Chapter 3

Fields and Vector Spaces

3.1 Elementary Properties of Fields

3.1.1 The Definition of a Field

In the previous chapter, we noted unecessarily that one of the main concerns of algebra is the business of solving equations. Beginning with the simplest, most trivial equation, the equation $ax = b$, we see that there is a subtle point. We are used to considering equations which involve integers. To divide three apples among 4 persons, we have to consider the equation $4x = 3$. This doesn't have an integer solution. More generally, we obviously cannot function without all quotients p/q, where p, q are integers and $q \neq 0$. The set $\mathbb{Q}$ of all such quotients is the set of *rational numbers*. Recall that addition and multiplication in $\mathbb{Q}$ is defined by:

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}, \tag{3.1}$$

and

$$\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}. \tag{3.2}$$

Clearly the sum and product of two rational numbers is another rational number.

Next suppose we have been given the less trivial job of solving two linear equations in two unknowns, say x an y. These equations might be written out as

$$\begin{aligned} ax + by &= m \\ cx + dy &= n. \end{aligned}$$

Assuming $ad - bc \neq 0$, there is a unique solution which is expressed as

$$\begin{aligned} x &= \frac{dm - bn}{ad - bc} \\ y &= \frac{-cm + an}{ad - bc}. \end{aligned}$$

(In fact we derived this in the previous chapter.) The main point of this is that to express the solutions of such a linear system, one needs all the available algebraic operations: addition, subtraction, multiplication and division. A set, such as the rationals or reals, where all these operations exist is called a field.

Before defining the notion of a field, we need to define the notion of a *binary operation* on a set. Addition and multiplication on the set of integers, $\mathbb{Z}$, are two basic examples of binary operations. Let S be any set, finite or infinite. Recall that the Cartesian product of S with itself is the set $S \times S$ of all ordered pairs (x, y) of elements $x, y \in S$. Note, we call (x, y) an ordered pair since $(x, y) \neq (y, x)$ unless $x = y$. Thus,

$$S \times S = \{(x, y) \mid x, y \in S\}.$$

Definition 3.1. A *binary operation* on S is a function $F : S \times S \to S$, that is, a function F whose domain is $S \times S$ which takes its values $F(x, y)$ in S.

Note: when A and B are sets, we will write $F : A \to B$ to indicate that F is a function with domain A and values in B. Also, we often express a binary operation by writing something like $x \cdot y$ or $x * y$ for $F(x, y)$. So, for example, the operation of addition on $\mathbb{Z}$ may be thought of as being a binary operation $+$ on $\mathbb{Z}$ such that $+(x, y) = x + y$. We also need the notion of a subset being closed with respect to a binary operation.

Definition 3.2. Let F be a binary operation on a set S. A subset T of S such that $F(x, y) \in T$ for all $x, y \in T$ is said to be *closed under the binary operation.*

For example, the positive integers are closed underboth addition and multiplication. The odd integers are closed under multiplication, but not closed under addition.

We now define the notion of a field.

Definition 3.3. Assume given a set $\mathbb{F}$ with two binary operations called addition and multiplication. The sum and product of two elements $a, b \in \mathbb{F}$ will be denoted by $a + b$ and ab respectively. Suppose addition and multiplication satisfy the following properties:

(i) $a + b = b + a$ (addition is commutative);

(ii) $(a + b) + c = a + (b + c)$ (addition is associative);

(iii) $ab = ba$ (multiplication is commutative);

(iv) $a(bc) = (ab)c$ (multiplication is associative);

(v) $a(b + c) = ab + ac$ (multiplication is distributive);

(vi) $\mathbb{F}$ contains an additive identity 0 and a multiplicative identity 1 distinct from 0; the additive and multiplicative identities have the property that $a + 0 = a$ and $1a = a$ for every $a \in \mathbb{F}$;

(vii) for every $a \in \mathbb{F}$, there is an element $-a$ called the *additive inverse* of a such that $a + (-a) = 0$; and

(viii) for every $a \neq 0$ in $\mathbb{F}$, there is an element a^{-1}, called the *multiplicative inverse* of a such that $aa^{-1} = 1$.

Then $\mathbb{F}$ is called a *field.*

Note that we will often express $a+(-b)$ as $a-b$. In particular, $a-a = 0$. In any field $\mathbb{F}$, $a0 = 0$ for all a. For

$$a0 = a(0 + 0) = a0 + a0,$$

so adding $-a0$ to both sides and using the associativity of addition, we get

$$0 = a0 - a0 = (a0 + a0) - a0 = a0 + (a0 - a0) = a0 + 0 = a0.$$

Hence $a0 = 0$ for all $a \in \mathbb{F}$.

Using this fact, we next show

Proposition 3.1. *In any field $\mathbb{F}$, whenever $ab = 0$, either a or b is zero. Put another way, if neither a nor b is zero, then $ab \neq 0$.*

Proof. Suppose $a \neq 0$ and $b \neq 0$. If $ab = 0$, it follows that

$$0 = a^{-1}0 = a^{-1}(ab) = (a^{-1}a)b = 1b = b.$$

This is a contradiction, so $ab \neq 0$. □

The conclusion that $ab = 0$ implies either a or b is zero is one of the field properties that is used repeatedly. We also have

Proposition 3.2. *In any field $\mathbb{F}$, the additive and multiplicative identities are unique. Moreover, the additive and multiplicative inverses are also unique.*

Proof. We will show 0 is unique. The proof that 1 is unique is is done in exactly the same way after replacing addition by multiplication. Let 0 and $0'$ be two additive identities. Then

$$0' = 0' + 0 = 0$$

so 0 is indeed unique. We next show additive inverses are unique. Let $a \in \mathbb{F}$ have two additive inverses b and c. Using associativity, we see that

$$b = b + 0 = b + (a + c) = (b + a) + c = 0 + c = c.$$

Thus $b = c$. The rest of the proof is similar. □

3.1.2 Arbitrary Sums and Products

In a field, we can take the sum and product of any finite number of elements. However, we have to say how to define and interpret expressions such as

$$\sum_{i=1}^{k} x_i \quad \text{and} \quad \prod_{i=1}^{k} x_i.$$

Suppose we want to define the sum $x_1 + x_2 + \cdots + x_n$ of n arbitrary elements of a field $\mathbb{F}$. We do this by induction. Suppose $x_1 + x_2 + \cdots + x_{n-1}$ has been defined, and put

$$x_1 + x_2 + \cdots + x_{n-1} + x_n = (x_1 + x_2 + \cdots + x_{n-1}) + x_n.$$

Likewise, put

$$x_1 x_2 \cdots x_n = (x_1 x_2 \cdots x_{n-1}) x_n.$$

In fact, in the above sum and product, the parens can be put anywhere, as we now show.

Proposition 3.3. *In any field $\mathbb{F}$,*

$$x_1 + x_2 + \cdots + x_{n-1} + x_n = \big(\sum_{i=1}^{r} x_i\big) + \big(\sum_{i=r+1}^{n} x_i\big),$$

for any r with $1 \leq r < n$. Similarly,

$$x_1x_2\cdots x_n = \big(\prod_{i=1}^{r} x_i\big)\big(\prod_{j=r+1}^{n} x_j\big),$$

for all r with $1 \leq r \leq n-1$.

Proof. We will give the proof for sums and leave products to the reader, as the details in both cases are the same. We use induction on n. There is nothing to show for $n = 1$, so suppose $n > 1$ and the result is true for $n - 1$. If $r = n - 1$, there is also nothing to show. Thus assume $r < n - 1$. Then

$$\begin{aligned} x_1 + x_2 + \cdots + x_{n-1} + x_n &= (x_1 + x_2 + \cdots + x_{n-1}) + x_n \\ &= \big(\sum_{i=1}^{r} x_i + \sum_{j=r+1}^{n-1} x_j\big) + x_n \\ &= \sum_{i=1}^{r} x_i + \big(\sum_{j=r+1}^{n-1} x_j + x_n\big) \\ &= \big(\sum_{i=1}^{r} x_i\big) + \big(\sum_{j=r+1}^{n} x_j\big) \end{aligned}$$

Hence the result is true for n, which completes the proof. □

3.1.3 Examples

We now give some examples.

