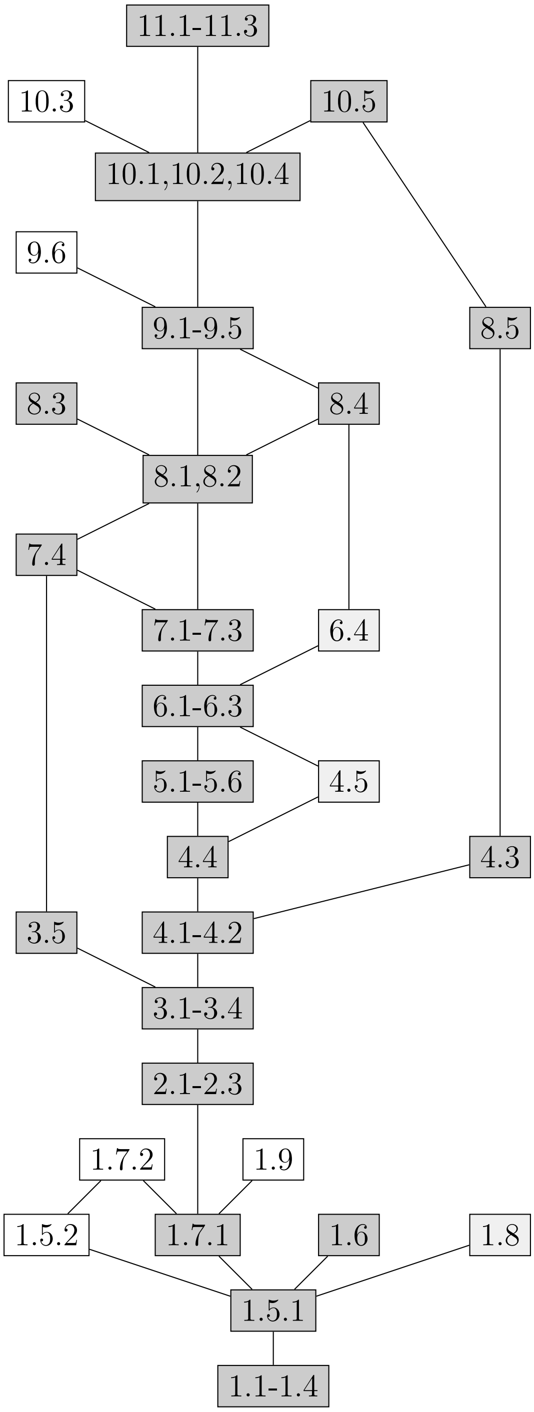
Linear Algebra I

Ronald van Luijk, 2018

With many parts from “Linear Algebra I” by Michael Stoll, 2007

Contents

Dependencies among sections



CHAPTER 1

Euclidean space: lines and hyperplanes

This chapter deals, for any non-negative integer n, with *Euclidean n-space* $\mathbb{R}^n$, which is the set of all (ordered) sequences of n real numbers, together with a distance that we will define. We make it slightly more general, so that we can also apply our theory to, for example, the rational numbers instead of the real numbers: instead of just the set $\mathbb{R}$ of real numbers, we consider any *subfield* of $\mathbb{R}$. At this stage, it suffices to say that a *subfield* of $\mathbb{R}$ is a nonempty subset $F \subset \mathbb{R}$ containing 0 and 1, in which we can multiply, add, subtract, and divide (except by 0); that is, for any $x, y \in F$, also the elements $xy, x + y, x - y$ (and x/y if $y \neq 0$) are contained in F. We refer the interested reader to Appendix B for a more precise definition of a field in general.

Therefore, for this entire chapter (and only this chapter), we let F denote a subfield of $\mathbb{R}$, such as the field $\mathbb{R}$ itself or the field $\mathbb{Q}$ of rational numbers. Furthermore, we let n denote a non-negative integer.

1.1. Definition

An *n-tuple* is an ordered sequence of n objects. We let F^n denote the set of all n-tuples of elements of F. For example, the sequence

$$\big(-17, 0, 3, 1+\sqrt{2}, e^\pi \big)$$

is an element of $\mathbb{R}^5$. The five numbers are separated by commas. In general, if we have n numbers $x_1, x_2, \ldots, x_n \in F$, then

$$x = (x_1, x_2, \ldots, x_n)$$

is an element of F^n. Again, the numbers are separated by commas. Such n-tuples are called *vectors*; the numbers in a vector are called *coordinates*. In other words, the i-th coordinate of the vector $x = (x_1, x_2, \ldots, x_n)$ is the number x_i.

We define an *addition* by adding two elements of F^n coordinate-wise:

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

For example, the sequence $(12, 14, 16, 18, 20, 22, 24)$ is an element of $\mathbb{R}^7$ and we have

$$(12, 14, 16, 18, 20, 22, 24) + (11, 12, 13, 14, 13, 12, 11) = (23, 26, 29, 32, 33, 34, 35).$$

Unsurprisingly, we also define a coordinate-wise *subtraction*:

$$(x_1, x_2, \ldots, x_n) \ominus (y_1, y_2, \ldots, y_n) = (x_1 - y_1, x_2 - y_2, \ldots, x_n - y_n).$$

Until the end of this section, we denote the sum and the difference of two elements $x, y \in F^n$ by $x \oplus y$ and $x \ominus y$, respectively, in order to distinguish them from the usual addition and subtraction of two numbers. Similarly, we define a *scalar multiplication*: for any element $\lambda \in F$, we set

$$\lambda \odot (x_1, x_2, \ldots, x_n) = (\lambda x_1, \lambda x_2, \ldots, \lambda x_n).$$

This is called scalar multiplication because the elements of F^n are scaled; the elements of F, by which we scale, are often called *scalars*. We abbreviate the special vector $(0, 0, \ldots, 0)$ consisting of only zeros by 0, and for any vector $x \in F^n$, we abbreviate the vector $0 \ominus x$ by $-x$. In other words, we have

$$-(x_1, x_2, \ldots, x_n) = (-x_1, -x_2, \ldots, -x_n).$$

Because our new operations are all defined coordinate-wise, they obviously satisfy the following properties:

(1) for all $x, y \in F^n$, we have $x \oplus y = y \oplus x$;
(2) for all $x, y, z \in F^n$, we have $(x \oplus y) \oplus z = x \oplus (y \oplus z)$;
(3) for all $x \in F^n$, we have $0 \oplus x = x$ and $1 \odot x = x$;
(4) for all $x \in F^n$, we have $(-1) \odot x = -x$ and $x \oplus (-x) = 0$;
(5) for all $x, y, z \in F^n$, we have $x \oplus y = z$ if and only if $y = z \ominus x$;
(6) for all $x, y \in F^n$, we have $x \ominus y = x \oplus (-y)$;
(7) for all $\lambda, \mu \in F$ and $x \in F^n$, we have $\lambda \odot (\mu \odot x) = (\lambda \cdot \mu) \odot x$;
(8) for all $\lambda, \mu \in F$ and $x \in F^n$, we have $(\lambda + \mu) \odot x = (\lambda \odot x) \oplus (\mu \odot x)$;
(9) for all $\lambda \in F$ and $x, y \in F^n$, we have $\lambda \odot (x \oplus y) = (\lambda \odot x) \oplus (\lambda \odot y)$.

In fact, in the last two properties, we may also replace $+$ and $\oplus$ by $-$ and $\ominus$, respectively, but the properties that we then obtain follow from the properties above. All these properties together mean that the operations $\oplus$, $\ominus$, and $\odot$ really behave like the usual addition, subtraction, and multiplication, as long as we remember that the **scalar multiplication is a multiplication of a scalar with a vector, and *not* of two vectors!**.

We therefore will usually leave out the circle in the notation: instead of $x \oplus y$ and $x \ominus y$ we write $x + y$ and $x - y$, and instead of $\lambda \odot x$ we write $\lambda \cdot x$ or even λx.

As usual, scalar multiplication takes priority over addition and subtraction, so when we write $\lambda x \pm \mu y$ with $\lambda, \mu \in F$ and $x, y \in F^n$, we mean $(\lambda x) \pm (\mu y)$. Also as usual, when we have vectors $x_1, x_2, \ldots, x_t \in F^n$, the expression $x_1 \pm x_2 \pm x_3 \pm \cdots \pm x_t$ should be read from left to right, so it stands for

$$\underbrace{(\ldots((}_{t-2} x_1 \pm x_2) \pm x_3) \pm \cdots) \pm x_t.$$

If all the signs in the expression are positive (+), then any other way of putting the parentheses would yield the same by property (2) above.

1.2. Euclidean plane and Euclidean space

For $n = 2$ or $n = 3$ we can identify $\mathbb{R}^n$ with the pointed plane or three-dimensional space, respectively. We say *pointed* because they come with a special point, namely 0. For instance, for $\mathbb{R}^2$ we take an orthogonal coordinate system in the plane, with 0 at the origin, and with equal unit lengths along the two coordinate axes. Then the vector $p = (p_1, p_2) \in \mathbb{R}^2$, which is by definition nothing but a pair of real numbers, corresponds with the point in the plane whose coordinates are p_1 and p_2. In this way, the vectors get a geometric interpretation. We can similarly identify $\mathbb{R}^3$ with three-dimensional space. We will often make these identifications and talk about points as if they are vectors, and vice versa. By doing so, we can now add points in the plane, as well as in space! Figure 1.1 shows the two points $p = (3, 1)$ and $q = (1, 2)$ in $\mathbb{R}^2$, as well as the points $0, -p, 2p, p + q$, and $q - p$.

For $n = 2$ or $n = 3$, we may also represent vectors by arrows in the plane or space, respectively. In the plane, the *arrow* from the point $p = (p_1, p_2)$ to the

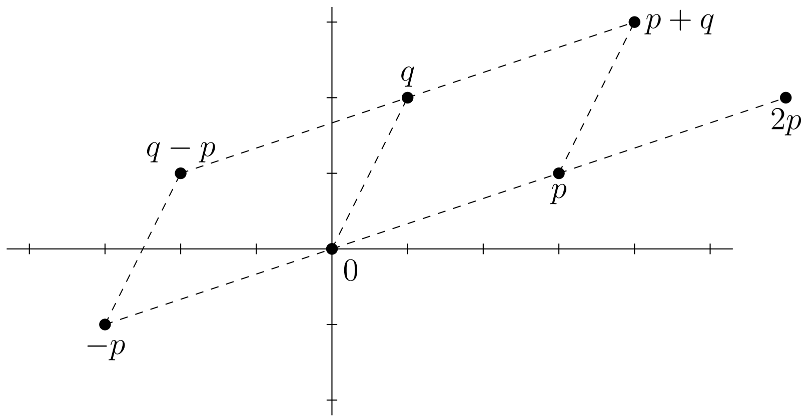


FIGURE 1.1. Two points p and q in $\mathbb{R}^2$, as well as $0, -p, 2p, p+q$, and $q-p$

point $q = (q_1, q_2)$ *represents* the vector $v = (q_1 - p_1, q_2 - p_2) = q - p$. (A careful reader notes that here we do indeed identify points and vectors.) We say that the point p is the tail of the arrow and the point q is the head. Note the distinction we make between an arrow and a vector, the latter of which is by definition just a sequence of real numbers. Many different arrows may represent the same vector v, but all these arrows have the same direction and the same length, which together narrow down the vector. One arrow is special, namely the one with 0 as its tail; the head of this arrow is precisely the point $q - p$, which is the point identified with v! See Figure 1.2, in which the arrows are labeled by the name of the vector v they represent, and the points are labeled either by their own name (p and q), or the name of the vector they correspond with (v or 0). Note that besides $v = q - p$, we (obviously) also have $q = p + v$.

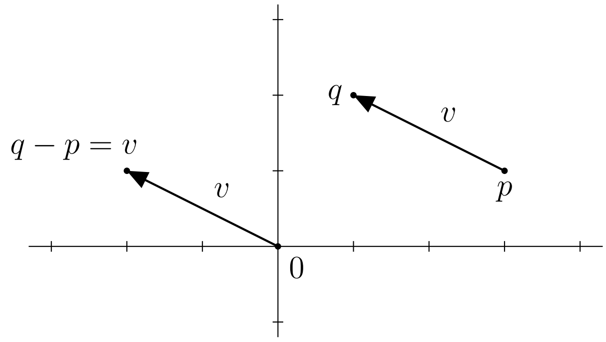
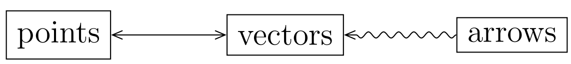


FIGURE 1.2. Two arrows representing the same vector $v = (-2, 1)$

Of course we can do the same for $\mathbb{R}^3$. For example, take the points $p = (3, 1, -4)$ and $q = (-1, 2, 1)$ and set $v = q - p$. Then we have $v = (-4, 1, 5)$. The arrow from p to q has the same direction and length as the arrow from 0 to the point $(-4, 1, 5)$. Both these arrows represent the vector v.

Note that we now have three notions: *points*, *vectors*, and *arrows*.



Vectors and points can be identified with each other, and arrows represent vectors (and thus points).

We can now interpret negation, scalar multiples, sums, and differences of vectors (as defined in Section 1.1) geometrically, namely in terms of points and arrows.

For points this was already depicted in Figure 1.1. If p is a point in $\mathbb{R}^2$, then $-p$ is obtained from p by rotating it 180 degrees around 0; for any real number $\lambda > 0$, the point λp is on the half line from 0 through p with distance to 0 equal to (λ times the distance from p to 0). For any points p and q in $\mathbb{R}^2$ such that $0, p$, and q are not collinear, the points $p+q$ and $q-p$ are such that the four points 0, p, $p+q$, and q are the vertices of a parallelogram with p and q opposite vertices, and the four points 0, $-p$, $q-p$, q are the vertices of a parallelogram with $-p$ and q opposite vertices.

In terms of arrows we get the following. If a vector v is represented by a certain arrow, then $-v$ is represented by any arrow with the same length but opposite direction; furthermore, for any positive $\lambda \in \mathbb{R}$, the vector λv is represented by the arrow obtained by scaling the arrow representing v by a factor λ.

If v and w are represented by two arrows that have common tail p, then these two arrows are the sides of a unique parallelogram; the vector $v + w$ is represented by a diagonal in this parallelogram, namely the arrow that also has p as tail and whose head is the opposite point in the parallelogram. An equivalent description for $v+w$ is to take two arrows, for which the head of the one representing v equals the tail of the one representing w; then $v + w$ is represented by the arrow from the tail of the first to the head of the second. See Figure 1.3.

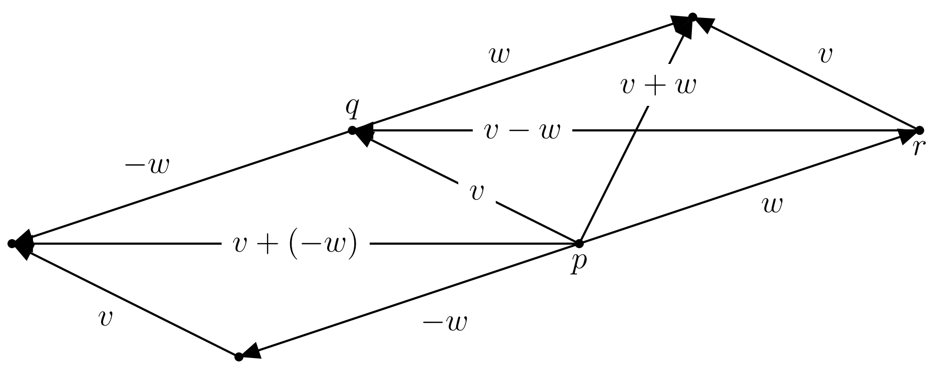


FIGURE 1.3. Geometric interpretation of addition and subtraction

The description of laying the arrows head-to-tail generalises well to the addition of more than two vectors. Let $v_1, v_2, \ldots, v_t$ in $\mathbb{R}^2$ or $\mathbb{R}^3$ be vectors and $p_0, p_1, \ldots, p_t$ points such that v_i is represented by the arrow from p_{i-1} to p_i. Then the sum $v_1 + v_2 + \cdots + v_t$ is represented by the arrow from p_0 to p_t. See Figure 1.4.

For the same v and w, still represented by arrows with common tail p and with heads q and r, respectively, the difference $v-w$ is represented by the other diagonal in the same parallelogram, namely the arrow from r to q. Another construction for $v-w$ is to write this difference as the sum $v+(-w)$, which can be constructed as described above. See Figure 1.3.

Representing vectors by arrows is very convenient in physics. In classical mechanics, for example, we identify forces applied on a body with vectors, which are often depicted by arrows. The total force applied on a body is then the sum of all the forces in the sense that we have defined it.

Motivated by the case $n = 2$ and $n = 3$, we will sometimes refer to vectors in $\mathbb{R}^n$ as *points* in general. Just as arrows in $\mathbb{R}^2$ and $\mathbb{R}^3$ are uniquely determined by their head and tail, for general n we define an *arrow* to be a pair (p, q) of points $p, q \in \mathbb{R}^n$ and we say that this arrow *represents* the vector $q-p$. The points p and q are the *tail* and the *head* of the arrow (p, q).

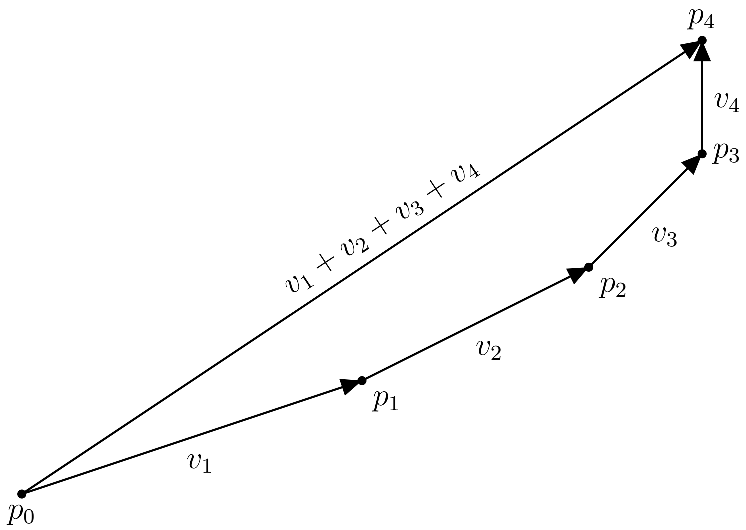


FIGURE 1.4. Adding four vectors

Exercises

1.2.1. Compute the sum of the given vectors v and w in $\mathbb{R}^2$ and draw a corresponding picture by identifying the vectors with points or representing them by arrows (or both) in $\mathbb{R}^2$.

(1) $v = (-2, 5)$ and $w = (7, 1)$,
(2) $v = 2 \cdot (-3, 2)$ and $w = (1, 3) + (-2, 4)$,
(3) $v = (-3, 4)$ and $w = (4, 3)$,
(4) $v = (-3, 4)$ and $w = (8, 6)$,
(5) $v = w = (5, 3)$.

1.2.2. Let $p, q, r, s \in \mathbb{R}^2$ be the vertices of a parallelogram, with p and r opposite vertices. Show that $p + r = q + s$.

1.2.3. Let $p, q \in \mathbb{R}^2$ be two points such that $0, p$, and q are not collinear. How many parallelograms are there with $0, p$, and q as three of the vertices? For each of these parallelograms, express the fourth vertex in terms of p and q.

1.3. The standard scalar product

We now define the *(standard) scalar product*[1] on F^n.

Definition 1.1. For any two vectors $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ in F^n we define the *standard scalar product* of x and y as

$$\langle x, y \rangle = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

We will often leave out the word 'standard'. The scalar product derives its name from the fact that $\langle x, y \rangle$ is a scalar, that is, an element of F. In LaTeX , the scalar product is *not* written `$<x,y>$`, but `$\langle x,y \rangle$`!

[1] The scalar *product* should not be confused with the scalar *multiplication*; the scalar multiplication takes a scalar $\lambda \in F$ and a vector $x \in F^n$, and yields a vector λx, while the scalar product takes two vectors $x, y \in F^n$ and yields a scalar $\langle x, y \rangle$.

Warning 1.2. While the name scalar product and the notation $\langle x, y\rangle$ for it are standard, in other pieces of literature, the standard scalar product is also often called the *(standard) inner product*, or the *dot product*, in which case it may get denoted by $x \cdot y$. Also, in other pieces of literature, the notation $\langle x, y\rangle$ may be used for other notions. One should therefore always check which meaning of the notation $\langle x, y\rangle$ is used.[2]

Example 1.3. Suppose we have $x = (3, 4, -2)$ and $y = (2, -1, 5)$ in $\mathbb{R}^3$. Then we get

$$\langle x, y\rangle = 3 \cdot 2 + 4 \cdot (-1) + (-2) \cdot 5 = 6 + (-4) + (-10) = -8.$$

The scalar product satisfies the following useful properties.

Proposition 1.4. *Let $\lambda \in F$ be an element and let $x, y, z \in F^n$ be elements. Then the following identities hold.*

(1) $\langle x, y\rangle = \langle y, x\rangle$,
(2) $\langle \lambda x, y\rangle = \lambda \cdot \langle x, y\rangle = \langle x, \lambda y\rangle$,
(3) $\langle x, y + z\rangle = \langle x, y\rangle + \langle x, z\rangle$.

Proof. Write x and y as

$$x = (x_1, x_2, \dots, x_n) \qquad \text{and} \qquad y = (y_1, y_2, \dots, y_n).$$

Then $x_1, \dots, x_n$ and $y_1, \dots, y_n$ are real numbers, so we obviously have $x_iy_i = y_ix_i$ for all integers i with $1 \le i \le n$. This implies

$$\langle x, y\rangle = x_1y_1 + x_2y_2 + \cdots + x_ny_n = y_1x_1 + y_2x_2 + \cdots + y_nx_n = \langle y, x\rangle,$$

which proves identity (1).

For identity (2), note that we have $\lambda x = (\lambda x_1, \lambda x_2, \dots, \lambda x_n)$, so

$$\begin{aligned}\langle \lambda x, y\rangle &= (\lambda x_1)y_1 + (\lambda x_2)y_2 + \ldots + (\lambda x_n)y_n \\ &= \lambda \cdot (x_1y_1 + x_2y_2 + \cdots + x_ny_n) = \lambda \cdot \langle x, y\rangle,\end{aligned}$$

which proves the first equality of (2). Combining it with (1) gives

$$\lambda \cdot \langle x, y\rangle = \lambda \cdot \langle y, x\rangle = \langle \lambda y, x\rangle = \langle x, \lambda y\rangle,$$

which proves the second equality of (2).

For identity (3), we write z as $z = (z_1, z_2, \dots, z_n)$. Then we have

$$\begin{aligned}\langle x, y + z\rangle &= x_1(y_1 + z_1) + x_2(y_2 + z_2) + \ldots + x_n(y_n + z_n) \\ &= (x_1y_1 + \ldots + x_ny_n) + (x_1z_1 + \ldots + x_nz_n) = \langle x, y\rangle + \langle x, z\rangle,\end{aligned}$$

which proves identity (3). □

Note that the equality $\langle x+y, z\rangle = \langle x, z\rangle+\langle y, z\rangle$ follows from properties (1) and (3). From the properties above, it also follows that we have $\langle x, y - z\rangle = \langle x, y\rangle - \langle x, z\rangle$ for all vectors $x, y, z \in F^n$; of course this is also easy to check directly.

Example 1.5. Let $L \subset \mathbb{R}^2$ be the line of all points $(x, y) \in \mathbb{R}^2$ that satisfy $3x + 5y = 7$. For the vector $a = (3, 5)$ and $v = (x, y)$, we have

$$\langle a, v\rangle = 3x + 5y,$$

[2] In fact, this warning holds for any notation...

so we can also write L as the set of all points $v \in \mathbb{R}^2$ that satisfy $\langle a, v \rangle = 7$.

Example 1.6. Let $V \subset \mathbb{R}^3$ be a plane. Then there are constants $p, q, r, b \in \mathbb{R}$, with p, q, r not all 0, such that V is given by

$$V = \{(x, y, z) \in \mathbb{R}^3 \ : \ px + qy + rz = b\}.$$

If we set $a = (p, q, r) \in \mathbb{R}^3$, then we can also write this as

$$V = \{v \in \mathbb{R}^3 : \langle a, v \rangle = b\}.$$

In examples 1.5 and 1.6, we used the terms *line* and *plane* without an exact definition. Lines in $\mathbb{R}^2$ and planes in $\mathbb{R}^3$ are examples of hyperplanes, which we define now.

Definition 1.7. A *hyperplane* in F^n is a subset of F^n that equals

$$\{\, v \in F^n \ : \ \langle a, v \rangle = b \,\}$$

for some nonzero vector $a \in F^n$ and some constant $b \in F$. A hyperplane in F^3 is also called a *plane*; a hyperplane in F^2 is also called a *line*.

Example 1.8. Let $H \subset \mathbb{R}^5$ be the subset of all quintuples $(x_1, x_2, x_3, x_4, x_5)$ of real numbers that satisfy

$$x_1 - x_2 + 3x_3 - 17x_4 - \tfrac{1}{2}x_5 = 13.$$

This can also be written as

$$H = \{\, x \in \mathbb{R}^5 \ : \ \langle a, x \rangle = 13 \,\}$$

where $a = (1, -1, 3, -17, -\frac{1}{2})$ is the vector of coefficients of the left-hand side of the equation, so H is a hyperplane.

As in this example, in general a hyperplane in F^n is a subset of F^n that is given by one linear equation $a_1x_1 + \ldots + a_nx_n = b$, with $a_1, \ldots, a_n, b \in F$. For any nonzero scalar λ, the equation $\langle a, x \rangle = b$ is equivalent with $\langle \lambda a, x \rangle = \lambda b$, so the hyperplane defined by $a \in F^n$ and $b \in F$ is also defined by λa and λb.

As mentioned above, a hyperplane in F^2 is nothing but a line in F^2. The following proposition states that instead of giving an equation for it, we can also describe the line in a different way: by specifying two vectors v and w. See Figure 1.5.

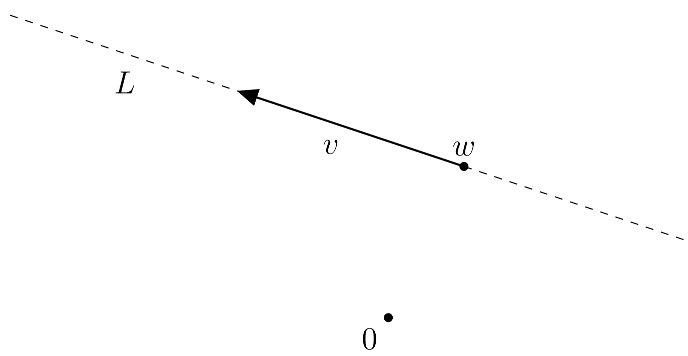


FIGURE 1.5. Parametrisation of the line L

Proposition 1.9. *For every line $L \subset F^2$, there are vectors $v, w \in F^2$, with v nonzero, such that we have*

$$L = \{\, w + \lambda v \in F^2 \ : \ \lambda \in F \,\}.$$

Conversely, for every vectors $v, w \in F^2$ with v nonzero, the set

$$\{\, w + \lambda v \in F^2 \ : \ \lambda \in F \,\}$$

is a line.

Proof. For the first statement, suppose L is a line. Then by definition there are $p, q, b \in F$, with p, q not both zero, such that L is the set of all points $(x, y) \in \mathbb{R}^2$ that satisfy $px + qy = b$. Let $w = (x_0, y_0)$ be a point of L, which exists because we can fix x_0 and solve for y_0 if q is nonzero, or the other way around if p is nonzero. Set $v = (-q, p)$. We denote the set $\{\, w + \lambda v \in F^2 \ : \ \lambda \in F \,\}$ of the proposition by M.

Since we have $px_0 + qy_0 = b$, we can write the equation for L as

$$p(x - x_0) + q(y - y_0) = 0. \tag{1.1}$$

To prove the equality $L = M$, we first prove the inclusion $L \subset M$. Let $z = (x, y) \in L$ be any point. We claim that there is a $\lambda \in F$ with $x - x_0 = -q\lambda$ and $y - y_0 = p\lambda$. Indeed, if $p \neq 0$, then we can set $\lambda = (y - y_0)/p$; using $y - y_0 = \lambda p$, equation (1.1) yields $x - x_0 = -q\lambda$. If instead we have $p = 0$, then $q \neq 0$, and we set $\lambda = -(x - x_0)/q$ to find $y - y_0 = p\lambda = 0$. This proves our claim, which implies $z = (x, y) = (x_0 - \lambda q, y_0 + \lambda p) = w + \lambda v \in M$, so we have $L \subset M$.

For the opposite inclusion, it is clear that for every scalar $\lambda \in F$, the point $w + \lambda v = (x_0 - \lambda q, y_0 + \lambda p)$ satisfies (1.1) and is therefore contained in L, so we have $M \subset L$. This finishes the proof of the first statement.

For the converse statement, let $p, q, x_0, y_0 \in F$ be such that $v = (-q, p)$ and $w = (x_0, y_0)$. Set $a = (p, q) \in F^2$ and $b = px_0 + qy_0 \in F$. We denote the set $\{\, w + \lambda v \in F^2 \ : \ \lambda \in F \,\}$ of the proposition by M. Let L be the line $\{\, u \in F^2 \ : \ \langle a, u \rangle = b \,\}$. Now the second and third paragraph of this proof can be repeated as a proof for the second statement of the proposition. □

We say that Proposition 1.9 gives a *parametrisation* of the line L, because for each scalar $\lambda \in F$ (the parameter) we get a point on L, and this yields a bijection (see Appendix A) between F and L.

Example 1.10. The points $(x, y) \in \mathbb{R}^2$ that satisfy $y = 2x + 1$ are exactly the points of the form $(0, 1) + \lambda(1, 2)$ with $\lambda \in \mathbb{R}$.

Inspired by the description of a line in Proposition 1.9, we define the notion of a line in F^n for general n. By Proposition 1.9, the following definition is equivalent with Definition 1.7 for $n = 2$.

Definition 1.11. A *line* in F^n is a subset of F^n that equals

$$\{\, w + \lambda v \ : \ \lambda \in F \,\}$$

for some vectors $v, w \in F^n$ with $v \neq 0$.

Proposition 1.12. *Let $p, q \in F^n$ be two distinct points. Then there is a unique line L that contains both. Moreover, every hyperplane that contains p and q also contains L.*

Proof. The existence of such a line is clear, as we can take $w = p$ and $v = q - p$ in Definition 1.11. The line L determined by these vectors contains $p = w + 0 \cdot v$ and $q = w + 1 \cdot v$. Conversely, suppose $v \neq 0$ and w are vectors such that the line $L' = \{ w + \lambda v \ : \ \lambda \in F \}$ contains p and q. Then there are $\mu, \nu \in F$ with $w + \mu v = p$ and $w + \nu v = q$. Subtracting these identities yields $q - p = (\nu - \mu)v$. Since p and q are distinct, we have $\nu - \mu \neq 0$. We write $c = (\nu - \mu)^{-1} \in F$. Then $v = c(q - p)$, and for every $\lambda \in F$ we have

$$w + \lambda v = p - \mu v + \lambda v = p + (\lambda - \mu)c(q - p) \in L.$$

This shows $L' \subset L$. The opposite inclusion $L \subset L'$ follows from the fact that for each $\lambda \in F$, we have $p + \lambda(q - p) = w + (\mu + \lambda c^{-1})v \in L'$. Hence, we find $L = L'$, which proves the first statement.

Let $a \in F^n$ be nonzero and $b \in F$ a constant and suppose that the hyperplane $H = \{v \in F^n \ : \ \langle a, v\rangle = b\}$ contains p and q. Then we have $\langle a, q-p\rangle = \langle a, q\rangle - \langle a, p\rangle = b - b = 0$. Hence, for each $\lambda \in F$, and the corresponding point $x = p + \lambda(q - p) \in L$, we have $\langle a, x\rangle = \langle a, p\rangle + \lambda\langle a, q - p\rangle = b + 0 = b$. This implies $x \in H$ and therefore $L \subset H$, which proves the second statement. $\square$

Notation 1.13. For every vector $a \in F^n$, we let $L(a)$ denote the set $\{\lambda a \ : \ \lambda \in F\}$ of all scalar multiples of a. If a is nonzero, then $L(a)$ is the line through 0 and a.

Exercises

1.3.1. For each of the pairs (v, w) given in Exercise 1.2.1, compute the scalar product $\langle v, w\rangle$.

1.3.2. For each of the following lines in $\mathbb{R}^2$, find vectors $v, w \in \mathbb{R}^2$, such that the line is given as in Proposition 1.9. Also find a vector $a \in \mathbb{R}^2$ and a number $b \in \mathbb{R}$, such that the line is given as in Definition 1.7.
(1) The line $\{(x, y) \in \mathbb{R}^2 \ : \ y = -3x + 4\}$.
(2) The line $\{(x, y) \in \mathbb{R}^2 \ : \ 2y = x - 7\}$.
(3) The line $\{(x, y) \in \mathbb{R}^2 \ : \ x - y = 2\}$.
(4) The line $\{v \in \mathbb{R}^2 \ : \ \langle c, v\rangle = 2\}$, with $c = (1, 2)$.
(5) The line through the points $(1, 1)$ and $(2, 3)$.

1.3.3. Write the following equations for lines in $\mathbb{R}^2$ with coordinates x_1 and x_2 in the form $\langle a, x\rangle = c$, that is, specify a vector a and a constant c in each case, such that the line equals the set $\{x \in \mathbb{R}^2 \ : \ \langle a, x\rangle = c\}$.
(1) $L_1 \colon 2x_1 + 3x_2 = 0$,
(2) $L_2 \colon x_2 = 3x_1 - 1$,
(3) $L_3 \colon 2(x_1 + x_2) = 3$,
(4) $L_4 \colon x_1 - x_2 = 2x_2 - 3$,
(5) $L_5 \colon x_1 = 4 - 3x_1$,
(6) $L_6 \colon x_1 - x_2 = x_1 + x_2$,
(7) $L_7 \colon 6x_1 - 2x_2 = 7$.

1.3.4. Let $V \subset \mathbb{R}^3$ be the subset given by

$$V = \{(x_1, x_2, x_3) \ : \ x_1 - 3x_2 + 3 = x_1 + x_2 + x_3 - 2\}.$$

Show that V is a plane as defined in Definition 1.7.

1.3.5. For each pair of points p and q below, determine vectors v, w, such that the line through p and q equals $\{w + \lambda v \,:\, \lambda \in F\}$.

(1) $p = (1, 0)$ and $q = (2, 1)$,
(2) $p = (1, 1, 1)$ and $q = (3, 1, -2)$,
(3) $p = (1, -1, 1, -1)$ and $q = (1, 2, 3, 4)$.

1.3.6. Let $a = (1, 2, -1)$ and $a' = (-1, 0, 1)$ be vectors in $\mathbb{R}^3$. Show that the intersection of the hyperplanes

$$H = \{v \in \mathbb{R}^3 \,:\, \langle a, v\rangle = 4\} \qquad \text{and} \qquad H' = \{v \in \mathbb{R}^3 \,:\, \langle a', v\rangle = 0\}$$

is a line as defined in Definition 1.11.

1.3.7. Let $p, q \in \mathbb{R}^n$ be distinct points. Show that the line through p and q (cf. Proposition 1.12) equals

$$\{\lambda p + \mu q \,:\, \lambda, \mu \in \mathbb{R} \text{ with } \lambda + \mu = 1\}.$$

1.4. Angles, orthogonality, and normal vectors

As in Section 1.2, we identify $\mathbb{R}^2$ and $\mathbb{R}^3$ with the Euclidean plane and Euclidean three-space: vectors correspond with points, and vectors can also be represented by arrows. In the plane and three-space, we have our usual notions of length, angle, and orthogonality. (Two intersecting lines are called *orthogonal*, or *perpendicular*, if the angle between them is $\pi/2$, or $90°$.) We will generalise these notions to F^n in the remaining sections of this chapter[3].

Because our field F is a subset of $\mathbb{R}$, we can talk about elements being 'positive' or 'negative' and 'smaller' or 'bigger' than other elements. This is used in the following proposition.

Proposition 1.14. *For every element $x \in F^n$ we have $\langle x, x\rangle \geq 0$, and equality holds if and only if $x = 0$.*

Proof. Write x as $x = (x_1, x_2, \ldots, x_n)$. Then $\langle x, x\rangle = x_1^2 + x_2^2 + \cdots + x_n^2$. Since squares of real numbers are non-negative, this sum of squares is also non-negative and it equals 0 if and only if each terms equals 0, so if and only if $x_i = 0$ for all i with $1 \leq i \leq n$. □

The vector $x = (x_1, x_2, x_3) \in \mathbb{R}^3$ is represented by the arrow from the point $(0, 0, 0)$ to the point (x_1, x_2, x_3); by Pythagoras' Theorem, the length of this arrow is $\sqrt{x_1^2 + x_2^2 + x_3^2}$, which equals $\sqrt{\langle x, x\rangle}$. See Figure 1.6, which is the only figure in this chapter where edges and arrows are labeled by their lengths, rather than the names of the vectors they represent. Any other arrow representing x has the same length. Similarly, the length of any arrow representing a vector $x \in \mathbb{R}^2$ equals $\sqrt{\langle x, x\rangle}$. We define the length of a vector in F^n for general $n \geq 0$ accordingly.

Definition 1.15. For any element $x \in F^n$ we define the *length* $\|x\|$ of x as

$$\|x\| = \sqrt{\langle x, x\rangle}.$$

[3]Those readers that adhere to the point of view that even for $n = 2$ and $n = 3$, we have not carefully defined these notions, have a good point and may skip the paragraph before Definition 1.15, as well as Proposition 1.19. They may take our definitions for general $n \geq 0$ as definitions for $n = 2$ and $n = 3$ as well.

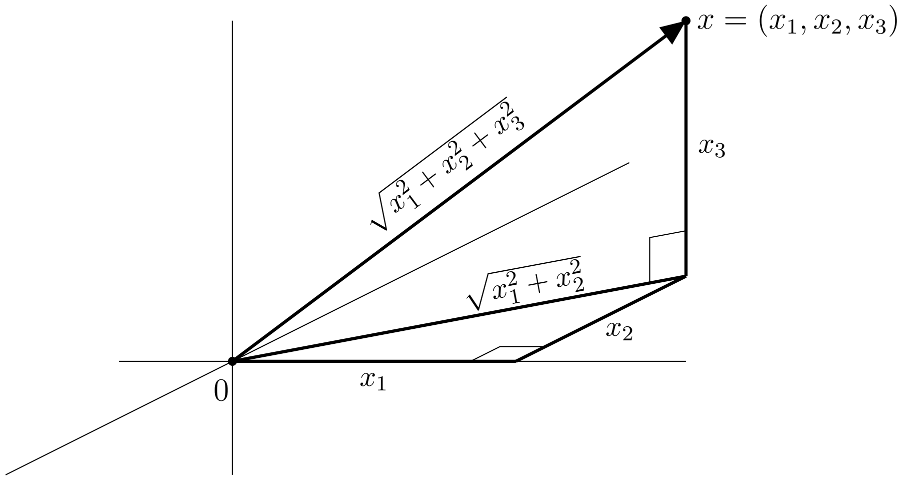


FIGURE 1.6. The length of an arrow

Note that by Proposition 1.14, we can indeed take the square root in $\mathbb{R}$, but the length $\|x\|$ may not be an element of F. For instance, the vector $(1, 1) \in \mathbb{Q}^2$ has length $\sqrt{2}$, which is not contained in $\mathbb{Q}$. As we have just seen, the length of a vector in $\mathbb{R}^2$ or $\mathbb{R}^3$ equals the length of any arrow representing it.

Example 1.16. The vector $(1, -2, 2, 3)$ in $\mathbb{R}^4$ has length $\sqrt{1 + 4 + 4 + 9} = 3\sqrt{2}$.

Lemma 1.17. *For all $\lambda \in F$ and $x \in F^n$ we have $\|\lambda x\| = |\lambda| \cdot \|x\|$.*

Proof. This follows immediately from the identity $\langle \lambda x, \lambda x\rangle = \lambda^2 \cdot \langle x, x\rangle$ and the fact that $\sqrt{\lambda^2} = |\lambda|$. □

In $\mathbb{R}^2$ and $\mathbb{R}^3$, the distance between two points x, y equals $\|x - y\|$. We will use the same phrasing in F^n.

Definition 1.18. The *distance* between two points $x, y \in F^n$ is defined as $\|x-y\|$. It is sometimes written as $d(x, y)$.

Proposition 1.19. *Suppose $n = 2$ or $n = 3$. Let v, w be two nonzero elements in $\mathbb{R}^n$ and let $\alpha \in [0, \pi]$ be the angle between the arrow from 0 to v and the arrow from 0 to w. Then we have*

$$\cos \alpha = \frac{\langle v, w\rangle}{\|v\| \cdot \|w\|}. \tag{1.2}$$

The arrows are orthogonal to each other if and only if $\langle v, w\rangle = 0$.

Proof. Because we have $n = 2$ or $n = 3$, the new definition of length coincides with the usual notion of length and we can use ordinary geometry. The arrows from 0 to v, from 0 to w, and from v to w form a triangle in which α is the angle at 0. The arrows represent the vectors v, w, and $w - v$, respectively. See Figure 1.7. By the cosine rule, we find that the length $\|w - v\|$ of the side opposite the angle α satisfies

$$\|w - v\|^2 = \|v\|^2 + \|w\|^2 - 2 \cdot \|v\| \cdot \|w\| \cdot \cos \alpha.$$

We also have

$$\|w-v\|^2 = \langle w-v, w-v\rangle = \langle w,w\rangle - 2\langle v,w\rangle + \langle v,v\rangle = \|v\|^2 + \|w\|^2 - 2\langle v,w\rangle.$$

Equating the two right-hand sides yields the desired equation. The arrows are orthogonal if and only if we have $\cos\alpha = 0$, so if and only if $\langle v,w\rangle = 0$. $\square$

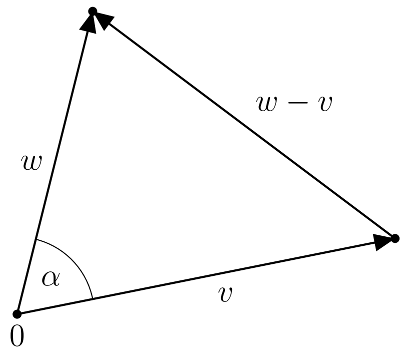


FIGURE 1.7. The cosine rule

Example 1.20. Let l and m be the lines in the (x,y)-plane $\mathbb{R}^2$, given by $y = ax+b$ and $y = cx+d$, respectively, for some $a,b,c,d \in \mathbb{R}$. Then their directions are the same as those of the line l' through 0 and $(1,a)$ and the line m' through 0 and $(1,c)$, respectively. By Proposition 1.19, the lines l' and m', and thus l and m, are orthogonal to each other if and only if $0 = \langle (1,a),(1,c)\rangle = 1+ac$, so if and only if $ac=-1$. See Figure 1.8.

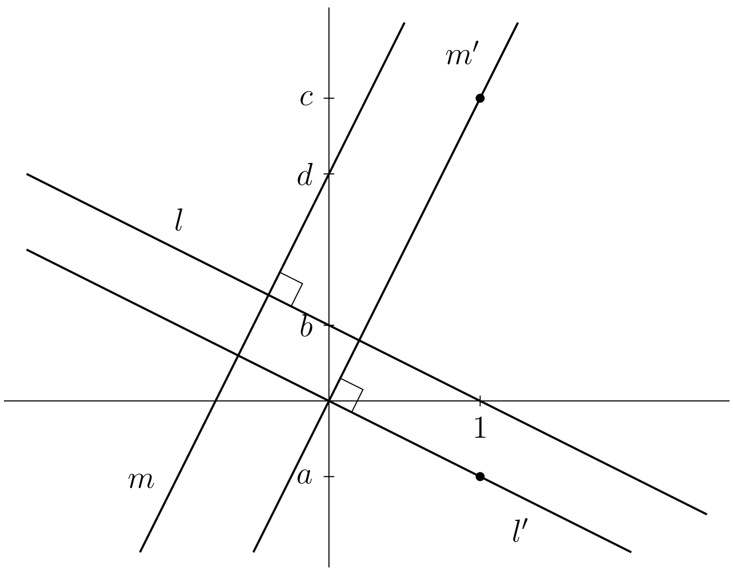


FIGURE 1.8. Orthogonal lines in $\mathbb{R}^2$

Inspired by Proposition 1.19, we define orthogonality for vectors in $\mathbb{R}^n$.

Definition 1.21. We say that two vectors $v,w \in F^n$ are *orthogonal*, or *perpendicular* to each other, when $\langle v,w\rangle = 0$; we then write $v \perp w$.

Warning 1.22. Let $v, w \in F^n$ be vectors, which by definition are just n-tuples of elements in F. If we want to think of them geometrically, then we can think of them as points or we can represent them by arrows. If we want to interpret the notion orthogonality geometrically, then we should represent v and w by *arrows*: Proposition 1.19 states for $n \in \{2, 3\}$ that the vectors v and w are orthogonal if and only if any two *arrows* with a common tail that represent them, are orthogonal to each other.

Note that the zero vector is orthogonal to every vector. With Definitions 1.15 and 1.21 we immediately have the following analogon of Pythagoras' Theorem.

Proposition 1.23. *Two vectors $v, w \in F^n$ are orthogonal if and only if they satisfy $\|v-w\|^2 = \|v\|^2+\|w\|^2$, and if and only if they satisfy $\|v+w\|^2 = \|v\|^2+\|w\|^2$.*

Proof. We have

$$\|v \pm w\|^2 = \langle v \pm w, v \pm w\rangle = \langle v, v\rangle \pm 2\langle v, w\rangle + \langle w, w\rangle = \|v\|^2 + \|w\|^2 \pm 2\langle v, w\rangle.$$

The right-most side equals $\|v\|^2 + \|w\|^2$ if and only if $\langle v, w\rangle = 0$, so if and only if v and w are orthogonal. □

Definition 1.24. For any subset $S \subset F^n$, we let $S^\perp$ denote the set of those elements of F^n that are orthogonal to all elements of S, that is,

$$S^\perp = \{\, x \in F^n \ :\ \langle s, x\rangle = 0 \text{ for all } s \in S\,\}.$$

For every element $a \in F^n$ we define $a^\perp$ as $\{a\}^\perp$. We leave it as an exercise to show that if a is nonzero, then we have $a^\perp = L(a)^\perp$.

Lemma 1.25. *Let $S \subset F^n$ be any subset. Then the following statements hold.*

(1) *For every $x, y \in S^\perp$, we have $x + y \in S^\perp$.*
(2) *For every $x \in S^\perp$ and every $\lambda \in F$, we have $\lambda x \in S^\perp$.*

Proof. Suppose $x, y \in S^\perp$ and $\lambda \in F$. Take any element $s \in S$. By definition of $S^\perp$ we have $\langle s, x\rangle = \langle s, y\rangle = 0$, so we find $\langle s, x + y\rangle = \langle s, x\rangle + \langle s, y\rangle = 0 + 0 = 0$ and $\langle s, \lambda x\rangle = \lambda\langle s, x\rangle = 0$. Since this holds for all $s \in S$, we conclude $x + y \in S^\perp$ and $\lambda x \in S^\perp$. □

By definition, every nonzero vector $a \in F^n$ is orthogonal to every element in the hyperplane $a^\perp$. As mentioned in Warning 1.22, in $\mathbb{R}^2$ and $\mathbb{R}^3$ we think of this as the arrow from 0 to (the point identified with) a being orthogonal to every arrow from 0 to an element of $a^\perp$. Since $a^\perp$ contains 0, these last arrows have both their tail and their head contained in the hyperplane $a^\perp$. Therefore, when we consider a hyperplane H that does not contain 0, the natural analog is to be orthogonal to every arrow that has both its tail and its head contained in H. As the arrow from $p \in H$ to $q \in H$ represents the vector $q - p \in F^n$, this motivates the following definition.

Definition 1.26. Let $S \subset F^n$ be a subset. We say that a vector $z \in F^n$ is *normal to S* when for all $p, q \in S$ we have $\langle q - p, z\rangle = 0$. In this case, we also say that z is a *normal* of S.

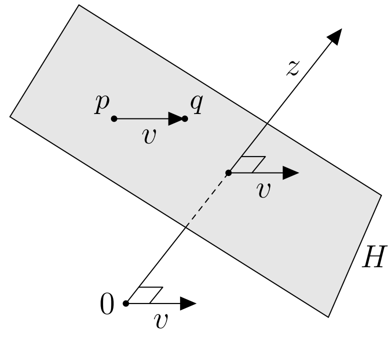


FIGURE 1.9. Normal z to a hyperplane H

See Figure 1.9, in which $S = H \subset \mathbb{R}^3$ is a (hyper-)plane that contains two points p and q, and the vector $v = q - p$ is represented by three arrows: one from p to q, one with its tail at the intersection point of H with $L(a)$, and one with its tail at 0. The first two arrows are contained in H.

Note that the zero vector $0 \in F^n$ is normal to every subset of F^n. We leave it as an exercise to show that every element of $S^\perp$ is a normal to S, and, if S contains 0, then a vector $z \in F^n$ is normal to S if and only if we have $z \in S^\perp$ (see Exercise 1.4.6).

Example 1.27. Let $L \subset \mathbb{R}^2$ be the line $L = \{\lambda \cdot (1,1) \ : \ \lambda \in \mathbb{R}\}$, consisting of all points (x,y) satisfying $y = x$. The set $L^\perp$ consists of all points $v = (a,b)$ that satisfy $0 = \langle v, (\lambda,\lambda)\rangle = \lambda(a+b)$ for all $\lambda \in \mathbb{R}$; this is equivalent to $b = -a$ and to $v = a \cdot (1,-1)$, so $L^\perp = \{a \cdot (1,-1) \ : \ a \in \mathbb{R}\}$ is a line. As mentioned above, since L contains 0, the set $L^\perp$ consists of all elements that are normal to L.

Example 1.28. Let $M \subset \mathbb{R}^2$ be the line $M = \{(0,1) + \lambda \cdot (1,1) \ : \ \lambda \in \mathbb{R}\}$, consisting of all points (x,y) satisfying $y = x+1$. The set $M^\perp$ consists of all points $v = (a,b)$ that satisfy $0 = \langle v, (\lambda, \lambda+1)\rangle = \lambda(a+b) + b$ for all $\lambda \in \mathbb{R}$; this is equivalent to $a = b = 0$ and $v = 0$, so $M^\perp = \{0\}$. In this case, the line M does not contain 0. We leave it to the reader to verify that every multiple of the vector $(1,-1)$ is a normal to M.

Proposition 1.29. *Let $a \in F^n$ be a nonzero vector and $b \in F$ a constant. Then a is normal to the hyperplane $H = \{\, x \in F^n \ : \ \langle a, x\rangle = b \,\}$.*

Proof. For every two elements $p, q \in H$ we have $\langle p, a\rangle = \langle q, a\rangle = b$, so we find $\langle q-p, a\rangle = \langle q, a\rangle - \langle p, a\rangle = b - b = 0$. This implies that a is normal to H. □

Corollary 1.36 of the next section implies the converse of Proposition 1.29: for every nonzero normal a' of a hyperplane H there is a constant $b' \in F$ such that

$$H = \{\, x \in F^n \ : \ \langle a', x\rangle = b' \,\}.$$

Exercises

1.4.1. Let a and b be the lengths of the sides of a parallelogram and c and d the lengths of its diagonals. Prove that $c^2 + d^2 = 2(a^2 + b^2)$.

1.4.2.

(1) Show that two vectors $v, w \in \mathbb{R}^n$ have the same length if and only if $v - w$ and $v + w$ are orthogonal.

(2) Prove that the diagonals of a parallelogram are orthogonal to each other if and only if all sides have the same length.

1.4.3. Let $a \in F^n$ be nonzero. Show that we have $a^\perp = L(a)^\perp$.

1.4.4. Determine the angle between the lines $L(a)$ and $L(b)$ with $a = (2, 1, 3)$ and $b = (-1, 3, 2)$.

1.4.5. True or False? If true, explain why. If false, give a counterexample.

(1) If $a \in \mathbb{R}^2$ is a nonzero vector, then the lines $\{x \in \mathbb{R}^2 \, : \, \langle a, x\rangle = 0\}$ and $\{x \in \mathbb{R}^2 \, : \, \langle a, x\rangle = 1\}$ in $\mathbb{R}^2$ are parallel.

(2) If $a, b \in \mathbb{R}^2$ are nonzero vectors and $a \neq b$, then the lines $\{x \in \mathbb{R}^2 \, : \, \langle a, x\rangle = 0\}$ and $\{x \in \mathbb{R}^2 \, : \, \langle b, x\rangle = 1\}$ in $\mathbb{R}^2$ are not parallel.

(3) For each vector $v \in \mathbb{R}^2$ we have $\langle 0, v\rangle = 0$. (What do the zeros in this statement refer to?)

1.4.6. Let $S \subset F^n$ be a subset.

(1) Show that every element in $S^\perp$ is a normal to S.

(2) Assume that S contains the zero element 0. Show that every normal to S is contained in $S^\perp$.

