MA136

Introduction to Abstract Algebra

Samir Siksek

Mathematics Institute
University of Warwick

Contents

CHAPTER I

Prologue

I.1. Who Am I?

I Samir Siksek have the immense pleasure of introducing you to three heroes of abstract algebra: groups, rings and fields. I am not an algebraist, but I have nothing but love, admiration and enthusiasm for the subject. Some of my best friends are algebraists.

I.2. A Jolly Good Read!

Abstract algebra is about patterns. You see one pattern repeating itself across mathematics and you try to extract the essential elements of that pattern and turn them into a definition. This process gives you groups, rings, fields, vector spaces, etc. You then study each of these new algebraic objects and become familiar with it. After that, when you spot one of these patterns in a new context, you'll say 'Aha! I know what that is, and what to do with it'.

three tips

Abstract algebra is incredibly useful, but to get any benefit from it you need to develop three essential habits:

(i) Study as many different examples as you can. The examples are as important as the theorems and definitions. There is absolutely no use in knowing the definition of a group if you're not familiar with the standard examples.

(2) Do calculations. Use calculations with matrices, permutations, symmetries, etc. to test your ideas. Calculations will lead you to counterexamples that can correct any erroneous ideas that you have. But also with practice, you will find that calculations often contain the germ of the proof you're looking for.

(c) Think geometrically and draw pictures. The true meaning of most mathematical concepts is geometric. If you spend all your time manipulating symbols (i.e. doing algebra) without understanding the relation to the geometric meaning, then you will have very little in terms of mathematical insight.

The three habits will not only help you learn the subject and apply it, you will develop great mathematical taste.

I.3. Proofs

When I was a student I found it very hard to follow proofs in books and lectures. So when I read a theorem, I would put down the book and try out a few examples. After that I would try to prove the theorem myself. After I finished (or if I failed) I would look at the proof in the book and compare. I heartily recommend this strategy. You'll gain a great understanding of the subject. You'll also get really good practice for the exam, where you may asked to prove statements that you haven't seen before.

I.4. Acknowledgements and Corrections

I thank Jonathan Addison, Alex Best, George Christofi, Jenny Cooley, John Cremona, Edward Day, Harry Graham, Darij Grinberg, Christian Fieldhouse, Giles Hutchings, Roderick Mansel, Dave McCormick, Joseph Miller, Xiao Lin, Ghaleo Tsoi Kwok-Wing, Joe O'Sullivan, James Soffe and Esther Turner for suggesting corrections to previous versions of these notes.

Please email me your comments, misprints and corrections. My address is `samir.siksek@gmail.com`.

Are previous exam papers available? Previous exam papers are available from the module page in the undergraduate handbook, together with the solutions.

Are we required to know the proofs taken during the lectures or found in the lecture notes? Yes, theorems, definitions, proofs and homework questions. I love bookwork.

Can questions of a similar type to part C appear in the exam? Yes they can! The exam will NOT be the academic equivalent of a chainsaw massacre, but certainly some interesting questions must appear, otherwise your maths degree would be trivial.

The exam is tomorrow/next week/within six months. I'm running around like a headless chicken and stressing all my friends because I can't do a homework question. Can I knock on your door and ask you about it? Don't worry, I've already branched out into agony-aunting. Yes come and ask; I promise not to set the dogs on you.

After Warwick I plan to devote my life to drunkenness and anti-social behaviour. Whilst I'm here, I want to enjoy mathematics to the full. Pleeeeeease set us obscene amounts of homework? *We must be careful.* If you do too much homework, you'll suffer severe withdrawal symptoms once the term is over, and there's no telling what you might do to yourself. I simply can't have that on my conscience. I'll therefore limit the homework to one sheet per week. It cuts me deep to be so hard on you, but sometimes you have to be tough to be kind.

I can't get hold of a pitchfork, and I'm worried that a torch would set off the fire alarm. How can I make constructive criticisms? Constructive criticisms are welcome, face-to-face or by email.

CHAPTER II

FAQ

Why is this FAQ upside down? This is to improve the chances that you will notice it and read it.

Your lectures are excruciatingly boring. Besides, 12 noon is a perversely early time to schedule a lecture and no self-respecting student can be expected to be awake yet. Do I really have to attend your lectures, or can I make do with these lecture notes? I'll take that as an endorsement of the greatness of my writing skills rather than a criticism of my lecturing skills. I love improvisation during lectures. So the material we cover in the lectures will not be identical to that in these printed notes. You need to come to the lectures and make your own notes. *The exam will be based on the contents of the lectures as well as these printed lecture notes.*

I was ill and missed a homework [2] **deadline.** These matters are handled by the Undergraduate Office.

How is this course assessed? 15% for four homework assignments and 85% for a one hour exam in term 3.

My copy of homework assignment x is lost/stolen. Where can I get another copy? You can get all the assignments from the course page in the undergraduate handbook.

The questions on the homework sheets are divided into parts A, B, C. Are the questions in part C optional? You are asked to hand-in both parts A and B but not part C. Part C questions should be attempted by students who hope to obtain a First or a II:1. Students who hope to get at II:2 or a Third should avoid attempting part C questions at all cost.

Do you subscribe to the illustrious Warwick tradition of setting the same exam every year? You're paying £9000 a year. You hardly expect me to rip you off with a second-hand exam? Come on, how low do you think I am?

[2] **home·work** /ˈhəʊmwɜːk/ *Noun.* A mandatory regular course of mental stimulation designed to vanquish intellectual impotence.

CHAPTER III

Algebraic Reorientation

III.1. Sets

Sets are a basic notation for most of modern pure mathematics, but life is too short to spend too much time on them. A *set* is simply a collection of objects. We use curly brackets to denote sets. For example, if I write

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

then I'm 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

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

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

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

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

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

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

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

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

then $C = \{4\}$.

If we write

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

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:

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

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

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

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

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

Here are some other sets that you're meant to know:

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

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

Examples of elements of $\mathbb{Q}$ are 0, 5, $-7/11$, $3/2$, $6/4$ (the last two being the same element). From *Foundations* you should 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

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

Tip

While we're on the subject of notation, compare the following two statements:

- Some positive real numbers are irrational.
- $\exists x \in \mathbb{R}$ s.t. $x > 0 \ \wedge \ x \notin \mathbb{Q}$.

The two statements say exactly the same thing. A professional mathematician prefers the first, and an amateur prefers the second. Use only as much mathematical notation as needed to make your ideas transparent and precise, but no more [1].

III.2. Binary Operations

Let S be a set. A *binary operation* on S is a rule which for every two elements of S gives another element of S. For example, 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.

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

[1] Coming soon to a supervision area near you: $\boxplus$, $\cup$, and $\sqcup\!\!\sqcup$.

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.*

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

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

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 $5+(-5)=0 \notin \mathbb{R}^*$.

III.3. Vector Operations

We define *Euclidean n-space* as

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

Thus $\mathbb{R}^2$ is the set of vectors in the plane, and $\mathbb{R}^3$ is the set of vectors in 3-space. Addition is a binary operation on $\mathbb{R}^n$, and so is subtraction. What about multiplication by a scalar? 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

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

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 and the cross product? The dot product is defined on $\mathbb{R}^n$ for all n. If $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ 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. The cross product is defined on $\mathbb{R}^3$ only. If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$ the $\mathbf{x} \times \mathbf{y}$ is again in $\mathbb{R}^3$. So the cross product is a binary operation on $\mathbb{R}^3$.

III.4. Operations on Polynomials

We shall write $\mathbb{R}[x]$ for the set of polynomials in x with real coefficients, $\mathbb{C}[x]$ for the set of polynomials in x with complex coefficients, $\mathbb{Q}[x]$ for the set of polynomials in x with rational coefficients, and $\mathbb{Z}[x]$ for the set of polynomials in x with integer coefficients. All these are closed under addition, multiplication and subtraction, but not division; for example $x/(x+1)$ is not a polynomial.

III.5. Composition of Functions

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.

Example III.1. Let

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

and

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

Then

$$(f \circ g)(x) = f(g(x)) = f(3x+2) = (3x+2)^2 - 5 = 9x^2 + 12x - 1,$$
$$(g \circ f)(x) = g(f(x)) = g(x^2 - 5) = 3(x^2 - 5) + 2 = 3x^2 - 13.$$

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. ♢

Note that in the example we started with functions $\mathbb{R} \to \mathbb{R}$ and composed to obtain functions $\mathbb{R} \to \mathbb{R}$. Likewise, in the above definition, 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 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.

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

Lemma III.2. *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. To stop ourself from getting muddled, let $k = g \circ h$ and $\ell = f \circ g$. Note that $k(x) = g(h(x))$ and $\ell(x) = f(g(x))$. So

$$(f \circ (g \circ h))(x) = (f \circ k)(x) = f(k(x)) = f(g(h(x))).$$

Also

$$((f \circ g) \circ h)(x) = (\ell \circ h)(x) = \ell(h(x)) = f(g(h(x))).$$

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

III.6. Composition 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 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 table* which tells us for any pair of elements of S what the third element is.

Example III.3. Let $S = \{a, b, c\}$. Let $\circ$ be the binary operation on S with the following composition table:

$\circ$	a	b	c
a	b	c	a
b	a	c	a
c	b	b	c

The result of the composition $a \circ b$, 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,

$$a \circ b = c, \qquad b \circ a = a, \qquad c \circ b = b, \qquad a \circ a = b, \ldots.$$

You might think that this example is somewhat contrived, and you're absolutely right. But later on we'll meet more natural composition tables that arise from studying groups, permutations, etc. $\Diamond$

III.7. Commutativity and Associativity

Definition. Let S be a set and $\circ$ a binary operation [1] on S. We say that the binary operation $\circ$ is *commutative on* S if $a \circ b = b \circ a$ for all $a, b \in S$. We say that the binary operation $\circ$ is *associative on* S if $(a \circ b) \circ c = a \circ (b \circ c)$ for all $a, b, c \in S$.

Example III.4. Addition and multiplication on $\mathbb{R}$ (or $\mathbb{C}$ or $\mathbb{R}[x]$ or $\ldots$) are both commutative and associative. When operations are commutative and associative, order and bracketing do not matter:

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

Of course subtraction is neither commutative nor associative (write some examples). $\Diamond$

Example III.5. 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

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

[1] Here $\circ$ doesn't have to be composition of functions. Simply any binary operation on any set.

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

Example III.6. Let $S = \{a, b, c\}$ and let $\circ$ be the binary operation given by the composition table in Example III.3. Then $\circ$ is not commutative; for example

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

It is also not associative; for example

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

◊

Example III.7. Composition of functions from a set A to itself is associative but not commutative. We know that it is associative from Lemma III.2. We know that it isn't commutative by Example III.1. When a binary operation is associative bracketing doesn't matter. For example,

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

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

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

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 ◊

Example III.8. 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 $\circ$ be the binary operation on S with the following composition table:

$\circ$	a	b	c
a	b	c	a
b	c	c	a
c	a	a	c

Note that $\circ$ 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,

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

◊

Exercise III.9. 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 $a \circ b = a/b$.
(b) $A = \{1, 2, 3, 4, \ldots\}$ is the set of positive integers and $a \circ b = a^b$.

(c) $A = \{\ldots, 1/8, 1/4, 1/2, 1, 2, 4, 8, \ldots\}$ is the set of powers of 2 and $a \circ b = ab$.
(d) $A = \mathbb{C}$ is the set of complex numbers and $a \circ b = |a - b|$.

III.8. Where are the Proofs?

You might be somewhat perturbed by the cavalier way I'm stating things without proving them. In mathematics we have to start with some assumptions (sometimes called *axioms*) and then prove things from there. A reasonable starting point is the properties of the real numbers. These we assume. What are they?

For all real numbers a, b, c

(i) $a + b = b + a$ (addition is commutative)
(ii) $(a + b) + c = a + (b + c)$ (addition is associative)
(iii) $a + 0 = a$ (0 is the additive identity element)
(iv) there is a real number $-a$ (the additive inverse of a) such that $a + (-a) = 0$.
(v) $ab = ba$ (multiplication is commutative)
(vi) $(ab)c = a(bc)$ (multiplication is associative)
(vii) $a(b + c) = ab + ac$ (multiplication distributes over addition)
(viii) $a \cdot 1 = a$ (1 is the multiplicative identity element)
(ix) if $a \neq 0$, there is a real number denoted by a^{-1} (the multiplicative inverse of a) such that $a \cdot a^{-1} = 1$.

We have not exhausted the properties of real numbers. For example, we can add

(x) If $a \geq b$ then $a + c \geq b + c$.
(xi) If $a \geq b$ and $c > 0$ then $ac \geq bc$. If $a \geq b$ and $c < 0$ then $ac \leq bc$.

One particularly important property that we will not write down, but which you will come to admire in the analysis courses is 'The Completeness Axiom'.

These properties are a reasonable starting point. We should be able to prove all the facts that we have been stating starting from here. For example, let us prove that multiplication of complex numbers is commutative. In other words, we want to show that if α and β are complex numbers then $\alpha\beta = \beta\alpha$. So suppose that α and β are complex numbers. Write $\alpha = a + bi$ and $\beta = c + di$ where a, b, c, d are real numbers. Then by the definition of multiplication

$$\alpha\beta = (ac - bd) + (ad + bc)i, \qquad \beta\alpha = (ca - db) + (da + cb)i.$$

But $ac = ca$, $bd = db$, $ad = da$, $bc = cb$. How do we know this; isn't this the same as what we want to prove? No, not really. We know this because a, b, c, d are real numbers and we are using the commutativity of multiplication for real numbers which we already decided to assume. It follows that $\alpha\beta = \beta\alpha$ which we wanted to prove.

Exercise III.10. You know that if $a, b \in \mathbb{R}$ and $ab = 0$ then either $a = 0$ or $b = 0$. Explain how this follows from property (ix) above.

III.9. The Quaternionic Number System (do not read)

This section is non-examinable; do not read it. It is here for the benefit of those who believe that the above discussion of commutativity of complex numbers is overly pedantic. "Why should multiplication not be commutative? After all, it is just multiplication. You are wasting time on contrived pedanticisms". For your benefit I will briefly exhibit the quaternionic number system where multiplication is not commutative. Quaternions were fashionable in the late 19th century and had substantial physical applications. Eventually it was discovered that vectors do a better job of just about anything you could do with quaternions, and they fell out of fashion.

Remember that the complex numbers are of the form $a + bi$ where a, b are real and i is a symbol satisfying $i^2 = -1$. Well, quaternions are of the form $a + bi + cj + dk$ where a, b, c, d are real and i, j, k are symbols satisfying

$$i^2 = j^2 = k^2 = -1, \qquad ij = -ji = k, \qquad jk = -kj = i, \qquad ki = -ik = j.$$

You can already see that quaternionic multiplication is not commutative, since $ij \neq ji$. You might also calculate $(1 + i)(1 + j)$ and $(1 + j)(1 + i)$.

Here is a standard description of the discovery of quaternions, which I've copied and pasted from Wikipedia:

> [Sir William Rowan] Hamilton knew that the complex numbers could be viewed as points in a plane, and he was looking for a way to do the same for points in space. Points in space can be represented by their coordinates, which are triples of numbers, and for many years Hamilton had known how to add and subtract triples of numbers. But he had been stuck on the problem of multiplication and division: He did not know how to take the quotient of two points in space.
>
> The breakthrough finally came on Monday 16 October 1843 in Dublin, when Hamilton was on his way to the Royal Irish Academy where he was going to preside at a council meeting. While walking along the towpath of the Royal Canal with his wife, the concept behind quaternions was taking shape in his mind. Hamilton could not resist the impulse to carve the formulae for the quaternions
>
> $$i^2 = j^2 = k^2 = ijk = -1$$
>
> into the stone of Brougham Bridge as he passed by it ... Since 1989, the Department of Mathematics of the

National University of Ireland, Maynooth has organized a pilgrimage, where scientists (including physicists Murray Gell-Mann in 2002, Steven Weinberg in 2005, and mathematician Andrew Wiles in 2003) take a walk from Dunsink Observatory to the Royal Canal bridge where, unfortunately, no trace of Hamilton's carving remains.

You see, even though the quaternions have been consigned to the compost heap of algebra, Hamilton's graffiti became history's most celebrated act of mathematical vandalism. *There is a great moral to this, but I can't find it.*

CHAPTER IV

Matrices—Read On Your Own

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. In this chapter I summarize what you need to know. We'll not cover this chapter in the lectures; I expect you to read it on your own. Even if you think you know all about matrices I advise you to read this chapter: do you know why matrix multiplication is defined the way it is? Do you know why matrix multiplication is associative?

IV.1. What are Matrices?

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

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

Example IV.1. 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 2×3 because it has 2 rows and 3 columns. Likewise B has size 3×2 and C has size 3×3. ◊

Displaying a matrix A by 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 $m \times n$ (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}.

Example IV.2. 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. ♢

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.

Example IV.3.

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

♢

IV.2. Matrix Operations

Definition. 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.

Example IV.4. 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

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

$\Diamond$

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 no confusion is feared.

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.

Example IV.5. 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**. $\Diamond$

Example IV.6. 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. $\Diamond$

Remark. 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 $M_{n\times n}(\mathbb{R})$.

Exercise IV.7. Commutativity—What can go wrong?

- 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 $AB = BA$.

IV.3. Where do matrices come from?

No doubt you have at some point wondered where matrices come from, and what is the reason for the weird definition of matrix multiplication. It is possible that your A-Level teachers didn't want to tell you. Because I am a really sporting kind of person and I love you very much, I am telling you some secrets of the trade. *I ♡ you*

Matrices originate from linear substitutions. Let a, b, c, d be fixed numbers, x, y some variables, and define x', y' by the linear substitutions

$$\begin{aligned} x' &= ax + by \\ y' &= cx + dy. \end{aligned} \tag{IV.1}$$

The definition of matrix multiplication allows us to express this pair of equations as one matrix equation

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{IV.2}$$

You should multiply out this matrix equation and see that it is the same as the pair of equations (IV.1).