First of all, it's easy to see that the rational numbers satisfy all the field axioms, so $\mathbb{Q}$ is a field. In fact, verifying the field axioms for $\mathbb{Q}$ simply boils down to the basic arithmetic properties of the integers: associativity, commutativity and distributivity and the existence of 0 and 1. Indeed, all one needs to do is to use (3.1) and (3.2) to prove the field axioms for $\mathbb{Q}$ from these properties of the integers.

The integers $\mathbb{Z}$ are not a field, since field axiom (viii) isn't satisfied by $\mathbb{Z}$. Indeed, the only integers which have multiplicative inverses are ± 1.

The second example of a field is the set of real numbers $\mathbb{R}$. The construction of the real numbers is actually somewhat technical, so we will skip it. For most purposes, it suffices to think of $\mathbb{R}$ as being the set of all decimal expansions

$$a_1a_2\cdots a_r.b_1b_2\cdots,$$

where all a_i and b_j are integers between 0 and 9. Note that there can be infinitely many b_j to the right of the decimal point. We also have to make appropriate identifications for repeating decimals such as $1 = .999999\ldots.$ A very useful fact is that $\mathbb{R}$ is ordered; that is, any real number x is either positive , negative or 0, and the product of two numbers with the same sign is positive. This makes it possible to solve systems of linear inequalities such as $a_1x_1 + a_2x_2 + \cdots + a_nx_n > c$. In addition, the reals have what is called the *Archimedian property*: if $a, b > 0$, then there exists an $x > 0$ so that $ax > b$.

The third basic field is $\mathbb{C}$, the field of complex numbers. This is a very important field. We will discuss it in the next section.

3.1.4 An Algebraic Number Field

Many examples of fields arise by extending an already given field. We will now give an example of a field called an algebraic number field which is obtained by adjoining the square root of an integer to the rationals $\mathbb{Q}$. Let us first recall the

Theorem 3.4 (Fundamental Theorem of Arithmetic). *Let m be an integer greater than 1. Then m can be factored $m = p_1p_2\cdots p_k$, where $p_1, p_2, \ldots, p_k$ are primes. Moreover, this factorization is unique up to the order of the factors.*

Recall that a positive integer p is called *prime* if $p > 1$ and its only positive factors are 1 and itself. For a proof of the Fundamental Theorem of Arithmetic, the reader is referred to a text on elementary number theory. We say that a positive integer m is *square free* if its prime factorization has no repeated factors. For example, $10 = 2 \cdot 5$ is square free while $12 = 4 \cdot 3$ isn't.

Let $m \in \mathbb{Z}$ be positive and square free, and let $\mathbb{Q}(\sqrt{m})$ denote the set of all real numbers of the form $a + b\sqrt{m}$, where a and b are arbitrary rational numbers. It is easy to see that sums and products of elements of $\mathbb{Q}(\sqrt{m})$ give elements of $\mathbb{Q}(\sqrt{m})$. Clearly 0 and 1 are elements of $\mathbb{Q}(\sqrt{m})$. Hence, assuming the field axioms for $\mathbb{R}$ allows us to conclude without any effort that all but one of the field axioms are satisfied in $\mathbb{Q}(\sqrt{m})$. We still have to prove that any non zero element of $\mathbb{Q}(\sqrt{m})$ has a multiplicative inverse.

So assume $a + b\sqrt{m} \neq 0$. Thus at least one of a or b is non zero. By clearing away the denominators, we can assume the a and b are integers (why?). Furthermore, we can assume they don't have any common prime factors; that is, a and b are *relatively prime.* (This will also mean that both

a and b are non zero.) The trick is to notice that

$$(a+b\sqrt{m})(a-b\sqrt{m}) = a^2 - mb^2.$$

Hence

$$\frac{1}{a+b\sqrt{m}} = \frac{a-b\sqrt{m}}{a^2-mb^2}.$$

Thus, if $a^2 - mb^2 \neq 0$, then $(a+b\sqrt{m})^{-1}$ exists in $\mathbb{R}$ and is an element of $\mathbb{Q}(\sqrt{m})$.

To see that indeed $a^2 - mb^2 \neq 0$, suppose to the contrary. Then

$$a^2 = mb^2.$$

But this implies that m divides a^2, hence any prime factor p_i of m has to divide a itself. In other words, m divides a. Given that, we may cancel m on both sides and get an equation

$$cm = b^2,$$

where c is an integer. Repeating the argument, any prime factor of m has to divide b^2, hence b. The upshot of this is that the original assumption that a and b had no common factor has been violated, so the equation $a^2 = mb^2$ is impossible. Therefore we have proven

Proposition 3.5. *If m is a square free positive integer, then $\mathbb{Q}(\sqrt{m})$ is a field.*

The field $\mathbb{Q}(\sqrt{m})$ is in fact the smallest field containing both the rationals $\mathbb{Q}$ and $\sqrt{m}$.

3.1.5 The Integers Modulo a Prime p

As we have just seen, the real number field $\mathbb{R}$ admits a number of subfields. These fields are all infinite. However, not all fields are infinite. In fact the fields we will now define, the *prime fields*, are the first examples we will see of an important family known as the Galois fields.

Definition 3.4. A field with only a finite number of elements is called a *Galois field*.

The prime fields are so named because the number of elements of a prime field is a prime. In fact, it turns out that the number of elements in any Galois field $\mathbb{F}$ is a prime power, i.e. is p^n for some prime p. Furthermore, we will show in due course that for every prime p and integer $n > 0$, there eists a Galois field with p^n elements, which is unique up to isomorphism.

The simplest example of a prime field occurs when $p = 2$. (This is also the most important example for computer scientists as will soon be clear.) Recall that every field has an additive identity 0 and a multiplicative identity 1. Thus every field has at least two elements. Now define the field $\mathbb{F}_2$ to be $\{0, 1\}$, where 0 is the additive identity and 1 is the multiplicative identity. Then addition and multiplication for $\mathbb{F}_2$ are partly determined by the field axioms. That is, we have to have the following:

$$0 + 0 = 0, \quad 0 + 1 = 1 + 0 = 1, \quad 1 + 1 = 0, \quad 0 \cdot 1 = 1 \cdot 0 = 0, \quad 1 \cdot 1 = 1.$$

Thus it remains to determine $1 + 1$. But 1 has to have an additive inverse x such that $1 + x = 0$, and $x \neq 0$, since otherwise $1 + x = 1$, contradicting $0 \neq 1$. Hence the only possibility is that $x = 1$, so we are forced to define $1 + 1 = 0$. Now we state the first result.

Proposition 3.6. *$\mathbb{F}_2$ is a field.*

Proof. By going through all possible cases, we could check that the 8 field axioms are true. But we can also obtain $\mathbb{F}_2$ via arithmetic modulo 2, as we will see below. □

Notice that we can think of the two elements of $\mathbb{F}_2$ as representing on (1) and off (0). Adding 1 causes the state to change: off becomes on and on becomes off. What turns out to be useful is that a sequence of 0's and 1's can now be viewed in a much more useful way as a sequence of elements of $\mathbb{F}_2$. This fact plays a very important role in coding theory and information theory, two disciplines that all sorts of everyday things such as PC's, CD players, modems etc. couldn't exist without. We will study linear coding theory in some detail in Chapter **??**.

Next consider a prime $p > 1$. We now define the field $\mathbb{F}_p$ with p elements. We will use modular arithmetic to make $\mathbb{F}_p$ into a Galois field. Doing modular arithmetic is like telling time on a clock. In fact, telling time on a clock requires addition modulo 12. In general, modular arithmetic can be succinctly described as addition and multiplication with remainders.

Put

$$\mathbb{F}_p = \{0, 1, 2, \dots p - 1\}. \tag{3.3}$$

We have to define addition and multiplication so that all eight field axioms are satisfied. To add two elements a and b in $\mathbb{F}_p$, first take their sum in the usual way to get the integer $a + b$. If $a + b < p$, then we define their sum in

$\mathbb{F}_p$ to be $a+b$. However, if $a+b \geq p$, we need to use *division with remainder.* This is the principle explained in the next Proposition.

