Appendix

Basic Mathematical Terminology

This appendix is presented as a series of notes which summarize most of the mathematical terminology needed in this book. We present the material in a more formal manner than we did in Chapter 1 and the rest of the book.

A.1. Sets

Here we recap some basic definitions etc. which we list here for completeness.

Definition 100.1 (Set Union, Intersection, Difference and Cartesian Product)**.** *For two sets A, B we define the union, intersection, difference and Cartesian product by*

$$\begin{aligned} A \cup B &= \{x : x \in A \textit{ or } x \in B\}, \\ A \cap B &= \{x : x \in A \textit{ and } x \in B\}, \\ A \setminus B &= \{x : x \in A \textit{ and } x \notin B\}, \\ A \times B &= \{(x, y) : x \in A \textit{ and } y \in B\}. \end{aligned}$$

The statement $A \subseteq B$ means that for all $x \in A$ it follows that $x \in B$.

Using these definitions one can prove in a standard way all the basic results of set theory that one shows in school using Venn diagrams.

Lemma 100.2. *If $A \subseteq B$ and $B \subseteq C$ then $A \subseteq C$.*

Proof. Let x be an element of A; we wish to show that x is an element of C. Now as $A \subseteq B$ we have that $x \in B$, and as $B \subseteq C$ we then deduce that $x \in C$. □

Notice that this is a proof whereas an argument using Venn diagrams to demonstrate something is not a proof. Using Venn diagrams to show something merely shows you were not clever enough to come up with a picture which proved the result false.

There are some standard sets which will be of interest in our discussions: $\mathbb{N}$ the set of natural numbers, $\{0, 1, 2, 3, 4, \ldots\}$; $\mathbb{Z}$ the set of integers, $\{0, \pm 1, \pm 2, \pm 3, \pm 4, \ldots\}$; $\mathbb{Q}$ the set of rational numbers, $\{p/q : p \in \mathbb{Z}, q \in \mathbb{N} \setminus \{0\}\}$; $\mathbb{R}$ the set of real numbers; $\mathbb{C}$ the set of complex numbers.

A.2. Relations

Next we define relations and some properties that they have. Relations, especially equivalence relations, play an important part in algebra and it is worth considering them at this stage so it is easier to understand what is going on later.

Definition 100.3 (Relation)**.** *A (binary) relation on a set A is a subset of the Cartesian product $A \times A$.*

This we explain with an example: Consider the relationship "less than or equal to" between natural numbers. This obviously gives us the set

$$\mathsf{LE} = \{(x, y) : x, y \in \mathbb{N},\ x \text{ is less than or equal to } y\}.$$

N.P. Smart, *Cryptography Made Simple*, Information Security and Cryptography, DOI 10.1007/978-3-319-21936-3

In much the same way every relationship that you have met before can be written in this set-theoretic way. An even better way to put the above is to define the relation "less than or equal to" to be the set

$$\mathsf{LE} = \{(x, y) : x, y \in \mathbb{N},\ x - y \notin \mathbb{N} \setminus \{0\}\}.$$

Obviously this is a very cumbersome notation so for a relation R on a set S we write

$$x\ R\ y$$

if $(x, y) \in R$, i.e. if we now write $\leq$ for LE we obtain the usual notation $1 \leq 2$ etc. Relations which are of interest in mathematics usually satisfy one or more of the following four properties.

Definition 100.4 (Properties of Relations)**.**

- *A relation R on a set S is reflexive if for all $x \in S$ we have $(x, x) \in R$.*
- *A relation R on a set S is symmetric if $(x, y) \in R$ implies that $(y, x) \in R$.*
- *A relation R on a set S is anti-symmetric if $(x, y) \in R$ and $(y, x) \in R$ implies that $x = y$.*
- *A relation R on a set S is transitive if $(x, y) \in R$ and $(y, z) \in R$ implies that $(x, z) \in R$.*

We return to our example of $\leq$. This relation $\leq$ is certainly reflexive as $x \leq x$ for all $x \in \mathbb{N}$. It is not symmetric as $x \leq y$ does not imply that $y \leq x$, however it is anti-symmetric as $x \leq y$ and $y \leq x$ imply that $x = y$. You should note that it is transitive as well.

Relations like $\leq$ occur so frequently that we give them a name.

Definition 100.5 (Partial Order Relation)**.** *A relation which is a partial order relation if it is reflexive, transitive and anti-symmetric.*

Definition 100.6 (Total Order Relation)**.** *A relation which is transitive and anti-symmetric and for which for all x and y, with $x \neq y$, we have either $(x, y) \in R$ or $(y, x) \in R$ is called a total order relation.*

Whilst every total order relation is a partial order relation, the converse is not true. For example consider the relation of

$$\mathsf{div} = \{(x, y) : x, y \in \mathbb{N}, x \text{ divides } y\}.$$

This is clearly a partial ordering, since

- It is reflexive, as x divides x.
- It is transitive, as x divides y and y divides z implies x divides z.
- It is anti-symmetric, as if x divides y and y divides x then $x = y$.

But it is clearly not a total order as 3 does not divide 4 and 4 does not divide 3.

Another important type of relationship is that of an equivalence relation.

Definition 100.7 (Equivalence Relation)**.** *A relation which is reflexive, symmetric and transitive is called an equivalence relation.*

The obvious example of $\mathbb{N}$ and the relation "is equal to" is an equivalence relation and hence gives this type of relation its name. One of the major problems in any science is that of classification of sets of objects. This amounts to placing the objects into mutually disjoint subsets. An equivalence relation allows us to place elements into disjoint subsets. Each of these subsets is called an equivalence class. If the properties we are interested in are constant over each equivalence class then we may as well restrict our attention to the equivalence classes themselves. This often leads to greater understanding. In the jargon this process is called factoring out by the equivalence relation. It occurs frequently in algebra to define new objects from old, e.g. quotient groups. The following example is probably the most familiar; being a description of modular arithmetic.

Let m be a fixed positive integer. Consider the equivalence relation on $\mathbb{Z}$ which says x is related to y if $(x - y)$ is divisible by m. This is an equivalence relation, which you should check. The equivalence classes we denote by

$$\begin{aligned}\overline{0} &= \{\ldots, -2\cdot m, -m, 0, m, 2\cdot m, \ldots\},\\ \overline{1} &= \{\ldots, -2\cdot m+1, -m+1, 1, m+1, 2\cdot m+1, \ldots\},\\ \cdots &\quad \cdots\\ \overline{m-1} &= \{\ldots, -m-1, -1, m-1, 2\cdot m-1, 3\cdot m-1, \ldots\}.\end{aligned}$$

Note that there are m distinct equivalence classes, one for each of the possible remainders on division by m. The classes are often called the residue classes modulo m. The resulting set $\{\overline{0}, \ldots, \overline{m-1}\}$ is often denoted by $\mathbb{Z}/m\mathbb{Z}$ as we have divided out by all multiples of m. If m is a prime number, say p, then the resulting set is often denoted $\mathbb{F}_p$ as the resulting object is a field.

A.3. Functions

We give two definitions of functions; the first is wordy and is easier to get hold of, the second is set-theoretic.

Definition 100.8 (Function – v1)**.** *A function is a rule which maps the elements of one set, the domain, to those of another, the codomain. Each element in the domain must map to one and only one element in the codomain (a.k.a. the range of the function).*

The point here is that the function is not just the rule, e.g. $f(x) = x^2$, but also the two sets that one is using. A few examples will suffice.

(1) The rule $f(x) = \sqrt{x}$ is not a function from $\mathbb{R}$ to $\mathbb{R}$ since the square root of a negative number is not in $\mathbb{R}$. It is also not a function (depending on how you define the $\sqrt{}$ symbol) from $\mathbb{R}_{\geq 0}$ to $\mathbb{R}$ since every element of the domain has two square roots in the codomain. But it is a function from $\mathbb{R}_{\geq 0}$ to $\mathbb{R}_{\geq 0}$.
(2) The rule $f(x) = 1/x$ is not a function from $\mathbb{R}$ to $\mathbb{R}$ but it is a function from $\mathbb{R} \setminus \{0\}$ to $\mathbb{R}$.
(3) Note that not every element of the codomain need have an element mapping to it. Hence, the rule $f(x) = x^2$ taking elements of $\mathbb{R}$ to elements of $\mathbb{R}$ is a function.

Our definition of a function is unsatisfactory as it would also require a definition of what a rule is. In keeping with the spirit of everything else we have done we give a set-theoretic description.

Definition 100.9 (Function – v2)**.** *A function from the set A to the set B is a subset F of $A \times B$ such that:*

(1) *If $(x, y) \in F$ and $(x, z) \in F$ then $y = z$.*
(2) *For all $x \in A$ there exists a $y \in B$ such that $(x, y) \in F$.*

The set A is called the domain, the set B the codomain. The first condition means that each element in the domain maps to at most one element in the codomain. The second condition means that each element of the domain maps to at least one element in the codomain. Given a function f from A to B and an element x of A then we denote by $f(x)$ the unique element in B such that $(x, f(x)) \in f$.

Composition of Functions: One can compose functions, if the definitions make sense. Say one has a function f from A to B and a function g from B to C, then the function $g \circ f$ is the function with domain A and codomain C consisting of the elements $(x, g(f(x)))$.