1.4.7. What would be a good definition for a line and a hyperplane (neither necessarily containing 0) to be orthogonal?

1.4.8. What would be a good definition for two lines (neither necessarily containing 0) to be parallel?

1.4.9. What would be a good definition for two hyperplanes (neither necessarily containing 0) to be parallel?

1.4.10. Let $a, v \in F^n$ be nonzero vectors, $p \in F^n$ any point, and $b \in F$ a scalar. Let $L \subset F^n$ be the line given by

$$L = \{p + tv \; : \; t \in F\}$$

and let $H \subset F^n$ be the hyperplane given by

$$H = \{x \in F^n \; : \; \langle a, x\rangle = b\}.$$

(1) Show that $L \cap H$ consists of exactly one point if $v \notin a^\perp$.

(2) Show that $L \cap H = \emptyset$ if $v \in a^\perp$ and $p \notin H$.

(3) Show that $L \subset H$ if $v \in a^\perp$ and $p \in H$.

1.5. Orthogonal projections and normality

Note that our field F is still assumed to be a subset of $\mathbb{R}$.

1.5.1. Projecting onto lines and hyperplanes containing zero.

Proposition 1.30. *Let $a \in F^n$ be a vector. Then every element $v \in F^n$ can be written uniquely as a sum $v = v_1 + v_2$ of an element $v_1 \in L(a)$ and an element $v_2 \in a^\perp$. Moreover, if a is nonzero, then we have $v_1 = \lambda a$ with $\lambda = \langle a, v\rangle \cdot \|a\|^{-2}$.*

Proof. For $a = 0$ the statement is trivial, as we have $0^\perp = F^n$, so we may assume a is nonzero. Then we have $\langle a, a\rangle \neq 0$. See Figure 1.10. Let $v \in F^n$ be a vector. Let $v_1 \in L(a)$ and $v_2 \in F^n$ be such that $v = v_1 + v_2$. Then there is a $\lambda \in F$ with $v_1 = \lambda a$ and we have $\langle a, v_2\rangle = \langle a, v\rangle - \lambda\langle a, a\rangle$; this implies that we have $\langle a, v_2\rangle = 0$ if and only if $\langle a, v\rangle = \lambda\langle a, a\rangle = \lambda\|a\|^2$, that is, if and only if $\lambda = \frac{\langle a,v\rangle}{\|a\|^2}$. Hence, this λ corresponds to unique elements $v_1 \in L(a)$ and $v_2 \in a^\perp$ with $v = v_1 + v_2$. $\square$

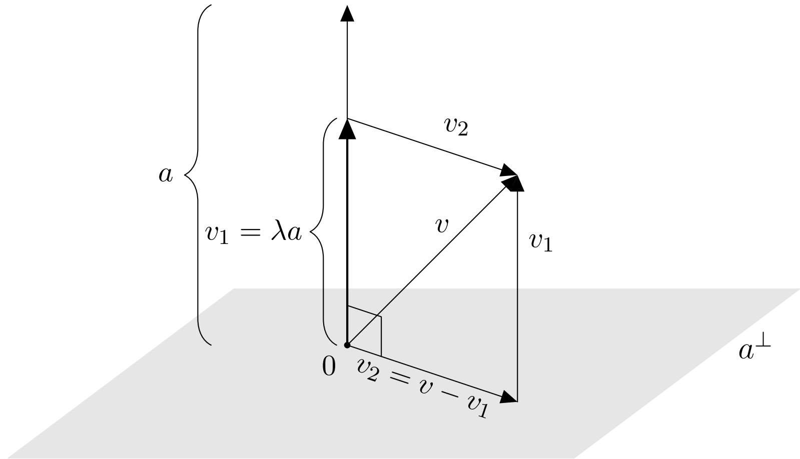


FIGURE 1.10. Decomposing the vector v as the sum of a multiple v_1 of the vector a and a vector v_2 orthogonal to a

Definition 1.31. Using the same notation as in Proposition 1.30 and assuming a is nonzero, we call v_1 the *orthogonal projection* of v onto a or onto $L = L(a)$, and we call v_2 the orthogonal projection of v onto the hyperplane $H = a^\perp$. We let

$$\pi_L \colon F^n \to F^n \qquad \text{and} \qquad \pi_H \colon F^n \to F^n$$

be the maps[4] that send v to these orthogonal projections of v on L and H, respectively, so $\pi_L(v) = v_1$ and $\pi_H(v) = v_2$. These maps are also called the orthogonal projections onto L and H, respectively.

We will also write π_a for π_L, and of course $\pi_{a^\perp}$ for π_H. Note that by Proposition 1.30 these maps are well defined and we have

$$\pi_a(v) = \frac{\langle a, v\rangle}{\langle a, a\rangle} \cdot a, \qquad \pi_{a^\perp}(v) = v - \frac{\langle a, v\rangle}{\langle a, a\rangle} \cdot a. \tag{1.3}$$

Example 1.32. Take $a = (2,1) \in \mathbb{R}^2$. Then the hyperplane $a^\perp$ is the line consisting of all points $(x_1, x_2) \in \mathbb{R}^2$ satisfying $2x_1 + x_2 = 0$. To write the vector $v = (3,4)$ as a sum $v = v_1 + v_2$ with v_1 a multiple of a and $v_2 \in a^\perp$, we compute

$$\lambda = \frac{\langle a, v\rangle}{\langle a, a\rangle} = \frac{10}{5} = 2,$$

so we get $\pi_a(v) = v_1 = 2a = (4,2)$ and thus $\pi_{a^\perp}(v) = v_2 = v - v_1 = (-1,2)$. Indeed, we have $v_2 \in a^\perp$.

Example 1.33. Take $a = (1,1,1) \in \mathbb{R}^3$. Then the hyperplane $H = a^\perp$ is the set

$$H = \{\, x \in \mathbb{R}^3 \; : \; \langle a, x\rangle = 0 \,\} = \{\, (x_1, x_2, x_3) \in \mathbb{R}^3 \; : \; x_1 + x_2 + x_3 = 0 \,\}.$$

To write the vector $v = (2,1,3)$ as a sum $v = v_1 + v_2$ with v_1 a multiple of a and $v_2 \in H$, we compute

$$\lambda = \frac{\langle a, v\rangle}{\langle a, a\rangle} = \frac{6}{3} = 2,$$

[4]For a review on maps, see Appendix A.

so we get $\pi_a(v) = v_1 = 2a = (2, 2, 2)$ and thus

$$\pi_H(v) = v_2 = v - v_1 = (2, 1, 3) - (2, 2, 2) = (0, -1, 1).$$

Indeed, we have $v_2 \in H$.

In fact, we can do the same for every element in $\mathbb{R}^3$. We find that we can write $x = (x_1, x_2, x_3)$ as $x = x' + x''$ with

$$x' = \frac{x_1 + x_2 + x_3}{3} \cdot a = \pi_a(x)$$

and

$$x'' = \left(\frac{2x_1 - x_2 - x_3}{3}, \frac{-x_1 + 2x_2 - x_3}{3}, \frac{-x_1 - x_2 + 2x_3}{3} \right) = \pi_H(x) \in H.$$

Verify this and derive it yourself!

Example 1.34. Suppose an object T is moving along an inclined straight path in $\mathbb{R}^3$. Gravity exerts a force f on T, which corresponds to a vector. The force f can be written uniquely as the sum of two components: a force along the path and a force perpendicular to the path. The acceleration due to gravity depends on the component along the path. If we take the zero of Euclidean space to be at the object T, and the path is decribed by a line L, then the component along the path is exactly the orthogonal projection $\pi_L(f)$ of f onto L. See Figure 1.11.

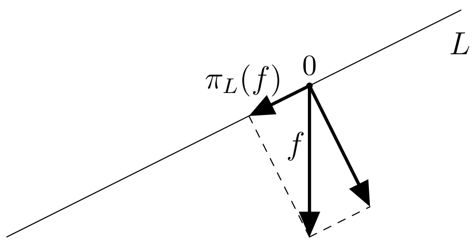


FIGURE 1.11. Two components of a force: one along the path and one perpendicular to it

We have already seen that for every vector $a \in F^n$ we have $L(a)^\perp = a^\perp$, so the operation $S \rightsquigarrow S^\perp$ sends the line $L(a)$ to the hyperplane $a^\perp$. The following proposition shows that the opposite holds as well.

Proposition 1.35. *Let $a \in F^n$ be a vector. Then we have $(a^\perp)^\perp = L(a)$.*

Proof. For every $\lambda \in F$ and every $t \in a^\perp$, we have $\langle \lambda a, t \rangle = \lambda \langle a, t \rangle = 0$, so we find $L(a) \subset (a^\perp)^\perp$. For the opposite inclusion, let $v \in (a^\perp)^\perp$ be arbitrary and let $v_1 \in L(a)$ and $v_2 \in a^\perp$ be such that $v = v_1 + v_2$ (as in Proposition 1.30). Then by the inclusion above we have $v_1 \in (a^\perp)^\perp$, so by Lemma 1.25 we find $v_2 = v - v_1 = (a^\perp)^\perp$. Hence, the element v_2 is orthogonal to every element in $a^\perp$, and in particular to itself, which implies $v_2 = 0$. We conclude $v - v_1 = 0$, so $v = v_1 \in L(a)$. This implies $(a^\perp)^\perp \subset L(a)$, which proves the proposition. $\square$

For generalisations of Proposition 1.35, see Proposition 8.20 and Exercise 8.2.4[5] (cf. Proposition 3.33, Remark 3.34). The following corollary shows that every hyperplane is determined by a nonzero normal to it and a point contained in it. Despite the name of this subsection, this corollary, and one of the examples following it, is not restricted to hyperplanes that contain the element 0.

Corollary 1.36. *Let $a, z \in F^n$ be nonzero vectors. Let $b \in F$ be a scalar and set*

$$H = \{\, x \in F^n \;:\; \langle a, x\rangle = b \,\}.$$

Let $p \in H$ be a point. Then the following statements hold.

(1) *The vector z is normal to H if and only if z is a multiple of a.*
(2) *If z is normal to H, then we have*

$$H = \{\, x \in F^n \;:\; \langle z, x\rangle = \langle z, p\rangle \,\} = \{\, x \in F^n \;:\; x - p \in z^\perp \,\}.$$

Proof. We first prove the 'if'-part of (1). Suppose $z = \lambda a$ for some $\lambda \in F$. Then λ is nonzero, and the equation $\langle a, x\rangle = b$ is equivalent with $\langle z, x\rangle = \lambda b$. Hence, by Proposition 1.29, applied to $z = \lambda a$, we find that z is normal to H. For the 'only if'-part and part (2), suppose z is normal to H. We translate H by subtracting p from each point in H, and obtain[6]

$$H' = \{\, y \in F^n \;:\; y + p \in H \,\}.$$

Since p is contained in H, we have $\langle a, p\rangle = b$, so we find

$$H' = \{\, y \in F^n \;:\; \langle a, y + p\rangle = \langle a, p\rangle \,\} = \{\, y \in F^n \;:\; \langle a, y\rangle = 0 \,\} = a^\perp.$$

On the other hand, for every $y \in H'$, we have $y+p \in H$, so by definition of normality, z is orthogonal to $(y+p)-p = y$. This implies $z \in H'^\perp = (a^\perp)^\perp = L(a)$ by Proposition 1.35, so z is indeed a multiple of a, which finishes the proof of (1).

This also implies that $H' = a^\perp$ equals $z^\perp$, so we get

$$\begin{aligned} H &= \{\, x \in F^n \;:\; x - p \in H' \,\} = \{\, x \in F^n \;:\; x - p \in z^\perp \,\} \\ &= \{\, x \in F^n \;:\; \langle z, x - p\rangle = 0 \,\} = \{\, x \in F^n \;:\; \langle z, x\rangle = \langle z, p\rangle \,\}. \end{aligned}$$

□

Example 1.37. If $H \subset F^n$ is a hyperplane that contains 0, and $a \in F^n$ is a nonzero normal of H, then we have $H = a^\perp$ by Corollary 1.36.

Example 1.38. Suppose $V \subset \mathbb{R}^3$ is a plane that contains the points

$$p_1 = (1, 0, 1), \qquad p_2 = (2, -1, 0), \qquad \text{and} \qquad p_3 = (1, 1, 1).$$

A priori, we do not know if such a plane exists. If a vector $a = (a_1, a_2, a_3) \in \mathbb{R}^3$ is a normal of V, then we have

$$0 = \langle p_2 - p_1, a\rangle = a_1 - a_2 - a_3 \qquad \text{and} \qquad 0 = \langle p_3 - p_1, a\rangle = a_2,$$

[5] The proof of Proposition 1.35 relies on Proposition 1.30, which is itself proved by explicitly computing the scalar λ. Therefore, one might qualify both these proofs as computational. In this book, we try to avoid computational proofs when more enlightening arguments are available. Proposition 8.20, which uses the notion of dimension, provides an independent non-computational proof of a generalisation of Proposition 1.35 (see Exercise 8.2.4).

[6] Make sure you understand why this is what we obtain, including the plus-sign in $y + p$.

which is equivalent with $a_1 = a_3$ and $a_2 = 0$, and thus with $a = a_3 \cdot (1, 0, 1)$. Taking $a_3 = 1$, we find that the vector $a = (1, 0, 1)$ is a normal of V and as we have $\langle a, p_1 \rangle = 2$, the plane V equals

$$\{ \, x \in \mathbb{R}^3 \ : \ \langle a, x \rangle = 2 \, \} \tag{1.4}$$

by Corollary 1.36, at least if V exists. It follows from $\langle p_2 - p_1, a \rangle = \langle p_3 - p_1, a \rangle = 0$ that $\langle p_2, a \rangle = \langle p_1, a \rangle = 2$ and $\langle p_3, a \rangle = \langle p_1, a \rangle = 2$, so the plane in (1.4) contains p_1, p_2, and p_3. This shows that V does indeed exist and is uniquely determined by the fact that it contains p_1, p_2, and p_3.

Remark 1.39. In a later chapter, we will see that any three points in $\mathbb{R}^3$ that are not on one line determine a unique plane containing these points.

Remark 1.40. If $W \subset F^n$ is a line containing 0, and $a \in W$ is a nonzero element, then $W = L(a)$ by Proposition 1.12. If $W \subset F^n$ is a hyperplane containing 0, and $a \in W$ is a nonzero normal of W, then $W = a^\perp$ by Corollary 1.36.

Corollary 1.41. *Let $W \subset F^n$ be a line or a hyperplane and assume $0 \in W$. Then we have $(W^\perp)^\perp = W$.*

Proof. If W is a line and $a \in W$ is a nonzero element, then we have $W = L(a)$ by Proposition 1.12; then we get $W^\perp = a^\perp$, and the equality $(W^\perp)^\perp = W$ follows from Proposition 1.35. If W is a hyperplane and $a \in F^n$ is a nonzero normal of W, then $W = a^\perp$ by Corollary 1.36; then we get $W^\perp = (a^\perp)^\perp = L(a)$ by Proposition 1.35, so we also find $(W^\perp)^\perp = L(a)^\perp = a^\perp = W$. □

In the definition of orthogonal projections, the roles of the line $L(a)$ and the hyperplane $a^\perp$ seem different. The following proposition characterises the orthogonal projection completely analogous for lines and hyperplanes containing 0 (cf. Figure 1.12). Proposition 1.43 generalises this to general lines and hyperplanes, which allows us to define the orthogonal projection of a point to any line or hyperplane.

Proposition 1.42. *Let $W \subset F^n$ be a line or a hyperplane, suppose $0 \in W$, and let $v \in F^n$ be an element. Then there is a unique element $z \in W$ such that $v - z \in W^\perp$. This element z equals $\pi_W(v)$.*

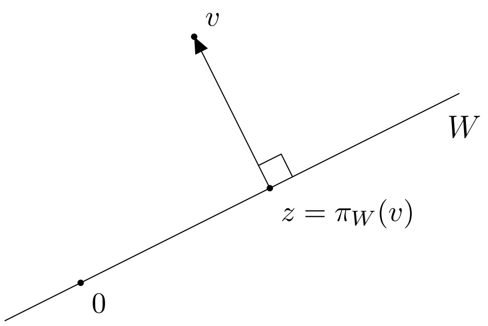


FIGURE 1.12. Orthogonal projection of v onto a line or hyperplane W with $0 \in W$

Proof. We have two cases. If W is a line, then we take any nonzero $a \in W$, so that we have $W = L(a)$ and $W^\perp = L(a)^\perp = a^\perp$. Then, by Proposition 1.30, there is a unique element $z \in W$ such that $v - z \in W^\perp$, namely $z = \pi_a(v)$.

If W is a hyperplane, then we take any nonzero normal a to W, so that we have $W = a^\perp$, and then $W^\perp = L(a)$ by Proposition 1.35. Then, again by Proposition 1.30, there is a unique element $z \in W$ such that $v - z \in W^\perp$, namely $z = \pi_{a^\perp}(v)$. □

Exercises

1.5.1. Show that there is a unique plane $V \subset \mathbb{R}^3$ containing the points

$$p_1 = (1,0,2), \qquad p_2 = (-1,2,2), \qquad \text{and} \qquad p_3 = (1,1,1).$$

Determine a vector $a \in \mathbb{R}^3$ and a number $b \in \mathbb{R}$ such that

$$V = \{x \in \mathbb{R}^3 \, : \, \langle a, x\rangle = b\}.$$

1.5.2. Take $a = (2,1) \in \mathbb{R}^2$ and $v = (4,5) \in \mathbb{R}^2$. Find $v_1 \in L(a)$ and $v_2 \in a^\perp$ such that $v = v_1 + v_2$.

1.5.3. Take $a = (2,1) \in \mathbb{R}^2$ and $v = (x_1, x_2) \in \mathbb{R}^2$. Find $v_1 \in L(a)$ and $v_2 \in a^\perp$ such that $v = v_1 + v_2$.

1.5.4. Take $a = (-1,2,1) \in \mathbb{R}^3$ and set $V = a^\perp \subset \mathbb{R}^3$. Find the orthogonal projections of the element $x = (x_1, x_2, x_3) \in \mathbb{R}^3$ onto $L(a)$ and V.

1.5.5. Show that for every subset $S \subset F^n$ we have $S \cap S^\perp \subset \{0\}$.

1.5.6. Let $W \subset F^n$ be a line or a hyperplane, and assume $0 \in W$. Use (1.3) to show that
(1) for every $x, y \in F^n$ we have $\pi_W(x+y) = \pi_W(x) + \pi_W(y)$, and
(2) for every $x \in F^n$ and every $\lambda \in F$ we have $\pi_W(\lambda x) = \lambda \pi_W(x)$.

1.5.7. Let $W \subset F^n$ be a line or a hyperplane, and assume $0 \in W$.
(1) Show that there exists a nonzero $a \in F^n$ such that $W = L(a)$ or $W = a^\perp$.
(2) Show that for every $v, w \in W$ we have $v + w \in W$.

1.5.8. This exercise proves the same as Exercise 1.5.6, but without formulas. Let $W \subset F^n$ be a line or a hyperplane, and assume $0 \in W$. Use Exercise 1.5.7, Proposition 1.42, and Lemma 1.25 to show that
(1) for every $x, y \in F^n$ we have $\pi_W(x+y) = \pi_W(x) + \pi_W(y)$, and
(2) for every $x \in F^n$ and every $\lambda \in F$ we have $\pi_W(\lambda x) = \lambda \pi_W(x)$.

1.5.9. Let $W \subset F^n$ be a line or a hyperplane, and assume $0 \in W$. Let $p \in W$ and $v \in F^n$ be points. Prove that we have $\pi_W(v - p) = \pi_W(v) - p$. See Proposition 1.43 for a generalisation.

1.5.10. Let $a \in F^n$ be nonzero and set $L = L(a)$. Let $q \in F^n$ be a point and let $H \subset F^n$ be the hyperplane with normal $a \in F^n$ and containing the point q.
(1) Show that the line L intersects the hyperplane H in a unique point, say p (see Exercise 1.4.10).
(2) Show that for every point $x \in H$ we have $\pi_L(x) = p$.

1.5.11. (1) Let $p, q, r, s \in \mathbb{R}^2$ be four distinct points. Show that the line through p and q is perpendicular to the line through r and s if and only if

$$\langle p, r\rangle + \langle q, s\rangle = \langle p, s\rangle + \langle q, r\rangle.$$

(2) Let $p, q, r \in \mathbb{R}^2$ be three points that are not all on a line. Then the *altitudes* of the triangle with vertices p, q, and r are the lines through one of the three points, orthogonal to the line through the other two points.
Prove that the three altitudes in a triangle go through one point. This point is called the *orthocenter* of the triangle. [Hint: let p, q, r be the vertices of the triangle and let s be the intersection of two of the three altitudes. Be careful with the case that s coincides with one of the vertices.]

1.5.2. Projecting onto arbitrary lines and hyperplanes.

We now generalise Proposition 1.42 to arbitrary lines and hyperplanes, not necessarily containing 0.

Proposition 1.43. *Let $W \subset F^n$ be a line or a hyperplane, and let $v \in F^n$ be an element. Then there is a unique element $z \in W$ such that $v - z$ is normal to W. Moreover, if $p \in W$ is any point, then $W' = \{x - p : x \in W\}$ contains 0 and we have*

$$z - p = \pi_{W'}(v - p).$$

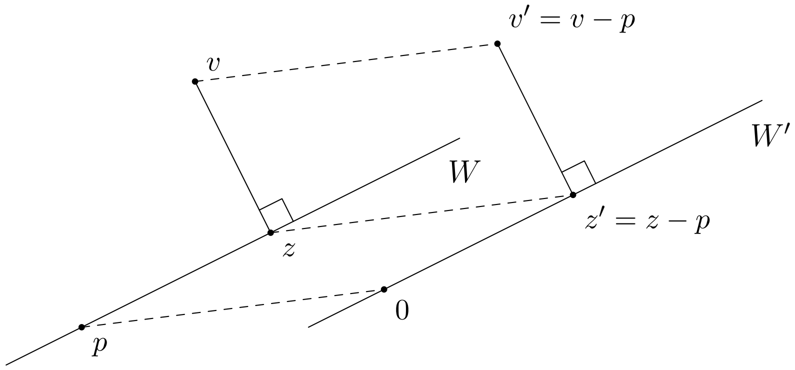


FIGURE 1.13. Orthogonal projection of v onto a general line or hyperplane W

Proof. We start with the special case that W contains 0 and we have $p = 0$. Since W contains 0, a vector $x \in F^n$ is contained in $W^\perp$ if and only if x is normal to W (see Exercise 1.4.6), so this special case is exactly Proposition 1.42. Now let W be an arbitrary line or hypersurface and let $p \in W$ be an element. See Figure 1.13. For any vector $z \in F^n$, each of the two conditions

(i) $z \in W$, and
(ii) $v - z$ is normal to W

is satisfied if and only if it is satisfied after replacing v, z, and W by $v' = v - p$, $z' = z - p$, and W', respectively. The hyperplane W' contains 0, so from the special case above, we find that there is indeed a unique vector $z \in F^n$ satisfying (i) and (ii), and the elements $v' = v - p$ and $z' = z - p$ satisfy $z' = \pi_{W'}(v')$, which implies the final statement of the proposition. □

Proposition 1.43 can be used to define the orthogonal projection onto any line or hyperplane $W \subset F^n$.

Definition 1.44. Let $W \subset F^n$ be a line or a hyperplane. The *orthogonal projection* $\pi_W : F^n \to F^n$ onto W is the map that sends $v \in F^n$ to the unique element z of Proposition 1.43, that is,

$$\pi_W(v) = p + \pi_{W'}(v - p).$$

When W contains 0, Proposition 1.42 shows that this new definition of the orthogonal projection agrees with Definition 1.31, because in this case, the vector $v - z$ is normal to W if and only if $v - z \in W^\perp$ (see Exercise 1.4.6).

It follows from Definition 1.44 that if we want to project v onto a line or hyperplane W that does not contain 0, then we may first translate everything so that the resulting line or hyperplane does contain 0, then project orthogonally, and finally translate back.

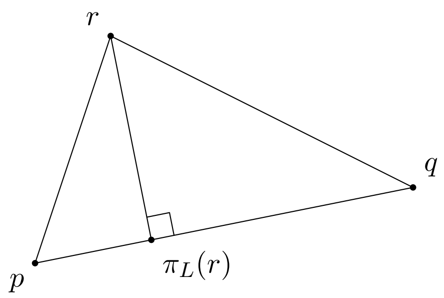


FIGURE 1.14. Altitude of a triangle

Exercises

1.5.12. Let $H \subset \mathbb{R}^3$ be a hyperplane with normal $a = (1, 2, 1)$ that contains the point $p = (1, 1, 1)$. Find the orthogonal projection of the point $q = (0, 0, 0)$ onto H.

1.5.13. Let $p, q, r \in \mathbb{R}^2$ be three points that are not all on a line. Show that the altitude through r intersects the line L through p and q in the point

$$\pi_L(r) = p + \frac{\langle r - p, q - p\rangle}{\|q - p\|^2} \cdot (q - p).$$

See Figure 1.14.

1.6. Distances

Lemma 1.45. *Let $a, v \in F^n$ be elements with $a \neq 0$. Set $L = L(a)$ and $H = a^\perp$. Let $v_1 = \pi_L(v) \in L$ and $v_2 = \pi_H(v) \in H$ be the orthogonal projections of v on L and H, respectively. Then the lengths of v_1 and v_2 satisfy*

$$\|v_1\| = \frac{|\langle a, v\rangle|}{\|a\|} \qquad \text{and} \qquad \|v_2\|^2 = \|v\|^2 - \|v_1\|^2 = \|v\|^2 - \frac{\langle a, v\rangle^2}{\|a\|^2}.$$

Moreover, for any $x \in L$ we have $d(v, x) \geq d(v, v_1) = \|v_2\|$ and for any $y \in H$ we have $d(v, y) \geq d(v, v_2) = \|v_1\|$.

Proof. By (1.3) we have $v_1 = \lambda a$ with $\lambda = \frac{\langle a,v\rangle}{\|a\|^2}$. Lemma 1.17 then yields

$$\|v_1\| = |\lambda| \cdot \|a\| = \frac{|\langle a, v\rangle|}{\|a\|}.$$

Since v_1 and v_2 are orthogonal, and $v_1 + v_2 = v$, we find from Proposition 1.23 (Pythagoras) that we have

$$\|v_2\|^2 = \|v\|^2 - \|v_1\|^2 = \|v\|^2 - \frac{\langle a, v\rangle^2}{\|a\|^2},$$

Suppose $x \in L$. we can write $v - x$ as the sum $(v - v_1) + (v_1 - x)$ of two orthogonal vectors (see Figure 1.15), so that, again by Proposition 1.23, we have

$$d(v, x)^2 = \|v - x\|^2 = \|v - v_1\|^2 + \|v_1 - x\|^2 \geq \|v - v_1\|^2 = \|v_2\|^2.$$

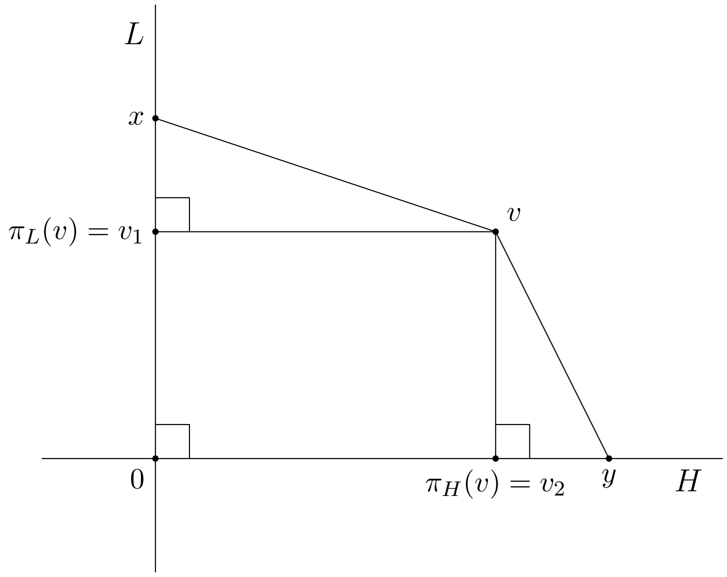


FIGURE 1.15. Distance from v to points on L and H

Because distances and lengths are non-negative, this proves the first part of the last statement. The second part follows similarly by writing $v - y$ as $(v - v_2) + (v_2 - y)$. $\square$

Lemma 1.45 shows that if $a \in F^n$ is a nonzero vector and W is either the line $L(a)$ or the hyperplane $a^\perp$, then the distance $d(v, x) = \|v - x\|$ from v to any point $x \in W$ is at least the distance from v to the orthogonal projection of v on W. This shows that the minimum in the following definition exists, at least if W contains 0. Of course the same holds when W does not contain 0, as we can translate W and v, and translation does not affect the distances between points. So the following definition makes sense.

Definition 1.46. Suppose $W \subset F^n$ is either a line or a hyperplane. For any $v \in F^n$, we define the distance $d(v, W)$ from v to W to be the minimal distance from v to any point in W, that is,

$$d(v, W) = \min_{w \in W} d(v, w) = \min_{w \in W} \|v - w\|.$$

Proposition 1.47. *Let $a, v \in F^n$ be elements with $a \neq 0$. Then we have*

$$d(v, a^\perp) = d(v, \pi_{a^\perp}(v)) = \frac{|\langle a, v\rangle|}{\|a\|} \qquad \textit{and}$$

$$d(v, L(a)) = d(v, \pi_{L(a)}(v)) = \sqrt{\|v\|^2 - \tfrac{\langle a, v\rangle^2}{\|a\|^2}}.$$

Proof. Let v_1 and v_2 be the orthogonal projections of v onto $L(a)$ and $a^\perp$, respectively. Then from Lemma 1.45 we obtain

$$d(v, a^\perp) = d(v, \pi_{a^\perp}(v)) = \|v - v_2\| = \|v_1\| = \frac{|\langle a, v\rangle|}{\|a\|} \qquad \text{and}$$

$$d(v, L(a)) = d(v, \pi_{L(a)}(v)) = \|v - v_1\| = \|v_2\| = \sqrt{\|v\|^2 - \tfrac{\langle a, v\rangle^2}{\|a\|^2}}.$$

$\square$

Note that $L(a)$ and $a^\perp$ contain 0, so Proposition 1.47 states that if a line or hyperplane W contains 0, then the distance from a point v to W is the distance from v to the nearest point on W, which is the orthogonal projection $\pi_W(v)$ of v onto W. Exercise 1.6.11 shows that the same is true for any line or hyperplane (see Proposition 1.43 and the subsequent paragraph for the definition of orthogonal projection onto general lines and hyperplanes).

In order to find the distance to a line or hyperplane that does *not* contain 0, it is usually easiest to first apply an appropriate translation (which does not affect distances between points) to make sure the line or hyperplane *does* contain 0 (cf. Examples 1.50 and 1.51).

Example 1.48. We continue Example 1.33. We find that the distance $d(v, L(a))$ from v to $L(a)$ equals $\|v_2\| = \sqrt{2}$ and we find that the distance from v to H equals $d(v, H) = \|v_1\| = 2\sqrt{3}$. We leave it as an exercise to use the general description of $\pi_a(x)$ and $\pi_H(x)$ in Example 1.33 to find the distances from $x = (x_1, x_2, x_3)$ to $L(a)$ and $H = a^\perp$.

Example 1.49. Consider the point $p = (2, 1, 1)$ and the plane

$$V = \{\, (x_1, x_2, x_3) \in \mathbb{R}^3 \;:\; x_1 - 2x_2 + 3x_3 = 0 \,\}$$

in $\mathbb{R}^3$. We compute the distance from p to V. The normal vector $a = (1, -2, 3)$ of V satisfies $\langle a, a\rangle = 14$. Since we have $V = a^\perp$, by Proposition 1.47, the distance $d(p, V)$ from p to V equals the length of the orthogonal projection of p on a. This projection is λa with $\lambda = \langle a, p\rangle \cdot \|a\|^{-2} = \frac{3}{14}$. Therefore, the distance we want equals $\|\lambda a\| = \frac{3}{14}\sqrt{14}$.

Example 1.50. Consider the vector $a = (1, -2, 3)$, the point $p = (2, 1, 1)$ and the plane

$$W = \{\, x \in \mathbb{R}^3 \;:\; \langle a, x\rangle = 1 \,\}$$

in $\mathbb{R}^3$. We will compute the distance from p to W. Since W does not contain 0, it is not a subspace and our results do not apply directly. Note that the point $q = (2, -1, -1)$ is contained in W. We translate the whole configuration by $-q$ and obtain the point $p' = p - q = (0, 2, 2)$ and the plane[7]

$$\begin{aligned} W' &= \{\, x - q \;:\; x \in W \,\} \\ &= \{\, x \in \mathbb{R}^3 \;:\; x + q \in W \,\} \\ &= \{\, x \in \mathbb{R}^3 \;:\; \langle a, x + q\rangle = 1 \,\} \\ &= \{\, x \in \mathbb{R}^3 \;:\; \langle a, x\rangle = 0 \,\} = a^\perp, \end{aligned}$$

which *does* contain 0 (by construction, of course, because it is the image of $q \in W$ under the translation). By Proposition 1.47, the distance $d(p', W')$ from p' to W' equals the length of the orthogonal projection of p' on a. This projection is λa with $\lambda = \langle a, p'\rangle \cdot \|a\|^{-2} = \frac{1}{7}$. Therefore, the distance we want equals $d(p, W) = d(p', W') = \|\lambda a\| = \frac{1}{7}\sqrt{14}$.

Example 1.51. Let $L \subset \mathbb{R}^3$ be the line through the points $p = (1, -1, 2)$ and $q = (2, -2, 1)$. We will find the distance from the point $v = (1, 1, 1)$ to L. First we translate the whole configuration by $-p$ to obtain the point $v' = v - p = (0, 2, -1)$ and the line L' through the points 0 and $q - p = (1, -1, -1)$.

[7]Note the plus sign in the derived equation $\langle a, x+q\rangle = 1$ for W' and make sure you understand why it is there!

If we set $a = q - p$, then we have $L' = L(a)$ (which is why we translated in the first place) and the distance $d(v, L) = d(v', L')$ is the length of the orthogonal projection of v' onto the hyperplane $a^\perp$. We can compute this directly with Proposition 1.47. It satisfies

$$d(v', L')^2 = \|v'\|^2 - \frac{\langle a, v'\rangle^2}{\|a\|^2} = 5 - \frac{(-1)^2}{3} = \frac{14}{3},$$

so we have $d(v, L) = d(v', L') = \sqrt{\frac{14}{3}} = \frac{1}{3}\sqrt{42}$. Alternatively, in order to determine the orthogonal projection of v' onto $a^\perp$, it is easiest to first compute the orthogonal projection of v' onto $L(a)$, which is λa with $\lambda = \frac{\langle a,v'\rangle}{\|a\|^2} = -\frac{1}{3}$. Then the orthogonal projection of v' onto $a^\perp$ equals $v' - (-\frac{1}{3}a) = (\frac{1}{3}, \frac{5}{3}, -\frac{4}{3})$ and the length of this vector is indeed $\frac{1}{3}\sqrt{42}$.

Exercises

1.6.1. Take $a = (2, 1) \in \mathbb{R}^2$ and $p = (4, 5) \in \mathbb{R}^2$. Find the distances from p to $L(a)$ and $a^\perp$.

1.6.2. Take $a = (2, 1) \in \mathbb{R}^2$ and $p = (x, y) \in \mathbb{R}^2$. Find the distances from p to $L(a)$ and $a^\perp$.

1.6.3. Compute the distance from the point $(1, 1, 1, 1) \in \mathbb{R}^4$ to the line $L(a)$ with $a = (1, 2, 3, 4)$.

1.6.4. Given the vectors $p = (1, 2, 3)$ and $w = (2, 1, 5)$, let L be the line consisting of all points of the form $p + \lambda w$ for some $\lambda \in \mathbb{R}$. Compute the distance $d(v, L)$ for $v = (2, 1, 3)$.

1.6.5. Suppose that $V \subset \mathbb{R}^3$ is a plane that contains the points

$$p_1 = (1, 2, -1), \qquad p_2 = (1, 0, 1), \qquad \text{and} \qquad p_3 = (-2, 3, 1).$$

Determine the distance from the point $q = (2, 2, 1)$ to V.

1.6.6. Let $a_1, a_2, a_3 \in \mathbb{R}$ be such that $a_1^2 + a_2^2 + a_3^2 = 1$, and let $f\colon \mathbb{R}^3 \to \mathbb{R}$ be the function that sends $x = (x_1, x_2, x_3)$ to $a_1x_1 + a_2x_2 + a_3x_3$.

(1) Show that the distance from any point p to the plane in $\mathbb{R}^3$ given by $f(x) = 0$ equals $|f(p)|$.

(2) Suppose $b \in \mathbb{R}$. Show that the distance from any point p to the plane in $\mathbb{R}^3$ given by $f(x) = b$ equals $|f(p) - b|$.

1.6.7. Finish Example 1.48 by computing the distances from a general point $x \in \mathbb{R}^3$ to the line $L(a)$ and to the hyperplane $a^\perp$ with $a = (1, 1, 1)$.

1.6.8. Given $a = (a_1, a_2, a_3)$ and $b = (b_1, b_2, b_3)$ in $\mathbb{R}^3$, the *cross product* of a and b is the vector

$$a \times b = (a_2b_3 - a_3b_2, a_3b_1 - a_1b_3, a_1b_2 - a_2b_1).$$

(1) Show that $a \times b$ is perpendicular to a and b.
(2) Show $\|a \times b\|^2 = \|a\|^2\, \|b\|^2 - \langle a, b\rangle^2$.
(3) Show $\|a \times b\| = \|a\|\, \|b\| \sin(\theta)$, where θ is the angle between a and b.
(4) Show that the area of the parallelogram spanned by a and b equals $\|a \times b\|$.
(5) Show that the distance from a point $c \in \mathbb{R}^3$ to the plane containing $0, a$, and b equals

$$\frac{|\langle a \times b, c\rangle|}{\|a \times b\|}.$$

(6) Show that the volume of the parallelepiped spanned by vectors $a, b, c \in \mathbb{R}^3$ equals $|\langle a \times b, c\rangle|$.

1.6.9. Let $L \subset \mathbb{R}^3$ be the line through two distinct points $p, q \in \mathbb{R}^3$ and set $v = q - p$. Show that for every point $r \in \mathbb{R}^3$ the distance $d(r, L)$ from r to L equals

$$\frac{\|v \times (r - p)\|}{\|v\|}$$

(see Exercise 1.6.8).

1.6.10. Let $H \subset \mathbb{R}^4$ be the hyperplane with normal $a = (1, -1, 1, -1)$ and containing the point $q = (1, 2, -1, -3)$. Determine the distance from the point $(2, 1, -3, 1)$ to H.

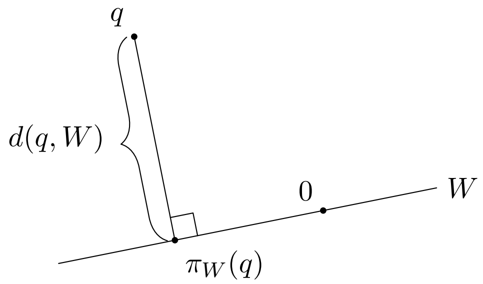


FIGURE 1.16. Distance from q to W

1.6.11. Let $W \subset F^n$ be a line or a hyperplane, not necessarily containing 0, and let $q \in F^n$ be a point. In Proposition 1.43 and the subsequent paragraph, we defined the orthogonal projection $\pi_W(q)$ of q onto W. Proposition 1.47 states that if W contains 0, then $\pi_W(q)$ is the nearest point to q on W. Show that this is true in general, that is, we have

$$d(q, W) = d(q, \pi_W(q)) = \|q - \pi_W(q)\|.$$

See Figure 1.16.

1.7. Reflections

If $H \subset \mathbb{R}^3$ is a plane, and $v \in \mathbb{R}^3$ is a point, then, roughly speaking, the reflection of v in H is the point $\tilde{v}$ on the other side of H that is just as far from H and for which the vector $\tilde{v} - v$ is normal to H (see Figure 1.17). This is made precise in Exercise 1.7.8 for general hyperplanes in F^n, but we will use a slightly different description.

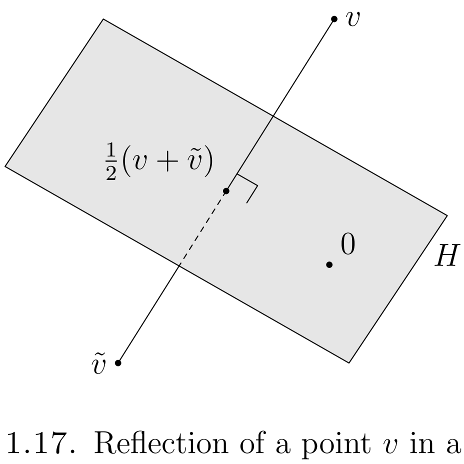


FIGURE 1.17. Reflection of a point v in a plane H

Note that in our rough description above, the element $\tilde{v}$ being just as far from H as v, yet on the other side of H, means that the midpoint $\frac{1}{2}(v + \tilde{v})$ between v and

$\tilde{v}$ is on H. This allows us to formulate an equivalent description of $\tilde{v}$, which avoids the notion of distance. Proposition 1.52 makes this precise, and also applies to lines.

1.7.1. Reflecting in lines and hyperplanes containing zero.

In this subsection, we let W denote a line or a hyperplane with $0 \in W$.

Proposition 1.52. *Let $v \in F^n$ be a point. Then there is a unique vector $\tilde{v} \in F^n$ such that*

(1) *the vector $v - \tilde{v}$ is normal to W, and*
(2) *we have $\frac{1}{2}(v + \tilde{v}) \in W$.*

This point equals $2\pi_W(v) - v$.

Proof. Let $\tilde{v} \in F^n$ be arbitrary and set $z = \frac{1}{2}(v + \tilde{v})$. Then $v - z = \frac{1}{2}(v - \tilde{v})$ is normal to W if and only if $v - \tilde{v}$ is. Since W contains 0, this happens if and only if $z - v \in W^\perp$ (see Exercise 1.4.6). Hence, by Proposition 1.42, the element $\tilde{v}$ satisfies the two conditions if and only if we have $z = \pi_W(v)$, that is, $\tilde{v} = 2\pi_W(v) - v$. □

Definition 1.53. The *reflection* in W is the map $s_W \colon F^n \to F^n$ that sends a vector $v \in F^n$ to the unique element $\tilde{v}$ of Proposition 1.52, so

$$s_W(v) = 2\pi_W(v) - v. \tag{1.5}$$

Note that the identity (1.5) is equivalent to the identity $s_W(v) - v = 2(\pi_W(v) - v)$, so the vectors $s_W(v) - v$ and $\pi_W(v) - v$ are both normal to W and the former is the double of the latter. In fact, this last vector equals $-\pi_{W^\perp}(v)$ by the identity $v = \pi_W(v) + \pi_{W^\perp}(v)$, so we also have

$$s_W(v) = v - 2\pi_{W^\perp}(v) \tag{1.6}$$

and

$$s_W(v) = \pi_W(v) - \pi_{W^\perp}(v). \tag{1.7}$$

From this last identity and the uniqueness mentioned in Proposition 1.30 we find the orthogonal projections of the point $s_W(v)$ onto W and $W^\perp$. They satisfy

$$\pi_W(s_W(v)) = \pi_W(v) \qquad \text{and} \qquad \pi_{W^\perp}(s_W(v)) = -\pi_{W^\perp}(v),$$

so the vector v and its reflection $s_W(v)$ in W have the same projection onto W, and the opposite projection onto $W^\perp$. This implies the useful properties

$$s_W(s_W(v)) = v, \tag{1.8}$$

$$s_W(v) = -s_{W^\perp}(v), \tag{1.9}$$

$$d(v, W) = d(s_W(v), W). \tag{1.10}$$

To make it more concrete, let $a \in \mathbb{R}^n$ be nonzero and set $L = L(a)$ and $H = a^\perp$. Let $v \in \mathbb{R}^n$ be a point and let $v_1 = \pi_a(v)$ and $v_2 = \pi_H(v)$ be its orthogonal projections on L and H, respectively. By Proposition 1.30, we have $v_1 = \lambda a$ with $\lambda = \frac{\langle a, v\rangle}{\|a\|^2}$, so we find

$$s_H(v) = v - 2v_1 = v - 2\frac{\langle a, v\rangle}{\|a\|^2} \cdot a \tag{1.11}$$

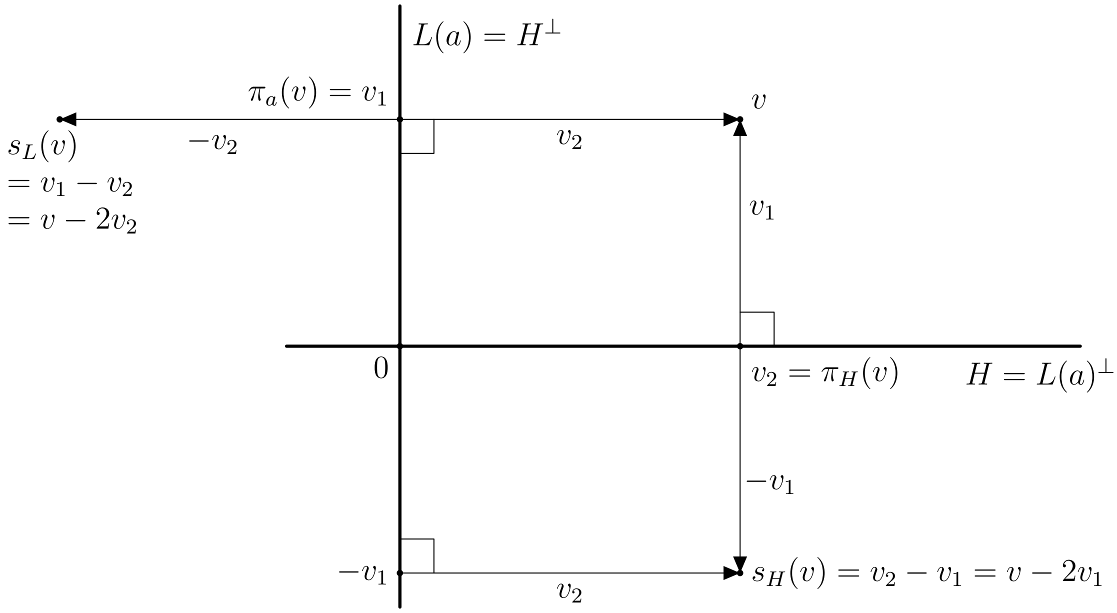


FIGURE 1.18. Reflection of v in $L = L(a)$ and in $H = a^\perp$

and $s_L(v) = -s_H(v)$. See Figure 1.18 for a schematic depiction of this, with H drawn as a line (which it would be in $\mathbb{R}^2$). Figure 1.19 shows the same in $\mathbb{R}^3$, this time with the plane H actually drawn as a plane. It is a useful exercise to identify identity (1.5), which can be rewritten as $s_W(v) - v = 2(\pi_W(v) - v)$, and the equivalent identities (1.6) and (1.7) in both figures (for both $W = L$ and $W = H$, and for the various points shown)!

We still consider $H \subset \mathbb{R}^3$, as in Figure 1.19. For $v \in H$ we have $\pi_H(v) = v$ and $\pi_L(v) = 0$, so $s_H(v) = v$ and $s_L(v) = -v$. This means that on H, the reflection in the line L corresponds to rotation around 0 over 180 degrees. We leave it as an exercise to show that on the whole of $\mathbb{R}^3$, the reflection in the line L is the same as rotation around the line over 180 degrees.

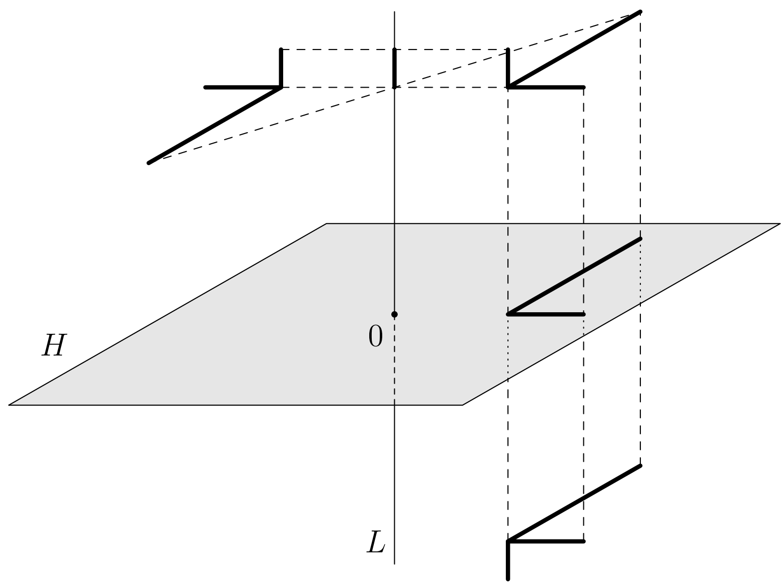


FIGURE 1.19. An object with its orthogonal projections on L and H, and its reflections in L and H

Example 1.54. Let $H \subset \mathbb{R}^3$ be the plane through 0 with normal $a = (0, 0, 1)$, and set $L = L(a)$. For any point $v = (x, y, z)$, the orthogonal projection $\pi_L(v)$ equals $(0, 0, z)$, so we find $s_H(v) = (x, y, -z)$ and $s_L(v) = (-x, -y, z)$.

Example 1.55. Let $M \subset \mathbb{R}^2$ be the line consisting of all points (x, y) satisfying $y = -2x$. Then $M = a^{\perp}$ for $a = (2, 1)$, that is, a is a normal of M. The reflection of the point $p = (3, 4)$ in M is

$$s_M(p) = p - 2\pi_a(p) = p - 2\frac{\langle p, a\rangle}{\langle a, a\rangle}a = p - 2 \cdot \frac{10}{5} \cdot a = p - 4a = (-5, 0).$$

Draw a picture to verify this.

Exercises

1.7.1. Let $L \subset \mathbb{R}^2$ be the line of all points (x_1, x_2) satisfying $x_2 = 2x_1$. Determine the reflection of the point $(5, 0)$ in L.

1.7.2. Let $L \subset \mathbb{R}^2$ be the line of all points (x_1, x_2) satisfying $x_2 = 2x_1$. Determine the reflection of the point (z_1, z_2) in L for all $z_1, z_2 \in \mathbb{R}$.

1.7.3. Let $V \subset \mathbb{R}^3$ be the plane through 0 that has $a = (3, 0, 4)$ as normal. Determine the reflections of the point $(1, 2, -1)$ in V and $L(a)$.

1.7.4. Let $W \subset F^n$ be a line or a hyperplane, and assume $0 \in W$. Use Exercise 1.5.8 to show that
(1) for every $x, y \in F^n$ we have $s_W(x + y) = s_W(x) + s_W(y)$, and
(2) for every $x \in F^n$ and every $\lambda \in F$ we have $s_W(\lambda x) = \lambda s_W(x)$.

1.7.5. Let $a \in F^n$ be nonzero and set $L = L(a)$. Let $p \in L$ be a point, and let $H \subset F^n$ be the hyperplane with normal $a \in F^n$ and containing the point p.
(1) Show that for every point $v \in H$, we have $s_L(v) - p = -(v - p)$ (see Exercise 1.5.10).
(2) Conclude that for $n = 3$ the restriction of the reflection s_L to H coincides with rotation within H around p over 180 degrees.
(3) Conclude that for $n = 3$ the reflection s_L in L coincides with rotation around the line L over 180 degrees (cf. Figure 1.19).

1.7.2. Reflecting in arbitrary lines and hyperplanes.

In this subsection, we generalise reflections to arbitrary lines and hyperplanes, not necessarily containing 0. It relies on orthogonal projections, which for general lines and hyperplanes are defined in Definition 1.44. In this subsection, we no longer assume that W is a line or a hyperplane containing 0.

Proposition 1.56. *Let $W \subset F^n$ be a line or a hyperplane, and $v \in F^n$ a point. Then there is a unique vector $\tilde{v} \in F^n$ such that*

(1) *the vector $v - \tilde{v}$ is normal to W, and*
(2) *we have $\frac{1}{2}(v + \tilde{v}) \in W$.*

Moreover, this point equals $2\pi_W(v) - v$.

Proof. Let $\tilde{v} \in F^n$ be arbitrary and set $z = \frac{1}{2}(v + \tilde{v})$. Then $v - z = \frac{1}{2}(v - \tilde{v})$ is normal to W if and only if $v - \tilde{v}$ is. Hence, $\tilde{v}$ satisfies the two conditions if and only if we have $z = \pi_W(v)$, that is, $\tilde{v} = 2\pi_W(v) - v$. □

Definition 1.57. Let $W \subset F^n$ be a line or a hyperplane. The *reflection*

$$s_W : F^n \to F^n$$

is the map that sends a vector $v \in F^n$ to the unique element $\tilde{v}$ of Proposition 1.56, that is,

$$s_W(v) = 2\pi_W(v) - v.$$

Clearly, this is consistent with Definition 1.53 for lines and hyperplanes that contain 0.

Warning 1.58. The reflection s_W in W is defined in terms of the projection π_W, just as in (1.5) for the special case that W contains 0. Note, however, that the alternative descriptions (1.6) and (1.7) only hold in this special case.

