see that it isn't because $1 \notin 2\mathbb{Z}$. $\Diamond$

16.3.5 Example In view of the previous example, let's try to discover if $\mathbb{Z}$ has any subrings other than itself. Let S be a subring of $\mathbb{Z}$. We know that $0,\ 1 \in S$. Also, by (b) we know that $2 = 1+1 \in S$. Repeating the argument, $3 = 2+1 \in S$ and so on. By induction we know that $0,1,2,\ldots$ are all in S. But by (c), if $a \in S$ then $-a \in S$. So $\ldots,-3,-2,-1$ are also in S. Hence $\mathbb{Z}$ is contained in S. But S is a subset of $\mathbb{Z}$. So they must be equal: $S = \mathbb{Z}$.

Therefore, the only subring of $\mathbb{Z}$ is $\mathbb{Z}$ itself. By contrast, in Section 10.3 we saw that $\mathbb{Z}$ has infinitely many subgroups. $\Diamond$

16.3.6 Exercise Let m be an integer satisfying $m \geq 2$. Show that the only subring of $\mathbb{Z}/m\mathbb{Z}$ is $\mathbb{Z}/m\mathbb{Z}$ itself. $\Diamond$

The easiest way to show that a set is a ring is to show that it is a subring of a known ring. If you do this, you only have four properties to check (a),(b),(c),(d). If you don't do this, you'll have eight properties to check (i)–(viii). The following two examples will help you appreciate this principle.

16.3.7 Example Let

$$\mathbb{Z}[i] = \{a+bi : a,b \in \mathbb{Z}\}.$$

The set $\mathbb{Z}[i]$ is called the set of *Gaussian integers.* How can we show that $\mathbb{Z}[i]$ is a ring?

We can try checking the eight defining properties of a ring. However, we note that $\mathbb{Z}[i]$ is contained in $\mathbb{C}$. Indeed, it is the set of complex numbers where the real and imaginary parts are integers. So let's prove that $\mathbb{Z}[i]$ is a subring of $\mathbb{C}$.

Now $0 = 0 + 0i$, $1 = 1 + 0i$ are clearly in $\mathbb{Z}[i]$. Suppose α, $\beta \in \mathbb{Z}[i]$. Write

$$\alpha = a_1 + a_2 i, \qquad \beta = b_1 + b_2 i,$$

where a_1, a_2, b_1, b_2 are integers. To apply Theorem 16.3.3 we need to check that $\alpha + \beta$, $-\alpha$ and $\alpha\beta$ are in $\mathbb{Z}[i]$. We note that

$$\alpha + \beta = (a_1 + b_1) + (a_2 + b_2)i, \qquad -\alpha = -a_1 + (-a_2)i,$$

and

$$\alpha\beta = (a_1 a_2 - b_1 b_2) + (a_1 b_2 + a_2 b_1)i.$$

Since we want to show that $\alpha + \beta$, $-\alpha$ and $\alpha\beta$ are in $\mathbb{Z}[i]$, we want to show that their real and imaginary parts are integers. Now as a_1, a_2, b_1, b_2 are integers, so are

$$a_1 + b_1, \quad a_2 + b_2, \quad -a_1, \quad -b_1, \quad a_1 a_2 - b_1 b_2, \quad a_1 b_2 + a_2 b_1.$$

Hence $\alpha + \beta$, $-\alpha$ and $\alpha\beta$ are in $\mathbb{Z}[i]$. By Theorem 16.3.3, we see that $\mathbb{Z}[i]$ is a subring of $\mathbb{C}$. Since $\mathbb{Z}[i]$ is a subring, it is a ring! ♢

16.3.8 Exercise Let S be a subring of $\mathbb{Z}[i]$. Suppose $i \in S$. Show that $S = \mathbb{Z}[i]$.

16.3.9 Example Let

$$S = \left\{ \frac{a}{2^r} : a, r \in \mathbb{Z}, r \geq 0 \right\}.$$

Then S is a ring.