Lemma 100.10. *Let f be a function from A to B, let g be a function from B to C and let h be a function from C to D, then we have*

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

PROOF. Let (a, d) belong to $(h \circ g) \circ f$. Then there exists an $(a, b) \in f$ and a $(b, d) \in (h \circ g)$ for some $b \in B$, by definition of composition of functions. Again by definition there exists a $c \in C$

such that $(b,c) \in g$ and $(c,d) \in h$. Hence $(a,c) \in (g \circ f)$, which shows $(a,d) \in h \circ (g \circ f)$. Hence

$$(h \circ g) \circ f \subseteq h \circ (g \circ f).$$

Similarly one can show the other inclusion. □

One function, the identity function, is particularly important.

Definition 100.11 (Identity Function)**.** *The identity function* id_A *from a set* A *to itself, is the set* $\{(x,x) : x \in A\}$.

Lemma 100.12. *For any function* f *from* A *to* B *we have*

$$f \circ \mathrm{id}_A = \mathrm{id}_B \circ f = f.$$

Proof. Let x be an element of A, then

$$(f \circ \mathrm{id}_A)(x) = f(\mathrm{id}_A(x)) = f(x) = \mathrm{id}_B(f(x)) = (\mathrm{id}_B \circ f)(x).$$

□

Injective, Surjective and Bijective Functions: Two properties that we shall use all the time are the following.

Definition 100.13 (Injective and Surjective)**.**

A function f *from* A *to* B *is said to be injective (or 1:1) if for any two elements,* x, y *of* A *with* $f(x) = f(y)$ *we have* $x = y$.

A function f *from* A *to* B *is said to be surjective (or onto) if for every element* $b \in B$ *there exists an element* $a \in A$ *such that* $f(a) = b$.

A function which is both injective and surjective is called bijective (or a 1:1 correspondence). We shall now give some examples.

(1) The function from $\mathbb{R}$ to $\mathbb{R}$ given by $f(x) = x + 2$ is bijective.
(2) The function from $\mathbb{N}$ to $\mathbb{N}$ given by $f(x) = x + 2$ is injective but not surjective as the elements $\{0, 1\}$ are not the image of anything.
(3) The function from $\mathbb{R}$ to $\mathbb{R}_{\geq 0}$ given by $f(x) = x^2$ is surjective as every non-negative real number has a square root in $\mathbb{R}$ but it is not injective as if $x^2 = y^2$ then we could have $x = -y$.

The following gives us a good reason to study bijective functions.

Lemma 100.14. *A function* $f : A \rightarrow B$ *is bijective if and only if there exists a function* $g : B \rightarrow A$ *such that* $f \circ g$ *and* $g \circ f$ *are the identity function.*

We leave the proof of this lemma as an exercise. Note that applying this lemma to the resulting g means that g is also bijective. Such a function as g in the above lemma is called the inverse of f and is usually denoted f^{-1}. Note that a function only has an inverse if it is bijective.

A.4. Permutations

We let A be a finite set of cardinality n; without loss of generality we can assume that $A = \{1, 2, \ldots, n\}$. A bijective function from A to A is called a permutation. The set of all permutations on a set of cardinality n is denoted by S_n.

Suppose $A = \{1, 2, 3\}$, then we have the permutation $f(1) = 2$, $f(2) = 3$ and $f(3) = 1$. This is a very cumbersome way to write a permutation. Mathematicians (being lazy people) have invented the following notation: the function f above is written as

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

What should be noted about this notation (which applies for arbitrary n) is that all the numbers between 1 and n occur exactly once on each row. The first row is always given as the numbers 1 to n in increasing order. Any such matrix with these properties represents a permutation, and all permutations can be represented by such a matrix. This leads us to the following elementary result.

Lemma 100.15. *The cardinality of the set S_n is $n!$.*

PROOF. This is a well-known argument. There are n choices for the first element in the second row of the above matrix. Then there are $n-1$ choices for the second element in the second row and so on. □

If σ is a permutation on a set S then we usually think of σ acting on the set. So if $s \in S$ then we write s^σ or $\sigma(s)$ for the action of σ on the element s.

Suppose we define the permutations

$$\begin{aligned} g &= \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \\ f &= \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}. \end{aligned}$$

As permutations are nothing but functions we can compose them. Remembering that $g \circ f$ means apply the function f and then apply the function g we see that

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$$

means $1 \rightarrow 3 \rightarrow 1$, $2 \rightarrow 2 \rightarrow 3$ and $3 \rightarrow 1 \rightarrow 2$. Hence, the result of composing the above two permutations is

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}. \tag{24}$$

However, this can cause confusion when using our "acting on a set" notation above. For example

$$1^{g \circ f} = g(f(1)) = 3$$

so we are unable to read the permutation from left to right. However, if we use another notation, say $\cdot$, to mean

$$f \cdot g = g \circ f$$

then we are able to read the expression from left to right. We shall call this operation multiplying permutations.

Cycle Notation: Mathematicians, as we said, are by nature lazy people and this notation we have introduced is still a little too much. For instance we always write down the numbers $1, \dots, n$ in the top row of each matrix to represent a permutation. Also some columns are redundant, for instance the first column of the permutation in equation (24). We now introduce another notation for permutations which is concise and clear. We first need to define what a cycle is.

Definition 100.16 (Cycle)**.** *By a cycle or n-cycle we mean the object $(x_1, \dots, x_n)$ with distinct $x_i \in \mathbb{N} \setminus \{0\}$. This represents the permutation $f(x_1) = x_2$, $f(x_2) = x_3$, $\dots, f(x_{n-1}) = x_n$, $f(x_n) = x_1$ and for $x \notin \{x_1, \dots, x_n\}$ we have $f(x) = x$.*

For instance we have

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} = (1, 2, 3) = (2, 3, 1) = (3, 1, 2).$$

Notice that a cycle is not a unique way of representing a permutation. Most permutations cannot be written as a single cycle, but they can be written as a product of cycles. For example we have

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} = (1,3) \circ (2) = (3,1) \circ (2).$$

The identity permutation is represented by (). Again, as mathematicians are lazy we always write $(1,3) \circ (2) = (1,3)$. This can lead to ambiguities as $(1,2)$ could represent a function from

$$\{1,2\} \text{ to } \{1,2\}$$

or

$$\{1,2,\dots,n\} \text{ to } \{1,2,\dots,n\}.$$

However, which function it represents is usually clear from the context.

Two cycles $(x_1,\dots,x_n)$ and $(y_1,\dots,y_n)$ are called disjoint if $\{x_1,\dots,x_n\} \cap \{y_1,\dots,y_n\} = \emptyset$. It is easy to show that if σ and τ are two disjoint cycles then

$$\sigma \cdot \tau = \tau \cdot \sigma.$$

Note that this is not true for cycles which are not disjoint, e.g.

$$(1,2,3,4) \cdot (3,5) = (1,2,5,3,4) \neq (1,2,3,5,4) = (3,5) \cdot (1,2,3,4).$$

Our action of permutations on the underlying set can now be read easily from left to right,

$$2^{(1,2,3,4)\cdot(3,5)} = 3^{(3,5)} = 5 = 2^{(1,2,5,3,4)},$$

as the permutation $(1,2,3,4)$ maps 2 to 3 and the permutation $(3,5)$ maps 3 to 5.

What really makes disjoint cycles interesting is the following.

Lemma 100.17. *Every permutation can be written as a product of disjoint cycles.*

Proof. Let σ be a permutation on $\{1,\dots,n\}$. Let σ_1 denote the cycle

$$(1, \sigma(1), \sigma(\sigma(1)), \dots, \sigma(\dots\sigma(1)\dots)),$$

where we keep applying σ until we get back to 1. We then take an element x of $\{1,\dots,n\}$ such that $\sigma_1(x) = x$, if one exists, and consider the cycle σ_2 given by

$$(x, \sigma(x), \sigma(\sigma(x)), \dots, \sigma(\dots\sigma(x)\dots)).$$

We then take an element of $\{1,\dots,n\}$ which is fixed by σ_1 and σ_2 to create a cycle σ_3. We continue this way until we have used all elements of $\{1,\dots,n\}$. The resulting cycles $\sigma_1,\dots,\sigma_t$ are obviously disjoint and their product is equal to the cycle σ. □

What is nice about this proof is that it is constructive. Given a permutation we can follow the procedure in the proof to obtain the permutation as a product of disjoint cycles. Consider the permutation

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

We have $\sigma(1) = 2,\ \sigma(2) = 3,\ \sigma(3) = 7$ and $\sigma(7) = 1$ so the first cycle is

$$\sigma_1 = (1,2,3,7).$$

The next element of $\{1,\dots,9\}$ which we have not yet considered is 4. We have $\sigma(4) = 6$ and $\sigma(6) = 4$ so $\sigma_2 = (4,6)$. Continuing in this way we find $\sigma_3 = (5,8)$ and $\sigma_4 = (9)$. Hence we have

$$\sigma = (1,2,3,7)(4,6)(5,8)(9) = (1,2,3,7)(4,6)(5,8).$$

A.5. Operations

In mathematics one meets lots of binary operations: ordinary addition and multiplication, composition of functions, matrix addition and multiplication, multiplication of permutations, etc., the list is somewhat endless. All of these binary operations have a lot in common; they also have many differences, for instance, for two real numbers x and y we have $x \cdot y = y \cdot x$, but for two 2×2 matrices with real entries, A and B, it is not true that we always have $A \cdot B = B \cdot A$. To study the similarities and differences between these operations we formalize the concept below. We then prove some results which are true of operations given some basic properties, these results can then be applied to any of the operations above which satisfy the given properties. Hence our abstraction will allow us to prove results in many areas at once.