Proposition 1.59. *Let $W \subset F^n$ be a line or a hyperplane, and $p \in F^n$ a point. Then the hyperplane $W' = \{x - p \ : \ x \in W\}$ contains* 0 *and we have*

$$s_W(v) - p = s_{W'}(v - p).$$

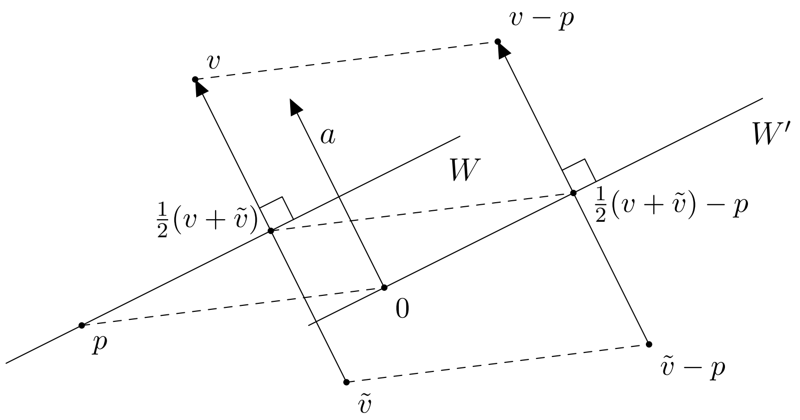


FIGURE 1.20. Reflection of v in a line or hyperplane W

Proof. We have $s_W(v) = 2\pi_W(v) - v$ and $s_{W'}(v - p) = 2\pi_{W'}(v - p) - (v - p)$ by Definition 1.57. Hence, the proposition follows from the fact that we have $\pi_W(v) = p + \pi_{W'}(v - p)$ by Definition 1.44. □

Proposition 1.59 states that if we want to reflect v in a line or hyperplane that does not contain 0, then we may first translate everything so that the resulting line or hyperplane does contain 0, then we reflect, and then we translate back. See Figure 1.20 and the end of Subsection 1.5.2.

Example 1.60. Consider the vector $a = (-1, 2, 3) \in \mathbb{R}^3$ and the plane

$$V = \{ v \in \mathbb{R}^3 \ : \ \langle a, v \rangle = 2\}.$$

We will compute the reflection of the point $q = (0, 3, 1)$ in V. Note that $p = (0, 1, 0)$ is contained in V, and set $q' = q - p = (0, 2, 1)$ and

$$V' = \{v - p \ : \ v \in V\}.$$

The vector a is normal to the plane V' and V' contains 0, so we have $V' = a^\perp$. The projection $\pi_a(q')$ of q' onto $L(a)$ is λa with $\lambda = \frac{\langle a,q'\rangle}{\langle a,a\rangle} = \frac{1}{2}$. Hence, we have

$$s_{V'}(q') = 2\pi_{a^\perp}(q') - q' = q' - 2\pi_a(q') = q' - 2\lambda a = q' - a = (1, 0, -2).$$

Hence, we have $s_V(q) = s_{V'}(q') + p = (1, 1, -2)$.

Exercises

1.7.6. Let $V \subset \mathbb{R}^3$ be the plane that has normal $a = (1, 2, -1)$ and that goes through the point $p = (1, 1, 1)$. Determine the reflection of the point $(1, 0, 0)$ in V.

1.7.7. Let $p, q \in \mathbb{R}^n$ be two different points. Let $V \subset \mathbb{R}^n$ be the set of all points in $\mathbb{R}^n$ that have the same distance to p as to q, that is,

$$V = \{\, v \in \mathbb{R}^n \;:\; \|v - p\| = \|v - q\| \,\}.$$

(1) Show that V is the hyperplane of all $v \in \mathbb{R}^n$ that satisfy

$$\langle q - p, v\rangle = \frac{1}{2}(\|q\|^2 - \|p\|^2).$$

(2) Show $q - p$ is a normal of V and that the point $\frac{1}{2}(p + q)$ is contained in V.
(3) Show that the reflection of p in V is q.

1.7.8. Let $H \subset F^n$ be a hyperplane and $v \in F^n$ a point that is not contained in H. Show that there is a unique vector $\tilde{v} \in F^n$ such that

(1) $v \neq \tilde{v}$,
(2) the vector $v - \tilde{v}$ is normal to H, and
(3) we have $d(v, H) = d(\tilde{v}, H)$.

Show that this vector $\tilde{v}$ is the reflection of v in H.

1.7.9. Let $p, q \in \mathbb{R}^3$ be two distinct points, and let L be the line through p and q. Let $H \subset \mathbb{R}^3$ be the plane through p that is orthogonal to L, that is, the vector $a = q - p$ is normal to H.

(1) Show that for every $v \in H$ we have $v - p \in a^\perp$.
(2) Show that for every $v \in H$ we have $\pi_L(v) = p$.
(3) Show that for every $v \in H$ we have $s_L(v) - p = -(v - p)$.
(4) Conclude that the restriction of the reflection s_L to H coincides with rotation within H around p over 180 degrees.
(5) Conclude that the reflection s_L in L coincides with rotation around the line L over 180 degrees (cf. Figure 1.19).

1.8. Cauchy-Schwarz

We would like to *define* the angle between two vectors in $\mathbb{R}^n$ by letting the angle $\alpha \in [0, \pi]$ between two nonzero vectors v, w be determined by (1.2). However, before we can do that, we need to know that the value on the right-hand side of (1.2) lies in the interval $[-1, 1]$. We will first prove that this is indeed the case.

Proposition 1.61 (Cauchy-Schwarz). *For all vectors $v, w \in F^n$ we have*

$$|\langle v, w\rangle| \leq \|v\| \cdot \|w\|$$

and equality holds if and only if there are $\lambda, \mu \in F$, not both zero, such that $\lambda v + \mu w = 0$.

Proof. If $v = 0$, then we automatically have equality, and for $\lambda = 1$ and $\mu = 0$ we have $\lambda v + \mu w = 0$. Suppose $v \neq 0$. Let z be the orthogonal projection of w onto $v^\perp$ (see Definition 1.31, so our vectors v, w, z correspond to a, v, v_2 of

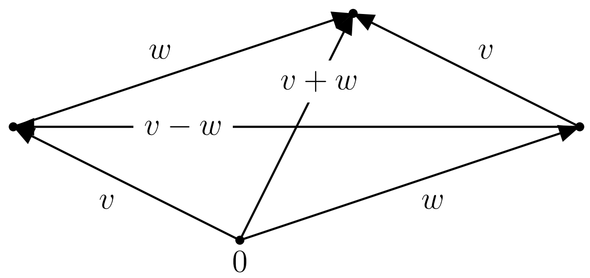


FIGURE 1.21. Arrows representing the vectors v, w and $v \pm w$ make a triangle

Proposition 1.30, respectively). Then by Proposition 1.30 we have

$$\|z\|^2 = \|w\|^2 - \frac{\langle v, w\rangle^2}{\|v\|^2}.$$

From $\|z\|^2 \geq 0$ we conclude $\langle v, w\rangle^2 \leq \|v\|^2 \cdot \|w\|^2$, which implies the inequality, as lengths are non-negative. We have equality if and only if $z = 0$, so if and only if $w = \lambda v$ for some $\lambda \in F$, in which case we have $\lambda v + (-1) \cdot w = 0$. Conversely, if we have $\lambda v + \mu w = 0$ with λ and μ not both zero, then we have $\mu \neq 0$, for otherwise $\lambda v = 0$ would imply $\lambda = 0$; therefore, we have $w = -\lambda\mu^{-1}v$, so w is a multiple of v and the inequality is an equality. □

The triangle inequality usually refers to the inequality $c \leq a + b$ for the sides a, b, c of a triangle in $\mathbb{R}^2$ or $\mathbb{R}^3$. Proposition 1.62 generalises this to F^n. See Figure 1.21.

Proposition 1.62 (Triangle inequality). *For all vectors $v, w \in F^n$ we have*

$$\|v + w\| \leq \|v\| + \|w\|$$

and equality holds if and only if there are non-negative scalars $\lambda, \mu \in F$, not both zero, such that $\lambda v = \mu w$.

Proof. By the inequality of Cauchy-Schwarz, Proposition 1.61, we have

$$\begin{aligned}\|v + w\|^2 &= \langle v + w, v + w\rangle = \langle v, v\rangle + 2\langle v, w\rangle + \langle w, w\rangle \\ &= \|v\|^2 + 2\langle v, w\rangle + \|w\|^2 \leq \|v\|^2 + 2 \cdot \|v\| \cdot \|w\| + \|w\|^2 = (\|v\| + \|w\|)^2.\end{aligned}$$

Since all lengths are non-negative, we may take square roots to find the desired inequality. Equality holds if and only if $\langle v, w\rangle = \|v\| \cdot \|w\|$.

If $v = 0$ or $w = 0$, then clearly equality holds and there exist λ and μ as claimed: take one of them to be 1 and the other 0, depending on whether v or w equals 0. For the remaining case, we suppose $v \neq 0$ and $w \neq 0$.

Suppose equality holds in the triangle inequality. Then $\langle v, w\rangle = \|v\| \cdot \|w\|$, so by Proposition 1.61 there exist $\lambda', \mu' \in F$, not both zero, with $\lambda' v + \mu' w = 0$. Since v and w are nonzero, both λ' and μ' are nonzero. For $\lambda = 1$ and $\mu = -\mu'/\lambda'$ we have $v = \lambda v = \mu w$, and from

$$\|v\| \cdot \|w\| = \langle v, w\rangle = \langle \mu w, w\rangle = \mu\|w\|^2$$

we conclude $\mu \geq 0$.

Conversely, suppose $\lambda, \mu \geq 0$, not both zero, and $\lambda v = \mu w$. Then λ and μ are both nonzero, because v and w are nonzero. With $\nu = \mu/\lambda > 0$, we find $v = \nu w$, so we have $\langle v, w\rangle = \langle \nu w, w\rangle = \nu\|w\|^2 = |\nu| \cdot \|w| \cdot \|w\| = \|v\| \cdot \|w\|$, which implies that equality holds in the triangle inequality. □

Definition 1.63. For all nonzero vectors $v, w \in F^n$, we define the *angle* between v and w to be the unique real number $\alpha \in [0, \pi]$ that satisfies

$$\cos\alpha = \frac{\langle v, w\rangle}{\|v\| \cdot \|w\|}. \tag{1.12}$$

Note that the angle α between v and w is well defined, as by Proposition 1.61, the right-hand side of (1.12) lies between -1 and 1. By Proposition 1.19, the angle also corresponds with the usual notion of angle in $\mathbb{R}^2$ and $\mathbb{R}^3$ in the sense that the angle between v and w equals the angle between the two arrows that represent v and w and that have 0 as tail. Finally, Definitions 1.21 and 1.63 imply that two nonzero vectors v and w in F^n are orthogonal if and only if the angle between them is $\pi/2$.

Example 1.64. For $v = (3, 0)$ and $w = (2, 2)$ in $\mathbb{R}^2$ we have $\langle v, w\rangle = 6$, while $\|v\| = 3$ and $\|w\| = 2\sqrt{2}$. Therefore, the angle θ between v and w satisfies $\cos\theta = 6/(3 \cdot 2\sqrt{2}) = \frac{1}{2}\sqrt{2}$, so we have $\theta = \pi/4$.

Example 1.65. For $v = (1, 1, 1, 1)$ and $w = (1, 2, 3, 4)$ in $\mathbb{R}^4$ we have $\langle v, w\rangle = 10$, while $\|v\| = 2$ and $\|w\| = \sqrt{30}$. Therefore, the angle θ between v and w satisfies $\cos\theta = 10/(2 \cdot \sqrt{30}) = \frac{1}{6}\sqrt{30}$, so $\theta = \arccos\left(\frac{1}{6}\sqrt{30}\right)$.

Exercises

1.8.1. Prove that for all $v, w \in \mathbb{R}^n$ we have $\|v - w\| \leq \|v\| + \|w\|$. When does equality hold?

1.8.2. Prove the cosine rule in $\mathbb{R}^n$.

1.8.3. Suppose $v, w \in F^n$ are nonzero, and let α be the angle between v and w.
  (1) Prove that $\alpha = 0$ if and only if there are positive $\lambda, \mu \in F$ with $\lambda v = \mu w$.
  (2) Prove that $\alpha = \pi$ if and only if there are $\lambda, \mu \in F$ with $\lambda < 0$ and $\mu > 0$ and $\lambda v = \mu w$.

1.8.4. Determine the angle between the vectors $(1, -1, 2)$ and $(-2, 1, 1)$ in $\mathbb{R}^3$.

1.8.5. Let $p, q, r \in \mathbb{R}^n$ be three points. Show that p, q, and r are collinear (they lie on one line) if and only if we have

$$\langle p - r, q - r\rangle^2 = \langle p - r, p - r\rangle \cdot \langle q - r, q - r\rangle.$$

1.8.6. Determine the angle between the vectors $(1, -1, 1, -1)$ and $(1, 0, 1, 1)$ in $\mathbb{R}^4$.

1.8.7. The angle between two hyperplanes is defined as the angle between their normal vectors. Determine the angle between the hyperplanes in $\mathbb{R}^4$ given by $x_1 - 2x_2 + x_3 - x_4 = 2$ and $3x_1 - x_2 + 2x_3 - 2x_4 = -1$, respectively.

1.9. What is next?

We have seen that $\mathbb{R}^n$ is a set with an addition, a subtraction, and a scalar multiplication, satisfying the properties mentioned in Section 1.1. This makes $\mathbb{R}^n$ our first example of a *vector space*, which we will define in the next chapter. In fact, a vector space is nothing but a set together with an addition and a scalar multiplication satisfying a priori only some of those same properties. The subtraction and the other properties will then come for free! Because we only have addition and scalar multiplication, all our operations are linear, which is why the study of vector spaces is called *linear algebra*.

In the next chapter we will see many more examples of vector spaces, such as the space of all functions from $\mathbb{R}$ to $\mathbb{R}$. Lines and hyperplanes in $\mathbb{R}^n$ that contain 0 are vector spaces as well. In fact, so is the zero space $\{0\} \subset \mathbb{R}^n$. Because these are all contained in $\mathbb{R}^n$, we call them *subspaces* of $\mathbb{R}^n$.

One of the most important notions in linear algebra is the notion of *dimension*, which we will define for general vector spaces in Chapter 7. It will not come as a surprise that our examples $\mathbb{R}^1$, $\mathbb{R}^2$, and $\mathbb{R}^3$ have dimension 1, 2, and 3, respectively. Indeed, the vector space $\mathbb{R}^n$ has dimension n. Lines (containing 0) have dimension 1, and every hyperplane in $\mathbb{R}^n$ (containing 0) has dimension $n-1$, which means that planes in $\mathbb{R}^3$ have dimension 2, as one would expect.

For $\mathbb{R}^n$ with $n \leq 3$ this covers all dimensions, so every subspace of $\mathbb{R}^n$ with $n \leq 3$ is either $\{0\}$ or $\mathbb{R}^n$ itself, or a line or a hyperspace, and these last two notions are the same in $\mathbb{R}^2$. For $n \geq 4$, however, these are far from all subspaces of $\mathbb{R}^n$, which exist for any dimension between 0 and n. All of them are intersections of hyperplanes containing 0.

The theory of linear algebra allows us to generalise some of the important results of this chapter about lines and hyperplanes to all subspaces of $\mathbb{R}^n$. For example, in Proposition 1.30 and Corollary 1.41, we have seen that for every line or hypersurface $W \subset \mathbb{R}^n$ containing 0, we can write every $v \in \mathbb{R}^n$ uniquely as $v = v_1 + v_2$ with $v_1 \in W$ and $v_2 \in W^\perp$. This does indeed hold for any subspace W of $\mathbb{R}^n$ (see Corollary 8.24). Moreover, for every subspace $W \subset \mathbb{R}^n$ we have $(W^\perp)^\perp = W$ (see Proposition 8.20), thus generalising Proposition 1.35. Both results make extensive use of theorems about dimensions. These two results can be used to compute the intersection of any two subspaces, or to solve any system of linear equations. The last of these two results can also be used to parametrise any subspace and translates thereof, including hyperplanes. In this chapter, we have only done this for lines (see Proposition 1.9). Orthogonal projections and reflections can also be defined with respect to any subspace of $\mathbb{R}^n$, just like distances from points to any (translate of a) subspace.

But linear algebra can be applied to many more vector spaces than only those contained in $\mathbb{R}^n$. For example, the set of all functions from $\mathbb{R}$ to $\mathbb{R}$ is a vector space of infinite dimension, to which our theory will apply just as easily as to $\mathbb{R}^n$! Therefore, most of this book will be about general vector spaces. As mentioned before, the space $\mathbb{R}^n$ of this first chapter is just one example.

As opposed to what we did in this chapter, we will also consider fields F that are not contained in $\mathbb{R}$. This allows examples over the field $\mathbb{C}$ of complex numbers and even over the field $\mathbb{F}_2 = \{0, 1\}$ of two elements (in which we have $1+1 = 0$), which is widely used in cryptography. The precise definition of a field (and of $\mathbb{C}$ and $\mathbb{F}_2$) is given in Appendix B, but, if wanted, readers can skip this definition and think of a field as just $\mathbb{R}$ (or as a subset of $\mathbb{R}$ containing 0 and 1 in which we can add, subtract, multiply, and divide by any nonzero element). They will still be able to learn linear algebra from this book by skipping a few examples, exercises, and remarks about fields such as $\mathbb{F}_2$, which are indicated by the symbol ††.

The real strength of linear algebra comes from the understanding of *linear maps*, which are functions between vector spaces that preserve the linear structure (the addition and the scalar multiplication) of the spaces. Linear maps are defined in Chapter 4. Matrices are a convenient way to describe maps from F^n to F^m and to do explicit computations. They are defined in Chapter 5. The last chapters of this book are dedicated to understanding various aspects of linear maps.

CHAPTER 2

Vector spaces

In Section 1.1 we have seen that the newly defined addition ($\oplus$) and scalar multiplication ($\odot$) on Euclidean space $\mathbb{R}^n$ behave so closely to the regular addition and multiplication, that we use the regular notations ($+$ and $\cdot$) for them. Although much of Chapter 1 also relies on the scalar product, we can prove many interesting theorems about Euclidean space using just the addition and scalar multiplication and the fact that they satisfy the properties (1)-(9) mentioned in Section 1.1.

It turns out that in mathematics we encounter many other sets V where one could define an interesting new addition and a scalar multiplication satisfying the same properties (1)-(9) of Section 1.1. Any proof of a fact about Euclidean space $\mathbb{R}^n$ that only makes use of these properties of addition and scalar multiplication is then also a proof of the analogous fact for V.

Rather than stating these facts and their proofs for all sets with an addition and a scalar multiplication separately, we define the abstract notion of a *vector space*, which is a set in which we can *add* and *scale* elements, and where the addition and scaling satisfy eight simple rules, called *axioms*. Euclidean space $\mathbb{R}^n$ then becomes merely an *example* of a vector space.

Linear algebra is the study of these abstract vector spaces in general and starts with proving that the properties (1)-(9) of Section 1.1 follow from the axioms. By proving theorems using only these axioms and all the rules that follow from them, we prove these theorems for all vector spaces at once.

As mentioned, in Chapter 1 we have seen the first examples, namely $V = F^n$ for any subfield F of $\mathbb{R}$, that is, for any subset $F \subset \mathbb{R}$ containing 0 and 1 in which we can add, multiply, subtract, and divide (except by 0). The scaling, or scalar multiplication, scales elements of V by elements of F. For the rest of this book, we do not require that F is a subset of $\mathbb{R}$. All we require from our scaling factors, or *scalars*, is that they form a *field*, which means that –roughly speaking– they form a set in which we can somehow add, subtract, and multiply elements, and divide by any nonzero element. See Appendix B for a precise definition of fields.

For the rest of this book, we let F denote a field; elements of F are called *scalars*.

So as not to force all readers to first study the theory of fields, this book is set up to allow some simplifications.

- Readers may assume that F is (contained in) the field $\mathbb{C}$ of complex numbers, in which case they should skip all examples, exercises, and remarks indicated by ††.
- Readers may assume that F is (contained in) the field $\mathbb{R}$ of real numbers, in which case they should skip all examples, exercises, and remarks indicated by † and ††.

Under these simplifying asumptions, the definition of a field reduces precisely to F being a subset of $\mathbb{R}$ or $\mathbb{C}$ that contains 0 and 1 and in which we can add, subtract, and multiply elements, and divide by any nonzero element. Examples are $\mathbb{R}$ and $\mathbb{C}$ themselves, and the field $\mathbb{Q}$ of rational numbers.

We will often use the field $\mathbb{R}$ of real numbers in our examples, but by allowing ourselves to work with general fields, we also cover linear algebra over the field $\mathbb{C}$ of complex numbers, and over finite fields, such as the field $\mathbb{F}_2 = \{0, 1\}$ of two elements (with $1 + 1 = 0$), which has important applications in computer science, cryptography, and coding theory. For the definitions of $\mathbb{C}$ and $\mathbb{F}_2$, see Appendix B.

2.1. Definition of a vector space

Roughly speaking, a vector space over the field F is just a set V of which we can add any two elements to get a new element of V, and of which we can scale any element by an element of F. The addition and scalar multiplication have to satisfy some rules, and the exact definition of a vector space is as follows.

Definition 2.1. A *vector space* or *linear space* over F, or an *F-vector space*, is a set V with a distinguished zero element $0_V \in V$, together with an operation $\oplus$ ('addition'[1]) that assigns to two elements $x, y \in V$ their sum $x \oplus y \in V$, and an operation $\odot$ ('scalar multiplication'[2]) that assigns to a scalar $\lambda \in F$ and an element $x \in V$ the scaled multiple $\lambda \odot x \in V$ of x, such that these operations satisfy the following axioms.

(1) For all $x, y \in V$, $\boxed{x \oplus y = y \oplus x}$ (addition is commutative).

(2) For all $x, y, z \in V$, $\boxed{(x \oplus y) \oplus z = x \oplus (y \oplus z)}$ (addition is associative).

(3) For all $x \in V$, $\boxed{x \oplus 0_V = x}$ (adding the zero element does nothing).

(4) For every $x \in V$, there is an $x' \in V$ such that $\boxed{x \oplus x' = 0_V}$ (existence of negatives).

(5) For all $\lambda, \mu \in F$ and $x \in V$, $\boxed{\lambda \odot (\mu \odot x) = (\lambda \cdot \mu) \odot x}$ (scalar multiplication is associative).

(6) For all $x \in V$, $\boxed{1 \odot x = x}$ (multiplication by 1 is the identity).

(7) For all $\lambda \in F$ and $x, y \in V$, $\boxed{\lambda \odot (x \oplus y) = (\lambda \odot x) \oplus (\lambda \odot y)}$ (distributivity I).

(8) For all $\lambda, \mu \in F$ and $x \in V$, $\boxed{(\lambda + \mu) \odot x = (\lambda \odot x) \oplus (\mu \odot x)}$ (distributivity II).

The elements of a vector space are usually called *vectors*. A *real* vector space is a vector space over the field $\mathbb{R}$ of real numbers and (†) a *complex* vector space is a vector space over the field $\mathbb{C}$ of complex numbers.

Remarks 2.2.

(1) Instead of writing $(V, 0_V, \oplus, \odot)$ (which is the complete data for a vector space), we usually just write V, with the zero element, the addition, and scalar multiplication being understood.

[1] Addition is a function $V \times V \to V$ that sends the pair (x, y) to $x \oplus y$.

[2] Scalar multiplication is a function $F \times V \to V$ that sends the pair (λ, x) to $\lambda \odot x$.

(2) We will often leave out the subscript V in 0_V, and just write 0 for the zero of the vectorspace. It is crucial to always distinguish this from the zero of the field F, even though both may be written as 0; it should always be clear from the context which zero is meant.

(3) For now, we denote the addition and scalar multiplication of a vector space by the symbols $\oplus$ and $\odot$, in order to distinguish them from the addition and multiplication in F. Soon, we will see that they behave so much like the usual addition and scaling, that we drop the circles in the notation.

Exercises

2.1.1. Suppose that F is contained in $\mathbb{R}$. Show that F^n together with the zero element and the coordinate-wise addition $\oplus$ and scalar multiplication $\odot$ as defined in Section 1.1 is a vector space. [In Example 2.5 we will generalise this to general fields.]

2.1.2. Let $V \subset \mathbb{R}^3$ be the set of all triples $(x_1, x_2, x_3) \in \mathbb{R}^3$ with $x_1+x_2+x_3 = 0$. Is V, together with the usual coordinate-wise addition and scalar multiplication, and the zero vector of $\mathbb{R}^3$, a vector space?

2.1.3. Let $V \subset \mathbb{R}^3$ be the set of all triples $(x_1, x_2, x_3) \in \mathbb{R}^3$ with $x_1+x_2+x_3 = 1$. Is V, together with the usual coordinate-wise addition and scalar multiplication, and the zero vector of $\mathbb{R}^3$, a vector space?

2.1.4. Let V be a vector space over F. In the following table, with a and b elements of F or V as given, indicate whether the elements $a \odot b$ and $a \oplus b$ are defined and, if so, whether they are contained in F or in V.

a	b	$a \odot b$	$a \oplus b$
F	F		
F	V		
V	V		

2.2. Examples

Recall that F is a field (see the beginning of this chapter).

Example 2.3. The simplest (and perhaps least interesting) example of a vector space over F is $V = \{0\}$, with addition given by $0 \oplus 0 = 0$ and scalar multiplication by $\lambda \odot 0 = 0$ for all $\lambda \in F$ (these are the only possible choices). Trivial as it may seem, this vector space, called the *zero space*, is important. It plays a role in linear algebra similar to the role played by the empty set in set theory.

Example 2.4. The next (still not very interesting) example is $V = F$ over itself, with addition, multiplication, and the zero being the ones that make F into a field. The axioms above in this case just reduce to the rules for addition and multiplication in F (see Appendix B).

Example 2.5. Now we come to a very important example, which is *the* model of a vector space over F. For F contained in $\mathbb{R}$, it was already studied extensively in Chapter 1 (cf. Exercise 2.1.1). Let n be a non-negative integer. We consider the set $V = F^n$ of n-tuples of elements of F. As in Section 1.1, we define addition and scalar multiplication 'component-wise':

$$(x_1, x_2, \dots, x_n) \oplus (y_1, y_2, \dots, y_n) = (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n),$$
$$\lambda \odot (x_1, x_2, \dots, x_n) = (\lambda x_1, \lambda x_2, \dots, \lambda x_n).$$

Also as in Section 1.1, we set $0_V = (0, 0, \dots, 0)$.

Of course, we now have to *prove* that our eight axioms are satisfied by our choice of $(V, 0_V, \oplus, \odot)$. In this case, this is very easy[3], since everything reduces to addition and multiplication in the field F. As an example, let us spell out in complete detail that the first distributive law (7) and the existence of negatives (4) are satisfied. We leave the other properties as an exercise.

For the first distributive law (7), take $x, y \in F^n$ and write them as

$$x = (x_1, x_2, \dots, x_n) \qquad \text{and} \qquad y = (y_1, y_2, \dots, y_n).$$

Then we have

$$\begin{aligned}
\lambda \odot (x \oplus y) &= \lambda \odot \big((x_1, x_2, \dots, x_n) \oplus (y_1, y_2, \dots, y_n)\big) \\
&= \lambda \odot (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n) \\
&= \big(\lambda(x_1 + y_1), \lambda(x_2 + y_2), \dots, \lambda(x_n + y_n)\big) \\
&= (\lambda x_1 + \lambda y_1, \lambda x_2 + \lambda y_2, \dots, \lambda x_n + \lambda y_n) \\
&= (\lambda x_1, \lambda x_2, \dots, \lambda x_n) \oplus (\lambda y_1, \lambda y_2, \dots, \lambda y_n) \\
&= \big(\lambda \odot (x_1, x_2, \dots, x_n)\big) \oplus \big(\lambda \odot (y_1, y_2, \dots, y_n)\big) \\
&= (\lambda \odot x) \oplus (\lambda \odot y),
\end{aligned}$$

where the first three and the last three equalities follow from the definitions of x, y and the operations $\oplus$ and $\odot$; the middle equality follows from the fact that for each i we have $\lambda(x_i + y_i) = \lambda x_i + \lambda y_i$ by the distributive law for the field F. This proves the first distributive law (7) for F^n.

For the existence of negatives (4), take an element $x \in F^n$ and write it as $x = (x_1, x_2, \dots, x_n)$. For each i with $1 \le i \le n$, we can take the negative $-x_i$ of x_i in the field F, where we already know we can take negatives, and set

$$x' = (-x_1, -x_2, \dots, -x_n).$$

Then, of course, we have

$$\begin{aligned}
x \oplus x' &= (x_1, x_2, \dots, x_n) \oplus (-x_1, -x_2, \dots, -x_n) \\
&= \big(x_1 + (-x_1), x_2 + (-x_2), \dots, x_n + (-x_n)\big) = (0, 0, \dots, 0) = 0_V,
\end{aligned}$$

which proves, indeed, that for every $x \in F^n$ there is an $x' \in F^n$ with $x + x' = 0_V$.

For $n = 1$, this example reduces to the previous one (if one identifies each element $x \in F$ with the 1-tuple (x)); for $n = 0$, it reduces to the zero space. (Why? Well, like an empty product of numbers should have the value 1, an empty product of sets like F^0 has exactly one element, the empty tuple $()$, which we can call 0 here.)

In physics, more precisely in the theory of relativity, $\mathbb{R}^4$ is often interpreted as space with a fourth coordinate for time.

[3]In fact, in Section 1.1 (where the fact that F was contained in $\mathbb{R}$ was actually never used) we already claimed that all these properties follow directly from the fact that the operations are defined coordinate-wise.

Example 2.6. Let F^∞ denote the set of all infinite sequences $(a_n)_{n\geq 0}$ of elements in F. Similar to Example 2.5, we define the addition and scalar multiplication component-wise, so

$$(a_0, a_1, a_2, \ldots) \oplus (b_0, b_1, b_2, \ldots) = (a_0 + b_0, a_1 + b_1, a_2 + b_2, \ldots)$$

and

$$\lambda \odot (a_0, a_1, a_2, \ldots) = (\lambda a_0, \lambda a_1, \lambda a_2, \ldots).$$

Together with the zero vector consisting of only zeros, this is again a vector space, and checking that all eight axioms are satisfied is just as easy as in Example 2.5.

Example 2.7. Suppose $F = \mathbb{R}$. A *magic square* is a square of 3×3 real numbers such that the three column sums, the three row sums and the two diagonal sums are all equal. An example is the following.

$$\begin{array}{|c|c|c|} \hline 8 & 1 & 6 \\ \hline 3 & 5 & 7 \\ \hline 4 & 9 & 2 \\ \hline \end{array}$$

This magic square is well known, because it uses all integers from 1 to 9 exactly once. Less interesting magic squares are

$$A = \begin{array}{|c|c|c|} \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline \end{array}, \quad B = \begin{array}{|c|c|c|} \hline 0 & -1 & 1 \\ \hline 1 & 0 & -1 \\ \hline -1 & 1 & 0 \\ \hline \end{array}, \quad \text{and} \quad C = \begin{array}{|c|c|c|} \hline 1 & -1 & 0 \\ \hline -1 & 0 & 1 \\ \hline 0 & 1 & -1 \\ \hline \end{array}.$$

Note that if we multiply each of the nine numbers in a magic square X by the same number λ, then we obtain a new square, which we denote by $\lambda \odot X$. If all rows, columns, and diagonals of X add up to s, then those of $\lambda \odot X$ all add up to $\lambda \cdot s$, so $\lambda \odot X$ is a magic square as well. Moreover, if we have two magic squares X and Y, then we can make a new magic square, which we will denote by $X \oplus Y$, by letting the top-left number in $X \oplus Y$ be the sum of the top-left numbers in X and Y, et cetera; if the sums in X and Y are all s and all t, respectively, then the sums of $X \oplus Y$ are all $s + t$. Check this, and verify that $((5\odot A)\oplus B)\oplus(3\odot C)$ equals the well-known magic square above. As mentioned above, we will see that $\oplus$ and $\odot$ behave as addition and scalar multiplication, so we will also write this combination as $5A+B+3C$. We leave it as an exercise to show that the set of magic squares, together with this addition $\oplus$ and scalar multiplication $\odot$ is a vector space over $\mathbb{R}$, with the square of all zeros as zero vector.

Definition 2.8. For any two sets A and B, the set of all functions from A to B is denoted by both $\mathrm{Map}(A, B)$ and B^A.

Remark 2.9. Obviously, if f is a function from A to B and a is an element of A, then $f(a)$ is an element of B. In our notation, we will always be careful to distinguish between the function f and the element $f(a)$. For example, in the case $A = B = \mathbb{R}$, we will **not** say: “the function $f(x) = x^2$.” Correct would be “the function f that is given by $f(x) = x^2$ for all $x \in \mathbb{R}$.”

Example 2.10. Suppose $F = \mathbb{R}$. Consider the set $\mathrm{Map}(\mathbb{R}, \mathbb{R})$ of all functions from $\mathbb{R}$ to $\mathbb{R}$. The sum of two functions $f, g \in \mathrm{Map}(\mathbb{R}, \mathbb{R})$ is the function $f \oplus g$ that is given by

$$(f \oplus g)(x) = f(x) + g(x)$$

for all $x \in \mathbb{R}$. The scalar multiplication of a function $f \in \mathrm{Map}(\mathbb{R}, \mathbb{R})$ by a factor $\lambda \in \mathbb{R}$ is the function $\lambda \odot f$ that is given by

$$(\lambda \odot f)(x) = \lambda \cdot (f(x))$$

for all $x \in \mathbb{R}$. Of course, this is just the usual addition and scaling of functions, and soon we will use the usual notation $f + g$ and λf again. The operations obviously satisfy the eight axioms, but it is a good exercise to write this out in detail. As an example, let us prove that the addition is associative.

Let $f, g, h \in \mathrm{Map}(\mathbb{R}, \mathbb{R})$ be three functions. We want to show that $p = (f \oplus g) \oplus h$ and $q = f \oplus (g \oplus h)$ are the same function. The two functions both have domain and codomain $\mathbb{R}$, so it suffices to prove that for all $x \in \mathbb{R}$ we have $p(x) = q(x)$. Indeed, for all $x \in \mathbb{R}$ we have

$$p(x) = ((f \oplus g) \oplus h)(x) = (f \oplus g)(x) + h(x) = (f(x) + g(x)) + h(x)$$

and

$$q(x) = (f \oplus (g \oplus h))(x) = f(x) + (g \oplus h)(x) = f(x) + (g(x) + h(x)),$$

which implies $p(x) = q(x)$, because addition in $\mathbb{R}$ is associative. We leave it to the reader to finish the verification that $\mathrm{Map}(\mathbb{R}, \mathbb{R})$ is indeed a vector space over $\mathbb{R}$, with the constant zero function that sends every $x \in \mathbb{R}$ to $0 \in \mathbb{R}$ as zero. (For the first distributive law, see Example 2.11, which generalises this example.)

Example 2.11. This example generalises Example 2.10. Let X be a set. Consider the set $V = \mathrm{Map}(X, F) = F^X$ of all maps (or functions) from X to F. In order to get a vector space over F, we have to define addition and scalar multiplication. To define addition, for every pair of functions $f, g \colon X \to F$, we have to define a new function $f \oplus g \colon X \to F$. The only reasonable way to do this is *point-wise*:

$$(f \oplus g)(x) = f(x) + g(x).$$

In a similar way, we define scalar multiplication:

$$(\lambda \odot f)(x) = \lambda \cdot f(x).$$

We take the zero vector 0_V to be the constant zero function that sends each element $x \in X$ to $0 \in F$. We then have to check the axioms in order to verify that we really get a vector space. Let us do again the first distributive law as an example. We have to check the identity $\lambda \odot (f \oplus g) = (\lambda \odot f) \oplus (\lambda \odot g)$, which means that for all $x \in X$, we want

$$\big(\lambda \odot (f \oplus g)\big)(x) = ((\lambda \odot f) \oplus (\lambda \odot g))(x)\,.$$

So let $\lambda \in F$ and $f, g\colon X \to F$ be given, and take any $x \in X$. Then we get

$$\begin{aligned}\big(\lambda \odot (f \oplus g)\big)(x) &= \lambda \cdot \big((f \oplus g)(x)\big)\\ &= \lambda \cdot \big(f(x) + g(x)\big)\\ &= \lambda \cdot f(x) + \lambda \cdot g(x)\\ &= (\lambda \odot f)(x) + (\lambda \odot g)(x)\\ &= \big((\lambda \odot f) \oplus (\lambda \odot g)\big)(x)\,,\end{aligned}$$

where all equalities, except for the middle one, follow from the definitions of the operators $\oplus$ and $\odot$; the middle equality follows from the first distributive law for F. We leave it to the reader to finish the verification that $\mathrm{Map}(X, F)$ is indeed a vector space over F.

Remark 2.12. Note the parallelism of this proof with the one of Example 2.5. That parallelism goes much further. If we take X to be $I = \{1, 2, \ldots, n\}$, then the vector space $F^I = \mathrm{Map}(I, F)$ of maps from $\{1, 2, \ldots, n\}$ to F can be identified with F^n by letting such a map f correspond to the n-tuple

$$(f(1), f(2), \ldots, f(n)).$$

Under this identification, the addition on F^I corresponds with the addition on F^n, and the same is true for scalar multiplication. It is not a coincidence that the notations F^I and F^n are chosen so similar! See also Proposition A.1.

Similarly, if we take $X = \mathbb{Z}_{\geq 0}$, then the vector space $\mathrm{Map}(\mathbb{Z}_{\geq 0}, F)$ can be identified with the vector space F^∞ of Example 2.6 by letting a map $f\colon \mathbb{Z}_{\geq 0} \to F$ correspond to the sequence

$$(\ f(0), f(1), f(2), f(3), \ldots\ )\,.$$

Cf. Example C.3. Again, the addition and scalar multiplication on $\mathrm{Map}(\mathbb{Z}_{\geq 0}, F)$ correspond with those on F^∞.

What do we get when X is the empty set?

Example 2.13. A *polynomial* in the variable x over F is a formal sum

$$f = a_d x^d + a_{d-1} x^{d-1} + \cdots + a_2 x^2 + a_1 x + a_0$$

of a finite number of scalar multiples of integral powers x^i (with $i \geq 0$); the products $a_i x^i$ are called the *terms* of f and we say that $a_i \in F$ is the *coefficient* of the *monomial* x^i in f. Here, we have used some intimidation, as we did not explain what a *variable* is, nor a *formal sum*. Any feeling of discomfort caused by this intimidation should be taken as a sign of good taste; a reader with this feeling is encouraged to read Appendix D, and to match what is said there with what we say in this example and the next.

We let the zero vector 0 be the zero polynomial: the polynomial of which all coefficients are 0. The *degree* of a nonzero polynomial $f = \sum_{i=0}^{d} a_i x^i$ with $a_d \neq 0$ is d. By definition, the degree of 0 equals $-\infty$. Let $F[x]$ denote the set of all polynomials over F.

A real *polynomial* in the variable x is a polynomial in the variable x over $\mathbb{R}$, so $\mathbb{R}[x]$ denotes the set of all real polynomials in the variable x.

We define the addition of polynomials coefficientwise. In other words, we collect equal powers of x, so that the sum of the polynomials

$$f = a_d x^d + \cdots + a_2 x^2 + a_1 x + a_0 \qquad \text{and} \qquad g = b_d x^d + \cdots + b_2 x^2 + b_1 x + b_0$$

in $F[x]$ equals

$$f \oplus g = (a_d + b_d)x^d + \cdots + (a_2 + b_2)x^2 + (a_1 + b_1)x + (a_0 + b_0).$$

The scalar multiplication of f by $\lambda \in F$ is given by

$$\lambda \odot f = \lambda a_d x^d + \cdots + \lambda a_2 x^2 + \lambda a_1 x + \lambda a_0.$$

For example, the real polynomials

$$f = 3x^5 + 2x^4 - x^2 + \sqrt{5}$$

and

$$g = -x^4 + 7x^3 + 3x^2 - \sqrt{2}x + 1$$

have degrees 5 and 4, respectively, and their sum is

$$f \oplus g = 3x^5 + x^4 + 7x^3 + 2x^2 - \sqrt{2}x + (1 + \sqrt{5}).$$

As before, we are merely using the notation $\oplus$ to distinguish it from the usual addition of two real numbers, but we will soon write $f + g$ for this sum again.

Anybody who can prove that the previous examples are vector spaces, will have no problems showing that $F[x]$ is a vector space as well.

Remark 2.14. We can multiply the polynomials $f = \sum_{i=0}^{d} a_i x^i$ and $g = \sum_{j=0}^{e} b_j x^j$ over F by expanding the product and using $x^i \cdot x^j = x^{i+j}$, which gives

$$f \cdot g = \sum_{k=0}^{d+e} \left(\sum_{\substack{i,j \\ i+j=k}} a_i b_j \right) x^k.$$

However, this multiplication is *not* part of the vector space structure on $F[x]$. Moreover, we can also define the derivative f' of a polynomial $f = \sum_{i=0}^{d} a_i x^i$ by $f' = \sum_{i=1}^{d} i a_i x^{i-1}$. (††) Note that while this reminds us of the derivative in analysis, we need to define this explicitly, as analysis does not make any sense for some fields, such as $\mathbb{F}_2$.

Example 2.15. (††) There are other examples that may appear stranger. Let X be any set, and let V be the set of all subsets of X. (For example, if $X = \{a, b\}$, then V has the four elements $\emptyset$, $\{a\}$, $\{b\}$, $\{a, b\}$.) We define addition on V as the *symmetric difference*: $A \oplus B = (A \setminus B) \cup (B \setminus A)$ (this is the set of elements of X that are in exactly one of A and B). We define scalar multiplication by elements of $\mathbb{F}_2$ in the only possible way: $0 \odot A = \emptyset$, $1 \odot A = A$. These operations turn V into an $\mathbb{F}_2$-vector space, with the empty set as zero.

To prove this assertion, we can check the vector space axioms (this is an instructive exercise). An alternative (and perhaps more elegant) way is to note that subsets of X correspond to maps $X \to \mathbb{F}_2$ (a map f corresponds to the subset $\{x \in X : f(x) = 1\}$) — there is a *bijection* between V and $\mathbb{F}_2^X$ — and this correspondence translates the addition and scalar multiplication we have defined on V into those we had defined on $\mathbb{F}_2^X$ in Example 2.11.

Exercises

2.2.1. Show that the set of magic squares, together with the addition and scalar multiplication defined in Example 2.7, is a real vector space.

2.2.2. Let A, B, C be the magic squares as in Example 2.7. Prove that for each 3×3 magic square X, there are real numbers λ, μ, ν such that $X = \lambda A + \mu B + \nu C$.

***2.2.3.** Let $n \geq 1$ be an integer.

(1) Show that there exists a finite number of $n \times n$ 'basic' magic squares, such that every $n \times n$ magic square is a sum of scalar multiples of these basic magic squares.
(2) How many basic squares do you need for $n = 4$?
(3) How many do you need for general n?

2.2.4. In Example 2.5, the first distributive law and the existence of negatives were proved for F^n. Show that the other six axioms for vector spaces hold for F^n as well, so that F^n is indeed a vector space over F.

2.2.5. Let X be the set of all your family members. We define two functions f, g from X to $\mathbb{R}$ (see Example 2.11). For every family member x, we let $f(x)$ be the year in which x was born, and we let $g(x)$ be the age of x (in years) today. Is the function $f \oplus g$ constant?

2.2.6. Finish the proof of the fact that $\mathrm{Map}(\mathbb{R}, \mathbb{R})$ is a vector space (see Example 2.10).

2.2.7. In Example 2.11, the first distributive law was proved for F^X. Show that the other seven axioms for vector spaces hold for F^X as well, so that F^X is indeed a vector space over F.

2.2.8. Prove that the set $F[x]$ of polynomials over F, together with addition, scalar multiplication, and the zero as defined in Example 2.13 is a vector space.

2.2.9. Given the field F and the set V in the following cases, together with the implicit element 0, are the described addition and scalar multiplication well defined, and if so, do they determine a vector space? If they are well defined, but they do not determine a vector space, then which rule is not satisfied?

(1) The field $F = \mathbb{R}$ and the set V of all functions $[0, 1] \to \mathbb{R}_{>0}$, together with the usual addition and scalar multiplication.
(2) (††) Example 2.15.
(3) The field $F = \mathbb{Q}$ and the set $V = \mathbb{R}$ with the usual addition and multiplication.
(4) The field $\mathbb{R}$ and the set V of all functions $f\colon \mathbb{R} \to \mathbb{R}$ with $f(3) = 0$, together with the usual addition and scalar multiplication.
(5) The field $\mathbb{R}$ and the set V of all functions $f\colon \mathbb{R} \to \mathbb{R}$ with $f(3) = 1$, together with the usual addition and scalar multiplication.
(6) Any field F together with the subset
$$\{(x, y, z) \in F^3 \ : \ x + 2y - z = 0\},$$
with coordinatewise addition and scalar multiplication.
(7) The field $F = \mathbb{R}$ together with the subset
$$\{(x, y, z) \in \mathbb{R}^3 \ : \ x - z = 1\},$$
with coordinatewise addition and scalar multiplication.

2.2.10. Let $a \in \mathbb{R}^n$ be a vector. Show that the set $a^\perp$ is a vector space.

2.2.11. (††) Suppose the set X contains exactly n elements. Then how many elements does the vector space $\mathbb{F}_2^X$ of functions $X \to \mathbb{F}_2$ consist of?

2.2.12. We can generalise Example 2.11 further. Let V be a vector space over F. Let X be any set and let $V^X = \mathrm{Map}(X, V)$ be the set of all functions $f\colon X \to V$. Define an addition and scalar multiplication on V^X that makes it into a vector space.

2.2.13. Let V be a vector space over F, and $\mathrm{Map}(V, V)$ the vector space of all functions from V to itself (see Exercise 2.2.12). Let id_V denote the identity map on V.

For every $p \in V$, we let $c_p\colon V \to V$ denote the constant map that sends every $v \in V$ to p, and we write $T_p = \mathrm{id}_V + c_p$.

(1) Show that for every $v \in V$, we have $T_p(v) = v + p$.
[This is why we call T_p 'translation by p'.]
(2) Show that for every $p, q \in V$, the composition $T_q \circ T_p$ equals T_{p+q}.
(3) Show that T_p is a bijection, with inverse T_{-p}.

2.2.14. Let S be the set of all infinite sequences $(a_n)_{n\geq 0}$ of real numbers satisfying the recurrence relation

$$a_{n+2} = a_{n+1} + a_n \quad \text{for all } n \geq 0.$$

An example of an element in S is the sequence

$$(a_0, a_1, a_2, a_3, a_4, a_5, a_6, a_7, \ldots) = (0, 1, 1, 2, 3, 5, 8, 13, \ldots)$$

of Fibonacci numbers. Show that the (term-wise) sum of two sequences from S is again in S and that any (term-wise) scalar multiple of a sequence from S is again in S. Finally show that S (with this addition and scalar multiplication) is a real vector space.

2.2.15. Let U and V be vector spaces over the same field F. Consider the Cartesian product

$$W = U \times V = \{\, (u, v) \;:\; u \in U,\, v \in V \,\}.$$

Define an addition and scalar multiplication on W that makes it into a vector space.

2.2.16. Set $V = \mathbb{R}_{\geq 0}$, the set of non-negative real numbers. Define the operation $\oplus$ on V by $x \oplus y = \max(x, y)$ for all $x, y \in V$, and define a scalar multiplication by $\lambda \odot x = x$ for all $\lambda \in \mathbb{R}$ and $x \in V$. Is V, together with these operations, and the element $0 \in V$, a vector space?

***2.2.17.** For each of the eight axioms in Definition 2.1, try to find a system $(V, 0, +, \cdot)$ that does not satisfy that axiom, while it does satisfy the other seven.

2.3. Basic properties

Before we can continue, we have to deal with a few little things. The fact that we talk about 'addition' and (scalar) 'multiplication' might tempt us to use more of the rules that hold for the traditional addition and multiplication than just the eight axioms given in Definition 2.1. We will show that many such rules do indeed follow from the basic eight. The first is a cancellation rule.

Lemma 2.16. *If three elements x, y, z of a vector space V satisfy $x \oplus z = y \oplus z$, then we have $x = y$.*

Proof. Suppose $x, y, z \in V$ satisfy $x \oplus z = y \oplus z$. By axiom (4) there is a $z' \in V$ with $z \oplus z' = 0$. Using such z' we get

$$x = x \oplus 0 = x \oplus (z \oplus z') = (x \oplus z) \oplus z' = (y \oplus z) \oplus z' = y \oplus (z \oplus z') = y \oplus 0 = y,$$

where we use axioms (3), (2), (2), and (3) for the first, third, fifth, and seventh equality respectively. So $x = y$. $\square$

It follows immediately that a vector space has only one zero element, as stated in the next remark.

Proposition 2.17. *In a vector space V, there is only one zero element: if two elements $0' \in V$ and $z \in V$ satisfy $0' \oplus z = z$, then $0' = 0$.*

Proof. Exercise. □

Because of Proposition 2.17, we often leave the zero vector implicit when defining a specific vector space. For instance, in Example 2.7 we could have just defined the addition and scalar multiplication of magic squares; for this to be a vector space, the only choice for the zero is the magic square consisting of only zeros.

Proposition 2.18. *In any vector space V, there is a unique negative for each element.*

Proof. The way to show that there is only one element with a given property is to assume there are two and then to show they are equal. Take $x \in V$ and assume that $a, b \in V$ are both negatives of x, that is, $x \oplus a = 0$ and $x \oplus b = 0$. Then by commutativity we have

$$a \oplus x = x \oplus a = 0 = x \oplus b = b \oplus x,$$

so $a = b$ by Lemma 2.16. □

Notation 2.19. Since negatives are unique, given $x \in V$ we may write $-x$ for the unique element that satisfies $x \oplus (-x) = 0$. Now we can also define a subtraction: we write $x \ominus y$ for $x \oplus (-y)$.

Note that if F is contained in $\mathbb{R}$, then the subtraction on F^n that we just got for free, coincides with the subtraction that we defined in Section 1.1.

Here are some more harmless facts.

Proposition 2.20. *Let $(V, 0_V, \oplus, \odot)$ be a vector space over F.*

(1) *For all $x \in V$, we have $0 \odot x = 0_V$.*
(2) *For all $x \in V$, we have $(-1) \odot x = -x$.*
(3) *For all $\lambda \in F$, we have $\lambda \odot 0_V = 0_V$.*
(4) *We have $-0_V = 0_V$.*
(5) *For all $\lambda \in F$ and $x \in V$ such that $\lambda \odot x = 0_V$, we have $\lambda = 0$ or $x = 0_V$.*
(6) *For all $\lambda \in F$ and $x \in V$, we have $-(\lambda \odot x) = \lambda \odot (-x) = (-\lambda) \odot x$.*
(7) *For all $x, y, z \in V$, we have $z = x \ominus y$ if and only if $x = y \oplus z$.*

Proof. We prove (1), (2), and (5), and leave the rest as an exercise.

(1) We have

$$(0 \odot x) \oplus 0_V = 0 \odot x = (0 + 0) \odot x = (0 \odot x) \oplus (0 \odot x)$$

with the equalities following from axiom (3), the fact that $0 = 0 + 0$ in F, and axiom (8), respectively. The Cancellation Lemma 2.16 implies $0_V = 0 \odot x$.

(2) It suffices to show that $(-1) \odot x$ satisfies the property that defines $-x$ uniquely, that is, it suffices to show $x \oplus ((-1) \odot x) = 0_V$. This follows from axioms (6) and (8), and property (1) of this proposition:

$$x \oplus ((-1) \odot x) = (1 \odot x) \oplus ((-1) \odot x) = (1 + (-1)) \odot x = 0 \odot x = 0_V.$$

(5) Suppose $\lambda \in F$ and $x \in V$ satisfy $\lambda \odot x = 0_V$. If $\lambda = 0$, then we are done, so we assume $\lambda \neq 0$ without loss of generality. Then λ has a

multiplicative inverse λ^{-1} in the field F. We find

$$x = 1 \odot x = (\lambda^{-1} \cdot \lambda) \odot x = \lambda^{-1} \odot (\lambda \odot x) = \lambda^{-1} \odot 0_V = 0_V,$$

where the equalities following from axiom (6), the fact that $\lambda^{-1} \cdot \lambda = 1$ in F, axiom (5), the hypothesis $\lambda \odot x = 0_V$, and property (3) of this proposition, respectively.

□

The axioms of Definition 2.1 and the properties that we just proved, show that the addition, scalar multiplication, and subtraction in a vector space behave just like the usual addition, multiplication, and subtraction, as long as we remember that the **scalar multiplication is a multiplication of a scalar with a vector, and *not* of two vectors!** Therefore, from now on, we will just use the usual notation: instead of $x \oplus y$ and $x \ominus y$ we write $x + y$ and $x - y$, and instead of $\lambda \odot x$ we write $\lambda \cdot x$ or even λx.

From the context it should always be clear what the symbols mean. Suppose, for example, that V is a general vector space over F. If x is an element of V, and we see the equality

$$0 \cdot x = 0,$$

then we know that the dot does not indicate the multiplication in F, so it stands for the scalar multiplication of V. Therefore, the first zero is the zero element of F. The scaled multiple $0 \cdot x$ is an element of V, so the second zero is the zero element of V.