We'll show this following the same strategy as the previous example. First think of a ring that contains S. The elements of S are rational numbers whose denominator is a power of 2; for example

$$7 = \frac{7}{2^0}, \qquad \frac{-1}{2}, \qquad \frac{15}{8} = \frac{15}{2^3}$$

are elements of S. An obvious choice of a ring that contains S is $\mathbb{Q}$, the ring of rational numbers. So let's show that S is a subring of $\mathbb{Q}$. Clearly $0 = 0/2^0$ and $1 = 1/2^0$ are in S. Suppose α, β are elements of S. We can write

$$\alpha = \frac{a}{2^r}, \qquad \beta = \frac{b}{2^s},$$

where a, b, r, $s \in \mathbb{Z}$ and r, $s \geq 0$. We want to check that $\alpha + \beta$, $-\alpha$ and $\alpha\beta$ are in S. Note that

$$-\alpha = \frac{-a}{2^r}, \qquad \alpha\beta = \frac{ab}{2^{r+s}}.$$

Clearly $-\alpha$, $\alpha\beta$ are in S, since $-a$, $a + b$, r, $r + s$ are integers and r, $r + s \geq 0$. Now for the sum, we'll assume without loss of generality that $r \geq s$. Then

$$\alpha + \beta = \frac{a + 2^{r-s}b}{2^r}.$$

Now since a, b, r, s are integers and $r \geq s$, we have $a + 2^{r-s}b$ is also an integer. Clearly, $\alpha + \beta$ is in S. By Theorem 16.3.3, S is a subring and therefore a ring. $\Diamond$

16.3.10 Exercise Let

$$\mathbb{Z}[2i] = \{a + 2bi : a, b \in \mathbb{Z}\}.$$

Show that $\mathbb{Z}[2i]$ is a subring of $\mathbb{Z}[i]$. Is $\{2a + 2bi : a, b \in \mathbb{Z}\}$ a subring of $\mathbb{Z}[i]$?

16.3.11 Exercise Which of the following are subrings of $M_{2\times 2}(\mathbb{R})$? If so, are they commutative?

(i) $\left\{ \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} : a, b, c \in \mathbb{R} \right\}$.

(ii) $\left\{ \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} : a, b \in \mathbb{R} \right\}$.

(iii) $\left\{ \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} : a, b \in \mathbb{R} \right\}$.

(iv) $\left\{ \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} : a \in \mathbb{R}, b \in \mathbb{Z} \right\}$.

(v) $\left\{ \begin{pmatrix} a & b \\ -b & a \end{pmatrix} : a, b \in \mathbb{R} \right\}$.

(vi) $\{A \in M_{2\times 2}(\mathbb{R}) : \det(A) = 1\}$.

Lecture 17 - The Unit Group of a Ring

17.1 The Unit Group of a Ring

Recall that we defined $\mathbb{R}^*$, $\mathbb{Q}^*$, $\mathbb{C}^*$ be removing from $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{C}$ the zero element; e.g.

$$\mathbb{R}^* = \{a \in \mathbb{R} \mid a \neq 0\}.$$

We found that $\mathbb{R}^*$ is group with respect to multiplication. In Example 5.2.4 we tried to do the same with $\mathbb{Z}$ and failed to obtain a group. Note that $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{C}$ are rings and so is $\mathbb{Z}$. Given a ring, is there a naturally defined subset that is a group with respect to multiplication? It turns out that the answer is yes, and that for $\mathbb{R}$, $\mathbb{Q}$ and $\mathbb{C}$ we obtain $\mathbb{R}^*$, $\mathbb{Q}^*$, $\mathbb{C}^*$ as we'd expect. To define this subset, we need the concept of a unit.

17.1.1 Definition Let R be a ring. An element u is called a *unit* if there is some element v in R such that $uv = vu = 1$. In other words, an element u of R is a unit if it has a multiplicative inverse that belongs to R.

17.1.2 Example In any non-zero ring, 0 is a non-unit. ◊

17.1.3 Example In $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{C}$, every non-zero element has a multiplicative inverse. So the units are the non-zero elements. ◊

17.1.4 Example What are the units in $\mathbb{Z}$? Suppose u is a unit in $\mathbb{Z}$. Then there is some $v \in \mathbb{Z}$ such that $uv = vu = 1$. This means that $1/u$ is an integer. The only integers u such that $1/u$ is also an integer are ± 1. So the units in $\mathbb{Z}$ are ± 1. ◊

17.1.5 Example Recall that $\mathbb{R}[x]$ is the ring of polynomials with real coefficients. Then x is not a unit, since $1/x$ is not a polynomial. However, 2 is a unit, since $1/2$ is a polynomial in $\mathbb{R}[x]$ with real coefficients:

$$\frac{1}{2} = \frac{1}{2} + 0x.$$

◊

We can now answer the question posed above.

17.1.6 Definition Let R be a ring. We define the *unit group of* R to be the set [15]

$$R^* = \{a \in R \mid a \text{ is a unit in } R\}. \tag{26}$$

Just because we've called R^* the *unit group of* R doesn't get us out of checking that it is really a group.