Now suppose moreover that we define new quantities x'' and y'' by

$$\begin{aligned} x'' &= \alpha x' + \beta y' \\ y'' &= \gamma x' + \delta y', \end{aligned} \tag{IV.3}$$

where α, β, γ, δ are constants. Again we can rewrite this in matrix form as

$$\begin{pmatrix} x'' \\ y'' \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix}. \tag{IV.4}$$

What is the relation between the latest quantities x'', y'' and our first pair x, y? One way to get the answer is of course to substitute equations (IV.1) into (IV.3). This gives us

$$\begin{aligned} x'' &= (\alpha a + \beta c)x + (\alpha b + \beta d)y \\ y'' &= (\gamma a + \delta c)x + (\gamma b + \delta d)y. \end{aligned} \tag{IV.5}$$

This pair of equations can re-expressed in matrix form as

$$\begin{pmatrix} x'' \\ y'' \end{pmatrix} = \begin{pmatrix} \alpha a + \beta c & \alpha b + \beta d \\ \gamma a + \delta c & \gamma b + \delta d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{IV.6}$$

Another way to get x'', y'' in terms of x, y is to substitute matrix equation (IV.2) into matrix equation (IV.4):

$$\begin{pmatrix} x'' \\ y'' \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{IV.7}$$

If the definition of matrix multiplication is sensible, then we expect that matrix equations (IV.6) and (IV.7) to be consistent. In other words, we would want that

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \alpha a + \beta c & \alpha b + \beta d \\ \gamma a + \delta c & \gamma b + \delta d \end{pmatrix}.$$

A quick check using the definition of matrix multiplication shows that this is indeed the case.

IV.4. How to think about matrices?

Let $A \in M_{2\times 2}(\mathbb{R})$. In words, A is a 2×2 matrix with real entries. Write

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

For now, think of the elements of $\mathbb{R}^2$ as *column vectors*: any $\mathbf{u} \in \mathbb{R}^2$ can be written as

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

with x, y real numbers. Thus we're thinking of the elements of $\mathbb{R}^2$ as 2×1-matrices. Note that in equation (IV.2), the matrix A 'converts' the vector $\mathbf{u}$ to another vector $\mathbf{u}' = \begin{pmatrix} x' \\ y' \end{pmatrix}$.

Some mathematicians would think that the word 'converts' is not very mathematical. Instead they would think of the matrix A as defining a function

$$T_A : \mathbb{R}^2 \to \mathbb{R}^2, \qquad T_A(\mathbf{u}) = A\mathbf{u}.$$

Other (less pedantic) mathematicians would not distinguish between the matrix and the function it defines. One of the points of the previous section is that if $C = BA$ then $T_C = T_B \circ T_A$, so that matrix multiplication is really an instance of composition of functions.

Let us look at some examples of these functions T_A.

Example IV.8. 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.

Again, 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.

non-commutativity seen geometrically

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. Notice $(AB)\mathbf{u} = A(B\mathbf{u})$. This means stretch $\mathbf{u}$ by a factor of 2 in the x-direction, then reflect it in the line $y = x$. And $(BA)\mathbf{u} = B(A\mathbf{u})$, 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 IV.1. Therefore $AB \neq BA$. ♢

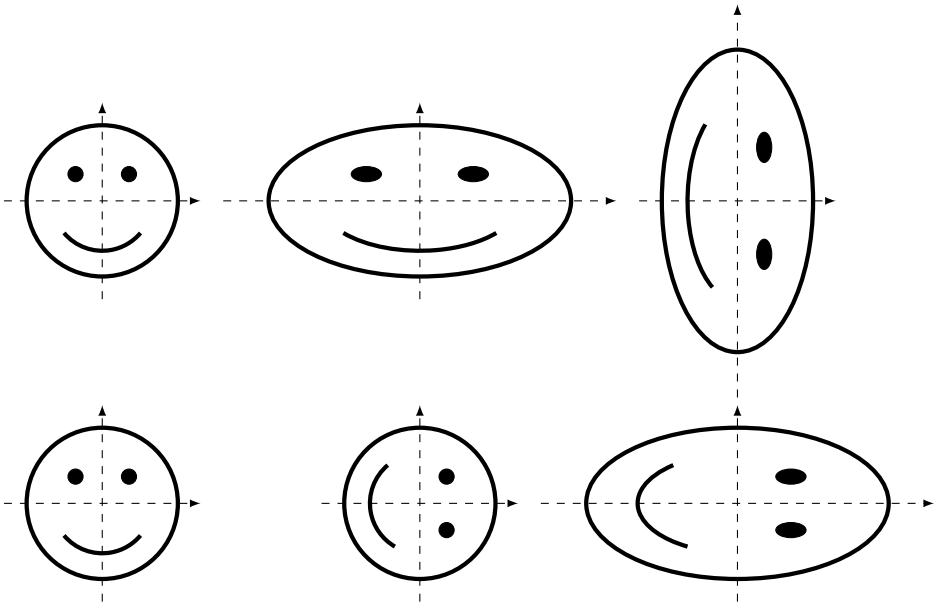


FIGURE IV.1. Non-commutativity of matrix multiplication. The matrix A represents reflection in the line $y = x$ and the matrix B represents stretching by a factor of 2 in the x-direction. On the top row we apply B first then A; the combined effect is represented by AB. On the bottom we apply A first 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 $AB \neq BA$.

Remark. Matrices don't give us all possible functions $\mathbb{R}^2 \to \mathbb{R}^2$. You will see 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? 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.

IV.4.1. Why is matrix multiplication associative? At the end of Example IV.8 there was a slight-of-hand that you might have noticed [1]. We assumed that matrix multiplication is associative when we wrote $(AB)\mathbf{u} = A(B\mathbf{u})$; here we're thinking of $\mathbf{u}$ as a 2×1-matrix. In fact matrix multiplication is associative whenever it is defined.

Theorem IV.9. *Let A be an $m \times n$ matrix, B be an $n \times p$ matrix and C a $p \times q$ matrix, then*

$$(AB)C = A(BC). \tag{IV.8}$$

PROOF. In terms of functions, (IV.8) is saying

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

This holds by Lemma III.2 [2]. □

Don't worry too much if this proof makes you uncomfortable! When you do *Linear Algebra* in Term 2 you will see a much more computational proof, but in my opinion the proof above is the most enlightening one. For now, you should be pleased if you have digested Example IV.8.

IV.5. Why Column Vectors?

You will have noticed that early on in these notes we were thinking of the elements of $\mathbb{R}^n$ as row vectors. But when we started talking about matrices as functions, we have taken a preference for column vectors as opposed to row vectors. Let us see how things are different if we stuck with row vectors. So for the moment think of elements of $\mathbb{R}^n$, $\mathbb{R}^m$ as row vectors. Let A be an $m \times n$ matrix.

[1] "well-done" if you did notice, and "learn to read more critically" if you haven't

[2] Here T_A, T_B, T_C respectively are functions $\mathbb{R}^n \to \mathbb{R}^m$, $\mathbb{R}^p \to \mathbb{R}^n$, $\mathbb{R}^q \to \mathbb{R}^p$.

If $\mathbf{u}$ is a (row) vector in $\mathbb{R}^n$ or $\mathbb{R}^m$ then $A\mathbf{u}$ is undefined. But we find that $\mathbf{u}A$ is defined if $\mathbf{u}$ is a (row) vector in $\mathbb{R}^m$ and gives a (row) vector in $\mathbb{R}^n$. Thus we get a function

$$S_A : \mathbb{R}^m \to \mathbb{R}^n$$

given by $S_A(\mathbf{u}) = \mathbf{u}A$. It is now a little harder to think of the matrix A as a function since we have written it on the right in the product $\mathbf{u}A$ (remember that when we thought of vectors as columns we wrote $A\mathbf{u}$).

Some mathematicians (particularly algebraists) write functions on the right, so instead of writing $f(x)$ they will write xf. They will be happy to think of matrices as functions on row vectors because they can write the matrix on the right [1]. Most mathematicians write functions on the left. They are happier to think of matrices as functions on column vectors because they can write the matrix on the left.

left versus right

Many of the abstract algebra textbooks you will see in the library write functions on the right. Don't be frightened by this! If you're uncomfortable with functions on the right, just translate by rewriting the theorems and examples in your notation.

IV.6. Multiplicative Identity and Multiplicative Inverse

We have mostly been focusing on 2×2 matrices, and we will continue to focus on them. One natural question to ask is what is the multiplicative identity for 2×2 matrices? You might be wondering what I mean 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, 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}? Does it even have an inverse? It is likely that you know the answers to these questions from school. If not don't worry, because we're about to discover the answers. If yes, please unremember the answers, because we want to work out the answers from scratch. We want to immerse ourselves in the thought process that went into discovering these answers.

temporary amnesia required

The first question is about the multiplicative identity. We haven't yet discovered what the multiplicative identity is, but let us denote it by I_2. What is the geometric meaning of I_2? Clearly we want I_2 to have the geometric meaning of 'do nothing', as opposed to reflect, stretch, project, etc. 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

[1] Algebraists have many idiosyncrasies that distinguish them from other mathematicians. I find most of these bewildering. However, they do have a very good point in the way they write functions. We said that BA means do A first and then B, because $(BA)\mathbf{u} = B(A\mathbf{u})$. However if you use row vectors then $\mathbf{u}(BA) = (\mathbf{u}B)A$, so BA means do B first then A, which seems entirely natural.

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'. Multiplying the two matrices on the right and equating the entries we obtain

$$ax + by = x, \qquad cx + dy = y.$$

We instantly 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

$$AI_2 = I_2 A = A \tag{IV.9}$$

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 $I_2 A = A$, so we've established (IV.9).

Before moving on to inverses, it is appropriate to ask in which world does the identity (IV.9) 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 (IV.9) 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 IV.8 and test if they're invertible.

Example IV.10. 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. 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

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

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

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

This is clearly absurd! Therefore, C is not invertible [1]. For a more graphic illustration of this fact, see Figure IV.2.

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

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

[1] In *Foundations*, 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.

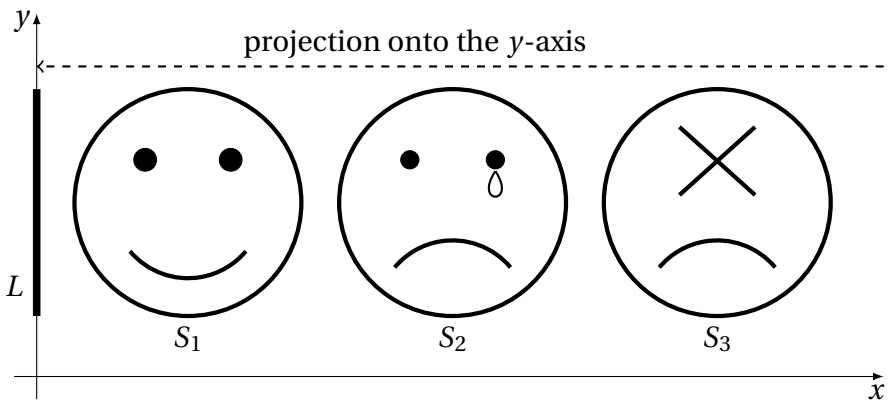


FIGURE IV.2. Some non-zero matrices don't have inverses. The matrix C represents projection onto the y-axis. Note that C sends the three 'smileys' S_1, S_2, S_3 to the line segment L. If C had an inverse, would this inverse send L to S_1, S_2 or S_3? We see that C^{-1} does not make any sense!

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 bottom-left entries don't match. So C is not invertible. $\Diamond$

IV.6.1. Discovering Invertibility. We will now work in generality. Let A be a 2×2 matrix and write

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

Suppose that A is invertible and write

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

We want to express A^{-1} in terms of A, or more precisely, x, y, z, w in terms of a, b, c, d. We want

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

Multiplying and equating entries we arrive at four equations:

$$ax + cy = 1 \tag{IV.10}$$

$$bx + dy = 0 \tag{IV.11}$$

$$az + cw = 0 \tag{IV.12}$$

$$bz + dw = 1. \tag{IV.13}$$

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

$$\text{(IV.14)} \qquad \begin{cases} (ad-bc)x = d, & (ad-bc)y = -b, \\ (ad-bc)z = -c, & (ad-bc)w = a. \end{cases}$$

Let's assume 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

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

Now check by multiplying that $A^{-1}A = I_2$ and also $AA^{-1} = I_2$.

What if $ad-bc = 0$? We assumed that A has an inverse A^{-1} and deduced (IV.14). If $ad-bc = 0$ then (IV.14) tells us that $a = b = c = d = 0$ and so $A = 0_{2\times 2}$ which certainly isn't invertible. This is a contradiction. Thus if $ad-bc = 0$ then A is not invertible. We've proved the following theorem.

Theorem IV.11. *A matrix* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ *is invertible if and only if* $ad-bc \neq 0$. *If so, the inverse is*

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

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

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

From Theorem IV.11 we know that a 2×2 matrix A is invertible if and only if $\det(A) \neq 0$.

Theorem IV.12. *(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 for the reader. Parts (a), (b) follow from the definition and effortless calculations (make sure you do them). For (c) note that

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

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

Exercise IV.13. (The Geometric Meaning of Determinant) You might be wondering (in fact should be wondering) about the geometric meaning of the determinant. This exercise answers your question. 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 IV.3). 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?

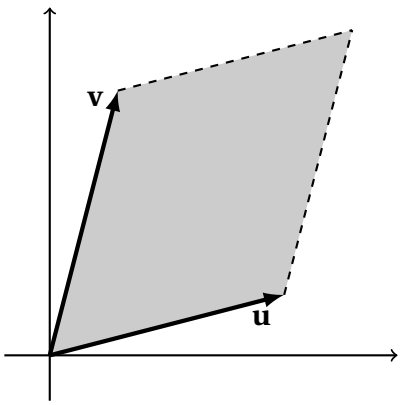


FIGURE IV.3. If $\mathbf{u}$ and $\mathbf{v}$ are the columns of A then the shaded area is $|\det(A)|$.

Exercise IV.14. 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.

IV.7. 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 [1]

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

Exercise IV.15. You know that R_θ represents anticlockwise rotation about the origin through angle θ. Describe in words the transformation associated to $-R_\theta$. (**Warning: don't be rash!**)

Gutted that the chapter on matrices is coming to an end? Cackle. You'll get to gorge (binge?) on them in *Linear Algebra*.

---

[1] Is this clever ... or lame: $\begin{pmatrix} \cos\frac{\pi}{4} & -\sin\frac{\pi}{4} \\ \sin\frac{\pi}{4} & \cos\frac{\pi}{4} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} =$  ?

CHAPTER V

Groups

V.1. The Definition of a Group

A *group* is a pair $(G, \circ)$ where G is a set and $\circ$ is a binary operation on G, such that the following four properties hold [2]

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

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

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

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

$$a \circ e = e \circ a = a;$$

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

$$a \circ b = b \circ a = e.$$

Remark. If $\circ$ is a binary operation then (i) automatically holds. So why did I list (i) in the definition? I've put it in 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.

V.2. First Examples (and Non-Examples)

Example V.1. $(\mathbb{R}, +)$ 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 $-a$: $a + (-a) = (-a) + a = 0$. $\Diamond$

Example V.2. Recall our definition of the natural numbers:

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

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).

[2] 99% of mathematicians call (i)–(iv) the "group axioms", and you can call them that if you wish. I call them the "defining properties of a group" since I think that the word axiom should be reserved for statements of 'universal truth'. An example of an axiom is: $a + b = b + a$ for any two integers a, b.

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. ♢

Example V.3. $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$ and $(\mathbb{C}, +)$ are groups. ♢

Example V.4. Recall we defined

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

Then $(\mathbb{R}^*, \cdot)$ is a group, where of course $\cdot$ 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 in Section XV.4. ♢

Example V.5. $(\mathbb{R}^2, +)$ is a group. Let's prove this. We're allowed to assume the usual properties of the real numbers (see Section III.8). 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.

If you got bored reading this then you have my sympathy; I was bored typing it. What matters is that you realize that the properties of addition in $\mathbb{R}^2$ are not there by divine covenant nor by magic; they simply follow from the definition of addition in $\mathbb{R}^2$ and corresponding properties of the real numbers. I can write down similar proofs for Examples V.6, V.7, and V.8, but I daren't try your patience with the interminable tedium. ♢

no magic yet

Example V.6. $(\mathbb{R}^n, +)$ is a group for any $n \geq 2$. ♢

Example V.7. $(\mathbb{R}[x], +)$ is a group. ♢

Example V.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. ♢

Example V.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). ♢

Example V.10. Let $B = \{1, i, -1, -i\}$, where $i = \sqrt{-1}$. Then $(B, \cdot)$ is another example of a finite group. ♢

Example V.11. Let $C = \{1, i\}$. Then $(C, \cdot)$ is not a group since it isn't closed; for example $i \cdot i = -1 \notin C$. ♢

V.3. Abelian Groups

We say that a group $(G, \circ)$ is *abelian* if (in addition to the defining properties (i)–(iv) of a group) it also satisfies

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

$$a \circ b = b \circ a.$$

Example V.12. 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.

V.4. 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 V.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.

shameless hard sell

The group of symmetries of a square is a great example of a group; it is finite, non-abelian and there is no additional structure. In future, this will be an excellent example to test any ideas you have on groups. Consider a square as in Figure V.1 with vertices labelled 1, 2, 3, 4 (note the vertex numbering goes in an anticlockwise direction). Let O be the centre of the square.

The *symmetries* of the square are the rotations and reflections that keep the square occupying exactly the same position (but might change where the vertex numbers are). Let ρ_0, ρ_1, ρ_2, ρ_3 be anticlockwise rotations of the square about O by 0°, 90°, 180° and 270°. In effect, ρ_0 means "do nothing". We think of two symmetries as being the same if they have the same effect on the square. A rotation about O of 360° has the same effect as ρ_0. A clockwise rotation about O of 90° has the same effect as ρ_3. We quickly see that ρ_0, ρ_1, ρ_2, ρ_3 are the only rotations that keep the square in exactly the same position.

What about reflections? There are four reflections that keep the square occupying exactly the same position, as in Figure V.1:

- 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.

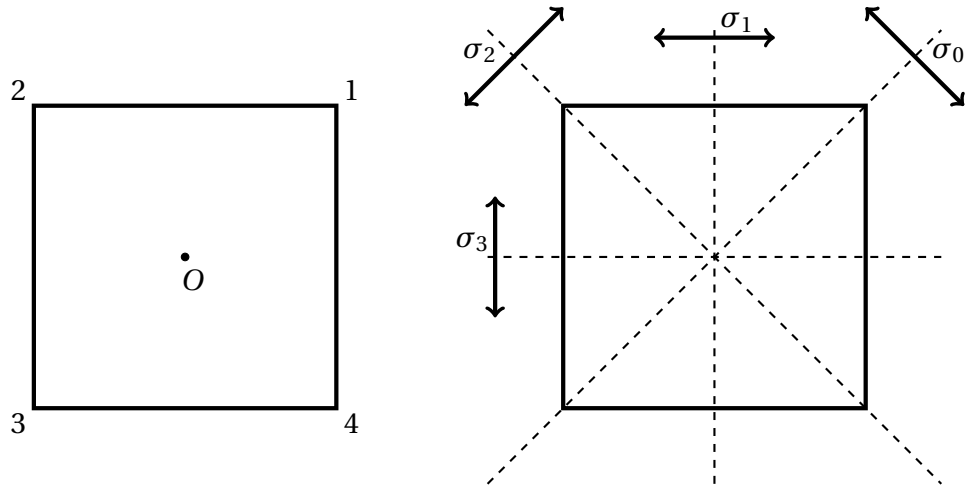


FIGURE V.1. Left: the square with vertices labelled 1, 2, 3, 4. Right: the reflections $\sigma_0, \sigma_1, \sigma_2, \sigma_3$.