Proposition 3.7. *Suppose a and b are non-negative integers with $b \neq 0$. Then one can uniquely express a as $a = qb + r$, where q is a non-negative integer and $0 \leq r < b$.*

Proof. If $a < b$, put $r = a$, and if $a = b$, put $r = 0$. If $a > b$, some positive multiple sb of b satisfies $sb > a$. Let s be the least positive integer such that this happens. Here we are using the assumption that every non-empty set of positive integers has a least element. Put $q = s - 1$. Thus $a = qb + (a - qb)$, so we have to show that $r = a - qb$ satisfies $0 \leq r < b$. Now $a - qb = a - (s-1)b \geq 0$ by definition. Also $a - qb = (a - sb) + b < b$ since $a < sb$. This finishes the proof. □

Thus, if $a + b \geq p$, write

$$a + b = qp + r,$$

where q is a nonnegative integer and r is an integer such that $0 \leq r < p$. Then the *sum of* a and b in $\mathbb{F}_p$ is defined to be r. This operation is called *addition modulo p.* It is a special case of *modular addition.*

To define the *product of* a and b in $\mathbb{F}_p$, we use (in exactly the same way) the remainder upon dividing ab by p.

Example 3.1. Let's carry out the definitions of addition and multiplication in $\mathbb{F}_3 = \{0, 1, 2\}$. Of course, 0 and 1 are always the identities, so all sums and products involving them are determined. To completely determine the addition, we therefore only have to define $1 + 1$, $1 + 2$ and $2 + 2$. First of all, $1 + 1 < 3$, so by definition, $1 + 1 = 2$. To find $2 + 2$, first take the usual sum 4, then express $4 = 3 + 1$ as in Proposition 3.7. The remainder is 1, so $2 + 2 = 1$ in $\mathbb{F}_3$. Similarly, $1 + 2 = 0$ in $\mathbb{F}_3$. Thus $-2 = 1$ and $-1 = 2$. To find all products, it remains to find $2 \cdot 2$. But $2 \cdot 2 = 4$ in usual arithmetic, so $2 \cdot 2 = 1$ in $\mathbb{F}_3$. Thus $2^{-1} = 2$. A good way to summarize addition and multiplication is to construct addition and multiplication tables. The addition table for $\mathbb{F}_3$ is

+	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

We suggest that the reader construct the multiplication table for $\mathbb{F}_3$.

For arbitrary primes, the existence of additive inverses is easy to see (the inverse of a is $p-a$), but it is not so obvious that multiplicative inverses always exist. To prove that they do, we will make a short diversion to prove the *pigeon hole principle.* First recall the following definition.

Definition 3.5. If X and Y are two sets, then a map $\phi : X \to Y$ is called *one to one* or *injective* if $\phi(x) = \phi(x')$ implies $x = x'$. Also, ϕ is *onto* or *surjective* if for every $y \in Y$, $y = \phi(x)$ for some $x \in X$. In other words, ϕ is surjective if the *image of* ϕ

$$\phi(X) = \{\phi(x) \mid x \in X\} \subset Y$$

is all of Y: $\phi(X) = Y$. If ϕ is both injective and surjective, it is said to be *bijective.*

If X is finite, recall that the number of elements of X is denoted by $|X|$.

Proposition 3.8 (The Pigeon Hole Principle). *Let X and Y be finite sets with $|X| = |Y|$, and suppose $\phi : X \to Y$ is a map. If ϕ is either injective or surjective, then ϕ is a bijection.*

Proof. If ϕ is injective, then X and its image $\phi(X)$ have the same number of elements. But this implies $\phi(X) = Y$, so ϕ is surjective, hence is a bijection. On the other hand, suppose ϕ is surjective, i.e. $\phi(X) = Y$. Then $|X| \geq |Y|$. But if $\phi(x) = \phi(x')$ where $x \neq x'$, then infact $|X| > |Y|$. This contradicts the assumption that $|X| = |Y|$, hence ϕ is bijective. □

We now return to the proof that every nonzero element a of $\mathbb{F}_p$ has an inverse a^{-1}. First, we show $\mathbb{F}_p$ satisfies the conclusion of Proposition 3.1:

Proposition 3.9. *Let p be a prime number. If $ab = 0$ in $\mathbb{F}_p$, then either $a = 0$ or $b = 0$ (or both).*

Proof. Since $ab = 0$ in $\mathbb{F}_p$ is the same thing as saying that p divides the usual product ab, the Proposition follows from the fact that if the prime number p divides ab, then it divides a or it divides b. This fact follows immediately from the Fundamental Theorem of Arithmetic (Theorem 3.4). □

This Proposition says that multiplication by a fixed non-zero element $a \in \mathbb{F}_p$ induces an injective map

$$\begin{aligned} \phi_a : \mathbb{F}_p \setminus \{0\} &\longrightarrow \mathbb{F}_p \setminus \{0\} \\ x &\longmapsto ax \end{aligned}$$

Here $\mathbb{F}_p \setminus \{0\}$ is the set $\mathbb{F}_p$ without 0. To see that ϕ_a is injective, let $\phi_a(x) = \phi_a(y)$, that is $ax = ay$. Thus $a(x - y) = 0$, so $x - y = 0$ since $a \neq 0$ (Proposition 3.9). Therefore $x = y$. Since $F_p \setminus \{0\}$ is a *finite* set, the Pigeon Hole Principle says that ϕ_a is a bijection. In particular, there exists an $x \in \mathbb{F}_p \setminus \{0\}$, such that $ax = 1$. Hence x is the required inverse of a.

We will skip the proofs that addition and multiplication defined on $\mathbb{F}_p$ using modular arithmetic satisfy the field axioms (i) through (v). Thus, putting the above facts together, we get

Theorem 3.10. *If p is a prime, then $\mathbb{F}_p$, as defined above, is a field.*

If the requirement of having multiplicative inverses is taken out of the definition of a field, the resulting system is called a *ring*. For example, $\mathbb{Z}_4$ is a ring, but not a field since, in $\mathbb{Z}_4$, $2 \cdot 2 = 0$. In fact, if q is a composite number, then $\mathbb{Z}_q$ (defined exactly as above) is a ring but not a field. Note that the integers $\mathbb{Z}$ also form a ring. We will take up ring theory in due course.

3.1.6 A Field with Four Elements

As we mentioned at the beginning of this section, there exist Galois fields of order p^n for every prime p and integer $n > 0$. We now write down the addition and multiplication tables for a field with four elements. However, the actual construction of this field will be left until Example 15.17. Let $\mathbb{F}_4 = \{0, 1, \alpha, \beta\}$. Define the addition table (omitting addition by 0) as follows.

$+$	1	α	β
1	0	β	α
α	β	0	1
β	α	1	0

The multiplication table (omitting the obvious cases) is defined by

$\cdot$	α	β
α	β	1
β	1	α

Then we have

Proposition 3.11. *The set $\mathbb{F}_4 = \{0, 1, \alpha, \beta\}$ having 0 and 1 as identities and addition and multiplication defined as above is a field.*

The verification of this Proposition requires that we check all the field axioms. Again, this has to be done by hand, and so we will omit it. This somewhat unsatisfactory situation will be fixed Chapter 15, where we will give the uniform construction of all the Galois fields and prove their uniqueness.

The reader may have noticed that the only possible way multiplication can be defined is to require that $\alpha^2 = \beta$ and $\beta^2 = \alpha$. It thus follows that $\alpha^3 = \beta^3 = 1$. Hence $\alpha^4 = \alpha$ and $\beta^4 = \beta$, so all elements of $\mathbb{F}_4$ satisfy the equation $x^4 - x = 0$ since 0 and 1 trivially do. Now by Section 3.1.9 below, we can view $x^4 - x$ as a polynomial in a variable x over the field $\mathbb{F}_2$, where we have the identity $x^4 - x = x^4 + x$. Thus, we can factor $x^4 - x = x(x+1)(x^2+x+1)$ (remember $1+1=0$ so $2x = 2x^2 = 0$). The elements α and β are therefore roots of $x(x+1)(x^2+x+1) = 0$, and the only possible way this can occur is when they satisfy $x^2+x+1=0$. This accounts for the definition of addition in $\mathbb{F}_4$. As we will see, there are appropriate generalizations of these statements for all Galois fields.

3.1.7 Some Elementary Number Theory

We now make some definitions from elementary number theory. For any integers a and b which are not both 0, let $d > 0$ be the largest integer which divides both a and b. We call d the *greatest common divisor* of a and b. The greatest common divisor, or simply, gcd of a and b is traditionally denoted (a, b). For example, $(4, 10) = 2$. The first fact is

Proposition 3.12. *Let a and b be integers which are not both 0, and let d be their gcd. Then there exist integers u and v such that $au + bv = d$. Conversely, if there exist integers u and v such that $au + bv = d$, then $d = (a, b)$.*

Proof. The proof is accomplished by repeated application of Proposition 3.7. We refer the reader to a book on number theory for complete details. □

Definition 3.6. Let a, b, c be integers. Then we say a is *congruent to b modulo c* if $a - b$ is divisible by c. If a is congruent to b modulo c, we write $a \equiv b \bmod c$.