17.1.7 Lemma Let $(R, +, \cdot)$ be a ring and let R^* be the subset defined in (26). Then $(R^*, \cdot)$ is a group.

Proof. We must first show that R^* is closed under multiplication. Suppose u_1, $u_2 \in R^*$. Thus u_1, u_2 are units of R, and so there are v_1, $v_2 \in R$ such that

$$u_1 v_1 = v_1 u_1 = 1, \qquad u_2 v_2 = v_2 u_2 = 1. \tag{27}$$

We want to show that $u_1 u_2$ is a unit. Note that $v_2 v_1 \in R$ since R is closed under multiplication (it's a ring after all). Moreover,

$$\begin{aligned}(u_1 u_2)(v_2 v_1) &= u_1(u_2 v_2)v_1 \qquad \text{associativity of multiplication}\\ &= u_1 \cdot 1 \cdot v_1 \qquad \text{since } u_2 v_2 = 1\\ &= 1 \qquad \text{since } u_1 v_1 = 1.\end{aligned}$$

Similarly $(v_2 v_1)(u_1 u_2) = 1$. Thus $u_1 u_2$ is a unit [16] in R, and so $u_1 u_2 \in R^*$. We've proved that R^* is closed under multiplication.

We want to show that multiplication is associative in R^*. But multiplication is associative in R since R is a ring. Therefore it is associative in R^*.

Since $1 \cdot 1 = 1$, 1 is a unit and so $1 \in R^*$.

Finally we want to show that every element in R^* has a multiplicative inverse that belongs to R^*. Suppose $u \in R^*$. Then $uv = vu = 1$ for some $v \in R$. Note that this makes v also a unit, and so $v \in R^*$. Thus u has a multiplicative inverse in R^*. This completes the proof that R^* is a group.

[15] Some mathematicians write $R^\times$ instead of R^*.

[16] Start again. We have u_1, u_2 are units and so satisfy (27) for some v_1, v_2 in R. We want to show that $u_1 u_2$ is a unit. **What is wrong with the following argument?**

$$(u_1 u_2)(v_1 v_2) = (u_1 v_1)(u_2 v_2) = 1 \cdot 1 = 1.$$

Similarly $(v_1 v_2)(u_1 u_2) = 1$. Thus $u_1 u_2$ is a unit.

17.1.8 Example Note that $\mathbb{R}^*$, $\mathbb{C}^*$, $\mathbb{Q}^*$ have exactly the same meaning as before. ♢

17.1.9 Example We showed that the units of $\mathbb{Z}$ are ± 1. Therefore the unit group of $\mathbb{Z}$ is

$$\mathbb{Z}^* = \{1, -1\}.$$

♢

17.1.10 Example Recall that $M_{2\times 2}(\mathbb{R})$ is the ring of 2×2 matrices with real entries. It is clear from the definition of a unit, that the units of $M_{2\times 2}(\mathbb{R})$ are the invertible matrices. In other words, they are the ones having non-zero determinant. Thus

$$(M_{2\times 2}(\mathbb{R}))^* = \mathrm{GL}_2(\mathbb{R}).$$

Similarly,

$$(M_{2\times 2}(\mathbb{Q}))^* = \mathrm{GL}_2(\mathbb{Q}), \qquad (M_{2\times 2}(\mathbb{C}))^* = \mathrm{GL}_2(\mathbb{C}).$$

What about the unit group of $M_{2\times 2}(\mathbb{Z})$? This is more complicated. For example, consider the matrix $A = \left(\begin{smallmatrix} 3 & 1 \\ 1 & 1 \end{smallmatrix}\right)$. The matrix A is invertible, and $A^{-1} = \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 3/2 \end{pmatrix}$. Although A is in $M_{2\times 2}(\mathbb{Z})$, its inverse is not in $M_{2\times 2}(\mathbb{Z})$, but it is in $M_{2\times 2}(\mathbb{Q})$ and $M_{2\times 2}(\mathbb{R})$. Thus A is a unit in $M_{2\times 2}(\mathbb{Q})$, and $M_{2\times 2}(\mathbb{R})$ but not in $M_{2\times 2}(\mathbb{Z})$. The problem is clear: when calculating the inverse of a matrix, we must divide by its determinant, and the result does not have to be an integer.