Thus the symmetries of a square form a set which we shall denote by

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

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. If we have two symmetries, how do we combine them? In other words, if $\alpha, \beta \in D_4$, what is $\alpha \circ \beta$? **We define $\alpha \circ \beta$ to be the symmetry we obtain by doing β first then α** (the order is very important and we'll talk more about this later). Thus $\rho_1 \circ \rho_2$ is anticlockwise rotation about O of 180° followed by anticlockwise rotation about O of 90°. This has the same effect as ρ_3. Thus we write $\rho_1 \circ \rho_2 = \rho_3$.

Let's try another composition: $\rho_1 \circ \sigma_0$. In other words, first reflect in the diagonal joining 1 and 3, then rotate anticlockwise about O by 90°. This has the same effect at σ_1 and we can write $\rho_1 \circ \sigma_0 = \sigma_1$. Also (draw pictures) $\sigma_1 \circ \sigma_0 = \rho_1$ and $\sigma_2 \circ \sigma_2 = \rho_0$ (note that ρ_0 means do nothing). Now we know how to do composition we can write out a composition table:

$\circ$	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 I've made any mistakes!

I want to convince you that $(D_4, \circ)$ is a group. The first thing we should ask about is closure. This is clear from the table; when you compose two elements of D_4 you get an element of D_4. Let's skip associativity for now and talk about the existence of the identity element. For this we don't need the table. 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. So we have to prove that $(f \circ g) \circ h = f \circ (g \circ h)$ for all $f, g, h \in D_4$. But there are 512 triples f, g, h, so that's a lot of checking. Is there a clever way? Yes there is, and it relies on thinking about the elements of D_4 as functions [1]. Remember that we labelled the vertices of the square with 1, 2, 3, 4 as in Figure V.1. Now ρ_1 takes vertex 1 to where vertex 2 was and vertex 2 to where vertex 3 was and vertex 3 to where vertex 4 was and vertex 4 to where vertex 1 was. We can think of ρ_1 as a function from $\{1, 2, 3, 4\}$ to itself as in Figure V.2.

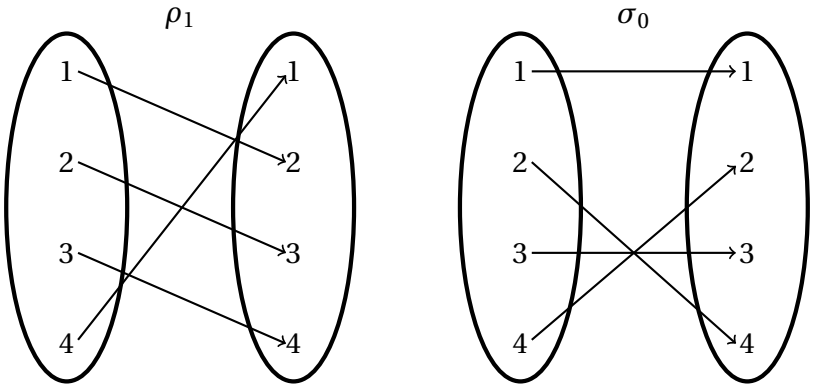


FIGURE V.2. ρ_1 (left) and σ_0 (right) considered as functions from $\{1, 2, 3, 4\}$ to itself.

In fact, this way every element of D_4 can be thought of as a function from $\{1, 2, 3, 4\}$ to itself. If we think like this, $\circ$ simply becomes composition of functions from D_4 to itself. We know that composition of functions is associative by Lemma III.2. Thus the binary operation $\circ$ on D_4 is *a key trick*

[1] You've seen this idea before: in the proof of Theorem IV.9 we showed that matrix multiplication is associative by thinking of matrices as functions and matrix multiplication as composition of functions.

associative. We have now checked all the conditions (i)–(iv) for a group, so $(D_4, \circ)$ is a group!

Remarks.

Moral

- Notice that by changing the way we looked at the elements of D_4, we saved ourselves from excruciatingly checking 512 laborious cases. This is a recurring theme in algebra, where a conceptual outlook saves you from much trouble.
- You might have found our definition of composition in D_4 strange: $\alpha \circ \beta$ means apply β first then α. 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.

non-abelian group

- D_4 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,

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

V.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; it's geometrically obvious. 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 V.3 for a complete list.

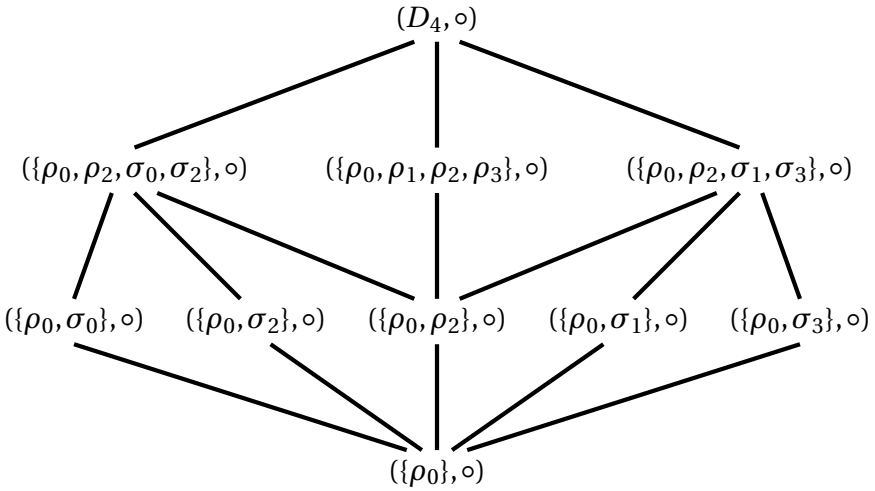


FIGURE V.3. The figure shows the subgroups of $(D_4,\circ)$ and how they fit inside each other.

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, but even then it will still be a little tedious!

Exercise V.13. 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. 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 [1]. Is it abelian? It has six subgroups; write them down.

Exercise V.14. Write down the symmetries of a triangle that is isoceles but not equilateral and a composition table for them. Do they form a group?

[1] 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 $S_{n!}$.

CHAPTER VI

First Theorems

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'?

Theorem VI.1. *Let* $(G,\circ)$ *be a group. Then* $(G,\circ)$ *has a unique identity element.*

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

$$a \circ e = e \circ a = a, \tag{VI.15}$$

and

$$a \circ e' = e' \circ a = a. \tag{VI.16}$$

Now let us try evaluating $e \circ e'$. If we let $a = e$ and use (VI.16) we find

$$e \circ e' = e.$$

But if we let $a = e'$ and use (VI.15) we find

$$e \circ e' = e'.$$

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

Theorem VI.2. *Let* $(G,\circ)$ *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 VI.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.

key trick!

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

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

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

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

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

VI.1. Getting Relaxed about Notation

It is quite tedious to keep writing $\circ$ for the group operation. If $(G, \circ)$ is a group and $a, b \in G$, we shall write ab for $a \circ b$, unless there is reason to fear confusion. For example if $(G, \circ) = (\mathbb{R}, +)$ then it is stupid 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.

If G is a group, and we're writing ab for $a \circ b$, 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. Now

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

which looks familiar. Moreover, if n is a positive integer we shall write

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

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

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

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

Thus writing a^3 in this context does not make any sense. ◊

Let's get back to groups. It's the associativity which makes it OK for us to write a^n, and you can convince yourself quickly that

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

You should realize that all this is happening inside the group G that contains the element a. In particular, $a^n \in G$ for all $n \in \mathbb{Z}$. How do we know this? For a start, G is closed under composition, so because $a \in G$, so is $a^2 = a \circ a$. Now that we know that a and a^2 are in G, we know that

$a^3 = a \circ a^2 \in G$ and so on. You can use induction to show that $a^n \in G$ for $n = 1, 2, 3, \ldots$. Also $1 \in G$ (because 1 is the symbol we're using for the identity element of G). And we've adopted the convention $a^0 = 1$, so $a^0 \in G$. We also want to check that $a^{-1}, a^{-2}, \ldots$ are in G. But $a^{-n} = (a^n)^{-1}$, and since a^n is already in G for positive n, so is its inverse.

an algebraic booby trap

What about $(ab)^n = a^n b^n$? Does this identity hold too? Let us think about this with $n = 2$. Now in the old notation [1]

$$(ab)^2 = a \circ b \circ a \circ b$$

and

$$a^2 b^2 = a \circ a \circ b \circ b.$$

Do these have to be the same? No, because the order of the middle two is different and since we're not assuming that our group is abelian we have no right to assume that $b \circ a = a \circ b$.

Example VI.4. 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 $\rho_1^2 \sigma_0^2 \neq (\rho_1 \sigma_0)^2$. ◊

Let us summarize our findings.

Theorem VI.5. *Let G be a group, and let $a \in G$. Then $a^n \in G$ for all $n \in \mathbb{Z}$. Moreover, if m, n are integers then*

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

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

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

Here is a crucial result that you should get used to.

Theorem VI.6. *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.

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

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

[1] Perhaps you think that I should write $(ab)^2 = (a \circ b) \circ (a \circ b)$, but because of associativity the bracketing does not matter. Whichever bracketing you apply to $a \circ b \circ a \circ b$ you will get the same result.

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

Never write a/b unless the group is abelian. This notation is ambiguous; does a/b mean $b^{-1}a$ or ab^{-1}? The two aren't the same in the non-abelian world. *pitfall*

Exercise VI.7. Use D_4 to give counterexamples to the following:

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

Exercise VI.8. Let G be a group satisfying $a^2 = 1$ for all a in G. Show that G is abelian.

VI.2. 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 VI.5. Here is the translation.

Theorem VI.9. *Let G be an (abelian) group with addition as the binary operation, and let $a \in G$. Then $na \in G$ for all $n \in \mathbb{Z}$. Moreover, if m, n are integers then*

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

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

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

CHAPTER VII

More Examples of Groups

Examples are an integral part of abstract algebra, and give it meaning and life. They are as important as the definitions and theorems. For this reason, I've crammed these notes with examples. Don't just flick through them saying, "yeah, yeah, that's obvious". Make a serious effort to study them, and know them for the exam. *And enjoy them.*

exam tip!

VII.1. Matrix Groups I

We saw that $(M_{2\times 2}(\mathbb{R}),+)$ is a group. This in fact is **not** an 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 IV.4 and in particular at Example IV.8. Can we obtain a group out of (say) 2×2 matrices under multiplication?

To answer, let's look back to Example V.4. There we obtained a multiplicative group from the real numbers by removing 0. Of course we removed 0 because it doesn't have a multiplicative inverse. It will not be enough for us to exclude the zero matrix, simply because there are non-zero matrices that do not have an inverse—see for example IV.10. What if we exclude all non-invertible matrices; do we get a group under multiplication?

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 IV.11). So $\mathrm{GL}_2(\mathbb{R})$ contains all the invertible 2×2 matrices (with real entries) and none of the non-invertible ones.

Theorem VII.1. $\mathrm{GL}_2(\mathbb{R})$ *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 IV.12). 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 IV.9.

The identity matrix $I_2 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ is in $\mathrm{GL}_2(\mathbb{R})$ 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 cooked up $\mathrm{GL}_2(\mathbb{R})$ 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 IV.12 that $\det(A^{-1}) \neq 0$ and indeed $\det(A^{-1}) = 1/\det(A)$. Moreover, A^{-1} is a 2×2 matrix with real entries. Hence $A^{-1} \in \mathrm{GL}_2(\mathbb{R})$. □

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...

Exercise VII.2. 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 XV.25.

VII.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. I have an excellent reason for writing $\mathbb{Z}/m\mathbb{Z}$ instead of $\mathbb{Z}/m$ and $\mathbb{Z}_m$. I want you to get used to the notation of quotient groups which we'll cover in Chapter XIII.

$\mathbb{Z}_m$ *vs.* $\mathbb{Z}/m\mathbb{Z}$

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

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

In otherwords, $\overline{a}$ consists of all integers congruent to a modulo m. From *Foundations* you know that

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

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

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

Example VII.3. The addition and multiplication tables for $\mathbb{Z}/6\mathbb{Z}$ are in Table VII.1.

◊

Exercise VII.4. 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 VII.1. The addition and multiplication tables for $\mathbb{Z}/6\mathbb{Z}$.

Theorem VII.5. *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 [1].

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

[1] 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 $-\overline{0}=\overline{0}$.

CHAPTER VIII

Orders and Lagrange's Theorem

We return to using multiplicative notation. In Theorem VI.5 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 $1, a, a^2, a^3$.

VIII.1. The Order of an Element

The above discussion leads us to the following definition.

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**.

Example VIII.1. 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? ◊

Example VIII.2. 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. ◊

Example VIII.3. When you saw the equation $a^n = 1$ in the above example, I'm sure you immediately remembered 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* that there are three cube roots of unity. These are $1, \zeta, \zeta^2$, where $\zeta = e^{2\pi i/3}$. See Figure VIII.1. 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 VIII.1. 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. ♢

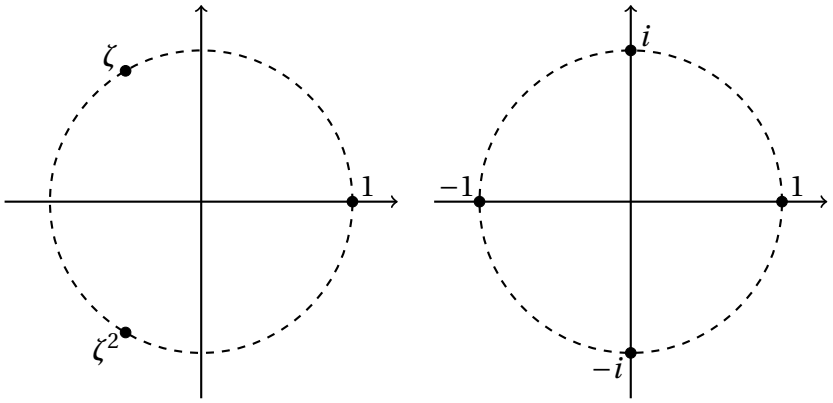


FIGURE VIII.1. On the left, the three cube roots of unity: here $\zeta = e^{2\pi i/3}$. On the right, the four fourth roots of unity. Note that $e^{2\pi i/4} = e^{\pi i/2} = i$, so the fourth roots of unity are 1, i, $i^2 = -1$, and $i^3 = -i$.

Exercise VIII.4. Write down and sketch the sixth roots of unity. What are their orders? Repeat with the eighth roots of unity.

Exercise VIII.5. $\mathbb{C}^*$ has plenty of elements of infinite order. Write down a few.

Exercise VIII.6. 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, the following observations will have dawned on you (given here in multiplicative notation).

Lemma VIII.7. *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. Using division with remainder, we may write

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

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

Exercise VIII.8. 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.

Example VIII.9. 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 $a = 0$. ◊

You're probably 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.

Example VIII.10. 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 VIII.7 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. ◊

Exercise VIII.11. Find the orders of the elements of $(\mathbb{Z}/4\mathbb{Z}, +)$ and $(\mathbb{Z}/5\mathbb{Z}, +)$.

VIII.2. Lagrange's Theorem—Version 1

Mathematics is unique in that supreme beauty goes hand in hand with tremendous power. Lagrange's Theorem is one of the loveliest examples of such a combination of qualities, and we're almost ready to meet it. *I know you're brimming with excitement, but please be a little patient; we need one more definition.* *self-control advised*

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 $\#G$.

Theorem VIII.12. *(Lagrange's Theorem—Version 1) Let G be a finite group, and let g be an element of G. The order of g divides the order of G.*

The proof of Lagrange's Theorem will have to wait till Chapter XII. Here's a useful corollary.

Corollary VIII.13. *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. □

Example VIII.14. 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. ◊

Example VIII.15. 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 Lagrange's Theorem. ◊

CHAPTER IX

Subgroups

Excruciating pain precedes orgasmic pleasure!

It will be a long time before you come to appreciate and enjoy groups. Abstract algebra goes from being mind-numbingly boring to being an acquired taste and then an exhilarating experience and finally—if you're not careful—a hopeless addiction. We're still in the mind-numbingly boring part of the journey; you should see this part as an initiation rite that can't be skipped.

IX.1. What Were They Again?

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

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

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.

a trivial but important example

Example IX.1. $\mathbb{R}^*$ 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. $\Diamond$

Example IX.2. $\mathbb{Z}$ 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. $\Diamond$

Example IX.3. $\mathbb{R}$ is a subgroup of $\mathbb{R}[x]$ since any real number can be viewed as a polynomial of degree 0. $\Diamond$

Example IX.4. $(\emptyset,+)$ 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. $\Diamond$

IX.2. Criterion for a Subgroup

Theorem IX.5. *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* $a^{-1} \in H$.

Let's delay the proof until after we've tried out the theorem.

Example IX.6. 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 IX.5, H is a subgroup of $\mathbb{R}^*$. ◊

Example IX.7. 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 $a \circ b$ where $\circ$ 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 IX.5 that $2\mathbb{Z}$ is a subgroup of $\mathbb{Z}$.

By contrast, the set of odd integers

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

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

Example IX.8. In Subsection V.4.1, we listed the ten subgroups of D_4. Go back to that list, and use Theorem IX.5 to verify that a couple of them are indeed subgroups. ◊

Example IX.9. 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 IX.1 will help you visualise this. But at this stage in your academic career, you are expected 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 $\mathbb{R}^2$. ♢

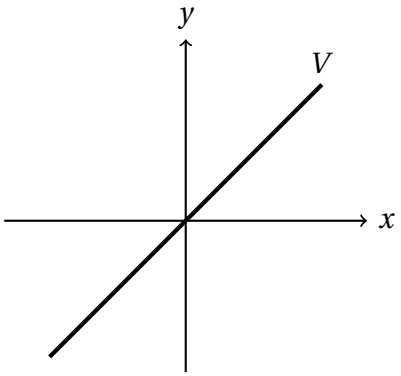


FIGURE IX.1. The set $V = \{(a,a) : a \in \mathbb{R}\}$ is the line $y = x$. It contains the identity element $(0,0)$, is closed under addition and negation. Therefore it is a subgroup of $\mathbb{R}^2$.