Definition 100.18 (Operation)**.** *A (binary) operation on a set A is a function from the domain $A \times A$ to the codomain A.*

So if $A = \mathbb{R}$ we could have the function $f(x, y) = x + y$. Writing $f(x, y)$ all the time can become a pain so we often write a symbol between the x and the y to denote the operation, e.g.

$$\begin{array}{ccc} x \cdot y & x + y & x \oplus y \\ x \circ y & x \odot y & x \diamond y \\ x \wedge y & x \vee y & x \star y. \end{array}$$

Most often we write $x + y$ and $x \cdot y$; we refer to the former as additive notation and the latter as multiplicative notation. One should bear in mind that we may not be actually referring to ordinary multiplication and addition when we use these terms/notations.

Associative and Commutative: Operations can satisfy various properties.

Definition 100.19 (Associative)**.** *An operation $\diamond$ is said to be associative if for all x, y and z we have*

$$(x \diamond y) \diamond z = x \diamond (y \diamond z).$$

Operations which are associative include all the examples mentioned above. Non-associative operations do exist (for example the subtraction operation on the integers is non-associative) but we shall not be interested in them much. Note that for an associative operation the expression

$$w \diamond x \diamond y \diamond z$$

is well defined; as long as we do not change the relative position of any of the terms it does not matter which operation we carry out first.

Definition 100.20 (Commutative)**.** *An operation $\vee$ is said to be commutative if for all x and y we have*

$$x \vee y = y \vee x.$$

Ordinary addition, multiplication and matrix addition are commutative, but multiplication of matrices and permutations are not.

Identities:

Definition 100.21 (Identity)**.** *An operation $\cdot$ on the set A is said to have an identity if there exists an element e of A such that for all x we have*

$$e \cdot x = x \cdot e = x.$$

The first thing we notice is that all the example operations above possess an identity, but ordinary subtraction on the set $\mathbb{R}$ does not possess an identity. The following shows that there can be at most one identity for any given operation.

Lemma 100.22. *If an identity exists then it is unique. It is then called "the" identity.*

Proof. Suppose there are two identities e and e'. As e is an identity we have $e \cdot e' = e'$ and as e' is an identity we have $e \cdot e' = e$. Hence, we have $e' = e \cdot e' = e$. □

Usually if we are using an additive notation then we denote the identity by 0 to correspond with the identity for ordinary addition, and if we are using the multiplicative notation then we denote the identity by either 1 or e.

Inverses:

Definition 100.23 (Inverses). *Let $+$ be an operation on a set A with identity 0. Let $x \in A$. If there is a $y \in A$ such that*

$$x + y = y + x = 0$$

then we call y an inverse of x.

In the additive notation it is usual to write the inverse of x as $-x$. In the multiplicative notation it is usual to write the inverse as x^{-1}.

All elements in $\mathbb{R}$ have inverses with respect to ordinary addition. All elements in $\mathbb{R}$ except zero have inverses with respect to ordinary multiplication. Every permutation has an inverse with respect to multiplication of permutations. However, only square matrices of non-zero determinant have inverses with respect to matrix multiplication. The next result shows that an element can have at most one inverse assuming the operation is associative.

Lemma 100.24. *Consider an associative operation on a set A with identity e. Let $x \in A$ have an inverse y, then this inverse is unique, we call it "the" inverse.*

Proof. Suppose there are two such inverses y and y', then

$$y = y \cdot e = y \cdot (x \cdot y') = (y \cdot x) \cdot y' = e \cdot y' = y'.$$

Note how we used the associativity property above. □

Lemma 100.25. *Consider an associative operation on a set A with an identity e. If $a, b, x \in A$ with $a \cdot x = b \cdot x$ then $a = b$.*

Proof. Let y denote the inverse of x, then we have

$$a = a \cdot e = a \cdot (x \cdot y) = (a \cdot x) \cdot y = (b \cdot x) \cdot y = b \cdot (x \cdot y) = b \cdot e = b.$$

□

We shall assume from now on that all operations we shall encounter are associative.

Powers: Say one wishes to perform the same operation over and over again, for example

$$x \vee x \vee x \vee \cdots \vee x \vee x.$$

If our operation is written additively then we write for $n \in \mathbb{N}$, $n \cdot x$ for $x + \cdots + x$, whilst if our operation is written multiplicatively we write x^n for $x \cdots x$. The following result can then be proved by induction.

Lemma 100.26 (Law of Powers). *For any operation $\circ$ which is associative we have*

$$g^m \circ g^n = g^{m+n}, \ (g^m)^n = g^{m \cdot n}.$$

We can extend the notation to all $n \in \mathbb{Z}$ if x has an inverse (and the operation an identity), by $(-n) \cdot x = n \cdot (-x)$ and $x^{-n} = (x^{-1})^n$. The following lemma is obvious, but often causes problems as it is slightly counter-intuitive. To get it in your brain consider the case of matrices.

Lemma 100.27. *Consider a set with an associative operation which has an identity, e. If $x, y \in G$ possess inverses then we have*

1. $(x^{-1})^{-1} = x$.
2. $(x \cdot y)^{-1} = y^{-1} \cdot x^{-1}$.

Proof. For the first we notice

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

Hence by definition of inverses the result follows. For the second we have

$$x \cdot y \cdot (y^{-1} \cdot x^{-1}) = x \cdot (y \cdot y^{-1}) \cdot x^{-1} = x \cdot e \cdot x^{-1} = x \cdot x^{-1} = e,$$

and again the result follows by the definition of inverses. □

We have the following dictionary to translate between additive and multiplicative notations:

Additive	Multiplicative
$x + y$	$x \cdot y$
0	1 or e
$-x$	x^{-1}
$n \cdot x$	x^n

A.6. Groups

Definition 100.28 (Group)**.** *A group is a set G with a binary operation $\circ$ such that*

(1) *$\circ$ is associative.*
(2) *$\circ$ has an identity element in G.*
(3) *Every element of G has an inverse.*

Note that we have not said that the binary operation is closed as this is implicit in our definition of what an operation is. If the operation is also commutative then we say that we have a commutative, or abelian, group. The following are all groups; as an exercise you should decide on the identity element, what the inverse of each element is, and which groups are abelian.

(1) The integers $\mathbb{Z}$ under addition (written $\mathbb{Z}^+$).
(2) The rationals $\mathbb{Q}$ under addition (written $\mathbb{Q}^+$).
(3) The reals $\mathbb{R}$ under addition (written $\mathbb{R}^+$).
(4) The complex numbers $\mathbb{C}$ under addition (written $\mathbb{C}^+$).
(5) The rationals (excluding zero) $\mathbb{Q} \setminus \{0\}$ under multiplication (written $\mathbb{Q}^*$).
(6) The reals (excluding zero) $\mathbb{R} \setminus \{0\}$ under multiplication (written $\mathbb{R}^*$).
(7) The complex numbers (excluding zero) $\mathbb{C} \setminus \{0\}$ under multiplication (written $\mathbb{C}^*$).
(8) The set of n-ary vectors over $\mathbb{Z}, \mathbb{Q}, \ldots$, etc. under vector addition.
(9) The set of $n \times m$ matrices with integer, rational, real or complex entries under matrix addition. This set is written $M_{n \times m}(\mathbb{Z})$, etc. however when $m = n$ we write $M_n(\mathbb{Z})$ instead of $M_{n \times n}(\mathbb{Z})$.
(10) The general linear group (the matrices of non-zero determinant) over the rationals, reals or complex numbers under matrix multiplication (written $\mathrm{GL}_n(\mathbb{Q})$, etc.).
(11) The special linear group (the matrices of determinant ±1) over the integers, rationals etc. (written $\mathrm{SL}_n(\mathbb{Z})$, etc.).
(12) The set of permutations on n elements, written S_n and often called the symmetric group on n letters.
(13) The set of continuous (differentiable) functions from $\mathbb{R}$ to $\mathbb{R}$ under pointwise addition.

The list is endless; a group is one of the most basic concepts in mathematics. However, not all mathematical objects are groups. Consider the following list of sets and operations which are not groups, you should also decide why.

(1) The natural numbers $\mathbb{N}$ under ordinary addition or multiplication.
(2) The integers $\mathbb{Z}$ under subtraction or multiplication.

We now give a number of definitions related to groups.

Definition 100.29 (Orders)**.**

The order of a group is the number of elements in the underlying set G and is denoted $|G|$ or $\#G$. The order of a group can be infinite.
The order of an element $g \in G$ is the least positive integer n such that $g^n = e$, if such an n exists; otherwise we say that g has infinite order.