Let's go back to the definition of a unit. Suppose $A \in M_{2\times 2}(\mathbb{Z})$ is a unit. Then there is a matrix $B \in M_{2\times 2}(\mathbb{Z})$ such that

$$AB = BA = I_2.$$

Taking determinants, are recalling that $\det(AB) = \det(A)\det(B)$ we find that

$$\det(A)\det(B) = 1.$$

Now $\det(A)$ and $\det(B)$ are integers because A and B have integer entries. Thus

$$\det(A) = \det(B) = 1, \qquad \text{or} \qquad \det(A) = \det(B) = -1.$$

Conversely if $A \in M_{2\times 2}(\mathbb{Z})$ has determinant ± 1, then its inverse will have integer entries and so A is a unit. We deduce that

$$(M_{2\times 2}(\mathbb{Z}))^* = \{A \in M_{2\times 2}(\mathbb{Z}) \mid \det(A) = \pm 1\}.$$

We define the group $\mathrm{GL}_2(\mathbb{Z})$ by

$$\mathrm{GL}_2(\mathbb{Z}) = \{A \in M_{2\times 2}(\mathbb{Z}) \mid \det(A) = \pm 1\}\,;$$

then $(M_{2\times 2}(\mathbb{Z}))^* = \mathrm{GL}_2(\mathbb{Z})$. In fact, for a *commutative* ring R we define

$$\mathrm{GL}_2(R) = \{A \in M_{2\times 2}(R) \mid \det(A) \in R^*\}\,.$$

You will easily see that this is consistent with the earlier definitions of $\mathrm{GL}_2(\mathbb{R})$, $\mathrm{GL}_2(\mathbb{C})$, $\mathrm{GL}_2(\mathbb{Q})$ and $\mathrm{GL}_2(\mathbb{Z})$, and that moreover, $(M_{2\times 2}(R))^* = \mathrm{GL}_2(R)$.

17.1.11 Example Let

$$S = \left\{ \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \mid a, b, c \in \mathbb{Z} \right\}.$$

We'll show that that S is a ring under the usual addition and multiplication of matrices abd then find S^*.

To show that S is a ring it is enough to show that it is a subring of $M_{2\times 2}(\mathbb{Z})$. We leave that as an exercise.

Let us compute the unit group. Suppose $A = \left(\begin{smallmatrix} a & b \\ 0 & c \end{smallmatrix}\right)$ is in S. To be unit it is not enough for this matrix to be invertible, we also want the inverse to belong to S. So we require the determinant ac to be non-zero and we want

$$A^{-1} = \frac{1}{ac}\begin{pmatrix} c & -b \\ 0 & a \end{pmatrix} = \begin{pmatrix} 1/a & -b/ac \\ 0 & 1/c \end{pmatrix}$$

to belong to S. Thus we want the integers a, b, c to satisfy

$$ac \neq 0, \qquad \frac{1}{a}, \frac{1}{c}, -\frac{b}{ac} \in \mathbb{Z}.$$

This happens precisely when $a = \pm 1$ and $c = \pm 1$. Thus

$$S^* = \left\{ \begin{pmatrix} \pm 1 & b \\ 0 & \pm 1 \end{pmatrix} : b \in \mathbb{Z} \right\}.$$

◊

17.1.12 Exercise In Example 16.3.9, we showed that

$$S = \left\{ \frac{a}{2^r} : a,\, r \in \mathbb{Z},\, r \geq 0 \right\}$$

is a ring. Find its unit group.

17.2 The Unit Group of the Gaussian Integers

The Gaussian integers $\mathbb{Z}[i]$ resemble the usual integers $\mathbb{Z}$ in many ways. For example, you know that every non-zero integer can be written as $\pm 1 \cdot p_1^{r_1} \dots p_n^{r_n}$ where the p_i are distinct primes, and this representation is unique (up to reordering the primes). This is the *Unique Factorization Theorem.* The Gaussian integers have their own *Unique Factorization Theorem*, which we don't have time to cover, but you can find out about this in *Algebra II.*

For now, we want to determine the unit group of $\mathbb{Z}[i]$. The most elegant way of doing this is via the *norm map.* We define the norm map $N : \mathbb{Z}[i] \to \mathbb{Z}$ by

$$N(a+bi) = a^2 + b^2, \qquad a, b \in \mathbb{Z}.$$

The norm map is multiplicative:

17.2.1 Lemma Let α, $\beta \in \mathbb{Z}[i]$. Then $N(\alpha\beta) = N(\alpha)N(\beta)$.

Proof. α and β are complex numbers, and you can see that $N(\alpha) = |\alpha|^2$. From the properties of the absolute value you know that $|\alpha\beta| = |\alpha| \cdot |\beta|$. The lemma follows. ♢

17.2.2 Theorem The unit group of $\mathbb{Z}[i]$ is $\{1, -1, i, -i\}$.

In other words, $(\mathbb{Z}[i])^* = U_4$, the group of fourth-roots of unity.