As usual, and as in Section 1.1, scalar multiplication takes priority over addition and subtraction, so when we write $\lambda x \pm \mu y$ with $\lambda, \mu \in F$ and $x, y \in V$, we mean $(\lambda x) \pm (\mu y)$. Also as usual, when we have t vectors $x_1, x_2, \ldots, x_t \in V$, the expression $x_1 \pm x_2 \pm x_3 \pm \cdots \pm x_t$ should be read from left to right, so it stands for

$$\underbrace{(\ldots((}_{t-2} x_1 \pm x_2) \pm x_3) \pm \cdots) \pm x_t.$$

If all the signs in the expression are positive (+), then any other way of putting the parentheses would yield the same by the fact that the addition is associative (axiom (2)). The *sum* of t vectors $x_1, \ldots, x_t$ is $x_1 + x_2 + \cdots + x_t$.

Exercises

2.3.1. Prove Proposition 2.17.

2.3.2. Finish the proof of Proposition 2.20.

2.3.3. Is the following statement correct? "Axiom (4) of Definition 2.1 is redundant because we already know by Proposition 2.20(2) that for each vector $x \in V$ the vector $-x = (-1) \odot x$ is also contained in V."

2.3.4. Let $(V, 0_V, \oplus, \odot)$ be a real vector space and define $x \ominus y = x \oplus (-y)$, as usual. Which of the vector space axioms are satisfied and which are not (in general), for $(V, 0_V, \ominus, \odot)$? NOTE. You are expected to give proofs for the axioms that hold and to give counterexamples for those that do not hold.

CHAPTER 3

Subspaces

Recall that F is a field (see the beginning of Chapter 2).

3.1. Definition and examples

In many applications, we do not want to consider all elements of a given vector space V, but only the elements of a certain subset. Usually, it is desirable that this subset is again a vector space (with the addition and scalar multiplication it 'inherits' from V). In order for this to be possible, a minimal requirement certainly is that addition and scalar multiplication make sense on the subset. Also, the zero vector of V has to be contained in U. (Can you explain why the zero vector of V is forced to be the zero vector in U?)

Definition 3.1. Let V be an F-vector space. A subset $U \subset V$ is called a *vector subspace* or *linear subspace* of V if it has the following properties.

(1) $0 \in U$.
(2) If $u_1, u_2 \in U$, then $u_1 + u_2 \in U$ ('*U is closed under addition*').
(3) If $\lambda \in F$ and $u \in U$, then $\lambda u \in U$ ('*U is closed under scalar multiplication*').

Here the addition and scalar multiplication are those of V. Often we will just say *subspace* without the words *linear* or *vector*.

Note that, given the third property, the first is equivalent to saying that U is non-empty. Indeed, let $u \in U$, then by (3), we have $0 = 0 \cdot u \in U$. Note that here the first 0 denotes the zero vector, while the second 0 denotes the scalar 0.

We should justify the name 'subspace'.

Lemma 3.2. *Let $(V, +, \cdot, 0)$ be an F-vector space. If $U \subset V$ is a linear subspace of V, then $(U, +, \cdot, 0)$ is again an F-vector space*[1].

Proof. By definition of what a linear subspace is, we really have well-defined addition and scalar multiplication maps on U. It remains to check the axioms. For the axioms that state 'for all $\dots$, $\boxed{\dots}$' and do not involve any existence statements, this is clear, since they hold (by assumption) even for all elements of V, so certainly for all elements of U. This covers all axioms but axiom (4). For axiom (4), we need that for all $u \in U$ there is an element $u' \in U$ with $u + u' = 0$. In the vector space V there is a unique such an element, namely $u' = -u = (-1)u$ (see Proposition 2.18, Notation 2.19, and Proposition 2.20).

[1] The operators $+$ and $\cdot$ for V are functions from $V \times V$ and $F \times V$, respectively, to V. The operators for U, also denoted by $+$ and $\cdot$, are strictly speaking the restrictions $+|_{U\times U}$ and $\cdot|_{F\times U}$ to $U \times U$ and $F \times U$ of these operators for V, with the codomain restricted from V to U as well.

This element $u' = -u$ is contained in U by the third property of linear subspaces (take $\lambda = -1 \in F$). □

It is time for some examples.

Example 3.3. Let V be a vector space. Then $\{0\} \subset V$ and V itself are linear subspaces of V.

Example 3.4. Let $V \subset \mathbb{R}^3$ be the set of all triples (x_1, x_2, x_3) satisfying $x_1 + x_2 + x_3 = 0$. Clearly the zero vector $0 \in \mathbb{R}^3$ is contained in V. Suppose we have elements $x, y \in V$ and write them as $x = (x_1, x_2, x_3)$ and $y = (y_1, y_2, y_3)$. Then by definition of V we have $x_1 + x_2 + x_3 = 0 = y_1 + y_2 + y_3$. Hence, if we write the sum $z = x + y$ as $z = (z_1, z_2, z_3)$, then we have $z_i = x_i + y_i$ for $i \in \{1, 2, 3\}$, so we get

$$z_1+z_2+z_3 = (x_1+y_1)+(x_2+y_2)+(x_3+y_3) = (x_1+x_2+x_3)+(y_1+y_2+y_3) = 0+0 = 0.$$

This implies that $z = x + y$ is also contained in V. We leave it as an exercise to show that for any $\lambda \in \mathbb{R}$ and any $x \in V$, we also have $\lambda x \in V$. This means that the subset $V \subset \mathbb{R}^3$ satisfies all three requirements for being a subspace, so V is a linear subspace of $\mathbb{R}^3$. In Section 3.2 we will generalise this example.

Example 3.5. Consider $V = \mathbb{R}^2$ and, for $b \in \mathbb{R}$, set

$$U_b = \{(x, y) \in \mathbb{R}^2 : x + y = b\}.$$

For which b is U_b a linear subspace?

We check the first condition. We have $0 = (0, 0) \in U_b$ if and only if $0 + 0 = b$, so U_b can only be a linear subspace when $b = 0$. The question remains whether U_b is indeed a subspace for $b = 0$. Let us check the other properties for U_0.

If we have $(x_1, y_1), (x_2, y_2) \in U_0$, then $x_1 + y_1 = 0$ and $x_2 + y_2 = 0$, so $(x_1+x_2)+(y_1+y_2) = 0$. This implies $(x_1, y_1)+(x_2, y_2) = (x_1+x_2, y_1+y_2) \in U_0$. This shows that U_0 is closed under addition.

For each $\lambda \in \mathbb{R}$ and $(x, y) \in U_0$, we have $x + y = 0$, so $\lambda x + \lambda y = \lambda(x + y) = 0$. This implies $\lambda(x, y) = (\lambda x, \lambda y) \in U_0$. This shows that U_0 is also closed under scalar multiplication. We conclude that U_0 is indeed a subspace.

The following example is a generalisation of Example 3.5. The scalar product and Proposition 1.4 allow us to write everything much more efficiently.

Example 3.6. Given a nonzero vector $a \in \mathbb{R}^2$ and a constant $b \in \mathbb{R}$, let $L \subset \mathbb{R}^2$ be the line consisting of all points $v \in \mathbb{R}^2$ satisfying $\langle a, v\rangle = b$. We wonder when L is a subspace of $\mathbb{R}^2$. The requirement $0 \in L$ forces $b = 0$.

Conversely, assume $b = 0$. Then for two elements $v, w \in L$ we have

$$\langle a, v + w\rangle = \langle a, v\rangle + \langle a, w\rangle = 2b = 0,$$

so $v + w \in L$. Similarly, for any $\lambda \in \mathbb{R}$ and $v \in L$, we have

$$\langle a, \lambda v\rangle = \lambda\langle a, v\rangle = \lambda \cdot b = 0.$$

So L is a subspace if and only if $b = 0$.

Example 3.7. Let X be a set, and $x \in X$ an element. Consider the subset

$$U_x = \{f \in F^X \ : \ f(x) = 0\}$$

of the vector space $F^X = \mathrm{Map}(X, F)$. Clearly the zero function 0 is contained in U_x, as we have $0(x) = 0$. For any two functions $f, g \in U_x$ we have $f(x) = g(x) = 0$, so also $(f+g)(x) = f(x)+g(x) = 0$, which implies $f+g \in U_x$. For any $\lambda \in F$ and any $f \in U_x$ we have $(\lambda f)(x) = \lambda \cdot f(x) = \lambda \cdot 0 = 0$, which implies $\lambda f \in U_x$. We conclude that U_x is a subspace of the vector space $\mathrm{Map}(X, F)$ over F.

Example 3.8. Consider $\mathrm{Map}(\mathbb{R}, \mathbb{R}) = \mathbb{R}^{\mathbb{R}}$, the set of real-valued functions on $\mathbb{R}$. You will learn in Analysis that if f and g are continuous functions, then $f + g$ is again continuous, and λf is continuous for any $\lambda \in \mathbb{R}$. Of course, the zero function $x \mapsto 0$ is continuous as well. Hence, the set of all continuous functions

$$\mathcal{C}(\mathbb{R}) = \{f \in \mathrm{Map}(\mathbb{R}, \mathbb{R}) \ : \ f \text{ is continuous}\}$$

is a linear subspace of $\mathrm{Map}(\mathbb{R}, \mathbb{R})$.

Similarly, you will learn that sums and scalar multiples of differentiable functions are again differentiable. Also, derivatives respect sums and scalar multiplication: $(f + g)' = f' + g'$, $(\lambda f)' = \lambda f'$. From this, we conclude that

$$\mathcal{C}^n(\mathbb{R}) = \{f \in \mathrm{Map}(\mathbb{R}, \mathbb{R}) \ : \ f \text{ is } n \text{ times differentiable and } f^{(n)} \text{ is continuous}\}$$

is again a linear subspace of $\mathrm{Map}(\mathbb{R}, \mathbb{R})$.

In a different direction, consider the set of all *periodic* functions with period 1:

$$U = \{f \in \mathrm{Map}(\mathbb{R}, \mathbb{R}) \ : \ f(x + 1) = f(x) \text{ for all } x \in \mathbb{R}\} .$$

The zero function is certainly periodic. If f and g are periodic, then

$$(f + g)(x + 1) = f(x + 1) + g(x + 1) = f(x) + g(x) = (f + g)(x) ,$$

so $f+g$ is again periodic. Similarly, λf is periodic (for $\lambda \in \mathbb{R}$). So U is a linear subspace of $\mathrm{Map}(\mathbb{R}, \mathbb{R})$.

Exercises

3.1.1. Let $V \subset \mathbb{R}^3$ be the set of all triples (x_1, x_2, x_3) satisfying $x_1 + 2x_2 - 3x_3 = 0$. Show that V is a linear subspace of $\mathbb{R}^3$.

3.1.2. Let $U \subset \mathbb{R}^3$ be the set of all triples (x_1, x_2, x_3) satisfying $x_1 + 2x_2 - 3x_3 = 1$. Is U a linear subspace of $\mathbb{R}^3$?

3.1.3. Let W be the set of all 3×3 magic squares whose row, column, and diagonal sums are all equal to 0. Is W a vector space?

3.1.4. Given an integer $d \geq 0$, let $\mathbb{R}[x]_d$ denote the set of real polynomials of degree at most d. Show that the addition of two polynomials $f, g \in \mathbb{R}[x]_d$ satisfies $f + g \in \mathbb{R}[x]_d$. Show also that any scalar multiple of a polynomial $f \in \mathbb{R}[x]_d$ is contained in $\mathbb{R}[x]_d$. Prove that $\mathbb{R}[x]_d$ is a vector space.

3.1.5. Let X be a set with elements $x_1, x_2 \in X$. Show that the set

$$U = \{ f \in F^X \ : \ f(x_1) = 2f(x_2) \}$$

is a subspace of F^X.

3.1.6. Let X be the interval $[0, 1] \subset \mathbb{R}$. Is the set

$$U = \{ f \in \mathrm{Map}(X, \mathbb{R}) \ : \ f(0) = f(1)^2 \}$$

a subspace of $\mathrm{Map}(X, \mathbb{R})$?

3.1.7. Which of the following are linear subspaces of the vector space $\mathbb{R}^2$?

(1) $U_1 = \{(x, y) \in \mathbb{R}^2 : y = -\sqrt{e^{\pi}}x\}$,
(2) $U_2 = \{(x, y) \in \mathbb{R}^2 : y = x^2\}$,

(3) $U_3 = \{(x, y) \in \mathbb{R}^2 : xy = 0\}$.

3.1.8. Which of the following are linear subspaces of the vector space V of all functions from $\mathbb{R}$ to $\mathbb{R}$?

(1) $U_1 = \{f \in V \ : \ f \text{ is continuous}\}$
(2) $U_2 = \{f \in V \ : \ f(3) = 0\}$
(3) $U_3 = \{f \in V \ : \ f \text{ is continuous or } f(3) = 0\}$
(4) $U_4 = \{f \in V \ : \ f \text{ is continuous and } f(3) = 0\}$
(5) $U_5 = \{f \in V \ : \ f(0) = 3\}$
(6) $U_6 = \{f \in V \ : \ f(0) \geq 0\}$

3.1.9. Let X be a set.

(1) Show that the set $F^{(X)}$ of all functions $f : X \to F$ that satisfy $f(x) = 0$ for all but finitely many $x \in X$ is a subspace of the vector space F^X.
(2) More generally, let V be a vector space over F. Show that the set $V^{(X)}$ of all functions $f : X \to V$ that satisfy $f(x) = 0$ for all but finitely many $x \in X$ is a subspace of the vector space V^X (cf. Exercise 2.2.12).

3.1.10. Let X be a set.

(1) Let $U \subset F^X$ be the subset of all functions $X \to F$ whose image is finite. Show that U is a subspace of F^X that contains $F^{(X)}$ of Exercise 3.1.9.
(2) More generally, let V be a vector space over F. Show that the set of all functions $f : X \to V$ with finite image is a subspace of the vector space V^X that contains $V^{(X)}$ of Exercise 3.1.9.

3.2. The standard scalar product (again)

In Section 1.3 we defined the (standard) scalar product[2] for fields that are contained in $\mathbb{R}$. That section actually never used the fact that the field was contained in $\mathbb{R}$, so we can quickly restate the definitions and results in the generality that we are working in now[3]. For this section, we let n be a non-negative integer.

Definition 3.9. For any two vectors $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ in F^n we define the *standard scalar product* of x and y as

$$\langle x, y \rangle = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

As mentioned in Section 1.3, we will often leave out the word 'standard', and the scalar product may, in other books, be called the *dot product*, in which case it may get denoted by $x \cdot y$. Some books may call it the *(standard) inner product* for any field, but we will only use that phrase for fields contained in $\mathbb{R}$.

Example 3.10. (††) Suppose we have $z = (1, 0, 1, 1, 0, 1, 0)$ in $\mathbb{F}_2^7$. Then we get

$$\begin{aligned} \langle z, z \rangle &= 1 \cdot 1 + 0 \cdot 0 + 1 \cdot 1 + 1 \cdot 1 + 0 \cdot 0 + 1 \cdot 1 + 0 \cdot 0 \\ &= 1 + 0 + 1 + 1 + 0 + 1 + 0 = 0. \end{aligned}$$

[2]See footnote 1 on page 9.

[3]For those readers that are assuming that F is contained in $\mathbb{R}$ (see the beginning of Chapter 2), the only things new in this section are Proposition 3.14 and the identity (3.1).

Proposition 3.11. *Let $\lambda \in F$ be a scalar and let $x, y, z \in F^n$ be elements. Then the following identities hold.*

(1) $\langle x, y\rangle = \langle y, x\rangle$,
(2) $\langle \lambda x, y\rangle = \lambda \cdot \langle x, y\rangle = \langle x, \lambda y\rangle$,
(3) $\langle x, y + z\rangle = \langle x, y\rangle + \langle x, z\rangle$.
(4) $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$.
(5) $\langle x, y - z\rangle = \langle x, y\rangle - \langle x, z\rangle$.
(6) $\langle x - y, z\rangle = \langle x, z\rangle - \langle y, z\rangle$.

Proof. See Proposition 1.4 for the first three identities. The last three follow from the first three. □

We also generalise the notion of hyperplanes and lines to general fields.

Definition 3.12. A *hyperplane* in F^n is a subset $H \subset F^n$ for which there exist a nonzero vector $a \in F^n$ and a constant $b \in F$ with

$$H = \{ v \in F^n \ : \ \langle a, v\rangle = b \}.$$

Definition 3.13. A *line* in F^n is a subset $L \subset F^n$ for which there exist vectors $a, v \in F^n$ with v nonzero and with

$$L = \{ a + \lambda v \ : \ \lambda \in F \}.$$

In Exercise 3.2.4 we will see when two hyperplanes or two lines are equal.

Proposition 3.14. *Let $W \subset F^n$ be a line or a hyperplane. Then W is a subspace if and only if it contains the element* 0.

Proof. Exercise. □

Inspired by Chapter 1, we define the notion of orthogonality to general fields, even though for fields that are not contained in $\mathbb{R}$, it has nothing to do with any angle being 90 degrees (see Definition 1.21 and Warning 3.17).

Definition 3.15. We say that two vectors $v, w \in F^n$ are *orthogonal* [4] to each other when $\langle v, w\rangle = 0$; we then write $v \perp w$.

Of course, now we also generalise the notation $S^\perp$ of Definition 1.24 to general fields.

Definition 3.16. For any subset $S \subset F^n$, we let $S^\perp$ denote the set of those elements of F^n that are orthogonal to all elements of S, that is,

$$S^\perp = \{ x \in F^n \ : \ \langle s, x\rangle = 0 \text{ for all } s \in S \}.$$

For every element $a \in F^n$ we define $a^\perp$ as $\{a\}^\perp$.

If $a \in F^n$ is nonzero, then $a^\perp$ is a hyperplane containing 0. By definition, the set $S^\perp$ is the intersection of all subspaces $a^\perp$ with $a \in S$, that is,

$$S^\perp = \bigcap_{a \in S} a^\perp. \tag{3.1}$$

[4] We reserve the word "perpendicular" for fields that are contained in $\mathbb{R}$.

This description will be used in the next section to show that $S^\perp$ is a linear subspace of F^n for any $S \subset F^n$, though it is also a nice exercise to prove this directly.

Warning 3.17. (†) Proposition 1.14 states that the only vector in $\mathbb{R}^n$ that is orthogonal to itself is 0. Over other fields, however, we may have $\langle v, v\rangle = 0$ for nonzero v. For instance, the vector $a = (1, i) \in \mathbb{C}^2$ satisfies $\langle a, a\rangle = 0$. The fact that a is orthogonal to itself, means that a is contained in the hyperplane $a^\perp$! (††) Also the vectors $w = (1, 1) \in \mathbb{F}_2^2$ and $z \in \mathbb{F}_2^7$ of Example 3.10 are orthogonal to themselves.

Exercises

3.2.1. Prove that for any two distinct points $p, q \in F^n$, there is a unique line that contains both (cf. Proposition 1.12).

3.2.2. Let $S \subset F^n$ be a subset. Prove that $S^\perp$ is a linear subspace (cf. Lemma 1.25).

3.2.3. Prove Proposition 3.14.

3.2.4. (1) Let $a, a' \in F^n$ be two nonzero vectors and $b, b' \in F$ two constants. Show that the hyperplanes

$$H_{a,b} = \{\, v \in F^n \ : \ \langle a, v\rangle = b \,\} \qquad \text{and} \qquad H_{a',b'} = \{\, v \in F^n \ : \ \langle a', v\rangle = b' \,\}$$

are equal if and only if there exists a nonzero $\lambda \in F$ such that $a' = \lambda a$ and $b' = \lambda b$.

(2) Let $a, a', v, v' \in F^n$ be vectors with v, v' nonzero. Show that the lines

$$L_{a,v} = \{\, a + \lambda v \ : \ \lambda \in F \,\} \qquad \text{and} \qquad L_{a',v'} = \{\, a' + \lambda v' \ : \ \lambda \in F \,\}$$

are equal if and only if we have $a' \in L$ and there exists a nonzero $\lambda \in F$ such that $v' = \lambda v$.

3.2.5. Let $a_1, \ldots, a_t \in F^n$ be vectors and $b_1, \ldots, b_t \in F$ constants. Let $V \subset F^n$ be the subset

$$V = \{x \in F^n \ : \ \langle a_1, x\rangle = b_1, \ \ldots, \ \langle a_t, x\rangle = b_t\}.$$

Show that with the same addition and scalar multiplication as F^n, the set V is a vector space if and only if $b_1 = \ldots = b_t = 0$.

3.3. Intersections

The following result can be used, for example, to show that, with U and $\mathcal{C}(\mathbb{R})$ as in Example 3.8, the intersection $U \cap \mathcal{C}(\mathbb{R})$ of all continuous periodic functions from $\mathbb{R}$ to $\mathbb{R}$ is again a linear subspace.

Lemma 3.18. *Let V be an F-vector space, and $U_1, U_2 \subset V$ linear subspaces of V. Then the intersection $U_1 \cap U_2$ is again a linear subspace of V.*
More generally, if $(U_i)_{i\in I}$ (with $I \neq \emptyset$) is any family of linear subspaces of V, then their intersection $U = \bigcap_{i\in I} U_i$ is again a linear subspace of V.

Proof. It is sufficient to prove the second statement (take $I = \{1, 2\}$ to obtain the first). We check the conditions.

(1) By assumption $0 \in U_i$ for all $i \in I$. So $0 \in U$.
(2) Let $x, y \in U$. Then $x, y \in U_i$ for all $i \in I$. Hence (since U_i is a subspace by assumption) $x + y \in U_i$ for all $i \in I$. But this means $x + y \in U$.

(3) Let $\lambda \in F$, $x \in U$. Then $x \in U_i$ for all $i \in I$. Hence (since U_i is a subspace by assumption) $\lambda x \in U_i$ for all $i \in I$. This means that $\lambda x \in U$.

We conclude that U is indeed a linear subspace. $\square$

Example 3.19. Consider the subspace $\mathcal{C}^2(\mathbb{R}) \subset \mathrm{Map}(\mathbb{R}, \mathbb{R})$ of all functions f from $\mathbb{R}$ to $\mathbb{R}$ that are twice differentiable and for which the second derivative f'' is continuous (see Example 3.8). Consider the sets

$$U = \{f \in \mathcal{C}^2(\mathbb{R}) \ : \ f'' = -f\}$$

and

$$V = \{f \in \mathrm{Map}(\mathbb{R}, \mathbb{R}) \ : \ f(0) = 0\}.$$

Since derivatives respect addition, we find that for all functions $f, g \in U$ we have

$$(f + g)'' = f'' + g'' = (-f) + (-g) = -(f + g),$$

so we obtain $f + g \in U$. Similarly, for any $\lambda \in \mathbb{R}$ and $f \in U$, we have $\lambda f \in U$. Since we also have $0 \in U$, we find that the set U of solutions to the *differential equation* $f'' = -f$ is a subspace of $\mathcal{C}^2(\mathbb{R})$. It contains, for example, the functions sine, cosine, and their sum. By Example 3.7, the set V is also a linear subspace, so by Lemma 3.18, the intersection $U \cap V$ is also a linear subspace of $\mathrm{Map}(\mathbb{R}, \mathbb{R})$. It is the set of solutions to the system of functional equations

$$f'' = -f \qquad \text{and} \qquad f(0) = 0.$$

The following proposition is a generalisation of Lemma 1.25 to all fields.

Proposition 3.20. *Let n be a non-negative integer, and $S \subset F^n$ a subset. Then $S^\perp$ is a linear subspace of F^n.*

Proof. We use the identity (3.1). For each $a \in S$, the hyperplane $a^\perp \subset F^n$ contains 0, so it is a subspace by Proposition 3.14. By Lemma 3.18 (with the index set I equal to S), the intersection $\bigcap_{a \in S} a^\perp$ is also a linear subspace. This intersection equals $S^\perp$ by (3.1). $\square$

Note that in general, if U_1 and U_2 are linear subspaces, then the union $U_1 \cup U_2$ is not (it is if and only if one of the two subspaces is contained in the other — exercise!).

Example 3.21. Consider the subspaces

$$U_1 = \{(x, 0) \in \mathbb{R}^2 \ : \ x \in \mathbb{R}\}, \qquad U_2 = \{(0, x) \in \mathbb{R}^2 \ : \ x \in \mathbb{R}\}.$$

The union $U = U_1 \cup U_2$ is not a subspace because the elements $u_1 = (1, 0)$ and $u_2 = (0, 1)$ are both contained in U, but their sum $u_1 + u_2 = (1, 1)$ is not.

Exercises

3.3.1. Suppose that U_1 and U_2 are linear subspaces of a vector space V. Show that $U_1 \cup U_2$ is a subspace of V if and only if $U_1 \subset U_2$ or $U_2 \subset U_1$.

3.3.2. Let H_1, H_2, H_3 be hyperplanes in $\mathbb{R}^3$ given by the equations

$$\langle (1, 0, 1), v\rangle = 2, \qquad \langle (-1, 2, 1), v\rangle = 0, \qquad \langle (1, 1, 1), v\rangle = 3,$$

respectively.

(1) Which of these hyperplanes is a subspace of $\mathbb{R}^3$?

(2) Show that the intersection $H_1 \cap H_2 \cap H_3$ contains exactly one element.

3.3.3. Give an example of a vector space V with two subsets U_1 and U_2, such that U_1 and U_2 are **not** subspaces of V, but their intersection $U_1 \cap U_2$ is.

3.3.4. Let n be a positive integer and let M denote the set of all magic $n \times n$ squares, that is, squares of $n \times n$ real numbers of which the n row sums, the n column sums, and the two diagonal sums are all equal. Let P denote the set of all n^2 positions in an $n \times n$ square.

(1) Show that M is a vector space over $\mathbb{R}$ with the position-wise addition and scalar multiplication.
(2) Suppose $p \in P$ is a position. Show that the set of magic squares with a 0 on position p is a subspace of M.
(3) Suppose $S \subset P$ is a subset. Show that the set of magic squares with a 0 on position p for each $p \in S$, is a subspace of M.

3.4. Linear hulls, linear combinations, and generators

Given a set S of vectors in a vector space V, we want to understand the smallest subspace of V that contains S. Let us look at a specific case first.

Example 3.22. Let V be a vector space over F, and let $v_1, v_2 \in V$ be two vectors. Suppose that W is any subspace of V that contains v_1 and v_2. According to the definition of linear subspaces, all scalar multiples of v_1 and v_2, and sums thereof are contained in W as well. This implies that every element of the form $\lambda_1 v_1 + \lambda_2 v_2$, with $\lambda_1, \lambda_2 \in F$, is contained in W. So for the set

$$U = \{\lambda_1 v_1 + \lambda_2 v_2 : \lambda_1, \lambda_2 \in F\}$$

we have $U \subset W$. On the other hand, U is itself a linear subspace:

(1) $0 = 0 \cdot v_1 + 0 \cdot v_2 \in U$,
(2) $(\lambda_1 v_1 + \lambda_2 v_2) + (\mu_1 v_1 + \mu_2 v_2) = (\lambda_1 + \mu_1)v_1 + (\lambda_2 + \mu_2)v_2 \in U$,
(3) $\lambda(\lambda_1 v_1 + \lambda_2 v_2) = (\lambda\lambda_1)v_1 + (\lambda\lambda_2)v_2 \in U$.

(Exercise: which of the vector space axioms have we used where?)
Therefore, U is the smallest linear subspace of V containing v_1 and v_2 in the following sense: U is a subspace containing v_1 and v_2, and every subspace $W \subset V$ containing v_1 and v_2 contains U.

This observation generalises.

Definition 3.23. Let V be an F-vector space with t elements $v_1, v_2, \ldots, v_t \in V$. The *linear combination* (or, more precisely, *F-linear combination*) of $v_1, v_2, \ldots, v_t$ with *coefficients* $\lambda_1, \lambda_2, \ldots, \lambda_t \in F$ is the element

$$v = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_t v_t.$$

If $t = 0$, then the only linear combination of no vectors is (by definition) $0 \in V$. If $S \subset V$ is any (possibly infinite) subset, then an *(F-)linear combination of S* is a linear combination of *finitely many* elements of S.

Definition 3.24. Let V be a vector space over F. If S is a subset of V, then $L(S)$ is the set of all linear combinations on S. If we want to indicate the field F of scalars, we write $L_F(S)$. For finitely many elements $v_1, v_2, \ldots, v_t \in V$, we also write $L(v_1, v_2, \ldots, v_t)$ instead of $L(\{v_1, v_2, \ldots, v_t\})$.

Remark 3.25. The set $L(v_1, v_2, \ldots, v_t)$ is defined as the set of all linear combinations on the set $S = \{v_1, v_2, \ldots, v_t\}$. It is true that this equals the set of all linear combinations of $v_1, v_2, \ldots, v_t$, but this relies on two subtleties. First of all, if some of the t vectors are equal, then the set S has fewer than t elements. Nonetheless, a linear combination of $v_1, v_2, \ldots, v_t$ is still a linear combination on S, as we can combine terms: if $v_i = v_j$, then $\lambda_i v_i + \lambda_j v_j = (\lambda_i + \lambda_j)v_i$. Second, the converse is also true. A linear combination on S may a priori not use all t vectors, but it is still also a linear combination of all $v_1, v_2, \ldots, v_t$, as we can just add coefficients zero for the vectors that were not used.

The linear combinations of one vector $a \in V$ are exactly its scalar multiples, so $L(a)$ is the set $\{\lambda a : \lambda \in F\}$ of all scalar multiples of a. Note that this is consistent with Notation 1.13.

Proposition 3.26. *Let V be a vector space with t elements $v_1, v_2, \ldots, v_t \in V$. Then the set $L(v_1, v_2, \ldots, v_t)$ is a linear subspace of V. More generally, let $S \subset V$ be a subset. Then $L(S)$ is a linear subspace of V.*

Proof. We start with the first statement. Write $U = L(v_1, v_2, \ldots, v_t)$. First of all, we have $0 \in U$, since $0 = 0v_1 + 0v_2 + \cdots + 0v_t$ (this even works for $t = 0$). To check that U is closed under addition, let $v = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_t v_t$ and $w = \mu_1 v_1 + \mu_2 v_2 + \cdots + \mu_t v_t$ be two elements of U. Then

$$\begin{aligned} v + w &= (\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_t v_t) + (\mu_1 v_1 + \mu_2 v_2 + \cdots + \mu_t v_t) \\ &= (\lambda_1 + \mu_1)v_1 + (\lambda_2 + \mu_2)v_2 + \cdots + (\lambda_t + \mu_t)v_t \end{aligned}$$

is again a linear combination of $v_1, v_2, \ldots, v_t$, so $v + w \in U$. Also, for $\lambda \in F$, the element

$$\begin{aligned} \lambda v &= \lambda(\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_t v_t) \\ &= (\lambda\lambda_1)v_1 + (\lambda\lambda_2)v_2 + \cdots + (\lambda\lambda_t)v_t \end{aligned}$$

is a linear combination of $v_1, v_2, \ldots, v_t$, so $\lambda v \in U$. We conclude that U is indeed a linear subspace of V.

For the general case, the only possible problem is with checking that the set of linear combinations on S is closed under addition, because two linear combinations might not be linear combinations of the same elements. For this, we observe that if v is a linear combination on the finite subset I of S and w is a linear combination on the finite subset J of S, then v and w can both be considered as linear combinations on the finite subset $I \cup J$ of S (just add coefficients zero); now our argument above applies. $\square$

For any subset S of a vector space V, the subspace $L(S)$ is called the *linear hull* or *linear span* of S, or the linear subspace *generated by* S. If $L(S) = V$, then we say that S is a *generating set* for V or that S *generates* V, or that the elements of S *generate* V. If V can be generated by a finite set S, then we say that V is *finitely generated*.

Be aware that, besides $L(S)$, there are various different notations for linear hulls in the literature, for example $\mathrm{Span}(S)$ or $\langle S \rangle$ (which in LaTeX is *not* written $<S>$, but $\langle S \rangle$!).

Example 3.27. Take the three vectors

$$e_1 = (1, 0, 0), \qquad e_2 = (0, 1, 0), \qquad \text{and} \qquad e_3 = (0, 0, 1)$$

in $\mathbb{R}^3$. Then for every vector $x = (x_1, x_2, x_3) \in \mathbb{R}^3$ we have $x = x_1e_1+x_2e_2+x_3e_3$, so every element in $\mathbb{R}^3$ is a linear combination of e_1, e_2, e_3. We conclude that $\mathbb{R}^3$ is contained in $L(e_1, e_2, e_3)$ and therefore $L(e_1, e_2, e_3) = \mathbb{R}^3$, so the set $\{e_1, e_2, e_3\}$ generates $\mathbb{R}^3$.

Definition 3.28. Let n be a positive integer. The *standard generators* of F^n are

$$\begin{aligned} e_1 &= (1, 0, 0, \dots, 0), \\ e_2 &= (0, 1, 0, \dots, 0), \\ &\vdots \\ e_i &= (0, 0, \dots, 0, 1, 0, \dots, 0), \\ &\vdots \\ e_n &= (0, 0, \dots, 0, 1), \end{aligned}$$

with e_i the vector in F^n whose i-th entry equals 1 while all other entries equal 0.

For every vector $x = (x_1, x_2, \dots, x_n) \in F^n$ we have $x = x_1e_1 + x_2e_2 + \cdots + x_ne_n$, so x is a linear combination of $e_1, e_2, \dots, e_n$. Therefore, as in the previous example, we find $L(E) = F^n$, so the set $E = \{e_1, e_2, \dots, e_n\}$ generates F^n, thus explaining the name *standard generators*.

Since every vector is a linear combination of itself ($v = 1 \cdot v$), it is clear that $L(S)$ contains S for every subset S of any vector space. The following lemma shows that $L(S)$ is the smallest linear subspace containing S.

Lemma 3.29. *Let V be an F-vector space and S a subset of V. Let U be any subspace of V that contains S. Then we have $L(S) \subset U$.*

Proof. Since U is a linear subspace that contains S, it also contains all scalar multiples of elements in S, as well as sums thereof. Hence, U contains all linear combinations on S, so we have $L(S) \subset U$. □

If U is a certain subspace of a vector space V, and we wish to show that U equals V, then, by Lemma 3.29, it suffices to show that U contains a generating set S for V.

Example 3.30. Consider the vectors

$$v_1 = (1, 0, 3), \quad v_2 = (0, 1, 2), \quad v_3 = (0, 0, 1)$$

in $\mathbb{R}^3$, and set $U = L(v_1, v_2, v_3)$. We wonder whether v_1, v_2, and v_3 generate $\mathbb{R}^3$, that is, whether $U = \mathbb{R}^3$. This is obviously equivalent to the question whether the inclusion $\mathbb{R}^3 \subset U$ holds. By Lemma 3.29, this is the case if and only if the standard generators e_1, e_2, e_3 are contained in U. Indeed, they are linear combinations of v_1, v_2, and v_3, as we have

$$e_1 = v_1 - 3v_3, \qquad e_2 = v_2 - 2v_3, \qquad \text{and} \qquad e_2 = v_3.$$

We conclude that v_1, v_2, and v_3 do indeed generate $\mathbb{R}^3$.

Example 3.31. Take $V = \mathbb{R}^4$ and consider $S = \{v_1, v_2, v_3\}$ with

$$v_1 = (1, 0, 1, 0), \qquad v_2 = (0, 1, 0, 1), \qquad v_3 = (1, 1, 1, 1).$$

For $a_1 = (1, 0, -1, 0)$ and $a_2 = (0, 1, 0, -1)$, the hyperplanes

$$H_1 = \{x \in \mathbb{R}^n \; : \; \langle x, a_1 \rangle = 0\}, \qquad \text{and} \qquad H_2 = \{x \in \mathbb{R}^n \; : \; \langle x, a_2 \rangle = 0\}$$

are subspaces (see Proposition 3.14) that both contain v_1, v_2, v_3. So certainly we have an inclusion $L(v_1, v_2, v_3) \subset H_1 \cap H_2 = \{a_1, a_2\}^\perp$.

Conversely, every element $x = (x_1, x_2, x_3, x_4)$ in the intersection $H_1 \cap H_2$ satisfies $\langle x, a_1 \rangle = 0$, so $x_1 = x_3$ and $\langle x, a_2 \rangle = 0$, so $x_2 = x_4$, which implies $x = x_1 v_1 + x_2 v_2$. We conclude $H_1 \cap H_2 \subset L(v_1, v_2)$, so we have

$$L(v_1, v_2, v_3) \subset H_1 \cap H_2 \subset L(v_1, v_2) \subset L(v_1, v_2, v_3).$$

As the first subspace equals the last, all these inclusions are equalities. We deduce the equality $L(S) = H_1 \cap H_2$, so S generates the intersection $H_1 \cap H_2$. In fact, we see that we do not need v_3, as also $\{v_1, v_2\}$ generates $H_1 \cap H_2$. In Section 8.3 we will see how to compute generators of intersections more systematically.

Lemma 3.32. *Let V be a vector space and S, T subsets of V satisfying $T \subset L(S)$ and $S \subset L(T)$. Then we have $L(S) = L(T)$.*

Proof. Applying Lemma 3.29 to S and $U = L(T)$, we obtain $L(S) \subset L(T)$. By symmetry we also have $L(T) \subset L(S)$, so we find $L(S) = L(T)$. ☐

In Proposition 3.20 we have seen that for any set $S \subset F^n$, the set $S^\perp$ is a linear subspace. The following proposition states a few more properties of $S^\perp$.

Proposition 3.33. *Let $n \geq 0$ be an integer, and S a subset of F^n. Then the following statements hold.*

(1) *For any subset $T \subset S$ we have $S^\perp \subset T^\perp$.*
(2) *We have $S^\perp = L(S)^\perp$.*
(3) *We have $L(S) \subset (S^\perp)^\perp$.*
(4) *For any subset $T \subset F^n$ we have $S^\perp \cap T^\perp = (S \cup T)^\perp$.*

Proof. We leave (1), (3), and (4) as an exercise to the reader. To prove (2), note that from $S \subset L(S)$ and (1) we have $L(S)^\perp \subset S^\perp$, so it suffices to prove the opposite inclusion. Suppose we have $x \in S^\perp$, so that $\langle s, x \rangle = 0$ for all $s \in S$. Now any element $t \in L(S)$ is a linear combination of elements in S, so there are elements $s_1, s_2, \ldots, s_n \in S$ and scalars $\lambda_1, \lambda_2, \ldots, \lambda_n \in F$ such that we have $t = \lambda_1 s_1 + \cdots + \lambda_n s_n$, which implies

$$\langle t, x \rangle = \langle \lambda_1 s_1 + \cdots + \lambda_n s_n, x \rangle = \lambda_1 \langle s_1, x \rangle + \cdots + \lambda_n \langle s_n, x \rangle = \lambda_1 \cdot 0 + \cdots + \lambda_n \cdot 0 = 0.$$

We conclude that we have $x \in L(S)^\perp$. ☐

Remark 3.34. Later we will see that the inclusion $L(S) \subset (S^\perp)^\perp$ of Proposition 3.33 is in fact an equality, so that for every subspace U we have $(U^\perp)^\perp = U$. See Proposition 8.20 and Exercise 8.2.4.

We finish this section with the vector space of polynomial functions.

Example 3.35. Let $F \subset \mathbb{C}$ be a field. Inside the vector space F^F of all functions from F to F. we consider the power functions $p_n \colon x \mapsto x^n$ Their linear hull $L(\{p_n : n \in \mathbb{Z}_{\geq 0}\}) \subset F^F$ is the linear subspace of *polynomial functions* from F to F, i.e, functions that are of the form

$$x \longmapsto a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

with $n \in \mathbb{Z}_{\geq 0}$ and $a_0, a_1, \ldots, a_n \in F$. By definition, the power functions p_n generate the subspace of polynomial functions, which we denote by $P(F)$. For $F = \mathbb{R}$, this subspace $P(\mathbb{R})$ is contained in the subspace $\mathcal{C}(\mathbb{R})$ of continuous functions.

Remark 3.36. Let $F \subset \mathbb{C}$ be a field. In Example 2.13 we defined *polynomials* over F as formal sums. These are a priori not the same as *polynomial functions*, but to any such polynomial $\sum_{n=0}^{d} a_n x^n$ we can associate the polynomial function that sends $x_0 \in F$ to $\sum_{n=0}^{d} a_n x_0^n$. This gives a map

$$\varphi\colon F[x] \to P(F) \subset F^F$$

that is clearly surjective. It is also injective, which follows from the theorem that a nonzero polynomial over F can not have more zeroes than its degree (see Exercise 11.3.8). Hence, there is a natural bijection between polynomials and polynomial functions. Under this bijection, also their derivatives, defined for polynomial functions in terms of the usual limits, and for abstract polynomials by Remark 2.14, coincide, so the difference between polynomials and polynomial functions will not cause any confusion over subfields of $\mathbb{C}$. In fact, By abuse of notation, the function $\varphi(f)$ is often also denoted by f. In Appendix D, we also define polynomial functions over general fields. In that context one should be careful, as the map above need not be injective.

Exercises

3.4.1. Prove Proposition 3.33.

3.4.2. Do the vectors

$$(1, 0, -1), \qquad (2, 1, 1), \qquad \text{and} \qquad (1, 0, 1)$$

generate $\mathbb{R}^3$?

3.4.3. Do the vectors

$$(1, 2, 3), \qquad (4, 5, 6), \qquad \text{and} \qquad (7, 8, 9)$$

generate $\mathbb{R}^3$?

3.4.4. Let $U \subset \mathbb{R}^4$ be the subspace generated by the vectors

$$(1, 2, 3, 4), \qquad (5, 6, 7, 8), \qquad \text{and} \qquad (9, 10, 11, 12).$$

What is the minimum number of vectors needed to generate U? As always, prove that your answer is correct.

3.4.5. Let X be a set. Consider the subspace $F^{(X)}$ of F^X consisting of all functions $f\colon X \to F$ that satisfy $f(x) = 0$ for all but finitely many $x \in X$ (cf. Exercise 3.1.9). For every $x \in X$ we define the function $e_x\colon X \to F$ by

$$e_x(z) = \begin{cases} 1 & \text{if } z = x, \\ 0 & \text{otherwise.} \end{cases}$$

Show that the set $\{e_x \ : \ x \in X\}$ generates $F^{(X)}$.

3.4.6. Does the equality $L(I \cap J) = L(I) \cap L(J)$ hold for all vector spaces V and subsets I and J of V?

3.4.7. We say that a function $f\colon \mathbb{R} \to \mathbb{R}$ is *even* if $f(-x) = f(x)$ for all $x \in \mathbb{R}$, and *odd* if $f(-x) = -f(x)$ for all $x \in \mathbb{R}$.

(1) Is the subset of $\mathbb{R}^\mathbb{R}$ consisting of all even functions a linear subspace?
(2) Is the subset of $\mathbb{R}^\mathbb{R}$ consisting of all odd functions a linear subspace?

3.4.8. Let V be a vector space and $S, T \subset V$ subsets. Show that the inclusion $L(S) \subset L(S \cup T)$ holds and that we have equality if and only if $T \subset L(S)$.

3.4.9. Let V be a vector space over F, containing vectors $v_1, v_2, \ldots, v_n \in V$. Set $W = L(v_1, v_2, \ldots, v_n)$. Using Lemma 3.32, give short proofs of the following equalities of subspaces.

(1) $W = L(v'_1, \ldots, v'_n)$ where for some fixed j and some nonzero scalar $\lambda \in F$ we have $v'_i = v_i$ for $i \neq j$ and $v'_j = \lambda v_j$ (the j-th vector is scaled by a nonzero factor λ).

(2) $W = L(v'_1, \ldots, v'_n)$ where for some fixed j, k with $j \neq k$ and some scalar $\lambda \in F$ we have $v'_i = v_i$ for $i \neq k$ and $v'_k = v_k + \lambda v_j$ (a scalar multiple of v_j is added to v_k).

(3) $W = L(v'_1, \ldots, v'_n)$ where for some fixed j and k we set $v'_i = v_i$ for $i \neq j, k$ and $v'_j = v_k$ and $v'_k = v_j$ (the elements v_j and v_k are switched),

3.4.10. Let V be an F-vector space and S a subset of V. Show that we have

$$L(S) = \bigcap \{U \subset V : U \text{ linear subspace of } V \text{ and } S \subset U\} .$$

[Note that the notation in this proposition means the intersection of all elements of the specified set: we intersect all linear subspaces containing S.]
[Note that in the extreme case $S = \emptyset$, we have to intersect *all* linear subspaces of V, so the above reduces to the (correct) statement $L(\emptyset) = \{0\}$.]

3.5. Sums of subspaces

We have seen that the intersection of linear subspaces is again a linear subspace, but the union usually is not, see Example 3.21. However, it is very useful to have a replacement for the union that has similar properties, but is a linear subspace. Note that the union of two (or more) sets is the smallest set that contains both (or all) of them. From this point of view, it is natural in the context of vector spaces to study the smallest subspace containing two given subspaces, which is the subspace generated by the union.

Definition 3.37. Let V be a vector space, $U_1, U_2 \subset V$ two linear subspaces. The *sum* of U_1 and U_2 is the linear subspace generated by $U_1 \cup U_2$:

$$U_1 + U_2 = L(U_1 \cup U_2) .$$

More generally, if $(U_i)_{i \in I}$ is a family of subspaces of V ($I = \emptyset$ is allowed here), then their *sum* is again

$$\sum_{i \in I} U_i = L\Big(\bigcup_{i \in I} U_i\Big) .$$

We want a more explicit description of these sums.

Lemma 3.38. *If U_1 and U_2 are linear subspaces of the vector space V, then we have*

$$U_1 + U_2 = \{u_1 + u_2 : u_1 \in U_1, u_2 \in U_2\} .$$

If $(U_i)_{i \in I}$ is a family of linear subspaces of V, then we have

$$\sum_{i \in I} U_i = \Big\{ \sum_{j \in J} u_j : J \subset I \text{ finite and } u_j \in U_j \text{ for all } j \in J \Big\} .$$

Proof. For each equality, it is clear that the set on the right-hand side is contained in the left-hand side (which is closed under addition). For the opposite inclusions, it suffices by Lemma 3.29 (applied with S equal to the unions $U_1 \cup U_2$ and $\bigcup_{i\in I} U_i$, respectively, which are obviously contained in the appropriate right-hand side) to show that the right-hand sides are linear subspaces.

We have $0 = 0 + 0$ (and $0 = \sum_{j\in\emptyset} u_j$), so 0 is an element of the right-hand side sets. Closure under scalar multiplication is easy to see. Indeed, for $u_1 \in U_1$ and $u_2 \in U_2$, we have

$$\lambda(u_1 + u_2) = \lambda u_1 + \lambda u_2\,,$$

and we have $\lambda u_1 \in U_1$, $\lambda u_2 \in U_2$, because U_1, U_2 are linear subspaces; hence, the element $\lambda(u_1+u_2)$ is also contained the right-hand side of the first equality of the lemma. Similarly, for every finite subset $J \subset I$ and elements $u_j \in U_j$ for each $j \in J$, we have

$$\lambda \sum_{j\in J} u_j = \sum_{j\in J} \lambda u_j\,,$$

and $\lambda u_j \in U_j$, since U_j is a linear subspace; hence, the element $\lambda \sum_{j\in J} u_j$ is also contained in the right-hand side of the second equality.

Finally, for $u_1, u_1' \in U_1$ and $u_2, u_2' \in U_2$, we have

$$(u_1 + u_2) + (u_1' + u_2') = (u_1 + u_1') + (u_2 + u_2')$$

with $u_1 + u_1' \in U_1$, $u_2 + u_2' \in U_2$. And for J_1, J_2 finite subsets of I, $u_j \in U_j$ for $j \in J_1$, $u_j' \in U_j$ for $j \in J_2$, we find

$$\Big(\sum_{j\in J_1} u_j\Big) + \Big(\sum_{j\in J_2} u_j'\Big) = \sum_{j\in J_1\cup J_2} v_j\,,$$

where we use $v_j = u_j \in U_j$ if $j \in J_1 \setminus J_2$, while $v_j = u_j' \in U_j$ if $j \in J_2 \setminus J_1$, and $v_j = u_j + u_j' \in U_j$ if $j \in J_1 \cap J_2$. This shows that the right-hand sides are also closed under addition, which implies that they are indeed subspaces. □

Alternative proof. Clearly the right-hand side is contained in the left-hand side, so it suffices to prove the opposite inclusions by showing that any linear combination of elements in the unions $U_1 \cup U_2$ and $\bigcup_{i\in I} U_i$, respectively, is contained in the appropriate right-hand side.

Suppose we have $v = \lambda_1 w_1 + \cdots + \lambda_s w_s$ with $w_i \in U_1 \cup U_2$. Then after reordering we may assume that for some non-negative integer $r \leq s$ we have $w_1, \ldots, w_r \in U_1$ and $w_{r+1}, \ldots, w_s \in U_2$. Then for $u_1 = \lambda_1 w_1 + \cdots + \lambda_r w_r \in U_1$ and $u_2 = \lambda_{r+1} w_{r+1} + \cdots + \lambda_s w_s \in U_2$ we have $v = u_1 + u_2$, as required.

Suppose we have $v = \lambda_1 w_1 + \cdots + \lambda_s w_s$ with $w_k \in \bigcup_{i\in I} U_i$ for each $1 \leq k \leq s$. Since the sum is finite, there is a finite subset $J \subset I$ such that $w_k \in \bigcup_{j\in J} U_j$ for each $1 \leq k \leq s$. After collecting those elements contained in the same subspace U_j together, we may write v as

$$v = \sum_{j\in J} \sum_{k=1}^{r_j} \lambda_{jk} w_{jk}$$

for scalars λ_{jk} and elements $w_{jk} \in U_j$. Then for $u_j = \sum_{k=1}^{r_j} \lambda_{jk} w_{jk} \in U_j$ we have $v = \sum_{j\in J} u_j$, as required. □

Example 3.39. The union $U_1 \cup U_2$ of Example 3.21 contains the vectors $e_1 = (1,0)$ and $e_2 = (0,1)$, so the sum $U_1 + U_2 = L(U_1 \cup U_2)$ contains $L(e_1, e_2) = \mathbb{R}^2$ and we conclude $U_1 + U_2 = \mathbb{R}^2$.

Example 3.40. Let V be the subspace of $\mathrm{Map}(\mathbb{R}, \mathbb{R})$ consisting of all continuous functions from $\mathbb{R}$ to $\mathbb{R}$. Set

$$U_0 = \{f \in V \ : \ f(0) = 0\}, \qquad U_1 = \{f \in V \ : \ f(1) = 0\}.$$

We now prove $U_0 + U_1 = V$. It suffices to show that every continuous function f can be written as $f = f_0 + f_1$ where f_0 and f_1 are continuous functions (depending on f) with $f_0(0) = f_1(1) = 0$. Indeed, if $f(0) \neq f(1)$, then we can take[5]

$$f_0 = \frac{f(1)}{f(1) - f(0)}(f - f(0)), \qquad f_1 = \frac{f(0)}{f(0) - f(1)}(f - f(1)),$$

while in the case $f(0) = f(1) = c$ we can take f_0 and f_1 that are given by

$$f_0(x) = c(f(x) + x - c) + (f(x) - c), \qquad f_1(x) = -c(f(x) + x - c - 1).$$

Note that in all cases we indeed have $f_0 \in U_0$ and $f_1 \in U_1$. This proves the claim.

The following lemma shows that the sum of two subspaces is generated by the union of any set of generators for one of the spaces and any set of generators for the other.

Lemma 3.41. *Suppose V is a vector space containing two subsets S and T. Then the equality $L(S) + L(T) = L(S \cup T)$ holds.*

Proof. Exercise. □

Definition 3.42. Let V be a vector space. Two linear subspaces $U_1, U_2 \subset V$ are said to be *complementary* (in V) if $U_1 \cap U_2 = \{0\}$ and $U_1 + U_2 = V$.

Example 3.43. Take $u = (1,0)$ and $u' = (2,1)$ in $\mathbb{R}^2$, and set $U = L(u)$ and $U' = L(u')$. We can write every $(x,y) \in \mathbb{R}^2$ as

$$(x,y) = (x - 2y, 0) + (2y, y) = (x - 2y) \cdot u + y \cdot u' \in U + U',$$

so $U + U' = \mathbb{R}^2$. Suppose $v \in U \cap U'$. Then there are $\lambda, \mu \in \mathbb{R}$ with

$$(\lambda, 0) = \lambda u = v = \mu u' = (2\mu, \mu),$$

which implies $\mu = 0$, so $v = 0$ and $U \cap U' = \{0\}$. We conclude that U and U' are complementary subspaces.

Lemma 3.44. *Let V be a vector space and U and U' subspaces of V. Then U and U' are complementary subspaces of V if and only if for every $v \in V$ there are unique elements $u \in U$ and $u' \in U'$ such that $v = u + u'$.*

Proof. First suppose U and U' are complementary subspaces. Let $v \in V$. Since $V = U + U'$, there certainly are $u \in U$ and $u' \in U'$ such that $v = u + u'$. Now

[5] Knowing that we have $f - f(a) \in U_a$ for $a \in \{0,1\}$, we found the mysterious choices for f_0 and f_1 by looking for $\lambda, \mu \in \mathbb{R}$ for which f equals $\lambda(f - f(0)) + \mu(f - f(1)) = (\lambda + \mu)f - (\lambda f(0) + \mu f(1))$ for all f; this yields two linear equations $\lambda + \mu = 1$ and $\lambda f(0) + \mu f(1) = 0$, which we can solve for λ and μ.

assume that also $v = w + w'$ with $w \in U$ and $w' \in U'$. Then $u + u' = w + w'$, so $u - w = w' - u' \in U \cap U'$, hence $u - w = w' - u' = 0$, and $u = w$, $u' = w'$.