Definition 100.30 (Cyclic Groups and Generators)**.**
A cyclic group G is a group which has an element g such that each element of G can be written in the form g^n for some $n \in \mathbb{Z}$ (in multiplicative notation). If this is the case then one can write $G = \langle g \rangle$ and one says that g is a generator of the group G.

Note that the only element in a group with order one is the identity element and if x is an element of a group then x and x^{-1} have the same order.

Lemma 100.31. *If $G = \langle g \rangle$ and g has finite order n then the order of G is n.*

Proof. Every element of G can be written as g^m for some $m \in \mathbb{Z}$, but as g has order n there are only n distinct such values, as

$$g^{n+1} = g^n \circ g = e \circ g = g.$$

So the group G has only n elements. □

Let us relate this back to the permutations which we introduced earlier. Recall that the set of permutations on a fixed set S forms a group under composition. It is easy to see that if $\sigma \in S_n$ is a k-cycle then σ has order k in S_n. One can also easily see that if σ is a product of disjoint cycles then the order of σ is the least common multiple of the orders of the constituent cycles.

A subset S of G is said to generate G if every element of G can be written as a product of elements of S. For instance

- the group S_3 is generated by the set $\{(1,2),(1,2,3)\}$,
- the group $\mathbb{Z}^+$ is generated by the element 1,
- the group $\mathbb{Q}^*$ is generated by the set of prime numbers, it therefore has an infinite number of generators.

Note that the order of a group says nothing about the number of generators it has, although the order is clearly a trivial upper bound on the number of generators.

An important set of finite groups which are easy to understand is groups obtained by considering the integers modulo a number m. Recall that we have $\mathbb{Z}/m\mathbb{Z} = \{0, 1, \ldots, m-1\}$. This is a group with respect to addition, when we take the non-negative remainder after forming the sum of two elements. It is not a group with respect to multiplication in general, even when we exclude 0. We can, however, get around this by setting

$$(\mathbb{Z}/m\mathbb{Z})^* = \{x \in \mathbb{Z}/m\mathbb{Z} : \gcd(m, x) = 1\}.$$

This latter set is a group with respect to multiplication, when we take the non-negative remainder after forming the product of two elements. The order of $(\mathbb{Z}/m\mathbb{Z})^*$ is denoted $\phi(m)$, the Euler ϕ function. This is an important function in the theory of numbers. As an example we have

$$\phi(p) = p - 1,$$

if p is a prime number. We shall return to this function later.

Subgroups: We now turn our attention to subgroups.

Definition 100.32 (Subgroup)**.** *A* subgroup *H of a group G is a subset of G which is also a group with respect to the operation of G. We write in this case $H < G$. A subgroup H is called* trivial *if it is equal to the whole group G, or is equal to the group consisting of just the identity element.*

Note that by this definition $\mathrm{GL}_n(\mathbb{R})$ is not a subgroup of $M_n(\mathbb{R})$, although $\mathrm{GL}_n(\mathbb{R}) \subset M_n(\mathbb{R})$. The operation on $\mathrm{GL}_n(\mathbb{R})$ is matrix multiplication whilst that on $M_n(\mathbb{R})$ is matrix addition. However we do have the subgroup chains:

$$\mathbb{Z}^+ < \mathbb{Q}^+ < \mathbb{R}^+ < \mathbb{C}^+,$$
$$\mathbb{Q}^* < \mathbb{R}^* < \mathbb{C}^*.$$

If we also identify $x \in \mathbb{Z}$ with the diagonal matrix $\mathrm{diag}(x, \ldots, x)$ then we also have that $\mathbb{Z}^+$ is a subgroup of $M_n(\mathbb{Z})$ and so on.

As an important example, consider the set $2\mathbb{Z}$ of even integers, which is a subgroup of $\mathbb{Z}^+$. If we write $\mathbb{Z}^+ = 1\mathbb{Z}$, then we have $n\mathbb{Z} < m\mathbb{Z}$ if and only if m divides n, where

$$m\mathbb{Z} = \{\ldots, -2 \cdot m, -m, 0, m, 2 \cdot m, \ldots\}.$$

We hence obtain various chains of subgroups of $\mathbb{Z}^+$,

$$18\mathbb{Z} < 6\mathbb{Z} < 2\mathbb{Z} < \mathbb{Z}^+,$$
$$18\mathbb{Z} < 9\mathbb{Z} < 3\mathbb{Z} < \mathbb{Z}^+,$$
$$18\mathbb{Z} < 6\mathbb{Z} < 3\mathbb{Z} < \mathbb{Z}^+.$$

We now show that these are the only such subgroups of $\mathbb{Z}^+$.

Lemma 100.33. *The only subgroups of $\mathbb{Z}^+$ are $n\mathbb{Z}$ for some positive integer n.*

PROOF. Let H be a subgroup of $\mathbb{Z}^+$. As H is non-empty it must contain an element x and its inverse $-x$. Hence H contains at least one positive element n. Let n denote the least such positive element of H. Hence $n\mathbb{Z} \subseteq H$.

Now let m denote an arbitrary non-zero element of H. By Euclidean division, there exist $q, r \in \mathbb{Z}$ with $0 \le r < n$ such that

$$m = q \cdot n + r.$$

Hence $r \in H$. By choice of n this must mean $r = 0$, since H is a group under addition. Therefore all elements of H are of the form $n \cdot q$, for some value of q, which is what was required. □

So every subgroup of $\mathbb{Z}^+$ is an infinite cyclic group. This last lemma combined with the earlier subgroup chains gives us a good definition of what a prime number is.

Definition 100.34 (Prime Number). *A prime number is a (positive) generator of a non-trivial subgroup H of $\mathbb{Z}^+$, for which no subgroup of $\mathbb{Z}^+$ contains H except $\mathbb{Z}^+$ and H itself.*

What is good about this definition is that we have not referred to the multiplicative structure of $\mathbb{Z}$ to define the primes. Also it is obvious that neither zero nor one is a prime number. You should convince yourself that this definition leads to the usual definition of primes in terms of divisibility. In addition the above definition allows one to generalize the notion of primality to other settings; for how this is done consult any standard textbook on abstract algebra.

Normal Subgroups and Cosets: A normal subgroup is particularly important in the theory of groups. The name should not be thought of as meaning that these are the subgroups that normally arise; the name is a historic accident. To define a normal subgroup we first need to define what is meant by conjugate elements.

Definition 100.35 (Conjugate). *Two elements x, y of a group G are said to be conjugate if there is an element $g \in G$ such that $x = g^{-1} \cdot y \cdot g$.*

It is obvious that two conjugate elements have the same order. As an exercise you should show that the conjugates in a group form an equivalence class under the conjugate relation. If N is a subgroup of G we define, for any $g \in G$,

$$g^{-1}Ng = \{g^{-1} \cdot x \cdot g : x \in N\},$$

which is another subgroup of G, called a conjugate of the subgroup N.

Definition 100.36 (Normal Subgroup)**.** *A subgroup $N < G$ is said to be a normal subgroup if $g^{-1}Ng \subseteq N$ for all $g \in G$. If this is the case then we write $N \lhd G$.*

For any group G we have $G \lhd G$ and $\{e\} \lhd G$ and if G is an abelian group then every subgroup of G is normal. The importance of normal subgroups comes from the fact that these are subgroups by which we can factor out. This is related to the cosets of a subgroup which we now go on to introduce.

Definition 100.37 (Cosets)**.** *Let G be a group and $H < G$ (H is not necessarily normal). Fix an element $g \in G$, then we define the left coset of H with respect to g to be the set*

$$gH = \{g \cdot h : h \in H\}.$$

Similarly we define the right coset of H with respect to g to be the set

$$Hg = \{h \cdot g : h \in H\}.$$

Let H denote a subgroup of G then one can show that the set of all left (or right) cosets of H in G forms a partition of G, but we leave this to the reader. In addition if $a, b \in G$ then $aH = bH$ if and only if $a \in bH$, which is also equivalent to $b \in aH$, a fact which we also leave to the reader to show. Note that we can have two equal cosets $aH = bH$ without having $a = b$.

What these latter facts show is that if we define the relation R_H on the group G with respect to the subgroup H by

$$(a, b) \in R_H \text{ if and only if } a = b \cdot h \text{ for some } h \in H,$$

then this relation is an equivalence relation. The equivalence classes are just the left cosets of H in G.

The number of left cosets of a subgroup H in G is denoted by $(G : H)_L$, the number of right cosets is denoted by $(G : H)_R$. We are now in a position to prove the most important theorem of elementary group theory, namely Lagrange's Theorem.

Theorem 100.38 (Lagrange's Theorem)**.** *Let H be a subgroup of a finite group G then*

$$\begin{aligned} |G| &= (G : H)_L \cdot |H| \\ &= (G : H)_R \cdot |H|. \end{aligned}$$

Before we prove this result we state some obvious important corollaries.

Corollary 100.39.

- *We have $(G : H)_L = (G : H)_R$; we denote this common number by $(G : H)$ and call it the index of the subgroup H in G.*
- *The order of a subgroup and the index of a subgroup both divide the order of the group.*
- *If G is a group of prime order, then G has only the subgroups G and $\langle e \rangle$.*

We now return to the proof of Lagrange's Theorem.