Proposition 3.13. *Let a, b, q be positive integers. Then the congruence equation $ax \equiv 1$ mod q has a solution if and only if $(a, q) = 1$.*

This proposition again implies that non-zero elements of $\mathbb{F}_p$ have multiplicative inverses. The following classical result of Fermat gives a formula for finding the inverse of any element in $\mathbb{F}_p$.

Fermat's Little Theorem: *Suppose p is a prime greater than 1. Then for any integer $a \not\equiv 0 \bmod p$, $a^{(p-1)} \equiv 1 \bmod p$.*

We will give a group theoretic proof of Fermat's Little Theorem in Chapter 13. Fermat's Little Theorem gives a well known test an integer m to be prime: in order that m be prime, $a^{(m-1)} \equiv 1 \bmod m$ has to hold for all integers a. One also gets a formula for the inverse in $\mathbb{F}_p$.

Proposition 3.14. *If p is a prime and $a \neq 0$ in $\mathbb{F}_p$, then the reduction modulo p of $a^{(p-2)}$ is the inverse of a in $\mathbb{F}_p$.*

For example, suppose we want to compute the inverse of 5 in $\mathbb{F}_{23}$. Since $5^{21} = 476837158203125$, we simply reduce 476837158203125 modulo 23, which gives 14. If you weren't able to do this calculation in your head, it is useful to have a math package such as Maple or Mathematica. Of course, $5 \cdot 14 = 70 = 3 \cdot 23 + 1$, which is easier to see than the above value of 5^{21}.

The proof of Fermat's Little Theorem rests on the following identity in $\mathbb{F}_p$:

$$(a+b)^p = a^p + b^p \tag{3.4}$$

for all $a, b \in \mathbb{F}_p$. This identity an application of the **Binomial Theorem**: Let x and y be any pair of commuting vaiables. Then, for any positive integer n,

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^{n-i} y^i, \tag{3.5}$$

where

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

The identity (3.4) does not in general hold in $\mathbb{Z}_n$ unless n is prime.

Note that Fermat's Little Theorem is not the fact known as Fermat's Last Theorem which Fermat famously stated without proof and which was finally proved some 350 years later by Andrew Wiles: namely, there are no integer solutions $m > 2$ of $a^m + b^m = c^m$ where $a, b, c \in \mathbb{Z}$ are all non zero. Amusingly, Fermat's Last Theorem is false in $\mathbb{F}_p$, since, by the Binomial Theorem,

$$(a+b)^p = a^p + b^p \tag{3.6}$$

for all $a, b \in \mathbb{F}_p$. Hence the sum of two pth powers is a pth power.

3.1.8 The Characteristic of a Field

If $\mathbb{F}$ is a finite field, then some multiple r of the identity $1 \in \mathbb{F}$ has to be 0. The reason for this is that since $\mathbb{F}$ is finite, the multiples $r1$ of 1 can't all be different. Hence there have to be $m > n$ such that $m1 = n1$ in $\mathbb{F}$. But this implies $(m-n)1 = 0$. Now I claim that the least positive integer r

such that $r1 = 0$ is a prime. For if r can be expressed as a product $r = st$, where s, t are positive integers, then, $r1 = (st)1 = (s1)(t1) = 0$. But, by the minimality of r, $s1 \neq 0$ and $t1 \neq 0$, so a factorization $r = st$ is impossible unless either s or t is 1. Therefore r is a prime, say $r = p$. One calls p the *characteristic of* $\mathbb{F}$. In general, one makes the following definition:

Definition 3.7. Let $\mathbb{F}$ be an arbitrary field. If some multiple $q1$ of 1 equals 0, we say that $\mathbb{F}$ has *positive characteristic*, and, in that case, the *characteristic of* $\mathbb{F}$ is defined to be the least positive integer q such that $q1 = 0$. If all multiples $q1$ are nonzero, we say $\mathbb{F}$ has *characteristic* 0.

Summarizing the above discussion, we state

Proposition 3.15. *If a field $\mathbb{F}$ has positive characteristic, then its characteristic is a prime. Otherwise its characteristic is zero.*

Clearly the characteristics of $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ are all 0. Moreover, the characteristic of any subfield of a field of characteristic 0 is also 0.

3.1.9 Polynomials

Let $\mathbb{F}$ be a field and suppose x denotes a variable. We will assume it makes sense to talk about the powers x^i, where i is any positive integer. Define $\mathbb{F}[x]$ to be the set of all polynomials

$$p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

with coefficients $a_i \in \mathbb{F}$ for each i, where n is an arbitrary non-negative integer. If $a_n \neq 0$, we say that f has degree n. Of course, if $a_i = 0$, we interpret $a_i x^i$ as being zero also. Addition of polynomials is defined by adding the coefficients of each x^i. We may also multiply two polynomials in the natural way using $x^i x^j = x^{i+j}$ and the distributive law.

Note that by definition, two polynomials $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ and $q(x) = b_k x^k + b_{k-1}x^{k-1} + \cdots + b_1 x + b_0$ are equal if and only if $a_i = b_i$ for each index i.

Exercises

Exercise 3.1. Prove that in any field $(-1)a = -a$.

Exercise 3.2. Find the characteristic of the field $\mathbb{F}_p$.

Exercise 3.3. Suppose the field $\mathbb{F}$ contains $\mathbb{F}_p$ as a subfield. Show that the characteristic of $\mathbb{F}$ is p.

Exercise 3.4. Show that if $\mathbb{F}$ is a finite field of characteristic p, then for any $a, b \in \mathbb{F}$, we have $(a+b)^p = a^p + b^p$. Also, show by example that this identity fails in $\mathbb{Z}_p$ when p isn't prime.

Exercise 3.5. Use (3.6) to show that $a^p = a$ in $\mathbb{F}_p$. Deduce Fermat's Little Theorem for this.

Exercise 3.6. Suppose $\mathbb{F}$ is a field of characteristic p. Show that if $a, b \in \mathbb{F}$ and $a^p = b^p$, then $a = b$.

Exercise 3.7. Show that $\mathbb{F}$ is a finite field of characteristic p, then $\mathbb{F}$ is *perfect*. That is, every element in $\mathbb{F}_p$ is a pth power. (Hint: use the pigeon hole principle.)

Exercise 3.8. Show directly that $\mathbb{F} = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$ is a field under the usual operations of addition and multiplication in $\mathbb{R}$. Also, find $(1-\sqrt{2})^{-1}$ and $(3-4\sqrt{2})^{-1}$.

Exercise 3.9. Describe addition and multiplication for the field $\mathbb{F}_p$ having p elements for $p = 5$. That is, construct addition and multiplication tables for $\mathbb{F}_5$ as in Example 1.1. Check that every element $a \neq 0$ has a multiplicative inverse.

Exercise 3.10. Use Fermat's Theorem to find 9^{-1} in $\mathbb{F}_{13}$. Use this to solve the equation $9x \equiv 15 \mod 13$.

Exercise 3.11. Find at least one primitive element β for $\mathbb{F}_{13}$? (Calculators should be used here.) Also, express 9^{-1} using this primitive element instead of Fermat's Theorem.

Exercise 3.12. Let $\mathbb{Z}$ denote the integers. Consider the set $\mathcal{Q}$ of all pairs (a, b) where $a, b \in \mathbb{Z}$ and $b \neq 0$. Consider two pairs (a, b) and (c, d) to be the same if $ad = bc$. Now define operations of addition and multiplication on $\mathcal{Q}$ as follows:

$$(a, b) + (c, d) = (ad + bc, bd) \quad \text{and} \quad (a, b)(c, d) = (ac, bd).$$

Show that $\mathcal{Q}$ is a field. Can you identify $\mathcal{Q}$?.

Exercise 3.13. Write out the addition and multiplication tables for $\mathbb{F}_6$. Is $\mathbb{F}_6$ is a field? If not, why not?

Exercise 3.14. Find both $-(6+6)$ and $(6+6)^{-1}$ in $\mathbb{F}_7$.