Proof. We want the units of $\mathbb{Z}[i]$. Let α be a unit. Then there is some $\beta \in \mathbb{Z}[i]$ such that [17] $\alpha\beta = 1$. Applying the norm map, and recalling that it is multiplicative, we see that

$$N(\alpha)N(\beta) = N(\alpha\beta) = N(1) = 1.$$

Now $N(\alpha)$ and $N(\beta)$ are in $\mathbb{Z}$ (go back to the definition of the norm map to see this), and they multiply to give 1. So

$$N(\alpha) = N(\beta) = 1, \qquad \text{or} \qquad N(\alpha) = N(\beta) = -1.$$

Write $\alpha = a + bi$ where a, b are in $\mathbb{Z}$. Then $a^2 + b^2 = N(\alpha) = \pm 1$. Of course -1 is impossible, so $a^2 + b^2 = 1$. But a, b are integers. So $(a, b) = (\pm 1, 0)$ or $(0, \pm 1)$. Hence $\alpha = a + bi = \pm 1$ or $\pm i$. Clearly ± 1, $\pm i$ are units. So the unit group is

$$\mathbb{Z}[i]^* = \{1, -1, i, -i\}.$$

[17] We could have written $\alpha\beta = \beta\alpha = 1$. But $\mathbb{Z}[i]$ is a *commutative* ring, so writing $\alpha\beta = 1$ is enough.

Remark. Compare the above proof to our determination of the unit group of $M_{2\times 2}(\mathbb{Z})$ in Example 17.1.10. I hope you agree that the similarities are striking!

17.2.3 Exercise In Exercise 16.3.10 you met the ring $\mathbb{Z}[2i]$. Find its unit group. (**Hint:** Show first that any unit in $\mathbb{Z}[2i]$ is a unit in $\mathbb{Z}[i]$.)

17.2.4 Exercise Let $\mathbb{Z}[\sqrt{2}] = \{a + b\sqrt{2} : a, b \in \mathbb{Z}\}$. Show that $\mathbb{Z}[\sqrt{2}]$ is a ring and that $1 + \sqrt{2}$ is a unit. What is its order as an element of the group $\mathbb{Z}[\sqrt{2}]^*$?

17.2.5 Exercise Let $\zeta = e^{2\pi i/3}$ (this is a cube root of unity). Check that $\overline{\zeta} = \zeta^2$. Let $\mathbb{Z}[\zeta] = \{a + b\zeta : a, b \in \mathbb{Z}\}$.

(i) Show that $\zeta^2 \in \mathbb{Z}[\zeta]$ (**Hint:** the sum of the cube roots of unity is ...).

(ii) Show that $\mathbb{Z}[\zeta]$ is a ring.

(iii) Show that ± 1, $\pm\zeta$ and $\pm\zeta^2$ are units in $\mathbb{Z}[\zeta]$.

(iv) (Harder) Show that $\mathbb{Z}[\zeta]^* = \{\pm 1, \pm\zeta, \pm\zeta^2\}$.

(v) Show that this group is cyclic.

17.3 Fields

A *field* $(F, +, \cdot)$ is a commutative ring which is not the zero ring such that every non-zero element is a unit. Thus a commutative ring F is a field if and only if its unit group is

$$F^* = \{a \in F \mid a \neq 0\}.$$

17.3.1 Example $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Q}$ are fields. ♢

17.3.2 Example $\mathbb{Z}$ is not a field, since for example $2 \in \mathbb{Z}$ is non-zero but not a unit. ♢

17.3.3 Example $\mathbb{R}[x]$ is not a field, since for example $x \in \mathbb{R}[x]$ is non-zero but not a unit. ♢

17.3.4 Example

$$\mathbb{Q}[i] = \{a + bi \mid a, b \in \mathbb{Q}\}$$

is a field as follows.

First we have to show that $\mathbb{Q}[i]$ is a commutative ring. For this it is enough to show that $\mathbb{Q}[i]$ is a subring of $\mathbb{C}$. It is clearly a subset of $\mathbb{C}$ that contains 0 and 1. Suppose α, $\beta \in \mathbb{Q}[i]$. We want to show that $\alpha + \beta$, $\alpha\beta$, $-\alpha$ are all in $\mathbb{Q}[i]$. Write

$$\alpha = a + bi, \qquad \beta = c + di$$

where a, b, c, $d \in \mathbb{Q}$. Then

$$\alpha + \beta = (a + c) + (b + d)i.$$

Since $\mathbb{Q}$ is closed under addition, $a + c$ and $b + d \in \mathbb{Q}$. So $\alpha + \beta \in \mathbb{Q}[i]$. Similarly, check for yourself that $\alpha\beta$ and $-\alpha$ are in $\mathbb{Q}[i]$. Thus $\mathbb{Q}[i]$ is a subring of $\mathbb{C}$ and so a ring [18].