Proof. We form the following collection of distinct left cosets of H in G which we define inductively. Put $g_1 = e$ and assume we are given i cosets by $g_1H, \ldots, g_iH$. Now take an element g_{i+1} not lying in any of the left cosets g_jH for $j \leq i$. After a finite number of such steps we have exhausted the elements of the group G. So we have a disjoint union of left cosets which cover the whole group.

$$G = \bigcup_{1 \leq i \leq (G:H)_L} g_iH.$$

We also have for each i, j that $|g_iH| = |g_jH|$, this follows from the fact that the map

$$\begin{aligned} H &\longrightarrow gH \\ h &\longmapsto g \cdot h \end{aligned}$$

is a bijective map on sets. Hence

$$|G| = \sum_{1 \leq i \leq (G:H)_L} |g_i H| = (G : H)_L |H|.$$

The other equality follows using the same argument. □

We can also deduce from the corollaries the following.

Lemma 100.40. *If G is a group of prime order then it is cyclic.*

Proof. If $g \in G$ is not the identity then $\langle g \rangle$ is a subgroup of G of order ≥ 2. But then it must have order $|G|$ and so G is cyclic. □

We can use Lagrange's Theorem to write down the subgroups of some small groups. For example, consider the group S_3: this has order 6 so by Lagrange's Theorem its subgroups must have order $1, 2, 3$ or 6. It is easy to see that the only subgroups are therefore:

- One subgroup of order 1; namely $\langle (1) \rangle$,
- Three subgroups of order 2; namely $\langle (1,2) \rangle$, $\langle (1,3) \rangle$ and $\langle (2,3) \rangle$,
- One subgroup of order 3; namely $\langle (1,2,3) \rangle$,
- One subgroup of order 6, which is S_3 obviously.

Factor or Quotient Groups: We let G be a group with a normal subgroup N. The following elementary lemma, whose proof we again leave to the reader, gives us our justification for looking at normal subgroups.

Lemma 100.41. *Let $H < G$ then the following are equivalent:*

1. *$xH = Hx$ for all $x \in G$.*
2. *$x^{-1}Hx = H$ for all $x \in G$.*
3. *$H \lhd G$.*
4. *$x^{-1} \cdot h \cdot x \in H$ for all $x \in G$ and $h \in H$.*

By G/N we denote the set of left cosets of N; note that these are the same as the right cosets of N. We note that two cosets, g_1N and g_2N are equal if and only if $g_1^{-1}g_2 \in N$.

We wish to turn G/N into a group, the so-called factor group or quotient group. Let g_1N and g_2N denote any two elements of G/N, then we define the product of their left cosets to be $(g_1g_2)N$.

We first need to show that this is a well-defined operation, i.e. if we replace g_1 by g_1' and g_2 by g_2' with $g_1^{-1}g_1' = n_1 \in N$ and $g_2^{-1} \cdot g_2' = n_2 \in N$ then our product still gives the same coset. In other words we wish to show

$$(g_1 \cdot g_2)N = (g_1' \cdot g_2')N.$$

Now let $x \in (g_1 \cdot g_2)N$, then $x = g_1 \cdot g_2 \cdot n$ for some $n \in N$. Then $x = g_1' \cdot n_1^{-1} \cdot g_2' \cdot n_2^{-1} \cdot n$. But as G is normal (left cosets = right cosets) we have $n_1'^{-1} \cdot g_2' = g_2' \cdot n_3$ for some $n_3 \in N$. Hence

$$x = g_1' \cdot g_2' \cdot n_3 \cdot n_2^{-1} \cdot n \in (g_1' \cdot g_2')N.$$

This proves the first inclusion; the other follows similarly. We conclude that our operation on G/N is well defined. One can also show that if N is an arbitrary subgroup of G and we define the operation on the cosets above then this is only a well-defined operation if N is a normal subgroup of G.

So we have a well-defined operation on G/N, we now need to show that this operation satisfies the axioms of a group:

- As an identity we take $eN = N$, since for all $g \in G$ we have
$$eN \cdot gN = (e \cdot g)N = gN.$$
- As an inverse of (gN) we take $g^{-1}N$ as
$$gN \cdot g^{-1}N = (g \cdot g^{-1})N = eN = N.$$

- Associativity follows from

$$\begin{aligned}(g_1N)\cdot(g_2N\cdot g_3N) &= g_1N\cdot((g_2\cdot g_3)N) = (g_1\cdot(g_2\cdot g_3))N\\ &= ((g_1\cdot g_2)\cdot g_3)N = ((g_1\cdot g_2)N)\cdot g_3N\\ &= (g_1N\cdot g_2N)\cdot(g_3N).\end{aligned}$$

We now present some examples.

(1) Let G be an arbitrary finite group of order greater than one; let H be a subgroup of G. Then $H = G$ and $H = \{e\}$ are always normal subgroups of G.
(2) If $H = G$ then there is only one coset and so we have $G/G = \{G\}$ is a group of order one.
(3) If $H = \{e\}$ then the cosets of H are the one-element subsets of G. That is $G/\{e\} = \{\{g\} : g \in G\}$.
(4) Put $G = S_3$ and $N = \{(1),(1,2,3),(1,3,2)\}$, then N is a normal subgroup of G. The cosets of N in G are N and $(1,2)N$ with

$$((1,2)N)^2 = (1,2)^2N = (1)N = N.$$

Hence $S_3/\langle(1,2,3)\rangle$ is a cyclic group of order 2.
(5) If G is abelian then every subgroup H of G is normal, so one can always form the quotient group G/H.
(6) Since $(\mathbb{Z},+)$ is abelian we have that $m\mathbb{Z}$ is always a normal subgroup. Forming the quotient group $\mathbb{Z}/m\mathbb{Z}$ we obtain the group of integers modulo m under addition.

Homomorphisms: Let G_1 and G_2 be two groups; we wish to look at the functions from G_1 to G_2. Obviously we could look at all such functions, however by doing this we would lose all the structure that the group laws give us. We restrict ourselves to maps which preserve these group laws.

Definition 100.42 (Homomorphism)**.** *A homomorphism from a group G_1 to a group G_2 is a function f with domain G_1 and codomain G_2 such that for all $x, y \in G_1$ we have*

$$f(x\cdot y) = f(x)\cdot f(y).$$

Note that multiplication on the left is with the operation of the group G_1 whilst the multiplication on the right is with respect to the operation of G_2. As examples we have

(1) The identity map $\mathrm{id}_G : G \to G$, where $\mathrm{id}_G(g) = g$ is a group homomorphism.
(2) Consider the function $\mathbb{R}^+ \to \mathbb{R}^*$ given by $f(x) = e^x$. This is a homomorphism as for all $x, y \in \mathbb{R}$ we have

$$e^{x+y} = e^x \cdot e^y.$$

(3) Consider the map from $\mathbb{C}^*$ to $\mathbb{R}^*$ given by $f(z) = |z|$. This is also a homomorphism.
(4) Consider the map from $\mathrm{GL}_n(\mathbb{C})$ to $\mathbb{C}^*$ given by $f(A) = \det(A)$; this is a group homomorphism as $\det(A\cdot B) = \det(A)\cdot\det(B)$ for any two elements of $\mathrm{GL}_n(\mathbb{C})$.

Two elementary properties of homomorphisms are summarized in the following lemma.

Lemma 100.43. *Let $f : G_1 \to G_2$ be a homomorphism of groups, then*

(1) $f(e_1) = e_2$.
(2) *For all $x \in G_1$ we have $f(x^{-1}) = (f(x))^{-1}$.*

Proof. For the first result we have $e_2\cdot f(x) = f(x) = f(e_1\cdot x) = f(e_1)\cdot f(x)$, and then from Lemma 100.25 we have $e_2 = f(e_1)$ as required. For the second result we have

$$f(x^{-1})\cdot f(x) = f(x^{-1}\cdot x) = f(e_1) = e_2,$$

so the result follows by definition. □

For any homomorphism f from G_1 to G_2 there are two special subgroups associated with f.

Definition 100.44 (Kernel and Image)**.**

- *The kernel of f is the set*

$$\mathrm{Ker}f = \{x \in G_1 : f(x) = e_2\}.$$

- *The image of f is the set*

$$\mathrm{Im}f = \{y \in G_2 : y = f(x),\ x \in G_1\}.$$

Lemma 100.45. Ker$f$ *is a normal subgroup of* G_1.

PROOF. We first show that it is a subgroup. It is certainly non-empty as $e_1 \in \mathrm{Ker}f$ as $f(e_1) = e_2$. Now if $x \in \mathrm{Ker}f$ then $f(x^{-1}) = f(x)^{-1} = e_2^{-1} = e_2$, hence $x^{-1} \in \mathrm{Ker}f$. Hence to show that Ker$f$ is a subgroup we only have to show that for all $x, y \in \mathrm{Ker}f$ we have $x \cdot y^{-1} \in \mathrm{Ker}f$. But this is easy as if $x, y \in \mathrm{Ker}f$ then we have

$$f(x \cdot y^{-1}) = f(x) \cdot f(y^{-1}) = e_2 \cdot e_2 = e_2,$$

and we are done.