Exercise 3.15. Let $\mathbb{F}$ be a field and suppose that $\mathbb{F}' \subset \mathbb{F}$ is a subfield, that is, $\mathbb{F}'$ is a field for the operations of $\mathbb{F}$. Show that $\mathbb{F}$ and $\mathbb{F}'$ have the same characteristic.

3.2 The Field of Complex Numbers

We will now introduce the field $\mathbb{C}$ of complex numbers. The complex numbers are incredibly rich. Without them, mathematics would be a far less interesting discipline. From our standpoint, the most notable fact about the complex numbers is that they form an *algebraically closed field*. That is, $\mathbb{C}$ contains all roots of any polynomial

$$x^n + a_1x^{n-1} + \ldots a_{n-1}x + a_n = 0$$

with complex coefficients. This statement, which is due to C. F. Gauss, is called the Fundamental Theorem of Algebra.

3.2.1 The Definition

The starting point for considering complex numbers is the problem that if a is a positive real number, then $x^2+a=0$ apparently doesn't have any roots. In order to give it roots, we have to make sense of an expression such as $\sqrt{-a}$. The solution turns turns out to be extremely natural. The real xy-plane $\mathbb{R}^2$ with its usual component-wise addition also has a multiplication such that certain points (namely points on the y-axis), when squared, give points on the negative x-axis. If we interpret the points on the x-axis as real numbers, this solves our problem. It also turns out that under this multiplication on $\mathbb{R}^2$, every nonzero pair $(a,b)^T$ has a multiplicative inverse. The upshot is that we obtain the field $\mathbb{C}$ of complex numbers. The marvelous and deep consequence of this definition is that $\mathbb{C}$ contains not only numbers such as $\sqrt{-a}$, it contains the roots of all polynomial equations with real coefficients.

Let us now give the details. The definition of multiplication on $\mathbb{R}^2$ is easy to state and has a natural geometric meaning discussed below. First of all, we will call the x-axis the *real axis*, and identify a point of the form $(a,0)^T$ with the real number a. That is, $(a,0)^T = a$. Hence multiplication on $\mathbb{R}$ can be reformulated as $ab = (a,0)^T \cdot (b,0)^T = (ab,0)^T$. We extend this multiplication to all of $\mathbb{R}^2$ by putting

$$(a,b)^T \cdot (c,d)^T = (ac-bd, ad+bc)^T. \tag{3.7}$$

(Note: do not confuse this with the dot product on $\mathbb{R}^2$.)

We now make the following definition.

Definition 3.8. Define $\mathbb{C}$ to be $\mathbb{R}^2$ with the usual component-wise addition (vector addition) and with the multiplication defined by (3.7).

Addition and multiplication are clearly binary operations. Notice that $(0, a)^T \cdot (0, a)^T = (-a^2, 0)^T$, so that $(0, a)^T$ is a square root of $-a^2$. It is customary to denote $(0, 1)^T$ by i so

$$i = \sqrt{-1}.$$

Since any point of $\mathbb{R}^2$ can be uniquely represented

$$(a, b)^T = a(1, 0)^T + b(0, 1)^T, \tag{3.8}$$

we can therefore write

$$(a, b)^T = a + ib.$$

In other words, by identifying the real number a with the vector $a(1, 0)^T$ on the real axis, we can express any element of $\mathbb{C}$ as a sum of a real number, its *real part*, and a multiple of i, its *imaginary part*. Thus multiplication takes the form

$$(a + ib)(c + id) = (ac - bd) + i(ad + bc).$$

Of course, $\mathbb{R}$ is explicitly given as a subset of $\mathbb{C}$, namely the real axis.

The Fundamental Theorem of Algebra is formally stated as follows:

Theorem 3.16. *A polynomial equation*

$$p(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0 = 0$$

with complex (but possibly real) coefficients has n complex roots.

There are many proofs of this theorem, but none of them are elementary enough to repeat here. Every known proof draws on some deep result from another field, such as complex analysis or topology.

An easy consequence is that given any polynomial $p(z)$ with complex coefficients, there exist $r_1, \ldots, r_n \in \mathbb{C}$ which are not necessarily all distinct such that

$$p(z) = (z - r_1)(z - r_2) \ldots (z - r_n).$$

We now prove

Theorem 3.17. $\mathbb{C}$ *is a field containing* $\mathbb{R}$ *as a subfield.*

Proof. The verification of this theorem is simply a computation. The real number 1 is the identity for multiplication in $\mathbb{C}$, and $0 = (0, 0)^T$ is the identity for addition. If $a + ib \neq 0$, then $a + ib$ has a multiplicative inverse, namely

$$(a + ib)^{-1} = \frac{a - ib}{a^2 + b^2}. \tag{3.9}$$

The other properties of a field follow easily from the fact that $\mathbb{R}$ is a field. □

3.2.2 The Geometry of $\mathbb{C}$

We now make some more definitions which lead to some beautiful geometric properties of $\mathbb{C}$. First of all, the *conjugate* $\overline{z}$ of $z = a + ib$ is defined by $\overline{z} = a - ib$. It is easy to check the following identities:

$$\begin{aligned} \overline{w+z} &= \overline{w} + \overline{z} \quad \text{and} && (3.10)\\ \overline{wz} &= \overline{w}\ \overline{z}. && (3.11) \end{aligned}$$

The real numbers are obviously the numbers which are equal to their conjugates. Complex conjugation is the transformation from $\mathbb{R}^2$ to itself which sends a point to its reflection through the real axis.

Formula (3.9) for $(a+ib)^{-1}$ above can now be expressed in a new way. Let $z = a + ib \neq 0$. Since $z\overline{z} = a^2 + b^2$, we get

$$z^{-1} = \frac{\overline{z}}{a^2 + b^2}.$$

Notice that the denominator of the above formula is the square of the length of z. The length of a complex number $z = a + ib$ is called its *modulus* and is denoted by $|z|$. Thus

$$|z| = (z\overline{z})^{1/2} = (a^2 + b^2)^{1/2}.$$

Since $\overline{wz} = \overline{w}\ \overline{z}$, we obtain the nice formula for the modulus of a product, namely

$$|wz| = |w||z|. \tag{3.12}$$

In particular, the product of two unit length complex numbers also has length one. Now the complex numbers of unit length are just those on the unit circle C=$\{x^2 + y^2 = 1\}$. Every point of C can be represented in the form $(\cos\theta, \sin\theta)$ for a unique angle θ such that $0 \leq \theta < 2\pi$. It is convenient to use a complex valued function of $\theta \in \mathbb{R}$ to express this. We define the *complex exponential* to be the function

$$e^{i\theta} := \cos\theta + i \sin\theta. \tag{3.13}$$

The following proposition is geometrically clear.

Proposition 3.18. *Any $z \in \mathbb{C}$ can be represented as $z = |z|e^{i\theta}$ for some $\theta \in \mathbb{R}$. θ is unique up to a multiple of 2π.*

The value of θ in $[0, 2\pi)$ such that $z = |z|e^{i\theta}$ is called the *argument* of z. The key property of the complex exponential is the identity

$$e^{i(\theta+\mu)} = e^{i\theta}e^{i\mu}, \tag{3.14}$$

which follows from the standard trigonometric formulas for the sine and cosine of the sum of two angles. (We will give a simple proof of this when we study rotations in the plane.) This gives complex multiplication a geometric interpretation. Writing $w = |w|e^{i\mu}$, we see that

$$wz = (|w|e^{i\mu})(|z|e^{i\theta}) = (|w||z|)(e^{i\mu}e^{i\theta}) = |wz|e^{i(\mu+\theta)}.$$

In other words, the product wz is obtained by multiplying the lengths of w and z and adding their arguments.

Exercises

Exercise 3.16. Find all solutions of the equation $z^3 + 1 = 0$ and interpret them as complex numbers. Do the same for $z^4 - 1 = 0$.

Exercise 3.17. Find all solutions of the linear system

$$\begin{aligned} ix_1 + 2x_2 + (1-i)x_3 &= 0 \\ -x_1 + ix_2 - (2+i)x_3 &= 0 \end{aligned}$$

Exercise 3.18. Suppose $p(x) \in \mathbb{R}[x]$. Show that the roots of $p(x) = 0$ occur in conjugate pairs, that is $\lambda, \mu \in \mathbb{C}$ where $\overline{\lambda} = \mu$.