Conversely, suppose that for every $v \in V$ there are unique $u \in U$, $u' \in U'$ such that $v = u + u'$. Then certainly we have $U + U' = V$. Now suppose $w \in U \cap U'$. Then we can write w in two ways as $w = u + u'$ with $u \in U$ and $u' \in U'$, namely with $u = w$ and $u' = 0$, as well as with $u = 0$ and $u' = w$. From uniqueness, we find that these two are the same, so $w = 0$ and $U \cap U' = \{0\}$. We conclude that U and U' are complementary subspaces. □

As it stands, we do not yet know if every subspace U of a vector space V has a complementary subspace in V. In Proposition 7.60 we will see that this is indeed the case, at least when V is finitely generated. The next proposition shows that it is true in an easy special case, namely when F is contained in $\mathbb{R}$ and U is the subspace of F^n generated by a nonzero element $a \in F^n$.

Corollary 3.45. *Suppose F is contained in $\mathbb{R}$. Let $n \geq 0$ be an integer and $a \in F^n$ a nonzero element. Then the subspaces $L(a)$ and*

$$a^\perp = \{\, x \in F^n \;:\; \langle a, x \rangle = 0 \,\}$$

are complementary subspaces of F^n.

Proof. Proposition 1.30 says that every $v \in F^n$ can be written uniquely as the sum of an element $v_1 \in L(a)$ and an element $v_2 \in a^\perp$. Hence, by Lemma 3.44, the spaces $L(a)$ and $a^\perp$ are complementary subspaces, which already finishes the proof.

Alternatively, we first conclude only $L(a) + a^\perp = F^n$ from Proposition 1.30. We also claim $L(a) \cap a^\perp = \{0\}$. Indeed, suppose that $w = \lambda a \in L(a)$ is also contained in $a^\perp$. Then we have $0 = \langle w, a \rangle = \lambda \langle a, a \rangle$. Since a is nonzero, we have $\langle a, a \rangle \neq 0$, so we conclude $\lambda = 0$, which means $w = 0$. □

Warning 3.46. If U and U' are complementary subspaces of a vector space V, then they are *not* setwise complements of each other! First of all, they are not disjoint, as we have $U \cap U' = \{0\} \neq \emptyset$. Second, we have $U \cup U' \neq V$ unless one one the subspaces is $\{0\}$ and the other is V.

Exercises

3.5.1. Prove Lemma 3.41.

3.5.2. State and prove a version of Lemma 3.41 for an arbitrary collection of $(S_i)_{i \in I}$ of subsets.

3.5.3. Suppose $U_1, U_2 \subset F^n$ are subspaces. Show that we have

$$(U_1 + U_2)^\perp = U_1^\perp \cap U_2^\perp.$$

3.5.4. Suppose V is a vector space with a subspace $U \subset V$. Suppose that $U_1, U_2 \subset V$ are subspaces of V that are contained in U. Show that the sum $U_1 + U_2$ is also contained in U.

3.5.5. Take $u = (1, 0)$ and $u' = (\alpha, 1)$ in $\mathbb{R}^2$, for any $\alpha \in \mathbb{R}$. Show that $U = L(u)$ and $U' = L(u')$ are complementary subspaces.

3.5.6. Let U_+ and U_- be the subspaces of $\mathbb{R}^{\mathbb{R}}$ of even and odd functions, respectively (cf. Exercise 3.4.7).

(1) Show that for any $f \in \mathbb{R}^{\mathbb{R}}$, the functions f_+ and f_- given by
$$f_+(x) = \frac{f(x)+f(-x)}{2} \qquad \text{and} \qquad f_-(x) = \frac{f(x)-f(-x)}{2}$$
are even and odd, respectively.

(2) Show that U_+ and U_- are complementary subspaces.

3.5.7. Are the subspaces U_0 and U_1 of Example 3.40 complementary subspaces?

3.5.8. True or false? For every subspaces U, V, W of a common vector space, we have $U \cap (V + W) = (U \cap V) + (U \cap W)$. Prove it, or give a counterexample.

3.5.9. Let W be a vector space with subspaces U_1, U_2, V_1, V_2 satisfying
$$U_1 \subset V_1 \qquad \text{and} \qquad U_2 \subset V_2.$$
Suppose that $U_1 + U_2 = W$ and $V_1 \cap V_2 = \{0\}$.

(1) Show that V_1 and V_2 are complementary subspaces in W, and that U_1 and U_2 are as well.

(2) Show that we have $U_1 = V_1$ and $U_2 = V_2$.

In the proof of Proposition 1.30 and Corollary 3.45, and the definition of reflection, we used the fact that a is nonzero to conclude that we have $\langle a, a\rangle \neq 0$. The following exercises show that, in these three cases, this is the only way in which we used that the ground field is $\mathbb{R}$. They give a generalisation to general fields.

3.5.10. Let $n \geq 0$ be an integer, and $a \in F^n$ an element with $\langle a, a\rangle \neq 0$. Show that for every element $v \in F^n$ there is a unique $\lambda \in F$ such that for $w = v - \lambda a$ we have $\langle a, w\rangle = 0$. Moreover, this λ equals $\frac{\langle a,v\rangle}{\langle a,a\rangle}$; we then have $\langle \lambda a, \lambda a\rangle = \frac{\langle a,v\rangle^2}{\langle a,a\rangle}$ and $w = v - \lambda a$ satisfies $\langle w, w\rangle = \langle v, v\rangle - \frac{\langle a,v\rangle^2}{\langle a,a\rangle}$.

3.5.11. Let $n \geq 0$ be an integer, and $a \in F^n$ an element with $\langle a, a\rangle \neq 0$. Show that the subspaces $L(a)$ and
$$a^\perp = \{\, x \in F^n \;:\; \langle a, x\rangle = 0 \,\}$$
are complementary subspaces of F^n.

3.5.12. Let $n \geq 0$ be an integer, and $a \in F^n$ an element with $\langle a, a\rangle \neq 0$. Set
$$H = a^\perp = \{\, x \in F^n \;:\; \langle a, x\rangle = 0 \,\}.$$
Then for any $v \in F^n$, we define the *reflection* of v in H to be
$$s_H(v) = v - 2\frac{\langle v, a\rangle}{\langle a, a\rangle}a.$$

(1) Show that the reflection of $s_H(v)$ in H equals v.

(2) Suppose that $s_H(w)$ is the reflection of a vector $w \in F^n$ and $s_H(x)$ is the reflection of the sum $x = v + w$. Show that $s_H(x) = s_H(v) + s_H(w)$. (A similar statement holds for the scalar multiplication instead of the sum; together, this shows that reflections are linear maps, as defined in the next section. See Example 4.16.)

CHAPTER 4

Linear maps

Recall that F is still a field (see the beginning of Chapter 2).

So far, we have defined the *objects* of our theory: vector spaces and their elements. Now we want to look at *relations* between vector spaces. These are provided by linear maps — maps between two vector spaces that preserve the linear structure.

4.1. Definition and examples

Among all maps between two vector spaces V and W, we want to single out those that are 'compatible with the linear structure.'

Definition 4.1. Let V and W be two F-vector spaces. A map $f\colon V \to W$ is called an *(F-)linear map* or a *homomorphism* if

(1) for all $v_1, v_2 \in V$, we have $f(v_1 + v_2) = f(v_1) + f(v_2)$, and
(2) for all $\lambda \in F$ and all $v \in V$, we have $f(\lambda v) = \lambda f(v)$.

The set of all linear maps from V to W is denoted by $\operatorname{Hom}(V, W)$.

A bijective homomorphism is called an *isomorphism*. Two vector spaces V and W are said to be *isomorphic*, written $V \cong W$, if there exists an isomorphism between them.

A linear map $f\colon V \to V$ is called an *endomorphism* of V; if f is in addition bijective, then it is called an *automorphism* of V. We recall (see Appendix A) that if $f\colon V \to V$ is an endomorphism and n is a positive integer, then we write

$$f^n = \underbrace{f \circ f \circ \cdots \circ f}_{n}$$

for the composition of n times applying f. The first examples of linear maps are given by the following proposition.

Proposition 4.2. *Let $n \geq 0$ be an integer. For every $a \in F^n$, the function*

$$F^n \to F, \quad x \mapsto \langle a, x\rangle$$

is a linear map.

Proof. This follows directly from Proposition 3.11. □

Obviously, the scalar product is in fact linear in both arguments, that is, if instead of the first argument, we fix the second argument to be $a \in F^n$, then also the map $F^n \to F, \quad x \mapsto \langle x, a\rangle$ is linear. This is why we call the scalar product *bilinear*.

Here are some simple properties of linear maps.

Lemma 4.3. *Let U, V, W be vector spaces over a field F.*

(1) *If $f\colon V \to W$ is linear, then $f(0) = 0$.*
(2) *If $f\colon V \to W$ is an isomorphism, then the inverse map f^{-1} is also an isomorphism.*
(3) *If $f\colon U \to V$ and $g\colon V \to W$ are linear maps, then $g \circ f\colon U \to W$ is also linear.*

Proof.

(1) This follows from either one of the two properties of linear maps. Using the first, we get
$$f(0) = f(0+0) = f(0) + f(0)$$
which by Lemma 2.16 implies $f(0) = 0$. Instead, we can also use the second property, which gives
$$f(0) = f(0 \cdot 0) = 0 \cdot f(0) = 0\,.$$
(Which of the zeros are scalars, which are vectors in V, in W?)
(2) The inverse map is certainly bijective; we have to show that it is linear. So take $w_1, w_2 \in W$ and set $v_1 = f^{-1}(w_1)$, $v_2 = f^{-1}(w_2)$. Then $f(v_1) = w_1$, $f(v_2) = w_2$, hence $f(v_1 + v_2) = w_1 + w_2$. This means that
$$f^{-1}(w_1 + w_2) = v_1 + v_2 = f^{-1}(w_1) + f^{-1}(w_2)\,.$$
The second property for being linear is checked in a similar way.
(3) Exercise.

□

Warning 4.4. Many people learn in high school that for all real numbers a, b, the function f from $\mathbb{R}$ to $\mathbb{R}$ given by $f(x) = ax + b$ is called linear. With our definition of linear functions, **this is only the case when** $b = 0$! Indeed, from Lemma 4.3 we find that if f is linear, then $b = f(0) = 0$. For $b = 0$, it is easy to see that f is indeed linear. (It also follows from Proposition 4.2 with $n = 1$.)

Lemma 4.5. *Let $f\colon V \to W$ be a linear map of F-vector spaces.*

(1) *For all $v, w \in V$ and $\lambda, \mu \in F$, we have $f(\lambda v - \mu w) = \lambda f(v) - \mu f(w)$.*
(2) *For all $v_1, v_2, \ldots, v_n \in V$ and $\lambda_1, \lambda_2, \ldots, \lambda_n \in F$ we have*
$$f(\lambda_1 v_1 + \cdots + \lambda_n v_n) = \lambda_1 f(v_1) + \cdots + \lambda_n f(v_n).$$
(3) *For any subset $S \subset V$ we have $f(L(S)) = L(f(S))$.*

Proof. Exercise. □

There are two important linear subspaces associated to any linear map: its image $\operatorname{im}(f)$ and its kernel, which is defined below.

Definition 4.6. Let $f\colon V \to W$ be a linear map. Then the *kernel* of f is defined to be
$$\ker(f) = \{v \in V : f(v) = 0\}\,.$$

Lemma 4.7. *Let $f\colon V \to W$ be a linear map.*

(1) *The kernel* $\ker(f)$ *is a linear subspace of* V*.*
(2) *The image* $\operatorname{im}(f)$ *is a linear subspace of* W*.*
(3) *The map f is injective if and only if* $\ker(f) = \{0\}$*.*

Proof.

(1) We have to check the three properties of subspaces for $\ker(f)$. By the previous remark, $f(0) = 0$, so $0 \in \ker(f)$. Now let $v_1, v_2 \in \ker(f)$. Then $f(v_1) = f(v_2) = 0$, so $f(v_1 + v_2) = f(v_1) + f(v_2) = 0 + 0 = 0$, and $v_1 + v_2 \in \ker(f)$. Finally, let λ be a scalar and $v \in \ker(f)$. Then $f(v) = 0$, so $f(\lambda v) = \lambda f(v) = \lambda \cdot 0 = 0$, and therefore $\lambda v \in \ker(f)$. We conclude that $\ker(f)$ is indeed a subspace.
(2) We check again the subspace properties. We have $f(0) = 0 \in \operatorname{im}(f)$. If $w_1, w_2 \in \operatorname{im}(f)$, then there are $v_1, v_2 \in V$ such that $f(v_1) = w_1$, $f(v_2) = w_2$, hence $w_1 + w_2 = f(v_1 + v_2) \in \operatorname{im}(f)$. If λ is a scalar and $w \in \operatorname{im}(f)$, then there is $v \in V$ such that $f(v) = w$, hence $\lambda w = f(\lambda v) \in \operatorname{im}(f)$. We conclude that $\operatorname{im}(f)$ is indeed a subspace.
(3) If f is injective, then there can be only one element of V that is mapped to $0 \in W$, and since we know that $f(0) = 0$, it follows that $\ker(f) = \{0\}$.

    For the converse, assume that $\ker(f) = \{0\}$, and let $v_1, v_2 \in V$ be such that $f(v_1) = f(v_2)$. Then $f(v_1 - v_2) = f(v_1) - f(v_2) = 0$, so $v_1 - v_2 \in \ker(f)$. By our assumption, this means that $v_1 - v_2 = 0$, hence $v_1 = v_2$. This shows that f is indeed injective.

□

Remark 4.8. If you want to show that a certain subset U in a vector space V is a linear subspace, it may be easier to find a linear map $f\colon V \to W$ such that $U = \ker(f)$ than to check the properties directly.

Example 4.9. Let $n \geq 0$ be an integer, and $a \in F^n$ an element. Then the kernel of the map

$$F^n \to F, \quad x \mapsto \langle a, x \rangle$$

is the set $a^\perp$.

The following lemma generalises the first two statements of Lemma 4.7.

Lemma 4.10. *Let $f\colon V \to W$ be a linear map.*

(1) *If $U \subset W$ is a linear subspace, then $f^{-1}(U)$ is a linear subspace of V; it contains* $\ker(f)$*.*
(2) *If $U \subset V$ is a linear subspace, then $f(U)$ is a linear subspace of W; it is contained in* $\operatorname{im}(f)$*.*

Proof. We leave it as an exercise to generalise the proofs of the first two statements of Lemma 4.7. □

It is time for some more examples of linear maps.

Example 4.11. Let V be any vector space. Then the unique map $f : V \to \{0\}$ to the zero space is linear. More generally, if W is another vector space, then $f : V \to W$, $v \mapsto 0$, is linear. It is called the *zero homomorphism;* often it is denoted by 0. Its kernel is all of V, its image is $\{0\} \subset W$.

Example 4.12. For any vector space V, the identity map id_V is linear; it is even an automorphism of V. Its kernel is trivial ($= \{0\}$); its image is all of V.

Example 4.13. If $V = F^n$, then all the *projection maps*

$$\pi_j : F^n \to F, \quad (x_1, \ldots, x_n) \mapsto x_j$$

are linear. (In fact, one can argue that the vector space structure on F^n is defined in exactly such a way as to make these maps linear.) This map π_j can also be given by $x \mapsto \langle x, e_j \rangle$, where e_j is the j-th standard generator of F^n. The image of π_j is F, so π_j is surjective; its kernel is $e_j^\perp$, which consists of all vectors of which the j-th coordinate is 0.

Example 4.14. Let V be a vector space over F, and $\lambda \in F$ an element. Then the map

$$V \to V, \quad v \mapsto \lambda v$$

is a linear map that is called *multiplication by* λ. It is sometimes denoted by $[\lambda]$, or just λ. Clearly, for two elements $\lambda, \mu \in F$, we have $[\lambda] \circ [\mu] = [\lambda\mu]$. If λ is nonzero, then $[\lambda]$ is an isomorphism, with inverse $[\lambda^{-1}]$.

Example 4.15. Take the vector $a = (1, 1, 1) \in \mathbb{R}^3$ and set

$$V = a^\perp = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \; : \; x_1 + x_2 + x_3 = 0\}.$$

Let $\psi : V \to \mathbb{R}^2$ denote the map that sends (x_1, x_2, x_3) to (x_1, x_2). Then clearly ψ is linear. For every $x = (x_1, x_2, x_3) \in V$ in the kernel of ψ we have $x_1 = x_2 = 0$, so from the definition of V we also get $x_3 = 0$, and therefore $x = 0$. It follows that $\ker(\psi) = \{0\}$, so ψ is injective. The map ψ is also surjective, so ψ is an isomorphism; its inverse sends $(x_1, x_2) \in \mathbb{R}^2$ to $(x_1, x_2, -x_1 - x_2)$.

Example 4.16. Suppose $V = \mathbb{R}^n$ and $a \in V$ is nonzero. Set $H = a^\perp$. Then the following maps from V to V are linear.

(1) The orthogonal projection $\pi_a : \mathbb{R}^n \to \mathbb{R}^n$ onto $L(a)$ given by

$$v \mapsto \frac{\langle v, a \rangle}{\langle a, a \rangle} a$$

(see Definition 1.31). Indeed, linearity follows from the fact that the scalar product with a is linear (see Proposition 4.2). Note that for the $a = e_j$, the j-th standard vector, and the projection map $\pi_j : \mathbb{R}^n \to \mathbb{R}$ on the j-th coordinate, we have

$$\pi_{e_j}(v) = \pi_j(v) \cdot e_j.$$

The kernel of π_a is $a^\perp$ and the image is $L(a)$.

(2) The orthogonal projection $\pi_H = \pi_{a^\perp} : \mathbb{R}^n \to \mathbb{R}^n$ onto H given by

$$v \mapsto v - \frac{\langle v, a \rangle}{\langle a, a \rangle} a = v - \pi_a(v)$$

(see Definition 1.31). Indeed, for checking addition, note that, by linearity of π_a, we have

$$\pi_{a^\perp}(v+w) = v+w-\pi_a(v+w) = v-\pi_a(v)+w-\pi_a(w) = \pi_{a^\perp}(v)+\pi_{a^\perp}(w).$$

The scalar multiplication follows similarly. The kernel of $\pi_{a^\perp}$ is $L(a)$ and the image is $a^\perp$.

(3) The reflection $s_H\colon \mathbb{R}^n \to \mathbb{R}^n$ in the hyperplane $H = a^\perp$ given by

$$v \mapsto v - 2\frac{\langle v, a\rangle}{\langle a, a\rangle}a$$

(see Definition 1.53 and the identity in (1.11)). The linearity is proven in the same way as for the projection onto $H = a^\perp$. The identity in (1.8) shows that $s_H \circ s_H = \mathrm{id}_V$, which implies that s_H is an isomorphism.

Example 4.17. Let V be the vector space of 3×3 magic squares (see Example 2.7). Then the map $r\colon V \to V$ that rotates the square over 90 degrees is linear. Another endomorphism is the map $c\colon V \to V$ that sends a square M to the constant square in which all entries are equal to the middle square of M. Check this for yourself! We leave it as an exercise to find the kernel and the image of these linear maps.

Example 4.18. For any two vector spaces V_1, V_2 over F, the projection maps $V_1 \times V_2 \to V_1$ and $V_1 \times V_2 \to V_2$ given by $(v_1, v_2) \mapsto v_1$ and $(v_1, v_2) \mapsto v_2$, respectively, are linear, cf. Exercise 2.2.15.

Exercises

4.1.1. Finish the proof of Lemma 4.3.

4.1.2. Prove Lemma 4.5.

4.1.3. Finish the proof of Lemma 4.7.

4.1.4. Which of the following maps between vector spaces are linear?
(1) $\mathbb{R}^3 \to \mathbb{R}^2$, $(x, y, z) \mapsto (x - 2y, z + 1)$,
(2) $\mathbb{R}^3 \to \mathbb{R}^3$, $(x, y, z) \mapsto (x^2, y^2, z^2)$,
(3) (†) $\mathbb{C}^3 \to \mathbb{C}^4$, $(x, y, z) \mapsto (x + 2y, x - 3z, y - z, x + 2y + z)$,
(4) $\mathbb{R}^3 \to V$, $(x, y, z) \mapsto xv_1 + yv_2 + zv_3$, for a vector space V over $\mathbb{R}$ with $v_1, v_2, v_3 \in V$,
(5) $P(\mathbb{R}) \to P(\mathbb{R})$, $f \mapsto f'$, where $P(\mathbb{R})$ is the vector space of real polynomials and f' is the derivative of f,
(6) $P \to \mathbb{R}^2$, $f \mapsto (f(2), f'(0))$.

4.1.5. Given the linear maps of Examples 4.17 and 4.18, what are their kernels and images?

4.1.6. Let $f\colon V \to W$ be a linear map of vector spaces. Show that the following are equivalent.
(1) The map f is surjective.
(2) For every subset $S \subset V$ with $L(S) = V$ we have $L(f(S)) = W$.
(3) There is a subset $S \subset V$ with $L(f(S)) = W$.

4.1.7. Let $\rho\colon \mathbb{R}^2 \to \mathbb{R}^2$ be rotation about the origin $(0, 0)$ over an angle θ.
(1) Show that ρ is a linear map. [You may assume that ρ sends parallelograms to parallelograms.]
(2) What are the images $\rho((1, 0))$ and $\rho((0, 1))$?
(3) Show that we have

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

4.1.8. Show that the reflection $s\colon \mathbb{R}^2 \to \mathbb{R}^2$ in the line given by $y = -x$ is a linear map. Give an explicit formula for s.

4.1.9. As before, let $F[x]$ be the vector space of polynomials over F.

(1) Given an element $a \in F$, we define the evaluation map $\mathrm{ev}_a\colon F[x] \to F$ that sends a polynomial $f = \sum_{i=0}^{d} c_i x^i$ to $f(a) = \sum_{i=0}^{d} c_i a^i$. Show that ev_a is linear.

(2) Show that the map $\varphi\colon F[x] \to F^F$ of Remark 3.36 and Exercise D.2.1 is given by
$$f \mapsto (a \mapsto \mathrm{ev}_a(f)),$$
and deduce that φ is linear.

4.1.10. Given the map
$$T\colon \mathbb{R}^2 \to \mathbb{R}^2, (x, y) \mapsto x(\tfrac{3}{5}, \tfrac{4}{5}) + y(\tfrac{4}{5}, -\tfrac{3}{5})$$
and the vectors $v_1 = (2, 1)$ and $v_2 = (-1, 2)$.

(1) Show that $T(v_1) = v_1$ and $T(v_2) = -v_2$.

(2) Show that T equals the reflection in the line given by $2y - x = 0$.

4.1.11. Give an explicit expression for the linear map $s\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by reflecting in the line $y = 3x$.

4.2. Linear maps form a vector space

If X is any set, and W an F-vector space, then we can add any two functions $f, g\colon X \to W$ point-wise, by defining the sum $f + g$ to be given by
$$(f + g)(x) = f(x) + g(x)$$
for every $x \in X$. Note that the last plus sign denotes addition in W. We will see that if X is itself a vector space over F, and f and g are linear maps, then the sum $f + g$ is linear as well. A similar statement holds for the point-wise scalar multiplication. With the language that we have set up so far, we can phrase this as follows.

Lemma 4.19. *Let V and W be two F-vector spaces. Then the set* $\mathrm{Hom}(V, W)$ *of all linear maps $V \to W$, with addition and scalar multiplication defined point-wise, forms an F-vector space.*

Proof. Using only the fact that W is a vector space, one checks that the vector space axioms hold for the set of *all* maps $V \to W$ (see Exercise 2.2.12). Hence it suffices to show that the linear maps form a linear subspace.

The zero map is a linear map, so it is contained in $\mathrm{Hom}(V, W)$. If $f, g\colon V \to W$ are two linear maps, we have to check that $f + g$ is again linear. So let $v_1, v_2 \in V$ be elements; then we have
$$\begin{aligned}(f + g)(v_1 + v_2) &= f(v_1 + v_2) + g(v_1 + v_2) = f(v_1) + f(v_2) + g(v_1) + g(v_2)\\ &= f(v_1) + g(v_1) + f(v_2) + g(v_2) = (f + g)(v_1) + (f + g)(v_2)\,.\end{aligned}$$
Similarly, if $\lambda \in F$ and $v \in V$, then we have
$$(f + g)(\lambda v) = f(\lambda v) + g(\lambda v) = \lambda f(v) + \lambda g(v) = \lambda\big(f(v) + g(v)\big) = \lambda \cdot (f + g)(v)\,.$$
We conclude that $f + g$ is indeed linear, so $\mathrm{Hom}(V, W)$ is closed under addition. Now let $\mu \in F$, and let $f\colon V \to W$ be linear. We have to check that μf is again

linear. So let $v_1, v_2 \in V$ be elements; then we have

$$\begin{aligned}(\mu f)(v_1 + v_2) &= \mu f(v_1 + v_2) = \mu\big(f(v_1) + f(v_2)\big) \\ &= \mu f(v_1) + \mu f(v_2) = (\mu f)(v_1) + (\mu f)(v_2)\,.\end{aligned}$$

Finally, let $\lambda \in F$ and $v \in V$. Then

$$(\mu f)(\lambda v) = \mu f(\lambda v) = \mu\big(\lambda f(v)\big) = (\mu\lambda) f(v) = \lambda\big(\mu f(v)\big) = \lambda \cdot (\mu f)(v)\,.$$

It follows that μf is indeed linear, so $\mathrm{Hom}(V, W)$ is also closed under scalar multiplication. It follows that $\mathrm{Hom}(V, W)$ is indeed a linear subspace. □

Example 4.20. Let $V = \mathrm{Map}(\mathbb{R}, \mathbb{R})$ be the vector space of functions from $\mathbb{R}$ to $\mathbb{R}$. For any $a \in \mathbb{R}$, we let $\mathrm{ev}_a \in \mathrm{Hom}(V, \mathbb{R})$ denote the evaluation map that sends a function $f \in V$ to $f(a)$. Then for two real numbers $a, b \in \mathbb{R}$, the map $\mathrm{ev}_a + \mathrm{ev}_b \in \mathrm{Hom}(V, \mathbb{R})$ sends the function f to $\mathrm{ev}_a(f) + \mathrm{ev}_b(f) = f(a) + f(b)$.

Example 4.21. Let $f, g \in \mathrm{Hom}(\mathbb{R}^3, \mathbb{R}^2)$ be given by

$$\begin{aligned} f\big((x, y, z)\big) &= (x - z, x + 2y), \\ g\big((x, y, z)\big) &= (y + z, y - z). \end{aligned}$$

Then the linear map $h = f + g$ is given by

$$h\big((x, y, z)\big) = (x + y, x + 3y - z).$$

Example 4.22. Let $\rho\colon \mathbb{R}^2 \to \mathbb{R}^2$ be the rotation around 0 over an angle $2\pi/3$. Then ρ sends $(x, y) \in \mathbb{R}^2$ to

$$(-\tfrac{1}{2}x - \tfrac{1}{2}\sqrt{3}y, \tfrac{1}{2}\sqrt{3}x - \tfrac{1}{2}y)$$

(see Exercise 4.1.7). The map $\rho^2 = \rho \circ \rho$ is rotation over $4\pi/3$, so we can use Exercise 4.1.7 to easily obtain an explicit formula for that as well. Instead, we use the above and compute

$$\begin{aligned} \rho^2\big((x, y)\big) = \rho\big(\rho(x)\big) &= \rho\big((-\tfrac{1}{2}x - \tfrac{1}{2}\sqrt{3}y, \tfrac{1}{2}\sqrt{3}x - \tfrac{1}{2}y)\big) \\ &= \big(-\tfrac{1}{2}(-\tfrac{1}{2}x - \tfrac{1}{2}\sqrt{3}y) - \tfrac{1}{2}\sqrt{3}(\tfrac{1}{2}\sqrt{3}x - \tfrac{1}{2}y), \\ &\qquad \tfrac{1}{2}\sqrt{3}(-\tfrac{1}{2}x - \tfrac{1}{2}\sqrt{3}y) - \tfrac{1}{2}(\tfrac{1}{2}\sqrt{3}x - \tfrac{1}{2}y)\big) \\ &= \big(-\tfrac{1}{2}x + \tfrac{1}{2}\sqrt{3}y, -\tfrac{1}{2}\sqrt{3}x - \tfrac{1}{2}y\big), \end{aligned}$$

which is indeed what Exercise 4.1.7 would have given. Adding the two expressions for ρ and ρ^2, we find that the sum $\rho + \rho^2$ sends a point $p = (x, y)$ to $(-x, -y) = -p$, so in fact, the map $\mathrm{id} + \rho + \rho^2$ is the zero map. We could have also seen this geometrically, as for each point $p \in \mathbb{R}^2$, the three points p, $\rho(p)$, and $\rho(\rho(p))$ are the vertices of an equilateral triangle with center 0, so their sum is 0.

Example 4.23. Suppose $V = \mathbb{R}^n$ and $a \in V$ is nonzero. Set $H = a^\perp$. Let π_a, π_H, and s_H be the orthogonal projection onto $L(a)$, the orthogonal projection onto H, and the reflection in H, respectively, as in Example 4.16. Then the linearity of the last two maps follows from the linearity of the first, as we have

$$\pi_H = \mathrm{id}_V - \pi_a, \qquad \text{and} \qquad s_H = \mathrm{id}_V - 2\pi_a = 2\pi_H - \mathrm{id}_V\,.$$

Note that this is in line with the fact that in Example 4.16 we used linearity of π_a to prove linearity of π_H and s_H.

Example 4.24. Suppose $V = \mathbb{R}^n$ and $a \in V$ is nonzero. Set $H = a^\perp$ and $L = L(a)$. Let π_a and π_H be the orthogonal projection onto $L(a)$ and H, respectively. Then the map $s_L = \mathrm{id}_V - 2\pi_H = 2\pi_a - \mathrm{id}_V$ is also linear. This is reflection in the line L. In $\mathbb{R}^3$, it is the same as rotation around L over 180 degrees.

Example 4.25. Let $M \subset \mathbb{R}^2$ be the line of all points $(x, y) \in \mathbb{R}^2$ with $x = y$. Then $a = (1, -1)$ is a normal of M, and the reflection s in M sends the point $(x, y) \in \mathbb{R}^2$ to (y, x). By the previous example, the orthogonal projection π_a onto $L(a)$ satisfies $s = \mathrm{id} - 2\pi_a$, so we have $\pi_a = \frac{1}{2}(\mathrm{id} - s)$. This means that π_a sends (x, y) to $\left(\frac{1}{2}(x-y), \frac{1}{2}(y-x)\right) = \frac{1}{2}(x-y)\cdot a$. The projection π_M onto the line M satisfies $\mathrm{id} + s = 2\pi_M$. This means that π_M sends (x, y) to $\left(\frac{1}{2}(x+y), \frac{1}{2}(x+y)\right)$. Draw pictures to convince yourself!

Example 4.26. If V is an F-vector space, and we multiply the identity id_V by the scalar λ, then we obtain the map $[\lambda]$ that is multiplication by λ.

The following proposition shows that composition of linear maps respects addition and scalar multiplication.

Proposition 4.27. *Let U, V, W be vector spaces over F. Let $f, f_1, f_2 \in \mathrm{Hom}(U, V)$ and $g, g_1, g_2 \in \mathrm{Hom}(V, W)$ be linear maps. Let $\lambda \in F$ be a scalar. Then we have*

$$\begin{aligned} g \circ (f_1 + f_2) &= (g \circ f_1) + (g \circ f_2), \\ (g_1 + g_2) \circ f &= (g_1 \circ f) + (g_2 \circ f), \\ g \circ (\lambda f) &= \lambda \cdot (g \circ f) = (\lambda g) \circ f. \end{aligned}$$

Proof. Let $u \in U$ be any element. To verify the first identity, we note that

$$\begin{aligned} (g \circ (f_1 + f_2))(u) &= g((f_1 + f_2)(u)) = g(f_1(u) + f_2(u)) = g(f_1(u)) + g(f_2(u)) \\ &= (g \circ f_1)(u) + (g \circ f_2)(u) = ((g \circ f_1) + (g \circ f_2))(u). \end{aligned}$$

Note that for the first and fourth equality we used the definition of composition, for the second and fifth equality we used the definition of addition of maps (to V and W, respectively), and for the third equality we used linearity of g. This proves the first identity, as it holds for all $u \in U$. For the second identity of the proposition, we have

$$\begin{aligned} ((g_1 + g_2) \circ f)(u) &= (g_1 + g_2)(f(u)) = g_1(f(u)) + g_2(f(u)) \\ &= (g_1 \circ f)(u) + (g_2 \circ f)(u) = ((g_1 \circ f) + (g_2 \circ f))(u), \end{aligned}$$

where the first and third equality follow from the definition of composition, and the second and fourth equality from the definition of addition of maps. Since this holds for all $u \in U$, it proves the second identity. We leave the last two identities as an exercise: only for one of the two, linearity of g is needed. $\square$

Warning 4.28. Note that composition of functions is not commutative in general. If V is a vector space and $f, g \in \mathrm{Hom}(V, V)$ are two endomorphisms of V, then we have

$$(f + g)^2 = (f + g) \circ (f + g) = f \circ f + f \circ g + g \circ f + g \circ g.$$

Since $f \circ g$ may not be equal to $g \circ f$, we can in general not simplify the right-hand side.

Example 4.29. Let $\mathcal{C}^\infty(\mathbb{R})$ denote the set of all functions from $\mathbb{R}$ to $\mathbb{R}$ that can be differentiated n times for every positive integer n. In other words, we set $\mathcal{C}^\infty(\mathbb{R}) = \bigcap_{n>0} \mathcal{C}^n(\mathbb{R})$ with $\mathcal{C}^n(\mathbb{R})$ as in Example 3.8. Let $D\colon \mathcal{C}^\infty(\mathbb{R}) \to \mathcal{C}^\infty(\mathbb{R})$ be the linear map that sends a function f to its derivative f'. Then for every integer $n > 0$, the map D^n sends a function f to its n-th derivative $f^{(n)}$. The maps $\mathrm{id} + D$ and $\mathrm{id} - D$ send a function f to $f + f'$ and $f - f'$, respectively. Of course, we can work out easily what the composition $(\mathrm{id} + D) \circ (\mathrm{id} - D)$ does to a function f, but with Proposition 4.27 we immediately find $(\mathrm{id} + D) \circ (\mathrm{id} - D) = \mathrm{id} - D^2$, so it sends f to $f - f^{(2)}$.

Example 4.30. Let V be a vector space and $\pi\colon V \to V$ an endomorphism.

(1) Suppose $\pi^2 = 0$. Then for $f = \mathrm{id} + \pi$ and $g = \mathrm{id} - \pi$ we have
$$f \circ g = g \circ f = \mathrm{id} - \pi^2 = \mathrm{id},$$
so $\mathrm{id} + \pi$ and $\mathrm{id} - \pi$ are each other's inverses, and therefore both bijective.
[A nonzero example is the map $\pi\colon \mathbb{R}^2 \to \mathbb{R}^2$ that sends (x, y) to $(0, x)$.]

(2) Suppose $\pi^2 = \pi$ (cf. Exercise 4.2.6). Then for $\pi' = \mathrm{id} - \pi$ we have
$$\pi'^2 = (\mathrm{id} - \pi) \circ (\mathrm{id} - \pi) = \mathrm{id} - \pi - \pi + \pi^2 = \mathrm{id} - \pi = \pi'.$$
[A nonzero example is $V = \mathbb{R}^n$ and, for some nonzero vector $a \in V$, the map π is the orthogonal projection onto the line $L(a)$; then π' is the orthogonal projection on the hyperplane $a^\perp$.]

Example 4.31. Let $P(\mathbb{R})$ be the vector space of polynomial functions on $\mathbb{R}$. Then the following maps are linear.

(1) Evaluation: given $a \in \mathbb{R}$, the map $\mathrm{ev}_a\colon P(\mathbb{R}) \to \mathbb{R}$, $p \mapsto p(a)$ is linear. The kernel of ev_a consists of all polynomials having a zero at a; the image is all of $\mathbb{R}$.

(2) Differentiation: $D\colon P(\mathbb{R}) \to P(\mathbb{R})$, $p \mapsto p'$ is linear.
The kernel of D consists of the constant polynomials; the image of D is $P(\mathbb{R})$ (see below).

(3) Definite integration: given $a < b$, the map
$$I_{a,b}\colon P(\mathbb{R}) \longrightarrow \mathbb{R}\,, \quad p \longmapsto \int_a^b p(x)\,dx$$
is linear.

(4) Indefinite integration: given $a \in \mathbb{R}$, the map
$$I_a\colon P(\mathbb{R}) \longrightarrow P(\mathbb{R})\,, \quad p \longmapsto \Big(x \mapsto \int_a^x p(t)\,dt\Big)$$
is linear. This map is injective; its image is the kernel of ev_a (see below).

(5) Translation: given $a \in \mathbb{R}$, the map
$$T_a\colon P(\mathbb{R}) \longrightarrow P(\mathbb{R})\,, \quad p \longmapsto \big(x \mapsto p(x + a)\big)$$
is linear. This map is an isomorphism: $T_a^{-1} = T_{-a}$.

The *Fundamental Theorem of Calculus* says that $D \circ I_a = \mathrm{id}_P$ and that for every $p \in P(\mathbb{R})$ we have $(I_{a,b} \circ D)(p) = p(b) - p(a)$ and $(I_a \circ D)(p) = p - p(a)$. This can now be written as $I_{a,b} \circ D = \mathrm{ev}_b - \mathrm{ev}_a$ and $I_a \circ D = \mathrm{id}_P - \mathrm{ev}_a$.

The relation $D \circ I_a = \mathrm{id}_P$ implies that I_a is injective and that D is surjective. This implies that $\mathrm{ev}_a \circ I_a = 0$, hence $\mathrm{im}(I_a) \subset \ker(\mathrm{ev}_a)$. On the other hand, if $p \in \ker(\mathrm{ev}_a)$, then $I_a(p') = p - p(a) = p$, so $p \in \mathrm{im}(I_a)$. Therefore we have shown that $\mathrm{im}(I_a) = \ker(\mathrm{ev}_a)$.

Let $C \subset P(\mathbb{R})$ be the subspace of constant polynomials, and let $Z_a \subset P(\mathbb{R})$ be the subspace of polynomials vanishing at $a \in \mathbb{R}$. Then $C = \ker(D)$ and $Z_a = \ker(\mathrm{ev}_a) = \mathrm{im}(I_a)$, and C and Z_a are complementary subspaces. The map D restricts to an isomorphism $Z_a \xrightarrow{\sim} P(\mathbb{R})$, and I_a restricts (on the target side) to an isomorphism $P(\mathbb{R}) \xrightarrow{\sim} Z_a$ (exercise!).

Exercises

4.2.1. Let V be the vector space of 3×3 magic squares (see Example 2.7). Let r and c be the endomorphisms of Example 4.17. Show that we have $\mathrm{id}_V + r^2 = 2c$.

4.2.2. As in Example 4.22, we let $\rho\colon \mathbb{R}^2 \to \mathbb{R}^2$ denote rotation around 0 over $2\pi/3$. Set $f = \rho - \mathrm{id}$ and $g = \rho + 2 \cdot \mathrm{id}$. [Suggestion: Draw some pictures of what these linear maps f and g do.]

(1) Use Example 4.22 to show that $f \circ g = g \circ f = -3 \cdot \mathrm{id}$.
(2) Conclude that f and g are isomorphisms.

4.2.3. Let $V \subset \mathbb{R}^3$ be the plane

$$V = \{\, (x, y, z) \in \mathbb{R}^3 \ : \ 2x - y + z = 0 \,\}.$$

(1) Give an explicit expression for the reflection $s\colon \mathbb{R}^3 \to \mathbb{R}^3$ in the plane V. [Hint: first find the images of the standard generators e_1, e_2, e_3.]
(2) Show that the subsets

$$U_+ = \{v \in \mathbb{R}^3 \ : \ s(v) = v\} \qquad \text{and} \qquad U_- = \{v \in \mathbb{R}^3 \ : \ s(v) = -v\}$$

are subspaces.
(3) Show $U_+ = V$ and $U_- = L(a)$ for some $a \in \mathbb{R}^3$.
(4) Show that U_+ and U_- are complementary subspaces.

4.2.4. This exercise generalises Exercises 3.5.6 and 4.2.3. Assume[1] that in F we have $2 \neq 0$, so that we can divide by 2. Let V be a vector space over F, and let $s\colon V \to V$ be a linear map satisfying $s(s(v)) = v$ for all $v \in V$ (for example, $s\colon \mathbb{R}^n \to \mathbb{R}^n$ is the reflection in some hyperplane). Set

$$V_+ = \{\, v \in V \ : \ s(v) = v \,\} \qquad \text{and} \qquad V_- = \{\, v \in V \ : \ s(v) = -v \,\}.$$

(1) Show that s is an isomorphism.
(2) Show that for every $v \in V$ we have

$$\frac{v + s(v)}{2} \in V_+ \qquad \text{and} \qquad \frac{v - s(v)}{2} \in V_-.$$

(3) Show that $\mathrm{id}_V + s$ has kernel V_- and image V_+.
(4) Show that $\mathrm{id}_V - s$ has kernel V_+ and image V_-.
(5) Show that V_+ and V_- are complementary subspaces in V.
(6) For what choice of s does Exercise 3.5.6 become a special case?

[1] For readers that assume F is contained in $\mathbb{R}$ or $\mathbb{C}$ (see beginning of Chapter 2), this assumption holds automatically.

4.2.5. Suppose V is a vector space with two complementary subspaces U and U', cf. Definition 3.42. Then for every $v \in V$ there are unique elements $u \in U$ and $u' \in U'$ with $v = u + u'$ by Lemma 3.44; let $\pi_U \colon V \to V$ denote the map that sends v to the corresponding element u. Note that π_U also depends on U', even though it is not referred to in the notation. We call π_U the projection of V onto U along U'.
(1) Show that π_U is linear.
(2) Show that π_U has image U and kernel $\ker \pi_U = U'$.
(3) Show that π_U satisfies $\pi_U \circ \pi_U = \pi_U$.
(4) Show that π_U is the unique endomorphism of V that is the identity on U and 0 on U'.
(5) Show that $\mathrm{id}_V - \pi_U$ is the projection of V onto U' along U.

4.2.6. Let V be a vector space and $\pi \colon V \to V$ an endomorphism that satisfies $\pi \circ \pi = \pi$. Set $U = \mathrm{im}(\pi)$ and $U' = \ker(\pi)$.
(1) Show that for every $v \in V$, we have $v - \pi(v) \in U'$.
(2) Show that U and U' are complementary subspaces in V.
(3) Show that π is the projection of V onto U along U'.
[For this reason, any endomorphism π satisfying $\pi^2 = \pi$ is often called a projection.]

4.2.7. Let V be the vector space of 3×3 magic squares, and let $c \colon V \to V$ be the endomorphism of Example 4.17. Show that we have $c^2 = c$, and use the Exercise 4.2.6 to show that V contains two complementary subspaces, namely the subspace of all constant squares and the subspace of all the magic squares of which the row, column, and diagonal sums are 0.

4.2.8. Let V be a vector space and $f \colon V \to V$ an endomorphism. Suppose that f is *nilpotent*, that is, there is a positive integer k such that $f^k = 0$. Show that the linear map $\mathrm{id} - f$ is an isomorphism.

4.3. Linear equations

Suppose $m, n > 0$ are integers and we have elements $a_{ij} \in F$ for $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$. Consider the system

$$\begin{cases} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ & & & & & & & \vdots & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{cases}$$

of m linear equations in n variables $x_1, \ldots, x_n$ over the field F. Let $f \colon F^n \to F^m$ be the map that sends $x = (x_1, \ldots, x_n) \in F^n$ to the vector

$$\begin{aligned} (&a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n, \\ &a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n, \ldots \\ &a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n) \in F^m, \end{aligned}$$

and set $b = (b_1, b_2, \ldots, b_m) \in F^m$. Then we can rewrite the system of equations as $f(x) = b$, to be solved for $x \in F^n$. The solution set equals

$$\{\, x \in F^n \ : \ f(x) = b \,\} = f^{-1}(b).$$

Since f is a linear map, we can use linear algebra to study these equations.

Definition 4.32. Let $f : V \to W$ be a linear map between two F-vector spaces. The equation

$$f(x) = 0\,,$$

to be solved for $x \in V$, is called a *homogeneous linear equation.* If $V = F^n$ and $W = F^m$ (with $m > 1$), we also speak of a *homogeneous system of linear equations.* (Since as above, the equation consists of m separate equations in F, coming from the coordinates of F^m.)

If $b \in W \setminus \{0\}$, then the equation

$$f(x) = b$$

(again to be solved for $x \in V$) is called an *inhomogeneous linear equation,* or in the case $V = F^n$ and $W = F^m$, an *inhomogeneous system of linear equations.* The equation or system of equations is called *consistent* if it has a solution, that is, if $b \in \operatorname{im}(f)$.

With the theory we have built so far, the following result is essentially trivial.

Theorem 4.33. *Let $f : V \to W$ be a linear map between two F-vector spaces.*

(1) *The solution set of the homogeneous linear equation $f(x) = 0$ is the linear subspace* $\ker f \subset V$.
(2) *Let $b \in W \setminus \{0\}$. If the inhomogeneous linear equation $f(x) = b$ is consistent, and $a \in V$ is a solution, then the set of all solutions is the set*

$$f^{-1}(b) = \{\ a + z\ :\ z \in \ker f\ \}. \tag{4.1}$$

Proof.

(1) By definition, the solution set $f^{-1}(0)$ is exactly the kernel of f.
(2) Let x be any solution and $z = x-a$. Then $f(z) = f(x)-f(a) = b-b = 0$, so $z \in \ker f$ and $x = a+z$. This shows the inclusion '$\subset$' in (4.1). Conversely, if $x = a + z$ for some $z \in \ker f$, then
$$f(x) = f(a + z) = f(a) + f(z) = b + 0 = b,$$
which proves the other inclusion '$\supset$'.

□

Example 4.34. As before, let $\mathbb{R}[x]$ denote the vector space of all real polynomials. Let $a, b \in \mathbb{R}$ be real numbers, and

$$X = \{\, g \in \mathbb{R}[x]\ :\ g(a) = b\,\}$$

the set of all polynomials that take the value b at a. If we let $\mathrm{ev}_a \colon \mathbb{R}[x] \to \mathbb{R}$ denote the linear map that sends g to $g(a)$, then X is the solution set of the linear equation $\mathrm{ev}_a(g) = b$ in the variable $g \in \mathbb{R}[x]$, so we have $X = \mathrm{ev}_a^{-1}(b)$. As the constant polynomial b is a solution, Theorem 4.33, applied to $f = \mathrm{ev}_a$, yields the (trivial) fact that $X = \{\, g + b\ :\ g \in \mathbb{R}[x],\ g(a) = 0\,\}$.

Example 4.35. Consider the *wave equation*

$$\frac{\partial^2 f}{\partial t^2} = c^2\,\frac{\partial^2 f}{\partial x^2}$$

for $f \in \mathcal{C}^2(\mathbb{R} \times [0,\pi])$, with boundary conditions $f(t,0) = f(t,\pi) = 0$ and initial conditions $f(0,x) = f_0(x)$ and $\frac{\partial f}{\partial t}(0,x) = 0$. If we ignore the first initial condition for a moment, we can consider this as a homogeneous linear equation, where we let

$$V = \{f \in \mathcal{C}^2(\mathbb{R}\times[0,\pi]) : \forall t \in \mathbb{R} : \ f(t,0) = f(t,\pi) = 0,\ \forall x \in \,]0,\pi[\, : \ \tfrac{\partial f}{\partial t}(0,x) = 0\}$$

and $W = \mathcal{C}(\mathbb{R} \times [0,\pi])$, and the linear map $V \to W$ is the *wave operator*

$$w : f \longmapsto \frac{\partial^2 f}{\partial t^2} - c^2 \frac{\partial^2 f}{\partial x^2} .$$

We can find fairly easily a bunch of solutions using the trick of 'separating the variables' — we look for solutions of the form $f(t,x) = g(t)h(x)$. This leads to an equation

$$\frac{1}{c^2}\frac{g''(t)}{g(t)} = \frac{h''(x)}{h(x)} ,$$

and the common value of both sides must be constant. The boundary conditions then force $h(x) = \sin kx$ (up to scaling) for some $k \geq 1$, and then $g(t) = \cos kct$ (again up to scaling). Since we know that the solution set is a linear subspace, we see that all linear combinations

$$f(t,x) = \sum_{k=1}^{n} a_k \cos kct \sin kx$$

are solutions. Such a solution has

$$f(0,x) = \sum_{k=1}^{n} a_k \sin kx ,$$

so if f_0 is of this form, we have found a (or the) solution to the original problem. Otherwise, we have to use some input from Analysis, which tells us that we can approximate f_0 by linear combinations as above and that the corresponding solutions will approximate the solution we are looking for.

Remark 4.36. Suppose $f \colon V \to W$ is a linear map of which you already know it is an isomorphism with inverse $g = f^{-1}$. Then for any $b \in W$, the (unique) solution to the linear equation $f(x) = b$ is of course just $x = g(b)$.

Exercises

4.3.1. For any $a, b, c \in \mathbb{R}$, we consider the system

$$\begin{cases} -x_1 + x_2 + x_3 & = & a \\ 2x_1 + x_2 - 2x_3 & = & b \\ x_1 + 2x_2 - x_3 & = & c \end{cases}$$

of linear equations in $x = (x_1, x_2, x_3)$.

(1) Show that for $a = b = c = 0$, the solution set equals

$$\{\,\lambda(1,0,1) \ : \ \lambda \in \mathbb{R}\,\}.$$

(2) Describe the solution set for $a = 1$, $b = 1$, $c = 2$.
(3) Describe the solution set for $a = 1$, $b = 0$, $c = 0$.

4.4. Characterising linear maps

In this section, we let n denote a non-negative integer, and we study the linear maps with F^n as domain. As before, we let $e_1, e_2, \dots, e_n$ denote the standard generators of F^n.

In Proposition 4.2 we saw that for every $a \in F^n$, the scalar product with a gives a linear map from F^n to F. The following proposition shows that all linear maps from F^n to F are of this form.

Proposition 4.37. *Let $f \colon F^n \to F$ be a linear map. Then there is a unique vector $a \in F^n$ such that for all $x \in F^n$ we have*

$$f(x) = \langle a, x \rangle.$$

Moreover, this vector a equals $\big(f(e_1), f(e_2), \dots, f(e_n)\big)$.

Proof. Suppose there exists such an element a and write $a = (a_1, a_2, \dots, a_n)$. Then for each i with $1 \le i \le n$ we have

$$f(e_i) = \langle a, e_i \rangle = a_1 \cdot 0 + \cdots + a_{i-1} \cdot 0 + a_i \cdot 1 + a_{i+1} \cdot 0 + \cdots + a_n \cdot 0 = a_i.$$

We conclude that $a = \big(f(e_1), f(e_2), \dots, f(e_n)\big)$, so a is completely determined by f and therefore unique, if it exists.

To show there is indeed an a as claimed, we take

$$a = \big(f(e_1), f(e_2), \dots, f(e_n)\big)$$

(we have no choice by the above) and show it satisfies $f(x) = \langle a, x \rangle$ for all $x \in F^n$, as required. Indeed, if we write $x = (x_1, x_2, \dots, x_n)$, then we find

$$f(x) = f(x_1 \cdot e_1 + \cdots + x_n \cdot e_n) = x_1 \cdot f(e_1) + \cdots + x_n \cdot f(e_n) = \langle x, a \rangle = \langle a, x \rangle.$$

$\square$

Propositions 4.2 and 4.37 give a bijection between the vector space F^n and the vector space $\mathrm{Hom}(F^n, F)$ of linear maps from F^n to F. The following proposition generalises the codomain from F to a general vector space W: there is a bijection between W^n and the vector space $\mathrm{Hom}(F^n, W)$ of linear maps from F^n to W (see Remark 4.40). In Exercise 4.4.5 we will see that this bijection is in fact an isomorphism.

Proposition 4.38. *Let W be a vector space over F. Then for every sequence $(w_1, w_2, \dots, w_n)$ of n vectors in W, there is a unique linear map $\varphi \colon F^n \to W$ such that for every $i \in \{1, \dots, n\}$ we have $\varphi(e_i) = w_i$. Moreover, this map φ sends the element $(x_1, \dots, x_n) \in F^n$ to $x_1 w_1 + \cdots + x_n w_n$.*

Proof. Suppose that φ is a linear map such that for every $i \in \{1, \dots, n\}$ we have $\varphi(e_i) = w_i$. Then for $x = (x_1, x_2, \dots, x_n) \in F^n$ we have

$$\varphi(x) = \varphi(x_1 e_1 + \cdots + x_n e_n) = x_1 \varphi(e_1) + \cdots + x_1 \varphi(e_n) = x_1 w_1 + \cdots + x_n w_n,$$

so φ is completely determined on all $x \in F^n$ by the vectors $w_1, w_2, \dots, w_n$ and therefore φ is unique, if it exists.