We now show that Ker$f$ is in fact a normal subgroup of G_1. We need to show that if $x \in \mathrm{Ker}f$ then $g^{-1} \cdot x \cdot g \in \mathrm{Ker}f$ for all $g \in G_1$. So let $x \in \mathrm{Ker}f$ and let $g \in G_1$, then we have

$$f(g^{-1} \cdot x \cdot g) = f(g^{-1}) \cdot f(x) \cdot f(g) = f(g)^{-1} \cdot e_2 \cdot f(g) = f(g)^{-1} \cdot f(g) = e_2,$$

so we are done. □

Lemma 100.46. Imf *is a subgroup of* G_2.

PROOF. Imf is certainly non-empty as $f(e_1) = e_2$. Now suppose $y \in \mathrm{Im}f$ so there is an $x \in G_2$ such that $f(x) = y$, then $y^{-1} = f(x)^{-1} = f(x^{-1})$ and $x^{-1} \in G_1$ so $y^{-1} \in \mathrm{Im}f$. Now suppose $y_1, y_2 \in \mathrm{Im}f$, hence for some x_1, x_2 we have

$$y_1 \cdot y_2^{-1} = f(x_1) \cdot f(x_2^{-1}) = f(x_1 \cdot x_2^{-1}).$$

Hence $\mathrm{Im}f < G_2$. □

It is clear that Imf in some sense measures whether the homomorphism f is surjective as f is surjective if and only if $\mathrm{Im}f = G_2$. Actually the set $G_2/\mathrm{Im}f$ is a better measure of the surjectivity of the function. On the other hand, Ker$f$ measures how far from injective f is, due to the following result.

Lemma 100.47. *A homomorphism, f, is injective if and only if* $\mathrm{Ker}f = \{e_1\}$.

PROOF. Assume f is injective, then we know that if $f(x) = e_2 = f(e_1)$ then $x = e_1$ and so $\mathrm{Ker}f = \{e_1\}$. Now assume that $\mathrm{Ker}f = \{e_1\}$ and let $x, y \in G_1$ be such that $f(x) = f(y)$. Then

$$f(x \cdot y^{-1}) = f(x) \cdot f(y^{-1}) = f(x) \cdot f(y)^{-1} = f(y) \cdot f(y)^{-1} = e_2.$$

So $x \cdot y^{-1} \in \mathrm{Ker}f$, but then $x \cdot y^{-1} = e_1$ and so $x = y$. So f is injective. □

Isomorphisms: Bijective homomorphisms allow us to categorize groups more effectively, as the following definition elaborates.

Definition 100.48 (Isomorphism). *A homomorphism f is said to be an isomorphism if it is bijective. Two groups are said to be isomorphic if there is an isomorphism between them, in which case we write $G_1 \cong G_2$.*

Note that this means that isomorphic groups have the same number of elements. Indeed for all intents and purposes one may as well assume that isomorphic groups are equal, since they look the same up to relabelling of elements. Isomorphisms satisfy the following properties.

- If $f : G_1 \to G_2$ and $g : G_2 \to G_3$ are isomorphisms then $g \circ f$ is also an isomorphism, i.e. isomorphisms are transitive.
- If $f : G_1 \to G_2$ is an isomorphism then so is $f^{-1} : G_2 \to G_1$, i.e. isomorphisms are symmetric.

- The identity map $\mathsf{id} : G \to G$ given by $\mathsf{id}(x) = x$ is an isomorphism, i.e. isomorphisms are reflexive.

From this we see that the relation "is isomorphic to" is an equivalence relation on the class of all groups. This justifies our notion of isomorphic being like equal.

Let G_1, G_2 be two groups, then we define the product group $G_1 \times G_2$ to be the set $G_1 \times G_2$ of ordered pairs (g_1, g_2) with $g_1 \in G_1$ and $g_2 \in G_2$. The group operation on $G_1 \times G_2$ is given componentwise:

$$(g_1, g_2) \circ (g_1', g_2') = (g_1 \circ g_1', g_2 \circ g_2').$$

The first $\circ$ refers to the group $G_1 \times G_2$, the second to the group G_1 and the third to the group G_2. Some well-known groups can actually be represented as product groups. For example, consider the map

$$\begin{aligned} \mathbb{C}^+ &\longrightarrow \mathbb{R}^+ \times \mathbb{R}^+ \\ z &\longmapsto (\mathrm{Re}(z), \mathrm{Im}(z)). \end{aligned}$$

This map is obviously a bijective homomorphism, hence we have $\mathbb{C}^+ \cong \mathbb{R}^+ \times \mathbb{R}^+$.

We now come to a crucial theorem which says that the concept of a quotient group is virtually equivalent to the concept of a homomorphic image.

Theorem 100.49 (First Isomorphism Theorem for Groups)**.** *Let f be a homomorphism from a group G_1 to a group G_2. Then*

$$G_1/\mathrm{Ker} f \cong \mathrm{Im} f.$$

The proof of this result can be found in any introductory text on abstract algebra. Note that $G_1/\mathrm{Ker} f$ makes sense as $\mathrm{Ker} f$ is a normal subgroup of G.

A.7. Rings

A ring is an additive finite abelian group with an extra operation, usually denoted by multiplication, such that the multiplication operation is associative and has an identity element. The addition and multiplication operations are linked via the distributive law,

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

If the multiplication operation is commutative then we say we have a commutative ring. The following are examples of rings.

- Integers under addition and multiplication of integers.
- Polynomials with coefficients in $\mathbb{Z}$, denoted $\mathbb{Z}[X]$, under polynomial addition and multiplication.
- Integers modulo a number m, denoted $\mathbb{Z}/m\mathbb{Z}$, under addition and multiplication modulo m.

Although one can consider subrings they turn out to be not so interesting. Of more interest are the ideals of the ring; these are additive subgroups $I < R$ such that

$$i \in I \text{ and } r \in R \text{ implies } i \cdot r \in I.$$

Examples of ideals in a ring are the principal ideals which are those additive subgroups generated by a single ring element. For example if $R = \mathbb{Z}$ then the principal ideals are the ideals $m\mathbb{Z}$, for each integer m.

Just as with normal subgroups and groups, where we formed the quotient group, with ideals and rings we can form the quotient ring. If we take $R = \mathbb{Z}$ and $I = m\mathbb{Z}$ for some integer m then the quotient ring is the ring $\mathbb{Z}/m\mathbb{Z}$ of integers modulo m under addition and multiplication modulo m. This leads us naturally to the Chinese Remainder Theorem.

Theorem 100.50 (CRT)**.** *Let $m = p_1^{z_1} \cdots p_t^{z_t}$ be the prime factorization of m, then the following map is a ring isomorphism*

$$f : \begin{array}{ccc} \mathbb{Z}/m\mathbb{Z} & \longrightarrow & \mathbb{Z}/p_1^{z_1}\mathbb{Z} \times \cdots \times \mathbb{Z}/p_t^{z_t}\mathbb{Z} \\ x & \longmapsto & (x \pmod{p_1^{z_1}}, \ldots, x \pmod{p_t^{z_t}}). \end{array}$$

Proof. This can be proved by induction on the number of prime factors of m. We leave the details to the interested reader. □

We shall now return to the Euler ϕ function mentioned earlier. Remember $\phi(n)$ denotes the order of the group $(\mathbb{Z}/n\mathbb{Z}^)*$. We would like to be able to calculate this value easily.

Lemma 100.51. *Let $m = p_1^{z_1} \cdots p_t^{z_t}$ be the prime factorization of m. Then we have*

$$\phi(m) = \phi(p_1^{z_1}) \cdots \phi(p_t^{z_t}).$$

Proof. This follows from the Chinese Remainder Theorem, as the ring isomorphism

$$\mathbb{Z}/m\mathbb{Z} \cong \mathbb{Z}/p_1^{z_1}\mathbb{Z} \times \cdots \times \mathbb{Z}/p_t^{z_t}\mathbb{Z}$$

induces a group isomorphism

$$(\mathbb{Z}/m\mathbb{Z})^* \cong (\mathbb{Z}/p_1^{z_1}\mathbb{Z})^* \times \cdots \times (\mathbb{Z}/p_t^{z_t}\mathbb{Z})^*.$$

□

To compute the Euler ϕ function all we now require is the following.

Lemma 100.52. *Let p be a prime number, then $\phi(p^e) = p^{e-1} \cdot (p-1)$.*

Proof. There are $p^e - 1$ elements of $\mathbb{Z}$ satisfying $1 \leq k < p^e$; of these we must eliminate those of the form $k = r \cdot p$ for some r. But $1 \leq r \cdot p < p^e$ implies $1 \leq r < p^{e-1}$, hence there are $p^{e-1} - 1$ possible values of r. So we obtain

$$\phi(p^e) = (p^e - 1) - (p^{e-1} - 1)$$

from which the result follows. □

An ideal I of a ring is called prime if $x \cdot y \in I$ implies either $x \in I$ or $y \in I$. Notice that the ideals $I = m\mathbb{Z}$ of the ring $\mathbb{Z}$ are prime if and only if m is plus or minus a prime number. The prime ideals are special as if we take the quotient of a ring by a prime ideal then we obtain a field. Hence, $\mathbb{Z}/p\mathbb{Z}$ is a field. This brings us naturally to the subject of fields.