Finally we have to show that every non-zero element of $\mathbb{Q}[i]$ is a unit. Suppose α is a non-zero element of $\mathbb{Q}[i]$. We can write $\alpha = a + bi$ where a, $b \in \mathbb{Q}$,

[18] Arguably, we could've made the proof more transparent by writing

$$\alpha = \frac{r}{s} + \frac{u}{v}i, \qquad \beta = \frac{k}{\ell} + \frac{m}{n}i,$$

where r, s, u, v, k, ℓ, m, n are integers and s, v, ℓ, n are non-zero. This would've worked, but it's probably better to get used to thinking of rational numbers as numbers in their own right.

and not both zero. We want to show that existence of some $\beta \in \mathbb{Q}[i]$ such that $\alpha\beta = \beta\alpha = 1$. In other words, we want to show that $1/\alpha$ is in $\mathbb{Q}[i]$. But we know how to compute $1/\alpha$. Recall that to divide complex numbers we multiply the numerator and denominator by the conjugate of the denominator:

$$\begin{aligned}
\frac{1}{\alpha} &= \frac{1}{a+bi} \\
&= \frac{1}{a+bi} \cdot \frac{a-bi}{a-bi} \\
&= \frac{a-bi}{a^2+b^2} \\
&= \frac{a}{a^2+b^2} - \frac{b}{a^2+b^2}i.
\end{aligned}$$

As a, b are rationals, so are $a/(a^2+b^2)$ and $b/(a^2+b^2)$. So $1/\alpha$ is in $\mathbb{Q}[i]$. Therefore $\mathbb{Q}[i]$ is a field. ♢

17.3.5 Exercise Let $\mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$. Show that $\mathbb{Q}[\sqrt{2}]$ is a field.

17.3.6 Exercise Let

$$F = \left\{ \begin{pmatrix} a & b \\ -b & a \end{pmatrix} : a, b \in \mathbb{R} \right\}.$$

(a) Show that F is a field (under the usual addition and multiplication of matrices). (**Hint:** Begin by showing that F is a subring of $M_{2\times 2}(\mathbb{R})$. You need to also show that F is commutative and that every non-zero element has an inverse in F.)

(b) Let $\phi : F \to \mathbb{C}$ be given by $\phi\begin{pmatrix} a & b \\ -b & a \end{pmatrix} = a + bi$. Show that ϕ is a bijection that satisfies $\phi(A+B) = \phi(A) + \phi(B)$ and $\phi(AB) = \phi(A)\phi(B)$.

(c) Show that

$$F' = \left\{ \begin{pmatrix} a & b \\ -b & a \end{pmatrix} : a, b \in \mathbb{C} \right\}$$

is not a field.

Lecture 18 - Congruences Revisited

We saw that there are two binary operations defined on $\mathbb{Z}/m\mathbb{Z}$, addition and multiplication. These make $(\mathbb{Z}/m\mathbb{Z}, +, \cdot)$ a commutative ring, and $(\mathbb{Z}/m\mathbb{Z}, +)$ a cyclic group of order m. We want to know about the unit group of $\mathbb{Z}/m\mathbb{Z}$.

18.1 Units in $\mathbb{Z}/m\mathbb{Z}$

18.1.1 Example What are the unit groups of $\mathbb{Z}/m\mathbb{Z}$ for $m = 2, 3, 4, 5, 6$.
To work this out just look at the multiplication table for $\mathbb{Z}/6\mathbb{Z}$ in Example 7.2.1 .
You'll see that

$$(\mathbb{Z}/6\mathbb{Z})^* = \left\{\overline{1}, \overline{5}\right\}.$$

In the same way you'll find that

$$\begin{aligned} (\mathbb{Z}/2\mathbb{Z})^* &= \left\{\overline{1}\right\}, & (\mathbb{Z}/3\mathbb{Z})^* &= \left\{\overline{1}, \overline{2}\right\}, \\ (\mathbb{Z}/4\mathbb{Z})^* &= \left\{\overline{1}, \overline{3}\right\}, & (\mathbb{Z}/5\mathbb{Z})^* &= \left\{\overline{1}, \overline{2}, \overline{3}, \overline{4}\right\}. \end{aligned}$$

In particular, $\mathbb{Z}/2\mathbb{Z}$, $\mathbb{Z}/3\mathbb{Z}$ and $\mathbb{Z}/5\mathbb{Z}$ are fields and $\mathbb{Z}/4\mathbb{Z}$, $\mathbb{Z}/6\mathbb{Z}$ are not fields. Can you make a general guess as to which $\mathbb{Z}/m\mathbb{Z}$ are fields and which aren't? Can you prove your guess? ♢