To show there is indeed a φ as claimed, we define the function $\varphi \colon F^n \to W$ by

$$\varphi(x) = x_1 w_1 + \cdots + x_n w_n$$

(we have no choice by the above). One easily checks that φ is linear. (Do this!) For i with $1 \leq i \leq n$, we have

$$\varphi(e_i) = 0 \cdot w_1 + \cdots + 0 \cdot w_{i-1} + 1 \cdot w_i + 0 \cdot w_{i+1} + \cdots + 0 \cdot w_n = w_i,$$

so φ indeed satisfies the requirements. $\square$

By construction, the image of the map φ of Proposition 4.38 consists of all linear combinations of $w_1, w_2, \ldots, w_n$, so it equals $L(w_1, \ldots, w_n)$; this implies that φ is surjective if and only if the elements $w_1, w_2, \ldots, w_n$ generate W.

Definition 4.39. For any F-vector space W, and a sequence $C = (w_1, w_2, \ldots, w_n)$ of n elements in W, we write φ_C for the linear map $\varphi\colon F^n \to W$ associated to C as in Proposition 4.38.

Remark 4.40. Suppose $W = F$. Then for any element $a \in F^n$, the associated map $\varphi_a\colon F^n \to F$ sends $x \in F^n$ to $\langle a, x\rangle$. Moreover, Proposition 4.37 is a special case of Proposition 4.38, which becomes clear from the rephrasing of these propositions in Exercises 4.4.5 and 4.4.4.

Exercises

4.4.1. For each of the problems 4.1.7, 4.1.8, 4.1.10, 4.1.11, and parts (3) and (4) of problem 4.1.4, give a vector space W, an integer n, and a sequence $C \in W^n$ such that the described linear map is φ_C.

4.4.2. Let $j \in \{1, \ldots, n\}$ be an integer, and let $\pi_j\colon F^n \to F$ be the projection on the j-th coordinate (see Example 4.13).

(1) For which vector space W, integer m, and a sequence $C \in W^m$ does π_j equal φ_C?

(2) For which element $a \in F^n$ is π_j given by $\pi_j(x) = \langle a, x\rangle$ for all $x \in F^n$?

4.4.3. In this exercise we characterise linear maps of which the *co*domain is F^m. For $1 \leq i \leq m$, let $\pi_i\colon F^m \to F$ denote the projection on the i-th coordinate, as in Example 4.13. Let V be a vector space over F.

(1) Let $f\colon V \to F^m$ be any map. Show that the map f is linear if and only if for every i, the composition $\pi_i \circ f\colon V \to F$ is linear.

(2) Conclude that for any linear maps $f_1, \ldots, f_m\colon V \to F$, the map

$$V \to F^m, \qquad v \mapsto \big(f_1(v), f_2(v), \ldots, f_m(v)\big)$$

is linear.

(3) Show that the associations above yield a bijection

$$\mathrm{Hom}(V, F^m) \to \mathrm{Hom}(V, F)^m.$$

(4) Show that this bijection is an isomorphism.

4.4.4. Prove that the map

$$F^n \to \mathrm{Hom}(F^n, F), \quad a \mapsto (x \mapsto \langle a, x\rangle)$$

is an isomorphism whose inverse sends the map $f \in \mathrm{Hom}(F^n, F)$ to the sequence $\big(f(e_1), f(e_2), \ldots, f(e_n)\big)$.

4.4.5. Let W be a vector space over F. Show that the map

$$W^n \to \mathrm{Hom}(F^n, W), \quad C \mapsto \varphi_C$$

is an isomorphism whose inverse sends the map $f \in \mathrm{Hom}(F^n, W)$ to the sequence $\big(f(e_1), f(e_2), \ldots, f(e_n)\big)$.

4.4.6. The scalar product on F^n is a map $F^n \times F^n \to F$, satisfying some conditions. In this exercise, we will generalise this to F^X for any set X. Note that if X is finite, then F^X and $F^{(X)}$ as in Exercise 3.1.9 are equal. In general, we have a map

$$F^X \times F^{(X)} \to F, \qquad (f, g) \mapsto \langle f, g\rangle = \sum_{x \in X} f(x)g(x),$$

where the sum contains only finitely many nonzero terms, because there are only finitely many $x \in X$ with $g(x) \neq 0$.

(1) Show that this generalised scalar product satisfies the conditions of Proposition 3.11.

(2) Show that there is an isomorphism

$$F^X \to \mathrm{Hom}(F^{(X)}, F)$$

that sends a vector $f \in F^X$ to the linear map $g \mapsto \langle f, g\rangle$.

4.4.7. This exercise generalises Proposition 4.38. Let X be a (not necessarily finite) set. Consider the subspace $F^{(X)}$ of F^X as in Exercise 3.1.9, and the elements e_x (for $x \in X$) as in Exercise 3.4.5. Let W be a vector space over F and let $C \in \mathrm{Map}(X, W) = W^X$ be a function from X to W. Set $w_x = C(x)$ for each $x \in X$.

(1) Show that there is a unique linear map $\varphi_C \colon F^{(X)} \to W$ that satisfies $\varphi_C(e_x) = w_x$ for every $x \in X$ and that this map is surjective if and only if the set $\{w_x : x \in X\}$ generates W.

(2) Show that there is an isomorphism

$$W^X \to \mathrm{Hom}(F^{(X)}, W)$$

that sends $C \in W^X$ to φ_C.

***4.4.8.** This exercise generalises several of the previous exercises. Let V and W be vector spaces over F, and let X be any set. Let $V^{(X)}$ be as in Exercise 3.1.9.

(1) Show that the map

$$\begin{aligned} \mathrm{Hom}(V, W)^X &\to \mathrm{Hom}(V^{(X)}, W) \\ f &\mapsto \Big(\sigma \mapsto \sum_{x \in X} f(x)(\sigma(x))\Big) \end{aligned}$$

is an isomorphism.

(2) Show that the map

$$\begin{aligned} \mathrm{Hom}(V, W)^X &\to \mathrm{Hom}(V, W^X) \\ f &\mapsto \big(v \mapsto (x \mapsto f(x)(v))\big) \end{aligned}$$

is an isomorphism.

(3) Show how Exercises 4.4.3, 4.4.6, and 4.4.7 are special cases of this.

4.5. Isomorphisms

If $f \colon V \to W$ is a isomorphism, then the two vector spaces V and W can for all practical purposes be identified through f. This is illustrated by the following proposition.

Proposition 4.41. *Suppose $\varphi\colon V \to V'$ and $\psi\colon W \to W'$ are isomorphisms of vector spaces. Suppose $f\colon V \to W$ is a linear map and set $f' = \psi \circ f \circ \varphi^{-1}\colon V' \to W'$. Then the diagram*

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "\varphi"'] & W \arrow[d, "\psi"] \\ V' \arrow[r, "f'"] & W' \end{tikzcd}$$

commutes, φ restricts to an isomorphism $\ker f \to \ker f'$, and ψ restricts to an isomorphism $\operatorname{im} f \to \operatorname{im} f'$.

Proof. Exercise. □

Example 4.42. Take the vector $a = (1,1,1) \in \mathbb{R}^3$ and set

$$V = a^\perp = \{(x_1,x_2,x_3) \in \mathbb{R}^3 \;:\; x_1+x_2+x_3 = 0\}.$$

Let $\psi\colon V \to \mathbb{R}^2$ be the isomorphism of Example 4.15 that sends (x_1,x_2,x_3) to (x_1,x_2). Its inverse ψ^{-1} sends (x_1,x_2) to $(x_1,x_2,-x_1-x_2)$. Let $r\colon V \to V$ be the linear map that sends (x_1,x_2,x_3) to (x_2,x_3,x_1) and set $r' = \psi \circ r \circ \psi^{-1}$. Then $r'\colon \mathbb{R}^2 \to \mathbb{R}^2$ sends the point (x_1,x_2) to $(x_2,-x_1-x_2)$. When we identify V with $\mathbb{R}^2$ through the map ψ, the map r corresponds with r'. For example, just like we have $r^3 = \mathrm{id}_V$, we also have $r'^3 = \mathrm{id}_{\mathbb{R}^2}$, which can easily be checked directly as well.

Exercises

4.5.1. Let $f\colon V \to V$ be an endomorphism of a vectorspace V. Let $\sigma\colon V \to W$ be a linear map. Suppose that f sends $\ker\sigma$ to itself, that is, $f(\ker\sigma) \subset \ker\sigma$. Show that f induces a well-defined endomorphism

$$\tilde{f}\colon \operatorname{im}\sigma \to \operatorname{im}\sigma$$

that sends the element $\sigma(z) \in \operatorname{im}\sigma$ to $\sigma(f(z))$ for every $z \in V$.

4.5.2. Suppose we have a diagram

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "\varphi"'] & W \arrow[d, "\psi"] \\ V' \arrow[r, "f'"] & W' \end{tikzcd}$$

of linear maps that commutes, that is, we have linear maps $\varphi\colon V \to V'$ and $\psi\colon W \to W'$ and $f\colon V \to W$ and $f'\colon V' \to W'$ satisfying $\psi \circ f = f' \circ \varphi$.

(1) Show that φ restricts to a linear map $\overline{\varphi}\colon \ker f \to \ker f'$.
(2) Show that ψ restricts to a linear map $\overline{\psi}\colon \operatorname{im} f \to \operatorname{im} f'$.
(3) Show that if φ is injective, then so is $\overline{\varphi}$.
(4) Show that if ψ is injective, then so is $\overline{\psi}$.
(5) Show that if φ is surjective, then so is $\overline{\psi}$.
(6) Show that if φ is surjective and ψ is injective, then $\overline{\varphi}$ is surjective.
(7) Give examples that show that neither of the two hypotheses can be left out of the previous statement.
(8) Prove Proposition 4.41.

4.5.3. Let V be a vector space and $\sigma\colon X \to Y$ any map of sets. Define the map

$$\sigma^*\colon V^Y = \operatorname{Map}(Y,V) \to \operatorname{Map}(X,V) = V^X$$

by $\sigma^*(f) = f \circ \sigma$.

(1) Show that σ^* is a linear map.
(2) Show that if σ is injective, then σ^* is surjective.
(3) Show that if σ is surjective, then σ^* is injective.
(4) Show that if σ is bijective, then σ^* is an isomorphism.

4.5.4.

(1) Suppose $\alpha\colon V' \to V$ is a linear map of vector spaces over F. Show that for every vector space W over F there is a linear map
$$\alpha^*\colon \operatorname{Hom}(V,W) \to \operatorname{Hom}(V',W)$$
that sends f to $f \circ \alpha$.
(2) Suppose $\beta\colon W \to W'$ is a linear map of vector spaces over F. Show that for every vector space V over F there is a linear map
$$\beta_*\colon \operatorname{Hom}(V,W) \to \operatorname{Hom}(V,W')$$
that sends f to $\beta \circ f$.
(3) Check that in Proposition 4.41 we have
$$f' = (\psi_* \circ (\varphi^{-1})^*)(f) = ((\varphi^{-1})^* \circ \psi_*)(f).$$

4.5.5. Suppose $\alpha, \alpha_1, \alpha_2\colon V' \to V$ and $\alpha'\colon V'' \to V'$ are linear maps of vector spaces over F. Let W be a vector space over F. With the notation of Exercise 4.5.4, show that we have the following.

(1) Show that $(\alpha \circ \alpha')^* = (\alpha')^* \circ \alpha^*$.
(2) Show that $(\alpha_1 + \alpha_2)^* = \alpha_1^* + \alpha_2^*$.
(3) Show that $(\lambda\alpha)^* = \lambda \cdot \alpha^*$ for any $\lambda \in F$.

4.5.6. Suppose $\beta, \beta_1, \beta_2\colon W \to W'$ and $\beta'\colon W' \to W''$ are linear maps of vector spaces over F. Let V be a vector space over F. With the notation of Exercise 4.5.4, show that we have the following.

(1) Show that $(\beta' \circ \beta)_* = \beta'_* \circ \beta_*$.
(2) Show that $(\beta_1 + \beta_2)_* = (\beta_1)_* + (\beta_2)_*$.
(3) Show that $(\lambda\beta)_* = \lambda \cdot \beta_*$ for any $\lambda \in F$.

4.5.7. Suppose $\varphi\colon V \to V'$ and $\psi\colon W \to W'$ are isomorphisms of vector spaces. Show that the linear map
$$(\varphi^{-1})^* \circ \psi_*\colon \operatorname{Hom}(V,W) \to \operatorname{Hom}(V',W'),$$
which sends f to $\psi \circ f \circ \varphi^{-1}$, is an isomorphism (see Proposition 4.41 and Exercise 4.5.4).

CHAPTER 5

Matrices

For this chapter, let m and n denote non-negative integers. Unless explicitly mentioned otherwise, the standard generators $e_1, e_2, \ldots, e_n$ are the standard generators of F^n.

Before we give the definition of a matrix in Section 5.1, we give some motivation for that definition. Let $\varphi\colon F^n \to F^m$ be a linear map. By Proposition 4.38, this map φ is uniquely determined by the images $w_1 = \varphi(e_1), \ldots, w_n = \varphi(e_n)$ in F^m of the n standard generators of F^n. If $C = (w_1, \ldots, w_n)$ is the sequence of these images, then φ equals φ_C as in Definition 4.39.

From a different viewpoint, we can interpret $\varphi\colon F^n \to F^m$ as a sequence of m linear maps $\varphi_1, \varphi_2, \ldots, \varphi_m\colon F^n \to F$, one for each coordinate of F^m, so that φ is given by (cf. Exercise 4.4.3)

$$x \mapsto \big(\varphi_1(x), \ldots, \varphi_m(x)\big).$$

Each of the m maps $\varphi_i\colon F^n \to F$ is given by $x \mapsto \langle v_i, x\rangle$ for some $v_i \in F^n$ (see Proposition 4.37), so φ is determined by the m vectors $v_1, \ldots, v_m \in F^n$.

We will see in Proposition 5.2 that if we write the n vectors $w_1, \ldots, w_n \in F^m$ as columns next to each other, then we obtain the same array of $m \times n$ elements of F as when we write the m vectors $v_1, \ldots, v_m \in F^n$ as rows underneath each other!

In other words, for every $i \in \{1, \ldots, m\}$ and every $j \in \{1, \ldots, n\}$, the i-th coordinate of w_j is equal to the j-th coordinate of v_i; this element equals $\varphi_i(e_j)$ and it is in the i-th row and the j-th column of the array just described. The map

$$A\colon \{1, \ldots, m\} \times \{1, \ldots, n\} \to F$$

that sends (i, j) to this element $\varphi_i(e_j)$ is called a matrix (see Definition 5.1), often written as the array

$$\begin{pmatrix} A(1,1) & A(1,2) & \cdots & A(1,n) \\ A(2,1) & A(2,2) & \cdots & A(2,n) \\ \vdots & \vdots & & \vdots \\ A(m,1) & A(m,2) & \cdots & A(m,n) \end{pmatrix} = \Big(A(i,j)\Big)_{1 \le i \le m, 1 \le j \le n}$$

described above. By abuse of language, we will also refer to such an array as a matrix.

5.1. Definition of matrices

Definition 5.1. Let F be a field and m, n non-negative integers. An $m \times n$ *matrix over* F is a function

$$A\colon \{1, \dots, m\} \times \{1, \dots, n\} \to F,$$

often written as an array

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} = (a_{ij})_{1 \le i \le m, 1 \le j \le n}$$

of *entries* or *coefficients* $a_{ij} = A(i, j) \in F$.

By abuse of language, we will often identify a matrix with its associated array and vice versa.

For $i \in \{1, \dots, m\}$, the vector $(a_{i1}, a_{i2}, \dots, a_{in})$ is a *row* of A, which is an element of F^n, and for $j \in \{1, \dots, n\}$, the vector

$$\begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}$$

is called a *column* of A, which is an element of F^m, be it written vertically here.

If we denote the j-th column by w_j (for $1 \le j \le n$), then we also write A as

$$A = \begin{pmatrix} | & | & & | \\ w_1 & w_2 & \cdots & w_n \\ | & | & & | \end{pmatrix}$$

where the vertical lines above and below w_j indicate that w_j is a vector that makes up the whole j-th column of the matrix. A similar notation can be used to indicate which rows the matrix A consists of.

The set of all $m \times n$ matrices with entries in F is denoted by $\mathrm{Mat}(m \times n, F)$. Note that as a boundary case, $m = 0$ or $n = 0$ (or both) is allowed; in this case $\mathrm{Mat}(m \times n, F)$ has only one element, which is an empty matrix.

If $m = n$, we sometimes write $\mathrm{Mat}(n, F)$ for $\mathrm{Mat}(n \times n, F)$. The matrix

$$I = I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} = (\delta_{ij})_{1 \le i, j \le n} .$$

is called the *identity matrix*.

5.2. Matrix associated to a linear map

Proposition 5.2. *Let $\varphi\colon F^n \to F^m$ be a linear map. For $1 \leq i \leq m$, let $v_i \in F^n$ be the vector for which the linear map $\varphi_i = \pi_i \circ \varphi\colon F^n \to F$ is given by $x \mapsto \langle v_i, x\rangle$. For $1 \leq j \leq n$, set $w_j = \varphi(e_j) \in F^m$. Then the $m \times n$ matrix*

$$\begin{pmatrix} | & | & & | \\ w_1 & w_2 & \cdots & w_n \\ | & | & & | \end{pmatrix}$$

with $w_1, \ldots, w_n$ as columns equals the matrix

$$\begin{pmatrix} -v_1- \\ -v_2- \\ \vdots \\ -v_m- \end{pmatrix}$$

with $v_1, \ldots, v_m$ as rows.

Proof. Consider any indices $1 \leq i \leq m$ and $1 \leq j \leq n$. The entry in row i and column j of the first matrix is the i-th coordinate of $w_j = \varphi(e_j)$, so this entry equals $(\pi_i \circ \varphi)(e_j) = \varphi_i(e_j) = \langle v_i, e_j\rangle$, which is the j-th coordinate of v_i, and thus the entry in row i and column j of the second matrix. The equality of the two matrices follows. $\square$

Remark 5.3. Note that if $\varphi\colon F^n \to F^m$ is a linear map, then the linear map $\varphi_i = \pi_i \circ \varphi\colon F^n \to F$ used in Proposition 5.2 is the map obtained from φ by only considering the i-th coordinate of the image in F^m.

Example 5.4. Let $\varphi\colon \mathbb{R}^4 \to \mathbb{R}^3$ be given by

$(x_1, x_2, x_3, x_4) \mapsto (x_1 + x_2 - 3x_3 + 2x_4, -2x_1 + 3x_3 - 5x_4, 7x_1 + 3x_2 - 2x_3 + 6x_4)$

and for $1 \leq j \leq 4$, set $w_j = \varphi(e_j) \in \mathbb{R}^3$. Then we have

$$w_1 = \begin{pmatrix} 1 \\ -2 \\ 7 \end{pmatrix}, \quad w_2 = \begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix}, \quad w_3 = \begin{pmatrix} -3 \\ 3 \\ -2 \end{pmatrix}, \quad w_4 = \begin{pmatrix} 2 \\ -5 \\ 6 \end{pmatrix},$$

where we have already conveniently written w_1, w_2, w_3, w_4 vertically. As in Proposition 5.2, for $1 \leq i \leq 3$, we let $v_i \in \mathbb{R}^4$ denote the vector for which the linear map $\varphi_i = \pi_i \circ \varphi\colon \mathbb{R}^4 \to \mathbb{R}$ is given by $x \mapsto \langle v_i, x\rangle$. For $i = 1$ we have $\varphi_1((x_1, x_2, x_3, x_4)) = x_1 + x_2 - 3x_3 + 2x_4$, so $v_1 = (1, 1, -3, 2)$. Similarly, we obtain

$$\begin{aligned} v_1 &= (1, 1, -3, 2), \\ v_2 &= (-2, 0, 3, -5), \\ v_3 &= (7, 3, -2, 6). \end{aligned}$$

Indeed, we find

$$\begin{pmatrix} | & | & | & | \\ w_1 & w_2 & w_3 & w_4 \\ | & | & | & | \end{pmatrix} = \begin{pmatrix} 1 & 1 & -3 & 2 \\ -2 & 0 & 3 & -5 \\ 7 & 3 & -2 & 6 \end{pmatrix} = \begin{pmatrix} -v_1- \\ -v_2- \\ -v_3- \end{pmatrix}.$$

For the rest of this section, let $\varphi \colon F^n \to F^m$ be a linear map and let the vectors $v_1, \ldots, v_m \in F^n$ and $w_1, \ldots, w_n \in F^m$ be as in Proposition 5.2. Let

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} = \begin{pmatrix} | & | & & | \\ w_1 & w_2 & \cdots & w_n \\ | & | & & | \end{pmatrix} = \begin{pmatrix} -v_1- \\ -v_2- \\ \vdots \\ -v_m- \end{pmatrix}$$

be the matrix associated to φ as in Proposition 5.2. Then for every vector

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in F^n,$$

the image $\varphi(x)$ can be written, by definition of $v_1, \ldots, v_m$ and $w_1, \ldots, w_n$, as

$$\varphi(x) = \begin{pmatrix} \langle v_1, x \rangle \\ \langle v_2, x \rangle \\ \vdots \\ \langle v_m, x \rangle \end{pmatrix} \qquad \text{and} \qquad \varphi(x) = x_1 w_1 + \ldots + x_n w_n$$

(see Proposition 4.38). If we write out either expression, we obtain

$$\varphi(x) = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix}. \tag{5.1}$$

Note that here we have written $\varphi(x)$ vertically, just like we may write the columns $w_1, \ldots, w_n$ vertically. This way the coordinates $\langle v_i, x \rangle$ are written underneath each other, analogous to how the rows v_i are written underneath each other in the matrix A. We will see later, in Remark 5.19, why in this context it is convenient to also write x vertically.

5.3. The product of a matrix and a vector

In the previous section, we started with a linear map φ and saw that we may associate a matrix to it. Conversely, we will see that every matrix defines a linear map. Motivated by (5.1), we define the product of any matrix $A \in \mathrm{Mat}(m \times n, F)$ and a vector $x \in F^n$ as follows.

Definition 5.5. For any matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \in \mathrm{Mat}(m \times n, F) \quad \text{and vector} \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in F^n$$

we define the product Ax as

$$Ax = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix}.$$

Note that here again, we have written x and Ax vertically.

Analogous to the previous section, if we let

$$v_i = (a_{i1}, a_{i2}, \dots, a_{in})$$

be the i-th row of A (for $1 \le i \le m$), then we can write Ax as

$$Ax = \begin{pmatrix} -v_1- \\ -v_2- \\ \vdots \\ -v_m- \end{pmatrix} \cdot x = \begin{pmatrix} \langle v_1, x\rangle \\ \langle v_2, x\rangle \\ \vdots \\ \langle v_m, x\rangle \end{pmatrix}, \tag{5.2}$$

so the entries of Ax are the scalar products of x with the row vectors of A. If we let

$$w_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}$$

denote the j-th column of A (for $1 \le j \le n$), then we can write Ax as

$$Ax = \begin{pmatrix} | & | & & | \\ w_1 & w_2 & \cdots & w_n \\ | & | & & | \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1 w_1 + x_2 w_2 + \ldots + x_n w_n, \tag{5.3}$$

so Ax is the linear combination of the column vectors of A with the entries of x as coefficients. Note that $Ae_j = w_j$.

Remark 5.6. Both points of view on the multiplication will prove useful: the coordinates of Ax being the scalar products of x with the rows of A on one hand, and Ax being a linear combination of the columns of A on the other hand.

Example 5.7. We have

$$\begin{pmatrix} 3 & 2 & 1 \\ -1 & 2 & 7 \\ -3 & 5 & -2 \end{pmatrix} \begin{pmatrix} 2 \\ -2 \\ -1 \end{pmatrix} = \begin{pmatrix} 3\cdot 2 + 2\cdot(-2) + 1\cdot(-1) \\ (-1)\cdot 2 + 2\cdot(-2) + 7\cdot(-1) \\ (-3)\cdot 2 + 5\cdot(-2) + (-2)\cdot(-1) \end{pmatrix} = \begin{pmatrix} 1 \\ -13 \\ -14 \end{pmatrix}.$$

Verify that the result does indeed correspond with the three scalar products of the vector $(2, -2, -1)$ with the rows of the 3×3 matrix. Also verify that the result equals the right linear combination of the columns.

Exercises

5.3.1. For the given matrix A and the vector x, determine Ax.

(1)

$$A = \begin{pmatrix} -2 & -3 & 1 \\ 1 & 1 & -2 \\ 0 & 1 & 1 \end{pmatrix} \quad \text{and} \quad x = \begin{pmatrix} -3 \\ -4 \\ 2 \end{pmatrix},$$

(2)

$$A = \begin{pmatrix} 1 & -3 & 2 \\ -2 & -4 & 2 \end{pmatrix} \quad \text{and} \quad x = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix},$$

(3)

$$A = \begin{pmatrix} 4 & 3 \\ 3 & -2 \\ -3 & -1 \\ -1 & 1 \end{pmatrix} \qquad \text{and} \qquad x = \begin{pmatrix} -2 \\ 3 \end{pmatrix}.$$

5.4. Linear maps associated to matrices

Definition 5.8. To any matrix $A \in \mathrm{Mat}(m \times n, F)$ we associate the function $f_A \colon F^n \to F^m$ given by

$$f_A(x) = Ax$$

for all $x \in F^n$.

Lemma 5.9. *Let A be an $m \times n$ matrix over F with columns $w_1, \ldots, w_n \in F^m$. Let $f_A \colon F^n \to F^m$ be the function associated to A. Then we have $f_A(e_j) = w_j$ for all $1 \le j \le n$, and f_A equals φ_C as in Definition 4.39 with $C = (w_1, \ldots, w_n)$.*

Proof. For every $x = (x_1, \ldots, x_n) \in F^n$, we have

$$f_A(x) = x_1 w_1 + \ldots + x_n w_n = \varphi_C(x),$$

so we obtain $f_A = \varphi_C$. In particular, we have $f_A(e_j) = w_j$ for all $1 \le j \le n$. $\square$

Note that Lemma 5.9 implies that for any $m \times n$ matrix A, the map f_A is linear and the j-th column of A equals $f_A(e_j)$ for any $j \in \{1, \ldots, n\}$. In fact, by Proposition 4.38, the function $f_A \colon F^n \to F^m$ is the unique linear map sending e_j to the j-th column of A.

Clearly, the linear map f_I associated to the matrix $I = I_n$ is the identity map $F^n \to F^n$.

Example 5.10. Let $A \in \mathrm{Mat}(3 \times 4, \mathbb{R})$ be the matrix

$$\begin{pmatrix} 3 & 2 & 0 & -1 \\ 1 & -2 & 5 & -3 \\ 0 & 1 & 4 & 7 \end{pmatrix}.$$

Then the map f_A sends

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \in \mathbb{R}^4 \qquad \text{to} \qquad \begin{pmatrix} 3x_1 & +2x_2 & & -x_4 \\ x_1 & -2x_2 & +5x_3 & -3x_4 \\ & x_2 & +4x_3 & +7x_4 \end{pmatrix} \in \mathbb{R}^3.$$

Proposition 5.11. *Let F be a field and m, n non-negative integers. Suppose $f \colon F^n \to F^m$ is a linear map. Then there is a unique matrix $A \in \mathrm{Mat}(m \times n, F)$ with $f = f_A$.*

Proof. Let A be the matrix associated to f as in Proposition 5.2, that is, define $w_j = f(e_j)$ for $1 \le j \le n$, and let A be the matrix of which the j-th column is w_j for each j. Then $f_A(e_j) = Ae_j = w_j = f(e_j)$ for all j, so $f = f_A$ by Proposition 4.38. Furthermore, any $m \times n$ matrix A' with $f_{A'} = f$ has its j-th column equal to $A'e_j = f_{A'}(e_j) = f(e_j) = w_j$ for all j, so $A' = A$. This finishes the proof. $\square$

Example 5.12. Let $\rho\colon \mathbb{R}^2 \to \mathbb{R}^2$ be the rotation about the origin $(0,0)$ over an angle θ. From Exercise 4.1.7, we know that ρ is given by

$$\rho\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x\cos\theta - y\sin\theta \\ x\sin\theta + y\cos\theta \end{pmatrix}.$$

We conclude that ρ corresponds to the matrix

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

Example 5.13. Let $s = s_L\colon \mathbb{R}^2 \to \mathbb{R}^2$ be the reflection in the line L given by $y = 2x$. Then s is linear and we can determine a 2×2 matrix A such that $s = f_A$. By Lemma 5.9, the columns of A are the images $f_A(e_1) = s(e_1)$ and $f_A(e_2) = s(e_2)$. Note that the vector $a = (2, -1)$ is a normal of L. For any vector $v \in \mathbb{R}^2$, the reflection of v in L is $s(v) = v - 2\lambda a$ with $\lambda = \frac{\langle v,a\rangle}{\langle a,a\rangle}$ (see (1.11) and Figure 1.18, where H plays the role of our L). We find

$$s(e_1) = e_1 - 2\cdot\tfrac{2}{5}\cdot a = \begin{pmatrix} -\frac{3}{5} \\ \frac{4}{5} \end{pmatrix} \qquad\text{and}\qquad s(e_2) = e_2 - 2\cdot\tfrac{-1}{5}\cdot a = \begin{pmatrix} \frac{4}{5} \\ \frac{3}{5} \end{pmatrix},$$

so we get

$$A = \begin{pmatrix} -\frac{3}{5} & \frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{pmatrix}.$$

Proposition 5.11 shows that the map

$$\mathrm{Mat}(m \times n, F) \to \mathrm{Hom}(F^n, F^m), \qquad A \mapsto f_A \tag{5.4}$$

is a bijection. Therefore, one often identifies a matrix A with the linear map f_A that the matrix induces. In this way we may refer to the kernel and image of f_A as the kernel and image of A and we write $\ker A = \ker f_A$ and $\operatorname{im} A = \operatorname{im} f_A$.

Exercises

5.4.1. For each of the linear maps of the form $f\colon F^n \to F^m$ of the exercises of Section 4.1, give a matrix M such that f is given by

$$x \mapsto Mx.$$

5.4.2. Given the matrix

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

and the linear map $f\colon \mathbb{R}^n \to \mathbb{R}^m, x \mapsto Mx$ for the corresponding m and n. What are m and n? Give vectors $w_1, \ldots, w_n$ such that f is also given by

$$f\big((x_1, x_2, \ldots, x_n)\big) = x_1 w_1 + \cdots + x_n w_n.$$

5.4.3. Determine the matrix M for which $f_M\colon \mathbb{R}^3 \to \mathbb{R}^3$ is reflection in the plane given by $x + 2y - z = 0$.

5.4.4. Given the following linear maps $\mathbb{R}^n \to \mathbb{R}^m$, determine a matrix A such that the map is also given by $x \mapsto Ax$.

(1) $f\colon \mathbb{R}^3 \to \mathbb{R}^4$, $(x, y, z) \mapsto (3x + 2y - z, -x - y + z, x - z, y + z)$,
(2) $g\colon \mathbb{R}^3 \to \mathbb{R}^3$, $(x, y, z) \mapsto (x + 2y - 3z, 2x - y + z, x + y + z)$,
(3) $h\colon \mathbb{R}^3 \to \mathbb{R}^2$, $(x, y, z) \mapsto x\cdot(1, 2) + y\cdot(2, -1) + z\cdot(-1, 3)$,
(4) $j\colon \mathbb{R}^2 \to \mathbb{R}^3$, $v \mapsto (\langle v, w_1\rangle, \langle v, w_2\rangle, \langle v, w_3\rangle)$, with $w_1 = (1, -1)$, $w_2 = (2, 3)$ and $w_3 = (-2, 4)$.

5.5. Addition and multiplication of matrices

We know that $\mathrm{Hom}(F^n, F^m)$ has the structure of a vector space (see Lemma 4.19). We can 'transport' this structure to $\mathrm{Mat}(m \times n, F)$ using the identification (5.4) of matrices and linear maps. The bijection (5.4) then becomes an isomorphism (see Exercise 5.5.11).

Definition 5.14. For $A, B \in \mathrm{Mat}(m \times n, F)$, we define $A + B$ to be the matrix corresponding to the linear map $f_A + f_B$ sending x to $Ax + Bx$. Similarly, for $\lambda \in F$, we define λA to be the matrix corresponding to the linear map λf_A sending x to $\lambda \cdot Ax$, so that $f_{A+B} = f_A + f_B$ and $f_{\lambda A} = \lambda f_A$.

It is a trivial verification to see that $(a_{ij})_{i,j} + (b_{ij})_{i,j} = (a_{ij} + b_{ij})_{i,j}$, that is, that addition of matrices is done coefficient-wise. Similarly, we see easily that $\lambda \cdot (a_{ij})_{i,j} = (\lambda a_{ij})_{i,j}$. With this addition and scalar multiplication, $\mathrm{Mat}(m \times n, F)$ becomes an F-vector space, and it is clear that it is 'the same' as (that is, isomorphic to) F^{mn} — the only difference is the arrangement of the coefficients in an array instead of in a sequence.

By Lemma 4.3, the composition of two linear maps is again linear. How is this reflected in terms of matrices?

Definition 5.15. Let $A \in \mathrm{Mat}(l \times m, F)$ and $B \in \mathrm{Mat}(m \times n, F)$. Then B gives a linear map $f_B \colon F^n \to F^m$, and A gives a linear map $f_A \colon F^m \to F^l$. We define the *product* AB to be the matrix corresponding to the composite linear map $f_A \circ f_B \colon F^n \xrightarrow{f_B} F^m \xrightarrow{f_A} F^l$. So AB will be a matrix in $\mathrm{Mat}(l \times n, F)$.

Remark 5.16. Note that for the product AB to exist, the number of columns of A has to equal the number of rows of B.

By Definition 5.15, the product AB satisfies $f_{AB} = f_A \circ f_B$, so we have

$$(AB)x = f_{AB}(x) = f_A(f_B(x)) = A(Bx) \tag{5.5}$$

for all $x \in F^n$. To express AB in terms of A and B, we let $v_1, v_2, \ldots, v_l$ denote the rows of A and $w_1, w_2, \ldots, w_n$ the columns of B. The relation (5.5) holds in particular for $x = e_k$, the k-th standard vector of F^n. Note that $(AB)e_k$ and Be_k are the k-th column of AB and B, respectively. Since the latter is w_k, we find that the k-th column of AB equals

$$(AB)e_k = A(Be_k) = Aw_k = \begin{pmatrix} \langle v_1, w_k \rangle \\ \langle v_2, w_k \rangle \\ \vdots \\ \langle v_l, w_k \rangle \end{pmatrix}.$$

We conclude

$$AB = \begin{pmatrix} -v_1- \\ -v_2- \\ \vdots \\ -v_l- \end{pmatrix} \begin{pmatrix} | & | & & | \\ w_1 & w_2 & \cdots & w_n \\ | & | & & | \end{pmatrix} = \begin{pmatrix} \langle v_1, w_1 \rangle & \langle v_1, w_2 \rangle & \cdots & \langle v_1, w_n \rangle \\ \langle v_2, w_1 \rangle & \langle v_2, w_2 \rangle & \cdots & \langle v_2, w_n \rangle \\ \vdots & \vdots & & \vdots \\ \langle v_l, w_1 \rangle & \langle v_l, w_2 \rangle & \cdots & \langle v_l, w_n \rangle \end{pmatrix}.$$

In other words, the (i, k)-th entry in the i-th row and the k-th column of the product AB is the scalar product $\langle v_i, w_k \rangle$ of the i-th row of A and the k-th row of B. With

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{l1} & a_{l2} & \cdots & a_{lm} \end{pmatrix} \qquad \text{and} \qquad B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix}$$

we get

$$v_i = (a_{i1}, a_{i2}, \ldots, a_{im}) \qquad \text{and} \qquad w_k = \begin{pmatrix} b_{1k} \\ b_{2k} \\ \vdots \\ b_{mk} \end{pmatrix},$$

so in terms of the entries of A and B, the (i,k)-th entry c_{ik} of the product AB equals

$$c_{ik} = \langle v_i, w_k \rangle = a_{i1}b_{1k} + a_{i2}b_{2k} + \cdots + a_{im}b_{mk} = \sum_{j=1}^{m} a_{ij}b_{jk}\,.$$

If we write the matrix A on the left of AB and the matrix B above AB, then the (i,k)-th entry c_{ik} of AB is the scalar product of the i-th row of A next to this entry and the k-th column of B above the entry.

(5.6)
$$\begin{array}{rcl} & & \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix} = B \\ A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{l1} & a_{l2} & \cdots & a_{lm} \end{pmatrix} & & \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{l1} & c_{l2} & \cdots & c_{ln} \end{pmatrix} = AB \end{array}$$

Example 5.17. To compute the product AB for the matrices

$$A = \begin{pmatrix} 1 & 3 & 5 & 7 \\ 9 & 11 & 13 & 15 \end{pmatrix} \qquad \text{and} \qquad B = \begin{pmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \\ 14 & 16 & 18 \\ 20 & 22 & 24 \end{pmatrix},$$

we write them (on scratch paper) diagonally with respect to each other.

$$\begin{array}{cc} & \begin{pmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \\ 14 & 16 & 18 \\ 20 & 22 & 24 \end{pmatrix} \\ \begin{pmatrix} 1 & 3 & 5 & 7 \\ 9 & 11 & 13 & 15 \end{pmatrix} & \begin{pmatrix} . & 268 & . \\ . & . & . \end{pmatrix} \end{array}$$

The product AB is a matrix with as many rows as A and as many columns as B, so it is a 2×3 matrix. The $(1,2)$-th entry of AB, for instance, is the scalar product of the first row of A and the second column of B, which equals

$$\langle (1,3,5,7), (4,10,16,22) \rangle = 1 \cdot 4 + 3 \cdot 10 + 5 \cdot 16 + 7 \cdot 22 = 268.$$

The other entries are computed similarly and we find

$$AB = \begin{pmatrix} 236 & 268 & 300 \\ 588 & 684 & 780 \end{pmatrix}.$$

Remark 5.18. The identity matrices act as a multiplicative identity:

$$I_m A = A = A I_n \quad \text{for } A \in \mathrm{Mat}(m \times n, F).$$

Remark 5.19. Suppose A is an $m \times n$ matrix and $x \in F^n$ is a vector. If we write x vertically and identify it with an $n \times 1$ matrix, then the product $A \cdot x$ of the matrix A and the vector x, as described in Section 5.3, corresponds with the matrix multiplication described in this section (assuming we also identify Ax with an $m \times 1$ matrix). This is why in this context it is convenient to write both x and Ax vertically.

Proposition 5.20. *The matrix multiplication is associative: for $A \in \mathrm{Mat}(k \times l, F)$ and $B \in \mathrm{Mat}(l \times m, F)$ and $C \in \mathrm{Mat}(m \times n, F)$, we have*

$$A(BC) = (AB)C.$$

Proof. The left-hand side is the unique matrix associated to the composition $f_A \circ (f_B \circ f_C)$, while the right-hand side is the unique matrix associated to the composition $(f_A \circ f_B) \circ f_C$. These composite maps are the same because of associativity of composition. In other words, we have

$$f_{A(BC)} = f_A \circ f_{BC} = f_A \circ (f_B \circ f_C) = (f_A \circ f_B) \circ f_C = f_{AB} \circ f_C = f_{(AB)C},$$

so $A(BC) = (AB)C$ by Proposition 5.11. □

Proposition 5.21. *The matrix multiplication is distributive with respect to addition:*

$$A(B+C) = AB + AC \quad \text{for } A \in \mathrm{Mat}(l \times m, F),\ B, C \in \mathrm{Mat}(m \times n, F);$$
$$(A+B)C = AC + BC \quad \text{for } A, B \in \mathrm{Mat}(l \times m, F),\ C \in \mathrm{Mat}(m \times n, F).$$

Proof. Exercise. □

If A is an $m \times n$ matrix, then for both the product AB and the product BA to exist, the matrix B has to be an $n \times m$ matrix. However, even if AB and BA both exist, we do not necessarily have $AB = BA$. In other words, matrix multiplication is *not* commutative in general. Furthermore, $AB = 0$ (where 0 denotes a *zero matrix* of suitable size) does *not* imply that $A = 0$ or $B = 0$. For a counterexample (to both properties), consider (over a field of characteristic $\neq 2$)

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}.$$

Then

$$AB = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = BA.$$

Definition 5.22. A matrix $A \in \mathrm{Mat}(m \times n, F)$ is called *invertible* if the linear map f_A corresponding to A is an isomorphism. The matrix corresponding to the inverse linear map f_A^{-1} is called the *inverse* of A and is denoted A^{-1}.

Note that the matrix associated to f_A^{-1} is unique by Proposition 5.11, and we have $f_{A^{-1}} = f_A^{-1}$. We will see in Exercise 6.3.5 and Corollary 8.9 that if the $m \times n$ matrix A is invertible, then $m = n$, so A is in fact a square matrix.

Lemma 5.23. *Let A be an $m \times n$ matrix and B an $n \times m$ matrix over F. If $AB = I_m$, then $f_A \colon F^n \to F^m$ is surjective and $f_B \colon F^m \to F^n$ is injective.*

Proof. The composition $f_A \circ f_B = f_{AB} = f_{I_m}$ is the identity and therefore both injective and surjective. It follows that f_A is surjective and f_B is injective. □

Remark 5.24. If matrices A and B satisfy $AB = I_m$ as in Lemma 5.23, then A is called a *left inverse* of B, and B is called a *right inverse* of A.

If A is an invertible $m \times n$ matrix, then we have $AA^{-1} = I_m$ and $A^{-1}A = I_n$, so A^{-1} is both a left and a right inverse of A. The following proposition shows that A^{-1} is uniquely determined by this property.

Proposition 5.25. *A matrix $A \in \mathrm{Mat}(m \times n, F)$ is invertible if and only if there exist matrices B and C such that $AB = I_m$ and $CA = I_n$. Any such matrices (if they **both** exist) satisfy $B = C = A^{-1}$.*

Proof. The "only if" part is obvious, as we can take $B = C = A^{-1}$ if A is invertible. For the "if"-part, suppose that there exist matrices B and C such that $AB = I_m$ and $CA = I_n$. Then by Lemma 5.23, applied to both identities, the linear map $f_A \colon F^n \to F^m$ is both injective and surjective, and therefore an isomorphism, so A is invertible. From $f_A \circ f_B = f_{AB} = \mathrm{id}_{F^m}$ and $f_C \circ f_A = f_{CA} = \mathrm{id}_{F^n}$, we conclude that f_B and f_C are the inverse of f_A, so $B = C = A^{-1}$. □

Proposition 5.26. *Suppose A and B are invertible matrices for which the product AB exists. Then AB is also invertible, and $(AB)^{-1} = B^{-1}A^{-1}$. (Note the reversal of the factors!)*

Proof. Suppose A is an $l \times m$ matrix and B is an $m \times n$ matrix. Then AB is an $l \times n$ matrix. Set $M = B^{-1}A^{-1}$. Then $M(AB) = B^{-1}A^{-1}AB = B^{-1}B = I_n$. We also have $(AB)M = ABB^{-1}A^{-1} = AA^{-1} = I_l$. Hence, M is indeed the inverse of the matrix AB. □

Notation 5.27. Let $A \in \mathrm{Mat}(n, F)$ be a square matrix. For any non-negative integer k, we write A^k for the product $A \cdot A \cdots A$ of k copies of A. If A is invertible and k is a negative integer then A^k denotes the matrix $(A^{-1})^{-k}$.

The usual rule $A^{k+l} = A^k \cdot A^l$ holds for all integers k, l, as long as all these powers are well defined (exercise). Note that because matrix multiplication is not commutative, we do *not* have $(AB)^k = A^k B^k$ in general.

Exercises

5.5.1. If matrices A and B have a product AB that is invertible, does this imply that A and B are invertible? Cf. Exercise 8.4.4.

5.5.2. Prove Proposition 5.21.

5.5.3. Let $A \in \mathrm{Mat}(n, F)$ be a square matrix.

(1) Show that for any non-negative integers k, l we have $A^{k+l} = A^k \cdot A^l$.
(2) Show that if A is invertible, the same holds for any integers k, l.

(3) Show that for every non-negative integer k, we have

$$f_{A^k} = \underbrace{f_A \circ f_A \circ \cdots \circ f_A}_{k}.$$

(4) Show that if A is invertible, the for every negative integer k, we have

$$f_{A^k} = \underbrace{f_A^{-1} \circ f_A^{-1} \circ \cdots \circ f_A^{-1}}_{-k}.$$

5.5.4. Let $\rho\colon \mathbb{R}^2 \to \mathbb{R}^2$ be rotation around 0 over an angle α, cf. Exercise 4.1.7. In Example 5.12 we showed that the matrix

$$A_\alpha = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$$

satisfies $\rho(v) = A_\alpha v$ for all $v \in \mathbb{R}^2$. Show that for all $\alpha, \beta \in \mathbb{R}$ we have

$$\begin{aligned} \cos(\alpha+\beta) &= \cos\alpha\cos\beta - \sin\alpha\sin\beta, \\ \sin(\alpha+\beta) &= \sin\alpha\cos\beta + \cos\alpha\sin\beta. \end{aligned}$$

5.5.5. For which $i, j \in \{1, \dots, 5\}$ does the product of the real matrices A_i and A_j exist and in which order?

$$A_1 = \begin{pmatrix} 1 & 1 & 1 \\ -1 & -2 & -1 \end{pmatrix}, \qquad A_2 = \begin{pmatrix} 2 & -1 & 1 & -4 \\ 3 & -1 & 2 & 4 \end{pmatrix}$$

$$A_3 = \begin{pmatrix} 2 & 3 & 4 \\ -1 & 0 & 2 \\ 3 & 2 & 1 \end{pmatrix}, \quad A_4 = \begin{pmatrix} -1 & -3 \\ 2 & -2 \\ 1 & 1 \end{pmatrix}, \quad A_5 = \begin{pmatrix} 1 & -2 \\ -3 & 2 \end{pmatrix}.$$

Determine those products.

5.5.6. For each $i \in \{1, , \dots, 5\}$, we define the linear map f_i by $x \mapsto A_i x$ with A_i as in Exercise 5.5.5.

(1) What are the domains and codomains of these functions?
(2) Which pairs of these maps can be composed and which product of the matrices belongs to each possible composition?
(3) Is there an order in which you can compose all maps, and if so, which product of matrices corresponds to this composition, and what are its domain and codomain?

5.5.7. Take the linear maps f and g of Exercise 5.4.4 and call the corresponding matrices A and B. In which order can you compose f and g? Write the composition in the same manner that f and g are given by substituting one in the other. Multiply the matrices A and B (in the appropriate order) and verify that this product does indeed correspond with the composition of the linear maps.

5.5.8. Let $A \in \mathrm{Mat}(l \times m, F)$ and $B \in \mathrm{Mat}(m \times n, F)$ be matrices over F. Show that the product $AB \in \mathrm{Mat}(l \times n, F)$ can be described as follows.

(1) The j-th column of AB is the linear combination of the columns of A with the entries of the j-th column of B as coefficients.
(2) The i-th row of AB is the linear combination of the rows of B with the entries of the i-th row of A as coefficients.

5.5.9. Suppose that A, B are matrices for which the product AB exists.

(1) Show that we have $\ker B \subset \ker AB$.
(2) Show that we have $\operatorname{im} AB \subset \operatorname{im} A$.

5.5.10. Give two matrices A and B that are not invertible, for which AB is an identity matrix.

5.5.11. Let F be a field and m, n non-negative integers. Show that the map

$$\mathrm{Mat}(m \times n, F) \to \mathrm{Hom}(F^n, F^m)$$

of (5.4) that sends A to f_A is an isomorphism. (The fact that this map is linear is almost true by definition, as we defined the addition and scalar product of matrices in terms of the addition and scalar product of the functions that are associated to them.)

5.5.12. Let F be a field and m, n non-negative integers. Some of the previous two sections can be summarized by the following diagram.

$$\begin{tikzcd}
(F^n)^m \arrow[r] \arrow[dr] & \mathrm{Mat}(m \times n, F) \arrow[d] & (F^m)^n \arrow[l] \arrow[dl] \\
& \mathrm{Hom}(F^n, F^m) &
\end{tikzcd}$$

Describe a natural isomorphism for each arrow, making the diagram commutative.

5.5.13. (Infinite matrices) As defined, an $m \times n$ matrix over a field F is a map from the set $\{1, 2, \ldots, m\} \times \{1, 2, \ldots, n\}$ to F (sending (i, j) to the (i, j)-th entry of the associated array in row i and column j). In general, for sets X and Y, we define an $X \times Y$ matrix over F to be a map $X \times Y \to F$. In other words, we set $\mathrm{Mat}(X \times Y, F) = \mathrm{Map}(X \times Y, F)$.

(1) Show that for each $M \in \mathrm{Mat}(X \times Y, F)$, there is a linear map
$$f_M \colon F^{(Y)} \to F^X, \qquad g \mapsto \left(x \mapsto \sum_{y \in Y} M(x, y) \cdot g(y) \right).$$
(2) Describe the map above both in terms of "row vectors" and "column vectors" as in Section 5.1, cf. Exercise 4.4.6.
(3) Show that there is an isomorphism
$$\mathrm{Mat}(X \times Y, F) \to \mathrm{Hom}(F^{(Y)}, F^X)$$
that sends a matrix M to the linear map f_M.

Note that, for any set W, two infinite matrices $N \in \mathrm{Mat}(W \times X, F)$ and $M \in \mathrm{Mat}(X \times Y, F)$ can, in general, not be multiplied together, just as the maps $F^{(Y)} \to F^X$ and $F^{(X)} \to F^W$ can not be composed.

5.6. Row space, column space, and transpose of a matrix

The following definition introduces the *transpose* $A^\top$ of a matrix A, which is the matrix we get from A by a 'reflection on the main diagonal.' This associated matrix occurs naturally in many applications, which can often be explained by Exercise 5.6.1.

Definition 5.28. Let $A = (a_{ij}) \in \mathrm{Mat}(m \times n, F)$ be a matrix. The *transpose* of A is the matrix
$$A^\top = (a_{ji})_{1 \le j \le n, 1 \le i \le m} \in \mathrm{Mat}(n \times m, F).$$

Example 5.29. For
$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{pmatrix}$$
we have
$$A^\top = \begin{pmatrix} 1 & 5 & 9 \\ 2 & 6 & 10 \\ 3 & 7 & 11 \\ 4 & 8 & 12 \end{pmatrix}.$$

In the next chapter, we will prove various statements about the rows of matrices. As the columns of A are the rows of $A^\top$, we will be able to use the transpose to conclude the analogs of these statements for columns as well.

Proposition 5.30. *Let F be a field, and l, m, n non-negative integers.*

(1) *For $A, B \in \mathrm{Mat}(m \times n, F)$ we have $(A+B)^\top = A^\top + B^\top$.*
(2) *For $A \in \mathrm{Mat}(m \times n, F)$ and $\lambda \in F$, we have $(\lambda A)^\top = \lambda \cdot A^\top$.*
(3) *For $A \in \mathrm{Mat}(l \times m, F)$ and $B \in \mathrm{Mat}(m \times n, F)$, we have $(AB)^\top = B^\top A^\top$ (note the reversal of factors!).*
(4) *If $A \in \mathrm{Mat}(m \times n, F)$ is invertible, then so is $A^\top$ and we have $(A^\top)^{-1} = (A^{-1})^\top$.*

Proof. The first two statements are obvious. For the third, let $v_1, \ldots, v_l$ be the rows of A and $w_1, \ldots, w_n$ the columns of B. Then the product AB is the $l \times n$ matrix whose (i,k)-th entry is $\langle v_i, w_k \rangle$. The rows of $B^\top$ are $w_1, \ldots, w_n$ and the columns of $A^\top$ are $v_1, \ldots, v_l$, so the (k,i)-th entry of the product $B^\top A^\top$ equals $\langle w_k, v_i \rangle = \langle v_i, w_k \rangle$ as well. This shows that $(AB)^\top = B^\top A^\top$. For a more abstract proof, see Exercise 5.6.1. The fourth statement follows from the third. □

Definition 5.31. The row space $R(A)$ of an $m \times n$ matrix $A \in \mathrm{Mat}(m \times n, F)$ is the subspace of F^n that is generated by the row vectors of A; the column space $C(A)$ is the subspace of F^m generated by the column vectors of A.

Clearly we have $R(A^\top) = C(A)$ and $C(A^\top) = R(A)$ for every matrix A.

Proposition 5.32. *Let $A \in \mathrm{Mat}(m \times n, F)$ be a matrix. Then we have*

$$\begin{aligned} \operatorname{im} A &= C(A) \subset F^m, \\ \ker A &= (R(A))^\perp \subset F^n, \\ \operatorname{im}(A^\top) &= R(A) \subset F^n, \\ \ker(A^\top) &= (C(A))^\perp \subset F^m. \end{aligned}$$

Proof. From (5.3), we see that the image $\operatorname{im} A$ consists of all linear combinations of the columns of A, which proves the first equality.

For the second, let $v_1, \ldots, v_m$ be the rows of A. Then $R(A) = L(v_1, \ldots, v_m)$. The map $f_A \colon F^n \to F^m$ is then given by $f_A(x) = (\langle v_1, x \rangle, \ldots, \langle v_m, x \rangle)$ for all $x \in F^n$ (here we have written $f_A(x)$ normally instead of vertically). Thus, for every $x \in F^n$ we have $f_A(x) = 0$ if and only if $\langle v_i, x \rangle = 0$ for all $1 \leq i \leq m$, so if and only if x is contained in

$$\{v_1, \ldots, v_m\}^\perp = L(v_1, \ldots, v_m)^\perp = (R(A))^\perp$$

(see Proposition 3.33(2)). We conclude $\ker A = (R(A))^\perp$, as stated.