3.3 Vector spaces

3.3.1 The notion of a vector space

In mathematics, there many situations in which one deals with sets of objects which can be added and multiplied by scalars, so that these two operations behave like vector addition and scalar multiplication in $\mathbb{R}^n$. A fundamental example of this is the set of all real valued functions whose domain is a closed interval $[a, b]$ in $\mathbb{R}$, which one frequently denotes as $\mathbb{R}^{[a,b]}$. Addition and scalar multiplication of functions is defined pointwise, as in calculus. That is, if f and g are functions on $[a, b]$, then $f + g$ is the function whose value at $x \in [a, b]$ is

$$(f + g)(x) = f(x) + g(x),$$

and if r is any real number, then rf is the function whose value at $x \in [a, b]$ is

$$(rf)(x) = rf(x).$$

The key point is that we have defined sums and scalar multiples so that the sum of $f, g \in \mathbb{R}^{[a,b]}$ and all scalar multiples of a single $f \in \mathbb{R}^{[a,b]}$ are also elements of $\mathbb{R}^{[a,b]}$. When a set S admits an addition (resp. scalar multiplication) with this property, we will say that S is *closed* under addition (resp. scalar multiplication).

A more familiar example is the set $C(a, b)$ of all continuous real valued functions on $[a, b]$. Since $C(a, b) \subset \mathbb{R}^{[a,b]}$, we will of course use the definitions of addition and scalar multiplication already given for $\mathbb{R}^{[a,b]}$. In order to know that $C(a, b)$ is closed under addition and scalar multiplication, we need to know that sums and scalar multiples of continuous functions are continuous. But this is guaranteed by a basic theorem usually discussed in calculus: the sum of two continuous functions is continuous and any scalar multiple of a continuous function is continuous. Hence

$f + g$ and rf belong to $C(a, b)$ for all f and g in $C(a, b)$ and any real scalar r.

We now give the definition of a vector space over a field $\mathbb{F}$. It will be clear that, under the definitions of addition and scalar multiplication given above, $\mathbb{R}^{[a,b]}$ is a vector space over $\mathbb{R}$.

Definition 3.9. Let $\mathbb{F}$ be a field and V a set. Assume that there is a binary operation on V called addition which assigns to each pair of elements $\mathbf{a}$ and $\mathbf{b}$ of V a unique sum $\mathbf{a} + \mathbf{b} \in V$. Assume also that there is a second operation, called scalar multiplication, which assigns to any $r \in \mathbb{F}$ and any

$\mathbf{a} \in V$ a unique scalar multiple $r\mathbf{a} \in V$. Suppose that addition and scalar multiplication satisfy the following axioms.

(1) Vector addition is commutative. That is, $\mathbf{a}+\mathbf{b} = \mathbf{b}+\mathbf{a}$ for all $\mathbf{a}, \mathbf{b} \in V$.

(2) Vector addition is also associative. That is, $(\mathbf{a}+\mathbf{b})+\mathbf{c} = \mathbf{a}+(\mathbf{b}+\mathbf{c})$ for all $\mathbf{a}, \mathbf{b}, \mathbf{c} \in V$.

(3) There is an additive identity $\mathbf{0} \in V$ so that $\mathbf{0}+\mathbf{a} = \mathbf{a}$ for all $\mathbf{a} \in V$.

(4) Every element of V has an additive inverse. That is, given $\mathbf{a} \in V$, there is an element denoted $-\mathbf{a} \in V$ so that $\mathbf{a}+(-\mathbf{a}) = \mathbf{0}$.

(5) $1\mathbf{a} = \mathbf{a}$, for all $\mathbf{a} \in V$.

(6) Scalar multiplication is associative. If $r, s \in \mathbb{F}$ and $\mathbf{a} \in V$, then $(rs)\mathbf{a} = r(s(\mathbf{a}))$.

(7) Scalar multiplication is distributive. If $r, s \in \mathbb{F}$ and $\mathbf{a}, \mathbf{b} \in V$, then $r(\mathbf{a}+\mathbf{b}) = r\mathbf{a}+r\mathbf{b}$, and $(r+s)\mathbf{a} = r\mathbf{a}+s\mathbf{a}$.

Then V is called a *vector space over* $\mathbb{F}$.

You will eventually come to realize that all of the above conditions are needed. Just as for fields, the additive identity $\mathbf{0}$ and additive inverses unique: each vector has exactly one negative. We will call $\mathbf{0}$ the *zero vector.*

Let's consider some more examples.

Example 3.2. The first example is the obvious one: if $n \geq 1$, then $\mathbb{R}^n$ with the usual component-wise addition and scalar multiplication is a real vector space, that is a vector space over $\mathbb{R}$. We usually call $\mathbb{R}^n$ real n-space.

Example 3.3. More generally, for any field $\mathbb{F}$ and $n \geq 1$, the set $\mathbb{F}^n$ of all n-tuples $(a_1, a_2, \dots, a_n)^T$ of elements of $\mathbb{F}$ can be made into a vector space over $\mathbb{F}$ in exactly the same way. That is,

$$(a_1, a_2, \dots, a_n)^T + (b_1, b_2, \dots, b_n)^T = (a_1+b_1, a_2+b_2, \dots, a_n+b_n)^T$$

and, for all $r \in \mathbb{F}$,

$$r(a_1, a_2, \dots, a_n)^T = (ra_1, ra_2, \dots, ra_n)^T.$$

Example 3.4. When $\mathbb{F} = \mathbb{F}_2$, the elements of $\mathbb{F}^n$ are called *n-bit strings.* For example, if $n = 4$, we have 4-bit strings such as 0000, 1000, 0100, 1100 and so forth. Since there are 4 places to put either a 0 or a 1, there are $2^4 = 16$ 4-bit strings. Binary strings have the nice property that each string is its own additive inverse. Also, the string 1111 changes the parity of each component. That is, $0101 + 1111 = 1010$. The space of n-bit strings are the fundamental objects of coding theory.

Example 3.5. Similarly, if $\mathbb{F} = Z_p$, we can consider the space of all p-ary strings $a_1 a_2 \dots a_n$ of elements of $\mathbb{F}_p$. Note that when we consider strings, we often, for simplicity, drop the commas. However, you have to remember that a string $a_1 a_2 \dots a_n$ can also be confused with a product in $\mathbb{F}$. The space $(\mathbb{F}_p)^n$ is frequently denoted as $V(n, p)$.

Example 3.6. This example generalizes $\mathbb{R}^{[a,b]}$. Let S be any set and define $\mathbb{R}^S$ to be the set of all real valued functions whose domain is S. We define addition and scalar multiplication pointwise, exactly as for $\mathbb{R}^{[a,b]}$. Then R^S is a vector space over $\mathbb{R}$. Notice that $\mathbb{R}^n$ is nothing but $\mathbb{R}^S$, where $S = \{1, 2, \dots, n\}$. This is because specifying the n-tuple $\mathbf{a} = (a_1, a_2, \dots a_n)^T \in \mathbb{R}^n$ is the same as defining a function $f_{\mathbf{a}} : S \to \mathbb{R}$ by setting $f_{\mathbf{a}}(i) = a_i$.

Example 3.7. The set $\mathcal{P}_n$ of all polynomials

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

with real coefficients having degree at most n is a real vector space under pointwise addition and scalar multiplication defined as above. Pointwise addition of two polynomials amounts to adding the coefficients of x^i in each polynomial, for every i. Scalar multiplication by r is multiplying each term $a_i x^i$ by r. Notice the similarity between these operations on polynomials and component-wise addition and scalar multiplication on $\mathbb{R}^{n+1}$.

$$(a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0) + (b_n x^n + b_{n-1} x^{n-1} + \cdots + b_1 x + b_0) =$$

$$(a_n + b_n) x^n + (a_{n-1} + b_{n-1}) x^{n-1} + \cdots + (a_1 + b_1) x + (a_0 + b_0),$$

while

$$\mathbf{a} + \mathbf{b} = (a_0, a_1, a_2, \dots a_n)^T + (b_0, b_1, b_2, \dots b_n)^T$$

$$= (a_0 + b_0, a_1 + b_1, a_2 + b_2, \dots a_n + b_n)^T.$$

In this sense, $\mathcal{P}_n$ and $\mathbb{R}^{n+1}$ are indistinguishable as vector spaces.

Example 3.8. Consider the differential equation

$$y'' + ay' + by = 0, \tag{3.15}$$

where a and b are real constants. This is an example of a homogeneous linear second order differential equation with constant coefficients. The set of twice differentiable functions on $\mathbb{R}$ which satisfy (3.15) is a real vector space.