18.1.2 Theorem Let $\overline{a} \in \mathbb{Z}/m\mathbb{Z}$. Then $\overline{a}$ is a unit in $\mathbb{Z}/m\mathbb{Z}$ if and only if $\gcd(a, m) = 1$. Thus

$$(\mathbb{Z}/m\mathbb{Z})^* = \{\overline{a} \mid 0 \leq a \leq m-1 \text{ and } \gcd(a, m) = 1\}.$$

Proof. Suppose $\overline{a}$ is a unit in $\mathbb{Z}/m\mathbb{Z}$. Then there is some $\overline{b}$ in $\mathbb{Z}/m\mathbb{Z}$ so that $ab \equiv 1 \pmod{m}$. Thus, there is some $k \in \mathbb{Z}$ such that $ab - 1 = km$. Write $g = \gcd(a, m)$. Then $g \mid a$ and $g \mid m$. So $g \mid (ab - km) = 1$. But this means that $g = 1$.

Conversely, suppose $\gcd(a, m) = 1$. By Bezout's Lemma (see *Foundations/Sets and Numbers*) we know that we can write $1 = ba + cm$ for some integers b, $c \in \mathbb{Z}$. Thus $ab \equiv 1 \pmod{m}$. Hence $\overline{a}$ is a unit.

18.1.3 Exercise Redo Example 18.1.1 using Theorem 18.1.2.

18.1.4 Example By Theorem 18.1.2, we know that $\overline{19}$ is invertible in $\mathbb{Z}/256\mathbb{Z}$. But the statement of the theorem does not tell us how to find the inverse. It would take us a very long to run through the elements $\overline{u} \in \mathbb{Z}/256\mathbb{Z}$ and check to see if $19u \equiv 1 \pmod{256}$. However, **the proof of the theorem does give us**

a recipe for finding the inverse. We know by factoring that $\gcd(19, 256) = 1$, but let's use Euclid's Algorithm [19] to write 1 as a linear combination of 19 and 256:

$$\begin{aligned} \mathbf{256} &= 13 \times \mathbf{19} + \mathbf{9} \\ \mathbf{19} &= 2 \times \mathbf{9} + \mathbf{1}. \end{aligned}$$

Thus

$$\mathbf{1} = \mathbf{19} - 2 \times \mathbf{9} = \mathbf{19} - 2 \times (\mathbf{256} - 13 \times \mathbf{19}) = (1 - 2 \times -13) \times \mathbf{19} - 2 \times \mathbf{256},$$

so

$$\mathbf{1} = 27 \times \mathbf{19} - 2 \times \mathbf{256}.$$

Hence $27 \times 19 \equiv 1 \pmod{256}$, so $\overline{27}$ is the inverse of $\overline{19}$ in $\mathbb{Z}/256\mathbb{Z}$. $\Diamond$

18.2 Fermat's Little Theorem

Through the computations you've done so far, you've probably conjectured the following.

18.2.1 Theorem Let p be a prime. Then $\mathbb{Z}/p\mathbb{Z}$ is a field. Therefore,

$$(\mathbb{Z}/p\mathbb{Z})^* = \{\overline{1}, \overline{2}, \ldots, \overline{p-1}\}.$$

Proof. We already know that $\mathbb{Z}/m\mathbb{Z}$ is a commutative ring for any integer $m \geq 2$. Now to show that $\mathbb{Z}/p\mathbb{Z}$ is a field, we must show that any non-zero $\overline{a} \in \mathbb{Z}/p\mathbb{Z}$ is invertible. But if $\overline{a} \in \mathbb{Z}/p\mathbb{Z}$ is non-zero, then a is one of $1, 2, \ldots, p-1$. Clearly a is not divisible by p. Since p is prime, $\gcd(a, p) = 1$. Hence by Theorem 18.1.2, $\overline{a}$ is invertible in $\mathbb{Z}/p\mathbb{Z}$. This shows that $\mathbb{Z}/p\mathbb{Z}$ is a field.

18.2.2 Exercise Prove the converse of Theorem 18.2.1: if $\mathbb{Z}/m\mathbb{Z}$ is a field then m is prime.

18.2.3 Theorem (Fermat's Little Theorem) Let p be a prime and a an integer such that $p \nmid a$. Then

$$a^{p-1} \equiv 1 \pmod{p}. \tag{28}$$

[19] It is easy to get muddled in the substitutions involved in Euclid's Algorithm. One way to reduce the muddle is to somehow distinguish the numbers you started with, here 256 and 19, and the remainders from the quotients. I did the distinguishing by writing the numbers we started with and the remainders in boldtype. In your calculations, you can underline them.