The last equations follow by applying the first two to $A^\top$. □

Remark 5.33. Let $U \subset F^n$ be a subspace of F^n. We can use Proposition 5.32 to reinterpret $U^\perp$. Let U be generated by the vectors $v_1, v_2, \ldots, v_m$.

Let $f\colon F^n \to F^m$ be the linear map given by

$$f(x) = \begin{pmatrix} \langle v_1, x\rangle \\ \langle v_2, x\rangle \\ \vdots \\ \langle v_m, x\rangle \end{pmatrix}.$$

Then the kernel of f equals $U^\perp$. The map f is also given by $x \mapsto Mx$, where M is the $m \times n$ matrix whose i-th row vector is v_i for all $i \leq m$.

Remark 5.34. We have expressed the product AB of matrices A and B in terms of the scalar products of the rows of A and the columns of B. Conversely, we can interpret the scalar product as product of matrices. Suppose we have vectors

$$a = (a_1, a_2, \dots, a_n) \qquad \text{and} \qquad b = (b_1, b_2, \dots, b_n)$$

in F^n. We can think of a and b as $n \times 1$ matrices (implicitly using that F^n and $\mathrm{Mat}(n \times 1, F)$ are isomorphic). Then the transpose $a^\top$ is a $1 \times n$ matrix and the matrix product

$$a^\top \cdot b = \begin{pmatrix} a_1 & a_2 & \dots & a_n \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = (a_1 b_1 + \dots + a_n b_n)$$

is the 1×1 matrix whose single entry equals the scalar product $\langle a, b\rangle$.

Remark 5.35. The vector space F^n is isomorphic to both $\mathrm{Mat}(n \times 1, F)$ and $\mathrm{Mat}(1 \times n, F)$. In this book, as in Remark 5.34, if we implicitly identify a vector $x \in F^n$ with a matrix, it will be identified with an $n \times 1$ matrix, that is, we then write x vertically.

Exercises

5.6.1. Let F be a field and m, n non-negative integers. For each $k \in \{m, n\}$, let $\varphi_k\colon F^k \to \mathrm{Hom}(F^k, F)$ denote the isomorphism that sends the vector $a \in F^k$ to the linear map $(x \mapsto \langle a, x\rangle)$ (see Propositions 4.2 and 4.37 and Exercise 4.4.4). To each linear map $f \in \mathrm{Hom}(F^n, F^m)$, we associate the linear map $f^*\colon \mathrm{Hom}(F^m, F) \to \mathrm{Hom}(F^n, F)$ that sends α to the composition $\alpha \circ f$ (see Exercise 4.5.3), and the linear map $f^\top = \varphi_n^{-1} \circ f^* \circ \varphi_m\colon F^m \to F^n$. (The notation $f^\top$ used in this exercise is not standard.)

$$\begin{tikzcd} \mathrm{Hom}(F^m, F) \arrow[r, "f^*"] & \mathrm{Hom}(F^n, F) \\ F^m \arrow[u, "\varphi_m"] \arrow[r, "f^\top"'] & F^n \arrow[u, "\varphi_n"'] \end{tikzcd}$$

Let A be an $m \times n$ matrix with rows $v_1, \dots, v_m$, and let $f_A\colon F^n \to F^m$ be the associated linear map. Let $j \in \{1, \dots, m\}$.

(1) Show that φ_m sends the j-th standard generator e_j to the projection map $\pi_j\colon F^m \to F$ onto the j-th coordinate.
(2) Show that $f_A^* \circ \varphi_m$ sends e_j to the map $F^n \to F$ that sends $x \in F^n$ to $\langle v_j, x\rangle$.
(3) Show that $f_A^\top$ sends e_j to v_j.

(4) Show that $f_A^\top$ is the map associated to the transpose $A^\top$ of A, that is, $f_A^\top = f_{A^\top}$.

(5) Use Exercise 4.5.5 to prove Proposition 5.30.

5.6.2. Suppose $M \in \mathrm{Mat}(m \times n, F)$ is a matrix and $x \in F^n$ and $y \in F^m$ are vectors. Show that we have

$$\langle Mx, y\rangle = \langle x, M^\top y\rangle.$$

5.6.3. For

$$a = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \qquad \text{and} \qquad b = \begin{pmatrix} -2 \\ 1 \\ 4 \\ 3 \end{pmatrix}$$

compute the matrix products $(a^\top) \cdot b$ and $a \cdot (b^\top)$.

CHAPTER 6

Computations with matrices

Matrices are very suitable for doing computations. Many applications require the understanding of the kernel of some matrix A. For example, the system

$$\left\{\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ & & & & & & & \vdots & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}\right.$$

of linear equations from the beginning of Section 4.3 can be written as $Ax = b$ with

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \in \mathrm{Mat}(m \times n, F) \qquad \text{and} \qquad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \in F^m$$

and the vector

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

of unknowns. The map $f\colon F^n \to F^m$ that was described in the beginning of Section 4.3 is the map $f_A\colon F^n \to F^m$ that sends x to Ax. The solution set equals

$$\{ \, x \in F^n \; : \; Ax = b \, \} = f_A^{-1}(b),$$

and if $a \in F^n$ satisfies $A \cdot a = b$, then by Theorem 4.33 this set equals

$$\{ \, a + z \; : \; z \in \ker A \, \}.$$

In patricular, for $b = 0$, the solution set equals $\ker A$.

If we replace A by a matrix A' that has the same row space, that is, $R(A) = R(A')$, then by Proposition 5.32 we also have $\ker A = \ker A'$. Our goal is to choose A' of a special form (the row echelon form of Section 6.2) that makes it easy to compute the kernel of A' (and thus the kernel of A). Exercise 3.4.9 gives us three operations that we can use on the rows of A to get from A to A' in small steps without changing the row space. These are described in Section 6.1.

As in the previous chapter, we let m and n denote non-negative integers.

6.1. Elementary row and column operations

The main tool for computations with matrices are the so-called 'elementary row and column operations,' described in Definition 6.2. We first give a motivation/analogue in terms of systems of linear equations.

Example 6.1. Consider the system

$$\left\{\begin{array}{rcrcrcl} & - & x_2 & + & x_3 & = & 0 \\ 2x_1 & + & 4x_2 & - & 6x_3 & = & 0 \\ 3x_1 & - & x_2 & - & 2x_3 & = & 0 \end{array}\right.$$

of linear equations over $\mathbb{R}$. To solve the system, we first choose an equation that involves x_1 in order to express x_1 in terms of the other variables, say the second equation. It yields $x_1 = -2x_2 + 3x_3$. We use this to eliminate the variable x_1 from the other equations, either by substituting $-2x_2 + 3x_3$ for x_1, or, equivalently, adding a multiple of the expression $2x_1 + 4x_2 - 6x_3$ from the second equation to the other equations, where the multiple is chosen such that the variable x_1 cancels out. By doing this, the first equation stays the same, that is, $-x_2+x_3 = 0$, while the third gives $-7x_2+7x_3 = 0$. These last two equations are equivalent, but if we had not realised that, we could use the first of these two to eliminate x_2 from the the second: this would give $0 = 0$, showing that the second is indeed trivially satisfied when the first one is. A careful reader checks that this shows not only that any solution of the original system is also a solution of the system

$$\left\{\begin{array}{rcrcrcl} x_1 & + & 2x_2 & - & 3x_3 & = & 0 \\ & & x_2 & - & x_3 & = & 0 \end{array}\right.$$

but also the other way around: every solution to this system is also a solution to the original system. From the second system we can easily describe all solutions. If $x = (x_1, x_2, x_3)$ is a solution, then the second equation allows us to express x_2 in terms of x_3, that is, $x_2 = x_3$, while the first equation lets us express x_1 in terms of x_2 and x_3, namely $x_1 = -2x_2 + 3x_3 = -2x_3 + 3x_3 = x_3$, so we get $x = (x_3, x_3, x_3) = x_3 \cdot (1, 1, 1)$. Hence the solution set is generated by $(1, 1, 1)$.

If we write

$$A = \begin{pmatrix} 0 & -1 & 1 \\ 2 & 4 & -6 \\ 3 & -1 & -2 \end{pmatrix} \qquad \text{and} \qquad x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix},$$

then the original system is equivalent to the homogeneous linear equation $Ax = 0$, so the solution set is $\ker A$. The second system is equivalent to the equation $A'x = 0$ with

$$A' = \begin{pmatrix} 1 & 2 & -3 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}.$$

The steps we took to get from the original system to the second, can be phrased in terms of matrices as follows. First switch the first two rows of A, so that the first row has a nonzero element in the first column. Then multiply the (new) first row by $\frac{1}{2}$, so that this nonzero element becomes 1. After this, subtract appropriate multiples of the first row from the other rows, so that the other rows have a 0 in the first column. This corresponds to eliminating x_1 from the other equations. After these steps, we obtain the matrix

$$\begin{pmatrix} 1 & 2 & -3 \\ 0 & -1 & 1 \\ 0 & -7 & 7 \end{pmatrix}.$$

We then leave the first row as is, and choose one of the other rows that has a nonzero element in the second column, say the second row. We scale it so that its first nonzero element is 1, and then subtract appropriate multiples of it from the rows below it (only the third row remains) in order to create zeros in the second column of those rows. After doing this, we do indeed obtain the matrix A'.

Note that also finding the solution set can be done as easily as above, as the shape of A' makes it easy to determine generators for its kernel as follows. The nonzero rows have a 1 as their first nonzero coordinate, and from bottom to top, each nonzero row can be used to express the coordinate corresponding to the column that contains this 1, in terms of the later coordinates. So for $x = (x_1, x_2, x_3) \in \ker A'$ we find again $x_2 = x_3$ and $x_1 = -2x_2 + 3x_3 = x_3$.

The example above uses so-called *Gaussian elimination* to solve a system of linear equations. In Section 8.5 we will see how to do this in general. In terms of matrices, we used three elementary operations that we now define.

Definition 6.2. Let A be a matrix with entries in a field F. We say that we perform an *elementary row operation* on A, if we

(1) multiply a row of A by some $\lambda \in F \setminus \{0\}$, or
(2) add a scalar multiple of a row of A to another (*not* the same) row of A, or
(3) interchange two rows of A.

We call two matrices A and A' *row equivalent* if A' can be obtained from A by a sequence of elementary row operations.

Note that the third type of operation is redundant, since it can be achieved by a sequence of operations of the first two types (exercise).

Let F be a field and m a positive integer. Let E_{ij} be the $m \times m$ matrix over F of which the only nonzero entry is a 1 in row i and column j. For $1 \leq i, j \leq m$ with $i \neq j$ and $\lambda \in F$, we define the elementary $m \times m$ matrices

$$\begin{aligned} L_i(\lambda) &= I_m + (\lambda - 1)E_{ii}, \\ M_{ij}(\lambda) &= I_m + \lambda E_{ij}, \\ N_{ij} &= I_m + E_{ij} + E_{ji} - E_{ii} - E_{jj}. \end{aligned}$$

One easily verifies that if A is an $m \times n$ matrix, then multiplying the i-th row of A by λ amounts to replacing A by $L_i(\lambda) \cdot A$, while adding λ times the j-th row of A to the i-th row of A amounts to replacing A by $M_{ij}(\lambda) \cdot A$ and switching the i-th and the j-th row amounts to replacing A by $N_{ij} \cdot A$.

The elementary matrices are invertible, which corresponds to the fact that all elementary row operations are invertible by an elementary row operation of the same type. Indeed, we have

$$L_i(\lambda) \cdot L_i(\lambda^{-1}) = I_m, \qquad M_{ij}(\lambda) \cdot M_{ij}(-\lambda) = I_m, \qquad \text{and} \qquad N_{ij}^2 = I_m.$$

This implies that row equivalence is indeed an equivalence.

We define *elementary column operations* and *column equivalence* in a similar way, replacing the word 'row' by 'column' each time it appears. While each row operation on a matrix $A \in \mathrm{Mat}(m \times n, F)$ corresponds to multiplying A by an elementary $m \times m$ matrix M from the left, yielding MA, each column operation corresponds to multiplying A by an elementary $n \times n$ matrix N from the right, yielding AN.

The following proposition shows that the elementary row operations do not change the row space and the kernel of a matrix.

Proposition 6.3. *If A and A' are row equivalent matrices, then we have*

$$R(A) = R(A') \qquad \text{and} \qquad \ker A = \ker A'.$$

Proof. Exercise 6.1.1. ☐

Proposition 6.4. *Suppose A and A' are row equivalent $m \times n$ matrices. If A' is obtained from A by a certain sequence of elementary row operations, then there is an invertible $m \times m$ matrix B, depending only on the sequence, such that $A' = BA$. Similarly, if A and A' are column equivalent, then there is an invertible $n \times n$ matrix C such that $A' = AC$.*

Proof. Let $A \in \mathrm{Mat}(m \times n, F)$. Let $B_1, B_2, \ldots, B_r$ be the elementary matrices corresponding to the row operations we have performed (in that order) on A to obtain A', then

$$A' = B_r\Big(B_{r-1}\cdots\big(B_2(B_1A)\big)\cdots\Big) = (B_rB_{r-1}\cdots B_2B_1)A\,,$$

and $B = B_rB_{r-1}\cdots B_2B_1$ is invertible as it is a product of invertible matrices. The statement on column operations is proved in the same way, or by applying the result on row operations to the transpose $A^\top$. ☐

Proposition 6.5. *Suppose $A \in \mathrm{Mat}(m \times n, F)$ is a matrix. Let A' be a matrix obtained from A by applying a sequence of elementary row and column operations. Then the following are true.*

(1) *If the sequence contains only row operations, then there is an isomorphism $\psi\colon F^m \to F^m$, depending only on the sequence, with $f_{A'} = \psi \circ f_A$.*
(2) *If the sequence contains only column operations, then there is an isomorphism $\varphi\colon F^n \to F^n$, depending only on the sequence, with $f_{A'} = f_A \circ \varphi$.*
(3) *There exist an isomorphism $\varphi\colon F^n \to F^n$, depending only on the subsequence of column operations, and an isomorphism $\psi\colon F^m \to F^m$, depending only on the subsequence of row operations, with $f_{A'} = \psi \circ f_A \circ \varphi$, so that the diagram*

$$\begin{tikzcd} F^n \arrow[r, "f_A"] & F^m \arrow[d, "\psi"] \\ F^n \arrow[u, "\varphi"] \arrow[r, "f_{A'}"'] & F^m \end{tikzcd}$$

is commutative.

Proof. Exercise. ☐

Corollary 6.6. *Let A and A' be row equivalent matrices. Then f_A is injective if and only if $f_{A'}$ is injective, and f_A is surjective if and only if $f_{A'}$ is surjective.*

Proof. By Proposition 6.5 there is an isomorphism ψ with $f_{A'} = \psi \circ f_A$. Indeed, the composition is surjective or injective if and only if f_A is, cf. Proposition 4.41. ☐

Exercises

6.1.1. (1) Let $v_1, v_2, \ldots, v_m \in \mathbb{R}^n$ be m vectors and consider the $m \times n$ matrix A whose rows are these vectors. Let A' be a matrix that is obtained from A by an elementary row operation. Show that for the rows $v_1', v_2', \ldots, v_m'$ of A' we have $L(v_1, \ldots, v_m) = L(v_1', \ldots, v_m')$ (cf. Exercise 3.4.9).
(2) Prove Proposition 6.3 (use Proposition 5.32).

6.1.2. Show that column equivalent matrices have the same column space, cf. Proposition 6.3.

6.1.3. In the following sequence of matrices, each is obtained from the previous by one or two elementary row operations. Find, for each $1 \leq i \leq 9$, a matrix B_i such that $A_i = B_i A_{i-1}$. Also find a matrix B such that $A_9 = BA_0$. You may write B as a product of other matrices without actually performing the multiplication.

$$A_0 = \begin{pmatrix} 2 & 5 & 4 & -3 & 1 \\ 1 & 3 & -2 & 2 & 1 \\ 0 & 4 & -1 & 0 & 3 \\ -1 & 2 & 2 & 3 & 1 \end{pmatrix} \qquad A_1 = \begin{pmatrix} 1 & 3 & -2 & 2 & 1 \\ 2 & 5 & 4 & -3 & 1 \\ 0 & 4 & -1 & 0 & 3 \\ -1 & 2 & 2 & 3 & 1 \end{pmatrix}$$

$$A_2 = \begin{pmatrix} 1 & 3 & -2 & 2 & 1 \\ 0 & -1 & 8 & -7 & -1 \\ 0 & 4 & -1 & 0 & 3 \\ 0 & 5 & 0 & 5 & 2 \end{pmatrix} \qquad A_3 = \begin{pmatrix} 1 & 3 & -2 & 2 & 1 \\ 0 & -1 & 8 & -7 & -1 \\ 0 & 0 & 31 & -28 & -1 \\ 0 & 0 & 40 & -30 & -3 \end{pmatrix}$$

$$A_4 = \begin{pmatrix} 1 & 3 & -2 & 2 & 1 \\ 0 & -1 & 8 & -7 & -1 \\ 0 & 0 & 31 & -28 & -1 \\ 0 & 0 & 9 & -2 & -2 \end{pmatrix} \qquad A_5 = \begin{pmatrix} 1 & 3 & -2 & 2 & 1 \\ 0 & -1 & 8 & -7 & -1 \\ 0 & 0 & 4 & -22 & 5 \\ 0 & 0 & 9 & -2 & -2 \end{pmatrix}$$

$$A_6 = \begin{pmatrix} 1 & 3 & -2 & 2 & 1 \\ 0 & -1 & 8 & -7 & -1 \\ 0 & 0 & 4 & -22 & 5 \\ 0 & 0 & 1 & 42 & -12 \end{pmatrix} \qquad A_7 = \begin{pmatrix} 1 & 3 & -2 & 2 & 1 \\ 0 & -1 & 8 & -7 & -1 \\ 0 & 0 & 1 & 42 & -12 \\ 0 & 0 & 4 & -22 & 5 \end{pmatrix}$$

$$A_8 = \begin{pmatrix} 1 & 3 & -2 & 2 & 1 \\ 0 & -1 & 8 & -7 & -1 \\ 0 & 0 & 1 & 42 & -12 \\ 0 & 0 & 0 & -190 & 53 \end{pmatrix} \qquad A_9 = \begin{pmatrix} 1 & 3 & -2 & 2 & 1 \\ 0 & 1 & -8 & 7 & 1 \\ 0 & 0 & 1 & 42 & -12 \\ 0 & 0 & 0 & 190 & -53 \end{pmatrix}$$

6.1.4. Show that row operations commute with column operations. In other words, if A is a matrix and A' is the matrix obtained from A by first applying a certain row operation and then a certain column operation, then applying the two operations in the opposite order to A yields the same matrix A'.

6.1.5. Prove Proposition 6.5.

6.1.6. Is Corollary 6.6 also true for column equivalent matrices A and A'? What about matrices A and A' that can be obtained from each other by a sequence of row *or* column operations?

6.2. Row echelon form

If we want to find generators for the kernel of an $m \times n$ matrix A or, equivalently, its associated linear map $f_A \colon F^n \to F^m$, then according to Proposition 6.3 we may replace A by any row equivalent matrix.

Example 6.7. We want generators for the kernel of the real matrix

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

We leave it to the reader to check that A is row equivalent to the matrix

$$A' = \begin{pmatrix} 1 & 0 & 3 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

(Start by multiplying the first row of A by -1 to obtain $v_1 = (1, -2, -1, -1)$ as first row and subtract v_1 and $2v_1$ from the second and third row, respectively.) Hence $\ker A = \ker A'$ by Proposition 6.3. Suppose $x = (x_1, x_2, x_3, x_4) \in \ker A'$. Then we have

$$A'x = \begin{pmatrix} 1 & 0 & 3 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} x_1 + 3x_3 \\ x_2 + 2x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

This yields three equations, namely

$$\begin{aligned} x_1 + 3x_3 &= 0, \\ x_2 + 2x_3 &= 0, \\ x_4 &= 0. \end{aligned}$$

It follows (using these equations from bottom to top) that $x_4 = 0$ and $x_2 = -2x_3$ and $x_1 = -3x_3$, so $x = x_3 \cdot (-3, -2, 1, 0)$. Hence, the vector $(-3, -2, 1, 0)$ generates the kernels of A' and A.

The matrix A' of Example 6.1 and the matrix A' of Example 6.7 have a shape that makes it easy to find generators for their kernels. Indeed, for $x \in \mathbb{R}^3$ or $x \in \mathbb{R}^4$ (for Example 6.1 and 6.7, respectively), each of the rows of A' gives a linear equation in the coordinates of x; the shape of A' allows us to use the i-th row (if it is nonzero) to express one of the coordinates, say x_{j_i}, in terms of other coordinates, while the equations associated to all lower rows do not involve this coordinate x_{j_i}.

This description of the shape of the two matrices A' of Example 6.1 and Example 6.7 shows that they are in row echelon form, as defined in Definition 6.8. In this section we will explain how to find, for any matrix A, a matrix in row echelon form that is row equivalent to A. In the next section we will see in full generality how to obtain generators for the kernel from the row echelon form.

Definition 6.8. A matrix is said to be in *row echelon form* when its nonzero rows (if they exist) are on top and its zero rows (if they exist) on the bottom and, moreover, the first nonzero entry in each nonzero row, the so-called *pivot* of that row, is farther to the right than the pivots in the rows above.[1]

Example 6.9. The matrix A_9 of Exercise 6.1.3 is in row echelon form. The following real matrices are all in row echelon form as well, with the last one

[1]Some books require the pivots to be equal to 1 for a matrix to be in row echelon form. We do not require this.

describing the most general shape with all pivots equal to 1.

$$\begin{pmatrix} 1 & 4 & -2 & 4 & 3 \\ 0 & 2 & 7 & 2 & 5 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & 4 & -2 & 4 \\ 0 & 5 & 7 & 2 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\begin{array}{c c} \begin{array}{c} 1 \\ 2 \\ \vdots \\ r \\ r+1 \\ \vdots \\ m \end{array} & \begin{pmatrix} 0\cdots 0 & 1 & *\cdots * & * & *\cdots * & * & *\cdots * \\ 0\cdots 0 & 0 & 0\cdots 0 & 1 & *\cdots * & * & *\cdots * \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0\cdots 0 & 0 & 0\cdots 0 & 0 & 0\cdots 0 & 1 & *\cdots * \\ 0\cdots 0 & 0 & 0\cdots 0 & 0 & 0\cdots 0 & 0 & 0\cdots 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0\cdots 0 & 0 & 0\cdots 0 & 0 & 0\cdots 0 & 0 & 0\cdots 0 \end{pmatrix} \\ & \begin{array}{ccccccc} & j_1 & & j_2 & \cdots & j_r & \end{array} \end{array}$$

To make the matrix $A = (a_{ij})_{i,j}$ in most general shape with all pivots equal to 1 more precise, note that there are integers $0 \leq r \leq m$ and $1 \leq j_1 < j_2 < \cdots < j_r \leq n$ where r is the number of nonzero rows and, for each $1 \leq i \leq r$, the number j_i denotes the column of the pivot in row i, so that we have $a_{ij} = 0$ if $i > r$ or ($i \leq r$ and $j < j_i$), and we have $a_{ij_i} = 1$ for $1 \leq i \leq r$.

Proposition 6.11 shows how every matrix can be brought into row echelon form by a sequence of elementary row operations, following the ideas of Example 6.1. The following example demonstrates all the required steps.

Example 6.10. Consider the matrix

$$A = \begin{pmatrix} 0 & 0 & -2 & 4 & 3 & -3 \\ 0 & 1 & 0 & 2 & 2 & -3 \\ 0 & 2 & 3 & -2 & 0 & -1 \\ 0 & 2 & 0 & 4 & 0 & -10 \end{pmatrix}.$$

The first column has no nonzero entries, so we look at the next column. We pick a row in which the second column contains a nonzero element, say the last row. We switch that row with the first to obtain the matrix

$$\begin{array}{c} R_4 \\ R_2 \\ R_3 \\ R_1 \end{array} \begin{pmatrix} 0 & 2 & 0 & 4 & 0 & -10 \\ 0 & 1 & 0 & 2 & 2 & -3 \\ 0 & 2 & 3 & -2 & 0 & -1 \\ 0 & 0 & -2 & 4 & 3 & -3 \end{pmatrix}.$$

Note that we have indicated how the rows of this matrix depend on the rows of the previous matrix. We now scale the first row to make its pivot equal to 1, that is, we multiply it by $\frac{1}{2}$. This yields

$$\begin{array}{c} \frac{1}{2}R_1 \\ R_2 \\ R_3 \\ R_4 \end{array} \begin{pmatrix} 0 & 1 & 0 & 2 & 0 & -5 \\ 0 & 1 & 0 & 2 & 2 & -3 \\ 0 & 2 & 3 & -2 & 0 & -1 \\ 0 & 0 & -2 & 4 & 3 & -3 \end{pmatrix}.$$

To make the entries under the pivot in the first row zero, we subtract the first row from the second row and twice the first row from the third row. This gives

$$\begin{array}{r}R_1\\R_2-R_1\\R_3-2R_1\\R_4\end{array}\begin{pmatrix}0&1&0&2&0&-5\\0&0&0&0&2&2\\0&0&3&-6&0&9\\0&0&-2&4&3&-3\end{pmatrix}.$$

We now leave this first row as is. We proceed to the next (third) column, and choose a row in which the corresponding element is nonzero, say the third row. We switch it with the second row to obtain

$$\begin{array}{r}R_1\\R_3\\R_2\\R_4\end{array}\begin{pmatrix}0&1&0&2&0&-5\\0&0&3&-6&0&9\\0&0&0&0&2&2\\0&0&-2&4&3&-3\end{pmatrix}.$$

We scale the new second row such that its pivot becomes 1, that is, we multiply it by $\frac{1}{3}$, which yields

$$\begin{array}{r}\\\frac{1}{3}R_2\\\\\\\end{array}\begin{pmatrix}0&1&0&2&0&-5\\0&0&1&-2&0&3\\0&0&0&0&2&2\\0&0&-2&4&3&-3\end{pmatrix}.$$

Note that we no longer indicate it if a row is unchanged from the previous matrix. To make the entries under this pivot zero, we add twice the second row to the last row. This gives

$$\begin{array}{r}\\\\\\R_4+2R_2\end{array}\begin{pmatrix}0&1&0&2&0&-5\\0&0&1&-2&0&3\\0&0&0&0&2&2\\0&0&0&0&3&3\end{pmatrix}.$$

We leave the second row as is, and proceed with the next column. This column has no nonzero elements in the remaining two rows, so we immediately continue with the next (fifth) column. The third row already contains a nonzero element in the fifth column, so we just scale this row to make the pivot 1, that is, we multiply it by $\frac{1}{2}$, and we obtain

$$\begin{array}{r}\\\\\frac{1}{2}R_3\\\\\end{array}\begin{pmatrix}0&1&0&2&0&-5\\0&0&1&-2&0&3\\0&0&0&0&1&1\\0&0&0&0&3&3\end{pmatrix}.$$

To make the entries under this third pivot zero, we subtract three times the third row from the last and get

$$\begin{array}{r}\\\\\\R_4-3R_3\end{array}\begin{pmatrix}0&1&0&2&0&-5\\0&0&1&-2&0&3\\0&0&0&0&1&1\\0&0&0&0&0&0\end{pmatrix}=A'.$$

Since the remaining rows (only the fourth is left) are all zero, we are done. Indeed, A' is in row echelon form.

The following procedure describes precisely how to bring a matrix into row echelon form in general. The input is a matrix A and the output is a matrix in row echelon form that is row equivalent to A. This algorithm is the key to most computations with matrices. It makes all pivots equal to 1.

Proposition 6.11 (Row Echelon Form Algorithm). *Let $A \in \mathrm{Mat}(m \times n, F)$ be a matrix. The following procedure applies successive elementary row operations to A and transforms it into a matrix A' in row echelon form.*

1. *Set $A' = A$, $r = 0$ and $j_0 = 0$. Write $A' = (a'_{ij})_{i,j}$.*
2. *[At this point, $a'_{ij} = 0$ if ($i > r$ and $j \leq j_r$) or ($1 \leq i \leq r$ and $1 \leq j < j_i$). Also, $a'_{ij_i} = 1$ for $1 \leq i \leq r$.]*
   If the $(r+1)$st up to the m-th rows of A' are zero, then stop.
3. *Find the smallest j such that there is some $a'_{ij} \neq 0$ with $r < i \leq m$. Replace r by $r + 1$, set $j_r = j$, and interchange the r-th and the i-th row of A' if $r \neq i$. Note that $j_r > j_{r-1}$.*
4. *Multiply the r-th row of A' by $(a'_{rj_r})^{-1}$.*
5. *For each $i = r + 1, \ldots, m$, add $-a'_{ij_r}$ times the r-th row of A' to the i-th row of A'.*
6. *Go to Step 2.*

Proof. The only changes that are done to A' are elementary row operations of the third, first and second kinds in steps 3, 4 and 5, respectively. Since in each pass through the loop, r increases, and we have to stop when $r = m$, the procedure certainly terminates. We have to show that when it stops, A' is in row echelon form.

We check that the claim made at the beginning of step 2 is always correct. It is trivially satisfied when we reach step 2 for the first time. We now assume it is correct when we are in step 2 and show that it is again true when we come back to step 2. Since the first r rows are not changed in the loop, the part of the statement referring to them is not affected. In step 3, we increase r and find j_r (for the new r) such that $a'_{ij} = 0$ if $i \geq r$ and $j < j_r$. By our assumption, we must have $j_r > j_{r-1}$. The following actions in steps 3 and 4 have the effect of producing an entry with value 1 at position (r, j_r). In step 5, we achieve that $a'_{ij_r} = 0$ for $i > r$. So $a'_{ij} = 0$ when ($i > r$ and $j \leq j_r$) and when ($i = r$ and $j < j_r$). This shows that the condition in step 2 is again satisfied.

So at the end of the algorithm, the statement in step 2 is true. Also, we have seen that $0 < j_1 < j_2 < \cdots < j_r$, hence A' has row echelon form when the procedure is finished. $\square$

Example 6.12. Consider the following real matrix.

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

Let us bring it into row echelon form.

Since the upper left entry is nonzero, we have $j_1 = 1$. We subtract 4 times the first row from the second and 7 times the first row from the third. This leads to

$$A' = \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -6 & -12 \end{pmatrix}.$$

Now we divide the second row by -3 and then add 6 times the new second row to the third. This gives

$$A'' = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix},$$

which is in row echelon form.

Example 6.13. In Examples 6.1, 6.7, and 6.10, the matrix A' is a matrix in row echelon form that is row equivalent to A.

Remark 6.14. The row space of A in Example 6.12 is spanned by its three rows. By Proposition 6.3, the row spaces of A and A'' are the same, so this space is also spanned by the two nonzero rows of A''. We will see in the next chapter that the space can not be generated by fewer elements. More generally, the number of nonzero rows in a matrix in row echelon form is the minimal number of vectors needed to span its row space (see Theorem 7.47 and Proposition 8.14).

Example 6.15 (Avoiding denominators)**.** The algorithm above may introduce more denominators than needed. For instance, it transforms the matrix

$$\begin{pmatrix} 22 & 5 \\ 9 & 2 \end{pmatrix}$$

in two rounds as

$$\begin{pmatrix} 22 & 5 \\ 9 & 2 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & \frac{5}{22} \\ 0 & -\frac{1}{22} \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & \frac{5}{22} \\ 0 & 1 \end{pmatrix}.$$

Instead of immediately dividing the first row by 22, we could first subtract a multiple of the second row from the first. We can continue to decrease the numbers in the first column by adding multiples of one row to the other. Eventually we end up with a 1 in the column, or, in general, with the greatest common divisor of the numbers involved.

$$\begin{pmatrix} 22 & 5 \\ 9 & 2 \end{pmatrix} \rightsquigarrow \begin{array}{r} R_1 - 2R_2 \\ R_2 \end{array} \begin{pmatrix} 4 & 1 \\ 9 & 2 \end{pmatrix} \rightsquigarrow \begin{array}{r} R_1 \\ R_2 - 2R_1 \end{array} \begin{pmatrix} 4 & 1 \\ 1 & 0 \end{pmatrix}$$
$$\rightsquigarrow \begin{array}{r} R_2 \\ R_1 \end{array} \begin{pmatrix} 1 & 0 \\ 4 & 1 \end{pmatrix} \rightsquigarrow \begin{array}{r} R_1 \\ R_2 - 4R_1 \end{array} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

We see that the 2×2 identity matrix is also a row echelon form for the original matrix.

Note that in Example 6.15 we indicated the row operations by writing on the left of each row of a matrix, the linear combination of the rows of the previous matrix that this row is equal to. This is necessary, because we do not follow the deterministic algorithm.

If you like living on the edge and taking risks, then you could write down the result of several row operations as one step, as long as you make sure it is the result of doing the operations one *after* another, not at the same time. For example, by applying the appropriate sequence of switching two rows, you can get any permutation of the rows. (Can you prove this?) You can also take one row v_i and add multiples of it to various *other* rows, as long as you keep v_i as a row in the new matrix. That way you only make steps that are reversible, either by permuting the rows back, or by subtracting the appropriate multiples of v_i from the other rows.

Warning 6.16. Make sure you do not accidentally perform two operations at the same time! If you start with the matrix

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 5 \end{pmatrix}$$

and add the first row to the second, while at the same time adding the second row to the first, you end up with

$$\begin{pmatrix} 4 & 7 \\ 4 & 7 \end{pmatrix}$$

which clearly does not have the same row space as A, so something has gone wrong. If you do it right, and first add the first row to the second, and then the second row to the first, we get

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 5 \end{pmatrix} \rightsquigarrow \quad \begin{array}{r} R_1 \\ R_2 + R_1 \end{array} \begin{pmatrix} 1 & 2 \\ 4 & 7 \end{pmatrix} \rightsquigarrow \quad \begin{array}{r} R_1 + R_2 \\ R_2 \end{array} \begin{pmatrix} 5 & 9 \\ 4 & 7 \end{pmatrix}.$$

What do you get if you perform these two operations in the opposite order?

We give one more example, where we avoid denominators all the way, except for the last step.

Example 6.17.

$$\begin{pmatrix} 3 & 5 & 2 & 2 \\ 1 & 3 & -4 & 3 \\ 2 & -2 & 5 & -1 \\ -1 & 3 & 1 & -3 \end{pmatrix} \rightsquigarrow \quad \begin{array}{r} R_2 \\ R_1 \\ R_3 \\ R_4 \end{array} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 3 & 5 & 2 & 2 \\ 2 & -2 & 5 & -1 \\ -1 & 3 & 1 & -3 \end{pmatrix}$$

$$\rightsquigarrow \quad \begin{array}{r} R_1 \\ R_2 - 3R_1 \\ R_3 - 2R_1 \\ R_4 + R_1 \end{array} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 0 & -4 & 14 & -7 \\ 0 & -8 & 13 & -7 \\ 0 & 6 & -3 & 0 \end{pmatrix} \rightsquigarrow \quad \begin{array}{r} R_1 \\ R_2 \\ R_3 \\ R_4 + R_2 \end{array} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 0 & -4 & 14 & -7 \\ 0 & -8 & 13 & -7 \\ 0 & 2 & 11 & -7 \end{pmatrix}$$

$$\rightsquigarrow \quad \begin{array}{r} R_1 \\ R_4 \\ R_3 \\ R_2 \end{array} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 0 & 2 & 11 & -7 \\ 0 & -8 & 13 & -7 \\ 0 & -4 & 14 & -7 \end{pmatrix} \rightsquigarrow \quad \begin{array}{r} R_1 \\ R_2 \\ R_3 + 4R_2 \\ R_4 + 2R_2 \end{array} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 0 & 2 & 11 & -7 \\ 0 & 0 & 57 & -35 \\ 0 & 0 & 36 & -21 \end{pmatrix}$$

$$\rightsquigarrow \quad \begin{array}{r} R_1 \\ R_2 \\ R_3 - R_4 \\ R_4 \end{array} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 0 & 2 & 11 & -7 \\ 0 & 0 & 21 & -14 \\ 0 & 0 & 36 & -21 \end{pmatrix} \rightsquigarrow \quad \begin{array}{r} R_1 \\ R_2 \\ R_3 \\ R_4 - R_3 \end{array} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 0 & 2 & 11 & -7 \\ 0 & 0 & 21 & -14 \\ 0 & 0 & 15 & -7 \end{pmatrix}$$

$$\rightsquigarrow \quad \begin{array}{r} R_1 \\ R_2 \\ R_3 - R_4 \\ R_4 \end{array} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 0 & 2 & 11 & -7 \\ 0 & 0 & 6 & -7 \\ 0 & 0 & 15 & -7 \end{pmatrix} \rightsquigarrow \quad \begin{array}{r} R_1 \\ R_2 \\ R_3 \\ R_4 - 2R_3 \end{array} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 0 & 2 & 11 & -7 \\ 0 & 0 & 6 & -7 \\ 0 & 0 & 3 & 7 \end{pmatrix}$$

$$\rightsquigarrow \quad \begin{array}{r} R_1 \\ R_2 \\ R_4 \\ R_3 \end{array} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 0 & 2 & 11 & -7 \\ 0 & 0 & 3 & 7 \\ 0 & 0 & 6 & -7 \end{pmatrix} \rightsquigarrow \quad \begin{array}{r} R_1 \\ R_2 \\ R_3 \\ R_4 - 2R_3 \end{array} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 0 & 2 & 11 & -7 \\ 0 & 0 & 3 & 7 \\ 0 & 0 & 0 & -21 \end{pmatrix}$$

$$\rightsquigarrow \begin{array}{r} R_1 \\ \frac{1}{2}R_2 \\ \frac{1}{3}R_3 \\ -\frac{1}{21}R_4 \end{array} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 0 & 1 & \frac{11}{2} & -\frac{7}{2} \\ 0 & 0 & 1 & \frac{7}{3} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Exercises

6.2.1. Find a row echelon form for each of the matrices in Exercise 5.5.5.

6.3. Generators for the kernel

If we want to compute generators for the kernel of a matrix $A \in \mathrm{Mat}(m \times n, F)$, then, according to Proposition 6.3, we may replace A by any row equivalent matrix. In particular, it suffices to understand how to determine generators for the kernel of matrices in row echelon form. We start with an example.

Example 6.18. Suppose M is the matrix (over $\mathbb{R}$)

$$\begin{pmatrix} 1 & 2 & -1 & 0 & 2 & 1 & -3 \\ 0 & 0 & 1 & -1 & 2 & -1 & 2 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

which is already in row echelon form with its pivots circled. Let v_1, v_2, v_3 denote its nonzero rows, which generate the row space $R(M)$. Suppose the vector $x = (x_1, x_2, x_3, x_4, x_5, x_6, x_7)$ is contained in

$$\ker M = R(M)^\perp = \{x \in \mathbb{R}^7 \;:\; \langle v_i, x\rangle = 0 \text{ for } i = 1, 2, 3\}.$$

Then the coordinates x_1, x_3, x_5, which belong to the columns with a pivot, are uniquely determined by the coordinates x_2, x_4, x_6, x_7, which belong to the columns without a pivot. Indeed, starting with the lowest nonzero row, the equation $\langle v_3, x\rangle = 0$ gives $x_5 + x_6 + x_7 = 0$, so

$$x_5 = -x_6 - x_7.$$

The equation $\langle v_2, x\rangle = 0$ then gives $x_3 - x_4 + 2x_5 - x_6 + 2x_7$, so

$$x_3 = x_4 - 2(-x_6 - x_7) + x_6 - 2x_7 = x_4 + 3x_6.$$

Finally, the equation $\langle v_1, x\rangle = 0$ gives

$$x_1 = -2x_2 + (x_4 + 3x_6) - 2(-x_6 - x_7) - x_6 + 3x_7 = -2x_2 + x_4 + 4x_6 + 5x_7.$$

Moreover, any choice for the values x_2, x_4, x_6, x_7, with these corresponding values for x_1, x_3, x_5, does indeed give an element of the kernel $\ker M$, as the equations $\langle v_i, x\rangle = 0$ for $1 \le i \le 3$ are automatically satisfied. With $q = x_2$, $r = x_4$,

$s = x_6$, and $t = x_7$, we may write

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{pmatrix} = \begin{pmatrix} -2q + r + 4s + 5t \\ q \\ r + 3s \\ r \\ -s - t \\ s \\ t \end{pmatrix} = q \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + r \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + s \begin{pmatrix} 4 \\ 0 \\ 3 \\ 0 \\ -1 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 5 \\ 0 \\ 0 \\ 0 \\ -1 \\ 0 \\ 1 \end{pmatrix}$$
$$= qw_2 + rw_4 + sw_6 + tw_7,$$

where

$$w_2 = \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad w_4 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad w_6 = \begin{pmatrix} 4 \\ 0 \\ 3 \\ 0 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \quad w_7 = \begin{pmatrix} 5 \\ 0 \\ 0 \\ 0 \\ -1 \\ 0 \\ 1 \end{pmatrix}.$$

This shows that the kernel $\ker M$ is generated by w_2, w_4, w_6, w_7, that is, we have $\ker M = L(w_2, w_4, w_6, w_7)$. In each w_k, we circled the coordinates that correspond to the columns of M with a pivot. Note that the non-circled coordinates in each w_k are all 0, except for one, the k-th coordinate, which equals 1. Conversely, for each of the columns of M without pivot, there is exactly one w_k with 1 for the (non-circled) coordinate corresponding to that column and 0 for all other coordinates belonging to a column without a pivot.

This could also be used to find w_2, w_4, w_6, w_7 directly: choose any column without a pivot, say the k-th, and set the k-th coordinate of a vector $w \in \mathbb{R}^7$ equal to 1, then set all other coordinates corresponding to columns without pivot equal to 0, and compute the remaining coordinates. For instance, for the sixth column, which has no pivot, we get a vector w of which the sixth entry is 1, and all other entries corresponding to columns without pivots are 0, that is,

$$w = \begin{pmatrix} * \\ 0 \\ * \\ 0 \\ * \\ 1 \\ 0 \end{pmatrix}.$$

The entries that correspond to columns with a pivot (so the first, third, and fifth) can now be computed using the equations $\langle v_i, w \rangle = 0$, starting with $i = 3$ and going down to $i = 1$. We find $w = w_6$ in this example.

The following proposition states that we can find generators for the kernel of any matrix in row echelon form in the same manner. In Proposition 7.22 we will see that the generators constructed in Proposition 6.19 actually form a so-called basis of the kernel.

Proposition 6.19. *Let $A \in \mathrm{Mat}(m \times n, F)$ be a matrix in row echelon form with r nonzero rows and let $j_1 < j_2 < \ldots < j_r$ be the indices of the columns with a pivot. Then for each $1 \leq k \leq n$ with $k \notin \{j_1, j_2, \ldots, j_r\}$, there is a unique vector $w_k \in \ker A$ such that*

1. *the k-th entry of w_k equals 1, and*
2. *the l-th entry of w_k equals 0 for all $1 \leq l \leq n$ with $l \notin \{k, j_1, j_2, \ldots, j_r\}$.*

Furthermore, the l-th entry of w_k equals 0 for all l with $k < l \leq n$, and the $n - r$ vectors w_k (for $1 \leq k \leq n$ with $k \notin \{j_1, j_2, \ldots, j_r\}$) generate the kernel $\ker A$.

Proof. The proof is completely analogous to Example 6.18 and is left to the reader. □

We can now also check efficiently whether the map associated to a matrix is injective.

Proposition 6.20. *Let $A \in \mathrm{Mat}(m \times n, F)$ be a matrix and A' a row equivalent matrix in row echelon form. Then the associated map $f_A \colon F^n \to F^m$ is injective if and only if A' has n nonzero rows or, equivalently, if and only if each column of A' contains a pivot.*

Proof. By Proposition 6.6, the map f_A is injective if and only if $f_{A'}$ is injective, so it suffices to do the case $A = A'$. By Lemma 4.7, the map f_A is injective if and only if the kernel $\ker f_A = \ker A$ is zero, which, according to Proposition 6.19, happens if and only if each of the n columns of A has a pivot, so if and only if there are exactly n nonzero rows. □

The following proposition explains which columns in a row echelon form of a matrix contain pivots.

Proposition 6.21. *Suppose A and A' are row equivalent $m \times n$ matrices with A' in row echelon form. Then for every $k \in \{1, \ldots, n\}$, the k-th column of A' contains a pivot if and only if the k-th column of A is not a linear combination of the previous columns of A.*

Proof. Let F be a field that A and A' are matrices over. Suppose the column vectors of an $m \times n$ matrix B over F are denoted by $v_1, v_2, \ldots, v_n$. Then the k-th column v_k of B is a linear combination of the previous columns if and only if there are $\lambda_1, \ldots, \lambda_{k-1}$ such that $v_k = \lambda_1 v_1 + \cdots + \lambda_{k-1} v_{k-1}$, that is, such that the element

$$(-\lambda_1, -\lambda_2, \ldots, -\lambda_{k-1}, 1, \underbrace{0, \ldots, 0}_{n-k})$$

is contained in the kernel of B. As A and A' have the same kernel by Proposition 6.3, the k-th column of A is a linear combination of the previous columns of A if and only if the k-th column of A' is a linear combination of the previous columns of A'. Thus, we have reduced to the case $A = A'$.

Let $v_1, v_2, \ldots, v_n$ denote the columns of A. If the k-th column v_k has a pivot, say in the i-th row, then the previous columns $v_1, \ldots, v_{k-1}$ have a 0 on that row, so clearly v_k is not a linear combination of $v_1, \ldots, v_{k-1}$. For the converse, if the k-th column does not contain a pivot, then by Proposition 6.19 there is an element $w_k \in \ker A$ whose k-th entry equals 1 and whose l-th entry equals 0 for $k < l \leq n$. By the above, that implies that v_k is indeed a linear combination of $v_1, v_2, \ldots, v_{k-1}$. □

Exercises

6.3.1. Prove Proposition 6.19.

6.3.2. Give generators for the kernels of each of the matrices in Exercise 5.5.5.

6.3.3. Give generators for the kernel of the matrix A in Example 6.10.

6.3.4. † Determine a row echelon form for the following matrices over $\mathbb{C}$ and give generators for their kernels.

$$\begin{pmatrix} 2+i & 1 & 1+i \\ 2 & 1-3i & 3-5i \end{pmatrix} \qquad \begin{pmatrix} 3 & 0 & 3 \\ 2 & 3 & 0 \\ 3 & 3 & 1 \end{pmatrix}$$

$$\begin{pmatrix} -1 & 0 & 0 & 1 & 2 \\ 2 & 1 & -1 & 0 & 2 \\ 0 & 0 & 0 & -1 & 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 2 & 2 & -2 \\ 2 & 3 & 1 & 0 \\ -2 & 0 & 2 & 1 \end{pmatrix}$$

6.3.5. Let $A \in \mathrm{Mat}(m \times n, F)$ be a matrix and $f_A \colon F^n \to F^m$ the associated linear map.
(1) Show that if f_A is injective, then $m \geq n$.
(2) Show that if A is invertible, then $m = n$ (cf. Corollary 8.9).

6.3.6. Consider the real matrix

$$A = \frac{1}{7} \cdot \begin{pmatrix} 5 & -4 & -2 & 2 \\ -4 & -1 & -4 & 4 \\ -2 & -4 & 5 & 2 \\ 2 & 4 & 2 & 5 \end{pmatrix}.$$

The map $f_A \colon \mathbb{R}^4 \to \mathbb{R}^4$ is the reflection in a hyperplane $H \subset \mathbb{R}^4$. Determine H.

6.4. Reduced row echelon form

While the row echelon form of a matrix is not unique, we will see that the reduced row echelon form below is (see Corollary 6.25).

Definition 6.22. A matrix $A = (a_{ij}) \in \mathrm{Mat}(m \times n, F)$ is in *reduced row echelon form,* if it is in row echelon form and in addition all pivots equal 1 and we have $a_{ij_k} = 0$ for all $1 \leq k \leq r$ and $i \neq k$. This means that the entries above the pivots are zero as well.

$$A = \begin{pmatrix} 0\cdots0 & 1 & *\cdots* & 0 & *\cdots* & 0 & *\cdots* \\ 0\cdots0 & 0 & 0\cdots0 & 1 & *\cdots* & 0 & *\cdots* \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0 & 1 & *\cdots* \\ 0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0 \end{pmatrix}$$

It is clear that every matrix can be transformed into reduced row echelon form by a sequence of elementary row operations — we only have to change Step 5 of the algorithm to

5. For each $i = 1, \ldots, r-1, r+1, \ldots, m$, add $-a'_{ij_r}$ times the r-th row of A' to the i-th row of A'.

Proposition 6.23. *Suppose that $A \in \mathrm{Mat}(m \times n, F)$ is a matrix in reduced row echelon form. Then the nonzero rows of A are uniquely determined by the row space $R(A)$.*

Proof. Let r be the number of nonzero rows of A and let $j_1 < j_2 < \ldots < j_r$ be the numbers of the columns with a pivot. Let $v_1, v_2, \ldots, v_r$ be the nonzero rows of A. Then the j_1-th, j_2-th, $\ldots, j_r$-th entries of the linear combination

$$\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_r v_r$$

are exactly the coefficients $\lambda_1, \lambda_2, \ldots, \lambda_r$. This implies that the nonzero vector in $R(A)$ with the most starting zeros is obtained by taking $\lambda_1 = \ldots = \lambda_{r-1} = 0$, so the vector v_r is the unique nonzero vector in $R(A)$ with the most starting zeros of which the first nonzero entry equals 1. Thus the row space $R(A)$ determines v_r and j_r uniquely. Similarly, v_{r-1} is the unique nonzero vector in $R(A)$ with the most starting zeros of which the j_r-th entry equals 0 and the first nonzero entry equals 1. This also uniquely determines j_{r-1}. By (downward) induction, v_i is the unique nonzero vector in $R(A)$ with the most starting zeros of which the j_{i+1}-th, $\ldots, j_r$-th entries equal 0 and the first nonzero entry, the j_i-th, equals 1. This process yields exactly the r nonzero rows of A and no more, as there are no nonzero vectors in $R(A)$ of which the j_1-th, j_2-th, $\ldots, j_r$-th entries are zero. This means that also r is determined uniquely by $R(A)$. □

Corollary 6.24. *Let $A, A' \in \mathrm{Mat}(m \times n, F)$ be two matrices. Then the following statements are equivalent.*

(1) *The matrices A and A' are row equivalent.*
(2) *The row spaces $R(A)$ and $R(A')$ are equal.*
(3) *For any matrices B and B' in reduced row echelon form that are row equivalent to A and A', respectively, we have $B = B'$.*

Proof. If A and A' are row equivalent, then the row spaces $R(A)$ and $R(A')$ are the same by Proposition 6.3, which proves (1) $\Rightarrow$ (2). For (2) $\Rightarrow$ (3), suppose that the row spaces $R(A)$ and $R(A')$ are equal. Let B and B' be any matrices in reduced row echelon form with B and B' row equivalent to A and A', respectively. By Proposition 6.3 we have $R(B) = R(A)$ and $R(B') = R(A')$, so we conclude $R(B) = R(B')$. Therefore, by Proposition 6.23, the nonzero rows of B and B' coincide, and as the matrices have the same size, they also have the same number of zero rows. This yields $B = B'$. The implication (3) $\Rightarrow$ (1) follows from the fact that if $B = B'$ is row equivalent to both A and A', then A and A' are row equivalent. □

Corollary 6.25. *The reduced row echelon form is unique in the sense that if a matrix A is row equivalent to two matrices B, B' that are both in reduced row echelon form, then $B = B'$.*

Proof. This follows from Corollary 6.24 by taking $A = A'$. □

In other words, the $m \times n$ matrices in reduced row echelon form give a complete system of representatives of the row equivalence classes.

Remark 6.26. It follows from Corollary 6.25 that the number r of nonzero rows in the reduced row echelon form of a matrix A is an invariant of A. It equals the number of nonzero rows in any row echelon form of A. We will see later that this number r equals the so-called *rank* of the matrix A, cf. Section 8.2.

The computation of generators of the kernel of a matrix A is easier when A is in *reduced* row echelon form. The reduced row echelon form for the matrix M of Example 6.18, for instance, is

$$\begin{pmatrix} 1 & 2 & 0 & -1 & 0 & -4 & -5 \\ 0 & 0 & 1 & -1 & 0 & -3 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The circled entries of w_6 of Example 6.18 are exactly the negatives of the elements $-4, -3, 1$ in the nonzero rows and the sixth column. A similar statement holds for the other generators w_2, w_4, and w_7. In terms of Proposition 6.19, with $A = (a_{ij})_{i,j}$ in reduced row echelon form: if $1 \leq k \leq n$ and $k \notin \{j_1, j_2, \ldots, j_r\}$, then the l-th entry of w_k is given by Proposition 6.19 for $l \notin \{j_1, j_2, \ldots, j_r\}$, while the j_i-th entry of w_k is $-a_{ik}$ for $1 \leq i \leq r$; this yields $w_k = e_k - \sum_{i=1}^{r} a_{ik} e_{j_i}$. This is summarized in the next proposition.

As for Proposition 6.19, we will see in Proposition 7.22 that the generators constructed in Proposition 6.27 actually form a so-called basis of the kernel.