Example IX.10. 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 IX.2. Note that W does satisfy the first two conditions (a), (b) for being a subgroup. However, it does not satisfy 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 the theorem is not always satisfied. ♢

Example IX.11. Let

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

union of subgroups not (always) subgroup

You know from Example IX.9 that V is a subgroup of $\mathbb{R}^2$ (and is the line $y = x$). It is just as easy to show 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 IX.5. However,

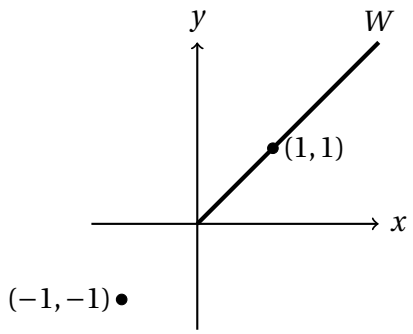


FIGURE IX.2. The ray $W = \{(a,a) : a \in \mathbb{R},\ a \geq 0\}$ is not a subgroup of $\mathbb{R}^2$. It contains the identity element $(0,0)$ and is closed under addition. The problem is with the existence of additive inverses; e.g. $(1,1)$ is in W but its inverse $(-1,-1)$ isn't in W.

$(1,1)$ 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 IX.3.

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

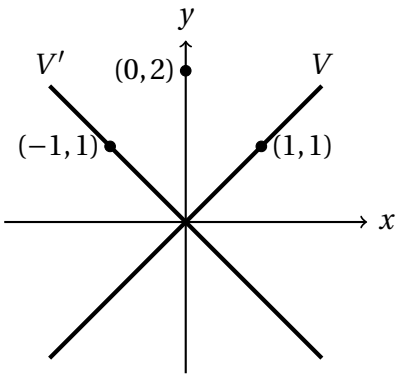


FIGURE IX.3. The lines $y = x$ and $y = -x$ are subgroups of $\mathbb{R}^2$. Their union is not.

Exercise IX.12. 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.

Example IX.13. 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 IX.4). 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. ♢

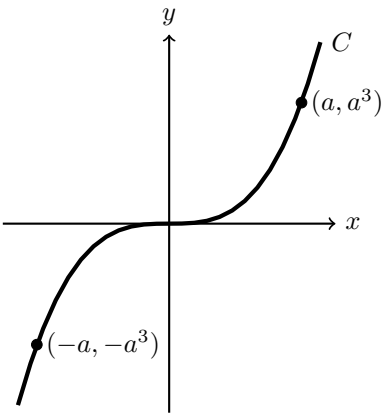


FIGURE IX.4. The set $C = \{(a, a^3) : a \in \mathbb{R}\}$ is the graph $y = x^3$. It satisfies conditions (a) and (c) for subgroups but not condition (b).

Example IX.14. $\mathbb{Z}^2$ is a subgroup of $\mathbb{R}^2$. ♢

Example IX.15. In Example IX.9 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 should've met at A-Level, 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 IX.5. 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:

$$\Pi \ : \ \mathbf{n} \cdot (\mathbf{x} - \mathbf{u}) = 0. \tag{IX.17}$$

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.

The plane Π in (IX.17) 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

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

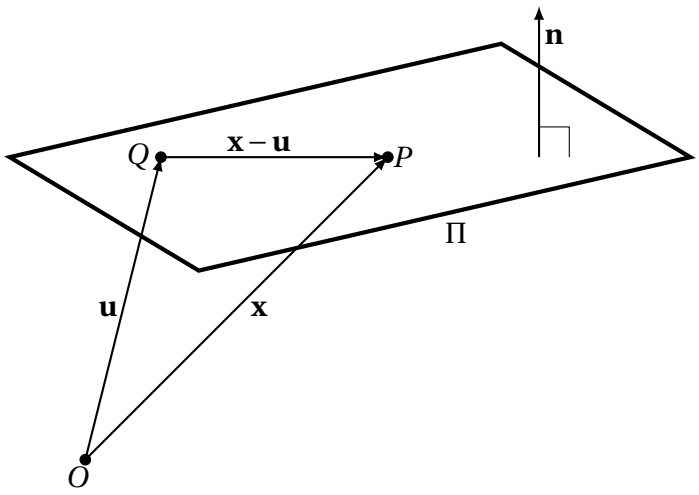


FIGURE IX.5. The *point-normal equation* of a plane. Here **n** is normal to the plane Π, Q is a fixed point on Π and **u** is its position vector. If P is any point on Π with position vector **x**, then **x**−**u** is parallel to the plane, and so **n**·(**x**−**u**) = 0.

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.* ◊

Exercise IX.16. Which lines in $\mathbb{R}^2$ define a subgroup? Justify your answer. *an important geometric exercise*

Example IX.17. 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 IX.6. Let us check that $\mathbb{S}$ is a subgroup of $\mathbb{C}^*$; it is clearly a subset. Of course the unit element of $\mathbb{C}^*$ is 1 and $|1| = 1$ so $1 \in \mathbb{S}$, which proves (a). Suppose $\alpha, \beta \in \mathbb{S}$. Then $|\alpha| = 1$ and $|\beta| = 1$.

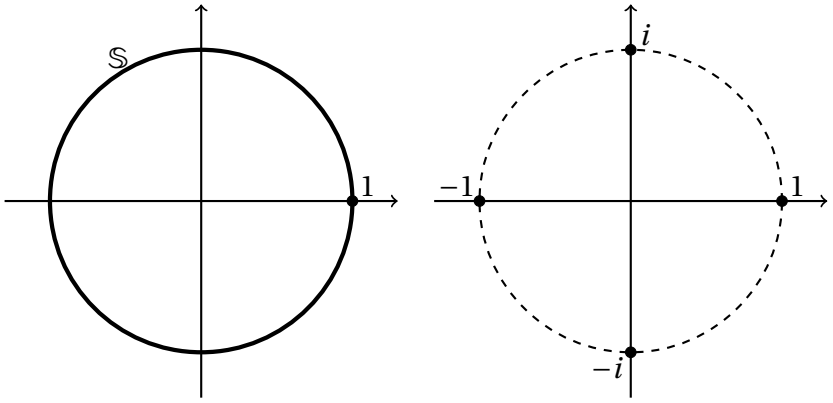


FIGURE IX.6. On the left, the group $\mathbb{S}$ which is just the unit circle. On the right, the subgroup of the fourth roots of unity.

From the properties of the absolute value [1] we have

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

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,

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

so $\alpha^{-1} \in \mathbb{S}$. By Theorem IX.5, $\mathbb{S}$ 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. ♢

Exercise IX.18. In the following, is H a subgroup of the group G? Give full justification. **Before you start answering:** You might be wondering why I don't specify the binary operation on G. Mathematicians generally don't; you're expected to figure it out from the context [2].

[1] At school you probably called $|\alpha|$ the *modulus* of α. Most mathematicians call $|\alpha|$ the *absolute value* of α.

[2] Devil's advocate: "Yes, I 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?"

In maths it is good to play the rôle of the devil's advocate, but not to the extent of renouncing good taste and common sense. Yes, there are binary operations other than addition that make the set of real numbers into a group. But if I wanted anything other than the usual or obvious operation I'd have told you so.

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

Exercise IX.19. Show that [1] the subgroups of $\mathbb{Z}/4\mathbb{Z}$ are $\{\overline{0}\}$, $\{\overline{0}, \overline{2}\}$ and $\mathbb{Z}/4\mathbb{Z}$.

Exercise IX.20. Show that the only subgroups of $\mathbb{Z}/3\mathbb{Z}$ are $\{\overline{0}\}$ and $\mathbb{Z}/3\mathbb{Z}$.

Exercise IX.21. Let

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

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

Exercise IX.22. 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 $\mathbb{C}^*$?

PROOF OF THEOREM IX.5. The theorem has an "if and only if" statement. It 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)".

if and only if

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 [2].

[1] 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.

[2] 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

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. But I'm already bored typing this proof, so I'll leave this part as a (mandatory) exercise. □

IX.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 VIII.1 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.

Lemma IX.23. *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.

Exercise IX.24. Is $U_2 \cup U_3$ a subgroup of $\mathbb{C}^*$?

---

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.

IX.4. Matrix Groups II

In Section VII.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.

Exercise IX.25. Let

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

Show that $\mathrm{SL}_2(\mathbb{R})$ is a group [1] (with respect to multiplication). This is known as the *special linear group* [2].

Exercise IX.26. 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* [3].

Now is a good time to revise Section IV.7 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

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

This is called the *special orthogonal group.*

Theorem IX.27. $\mathrm{SO}_2(\mathbb{R})$ *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 [4] $I_2 = R_0$, 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

[1] 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.

[2] If you've done Exercise IV.13 then you'll see that $\mathrm{SL}_2(\mathbb{R})$ consists of the matrices that preserve area and orientation.

[3] The modular group is probably the most interesting group in mathematics. Google it!

[4] In geometric terms, both I_2 and R_0 mean 'do nothing', so they must be equal.

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.

IX.5. Differential Equations

Let $\mathscr{C}$ be the set of infinitely differentiable real functions. This probably sounds scary, but to reassure you I'll just point out that $\mathscr{C}$ contains all polynomials, as well as $\sin t$, $\cos t$, e^t, e^{-t}. It is a fact that $\mathscr{C}$ is an additive group. Don't worry about the proof; it depends on properties of differentiability that you'll see eventually in analysis. Addition in $\mathscr{C}$ is done in a common sense way. For example, if $f(t) = t^2 + \sin(t)$ and $g(t) = 2t^2 - e^t$ then $f(t) + g(t) = 3t^2 + \sin(t) - e^t$. The identity element is 0.

Let's dive straight into an example. We define the following subset

$$H = \left\{ x(t) \in \mathscr{C} \,:\, t\frac{\mathrm{d}x}{\mathrm{d}t} - 2x = 0 \right\}.$$

In other words, H is the set of infinitely differentiable functions $x(t)$ that satisfy the differential equation

$$t\frac{\mathrm{d}x}{\mathrm{d}t} - 2x = 0. \tag{IX.18}$$

The function $x(t) = 0$ (which is the identity element of $\mathscr{C}$) clearly satisfies (IX.18) 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 $-x_1(t) \in H$ (easy exercise). So H is a subgroup of $\mathscr{C}$.

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.

Exercise IX.28. Which of the following differential equations define subgroups of $\mathscr{C}$?

(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) $\dfrac{\mathrm{d}\,x}{\mathrm{d}\,t} - x^2 = 0.$

IX.6. Non-Trivial and Proper Subgroups

It's very easy for you to prove the following proposition.

Proposition IX.29. *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 are boring, since they're always there. The interesting subgroups are the proper non-trivial subgroups.

Example IX.30. 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. $\Diamond$

Example IX.31. 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 $U_2 = \{1, -1\}$. ♢

Exercise IX.32. 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?

IX.7. Lagrange's Theorem—Version 2

Here is another version of Lagrange's Theorem. The relation between this version and the earlier one (Theorem VIII.12) will be explained once we have studied cyclic groups.

Theorem IX.33. *(Lagrange's Theorem—Version 2) Let G be a finite group, and H a subgroup of G. Then the order of H divides the order of G.*

Example IX.34. We saw in Example IX.31 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. ♢

Example IX.35. Recall that D_4 has order 8. In Figure V.3 we listed the ten subgroups of D_4. These have orders 1, 2, 4 and 8. This is consistent with Lagrange's Theorem. ♢

Exercise IX.36. 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.

Still hurting?

CHAPTER X

Cyclic Groups and Cyclic Subgroups

Cyclic groups are the simplest groups to understand.

Theorem X.1. *Let G be a group, and let g be an element of G. Write $\langle g\rangle$ for the set*

$$\langle g\rangle = \{g^n : n \in \mathbb{Z}\} = \{\ldots, g^{-2}, g^{-1}, 1, g, g^2, g^3, \ldots\}.$$

Then $\langle g\rangle$ is a subgroup of G.

PROOF. This is very easy to prove using Theorem IX.5. Have a go! □

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.

Example X.2. 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

$$\zeta^0 = 1, \zeta, \zeta^2, \ldots, \zeta^{n-1}.$$

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 X.1 lists the elements of U_6 and the subgroups they generate.

♢

Example X.3. 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

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 X.1. The six elements of U_6 and the cyclic subgroups they generate.

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 X.2. ♢

$\overline{a}$	$\langle \overline{a} \rangle$
$\overline{0}$	$\{\overline{0}\}$
$\overline{1}$	$\{\overline{0}, \overline{1}, \overline{2}, \overline{3}, \overline{4}, \overline{5}\}$
$\overline{2}$	$\{\overline{0}, \overline{2}, \overline{4}\}$
$\overline{3}$	$\{\overline{0}, \overline{3}\}$
$\overline{4}$	$\{\overline{0}, \overline{2}, \overline{4}\}$
$\overline{5}$	$\{\overline{0}, \overline{1}, \overline{2}, \overline{3}, \overline{4}, \overline{5}\}$

TABLE X.2. The six elements of $\mathbb{Z}/6\mathbb{Z}$ and the cyclic subgroups that they generate.

Example X.4. 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 V.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. ♢

Theorem X.5. *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.

Theorem X.6. *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.

a fundamental principle

PROOF. Observe that $\langle g \rangle$ is a set, and $\{1, g, \ldots, g^{n-1}\}$ 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,

$$\langle g \rangle = \{g^n : n \in \mathbb{Z}\} = \{\ldots, g^{-2}, g^{-1}, 1, g, g^2, g^3, \ldots\}.$$

That is $\langle g \rangle$ is the set of all powers of g. It is obvious that every element of $\{1, g, \ldots, g^{n-1}\}$ 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 $\{1, g, \ldots, g^{n-1}\}$. 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 $0, 1, 2, \ldots, n-1$.

The division algorithm is one of the most powerful ideas in algebra.

For this we will use the *division algorithm* which you met in *Foundations*. We can write

$$m = qn + r, \qquad q, r \in \mathbb{Z}, \quad 0 \leq r < n.$$

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 $0 \leq r < n$, we see that r is one of $0, 1, \ldots, n-1$. Therefore h is in $\{1, g, \ldots, g^{n-1}\}$. By our principle, we see that $\langle g \rangle = \{1, g, \ldots, g^{n-1}\}$. □

Exercise X.7. 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) $G = \mathrm{GL}_2(\mathbb{R})$, $g = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$.

Exercise X.8. 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.

Exercise X.9. 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?

Exercise X.10. 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.

Exercise X.11. Show that $\mathbb{S}$ is not cyclic.

X.1. Lagrange Revisited

You saw two versions of Lagrange's Theorem:

Theorem X.12. *(Lagrange's Theorem—Version 1) Let G be a finite group, and let g be an element of G. The order g divides the order of G.*

Theorem X.13. *(Lagrange's Theorem—Version 2) Let G be a finite group, and H a subgroup of G. Then the order of H divides the order of G.*

In fact Version 2 implies Version 1. Let us prove that.

Proposition X.14. *Version 2 of Lagrange's Theorem implies Version 1 of Lagrange's Theorem.*

PROOF. We assume Version 2 and deduce Version 1. Let G be a finite group and g an element of G. Suppose g has order n. By Theorem X.6, 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. □

This doesn't mean that we've proved Version 1 of Lagrange's Theorem. It does mean that once we prove Version 2, then we will have also proved Version 1.

Exercise X.15. 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 $G = \langle g \rangle$.

X.2. Subgroups of $\mathbb{Z}$

I'm feeling particularly inarticulate at the moment, so I can't explain why subgroups of $\mathbb{Z}$ are important. But nevertheless they are important and so we'll do them.

Trust me, I'm a doctor.

The first thing to note about $\mathbb{Z}$ is that it is cyclic. Does that mean that all elements of $\mathbb{Z}$ are powers of some element. No, because it is an additive group. If G is an additive group, and g is an element of G then

$$\langle g\rangle = \{ng : n \in \mathbb{Z}\} = \{\ldots, -2g, -g, 0, g, 2g, 3g, \ldots\}.$$

Thus $\mathbb{Z} = \langle 1\rangle$ and so it is cyclic. In fact, it is infinite and cyclic, unlike for example, U_n.

Lemma X.16. *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 IX.7. However, it is quicker to note that $k\mathbb{Z} = \langle k\rangle$, and so is a subgroup by Theorem X.1. □

Note that $0\mathbb{Z} = \{0\}$ has only the identity element. Also

$$\begin{aligned}(-k)\mathbb{Z} &= \{\ldots, -2(-k), -(-k), 0, -k, 2(-k), \ldots\}\\ &= \{\ldots, 2k, k, 0, -k, -2k, \ldots\}\\ &= \{\ldots, -2k, -k, 0, k, 2k, \ldots\}\\ &= k\mathbb{Z}\end{aligned}$$

because the order of elements in a set does not matter. In other words,

$$-\mathbb{Z} = \mathbb{Z}, \qquad -2\mathbb{Z} = 2\mathbb{Z}, \qquad -3\mathbb{Z} = 3\mathbb{Z}, \ldots.$$

So we have an infinite list of subgroups

$$\{0\}, \quad \mathbb{Z}, \quad 2\mathbb{Z}, \quad 3\mathbb{Z}, \quad 4\mathbb{Z}, \ldots$$

and we want to know if they're all the subgroups of $\mathbb{Z}$. The following theorem tells us that they are.

Theorem X.17. *Any subgroup of $\mathbb{Z}$ has the form $k\mathbb{Z}$ for some non-negative*[1] *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 $H = k\mathbb{Z}$.

Is this familiar? 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.

[1] The *non-negative integers* are $0, 1, 2, 3, \ldots$.

As k is in H, we know by Theorem VI.9 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, we can write

$$a = qk + r, \qquad q, r \text{ are integers and } 0 \leq r < k.$$

To remind: here q is the quotient of dividing a by k and r is the remainder. 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 $0 \leq r < k$, 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. *punchline*

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. □

Exercise X.18. The subgroups of $\mathbb{Z}^2$ are harder to describe. Write down a few.

Thrilled? Aren't you glad you trusted me?

CHAPTER XI

Isomorphisms

You must've noticed that 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 X.1 and X.2. In fact the groups U_m and $\mathbb{Z}/m\mathbb{Z}$ are isomorphic. What does this mean?

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 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.

Example XI.1. Define $\phi : \mathbb{Z}/m\mathbb{Z} \to U_m$ by the simple rule

$$\phi(\overline{a}) = \zeta^a, \qquad a = 0, 1, \ldots, m-1.$$

Then ϕ is a bijection and satisfies the magical 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. ♢

Example XI.2. 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 IX.27) defined by

$$\mathrm{SO}_2(\mathbb{R}) = \{R_\theta : \theta \in \mathbb{R}\},$$

and $\mathbb{S}$ is the circle group (page 55) 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. You should have no trouble guessing what the 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 $I_2, \mathcal{Z}, \ldots, \mathcal{Z}^{n-1}$ all satisfy the relationship $A^n = I_2$. ♢

Exercise XI.3. Let $\mathcal{Z} = R_{2\pi/6}$. Show that $\{1, \mathcal{Z}, \ldots, \mathcal{Z}^5\}$ 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.