Proof. We know that $a \equiv b \pmod{p}$ where b is one of $0, 1, 2, \ldots, p-1$. Now as $p \nmid a$, we see that $b \neq 0$. By Theorem 18.2.1, $\overline{b}$ is in the unit group of $\mathbb{Z}/p\mathbb{Z}$ which is

$$(\mathbb{Z}/p\mathbb{Z})^* = \{\overline{1}, \overline{2}, \ldots, \overline{p-1}\}.$$

The order of the group $(\mathbb{Z}/p\mathbb{Z})^*$ is clearly $p-1$. By Corollary 8.2.3 (the corollary to Lagrange's Theorem),

$$\overline{b}^{p-1} = 1.$$

Thus $b^{p-1} \equiv 1 \pmod{p}$. Since $a \equiv b \pmod{p}$, we obtain (28). ♢

Here's a fun application of Fermat's Little Theorem.

18.2.4 Example Here we'll compute $2^{1000} \pmod{13}$.

Since 13 is prime and $13 \nmid 2$, we know by Fermat's Little Theorem that $2^{12} \equiv 1 \pmod{13}$. Now by the Division Algorithm,

$$1000 = 83 \times 12 + 4.$$

Therefore,

$$2^{1000} = 2^{83 \times 12 + 4} = (2^{12})^{83} \times 2^4 \equiv 1^{83} \times 16 \equiv 3 \pmod{13}.$$

♢

18.3 Euler's Theorem

18.3.1 Definition Let $m \geq 1$. We denote the order of the group $(\mathbb{Z}/m\mathbb{Z})^*$ by $\varphi(m)$. The function φ is called *Euler's φ-function.*

18.3.2 Example We know that if p is a prime, then $(\mathbb{Z}/p\mathbb{Z})^* = \{\overline{1}, \overline{2}, \ldots, \overline{p-1}\}$, and so $\varphi(p) = p - 1$. ♢

18.3.3 Example We know that

$$(\mathbb{Z}/6\mathbb{Z})^* = \{\overline{1}, \overline{5}\},$$

and so $\varphi(6) = 2$. ♢

18.3.4 Example Let $n \geq 1$. Then $(\mathbb{Z}/2^n\mathbb{Z})^*$ consists of $\overline{a}$ with a in the range $0 \leq a \leq 2^n - 1$ that are coprime to 2^n. These are the odd integers a in the range $0 \leq a \leq 2^n - 1$. Thus

$$(\mathbb{Z}/2^n\mathbb{Z})^* = \{\overline{1}, \overline{3}, \dots, \overline{2^n - 1}\}.$$

Hence $\varphi(2^n) = 2^{n-1}$. $\Diamond$

18.3.5 Theorem (Euler's Theorem) Let m be an integer satisfying $m \geq 2$. Let a be an integer such that $\gcd(a, m) = 1$. Then

$$a^{\varphi(m)} \equiv 1 \pmod{m}.$$

Proof. This has almost the same proof as Fermat's Little Theorem. I'll leave the necessary modifications as an exercise.

You're probably wondering if there is a formula for $\varphi(m)$, and in fact there is.

18.3.6 Proposition Write

$$m = p_1^{r_1} \cdots p_k^{r_k}$$

where $p_1, \dots, p_k$ are distinct primes and $r_1, \dots, r_k$ are positive integers. Then

$$\varphi(m) = (p_1^{r_1} - p_1^{r_1 - 1}) \cdots (p_k^{r_k} - p_k^{r_k - 1}).$$

The proof of Proposition 18.3.6 is a little long and we'll not include it in these notes. I'll post a video of this but it's non-examinable.

18.3.7 Exercise Use Euler's Theorem to compute $2^{1000} \pmod{63}$.

18.3.8 Exercise It is known that $(\mathbb{Z}/m\mathbb{Z})^*$ is cyclic if $m = 2$, 4, p^a or $2p^a$ where p is an odd prime. For all other $m \geq 2$, the unit group $(\mathbb{Z}/m\mathbb{Z})^*$ is not cyclic. For more on this, do *Number Theory* in term 3. For now, check that $(\mathbb{Z}/7\mathbb{Z})^*$ is cyclic, but $(\mathbb{Z}/8\mathbb{Z})^*$ is not cyclic.

18.3.9 Exercise Use Lagrange's Theorem to show that $\varphi(m)$ is even for $m \geq 3$.