A.8. Fields

A field is essentially two abelian groups stuck together using the distributive law.

Definition 100.53 (Field)**.** *A field is an additive abelian group F, such that $F \setminus \{0\}$ also forms an abelian group with respect to another operation (which is usually written multiplicatively). The two operations, addition and multiplication, are linked via the distributive law:*

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

Many fields that one encounters have infinitely many elements. Every field either contains $\mathbb{Q}$ as a subfield, in which case we say it has characteristic zero, or it contains $\mathbb{F}_p$ as a subfield in which case we say it has characteristic p. The only fields with finitely many elements have p^r elements when p is a prime. We denote such fields by $\mathbb{F}_{p^r}$; for each value of r there is only one such field up to isomorphism. Such finite fields are often called Galois fields.

Let F be a field. We denote by $F[X]$ the ring of polynomials in a single variable X with coefficients in the field F. The set $F(X)$ of rational functions in X is the set of functions of the form

$$f(X)/g(X),$$

where $f(X), g(X) \in F[X]$ and $g(X)$ is not the zero polynomial. The set $F(X)$ is a field with respect to the obvious addition and multiplication. One should note the difference in the notation of the brackets, $F[X]$ and $F(X)$.

Let f be a polynomial of degree n with coefficients in $\mathbb{F}_p$ which is irreducible. Let θ denote a root of f. Consider the set

$$\mathbb{F}_p(\theta) = \{a_0 + a_1 \cdot \theta + \cdots + a_{n-1} \cdot \theta^{n-1} : a_i \in \mathbb{F}_p\}.$$

Given two elements of $\mathbb{F}_p(\theta)$ one adds them componentwise and multiplies them as polynomials in θ but then one takes the remainder of the result on division by $f(\theta)$. The set $\mathbb{F}_p(\theta)$ is a field; there are field-theoretic isomorphisms

$$\mathbb{F}_{p^n} \cong \mathbb{F}_p(\theta) \cong \mathbb{F}_p[X]/(f),$$

where (f) represents the ideal $\{f \cdot g : g \in \mathbb{F}_p[X]\}$.

Finite Field Example 1: To be more concrete let us look at the specific example given by choosing a value of $p = 3 \pmod 4$ and $f(X) = X^2 + 1$. Now since $p = 3 \pmod 4$ the polynomial f is irreducible over $\mathbb{F}_p[X]$ and so the quotient $\mathbb{F}_p[X]/(f)$ forms a field, which is isomorphic to $\mathbb{F}_{p^2}$. Let i denote a root of the polynomial X^2+1. The field $\mathbb{F}_{p^2} = \mathbb{F}_p(i)$ consists of numbers of the form $a + b \cdot i$, where a and b are integers modulo p. We add such numbers as

$$(a + b \cdot i) + (c + d \cdot i) = (a + c) + (b + d) \cdot i.$$

We multiply such numbers as

$$(a + b \cdot i) \cdot (c + d \cdot i) = (a \cdot c + (a \cdot d + b \cdot c) \cdot i + b \cdot d \cdot i^2) = (a \cdot c - b \cdot d) + (a \cdot d + b \cdot c) \cdot i.$$

Finite Field Example 2: Let θ denote a root of the polynomial $x^3 + 2$, then an element of $\mathbb{F}_{7^3} = \mathbb{F}_7(\theta)$ can be represented by

$$a + b \cdot \theta + c \cdot \theta^2,$$

where $a, b, c \in \mathbb{F}_7$. Multiplication of two such elements gives

$$\begin{aligned}(a + b \cdot \theta + c \cdot \theta^2) \cdot (a' + b' \cdot \theta + c' \cdot \theta^2) &= a \cdot a' + \theta \cdot (a' \cdot b + b' \cdot a) + \theta^2 \cdot (a \cdot c' + b \cdot b' + c \cdot a') \\ &\quad + \theta^3 \cdot (b \cdot c' + c \cdot b') + c \cdot c' \cdot \theta^4 \\ &= (a \cdot a' - 2 \cdot b \cdot c' - 2 \cdot c \cdot b') + \theta \cdot (a' \cdot b + b' \cdot a - 2 \cdot c \cdot c') \\ &\quad + \theta^2 \cdot (a \cdot c' + b \cdot b' + c \cdot a').\end{aligned}$$

A.9. Vector Spaces

Definition 100.54 (Vector Space)**.** *Given a field K, a vector space (or a K-vector space) V is an abelian group (also denoted V) and an external operation $\cdot : K \times V \to V$ (called scalar multiplication) which satisfies the following axioms: For all $\lambda, \mu \in K$ and all $\mathbf{x}, \mathbf{y} \in V$ we have*

(1) $\lambda \cdot (\mu \cdot \mathbf{x}) = (\lambda \cdot \mu) \cdot \mathbf{x}$.
(2) $(\lambda + \mu) \cdot \mathbf{x} = \lambda \cdot \mathbf{x} + \mu \cdot \mathbf{x}$.
(3) $1_K \cdot \mathbf{x} = \mathbf{x}$.
(4) $\lambda \cdot (\mathbf{x} + \mathbf{y}) = \lambda \cdot \mathbf{x} + \lambda \cdot \mathbf{y}$.

where 1_K denotes the multiplicative identity of K.

One often calls the elements of V the vectors and the elements of K the scalars. Note that we have not defined how to (or whether we can) multiply or divide two vectors. With a general vector space we are not interested in multiplying or dividing vectors, only in multiplying them with scalars. We shall start with some examples:

- For a given field K and an integer $n \geq 1$, let $V = K^n = K \times \cdots \times K$ be the n-fold Cartesian product. This is a vector space over K with respect to the usual addition of vectors and multiplication by scalars. The special case of $n = 1$ shows that any field is a vector space over itself. When $K = \mathbb{R}$ and $n = 2$ we obtain the familiar system of geometric vectors in the plane. When $n = 3$ and $K = \mathbb{R}$ we obtain 3-dimensional vectors. Hence you can already see the power of vector spaces as they allow us to consider n-dimensional space in a concrete way.
- Let K be a field and consider the set of polynomials over K, namely $K[X]$. This is a vector space with respect to addition of polynomials and multiplication by elements of K.
- Let K be a field and E any set at all. Define V to be the set of functions $f : E \to K$. Given $f, g \in V$ and $\lambda \in K$ one can define the sum $f + g$ and scalar product λf via
$$(f + g)(x) = f(x) + g(x) \text{ and } (\lambda \cdot f)(x) = \lambda \cdot f(x).$$
We leave the reader the simple task of checking that this is a vector space.
- The set of all continuous functions $f : \mathbb{R} \to \mathbb{R}$ is a vector space over $\mathbb{R}$. This follows from the fact that if f and g are continuous then so are $f + g$ and $\lambda \cdot f$ for any $\lambda \in \mathbb{R}$. Similarly the set of all differentiable functions $f : \mathbb{R} \to \mathbb{R}$ also forms a vector space.

Vector Sub-spaces: Let V be a K-vector space and let W be a subset of V. W is said to be a vector subspace (or just subspace) of V if

(1) W is a subgroup of V with respect to addition.
(2) W is closed under scalar multiplication.

By this last condition we mean $\lambda \cdot \mathbf{x} \in W$ for all $\mathbf{x} \in W$ and all $\lambda \in K$. What this means is that a vector subspace is a subset of V which is also a vector space with respect to the same addition and multiplication laws as V. There are always two trivial subspaces of a space, namely $\{\mathbf{0}\}$ and V itself. Here are some more examples:

- $V = K^n$ and $W = \{(\xi_1, \ldots, \xi_n) \in K^n : \xi_n = 0\}$.
- $V = K^n$ and $W = \{(\xi_1, \ldots, \xi_n) \in K^n : \xi_1 + \cdots + \xi_n = 0\}$.
- $V = K[X]$ and $W = \{f \in K[X] : f = 0 \text{ or } \deg f \leq 10\}$.
- $\mathbb{C}$ is a natural vector space over $\mathbb{Q}$, and $\mathbb{R}$ is a vector subspace of $\mathbb{C}$.
- Let V denote the set of all continuous functions from $\mathbb{R}$ to $\mathbb{R}$ and W the set of all differentiable functions from $\mathbb{R}$ to $\mathbb{R}$. Then W is a vector subspace of V.

Properties of Elements of Vector Spaces: Before we go any further we need to define certain properties which sets of elements of vector spaces can possess. For the following definitions let V be a K-vector space and let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and $\mathbf{x}$ denote elements of V.

Definition 100.55 (Linear Independence). *We have the following definitions related to linear independence of vectors.*

- $\mathbf{x}$ *is said to be a linear combination of* $\mathbf{x}_1, \ldots, \mathbf{x}_n$ *if there exists scalars* $\lambda_i \in K$ *such that*
$$\mathbf{x} = \lambda_1 \cdot \mathbf{x}_1 + \cdots + \lambda_n \cdot \mathbf{x}_n.$$
- *The elements* $\mathbf{x}_1, \ldots, \mathbf{x}_n$ *are said to be linearly independent if the relation*
$$\lambda_1 \cdot \mathbf{x}_1 + \cdots + \lambda_n \cdot \mathbf{x}_n = \mathbf{0}$$
implies that $\lambda_1 = \cdots = \lambda_n = 0$. *If* $\mathbf{x}_1, \ldots, \mathbf{x}_n$ *are not linearly independent then they are said to be linearly dependent.*
- *A subset A of a vector space is linearly independent or free if whenever* $\mathbf{x}_1, \ldots, \mathbf{x}_n$ *are finitely many elements of A, they are linearly independent.*
- *A subset A of a vector space V is said to span (or generate) V if every element of V is a linear combination of finitely many elements from A.*
- *If there exists a finite set of vectors spanning V then we say that V is finite-dimensional.*

We now give some examples of the last concept.