3.3.2 Inner product spaces

The set $C(a, b)$ of continuous real valued functions on the interval $[a, b]$ defined in the previous subsection is one of the most basic vector spaces in mathematics. Although $C(a, b)$ is much more complicated than $\mathbb{R}^n$, it has an important structure in common with $\mathbb{R}^n$ which lets us partially extend our intuition about $\mathbb{R}^n$ to $C(a, b)$. Namely, we can define an inner product (f, g) of $f, g \in C(a, b)$ by

$$(f, g) = \int_a^b f(t)g(t)\mathrm{d}t.$$

The first three axioms for the Euclidean inner product (dot product) on $\mathbb{R}^n$ are verified by applying standard facts about integration proved (or at least stated) in any calculus book. Recall that the last axiom requires that $(f, f) \geq 0$ and $(f, f) = 0$ only if $f = 0$. The verification of this requires some argument, and we leave it as an exercise in elementary real analysis.

If a real vector space admits an inner product, then the notions of length and distance can be introduced by just copying the definitions used for $\mathbb{R}^n$ in Chapter 1. The length $||f||$ of any $f \in C(a, b)$ is defined to be

$$||f|| := (f, f)^{1/2} = \big(\int_a^b f(t)^2\mathrm{d}t\big)^{1/2},$$

and the distance between $f, g \in C(a, b)$ is defined to be

$$d(f, g) = ||f - g|| = \big(\int_a^b (f(t) - g(t))^2\mathrm{d}t\big)^{1/2}.$$

Just as for the Euclidean inner product on $\mathbb{R}^n$, we can say two functions $f, g \in C(a, b)$ are *orthogonal* if $(f, g) = \int_a^b f(t)g(t)\mathrm{d}t = 0$. Then the tools we developed from the Euclidean inner product on $\mathbb{R}^n$ such as projections and orthogonal decompositions extend word by word to $C(a, b)$. For example, $\cos t$ and $\sin t$ are orthogonal on $[0, 2\pi]$ because $\int_0^{2\pi} \cos t \sin t\mathrm{d}t = 0$.

Although the notion of orthogonality for $C(a,b)$ doesn't have any obvious geometric meaning, it nevertheless enables us to extend our intuitive concept of orthogonality into a new situation. In fact, this extension turns out to be extremely important since it leads to the idea of expanding a function in terms of possibly infinitely many mutually orthogonal functions. These infinite series expansions are called Fourier series. For example, the functions $\cos mx$, $m = 0, 1, 2, \ldots$ are orthogonal on $[0, 2\pi]$, and the Fourier cosine series for $f \in C(0, 2\pi)$ has the form

$$f(x) = \sum_{m=0}^{\infty} a_m \cos mx,$$

where

$$a_m = \int_0^{2\pi} f(t) \cos mt \mathrm{d}t / \int_0^{\pi} \cos^2 mt \mathrm{d}t.$$

We call a_m the *Fourier coefficient* of f with respect to $\cos mt$. Notice that $a_m \cos mx$ is the projection of f on $\cos mx$. This series is an infinite version of the formula in Proposition 1.3.

If we only take finitely many terms of the above Fourier series, we obtain a *least squares approximation* to f.

Example 3.9. Suppose $[a,b] = [-1,1]$. Then the functions 1 and x are orthogonal. In fact, x^k and x^m are orthogonal if k is even and m is odd, or vice versa. Indeed,

$$(x^k, x^m) = \int_{-1}^{1} x^k \cdot x^m \mathrm{d}x = \int_{-1}^{1} x^{k+m} \mathrm{d}x = 0,$$

since $k+m$ is odd. On the other hand, the projection of x^2 on the constant function 1 is $r1$, where $r = \frac{1}{2}\int_{-1}^{1} 1 \cdot x^2 \mathrm{d}x = \frac{1}{3}$. Thus, $x^2 - 1/3$ is orthogonal to the constant function 1 on $[-1,1]$, and $x^2 = (x^2 - 1/3) + 1/3$ is an orthogonal decomposition of x^2 on $[-1,1]$.

Similarly, by arguing exactly as in §2, we immediately obtain a Cauchy-Schwartz inequality on $C(a,b)$.

Cauchy-Schwartz Inequality for $C(a,b)$. *For any $f, g \in C(a,b)$, the inequality*

$$|\int_a^b f(t)g(t)\mathrm{d}t| \leq \big(\int_a^b f(t)^2 \mathrm{d}t\big)^{1/2} \big(\int_a^b g(t)^2 \mathrm{d}t\big)^{1/2}$$

holds. Equality holds if and only if one of the functions is a constant multiple of the other.

3.3.3 Subspaces and Spanning Sets

We next consider the extremely important notion of a subspace.

Definition 3.10. Let V be vector space over a field $\mathbb{F}$. A non-empty subset W of V is called a *linear subspace of* V, or simply a *subspace*, provided $\mathbf{a} + \mathbf{b} \in W$ and $r\mathbf{a}$ are in W whenever $\mathbf{a}, \mathbf{b} \in W$ and $r \in \mathbb{F}$.

The following Proposition is immediate.

Proposition 3.19. *Every subspace W of V is a vector space over $\mathbb{F}$ in its own right.*

Proof. This is left as an exercise. □

Notice that every subspace of a vector space contains the zero vector $\mathbf{0}$ (why?). In fact, $\{\mathbf{0}\}$ is itself a subspace, called the *trivial subspace.* Hence, if the constant term of a homogeneous linear equation $ax + by + cz = d$ above is nonzero, then the solution set cannot be a subspace.

Here is a fundamental example of a subspace of $\mathbb{R}^3$.

Example 3.10. The solutions $(x, y, z)^T \in \mathbb{R}^3$ of a homogeneous linear equation $ax + by + cz = 0$, with $a, b, c \in \mathbb{R}$ make up the plane consisting of all vectors orthogonal to $(a, b, c)^T$. By the properties of the dot product, the sum of any two solutions is another solution, and any scalar multiple of a solution is a solution. Hence the solution set of a homogeneous linear equation in three variables is a subspace of $\mathbb{R}^3$. More generally, the solution set of a homogeneous linear equation in n variables with real coefficients is a subspace of $\mathbb{R}^n$. If the coefficients are in the field $\mathbb{F}$, then the solutions in $\mathbb{F}^n$ make up a subspace of $\mathbb{F}^n$.

The subspaces of $\mathbb{R}^2$ are easily described. They are $\{\mathbf{0}\}$, any line through $\mathbf{0}$ and $\mathbb{R}^2$ itself. We will consider the subspaces of $\mathbb{R}^3$ below. Try to guess what they are before reading further.

A basic method for constructing subspaces of a given vector space is to take linear combinations.

Definition 3.11. Let $\mathbf{v}_1, \dots, \mathbf{v}_k$ be vectors in V, and let $r_1, \dots, r_k$ be any elements of $\mathbb{F}$. Then the vector

$$\mathbf{w} = \sum_{i=1}^{k} r_i \mathbf{v}_i$$

is called a *linear combination* of $\mathbf{v}_1, \dots, \mathbf{v}_k$. A subspace W which consists of all linear combinations of an arbitrary collection of vectors in V, say $\mathbf{v}_1, \dots, \mathbf{v}_k$, is said to be spanned by $\mathbf{v}_1, \dots, \mathbf{v}_k$.

Remark: We only need to mention that a sum such as $\mathbf{w} = \sum_{i=1}^{k} r_i\mathbf{v}_i$ is defined inductively in the same way arbitrary sums in a field are defined (see Section 3.1.2). Moreover, vector sums can be grouped arbitrarily in the same way as sums and products in a field are in Proposition 3.3. The details are the same as Section 3.1.2.

Proposition 3.19 says that subspaces are closed under taking linear combinations. It also asserts the converse. The set of all linear combinations of a collection of vectors in V is a subspace of V. We will denote the subspace spanned by $\mathbf{v}_1, \dots, \mathbf{v}_k$ by $\mathrm{span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$.