Exercise XI.4. 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.

Tragically, the powers that be (who shall remain nameless) decided that this *Introduction to Abstract Algebra* should be a half-module, and so we won't have the time to explore the manifold pleasures of isomorphisms.

> "SAMIR, WHY HAVE THEY DENIED US THE CATS TO EXPERIENCE THESE PLEASURES? DON'T THEY LOVE US LIKE YOU DO? DO THEY WANT US TO FAIL AND TURN TO THE BOTTLE?"

Indignant? Paranoid? Seething with self-righteous rage?

Don't be a drama queen—there must be a perfectly innocent explanation.

CHAPTER XII

Cosets

Cosets are what we get when we 'shift' a subgroup by the elements of the group.

Definition. Let G be a group and H a subgroup. Let g be an element of G. We call the set

$$gH = \{gh : h \in H\}$$

a *left coset of* H in G and the set

$$Hg = \{hg : h \in H\}$$

a *right coset of* H in G.

Example XII.1. 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 V.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

$$\sigma_0 R = \sigma_1 R = \sigma_2 R = \sigma_3 R = \{\sigma_0, \sigma_1, \sigma_2, \sigma_3\}.$$

$\Diamond$

Exercise XII.2. Recall that $H = \{1, \sigma_2\}$ is also a subgroup of D_4. Compute its left cosets. Check that $\sigma_1 H \neq H\sigma_1$.

Of course, 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.

Example XII.3. $\mathbb{Z}$ is an additive group. The set of even integers $2\mathbb{Z}$ is a subgroup. What are its cosets? Let's compute a few:

$$0 + 2\mathbb{Z} = \{\ldots, 0 + (-4), 0 + (-2), 0 + 0, 0 + 2, 0 + 4, \ldots\} = \{\ldots, -4, -2, 0, 2, 4, \ldots\};$$
$$1 + 2\mathbb{Z} = \{\ldots, 1 + (-4), 1 + (-2), 1 + 0, 1 + 2, 1 + 4, \ldots\} = \{\ldots, -3, -1, 1, 3, 5, \ldots\};$$
$$2 + 2\mathbb{Z} = \{\ldots, 2 + (-4), 2 + (-2), 2 + 0, 2 + 2, 2 + 4, \ldots\} = \{\ldots, -4, -2, 0, 2, 4, \ldots\};$$
$$3 + 2\mathbb{Z} = \{\ldots, 3 + (-4), 3 + (-2), 3 + 0, 3 + 2, 3 + 4, \ldots\} = \{\ldots, -3, -1, 1, 3, 5, \ldots\}.$$

You'll quickly discover that

$$\cdots = -4 + 2\mathbb{Z} = -2 + 2\mathbb{Z} = 2\mathbb{Z} = 2 + 2\mathbb{Z} = 4 + 2\mathbb{Z} = \ldots$$

and

$$\cdots = -3 + 2\mathbb{Z} = -1 + 2\mathbb{Z} = 1 + 2\mathbb{Z} = 3 + 2\mathbb{Z} = \ldots.$$

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. ♢

Exercise XII.4. 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?

Exercise XII.5. 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 $\mathbb{R}^*$.

XII.1. Geometric Examples

Long ago (page 1) I told you:

> *You should get used to thinking geometrically, and to drawing pictures. The true meaning of most mathematical concepts is geometric. If you spend all your time manipulating symbols (i.e. doing algebra) without understanding the relation to the geometric meaning, then you will have very little in terms of mathematical insight.*

No doubt you have taken my advice on board and so there is no need for me to repeat it.

Example XII.6. 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 IX.17 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 XII.1.

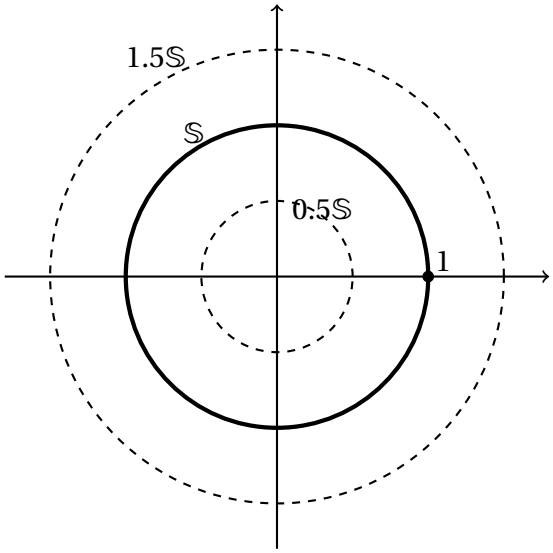


FIGURE XII.1. $\mathbb{S}$ and its cosets $0.5\mathbb{S}$ and $1.5\mathbb{S}$ in $\mathbb{C}^*$.

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). $\Diamond$

Example XII.7. In Exercise IX.16 I asked you 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

$$L\colon \mathbf{x} = \mathbf{w} + t\mathbf{v}.$$

This is a (slightly clumsy) school 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 XII.2

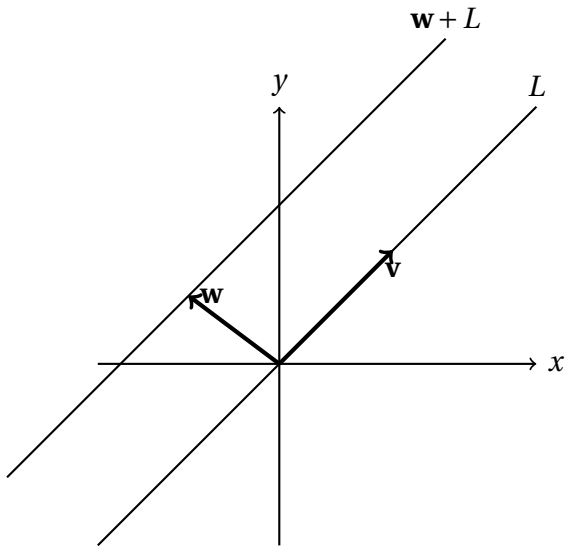


FIGURE XII.2. A line L defines a subgroup of $\mathbb{R}^2$ if and only if it passes through the origin. In that case, its cosets are the lines parallel to it.

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.* ♢

XII.2. Solving Equations

Cosets arise naturally when solving certain types of equations. It's difficult to make this precise at present. Instead I'll show you some examples so that you can see what I mean.

Example XII.8. If you did matrices at school, then you will probably know that a system of m linear equations in n variables can be written as a single matrix equation

$$A\mathbf{x} = \mathbf{b} \tag{XII.19}$$

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 (XII.19)? If (XII.19) 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 (XII.19) is precisely the coset $\mathbf{x}_0 + K$. ♢

Example XII.9. 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

$$a\frac{d^2x}{dt^2} + b\frac{dx}{dt} + cx = f(t), \tag{XII.20}$$

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*

$$a\frac{d^2x}{dt^2} + b\frac{dx}{dt} + cx = 0. \tag{XII.21}$$

Convince yourself that the solutions to the homogeneous equation (XII.21) form a group K with respect to addition (revise Section IX.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 (XII.20) and K? Again, if (XII.20) 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 (XII.20), then you can check that $x(t) - x_0(t)$ is a solution to the homogeneous equation (XII.21), and so is an element of K. It follows that the set of solutions to (XII.20) 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*. A lot of maths students never understand the First Isomorphism Theorem. They somehow don't realize that they've been using it for years when solving linear equations (and linear differential equations). Don't let that happen to you; after you meet the First Isomorphism Theorem, come back and review these two examples again.

a fate $almost$ *worse than death*

XII.3. Index

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.

Example XII.10. In Example XII.1, 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 $[D_4 : R] = 2$. ♢

Example XII.11. In Example XII.3 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 XII.4 then you'll know that $[\mathbb{Z}^2 : 2\mathbb{Z}^2] = 4$. ♢

Example XII.12. In Example XII.6, we found that the cosets of the circle group $\mathbb{S}$ in $\mathbb{C}^*$ are the circles centred at the origin. So the index $[\mathbb{C}^* : \mathbb{S}] = \infty$.

Example XII.13. 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 $[\mathbb{Z} : \{0\}] = \infty$. ♢

Exercise XII.14. Let G be a finite group. Let $\{1\}$ be the trivial subgroup consisting only of the identity element. Explain why $[G : \{1\}] = |G|$.

XII.4. 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.

Lemma XII.15. *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

A priceless tip!

finite. **The best way to show that two finite sets have the same number of elements is to set up a bijection 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

$$gh_1 = gh_2.$$

Now we *can't* say, 'divide by g'. If you've forgotten why, see the pitfall on page 40. By we *can* say multiply 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, $gH = \{gh : h \in H\}$, so $k = gh = \phi(h)$ for some h in H. □

A Highbrow Remark for the Cognoscenti. 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. If you find this interesting, have a look at cardinalities on `Wikipedia`. But only a brief look; trust me, set theory is as boring as hell. In any case, feel free to ignore this remark.

Example XII.16. Now is a good time to revisit the examples at the beginning of the chapter and make sure that Lemma XII.15 holds for them.

XII.5. The Second Innermost Secret of Cosets

Lemma XII.17. *Let G be a group and H be a subgroup. Let g_1, g_2 be elements of G. Then the cosets g_1H, g_2H are either equal or disjoint* [1].

[1] Two sets A, B are *disjoint* if they have no members in common. Another way of saying the same thing is: two sets A, B are disjoint if $A \cap B = \emptyset$. I'm now confused—have I said it in another way, or in the same way but with more notation?

Example XII.18. Look again at Example XII.6 and in particular Figure XII.1. 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. ◊

Geometric Epiphany I

Example XII.19. In Example XII.7, 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. ◊

Geometric Epiphany II

PROOF OF LEMMA XII.17. Suppose g_1H and g_2H are not disjoint. We want to show that they're equal. If you look again at the examples you'll see that $g_1H = g_2H$ doesn't necessarily mean that $g_1 = g_2$.

As g_1H and g_2H are not disjoint, they must have a common element. The elements of g_1H have the form g_1h_1 and the elements of g_2H have the form g_2h_2 where h_1, h_2 are in H. Thus there is a pair h_1, h_2 in H so that $g_1h_1 = g_2h_2$. In particular

$$g_1 = g_2h_2h_1^{-1}. \tag{XII.22}$$

We want to show that $g_1H = g_2H$. You'll no doubt recall that to prove two sets are equal we have to show that every element in either set is an element of the other set. Take an element of g_1H. This must have the form g_1h for some h in H. We want to show that g_1h is also an element of g_2H. Now note

$$\begin{aligned} g_1h &= (g_2h_2h_1^{-1})h \qquad \text{by (XII.22)} \\ &= g_2(h_2h_1^{-1}h). \end{aligned}$$

However, $h_2h_1^{-1}h$ is a product of elements of the subgroup H and therefore an element of H. Hence we've written g_1h in the form g_2h' where $h' = h_2h_1^{-1}h$ is an element of H. Thus every element of g_1H is again an element of g_2H. Similarly, every element of g_2H is an element of g_1H. Hence $g_1H = g_2H$. □

XII.6. Lagrange Super-Strength

I've stated Lagrange's Theorem a very long time ago, and kept you waiting for the proof ever since. Surely you consider this delay a deliberate act of unspeakable cruelty. It was indeed deliberate; I thought the prolonged wait would heighten the anticipation and make you appreciate and enjoy the proof even more. Alas, through an act of infinite selflessness, I've sacrificed my popularity to intensify your infatuation with the subject.

shedding bitter tears of remorse and penantly pleading for forgiveness

We now state an even stronger version of Lagrange's Theorem.

Theorem XII.20. *(Lagrange's Theorem—Version 3) Let G be a finite group and H a subgroup. Then*

$$|G| = [G : H] \cdot |H|.$$

This version is saying more than Version 2 of Lagrange's Theorem (Theorem IX.33). Version 2 says that $|H|$ divides $|G|$. Version 3 tells us that not only does $|H|$ divide $|G|$, but that the ratio is the index $[G:H]$. So if we prove this version of Lagrange's Theorem then we have also proved Version 2.

PROOF OF THEOREM XII.20. Let $g_1H, g_2H, \ldots, g_mH$ be the distinct left cosets of H. As they are distinct, we know by Lemma XII.17 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 XII.15,

$$|g_1H| = |g_2H| = \cdots = |g_mH| = |H|.$$

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. □

Pure ecstacy? Of course! Maths is about delayed gratification.

CHAPTER XIII

Quotient Groups

allow me to end your suffering

Taking quotients is one of the most powerful concepts in mathematics. It should also be one of the least painful to assimilate. Instead of revelling in quotients, most Warwickers go through three or four miserable years of being terrorised by them. The difficulties with quotients are purely psychological. To overcome them, you just need to study and visualize a good number of examples. *What are we waiting for?*

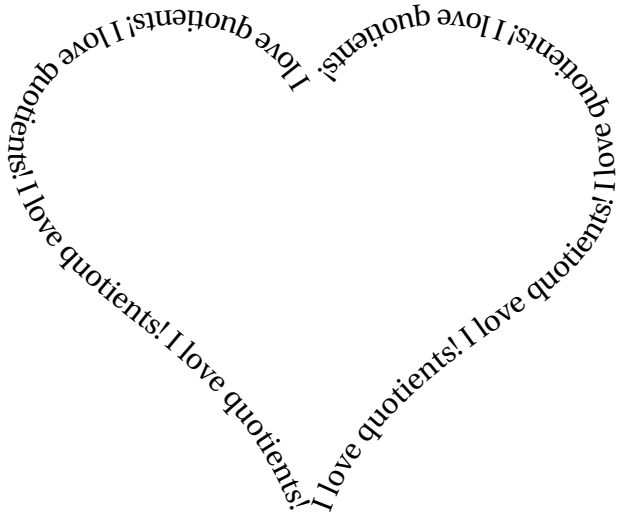


FIGURE XIII.1. Nurture a positive attitude to quotients—it won't let you down!

XIII.1. Congruences Modulo Subgroups

Let $(G,+)$ be an abelian group, where the binary operation is 'addition'. For example, G could be $\mathbb{R}$, $\mathbb{R}^2$, $\mathbb{C}$, $\mathbb{Z}$, $\mathbb{R}[x]$ etc. Let H be a subgroup. Let a, b be elements of G. We say that a, b are *congruent modulo* H if $a-b\in H$. In this case we write $a\equiv b \pmod{H}$.

Example XIII.1. 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. Let a, $b\in\mathbb{Z}$. Then $a\equiv b \pmod{m\mathbb{Z}}$ if and only $a-b$ is a multiple of m. In other words, $a\equiv b$

$(\text{mod } m\mathbb{Z})$ if and only if $a \equiv b \pmod{m}$. *The concept of congruence modulo subgroups is a generalization of the earlier concept of congruence modulo integers.* ♢

Example XIII.2. $\mathbb{Z}$ is a subgroup of $\mathbb{R}$. Two real numbers a, b are congruent modulo $\mathbb{Z}$ if and only if $a - b \in \mathbb{Z}$. This means that their difference is an integer. So, for example $1437.14 \equiv 0.14 \pmod{\mathbb{Z}}$. It may seem that congruence modulo $\mathbb{Z}$ is a stupid idea. 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 IX.17 we defined the circle group

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

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 $1 \in \mathbb{S}$. Of course, you want me to be less clumsy and say that f is periodic with period 1. Indeed $f(\phi) = f(\theta)$ if and only if $\phi = \theta + n$ where n is an integer. Now $\mathbb{Z}$ is a subgroup of $\mathbb{R}$. So we can rewrite that fact as $f(\phi) = f(\theta)$ if and only if $\phi \equiv \theta \pmod{\mathbb{Z}}$. ♢

Example XIII.3. Let $X = \{(a, 0) : a \in \mathbb{R}\}$. It's easy to show that X is a subgroup of $\mathbb{R}^2$, which is simply the x-axis. What does it mean for two points to be congruent modulo X? Suppose (a_1, b_1) and (a_2, b_2) are in $\mathbb{R}^2$. Then $(a_1, b_1) \equiv (a_2, b_2) \pmod{X}$ 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 two points are congruent modulo X if and only if they have the same y-coordinate. ♢

Example XIII.4. Let $G = \mathbb{R}[x]$. Let $H = \{f \in \mathbb{R}[x] : f(0) = 0\}$. It is an easy exercise to show that H is a subgroup of $\mathbb{R}[x]$. Now let's understand what it means for two polynomials to be congruent modulo H. Suppose g, $h \in \mathbb{R}[x]$. Write [1]

$$g = a_0 + a_1 x + \cdots + a_n x^n, \qquad h = b_0 + b_1 x + \cdots + b_n x^n,$$

where $a_0, \ldots, a_n$ and $b_0, \ldots, b_n$ are real numbers. Let $f = g - h$. Then $g \equiv h \pmod{H}$ if and only if $f(0) = 0$, which means $a_0 - b_0 = 0$. Therefore g and h are congruent modulo H if and only their constant terms are equal. ♢

[1] 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 $b_0 = 4$, $b_1 = -3$, $b_2 = 0$, $b_3 = 2$

Exercise XIII.5. Let $(G,+)$ be an abelian group. We know that $\{0\}$ and G are subgroups of G. What does it mean for a and b to congruent modulo $\{0\}$? What does it mean for a and b to be congruent modulo G?

XIII.2. Congruence Classes and Cosets

Let $(G,+)$ be an additive abelian group and H a subgroup. Let $a \in G$. We shall denote by $\overline{a}$ the *congruence class of a modulo H*; this is defined by

$$\overline{a} = \{b \in G : b \equiv a \pmod{H}\}.$$

In words, the congruence class of a modulo H is the set of all elements of G that are congruent to a modulo H.

Example XIII.6. If $G = \mathbb{Z}$ and $H = m\mathbb{Z}$, then $\overline{a}$ is simply the congruence class of a modulo m:

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

In *Foundations*, $\overline{a}$ is denoted by $[a]$. ◊

Lemma XIII.7. *Let $(G,+)$ be an additive abelian group and H a subgroup. Let $a \in G$ let $\overline{a}$ be the congruence class of a modulo H. Then*

$$\overline{a} = a + H.$$

PROOF. We know $\overline{a}$ is the set of $b \in G$ that are congruent to a modulo H. But $b \equiv a \pmod{H}$ is the same as saying $b - a \in H$ or $b \in a + H$. So $\overline{a} = a + H$. □

We made the set of congruence classes in $\mathbb{Z}$ modulo $m\mathbb{Z}$ into a group $\mathbb{Z}/m\mathbb{Z}$, and in the same way we can form a group out of the set of congruence classes in an additive abelian group G modulo a subgroup H.