- The vector space $V = K^n$ is finite-dimensional. Since if we let
$$\mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0)$$
be the n-tuple with 1 in the ith place and 0 elsewhere, then V is spanned by the vectors $\mathbf{e}_1, \dots, \mathbf{e}_n$. Note the analogy with the geometric plane.
- $\mathbb{C}$ is a finite-dimensional vector space over $\mathbb{R}$, and $\{1, \sqrt{-1}\}$ is a spanning set.
- $\mathbb{R}$ and $\mathbb{C}$ are not finite-dimensional vector spaces over $\mathbb{Q}$. This is obvious since $\mathbb{Q}$ has countably many elements, so any finite-dimensional subspace over $\mathbb{Q}$ will also have countably many elements. However it is a basic result in analysis that both $\mathbb{R}$ and $\mathbb{C}$ have uncountably many elements.

Now some examples about linear independence:

- In the vector space $V = K^n$ the n vectors $\mathbf{e}_1, \dots, \mathbf{e}_n$ defined earlier are linearly independent.
- In the vector space $\mathbb{R}^3$ the vectors $\mathbf{x}_1 = (1, 2, 3)$, $\mathbf{x}_2 = (-1, 0, 4)$ and $\mathbf{x}_3 = (2, 5, -1)$ are linearly independent.
- On the other hand, the vectors $\mathbf{y}_1 = (2, 4, -3)$, $\mathbf{y}_2 = (1, 1, 2)$ and $\mathbf{y}_3 = (2, 8, -17)$ are linearly dependent as we have $3 \cdot \mathbf{y}_1 - 4 \cdot \mathbf{y}_2 - \mathbf{y}_3 = \mathbf{0}$.
- In the vector space (and ring) $K[X]$ over the field K the infinite set of vectors
$$\{1, X, X^2, X^3, \dots\}$$
is linearly independent.

Dimension and Bases:

Definition 100.56 (Basis)**.** *A subset A of a vector space V which is linearly independent and spans the whole of V is called a basis.*

Given a basis, each element in V can be written in a unique way: for suppose $\mathbf{x}_1, \dots, \mathbf{x}_n$ is a basis and we can write $\mathbf{x}$ as a linear combination of the $\mathbf{x}_i$ in two ways i.e. $\mathbf{x} = \lambda_1 \cdot \mathbf{x}_1 + \cdots + \lambda_n \cdot \mathbf{x}_n$ and $\mathbf{x} = \mu_1 \cdot \mathbf{x}_1 + \cdots + \mu_n \cdot \mathbf{x}_n$. Then we have
$$\mathbf{0} = \mathbf{x} - \mathbf{x} = (\lambda_1 - \mu_1) \cdot \mathbf{x}_1 + \cdots + (\lambda_n - \mu_n) \cdot \mathbf{x}_n$$
and as the $\mathbf{x}_i$ are linearly independent we obtain $\lambda_i - \mu_i = 0$, i.e. $\lambda_i = \mu_i$. We have the following examples.

- The vectors $\mathbf{e}_1, \dots, \mathbf{e}_n$ of K^n introduced earlier form a basis of K^n. This basis is called the standard basis of K^n.
- The set $\{1, i\}$ is a basis of the vector space $\mathbb{C}$ over $\mathbb{R}$.
- The infinite set $\{1, X, X^2, X^2, \dots\}$ is a basis of the vector space $K[X]$.

By way of terminology we call the vector space $V = \{\mathbf{0}\}$ the trivial or zero vector space. All other vector spaces are called non-zero. To make the statements of the following theorems easier we shall say that the zero vector space has the basis set $\emptyset$.

Theorem 100.57. *Let V be a finite-dimensional vector space over a field K. Let C be a finite subset of V which spans V and let A be a subset of C which is linearly independent. Then V has a basis, B, such that $A \subseteq B \subseteq C$.*

PROOF. We can assume that V is non-zero. Consider the collection of all subsets of C which are linearly independent and contain A. Certainly such subsets exist since A is itself an example. So choose one such subset B with as many elements as possible. By construction B is linearly independent. We now show that B spans V.

Since C spans V we only have to show that every element $\mathbf{x} \in C$ is a linear combination of elements of B. This is trivial when $\mathbf{x} \in B$ so assume that $\mathbf{x} \notin B$. Then $B' = B \cup \{\mathbf{x}\}$ is a subset of C larger than B, whence B' is linearly dependent, by choice of B. If $\mathbf{x}_1, \ldots, \mathbf{x}_r$ are the distinct elements of B this means that there is a linear relation

$$\lambda_1 \cdot \mathbf{x}_1 + \cdots + \lambda_r \cdot \mathbf{x}_r + \lambda \cdot \mathbf{x} = \mathbf{0},$$

in which not all the scalars, λ_i, λ, are zero. In fact $\lambda \neq 0$, otherwise B would consist of linearly dependent vectors. So we may rearrange to express $\mathbf{x}$ as a linear combination of elements of B, as λ has an inverse in K. □

Corollary 100.58. *Every finite-dimensional vector space V has a basis.*

Proof. We can assume that V is non-zero. Let C denote a finite spanning set of V and let $A = \emptyset$ and then apply the above theorem. □

The last theorem and its corollary are true if we drop the assumption of finite-dimension. However then we require much more deep machinery to prove the result. The following result is crucial to the study of vector spaces as it allows us to define the dimension of a vector space. One should think of the dimension of a vector space as the same as the dimension of the 2-D or 3-D space one is used to.

Theorem 100.59. *Suppose a vector space V contains a spanning set of m elements and a linearly independent set of n elements. Then $m \geq n$.*

Proof. Let $A = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$ span V, and let $B = \{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$ be linearly independent and suppose that $m < n$. Hence we wish to derive a contradiction.

We successively replace the $\mathbf{x}$s by the $\mathbf{y}$s, as follows. Since A spans V, there exists scalars $\lambda_1, \ldots, \lambda_m$ such that

$$\mathbf{y}_1 = \lambda_1 \cdot \mathbf{x}_1 + \cdots + \lambda_m \cdot \mathbf{x}_m.$$

At least one of the scalars, say λ_1, is non-zero and we may express $\mathbf{x}_1$ in terms of $\mathbf{y}_1$ and $\mathbf{x}_2, \ldots, \mathbf{x}_m$. It is then clear that $A_1 = \{\mathbf{y}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$ spans V.

We repeat the process m times and conclude that $A_m = \{\mathbf{y}_1, \ldots, \mathbf{y}_m\}$ spans V. (One can formally dress this up as induction if one wants to be precise, which we will not bother with.)

By hypothesis $m < n$ and so A_m is not the whole of B and $\mathbf{y}_{m+1}$ is a linear combination of $\mathbf{y}_1, \ldots, \mathbf{y}_m$, as A_m spans V. This contradicts the fact that B is linearly independent. □

Let V be a finite-dimensional vector space. Suppose A is a basis of m elements and B a basis of n elements. By applying the above theorem twice (once to A and B and once to B and A) we deduce that $m = n$. From this we conclude the following theorem.

Theorem 100.60. *Let V be a finite-dimensional vector space. Then all bases of V have the same number of elements; we call this number the dimension of V (written $\dim V$).*

It is clear that $\dim K^n = n$. This agrees with our intuition that a vector with n components lives in an n-dimensional world, and that $\dim \mathbb{R}^3 = 3$. Note that when referring to dimension we sometimes need to be clear about the field of scalars. If we wish to emphasize the field of scalars we write $\dim_K V$. This can be important, for example if we consider the complex numbers we have

$$\dim_{\mathbb{C}} \mathbb{C} = 1, \ \dim_{\mathbb{R}} \mathbb{C} = 2, \ \dim_{\mathbb{Q}} \mathbb{C} = \infty.$$

The following results are left as exercises.

Theorem 100.61. *If V is a (non-zero) finite-dimensional vector space, of dimension n, then*

1. *Given any linearly independent subset A of V, there exists a basis B such that $A \subseteq B$.*
2. *Given any spanning set C of V, there exists a basis B such that $B \subseteq C$.*
3. *Every linearly independent set in V has $\leq n$ elements.*
4. *If a linearly independent set has exactly n elements then it is a basis.*
5. *Every spanning set has $\geq n$ elements.*
6. *If a spanning set has exactly n elements then it is a basis.*

Theorem 100.62. *Let W be a subspace of a finite-dimensional vector space V. Then* $\dim W \leq \dim V$*, with equality holding if and only if* $W = V$*.*

Index

N.P. Smart, *Cryptography Made Simple*, Information Security and Cryptography, DOI 10.1007/978-3-319-21936-3