As previously noted, lines L and planes P in $\mathbb{R}^3$ containing $\mathbf{0}$ are subspaces of $\mathbb{R}^3$. Every line L is by definition $\mathrm{span}\{\mathbf{a}\}$ for some (in fact, any) nonzero $\mathbf{a} \in L$. Is every plane P the span of a set of vectors? Well, if $\mathbf{a}$ and $\mathbf{b}$ are two non-collinear vectors in P, then $W := \mathrm{span}\{\mathbf{a}, \mathbf{b}\}$ is contained in P. The question remains as to whether $W = P$. To see why the answer is yes, as expected, you can argue as follows. Let $\mathbf{n}$ denote any non-zero normal to P, and take any $\mathbf{c} \in P$. The line through $\mathbf{a}$ and $\mathbf{0}$ and the line through $\mathbf{b}$ and $\mathbf{0}$ both lie on P. Now any vector of the form $\mathbf{c} + t\mathbf{b}$ is orthogonal to $\mathbf{n}$, so the line through $\mathbf{c}$ parallel to $\mathbf{b}$ also lies on P. This line meets the line through $\mathbf{a}$ and $\mathbf{0}$ at some $r\mathbf{a}$ (why?). Next construct $s\mathbf{b}$ in the same way by interchanging the roles of $\mathbf{a}$ and $\mathbf{b}$. Then clearly, $\mathbf{c} = r\mathbf{a} + s\mathbf{b}$, because $\mathbf{c}$ is the intersection of the line through $r\mathbf{a}$ parallel to $\mathbf{b}$ and the line through $s\mathbf{b}$ parallel to $\mathbf{a}$. Hence $\mathbf{c} \in \mathrm{span}\{\mathbf{a}, \mathbf{b}\}$, so $P = \mathrm{span}\{\mathbf{a}, \mathbf{b}\}$.

On the other hand, if $\mathbf{a}$ and $\mathbf{b}$ are two non-collinear vectors in $\mathbb{R}^3$, then $\mathbf{n} = \mathbf{a} \times \mathbf{b}$ is orthogonal to any linear combination of $\mathbf{a}$ and $\mathbf{b}$. Thus we obtain a homogeneous equation satisfied by exactly those vectors in $P = \mathrm{span}\{\mathbf{a}, \mathbf{b}\}$. (We just showed above that every vector orthogonal to $\mathbf{n}$ is on P.) If $\mathbf{n} = (r, s, t)^T$, then an equation is $rx + sy + tz = 0$.

Example 3.11. Let P be the plane spanned by $(1, 1, 2)^T$ and $(-1, 0, 1)^T$. Then $(1, 1, 2)^T \times (-1, 0, 1)^T = (1, -3, 1)^T$ is a normal to P, so an equation for P is $x - 3y + z = 0$.

3.3.4 Linear Systems and Matrices Over an Arbitrary Field

Although we developed the theory of linear systems over the reals, the only reason we didn't use an arbitrary field is that the definition hadn't yet been made. In fact, the material covered in Chapter 2 pertaining to linear systems and matrices goes through word for word when we use an arbitrary field $\mathbb{F}$. Thus we have $m \times n$ matrices over $\mathbb{F}$, which will be denoted by $\mathbb{F}^{m\times n}$, and linear systems $A\mathbf{x} = \mathbf{b}$, where $A \in \mathbb{F}^{m\times n}$, $\mathbf{x} \in \mathbb{F}^n$ and $\mathbf{b} \in \mathbb{F}^m$.

Row reduction, matrix inversion etc. all go through as for $\mathbb{R}$. We will not bother to restate all the results, but we will use them when needed. The matrix group $GL(n, \mathbb{R})$ is replaced by its counterpart $GL(n, \mathbb{F})$, which is also a matrix group. One thing to be careful of, however, is that in $\mathbb{R}^n$, if $\mathbf{x}^T\mathbf{x} = 0$, then $\mathbf{x} = \mathbf{0}$. This is false for most other fields such as $\mathbb{F}_p$. It is even false for $\mathbb{C}$ since $(1 \;\; i)\begin{pmatrix}1\\ i\end{pmatrix} = 1 + i^2 = 0$.

Exercises

Exercise 3.19. Let V be a vector space. Show that $0\mathbf{a} = \mathbf{0}$.

Exercise 3.20. Let V be a vector space. Show that for any $\mathbf{a} \in V$, the vector $(-1)\mathbf{a}$ is an additive inverse of V. In other words, prove the formula $(-1)\mathbf{a} = -\mathbf{a}$.

Exercise 3.21. Describe all subspaces of $\mathbb{R}^3$.

Exercise 3.22. Which of the following subsets of $\mathbb{R}^2$ is not a subspace?

(a) The line $x = y$;

(b) The unit circle;

(c) The line $2x + y = 1$;

s(d) The first octant $x, y \geq 0$.

Exercise 3.23. Prove that every line through the origin and plane through the origin in $\mathbb{R}^3$ are subspaces.

Exercise 3.24. Find all the subspaces of the vector space $V(n,p) = (\mathbb{F}_p)^n$ in the following cases:

(i) $n = p = 2$;

(ii) $n = 2$, $p = 3$; and

(iii) $n = 3$, $p = 2$.

Exercise 3.25. How many points lie on a line in $V(n,p)$? On a plane?

Exercise 3.26. Let $\mathbb{F} = \mathbb{F}_2$. Find all solutions in $\mathbb{F}^4$ of the equation $w+x+y+z=0$. Compare the number of solutions with the number of elements $\mathbb{F}^4$ itself has?

Exercise 3.27. Consider the real vector space $V = C(0, 2\pi)$ with the inner product defined in §3.3.2.

(a) Find the length of $\sin^2 t$ in V.

(b) Compute the inner product $(\cos t, \sin^2 t)$.

(c) Find the projection of $\sin^2 t$ on each of the functions $1, \cos t$, and $\sin t$ in V.

(d) Are 1, $\cos t$ and $\sin t$ mutually orthogonal as elements of V?

(e) How would you define the orthogonal projection of $\sin^2 t$ onto the subspace W of V spanned by $1, \cos t$, and $\sin t$?

(f) Describe the subspace W of part (e).

Exercise 3.28. Assume $f \in C(a, b)$. Recall that the average value of f over $[a, b]$ is defined to be

$$\frac{1}{b-a} \int_a^b f(t)dt.$$

Show that the average value of f over $[a, b]$ is the projection of f on 1. Does this suggest an interpretation of the average value?

Exercise 3.29. Let $f, g \in C(a, b)$. Give a formula for the scalar t which minimizes

$$||f - tg||^2 = \int_a^b (f(x) - tg(x))^2 dx.$$

Exercise 3.30. Find a spanning set for the plane $3x - y + 2z = 0$ in $\mathbb{R}^3$.

Exercise 3.31. Find an equation for the plane in $\mathbb{R}^3$ through the origin containing both $(1, 2, -1)^T$ and $(3, 0, 1)^T$.

Exercise 3.32. Let L be the line obtained by intersecting the two planes in the previous two exercises. Express L as $\text{span}\{\mathbf{a}\}$ for some $\mathbf{a}$.

Exercise 3.33. Describe all subspaces of $\mathbb{R}^4$ and $\mathbb{R}^5$.

3.4 Summary

The purpose of this chapter was to introduce two fundamental notions: fields and vector spaces. Fields are the number systems where we can add, subtract, multiply and divide in the usual sense. The basic examples were the rationals $\mathbb{Q}$, which form the smallest field containing the integers, the reals (which are hard to define, so we didn't), the prime fields $\mathbb{F}_p$, which are the systems which support modular arithmetic, and the queen of all fields, the complex numbers $\mathbb{C}$. The basic property of $\mathbb{C}$ is that it contains $\mathbb{R}$ and is algebraically closed. A vector space is what happens when a field is cloned. That is, we get the space $\mathbb{F}^n$ of n-tuples of elements in $\mathbb{F}$. In a vector space, we can add elements and operate on them by scalars. General vector spaces do not have a multiplication, although some specific examples do. Vector spaces V have subspaces, the most common example of a subspace being the set of all linear combinations of a subcollection of the vectors in V. We mentioned a special class of vector spaces over $\mathbb{R}$, namely inner product spaces. These spaces are just like $\mathbb{R}^n$ except for the fact that they are frequently not spanned by finite sets as $\mathbb{R}^n$ is. However, some of the properties we developed for $\mathbb{R}^n$, such as orthogonal projection and the Cauchy-Schwartz Inequalty, go through in the general case just as they did in $\mathbb{R}^n$.

For example, $C(a, b)$ is an inner product space that doesn't have this property. We also pointed out that the theory of linear systems and matrix theory, two themes that were carried out over $\mathbb{R}$ in Chapter 2, have identical versions over an arbitrary field.