Definition. Let $(G,+)$ be an additive abelian group and H a subgroup. We define the quotient group $(G/H,+)$ to the set of congruence classes (or the set of cosets)

$$G/H = \{\overline{a} : a \in G\}$$

with addition being defined by

$$\overline{a} + \overline{b} = \overline{a+b}. \tag{XIII.23}$$

As usual, we need 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 (XIII.23) is *well-defined*. These details will disrupt the flow of things and I've relegated them to Section XIII.6. For now, we want to focus on understanding quotient groups and how to think about them.

XIII.3. $\mathbb{R}/\mathbb{Z}$

In Example XIII.2 we looked at congruences in $\mathbb{R}$ modulo $\mathbb{Z}$. We now want to understand the group $\mathbb{R}/\mathbb{Z}$. Note that every real number is congruent modulo $\mathbb{Z}$ to a unique number in the half-open interval

$$[0,1) = \{x \in \mathbb{R} : 0 \le x < 1\}.$$

Therefore

$$\mathbb{R}/\mathbb{Z} = \{\overline{a} : a \in [0,1)\}.$$

So when we add $\overline{a} + \overline{b}$, we take the result of $a+b$, and and simplify by subtracting an integer if necessary to obtain $c \in [0,1)$, and then letting $\overline{a} + \overline{b} = \overline{c}$. For example,

$$\overline{0.7} + \overline{0.2} = \overline{0.9}, \qquad \overline{0.7} + \overline{0.4} = \overline{0.1}, \qquad \overline{0.3} - \overline{0.5} = \overline{0.8}.$$

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,

$$\hat{f} : \mathbb{R}/\mathbb{Z} \to \mathbb{S}, \qquad \hat{f}\left(\overline{\theta}\right) = e^{2\pi i\theta}.$$

Check that $\hat{f}$ is a bijection that satisfies

$$\hat{f}\left(\overline{\theta} + \overline{\phi}\right) = \hat{f}\left(\overline{\theta}\right)\hat{f}\left(\overline{\phi}\right).$$

In essence what this is saying is that the two groups $(\mathbb{R}/\mathbb{Z}, +)$ and $(\mathbb{S}, \cdot)$ are 'essentially the same'. Indeed, they're isomorphic. Now is a good time to look again at Chapter XI.

Here is how you should think about $\mathbb{R}/\mathbb{Z}$. We identified it with the interval $[0,1)$. Think of starting at $\overline{0.95}$ and moving up in small steps of $\overline{0.01}$:

$$\overline{0.95}, \overline{0.96}, \overline{0.97}, \overline{0.98}, \overline{0.99}, \overline{0.00}, \overline{0.01}, \overline{0.02}, \overline{0.03}, \ldots$$

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.

Exercise XIII.8. $\mathbb{R}/\mathbb{Z}$ has four elements of order 5. Find them.

Exercise XIII.9. Let $\alpha \in [0,1)$. Show that $\overline{\alpha}$ has finite order in $\mathbb{R}/\mathbb{Z}$ if and only if α is rational.

Exercise XIII.10. Show that $\hat{f}$ takes rational numbers to roots of unity.

XIII.4. $\mathbb{R}^2/\mathbb{Z}^2$

After $\mathbb{R}/\mathbb{Z}$ you shouldn't have any trouble imagining $\mathbb{R}^2/\mathbb{Z}^2$. You're allowed to shift any point in $\mathbb{R}^2$ by an integer multiple of **i** and an integer multiple of **j**. So you end up in the unit square:

$$\{(x,y) : 0 \le x < 1,\ 0 \le y < 1\}. \tag{XIII.24}$$

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 XIII.2

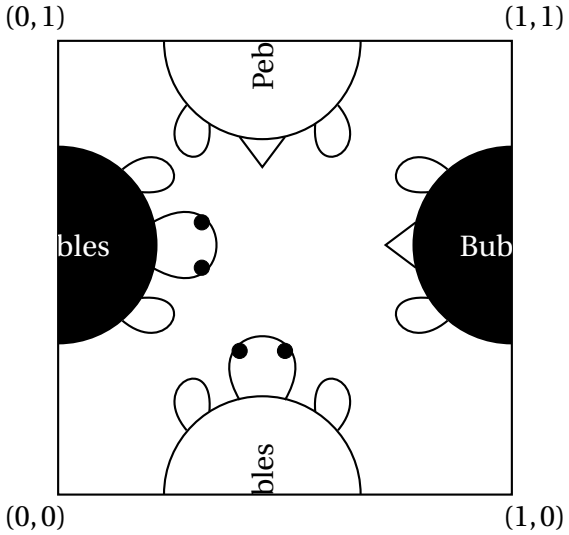


FIGURE XIII.2. $\mathbb{R}^2/\mathbb{Z}^2$ is really just the unit square with the top side glued to the bottom side, and the right side glued to the left side.

Example XIII.11. In this example, we'll find the elements of order 2 in $\mathbb{R}^2/\mathbb{Z}^2$. Suppose $\overline{(x,y)}$ is such an element, where x, y belong to the interval $[0,1)$. Then $\overline{(2x,2y)} = 0$ and so $2x$, $2y$ are integers. Hence

$$2x = \ldots, -1, 0, 1, 2, 3, \ldots, \qquad 2y = \ldots, -1, 0, 1, 2, 3, \ldots.$$

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

$$\overline{(x,y)} = \overline{(0,0)}, \quad \overline{(1/2,0)}, \quad \overline{(0,1/2)}, \quad \overline{(1/2,1/2)}.$$

However, the first of these has order 1. So the elements of order 2 in $\mathbb{R}^2/\mathbb{Z}^2$ are

$$\overline{(x,y)} = \overline{(1/2,0)}, \quad \overline{(0,1/2)}, \quad \overline{(1/2,1/2)}.$$

$\Diamond$

Exercise XIII.12. Find all elements of order 3 in $\mathbb{R}^2/\mathbb{Z}^2$ (there are eight of them).

Exercise XIII.13. In $\mathbb{Z}^2$ we let $\mathbf{i} = (1,0)$ and $\mathbf{j} = (0,1)$ as usual. Write $2\mathbb{Z}^2 = \{(2a,2b) : a,b \in \mathbb{Z}\}$. Convince yourself that $2\mathbb{Z}^2$ is a subgroup of $\mathbb{Z}^2$ of index 4 and that

$$\mathbb{Z}^2/2\mathbb{Z}^2 = \{\overline{\mathbf{0}}, \overline{\mathbf{i}}, \overline{\mathbf{j}}, \overline{\mathbf{i}+\mathbf{j}}\}.$$

Write down an addition table for $\mathbb{Z}^2/2\mathbb{Z}^2$.

Exercise XIII.14. How would you describe $\mathbb{C}/\mathbb{Z}[i]$? Is it really different from $\mathbb{R}^2/\mathbb{Z}^2$?

Exercise XIII.15. How would you describe $\mathbb{C}/\mathbb{Z}$? Find all elements of order 2.

XIII.5. $\mathbb{R}/\mathbb{Q}$

In this section, we shall briefly think about $\mathbb{R}/\mathbb{Q}$. In $\mathbb{R}/\mathbb{Z}$, we treat the integers as 'zero'. In $\mathbb{R}/\mathbb{Q}$, we treat the rationals as 'zero'. 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 (XIII.23). 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

$$\overline{a}+\overline{b} = \overline{a+b} = \overline{5/3} = \overline{0},$$

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

$$\overline{a}+\overline{b} = \overline{\pi+e}.$$

Can we simplify this? For example, is this equal to $\overline{0}$? 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 $\overline{0}$ or not.

XIII.6. Well-Defined and Proofs

We know that in $\mathbb{Z}/m\mathbb{Z}$, not only does addition make sense, but also multiplication makes sense. In $\mathbb{Z}/m\mathbb{Z}$ we defined multiplication by

$$\overline{a}\cdot\overline{b} = \overline{ab}. \tag{XIII.25}$$

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:

$$\overline{0.5}\times\overline{0.5} = \overline{0.25}, \qquad \overline{1.5}\times\overline{0.5} = \overline{0.75}.$$

There is a problem: in $\mathbb{R}/\mathbb{Z}$, the classes $\overline{1.5}$ and $\overline{0.5}$ are equal, but the classes $\overline{0.75}$ and $\overline{0.25}$ aren't. Multiplication doesn't make sense on $\mathbb{R}/\mathbb{Z}$.

The problem comes from the 'definition' for multiplication in (XIII.25). We're trying to define the product of $\overline{a}$ and $\overline{b}$ in terms of the representatives a, b of these classes. But each class 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* but it is worth looking at the proof again.

Lemma XIII.16. *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

$$\overline{ab} = \overline{a'b'}.$$

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

$$a'b' \equiv ab \pmod{m\mathbb{Z}},$$

which means

$$\overline{ab} = \overline{a'b'}.$$

□

Are we sure addition that addition is well-defined in $\mathbb{R}/\mathbb{Z}$ and the other quotient groups that we've been working with? The following lemma checks that.

Lemma XIII.17. *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*

$$\overline{a} = \overline{a'}, \qquad \overline{b} = \overline{b'},$$

then

$$\overline{a+b} = \overline{a'+b'}.$$

PROOF. Suppose $\overline{a} = \overline{a'}$ and $\overline{b} = \overline{b'}$ in G/H. Then $a - a' = h_1$ and $b - b' = h_2$ where h_1, $h_2 \in H$. Thus

$$(a+b) - (a'+b') = (a-a') + (b-b') = h_1 + h_2.$$

As H is a subgroup containing h_1 and h_2, we know that the sum $h_1 + h_2$ belongs to H. Thus the classes $\overline{a+b}$ and $\overline{a'+b'}$ are equal. □

To wrap up this chapter, we need to check one thing: that G/H is indeed a group.

Theorem XIII.18. *Let $(G,+)$ be an additive abelian group and H a subgroup. Then $(G/H,+)$ is an abelian group.*

PROOF. All we have to do is check the defining properties for abelian groups. I'll just check that addition is commutative and leave the rest to you. Suppose $a, b \in G$. Then

$$\begin{aligned} \overline{b}+\overline{a} &= \overline{b+a} && \text{from the definition of addition in } G/H \\ &= \overline{a+b} && b+a=a+b \text{ as } G \text{ is abelian} \\ &= \overline{a}+\overline{b} && \text{from the definition of addition in } G/H. \end{aligned} \tag{XIII.26}$$

□

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.

CHAPTER XIV

Symmetric Groups

The coolest groups have elements that are functions. Matrix groups are examples, and symmetric groups are other examples. It turns out that every finite group is a subgroup of one of the symmetric groups. So if we understand symmetric groups completely, then we'll understand finite groups completely. Have I given you hope that you'll have a complete understanding of finite groups by the end of the chapter? Sorry, *I was saying 'if…'*

XIV.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 $\mathrm{Map}(A)$ for the set of functions from A to itself. Then $\circ$ is a binary operation on $\mathrm{Map}(A)$. And it's natural to ask if this makes $\mathrm{Map}(A)$ into a group. After all, we know by Lemma III.2 that composition of functions is associative. The following example will help clarify these ideas.

Example XIV.1. Let $A = \{1,2\}$. You will quickly convince yourself that there are only four functions from A to itself, which are given in Figure XIV.1.

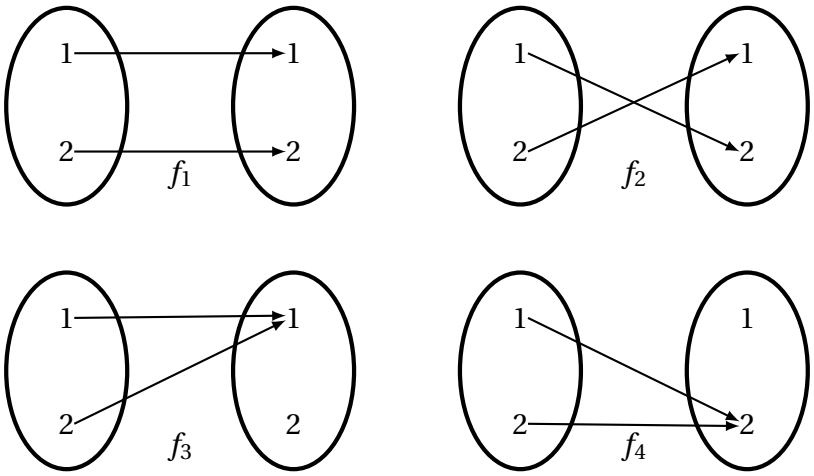


FIGURE XIV.1. f_1, f_2, f_3 and f_4 are the four functions from $\{1,2\}$ to itself.

Thus $\mathrm{Map}(A) = \{f_1, f_2, f_3, f_4\}$. Is $\mathrm{Map}(A)$ a group with respect to composition of functions? Here is the composition table for $\mathrm{Map}(A)$:

$\circ$	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 III.2. 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: $\{f_1, f_2\}$. If you've been paying attention in *Foundations* you will 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.

Now is a good time for you to revise Example IV.10. There you saw a that non-invertible matrix (which represented projection onto the y-axis) was also a non-bijective 'function' $\mathbb{R}^2 \to \mathbb{R}^2$. ◊

XIV.2. Injections, Surjections and Bijections

In this section we revise some stuff that you've done in Foundations. For the proofs you should refer back to your Foundations lecture notes.

Definition. Let A, B be sets, and let $f : A \to B$ be a function from A to B (also called a map or a mapping from A to B). We call A the *domain of* f and B the *codomain of* f or the *range of* f. We say f is *injective* if whenever a_1, $a_2 \in A$ satisfy $a_1 \neq a_2$ then $f(a_1) \neq f(a_2)$. In other words, distinct elements of A are mapped to distinct elements of B.

We say f is *surjective* if for every $b \in B$, there is some element $a \in A$ such that $f(a) = b$. In other words, every element of B is in the image.

We say f is *bijective* if it is injective and surjective.

Example XIV.2. See Figure XIV.2. Here f_1 is **not** a function, since $f_1(2) = b$ and $f_1(2) = c$. A function takes one element of the domain to exactly one element of the codomain. If you write $\sqrt{4} = \pm 2$, then you're thinking of $\sqrt{\ }$ as a *multifunction* and not a function. In terms of pictures as in the figure, for f to be a function, exactly one arrow originates at any one element of the domain.

f_2 is injective since there are exactly two distinct elements in its domain which are 1 and 2 and these get mapped to distinct elements c and a. However, f_2 is not surjective, since b is in the codomain, but $f(1) \neq b$

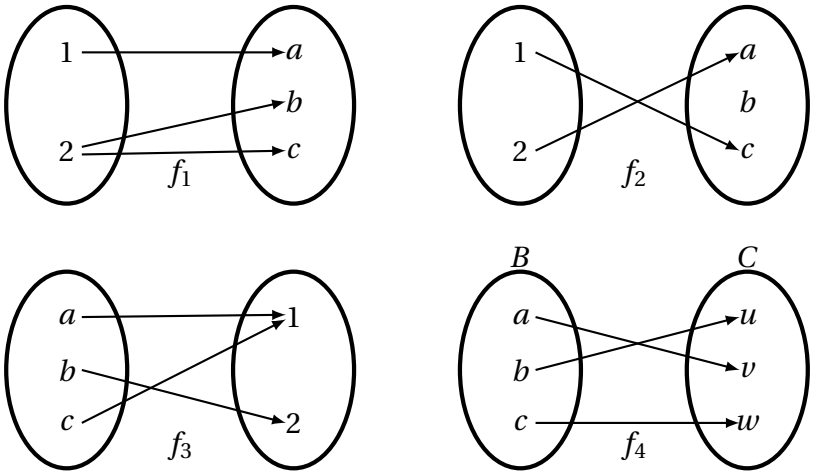


FIGURE XIV.2. f_1 is **not** a function, f_2 is injective but not surjective, f_3 is surjective but not injective and f_4 is a bijection.

and $f(2) \neq b$. In terms of pictures, surjective means that every element of the codomain is the end point of at least one arrow.

f_3 is surjective but not injective. In terms of pictures, injective means that no two arrows share the same end point.

bijection is relabelling

$f_4 : B \to C$ is a bijection. A bijection is merely an act of relabelling. The sets B and C are the same if we relabel a as v, b as u and c as w. ◊

Example XIV.3. Let f_1, f_2, f_3, f_4 be the functions $\{1,2\} \to \{1,2\}$ in Figure XIV.1. Then f_1 and f_2 are bijections. However, f_3 and f_4 are neither injective nor surjective. ◊

Remarks.

(i) Instead of saying that a function is injective, mathematicians sometimes say that it is *one-to-one* (also written $1-1$).
(ii) The definition of injective is often given in the contrapositive form: if $f(a_1) = f(a_2)$ then $a_1 = a_2$. The way we've phrased the definition is more helpful for this chapter, but you should get used to both forms.
(iii) Instead of saying that a function is surjective, mathematicians sometimes say that it is *onto*. I found this jarring when I first saw it. But I quickly got used to it.
(iv) A bijection is also called a *one-to-one correspondence*.

Here is a theorem you have seen in *Foundations* where it was probably called 'the pigeon-hole principle'.

Theorem XIV.4. *Let A be a finite set and let f be a function from A to itself. Then f is injective if and only if f is surjective.*

Example XIV.5. Look back at the functions in Figure XIV.1 for a very basic illustration of Theorem XIV.4. ◊

Example XIV.6. Theorem XIV.4 is true for *finite* sets only. For infinite sets it might or might not hold. Let $f_1 : \mathbb{N} \to \mathbb{N}$ be given by $f_1(x) = x + 1$. Then f_1 is not surjective since 0 is not in the image. However, f_1 is injective. See Figure XIV.3. By contrast, let $f_2 : \mathbb{Z} \to \mathbb{Z}$ be also given by $x \mapsto x + 1$. Then

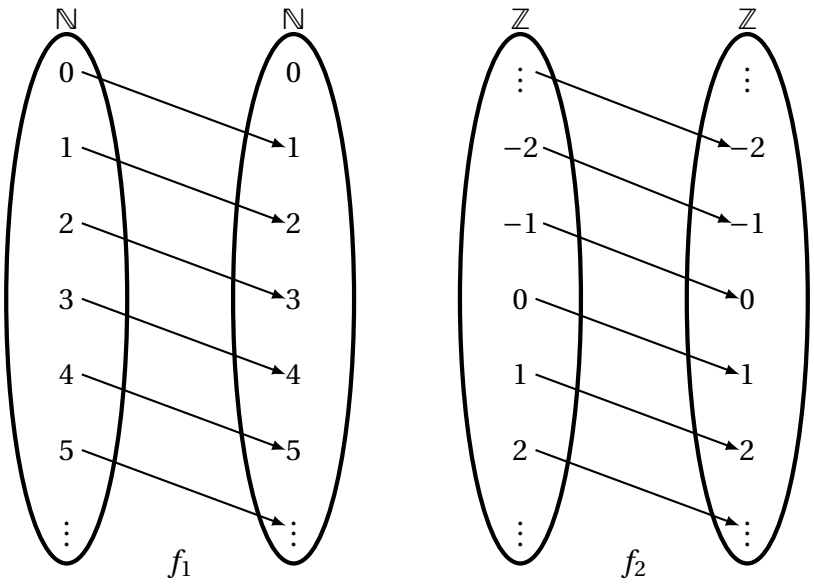


FIGURE XIV.3. $f_1 : \mathbb{N} \to \mathbb{N}$ and $f_2 : \mathbb{Z} \to \mathbb{Z}$ are both given by $x \mapsto x + 1$. The function f_1 is injective but not surjective. The function f_2 is injective and surjective; therefore it is a bijection. *The pigeon-hole principle holds only for finite sets!*

f_2 is a bijection. ◊

The following theorems collect together key results regarding bijections. Again you know all of this from *Foundations*.

Theorem XIV.7. *Let $f : A \to B$ and $g : B \to C$ be bijections. Then $g \circ f$ is a bijection $A \to C$.*

Definition. Let A be a set. The identity map on A is the map $\mathrm{id}_A : A \to A$ satisfying $\mathrm{id}_A(x) = x$ for all $x \in A$.

Let $f : A \to A$ be a function on A. We say that f is *invertible* if there exists a function $g : A \to A$ such that $f \circ g = g \circ f = \mathrm{id}_A$. If such a g exists then we call it the inverse of f and denote it by f^{-1}.

Of course you know from Foundations that if f has an inverse then it is unique.

Theorem XIV.8. *Let $f : A \to A$ be a function. Then f is invertible if and only if f is a bijection. If f is invertible then f^{-1} is also a bijection.*

XIV.3. The Symmetric Group

Let A be a set. We shall denote the set of bijections from A to itself by $\mathrm{Sym}(A)$.

Example XIV.9. In Example XIV.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 XIV.1. Hence $\mathrm{Sym}(A) = \{f_1, f_2\}$. Note that $f_1 = \mathrm{id}_A$. 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. ◊

Theorem XIV.10. *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 Theorem XIV.7, $\mathrm{Sym}(A)$ is closed under composition. Moreover, composition of functions is associative by Lemma III.2.

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 XIV.8. □

Example XIV.11. Let f_1, f_2 be as in Example XIV.6. Note that $f_1 \notin \mathrm{Sym}(\mathbb{N})$ since it is not a bijection. However $f_2 \in \mathrm{Sym}(\mathbb{Z})$. What is f_2^{-1}? It is simply the function $\mathbb{Z} \to \mathbb{Z}$ given by $x \mapsto x - 1$.

Let's calculate $g = f_2^3 = f_2 \circ f_2 \circ f_2$. Then

$$g(x) = f_2(f_2(f_2(x))) = f_2(f_2(x+1)) = f_2(x+2) = x+3.$$

It will be easy for you to show, for any integer n, that f_2^n is the function $\mathbb{Z} \to \mathbb{Z}$ satisfying $x \mapsto x + n$. In particular, $f_2^n \neq \mathrm{id}_A$ for $n \neq 0$. Thus f_2 is an element of infinite order in the group $\mathrm{Sym}(\mathbb{Z})$. ◊

Exercise XIV.12. 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.

Exercise XIV.13. 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?

XIV.4. 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 XIV.9 we found that S_2 is a group of order 2.

Theorem XIV.14. *S_n has order $n!$.*

PROOF. S_n is the set of bijections from $\{1,2,\dots,n\}$ to itself. So we want to count these bijections. Since the set $\{1,2,\dots,n\}$ is finite, Theorem XIV.4 tells us that a bijection from the set to itself is the same as an injection. So let's count the injections. Let f be an injection from $\{1,2,\dots,n\}$ to itself. Then $f(1)$ can be any of $1,2,\dots,n$; 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 XIV.1. 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.

Example XIV.15. 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 XIV.1. The first of these is the identity element. We noted in Example XIV.9 that $S_2 = \mathrm{Sym}(\{1,2\})$ is abelian. ◊

Example XIV.16. We know from Theorem XIV.14 that S_3 has 6 elements. These are

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$$
$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}.$$

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

$$\rho(1) = 3, \qquad \rho(2) = 1, \qquad \rho(3) = 2. \tag{XIV.27}$$

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 (XIV.27) 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

$$\rho^{-1} = \begin{pmatrix}1 & 2 & 3\\ 2 & 3 & 1\end{pmatrix}.$$

$\Diamond$

Exercise XIV.17. Write down a multiplication table for S_3 and determine the orders of all six elements checking that your answers are consistent with Lagrange's Theorem.

Exercise XIV.18. 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.

Exercise XIV.19. Show that S_n is non-abelian for $n \geq 3$.

Exercise XIV.20. Recall (page 34) that we interpreted elements of D_4 as functions from $\{1,2,3,4\}$ to itself. Go back and check that these are bijections. Thus D_4 is a subgroup of S_4.

XIV.4.1. What's Special About S_n? We started the chapter 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.

XIV.5. A Nice Application of Lagrange's Theorem

Let n, m be integers with $0 \leq m \leq n$. You met before the binomial coefficient

$$\binom{n}{m} = \frac{n!}{m!(n-m)!}.$$

It is not all obvious from this formula that the binomial coefficient is an integer. As an application of Lagrange's Theorem, we show that it is. Recall that S_n has order $n!$. If we can find a subgroup H of S_n of order $m!(n-m)!$, then

$$\binom{n}{m} = \frac{|S_n|}{|H|}.$$

We know by Lagrange's Theorem that the right-hand side is simply $[S_n : H]$, the number of cosets of H in S_n, and is therefore an integer. All we have to do now is give the subgroup H of order $m!(n-m)!$. Let H be the subset of S_n consisting of permutations σ such that σ permutes $\{1,2,\ldots,m\}$ and permutes $\{m+1, m+2, \ldots, n\}$. What do we mean by this? Write

$$A = \{1,2,\ldots,m\}, \qquad B = \{m+1, m+2, \ldots, n\}.$$

Note that

$$\{1,2,\ldots,n\} = A \cup B.$$

The elements of S_n are the bijections from the set $\{1,2,\ldots,n\}$ to itself. The elements of H are those elements σ of S_n that satisfy $\sigma(a) \in A$ for all $a \in A$, and $\sigma(b) \in B$ for all $b \in B$. It's an easy exercise to show that H is a subgroup of S_n. We want to check that its order is really $m!(n-m)!$. We count the elements of H in a similar way to the argument in the proof of Theorem XIV.14. Let σ be an element of H. Then $\sigma(1)$ can be any of $1,2,\ldots,m$. Once we've chosen $\sigma(1)$, we know $\sigma(2)$ can be any of $1,2,\ldots,m$ except for $\sigma(1)$. Thus there are m choices for $\sigma(1)$, $m-1$ choices for $\sigma(2)$

and so on. Until we reach $\sigma(m+1)$. This can be any element of B, and so there are $n-m$ choices for $\sigma(m+1)$. etc. You see that the order of H is

$$m(m-1)\cdots 1\cdot(n-m)(n-m-1)\cdots 1 = m!(n-m)!$$

Don't you just love maths?

Exercise XIV.21. To make sure you've understood the argument above, let $m=2$ and $n=4$, so that

$$A=\{1,2\}, \qquad B=\{3,4\}.$$

Now write down all permutations σ in S_4 that satisfy $\sigma(a)\in A$ for all $a\in A$ and $\sigma(b)\in B$ for all $b\in B$, and convince yourself that these form a group.

XIV.6. Cycle Notation

Let $a_1, a_2, \ldots, a_m$ be distinct elements of the set $\{1,2,\ldots,n\}$. By the notation

$$(a_1, a_2, \ldots, a_m) \tag{XIV.28}$$

we mean the element of S_n that takes a_1 to a_2, a_2 to a_3, …, 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 (XIV.28) is called a *cycle of length m*. A cycle of length 2 is called a *transposition*.

Example XIV.22. Let $\mu=(1,4,5)\in S_5$. The cycle μ is of length 3 and is illustrated in Figure XIV.4.

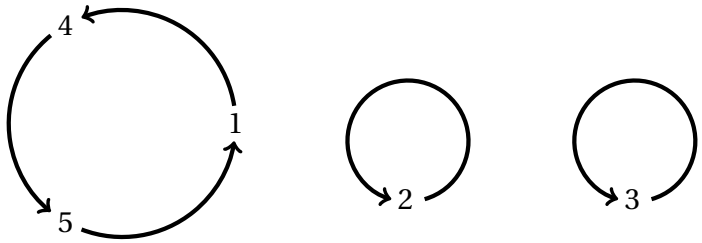


FIGURE XIV.4. The cycle $(1,4,5)\in S_5$.

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 XIV.5.

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 $(1,5)=(5,1)$.

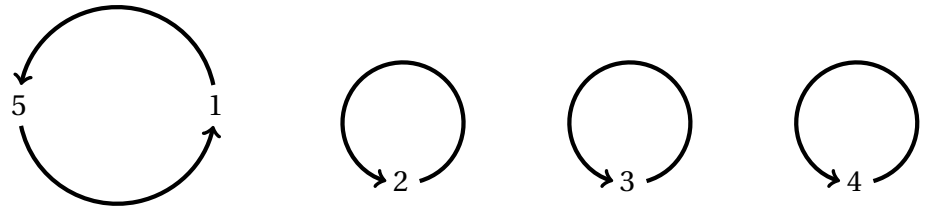


FIGURE XIV.5. The transposition $(1,5) \in S_5$. This merely swaps 1 and 5, and fixes all other elements.

Finally (1) is the cycle that takes 1 to itself and fixes all the other elements. Clearly (1) = (2) = (3) = (4) = (5) = id is nothing other than the identity permutation. ◊

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 will come as a pleasant surprize.

Theorem XIV.23. *Every permutation can be written as a product of disjoint cycles.*

What does *disjoint* mean? Two cycles $(a_1, a_2, \ldots, a_n)$ and $(b_1, b_2, \ldots, b_m)$ are said to be disjoint if $a_i \neq b_j$ for all i, j. What does *product* mean? The product of two permutations is of course their composition as functions. Before we prove the theorem, let's see an example where we write down a permutation as a product of cycles.

Example XIV.24. 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. ◊

Example XIV.25. 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 $\{1,2,\ldots,10\}$, 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 $S_{11}, S_{12}, \ldots$. 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:

$$\begin{aligned} \sigma: &\quad 1 \mapsto 3 \mapsto 10 \mapsto 9 \mapsto 1, \qquad 2 \mapsto 5 \mapsto 6 \mapsto 2; \\ \sigma^{-1}: &\quad 1 \leftarrow 3 \leftarrow 10 \leftarrow 9 \leftarrow 1, \qquad 2 \leftarrow 5 \leftarrow 6 \leftarrow 2. \end{aligned}$$

Therefore $\sigma^{-1} = (1,9,10,3)(2,6,5)$. Check for yourself that $\sigma\sigma^{-1}$ is indeed the identity permutation. ◊

Exercise XIV.26. Let ρ and τ be as given in Exercise XIV.18. Write ρ and τ as products of disjoint cycles.

Exercise XIV.27. 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 $(3,1,2)(5,4,6)$.

Exercise XIV.28. 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}, $\rho\tau$, $\tau\rho^2$.

Lemma XIV.29. *Disjoint cycles commute.* *as promised*

PROOF. Let σ and τ be disjoint cycle in S_n and write

$$\sigma = (a_1, a_2, \ldots, a_k), \qquad \tau = (b_1, b_2, \ldots, b_\ell).$$

Since σ and τ are disjoint $a_i \neq b_j$ for $i = 1, \ldots, k$ and $j = 1, \ldots, \ell$.

We want to show that $\sigma\tau = \tau\sigma$. This means that $\sigma\tau x = \tau\sigma x$ for all $x \in \{1, 2, \ldots, n\}$. 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 $i = 1, \ldots, k$. 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 $j = 1, \ldots, \ell$. This is similar to Case 2.

We conclude that $\sigma\tau = \tau\sigma$ as required. □

PROOF OF THEOREM XIV.23. Let ρ be an element of S_n. Consider the sequence

$$1, \quad \rho 1, \quad \rho^2 1, \quad \rho^3 1, \ldots$$

Every term in this infinite sequence is contained in the finite set $\{1, 2, \ldots, n\}$. 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 $1, \rho 1, \ldots, \rho^{u-1} 1$ are distinct.

Let μ_1 be the cycle of length u

$$\mu_1 = \left(1, \rho 1, \rho^2 1, \ldots, \rho^{u-1} 1\right).$$

It is clear that μ_1 has the same effect as ρ on the elements $1, \rho 1, \ldots, \rho^{u-1} 1$.

Now let a be the first element of the set $\{1, 2, \ldots, n\}$ not appearing in the list $1, \rho 1, \ldots, \rho^{u-1} 1$. Repeat the above argument with the sequence

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

We deduce the existence of a cycle

$$\mu_2 = \left(a, \rho a, \ldots, \rho^{v-1} a\right)$$

such that μ_2 and ρ have the same effect on the elements $a, \rho a, \ldots, \rho^{v-1}a$. Let us show that μ_1 and μ_2 are disjoint. Suppose otherwise. Then $\rho^i 1 = \rho^j a$ for some $0 \le i < u$ and $0 \le j < v$. 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

$$\rho = \mu_1\mu_2 \cdots \mu_r.$$

□

Exercise XIV.30. We will shortly meet the Alternating Groups, one of which is

$$A_3 = \{\mathrm{id}, (1,2,3), (1,3,2)\}.$$

Verify that A_3 is a subgroup of S_3, and write down its left cosets.

Exercise XIV.31. Verify that $H = \{\mathrm{id}, (1,2)\}$ is a subgroup of S_3, and write down its left cosets.

Exercise XIV.32.
(i) Use Lagrange's Theorem to show that S_4 does not have an element of order 5.
(ii) Let $\sigma = (a_1, a_2, \ldots, a_m)$ be a cycle of length m in S_n. Explain why σ has order m.
(iii) Now let $\rho = \sigma_1\sigma_2 \ldots \sigma_k$ where the σ_i are disjoint cycles of lengths m_i in S_n. Explain carefully why ρ has order $\mathrm{lcm}(m_1, m_2, \ldots, m_k)$.
(iv) Show that S_4 does not have elements of order 6. Could you have shown this using Lagrange's Theorem?

XIV.7. Permutations and Transpositions

Lemma XIV.33. *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

$$(a_1, a_2, \ldots, a_m) = (a_1, a_m) \cdots (a_1, a_3)(a_1, a_2). \tag{XIV.29}$$

□

Example XIV.34. Equation (XIV.29) 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 (XIV.29) 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. ♢

XIV.8. 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

$$\sigma(P_n) = \prod_{1 \leq i < j \leq n} (x_{\sigma i} - x_{\sigma j}).$$

Example XIV.35. 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 $\rho \in S_3$. ◊

Lemma XIV.36. *Let $\tau \in S_n$ be a transposition. Then $\tau(P_n) = -P_n$.*

The proof of the lemma is **not** hard. But it is very easy to get muddled in this proof. So before you read on, try out lots of examples and drink lots of coffee. Sit where no one can see you and try to prove it. If you manage it, *feel free to jump up and down from excitement*—you deserve it.

The examples we've done are quite basic, so let's do a more serious one.

Example XIV.37. 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 [1]:

(I)	$x_1 - x_2$,	$x_1 - x_4$,
(II)	$x_2 - x_3$,	$x_3 - x_4$,
(III)	$x_2 - x_5$,	$x_4 - x_5$,
(IV)	$x_2 - x_4$.	

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,

$$\tau(x_2 - x_4) = x_{\tau 2} - x_{\tau 4} = x_4 - x_2 = -(x_2 - x_4).$$

[1] The word "groups" here is used in its English language sense, not in its mathematical sense.

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: $\tau(P_5) = (-1)^3 P_5 = -P_5$. $\diamondsuit$

PROOF OF LEMMA XIV.36. 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. We pair off the other factors as follows:

$$\text{(I)} \quad \begin{cases} x_1 - x_\ell, & x_1 - x_m, \\ x_2 - x_\ell, & x_2 - x_m, \\ \vdots & \vdots \\ x_{\ell-1} - x_\ell, & x_{\ell-1} - x_m, \end{cases}$$

$$\text{(II)} \quad \begin{cases} x_\ell - x_{\ell+1}, & x_{\ell+1} - x_m, \\ x_\ell - x_{\ell+2}, & x_{\ell+2} - x_m, \\ \vdots & \vdots \\ x_\ell - x_{m-1}, & x_{m-1} - x_m, \end{cases}$$

$$\text{(III)} \quad \begin{cases} x_\ell - x_{m+1}, & x_m - x_{m+1}, \\ x_\ell - x_{m+2}, & x_m - x_{m+2}, \\ \vdots & \vdots \\ x_\ell - x_n, & x_m - x_n, \end{cases}$$

$$\text{(IV)} \quad \{x_\ell - x_m.$$

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 $\tau(P_n) = -P_n$. □

Lemma XIV.38. *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 XIV.33, that we can write every permutation as a product of transpositions. Every transposition changes the sign of P_n. The lemma follows. □

Example XIV.39. We have noted in Example XIV.34 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. ♢

You should now have no trouble in proving the following theorem.

Theorem XIV.40. *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****.*

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.

Example XIV.41. $(1,2,3,4)$ is an odd permutation because we can write it as the product of 3 transpositions:

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

Beware!

Indeed, a cycle of length n can be written as product of $n-1$ transpositions by (XIV.29). So 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 find that kind of reasoning disturbing, you have my sympathy. 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

$$A_n = \{\sigma \in S_n : \sigma \text{ is even}\}.$$

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.

Theorem XIV.42. *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 $\sigma, \rho \in A_n$ 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

$$\sigma = \tau_1\tau_2\ldots\tau_m$$

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. □

Example XIV.43. Recall that $S_2 = \{\mathrm{id}, (1,2)\}$. We see that $A_2 = \{\mathrm{id}\}$ is the trivial subgroup. ◊

Example XIV.44. 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. ◊

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.

Theorem XIV.45. *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 XII.17). So there aren't any other cosets and the proof is complete. □

Exercise XIV.46. Let ρ and τ be as given in Exercise XIV.18. Write ρ and τ as products of transpositions and state if they're even or odd.

Exercise XIV.47. Write down the elements of A_3 and check that it is cyclic (and hence abelian). Show that A_n is non-abelian for $n \geq 4$.

Exercise XIV.48. Let f be a polynomial in variables $x_1, x_2, \ldots, x_n$. Let σ be a permutation in S_n. We define $\sigma(f)$ to be the polynomial $f(x_{\sigma 1}, x_{\sigma 2}, \ldots, x_{\sigma n})$. For example, if $f = x_1 + x_2^2 + x_3x_4$ and $\sigma = (1,4)(2,3)$ then σ swaps x_1 and x_4, and swaps x_2 and x_3; thus $\sigma(f) = x_4 + x_3^2 + x_2x_1$. Compute $\sigma(f)$ for the following pairs f, σ:

(i) $f = x_1^2 - x_2x_3$, $\sigma = (1,2,3)$.
(ii) $f = x_1x_2 + x_3x_4$, $\sigma = (1,3)(2,4)$.

Exercise XIV.49. Let f be a polynomial in variables $x_1, \ldots, x_n$.

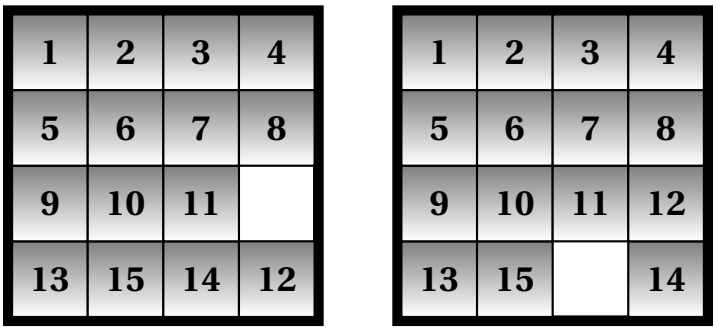
(a) Let H be a subgroup of S_n. We say that f is *H-invariant* if it satisfies the property that $\sigma(f) = f$ for all $\sigma \in H$. We say that f is *symmetric* if it is S_n-invariant. Find a polynomial in x_1, x_2, x_3, x_4 that is D_4-invariant but not symmetric.
(b) Define $\mathrm{Fix}(f) = \{\sigma \in S_n : \sigma(f) = f\}$. Show that $\mathrm{Fix}(f)$ is a subgroup of S_n. Write down $\mathrm{Fix}(f)$ for the following polynomials in $x_1, \ldots, x_4$:
  (i) $x_4^2 + x_1x_2x_3$.
  (ii) $x_1x_2 + x_3x_4$.

Exercise XIV.50. 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 $\rho = c_1c_2\cdots c_m$ as a product of transpositions, then** $\rho^{-1} = c_mc_{m-1}\ldots c_1$).

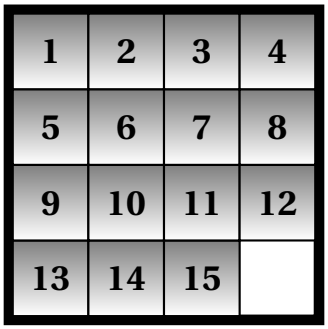
Exercise XIV.51. This exercise concerns the 15-tile puzzle. The puzzle consists of 15 square tiles (numbered $1, 2, \ldots, 15$) arranged in a 4×4 square with one position blank. The initial arrangement of the tiles is as follows:

1	2	3	4
5	6	7	8
9	10	11	12
13	15	14	

You can slide any tile adjacent to the blank into the position of the blank. So starting from the initial arrangement there are two possible moves:



In the 1880s—as a marketing ploy to improve the sales of the puzzle—Sam Lloyd (an amateur mathematician) offered $1000 to anyone who can reach:



Show that this is impossible! You might want to follow these hints and tips: *a mathematical scam*

(i) Think of the blank as a tile numbered 16. This way every rearrangement is a permutation on $1,\ldots,16$ and so an element of S_{16}.
(ii) Observe that every move is a transposition involving 16.
(iii) Observe that to go from the initial arrangement to the desired final arrangement, tile 16 must make the same number of moves down as up, and the same number of moves right as left.

Can you use your knowledge of maths to think of other ways of ripping people off? This of course is a purely intellectual exercise. As citizens of Warwick plc you are fine upright human beings who would not dream of putting such ideas into practice. But DON'T share these ideas with friends from less scrupulous universities. I don't have a particular two universities in mind. *applied maths!*

Riveted?

CHAPTER XV

Rings

If groups took your breath away, wait till you meet rings.

XV.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.

XV.2. Examples

Example XV.1. 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.

Example XV.2. 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. ◊

Theorem XV.3. *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. I leave you to ponder why the remaining properties (vi)–(ix) must be true. □

Example XV.4. 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,

$$\mathbf{a}\times\mathbf{b} = -\mathbf{b}\times\mathbf{a}. \tag{XV.30}$$

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 (XV.30) we see that $\mathbf{a} = -\mathbf{a}$ for all $\mathbf{a}\in\mathbb{R}^3$. This gives a contradiction. Therefore (viii) fails too. ◊

Example XV.5. 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

$$f \circ (g + h) = f(2x) = 4x^2; \qquad f \circ g + f \circ h = x^2 + x^2 = 2x^2.$$

$\Diamond$

Example XV.6. 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 $1 = 0$.

Example XV.7. 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 rather 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

$$(a_1, a_2) \times (b_1, b_2) = (a_1 b_1 - a_2 b_2, a_1 b_2 + a_2 b_1). \tag{XV.31}$$

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

$$(a_1 + ia_2) \times (b_1 + ib_2) = (a_1 b_1 - a_2 b_2) + i(a_1 b_2 + a_2 b_1). \tag{XV.32}$$

Notice that definitions (XV.31), (XV.32) 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 (XV.31)) is $(1, 0)$. Check for yourself that

$$(a_1, a_2) \times (1, 0) = (1, 0) \times (a_1, a_2) = (a_1, a_2).$$

This example is NOT IMPORTANT. Don't lose any sleep over it. $\Diamond$

XV.3. Subrings

Just as we have subgroups, so we have subrings.

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 $(R,+,\cdot)$).

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.

Example XV.8. $\mathbb{Z}$ 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 $\mathbb{R}[x]$. ♢

Theorem IX.5 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.

Theorem XV.9. *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. First re-read the proof of Theorem IX.5. Then prove this theorem on your own. It won't take you long. □

Example XV.10. In Example IX.7, 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 XV.9 we see that it isn't because $1 \notin 2\mathbb{Z}$. ♢

Example XV.11. 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 X.2 we saw that $\mathbb{Z}$ has infinitely many subgroups. ♢

Exercise XV.12. 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.

A very important tip!

Remark. 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.

Example XV.13. Let

$$\mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z}\}.$$

The set $\mathbb{Z}[i]$ is called the set of *Gaussian integers*. Show that $\mathbb{Z}[i]$ is a ring.
Answer. 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 $\alpha, \beta \in \mathbb{Z}[i]$. 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 XV.9 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 XV.9, we see that $\mathbb{Z}[i]$ is a subring of $\mathbb{C}$. Since $\mathbb{Z}[i]$ is a subring, it is a ring! ♢

Exercise XV.14. Let S be a subring of $\mathbb{Z}[i]$. Suppose $i \in S$. Show that $S = \mathbb{Z}[i]$.

Example XV.15. Let

$$S = \left\{ \frac{a}{2^r} : a, r \in \mathbb{Z}, r \geq 0 \right\}.$$

Show that S is a ring.
Answer. We shall follow 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 XV.9, S is a subring and therefore a ring. ◊

Exercise XV.16. 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 $\mathbb{Z}[i]$?

Exercise XV.17. 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) $\{A \in M_{2\times 2}(\mathbb{R}) : \det(A) = 1\}$.

XV.4. 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.

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

We found that $\mathbb{R}^*$ is group with respect to multiplication. In Example V.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.

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.

Example XV.18. In any non-zero ring, 0 is a non-unit. ◊

Example XV.19. In $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{C}$, every non-zero element has a multiplicative inverse. So the units are the non-zero elements. ◊

Example XV.20. 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. ◊

Example XV.21. 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.

Definition. Let R be a ring. We define the *unit group of* R to be the set [1]

$$R^* = \{a \in R : a \text{ is a unit in } R\}. \tag{XV.33}$$

Just because we've called R^* the *unit group of* R doesn't get us out of checking that it is really a group.

Lemma XV.22. *Let* $(R, +, \cdot)$ *be a ring and let* R^* *be the subset defined in* (XV.33). *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

$$u_1 v_1 = v_1 u_1 = 1, \qquad u_2 v_2 = v_2 u_2 = 1. \tag{XV.34}$$

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,

$$\begin{aligned}(u_1 u_2)(v_2 v_1) &= u_1 (u_2 v_2) v_1 && \text{associativity of multiplication}\\ &= u_1 \cdot 1 \cdot v_1 && \text{since } u_2 v_2 = 1\\ &= 1 && \text{since } u_1 v_1 = 1.\end{aligned}$$

Similarly $(v_2 v_1)(u_1 u_2) = 1$. Thus $u_1 u_2$ is a unit [2] in R, and so $u_1 u_2 \in R^*$. We've proved that R^* is closed under multiplication.

---

[1] Functorially-inclined mathematicians write $R^\times$ instead of R^*. I happen to be functorially-disinclined, but I do use their notation when I'm feeling pretentious.

[2] Start again. We have u_1, u_2 are units and so satisfy (XV.34) 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.

offence intended Do I distress you by repeatedly exhibiting such offences against mathematical decency? Do you feel that these notes are degenerating into page after page of perversion and blasphemy? I am sorry; I simply want you to join me in condemning these abominations.

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. □

Example XV.23. Note that $\mathbb{R}^*$, $\mathbb{C}^*$, $\mathbb{Q}^*$ have exactly the same meaning as before. ◊

Example XV.24. We showed that the units of $\mathbb{Z}$ are ± 1. Therefore the unit group of $\mathbb{Z}$ is

$$\mathbb{Z}^* = \{1, -1\}.$$

◊

Example XV.25. 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 $A^{-1} = \left(\begin{smallmatrix} 1/2 & -1/2 \\ -1/2 & 3/2 \end{smallmatrix}\right)$. 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

$$(M_{2\times 2}(\mathbb{Z}))^* = \{A \in M_{2\times 2}(\mathbb{Z}) : \det(A) = \pm 1\}.$$

We define the group $\mathrm{GL}_2(\mathbb{Z})$ by

$$\mathrm{GL}_2(\mathbb{Z}) = \{A \in M_{2\times 2}(\mathbb{Z}) : \det(A) = \pm 1\};$$

then $(M_{2\times 2}(\mathbb{Z}))^* = \mathrm{GL}_2(\mathbb{Z})$. In fact, for a *commutative* ring R we define

$$\mathrm{GL}_2(R) = \left\{A \in M_{2\times 2}(R) : \det(A) \in R^*\right\}.$$

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, $(M_{2\times 2}(R))^* = \mathrm{GL}_2(R)$.

Example XV.26. Let

$$S = \left\{\begin{pmatrix} a & b \\ 0 & c \end{pmatrix} : a, b, c \in \mathbb{Z}\right\}.$$

Show that S is a ring under the usual addition and multiplication of matrices. Compute S^*.

Answer: 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 easy 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

$$S^* = \left\{\begin{pmatrix} \pm 1 & b \\ 0 & \pm 1 \end{pmatrix} : b \in \mathbb{Z}\right\}.$$

$\Diamond$

Exercise XV.27. In Example XV.15, 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.

XV.5. 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 $\pm 1 \cdot p_1^{r_1} \ldots p_n^{r_n}$ 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 look forward to doing this in *Algebra II*. For now, we want to determine the unit

a profound example

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:

Lemma XV.28. *Let $\alpha, \beta \in \mathbb{Z}[i]$. 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. □

Theorem XV.29. *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 [1] $\alpha\beta = 1$. 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

$$\mathbb{Z}[i]^* = \{1, -1, i, -i\}.$$

□

Remark. Compare the above proof to our determination of the unit group of $M_{2\times 2}(\mathbb{Z})$ in Example XV.25. I hope you agree that the similarities are striking!

Exercise XV.30. In Exercise XV.16 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 $\mathbb{Z}[i]$.)

Exercise XV.31. 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 $\mathbb{Z}[\sqrt{2}]^*$?

Exercise XV.32. 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.

[1] We could have written $\alpha\beta = \beta\alpha = 1$. But $\mathbb{Z}[i]$ is a *commutative* ring, so writing $\alpha\beta = 1$ is enough.

(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.

CHAPTER XVI

Fields

A *field* $(F, +, \cdot)$ is a commutative 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

$$F^* = \{a \in F : a \neq 0\}.$$

Example XVI.1. $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Q}$ are fields. $\Diamond$

Example XVI.2. $\mathbb{Z}$ is not a field, since for example $2 \in \mathbb{Z}$ is non-zero but not a unit. $\Diamond$

Example XVI.3. $\mathbb{R}[x]$ is not a field, since for example $x \in \mathbb{R}[x]$ is non-zero but not a unit. $\Diamond$

Example XVI.4. Show that

$$\mathbb{Q}[i] = \{a + bi : a, b \in \mathbb{Q}\}$$

is a field.

Answer: 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 [2].

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, $b \in \mathbb{Q}$, 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

[2] We could've made the proof more tedious 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 why do it? Get used to thinking of rational numbers as numbers in their own right!

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. ♢

Exercise XVI.5. Let $\mathbb{Q}[\sqrt{2}] = \{a+b\sqrt{2} : a,b \in \mathbb{Q}\}$. Show that $\mathbb{Q}[\sqrt{2}]$ is a field.

Exercise XVI.6. Let

$$F = \left\{\left(\begin{smallmatrix} a & b \\ -b & a\end{smallmatrix}\right) : a,b \in \mathbb{R}\right\}.$$

(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 $\phi\left(\begin{smallmatrix} a & b \\ -b & a\end{smallmatrix}\right) = a+bi$. 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

$$F' = \left\{\left(\begin{smallmatrix} a & b \\ -b & a\end{smallmatrix}\right) : a,b \in \mathbb{C}\right\}$$

is not a field.

CHAPTER XVII

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 $\mathbb{Z}/m\mathbb{Z}$.

XVII.1. Units in $\mathbb{Z}/m\mathbb{Z}$

Example XVII.1. Find the unit groups of $\mathbb{Z}/m\mathbb{Z}$ for $m = 2, 3, 4, 5, 6$.
Answer: You don't have be very clever here! Just look at the multiplication table for $\mathbb{Z}/6\mathbb{Z}$ in Example VII.3 and you'll see that

$$(\mathbb{Z}/6\mathbb{Z})^* = \{\overline{1}, \overline{5}\}.$$

In the same way you'll find that

$$(\mathbb{Z}/2\mathbb{Z})^* = \{\overline{1}\}, \qquad (\mathbb{Z}/3\mathbb{Z})^* = \{\overline{1}, \overline{2}\},$$
$$(\mathbb{Z}/4\mathbb{Z})^* = \{\overline{1}, \overline{3}\}, \qquad (\mathbb{Z}/5\mathbb{Z})^* = \{\overline{1}, \overline{2}, \overline{3}, \overline{4}\}.$$

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? ◊

Theorem XVII.2. *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*

$$(\mathbb{Z}/m\mathbb{Z})^* = \{\overline{a} : 0 \le a \le m-1 \text{ and } \gcd(a, m) = 1\}.$$

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 Euclid's Algorithm, 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. □

Exercise XVII.3. Redo Example XVII.1 using Theorem XVII.2.

Example XVII.4. By Theorem XVII.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

crucial point

$\gcd(19,256) = 1$, but let's use Euclid's Algorithm [1] to write 1 as a linear combination of 19 and 256:

$$\mathbf{256} = 13 \times \mathbf{19} + \mathbf{9}$$
$$\mathbf{19} = 2 \times \mathbf{9} + \mathbf{1}.$$

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 $\mathbb{Z}/256\mathbb{Z}$. ◊

XVII.2. Fermat's Little Theorem

Through the computations you've done so far, you've probably conjectured the following.

Theorem XVII.5. *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 XVII.2, $\overline{a}$ is invertible in $\mathbb{Z}/p\mathbb{Z}$. This shows that $\mathbb{Z}/p\mathbb{Z}$ is a field. □

Exercise XVII.6. Prove the converse of Theorem XVII.5: if $\mathbb{Z}/m\mathbb{Z}$ is a field then m is prime.

Theorem XVII.7. *(Fermat's Little Theorem) Let p be a prime and a an integer such that $p \nmid a$. Then*

$$a^{p-1} \equiv 1 \pmod{p}. \tag{XVII.35}$$

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 XVII.5, $\overline{b}$ 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 VIII.13 (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 (XVII.35) □

[1] 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.

Here's a fun application of Fermat's Little Theorem.

Example XVII.8. Compute $2^{1000} \pmod{13}$.
Answer: 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,

$$2^{1000} = 2^{83\times 12+4} = (2^{12})^{83} \times 2^4 \equiv 1^{83} \times 16 \equiv 3 \pmod{13}.$$

$\Diamond$

XVII.3. Euler's Theorem

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.*

Example XVII.9. We know that if p is a prime, then $(\mathbb{Z}/p\mathbb{Z})^* = \{\overline{1}, \overline{2}, \ldots, \overline{p-1}\}$, and so $\varphi(p) = p - 1$. $\Diamond$

Example XVII.10. We know that

$$(\mathbb{Z}/6\mathbb{Z})^* = \{\overline{1}, \overline{5}\},$$

and so $\varphi(6) = 2$. $\Diamond$

Example XVII.11. 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

$$(\mathbb{Z}/2^n\mathbb{Z})^* = \{\overline{1}, \overline{3}, \ldots, \overline{2^n - 1}\}.$$

Hence $\varphi(2^n) = 2^{n-1}$. $\Diamond$

Theorem XVII.12. *(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 easy exercise. $\square$

You're probably wondering if there is a formula for $\varphi(m)$, and in fact there is.

Proposition XVII.13. *Write*

$$m = p_1^{r_1} \cdots p_k^{r_k}$$

where $p_1, \ldots, p_k$ are distinct primes and $r_1, \ldots, r_k$ 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 XVII.13 is not difficult, but it is a little long and we shall skip it.

Exercise XVII.14. Use Euler's Theorem to compute $2^{1000} \pmod{63}$.

Exercise XVII.15. 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.

Exercise XVII.16. Use Lagrange's Theorem to show that $\varphi(m)$ is even for $m \geq 3$.

XVII.4. *Vale Dicere*

With tear-filled eyes I say goodbye, and begin to suffer the heart-rending pangs of separation ... *a sincere outpouring of grief*

Exercise XVII.17. Write a 5000 word essay on how abstract algebra has changed your outlook on life, detailing the insights you have gained into the great challenges facing/menacing humanity [1].

Exercise XVII.18. "Veneration of abstract algebra is at the root of contemporary mathematical decline." Discuss.

[1] Here's a good way to start your essay: "Unbeknownst to me, I misspent the first 18 years of my life wallowing in a cesspool of intellectual stagnation, until week six of term 1 when the *Abstract Algebra* lectures started ...".