Proposition 6.27. *If $A = (a_{ij}) \in \mathrm{Mat}(m \times n, F)$ is a matrix in reduced row echelon form with r nonzero rows and pivots in the columns numbered $j_1 < \ldots < j_r$, then the kernel $\ker(A)$ is generated by the $n - r$ elements*

$$w_k = e_k - \sum_{\substack{1 \leq i \leq r \\ j_i < k}} a_{ik} e_{j_i}\,, \qquad \text{for } k \in \{1, \ldots, n\} \setminus \{j_1, \ldots, j_r\}\,,$$

where $e_1, \ldots, e_n$ are the standard generators of F^n.

Proof. We leave it as an exercise to show that this follows from Proposition 6.19. □

Proposition 6.27 gives a very efficient way of computing the kernel of a matrix. First bring it into reduced row echelon form using elementary row operations, and then write down generators for the kernel according to the given recipe, one generator for each column without pivot.

Exercises

6.4.1. Redo Exercises 6.3.2 and 6.3.4 using the reduced row echelon form.

CHAPTER 7

Linear independence and dimension

7.1. Linear independence

This section, like all others, has a large overlap with Stoll's notes [**S**], in particular with its chapter 6, which in turn follows essentially Chapter 3 in Jänich's book [**J**].

In the context of looking at linear hulls, it is a natural question whether we really need all the given vectors in order to generate their linear hull. Also (maybe in order to reduce waste…), it is interesting to consider *minimal* generating sets. These questions lead to the notions of linear independence and basis.

Definition 7.1. Let V be an F-vector space, and $v_1, v_2, \ldots, v_n \in V$. We say that $v_1, v_2, \ldots, v_n$ are *linearly independent*, if for all $\lambda_1, \lambda_2, \ldots, \lambda_n \in F$, the equality

$$\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n = 0$$

implies $\lambda_1 = \lambda_2 = \cdots = \lambda_n = 0$. ("The zero vector cannot be written as a nontrivial linear combination of $v_1, \ldots, v_n$.")

In a similar way we can define linear independence for arbitrary *collections* of elements of V. If I is any index set (not necessarily finite) and for each $i \in I$ we have an element $v_i \in V$, then we write the collection of all these elements as $(v_i)_{i\in I}$. The element i is called the *index* or *label* of v_i. Elements may occur multiple times, so for $i, j \in I$ with $i \neq j$, we may have $v_i = v_j$. For a more precise definition of (labeled) collections, see Appendix C.

Definition 7.2. A collection $(v_i)_{i\in I}$ of elements in V is *linearly independent* if for every finite subset $S \subset I$, the finite collection $(v_i)_{i\in S}$ is linearly independent, that is, for all (finite) collections $(\lambda_i)_{i\in S}$ of scalars in F, the equality $\sum_{i\in S} \lambda_i v_i = 0$ implies $\lambda_i = 0$ for all $i \in S$.

Note that for finite index sets $I = \{1, 2, \ldots, n\}$, Definitions 7.1 and 7.2 are equivalent, so we have no conflicting definitions. As a special case, the empty sequence or empty collection of vectors is considered to be linearly independent.

If we want to refer to the field of scalars F, we say that the given vectors are *F-linearly independent* or *linearly independent over F*.

If $v_1, v_2, \ldots, v_n$ (resp., $(v_i)_{i\in I}$) are not linearly independent, then we say that they are *linearly dependent*. An equation of the form $\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n = 0$ is called a *linear relation* among the elements $v_1, \ldots, v_n$; if the scalars $\lambda_1, \lambda_2, \ldots, \lambda_n$ are all zero, then we call it the trivial relation, otherwise a nontrivial relation.

Example 7.3. Let V be any vector space. If a collection $(v_i)_{i\in I}$ of elements of V contains the element $0_V \in V$, then the collection is linearly dependent. Furthermore, if there are $i, j \in I$ with $i \neq j$ and $v_i = v_j$, then the collection is linearly dependent as well.

Example 7.4. Let V be a vector space over a field F. Then for any $v \in V$, the one-element sequence v is linearly independent if and only if $v \neq 0$. Any two elements $v_1, v_2 \in V$ are linearly dependent if and only if there are $s, t \in F$, not both 0, such that $sv_1 + tv_2 = 0$. This is the case if and only if v_1 is a multiple of v_2 or v_2 is a multiple of v_1 (or both), because $s \neq 0$ implies $v_1 = -\frac{t}{s}v_2$, while $t \neq 0$ implies $v_2 = -\frac{s}{t}v_1$.

Example 7.5. For an easy example that the field of scalars matters in the context of linear independence, consider $1, i \in \mathbb{C}$, where $\mathbb{C}$ can be considered as a real or as a complex vector space. We then have that 1 and i are $\mathbb{R}$-linearly independent (essentially by definition of $\mathbb{C}$ — $0 = 0 \cdot 1 + 0 \cdot i$, and this representation is unique), whereas they are $\mathbb{C}$-linearly dependent — $i \cdot 1 + (-1) \cdot i = 0$.

Example 7.6. The vectors

$$v_1 = (1, 2, 3, 4), \qquad v_2 = (5, 6, 7, 8), \qquad v_3 = (9, 10, 11, 12)$$

in $\mathbb{R}^4$ are linearly dependent, as we have a linear relation $v_1 - 2v_2 + v_3 = 0$.

Example 7.7. Let F be a field and $V = F[x]$ be the vector space of all polynomials in the variable x over F (see Example 2.13 and Appendix D). For each $n \in \mathbb{Z}_{\geq 0}$ we have the monomial x^n. The collection $(x^n)_{n \in \mathbb{Z}_{\geq 0}}$ is linearly independent, because any finite subcollection is contained in $(1, x, x^2, \ldots, x^d)$ for some $d \in \mathbb{Z}_{\geq 0}$ and any relation

$$a_d x^d + a_{d-1} x^{d-1} + \cdots + a_1 x + a_0 = 0$$

(as polynomials) implies $a_d = a_{d-1} = \ldots = a_1 = a_0 = 0$.

Example 7.8. In $\mathcal{C}(\mathbb{R})$, the functions

$$x \longmapsto 1\,, \quad x \longmapsto \sin x\,, \quad x \longmapsto \cos x\,, \quad x \longmapsto \sin^2 x\,, \quad x \longmapsto \cos^2 x$$

are linearly dependent, since $1 - \sin^2 x - \cos^2 x = 0$ for all $x \in \mathbb{R}$.

On the other hand,

$$x \longmapsto 1\,, \quad x \longmapsto \sin x\,, \quad x \longmapsto \cos x$$

are linearly independent. To see this, assume that $\lambda + \mu \sin x + \nu \cos x = 0$ for all $x \in \mathbb{R}$. Plugging in $x = 0$, we obtain $\lambda + \nu = 0$. For $x = \pi$, we get $\lambda - \nu = 0$, which together imply $\lambda = \nu = 0$. Then taking $x = \pi/2$ shows that $\mu = 0$ as well.

Example 7.9. Consider the vectors

$$w_1 = (1, 1, 1), \quad w_2 = (1, 2, 4), \qquad w_3 = (1, 3, 9)$$

in $\mathbb{R}^3$ and suppose we have $\lambda_1 w_1 + \lambda_2 w_2 + \lambda_3 w_3 = 0$. Then we have

$$\begin{aligned} \lambda_1 + \lambda_2 + \lambda_3 &= 0, \\ \lambda_1 + 2\lambda_2 + 3\lambda_3 &= 0, \\ \lambda_1 + 4\lambda_2 + 9\lambda_3 &= 0. \end{aligned}$$

These equations imply $\lambda_1 = \lambda_2 = \lambda_3 = 0$ (exercise), so w_1, w_2, and w_3 are linearly independent.

Recall from Definition 4.39 that for any sequence $C = (w_1, \ldots, w_n)$ of n elements in a vector space W over a field F, we have a unique linear map $\varphi_C \colon F^n \to W$ that sends the j-th standard vector e_j to w_j; the map φ_C sends $(a_1, \ldots, a_n) \in F^n$ to $a_1 w_1 + \cdots + a_n w_n$.

Proposition 7.10. *Suppose that W is a vector space over the field F and suppose that $C = (w_1, w_2, \dots, w_n)$ a sequence of n vectors in W. Then the elements $w_1, w_2, \dots, w_n$ are linearly independent if and only if $\ker \varphi_C = \{0\}$.*

Proof. The kernel of φ_C consists of all the n-tuples $(\lambda_1, \dots, \lambda_n)$ that satisfy $\lambda_1 w_1 + \cdots + \lambda_n w_n = 0$, so indeed, we have $\ker \varphi_C = \{0\}$ if and only if the elements $w_1, w_2, \dots, w_n$ are linearly independent. □

In fact, the proof shows that the nontrivial linear relations on $w_1, \dots, w_n$ correspond exactly with the nonzero elements of the kernel of φ_C. A statement similar to Proposition 7.10 holds for arbitrary collections (Exercise 7.1.9). For $W = F^m$, we have the following corollary.

Corollary 7.11. *Let F be a field and m a non-negative integer. Then any vectors $w_1, w_2, \dots, w_n \in F^m$ are linearly independent if and only if the $m \times n$ matrix that has $w_1, w_2, \dots, w_n$ as columns has kernel $\{0\}$.*

Proof. The linear map $F^n \to F^m$ that sends e_j to $w_j \in F^m$ corresponds to the described matrix by Lemma 5.9 and Proposition 5.11, so this follows from Proposition 7.10. □

Example 7.12. Let $w_1, w_2, w_3 \in \mathbb{R}^3$ be as in Example 7.9. Then the map $\mathbb{R}^3 \to \mathbb{R}^3$ that sends e_j to w_j corresponds to the matrix

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 4 & 9 \end{pmatrix}$$

that has w_1, w_2, w_3 as columns. It is easily checked that the kernel of this matrix is zero, so it follows again that the vectors w_1, w_2, w_3 are linear independent. If we add the vector $w_4 = (1, 4, 16)$, then the vectors w_1, w_2, w_3, w_4 are linearly independent if and only if the matrix

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 4 & 9 & 16 \end{pmatrix}$$

has kernel zero. Its reduced row echelon form is

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -3 \\ 0 & 0 & 1 & 3 \end{pmatrix}$$

so the kernel is spanned by $(-1, 3, -3, 1)$ and we find the linear relation

$$-w_1 + 3w_2 - 3w_3 + w_4 = 0.$$

We conclude that the vectors w_1, w_2, w_3, w_4 are linearly dependent. Of course, we could have already concluded that from the fact that the matrix with w_1, w_2, w_3, w_4 as columns has more columns than rows, so not every column in the reduced row echelon form could have a pivot, cf. Proposition 6.20.

Lemma 7.13. *Let $f\colon V \to W$ be a linear map of vector spaces. Then any vectors $v_1, v_2, \dots, v_n \in V$ are linearly independent if their images $f(v_1), f(v_2), \dots, f(v_n)$ are. If f is injective, then the converse holds as well.*

Proof. Take any sequence $C = (v_1, v_2, \ldots, v_n)$ of vectors in V. Then, by Proposition 7.10, the map $\varphi_C \colon F^n \to V$ sending e_j to v_j for $1 \leq j \leq n$ is injective if and only if $v_1, v_2, \ldots, v_n$ are linearly independent. Similarly, the composition $f \circ \varphi_C \colon F^n \to W$, which sends e_j to $f(v_j)$, is injective if and only if $f(v_1), f(v_2), \ldots, f(v_n)$ are linearly independent. Therefore, the first statement follows from the fact that if $f \circ \varphi_C$ is injective, then so is φ_C. The second statement follows from the fact that if f is injective, then φ_C is injective if and only if the composition $f \circ \varphi_C$ is. $\square$

Alternative proof. Take any vectors $v_1, v_2, \ldots, v_n \in V$. Any nontrivial relation $\lambda_1 v_1 + \cdots + \lambda_n v_n = 0$ implies a nontrivial relation

$$\lambda_1 f(v_1) + \cdots + \lambda_n f(v_n) = f(\lambda_1 v_1 + \cdots + \lambda_n v_n) = f(0) = 0,$$

so if the elements $v_1, v_2, \ldots, v_n$ are linearly dependent, then so are the elements $f(v_1), f(v_2), \ldots, f(v_n)$. This is equivalent to the first statement.

Suppose that f is injective. Take linearly independent vectors $v_1, \ldots, v_n \in V$. Any linear relation

$$\lambda_1 f(v_1) + \cdots + \lambda_n f(v_n) = 0$$

implies $f(v) = 0$ with $v = \lambda_1 v_1 + \cdots + \lambda_n v_n$, so $v \in \ker f = \{0\}$ and thus $v = 0$. Since $v_1, \ldots, v_n$ are linearly independent, this implies $\lambda_1 = \ldots = \lambda_n = 0$, which implies that the elements $f(v_1), \ldots, f(v_n)$ are linearly independent as well. This proves the second statement. $\square$

From the finite case, it follows immediately that Lemma 7.13 holds for arbitrary collections as well (exercise).

Example 7.14. Let $V = \mathbb{R}[x]$ be the vector space of all real polynomials, containing the elements $f_1 = x^3 - x - 3$, $f_2 = x^2 + 4$, and $f_3 = x^2 + x + 1$. These polynomials all lie in the subspace $\mathbb{R}[x]_3$ of all polynomials of degree at most 3, so to check for linear independence, we may check it within $\mathbb{R}[x]_3$. This is obvious, but it also follows from Lemma 7.13, with f taken to be the inclusion $\mathbb{R}[x]_3 \to \mathbb{R}[x]$ sending any polynomial p to itself.

The linear map $c \colon \mathbb{R}[x]_3 \to \mathbb{R}^4$ that sends any polynomial $a_3x^3 + a_2x^2 + a_1x + a_0$ to the sequence (a_0, a_1, a_2, a_3) of its coefficients is injective (in fact, an isomorphism), so by Lemma 7.13, the polynomials f_1, f_2, and f_3 are linearly independent if and only if $c(f_1), c(f_2)$, and $c(f_3)$ are. The matrix that has these vectors as columns is

$$M = \begin{pmatrix} -3 & 4 & 1 \\ -1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix},$$

which is easily checked to have zero kernel, so $c(f_1), c(f_2)$, and $c(f_3)$ are linearly independent by Corollary 7.11, and therefore, so are f_1, f_2, and f_3.

Note that if we had looked for explicit $\lambda_1, \lambda_2, \lambda_3$ with $\lambda_1 f_1 + \lambda_2 f_2 + \lambda_3 f_3 = 0$, then collecting similar powers of x gives

$$(-3\lambda_1 + 4\lambda_2 + \lambda_3) + (-\lambda_1 + \lambda_3)x + (\lambda_2 + \lambda_3)x^2 + \lambda_1 x^3 = 0.$$

Each of the coefficients has to equal 0, which gives four equations, expressed by the equation

$$M \cdot \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} = 0$$

for the same matrix M. As we have seen before, we have $\ker M = \{0\}$, so the only solution is $\lambda_1 = \lambda_2 = \lambda_3 = 0$, and we conclude again that f_1, f_2, and f_3 are linearly independent.

Proposition 7.15. *Let V be a vector space.*

(1) *For any vectors $v_1, v_2, \ldots, v_n \in V$, the following statements are equivalent.*
   (a) *The vectors $v_1, v_2, \ldots, v_n$ are linearly dependent.*
   (b) *One of the vectors is a linear combination of the previous ones, that is, there is a $j \in \{1, 2, \ldots, n\}$ with $v_j \in L(v_1, \ldots, v_{j-1})$.*
   (c) *One of the vectors is a linear combination of the others, that is, there is a $j \in \{1, 2, \ldots, n\}$ with $v_j \in L(v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_n)$.*

(2) *An infinite sequence $v_1, v_2, v_3, \ldots$ of vectors in V is linearly dependent if and only if one of the vectors is a linear combination of the previous ones.*

(3) *Suppose I is any index set. Then a collection $(v_i)_{i \in I}$ of vectors in V is linearly dependent if and only if one of the vectors is a linear combination of (finitely many of) the others.*

Proof. We start with (1). Let us first assume that $v_1, v_2, \ldots, v_n$ are linearly dependent. Then there are scalars $\lambda_1, \lambda_2, \ldots, \lambda_n$, not all zero, such that

$$\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n = 0 .$$

Let j be the largest index such that $\lambda_j \neq 0$. Then

$$v_j = -\lambda_j^{-1}(\lambda_1 v_1 + \cdots + \lambda_{j-1} v_{j-1}) \in L(v_1, \ldots, v_{j-1}) .$$

This proves the implication (a) $\Rightarrow$ (b). The implication (b) $\Rightarrow$ (c) is trivial. For the implication (c) $\Rightarrow$ (a), assume that v_j is a linear combination of the others:

$$v_j = \lambda_1 v_1 + \cdots + \lambda_{j-1} v_{j-1} + \lambda_{j+1} v_{j+1} + \ldots + \lambda_n v_n$$

for some $\lambda_1, \lambda_2, \ldots, \lambda_{j-1}, \lambda_{j+1}, \ldots, \lambda_n$. Then

$$\lambda_1 v_1 + \cdots + \lambda_{j-1} v_{j-1} - v_j + \lambda_{j+1} v_{j+1} + \ldots + \lambda_n v_n = 0 ,$$

so the given vectors are linearly dependent. This proves part (1).

For (2) and (3), we recall that a collection $(v_i)_{i \in I}$ is linearly dependent if and only if for some finite subset $S \subset I$, the finite subcollection $(v_i)_{i \in S}$ is linearly dependent. For part (2) we finish the proof by noting that for every finite set S, there is an integer n such that we have $S \subset \{1, 2 \ldots, n\}$, so we can apply the equivalence (a) $\Leftrightarrow$ (b) of part (1). For part (3) we can just number the elements of S by $1, 2, \ldots, n = |S|$, and then apply the equivalence (a) $\Leftrightarrow$ (c) of part (1). ☐

Example 7.16. Consider the real polynomials

$$f_1 = 1, \quad f_2 = x + 2, \quad f_3 = x^2 - 2x + 3, \quad f_4 = 2x^4 - 2x^2 + 5$$

inside the real vector space $\mathbb{R}[x]$ (cf. Example 2.13 and Appendix D). The degree of each polynomial is higher than the degree of all the previous ones, so none of the polynomials is a linear combination of the previous ones and we conclude by Proposition 7.15 that the polynomials are linearly independent.

Example 7.17. Take the vectors

$$\begin{aligned} v_1 &= (1, 2, 1, -1, 2, 1, 0), \\ v_2 &= (0, 1, 1, 0, -1, -2, 3), \\ v_3 &= (0, 0, 0, 3, 3, -1, 2), \\ v_4 &= (0, 0, 0, 0, 0, 6, 4) \end{aligned}$$

in $\mathbb{Q}^7$. We consider them in opposite order, so v_4, v_3, v_2, v_1. Then for each vector, the first coordinate that is nonzero (namely the sixth, fourth, second, and first coordinate respectively), is zero for all previous vectors. This implies that no vector is a linear combination of the previous ones, so the vectors are linearly independent by Proposition 7.15.

Exercises

7.1.1. Which of the following sequences of vectors in $\mathbb{R}^3$ are linearly independent?
(1) $\big((1, 2, 3), (2, 1, -1), (-1, 1, 1)\big)$,
(2) $\big((1, 3, 2), (1, 1, 1), (-1, 3, 1)\big)$.

7.1.2. Are the real polynomials $3, x-1, x^2-3x+2, x^4-3x+13, x^7-x+14$ linearly independent?

7.1.3. Are the complex polynomials $x^7 - 2x + 1, 5x^2, 2x^4 - 5x^3, x, x^6 - 3x$ linearly independent?

7.1.4. Are the vectors

$$\begin{aligned} v_1 &= (1, 4, 2, 3, 5), \\ v_2 &= (-1, 7, 2, 3, 6), \\ v_3 &= (4, 2, 3, -3, 4), \\ v_4 &= (2, -3, 1, 4, 2), \\ v_5 &= (6, 5, 3, -2, -4), \\ v_6 &= (1, -7, 3, 2, 5) \end{aligned}$$

in $\mathbb{R}^5$ linearly independent? (Hint: do not start a huge computation)

7.1.5. Phrase and prove a version of part (2) of Proposition 7.15 for any collection of vectors indexed by a totally ordered set I.

7.1.6. Let V be a vector space, I an index set, and $(v_i)_{i\in I}$ a collection of elements of V. Let $j \in I$ be an index and suppose that the subcollection $(v_i)_{i\in I\setminus\{j\}}$ is linearly independent. Prove that the whole collection $(v_i)_{i\in I}$ is linearly independent if and only if we have

$$v_j \notin L\big((v_i)_{i\in I\setminus\{j\}}\big).$$

7.1.7. Let $V = \mathrm{Map}(\mathbb{R}, \mathbb{R})$ be the vector space of all functions from $\mathbb{R}$ to $\mathbb{R}$. Let $\mathcal{I}$ denote the set of all closed intervals $[a, b]$ in $\mathbb{R}$. For each interval $I \in \mathcal{I}$, we let h_I denote the function given by

$$h_I(x) = \begin{cases} 1 & \text{if } x \in I, \\ 0 & \text{if } x \notin I. \end{cases}$$

Is the collection $\big(h_I\big)_{I\in\mathcal{I}}$ linearly independent?

7.1.8. Let $n > 0$ be an integer and $a_0, \ldots, a_n \in \mathbb{R}$ real numbers. Let $f_0, \ldots, f_n \in \mathcal{C}(\mathbb{R})$ be continuous functions that satisfy

$$f_i(a_j) = \begin{cases} 1 & \text{if } j \leq i, \\ 0 & \text{otherwise.} \end{cases}$$

Show that the functions $f_0, f_1, \ldots, f_n$ are linearly independent.

7.1.9. Suppose W is a vector space over a field F, containing a (possibly infinite) collection $(w_i)_{i \in I}$ of elements. Let $\varphi \colon F^{(I)} \to W$ be the unique linear map sending the standard vector e_i to w_i for all $i \in I$ (see Exercise 4.4.7).

(1) Show that the collection $(w_i)_{i \in I}$ is linearly independent if and only if φ is injective. This is a generalisation of Proposition 7.10.

(2) Show that the collection $(w_i)_{i \in I}$ generates W if and only if φ is surjective. This is a reformulation of Exercise 4.4.7.

7.1.10. State and prove a generalisation of Lemma 7.13 for arbitrary collections of vectors.

7.2. Bases

Definition 7.18. Let V be a vector space. A *basis* is a collection $(v_i)_{i \in I}$ of vectors in V that is linearly independent and generates V, that is, $V = L\big((v_i)_{i \in I}\big)$.

In particular, a finite sequence $(v_1, v_2, \ldots, v_n)$ of elements of V is a basis for V if and only if $v_1, v_2, \ldots, v_n$ are linearly independent, and they generate V. We also say that the elements $v_1, v_2, \ldots, v_n$ *form* a basis for V.

Note that the elements of a basis $(v_1, v_2, \ldots, v_n)$ have a specific order. Also in the general case of arbitrary labeled collections, a basis $(v_i)_{i \in I}$ has a similar structure: for each index $i \in I$, we know which element is the i-th element.

Remark 7.19. Technically, we have not defined the notation $L\big((v_i)_{i \in I}\big)$ used in Definition 7.18, as we only defined the span of sets and finite sequences in Definition 3.24, not of (labeled) collections. Of course, though, the notation $L\big((v_i)_{i \in I}\big)$ stands for the set of all linear combinations of finite subcollections $(v_i)_{i \in S}$ with $S \subset I$ finite. This equals the span of the set $\{v_i \,:\, i \in I\}$ of elements in the collection (cf. Remark 3.25).

Example 7.20. The most basic example of a basis is the *canonical basis* or *standard basis* of F^n. This is $E = (e_1, e_2, \ldots, e_n)$, where

$$\begin{aligned} e_1 &= (1, 0, 0, \ldots, 0, 0) \\ e_2 &= (0, 1, 0, \ldots, 0, 0) \\ &\ \ \vdots \qquad\qquad \vdots \\ e_n &= (0, 0, 0, \ldots, 0, 1)\,. \end{aligned}$$

The standard generators $e_1, \ldots, e_n$ are therefore also called *standard basis vectors.*

Example 7.21. Let X be a *finite* set and F a field. For each $x \in X$, we define the function $f_x \colon X \to F$ that sends x to 1 and every other element of X to 0. Then the collection $(f_x)_{x \in X}$ is a basis for the vector space F^X. Compare this to the previous example. See Exercise 7.2.5 for infinite sets.

Proposition 7.22 (Basis of row space and kernel). *Let $A \in \mathrm{Mat}(m \times n, F)$ be a matrix in row echelon form with r nonzero rows. Then these r rows form a basis for the row space $R(A)$. The $n-r$ elements w_k (for all $1 \leq k \leq n$ for which the k-th column contains no pivot) of Proposition 6.19 (or Proposition 6.27 if A is in reduced row echelon form) form a basis of the kernel of A.*

Proof. Consider the r nonzero rows from bottom to top. Then, just as in Example 7.17, for each row, the first coordinate that is nonzero, is zero for all previous rows. This implies that no row is a linear combination of the previous ones, so the vectors are linearly independent by Proposition 7.15. These r rows generate the row space by definition, so they form a basis for $R(A)$.

For each k with $1 \leq k \leq n$, for which the k-th column of A contains no pivot, the element w_k has a 1 on the k-th coordinate, where all the other $n-r-1$ elements have a 0. This implies that none of the w_k is a linear combination of the others, so by Proposition 7.15, these $n-r$ elements are linearly independent. They generate the kernel by Proposition 6.19 (or 6.27), so they form a basis for $\ker A$. $\square$

Remark 7.23 (Basis of U and $U^\perp$ using rows)**.** We can use Proposition 7.22 to find a basis of a subspace U of F^n generated by elements $v_1, v_2, \ldots, v_m$. First we let A denote the $m \times n$ matrix of which the rows are $v_1, v_2, \ldots, v_m$. Then we apply a sequence of elementary row operations to A to obtain a matrix A' that is in row echelon form. Since the row spaces $R(A)$ and $R(A')$ are equal by Proposition 6.3, the nonzero rows of A' form a basis for $R(A') = R(A) = U$ by Proposition 7.22. Moreover, the subspace $U^\perp$ equals $\ker A = \ker A'$ by Propositions 5.32 and 6.3, so Proposition 7.22 also gives a basis for $U^\perp$.

Remark 7.23 puts generators of a subspace $U \subset F^n$ as rows in a matrix in order to find a basis for U and $U^\perp$. In Proposition 7.26 we will describe a method to find a basis for U that puts generators of U as columns in a matrix. We first phrase a useful lemma.

Lemma 7.24. *Suppose V is a vector space with elements $v_1, v_2, \ldots, v_n \in V$. Let $I \subset \{1, 2, \ldots, n\}$ be the set of all i for which v_i is* not *a linear combination of $v_1, \ldots, v_{i-1}$. Then the collection $(v_i)_{i \in I}$ is a basis for $L(v_1, v_2, \ldots, v_n)$.*

Proof. Set $U = L\big((v_i)_{i \in I}\big) \subset L(v_1, v_2, \ldots, v_n)$. By induction we show that $L(v_1, v_2, \ldots, v_j) \subset U$ for all integers $0 \leq j \leq n$. For $j = 0$ this is trivial, as we have $L(v_1, v_2, \ldots, v_j) = L(\emptyset) = \{0\}$. For $0 < j \leq n$ we have two cases. In the case $j \in I$ we clearly have $v_j \in U$. For $j \notin I$, the vector v_j is by definition a linear combination of $v_1, \ldots, v_{j-1}$, so we have $v_j \in L(v_1, \ldots, v_{j-1}) \subset U$ by the induction hypothesis. For $j = n$ we obtain $U = L(v_1, v_2, \ldots, v_n)$. It remains to show that the collection $(v_i)_{i \in I}$ is linearly independent, which follows from part (1) or from Proposition 7.15. $\square$

Example 7.25. Consider the matrix

$$A = \begin{pmatrix} 1 & 1 & 2 & 1 & 3 & 4 & 0 \\ 0 & 1 & -1 & 2 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 & 0 & 2 & -3 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

which is in row echelon form. By Proposition 6.21, the columns with a pivot, that is, the first, second, fourth, and sixth, are exactly the columns that are not a linear combination of the previous columns of A. From Lemma 7.24 we conclude that these four columns form a basis for the column space $C(A)$ of A.

We can combine Proposition 6.21 and Lemma 7.24 to make a method to determine a basis for the column space of a matrix.

Proposition 7.26 (Basis of column space). *Let A be an $m \times n$ matrix over a field F with columns $w_1, \ldots, w_n$. Let A' be a matrix in row echelon form that is row equivalent to A. Let $I \subset \{1, \ldots, n\}$ be the set of all indices of columns of A' with a pivot. Then the collection $(w_i)_{i \in I}$ is a basis for the column space $C(A) = L(w_1, \ldots, w_n)$ of A.*

Proof. By Proposition 6.21, the collection $(w_i)_{i \in I}$ consists of those columns w_i of A that are not a linear combination of the previous columns of A. By Lemma 7.24, this implies that this collection $(w_i)_{i \in I}$ is a basis for the space $L(w_1, \ldots, w_n) = C(A)$. $\square$

Remark 7.27 (Basis of U using columns)**.** We can use Proposition 7.26 to determine a basis of a subspace U of F^m generated by elements $w_1, w_2, \ldots, w_n$. First we let A denote the $m \times n$ matrix of which the columns are $w_1, w_2, \ldots, w_n$. Then we apply a sequence of elementary row operations to A to obtain a matrix A' that is in row echelon form, and we let I denote the set of all indices i with $1 \leq i \leq n$ for which the i-th column of A' contains a pivot. Then the collection $(w_i)_{i \in I}$ is a basis for $U = C(A)$.

An advantage of this method is that the basis we find consists entirely of vectors that we started with.

A summary of the idea behind this is the following. Note that row operations may change the column space, but the kernel is preserved, which means that linear relations among the columns of a matrix B are preserved among the columns of a row equivalent matrix B' (and vice versa). If B' is a matrix in row echelon form, the existence of linear relations can be read off easily from the pivots.

Example 7.28. Let us determine a basis for the subspace $U \subset \mathbb{R}^4$ generated by

$$\begin{aligned} v_1 &= (1, 0, 2, -1), \\ v_2 &= (0, 1, 0, 2), \\ v_3 &= (1, 2, 2, 3), \\ v_4 &= (1, -1, 0, 1), \\ v_5 &= (0, 3, 2, 2). \end{aligned}$$

The 4×5 matrix B with these vectors as columns has reduced row echelon form

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

The pivots are contained in columns 1, 2, and 4, so the first, second, and fourth column of B form a basis (v_1, v_2, v_4) for U. From the reduced row echelon form we can also read off the linear relations $v_3 = v_1 + 2v_2$ and $v_5 = v_1 + 2v_2 - v_4$, which correspond to the generators $(1, 2, -1, 0, 0)$ and $(1, 2, 0, -1, -1)$ of the kernel (cf. Proposition 6.19 or 6.27).

Recall from Definition 4.39, as in the previous section, that for any sequence $C = (w_1, \ldots, w_n)$ of n elements in a vector space W over a field F, we have a unique linear map $\varphi_C \colon F^n \to W$ that sends the j-th standard vector e_j to w_j; the map φ_C sends $(a_1, \ldots, a_n) \in F^n$ to $a_1 w_1 + \cdots + a_n w_n$.

Proposition 7.29. *Let W be a vector space over the field F and $C = (w_1, \ldots, w_n)$ a sequence of n vectors in W. Then C is a basis for W if and only if the map $\varphi_C \colon F^n \to W$ is an isomorphism.*

Proof. The map φ_C is injective if and only if $w_1, \ldots, w_n$ are linearly independent by Proposition 7.10. The map φ_C is surjective if and only if $w_1, \ldots, w_n$ generate W (see the remark below Proposition 4.38). The statement follows. $\square$

A statement similar to Proposition 7.29 holds for arbitrary collections (Exercise 7.2.6).

From Proposition 7.15 above, we see that the elements of a basis for V form a *minimal* generating set of V in the sense that we cannot leave out some element and still have a generating set. Lemma 7.30 states a consequence that makes bases special among all generating sets.

Lemma 7.30. *Suppose V is an F-vector space. Then a sequence $(v_1, v_2, \ldots, v_n)$ of elements in V is a basis for V if and only if for every $v \in V$, there are* unique *scalars $\lambda_1, \lambda_2, \ldots, \lambda_n \in F$ such that*

$$v = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n\,.$$

Proof. Set $C = (v_1, v_2, \ldots, v_n)$. Then by Proposition 7.29, the sequence C is basis for V if and only if φ_C is an isomorphism. On the other hand, φ_C is surjective if and only if for every $v \in V$, there are scalars $\lambda_1, \lambda_2, \ldots, \lambda_n \in F$

such that

$$v = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n,$$

and φ_C is injective if and only if such scalars are unique, if they exist. It follows that φ_C is bijective if and only if there are unique scalars satisfying the given equation. This proves the lemma. □

Alternative proof. Suppose that the sequence $(v_1, v_2, \ldots, v_n)$ is a basis for V. The *existence* of $(\lambda_1, \lambda_2, \ldots, \lambda_n) \in F^n$ such that

$$v = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n$$

follows from the fact that $v_1, v_2, \ldots, v_n$ generate V.

To show *uniqueness,* assume that $(\mu_1, \mu_2, \ldots, \mu_n) \in F^n$ also satisfy

$$v = \mu_1 v_1 + \mu_2 v_2 + \cdots + \mu_n v_n \,.$$

Taking the difference, we obtain

$$0 = (\lambda_1 - \mu_1)v_1 + (\lambda_2 - \mu_2)v_2 + \cdots + (\lambda_n - \mu_n)v_n \,.$$

Since $v_1, v_2, \ldots, v_n$ are linearly independent, it follows that

$$\lambda_1 - \mu_1 = \lambda_2 - \mu_2 = \cdots = \lambda_n - \mu_n = 0\,,$$

that is, $(\lambda_1, \ldots, \lambda_n) = (\mu_1, \ldots, \mu_n)$. This shows that the sequence $(\lambda_1, \ldots, \lambda_n)$ was indeed unique. The converse is left as an exercise. □

A statement similar to Lemma 7.30 holds for arbitrary collections (Exercise 7.2.7).

Proposition 7.31. *Let V and W be vector spaces, $f\colon V \to W$ a linear map, and let $v_1, \ldots, v_n \in V$ be vectors that generate V. Then*

(1) *f is surjective if and only if $L(f(v_1), \ldots, f(v_n)) = W$.*

Assume that $v_1, \ldots, v_n$ form a basis for V. Then

(2) *f is injective if and only if $f(v_1), \ldots, f(v_n)$ are linearly independent,*
(3) *f is an isomorphism if and only if $f(v_1), \ldots, f(v_n)$ is a basis of W.*

Proof. Set $C = (v_1, v_2, \ldots, v_n)$ and $D = (f(v_1), f(v_2), \ldots, f(v_n))$. Then the linear maps $\varphi_C\colon F^n \to V$ and $\varphi_D\colon F^n \to W$ are related by $\varphi_D = f \circ \varphi_C$. Since the elements $v_1, \ldots, v_n \in V$ generate V, the map φ_C is surjective. We conclude that f is surjective if and only if φ_D is surjective, which is the case if and only if $L(f(v_1), \ldots, f(v_n)) = W$. This proves (1). For (2), we note that φ_C is an isomorphism, because C is a basis for V. We conclude that f is injective if and only if φ_D is injective, which is the case if and only if $f(v_1), \ldots, f(v_n)$ are linearly independent. Statement (3) follows from (1) and (2). □

Just as for Lemma 7.30, we can also give an alternative proof straight from the definitions of 'generating' and 'linearly independent', without making use of the maps φ_C and φ_D. We leave this to the reader.

The following corollary follows directly from part (3) of Proposition 7.31 and implies that if $f\colon V \to W$ is an isomorphism, then $v_1, v_2, \ldots, v_n \in V$ form a basis for V if and only if their images $f(v_1), f(v_2), \ldots, f(v_n)$ form a basis for W.

Corollary 7.32. *Let $f\colon V \to W$ be a linear map of vector spaces and $v_1, v_2, \dots, v_n$ elements of V. Then any two of the following three statements together imply the third.*

(1) *The map f is an isomorphism.*
(2) *The elements $v_1, v_2, \dots, v_n$ form a basis for V.*
(3) *The elements $f(v_1), f(v_2), \dots, f(v_n)$ form a basis for W.*

Proof. If (2) holds, then (1) and (3) are equivalent by part (3) of Proposition 7.31. This proves the implications $(1) + (2) \Rightarrow (3)$ and $(2) + (3) \Rightarrow (1)$. Applying the first of these implications to f^{-1}, we deduce the remaining implication $(1) + (3) \Rightarrow (2)$. $\square$

Lemma 7.30, Proposition 7.31, and Corollary 7.32 also hold for arbitrary collections (see Exercises 7.2.2, 7.2.10, and 7.2.11).

Exercises

7.2.1. Determine a basis for the subspaces of $\mathbb{R}^n$ generated by
(1) $v_1 = (1,3), v_2 = (2,1), v_3 = (1,1)$,
(2) $v_1 = (1,3,1), v_2 = (2,1,2), v_3 = (1,1,1)$,
(3) $v_1 = (1,3,1), v_2 = (3,1,3), v_3 = (1,1,1)$,
(4) $v_1 = (1,2,3), v_2 = (4,5,6), v_3 = (7,8,9)$,
(5) $v_1 = (1,2,3,4), v_2 = (4,3,2,1), v_3 = (1,-1,1,-1)$,

7.2.2. Finish the alternative proof of Lemma 7.30.

7.2.3. For each of the matrices of Exercise 6.3.4, select some columns that form a basis for the column space of that matrix.

7.2.4. Consider the real polynomials

$$\begin{aligned}
f_1 &= x^2 + 1,\\
f_2 &= x^3 - x^2 + x,\\
f_3 &= x^4 + x - 7,\\
f_4 &= x^4 - 6,\\
f_5 &= x^3 + 2,\\
f_6 &= x^2 + x.
\end{aligned}$$

and the vectorspace $U \subset \mathbb{R}[x]$ they generate. Select some polynomials that form a basis for U.

7.2.5. This exercise generalises Example 7.21. Let X be any set and F a field. For each $x \in X$, we define the function $f_x\colon X \to F$ that sends x to 1 and every other element of X to 0.
(1) Give an example where the collection $(f_x)_{x\in X}$ is not a basis for F^X.
(2) Show that the collection $(f_x)_{x\in X}$ is a basis of the vector space $F^{(X)}$.

7.2.6. State and prove a generalisation of Proposition 7.29 for arbitrary collections of vectors, cf. Exercises 4.4.7 and 7.1.9.

7.2.7. State and prove an analog of Lemma 7.30 for arbitrary collections $(v_i)_{i\in I}$ of vectors in V.

7.2.8. (1) Use Proposition 4.38 to prove the following generalisation of Proposition 4.38 itself: "Let V and W be vector spaces over a field F, and let $B = (v_1, v_2, \dots, v_n)$ be a basis for V. Then for every sequence $w_1, \dots, w_n$ of vectors in W there is a unique linear map $f\colon V \to W$ such that $f(v_j) = w_j$ for all $j \in \{1, \dots, n\}$."

(2) Also state and prove an analog for arbitrary collections $(v_i)_{i\in I}$ (basis for V) and $(w_i)_{i\in I}$ (general elements in W).

7.2.9. (1) Prove a version of Lemma 7.24 for infinite sequences $v_1, v_2, v_3, \dots$.

(2) What about sequences $(v_i)_{i\in\mathbb{Z}} = \dots, v_{-1}, v_0, v_1, \dots$ that are infinite in both directions, with the hypothesis that I consist of all $i \in \mathbb{Z}$ for which v_i is not a linear combination of the previous elements?

The last exercises relate linear independence and generating on one hand to injectivity and surjectivity on the other. They are related to Lemmas 7.13, Proposition 7.31, and Corollary 7.32. Some parts require the existence of a basis. Appendix E shows that using Zorn's Lemma one can indeed prove that all vector spaces have a basis (cf. Warning 7.52). In these exercises, however, we will include it as an explicit hypothesis whenever it is needed.

7.2.10. State and prove an analog of Proposition 7.31 for an arbitrary collection $(v_i)_{i\in I}$ of vectors in V (also follows from Exercises 7.2.12, 7.2.13, and 7.2.14).

7.2.11. State and prove an analog of Corollary 7.32 for arbitrary collections $(v_i)_{i\in I}$ of vectors in V.

7.2.12. Let $f\colon V \to W$ be a linear map. Show that the following are equivalent.

(1) The map f is injective.
(2) For every non-negative integer n and every sequence $v_1, \dots, v_n \in V$ of linearly independent vectors, the images $f(v_1), \dots, f(v_n)$ are linearly independent in W.
(3) For every collection $(v_i)_{i\in I}$ of linearly independent vectors in V, the collection $(f(v_i))_{i\in I}$ of images is linearly independent in W.

Show that if V has a (not necessarily finite) basis, then these statements are also equivalent to the following.

(4) For all bases $(v_i)_{i\in I}$ for V, the collection $(f(v_i))_{i\in I}$ of images is linearly independent in W.
(5) There exists a basis $(v_i)_{i\in I}$ for V for which the collection $(f(v_i))_{i\in I}$ of images is linearly independent in W.

7.2.13. Let $f\colon V \to W$ be a linear map. Show that the following are equivalent.

(1) The map f is surjective.
(2) For every collection $(v_i)_{i\in I}$ of vectors that generates the space V, the collection $(f(v_i))_{i\in I}$ of their images generates W.
(3) There is a collection $(v_i)_{i\in I}$ of vectors in V for which the collection $(f(v_i))_{i\in I}$ of their images generates W.

Explain why the analog for finite sequences is missing among these statements by giving an example of a linear map $f\colon V \to W$ that is not surjective, but such that for all sequences $v_1, v_2, \dots, v_n$ of elements in V that generate V, the images $f(v_1), f(v_2), \dots, f(v_n)$ generate W.

7.2.14. Let $f\colon V \to W$ be a linear map and assume V has a (not necessarily finite) basis. Then the following are equivalent.

(1) The map f is an isomorphism.
(2) For every basis $(v_i)_{i\in I}$ for V, the collection $(f(v_i))_{i\in I}$ is a basis for W.
(3) There exists a basis $(v_i)_{i\in I}$ for V for which the collection $(f(v_i))_{i\in I}$ is a basis for W.

7.3. The basis extension theorem and dimension

Proposition 7.29 says that if $v_1, v_2, \dots, v_n$ form a basis for a vector space V, then V is isomorphic to the standard vector space F^n, so we can express everything in V in terms of F^n. Since we seem to know "everything" about a vector space as soon as we know a basis, it makes sense to use bases to measure the "size" of vector spaces. In order for this to make sense, we need to know that any two bases of a given vector space have the same size. The key to this (and many other important results) is the following.

Theorem 7.33 (Basis Extension Theorem). *Let V be a vector space, and let $v_1, \ldots, v_r, w_1, \ldots, w_s \in V$ be vectors such that $v_1, \ldots, v_r$ are linearly independent and $V = L(v_1, \ldots, v_r, w_1, \ldots, w_s)$. Let $I \subset \{1, \ldots, s\}$ be the set of indices i for which w_i is not a linear combination of $v_1, \ldots, v_r, w_1, \ldots, w_{i-1}$. Then $v_1, v_2, \ldots, v_r$ and $(w_i)_{i \in I}$ together form a basis for V.*

Proof. Because $v_1, v_2, \ldots, v_r$ are linearly independent, none of them are linear combinations of the other $r-1$ of them. Hence, this follows immediately from applying Lemma 7.24 to the elements $v_1, v_2, \ldots, v_r, w_1, w_2, \ldots, w_s$. □

The Basis Extension Theorem says that if we have a bunch of vectors that is 'too small' ($v_1, \ldots, v_r$ linearly independent, but not necessarily generating) and a larger bunch of vectors that is 'too large' ($v_1, \ldots, v_r, w_1, \ldots, w_s$ generating but not necessarily linearly independent), then there is a basis in between: by adding suitably chosen vectors from $w_1, \ldots, w_s$, we can extend $v_1, \ldots, v_r$ to a basis of V.

As we saw in its proof, Theorem 7.33 is nothing but a special case of Lemma 7.24, namely the case in which we already know that the first r vectors $v_1, \ldots, v_r$ are linearly independent. In a different direction, Proposition 7.26 is a specialisation of Lemma 7.24 as well, namely where we take $V = F^m$ (and Proposition 6.21 was used to define the set I more explicitly in terms of the columns that contain pivots). The following common specialisation is an explicit version of the Basis Extension Theorem for F^m.

Corollary 7.34. *Let $v_1, \ldots, v_r, w_1, \ldots, w_s \in F^m$ be elements such that $v_1, \ldots, v_r$ are linearly independent and set $V = L(v_1, \ldots, v_r, w_1, \ldots, w_s)$. Let A be the matrix with columns $v_1, \ldots, v_r, w_1, \ldots, w_s$, let A' be the reduced row echelon form of A, and let I be the set of all indices $1 \leq i \leq s$ for which the $(r+i)$-th column of A' has a pivot. Then $v_1, v_2, \ldots, v_r$ and $(w_i)_{i \in I}$ together form a basis for V.*

Proof. By Proposition 6.21, the collection $(w_i)_{i \in I}$ consists exactly of those columns w_i of A that are not a linear combination of the previous columns of A. By Theorem 7.33, this implies the desired conclusion. □

Example 7.35. Consider the vectors

$$v_1 = (1, 1, 2) \qquad \text{and} \qquad v_2 = (-1, 2, 4)$$

in $\mathbb{R}^3$. We will extend (v_1, v_2) to a basis for $\mathbb{R}^3$. Clearly, the vectors v_1, v_2, e_1, e_2, e_3 together generate $\mathbb{R}^3$, because the standard generators e_1, e_2, e_3 already generate $\mathbb{R}^3$ by themselves. We apply Corollary 7.34 and find that the matrix with v_1, v_2, e_1, e_2, e_3 as columns has reduced row echelon form

$$\begin{pmatrix} 1 & 0 & \frac{2}{3} & 0 & \frac{1}{6} \\ 0 & 1 & -\frac{1}{3} & 0 & \frac{1}{6} \\ 0 & 0 & 0 & 1 & -\frac{1}{2} \end{pmatrix}.$$

The pivots are in columns 1,2, and 4. Hence, the corresponding vectors v_1, v_2, e_2 form a basis for $\mathbb{R}^3$.

Example 7.36. Consider the real polynomials $f_1 = x^2 - 1$, $f_2 = x^3 - x$, and $f_3 = x^3 - 2x^2 - x + 1$ in the vector space $\mathbb{R}[x]_3$ of polynomials of degree at most 3. It is easy to check that these polynomials are linearly independent. On the other hand, the monomials $1, x, x^2, x^3$ generate $\mathbb{R}[x]_3$, so certainly

$$f_1, f_2, f_3, 1, x, x^2, x^3$$

generate $\mathbb{R}[x]_3$. By the Basis Extension Theorem we can extend f_1, f_2, f_3 to a basis by adding suitably chosen monomials. The monomials $1 = f_2 - 2f_1 - f_3$ and $x^2 = f_2 - f_1 - f_3$ are already contained in $L(f_1, f_2, f_3)$, so adding either of those to f_1, f_2, f_3 would cause nontrivial linear relations. The element x, however, is not contained in $L(f_1, f_2, f_3)$, because f_1, f_2, f_3, x are linearly independent (check this). We have

$$1 = f_2 - 2f_1 - f_3, \qquad x^2 = f_2 - f_1 - f_3, \qquad \text{and} \qquad x^3 = f_2 + x,$$

so the generators $1, x, x^2, x^3$ of $\mathbb{R}[x]_3$ are contained in $L(f_1, f_2, f_3, x)$, and therefore $L(f_1, f_2, f_3, x) = \mathbb{R}[x]_3$, so f_1, f_2, f_3, x generate $\mathbb{R}[x]_3$ and form a basis for $\mathbb{R}[x]_3$. We could have also added x^3 to f_1, f_2, f_3 to obtain a basis.

Example 7.37. Let us revisit the previous example. The linear map

$$\varphi\colon \mathbb{R}^4 \to \mathbb{R}[x]_3, \qquad (a_0, a_1, a_2, a_3) \mapsto a_3x^3 + a_2x^2 + a_1x + a_0$$

is an isomorphism, so φ and φ^{-1} send linearly independent vectors to linearly independent vectors (Lemma 7.13) and bases to bases (Corollary 7.32). Setting $v_i = \varphi^{-1}(f_i)$ for $i = 1, 2, 3$ and $w_j = \varphi^{-1}(x^j)$ for $j = 0, 1, 2, 3$, we get $w_j = e_j$ and

$$v_1 = \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \qquad v_2 = \begin{pmatrix} 0 \\ -1 \\ 0 \\ 1 \end{pmatrix}, \qquad \text{and} \qquad v_3 = \begin{pmatrix} 1 \\ -1 \\ -2 \\ 1 \end{pmatrix}.$$

We wish to extend v_1, v_2, v_3 to a basis of $\mathbb{R}^4$ by adding suitably chosen elements from $\{e_1, e_2, e_3, e_4\}$. In order to do so, we use Proposition 7.26 and Remark 7.27 and put the seven vectors as columns in a matrix

$$A = \begin{pmatrix} -1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & -1 & -1 & 0 & 1 & 0 & 0 \\ 1 & 0 & -2 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{pmatrix},$$

of which the reduced row echelon form equals

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

The pivots in the latter matrix are contained in columns 1, 2, 3, and 5, so by Proposition 7.26 and Remark 7.27, the corresponding columns v_1, v_2, v_3, e_2 of A form a basis for $C(A) = \mathbb{R}^4$. After applying φ, we find that (f_1, f_2, f_3, x) is a basis for $\mathbb{R}[x]_3$, which is exactly the basis we had found before.

Note that it was not a coincidence that the first three columns of the matrix in row echelon form contained a pivot, because we already knew that the elements v_1, v_2, v_3 are linearly independent, so none of these is a linear combination of the previous, cf. Proposition 6.21.

The Basis Extension Theorem implies another important statement, namely the Exchange Lemma. It says that if we have two finite bases of a vector space, then we can trade any vector of our choice in the first basis for a vector in the second basis in such a way as to still have a basis.

Lemma 7.38 (Exchange Lemma). *If $v_1, \dots, v_n$ and $w_1, \dots, w_m$ are two bases of a vector space V, then for each $i \in \{1, 2, \dots, n\}$ there is some $j \in \{1, 2, \dots, m\}$ such that $v_1, \dots, v_{i-1}, w_j, v_{i+1}, \dots, v_n$ is again a basis of V.*

Proof. Fix $i \in \{1, \dots, n\}$ and set $U = L(v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_n)$. As $v_1, \dots, v_n$ are linearly independent, we have $v_i \notin U$ by the equivalence (a) $\Leftrightarrow$ (c) of Proposition 7.15, so $U \subsetneq V$. This implies that there is some $j \in \{1, \dots, m\}$ such that $w_j \notin U$ (if we had $w_j \in U$ for all j, then we would have $V \subset U$). Choose such a j. Then by the equivalence (a) $\Leftrightarrow$ (b) of Proposition 7.15, the vectors $v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_n, w_j$ are linearly independent. We claim that they form a basis. Indeed, suppose they did not. Then by the Basis Extension Theorem applied to these n linearly independent vectors and the additional vector v_i (which together generate V), the elements $v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_n, w_j, v_i$ must form a basis. However, the vectors in this latter sequence are not linearly independent, since w_j is a linear combination of $v_1, \dots, v_n$ (another application of Proposition 7.15). This proves the claim. $\square$

Theorem 7.39. *If $v_1, v_2, \dots, v_n$ and $w_1, w_2, \dots, w_m$ are two bases of a vector space V, then $n = m$.*

Proof. Assume, without loss of generality, that $n > m$. By repeatedly applying the Exchange Lemma, we can successively replace v_1, v_2, $\dots$, v_n by some w_j and still have a basis. Since there are more v's than w's, the resulting sequence must have repetitions and therefore cannot be linearly independent, contradiction. $\square$

Theorem 7.39 implies that the following definition makes sense.

Definition 7.40. If a vector space V over a field F has a basis $(v_1, v_2, \dots, v_n)$, then $n \geq 0$ is called the *dimension* of V, written $n = \dim V = \dim_F V$, and we say that V is *finite-dimensional*. If V does not have a finite basis, then we write $\dim V = \infty$ and we say that V is *infinite-dimensional*.

Example 7.41. The empty sequence is a basis of the zero space, so $\dim \{0\} = 0$.

Example 7.42. The canonical basis of F^n has length n, so $\dim F^n = n$.

Example 7.43. Any line L in F^n that contains 0 is equal to $L(a)$ for some nonzero $a \in F^n$. The element a forms a basis for $L = L(a)$, so $\dim L = 1$.

Example 7.44. Let F be a field. The vector space $F[x]$ of all *polynomials* in the variable x with coefficients in F contains polynomials of arbitrarily high degree. The polynomials in any finite sequence $f_1, f_2, \dots, f_r$ have bounded degree, so they can not generate $F[x]$. This shows that no finite sequence of polynomials can form a basis for $F[x]$, so $\dim F[x] = \infty$.

Example 7.45. Let F be a field and $d \geq 0$ an integer. Then the monomials $1, x, x^2, \ldots, x^d$ form a basis for the vector space $F[x]_d$ of all polynomials of degree at most d (check this!), so $\dim F[x]_d = d + 1$.

Example 7.46. Let A be an $m \times n$ matrix with row echelon form A' and let r be the number of pivots in A', that is, the number of nonzero rows of A'. Then by Propositions 6.3 and 7.22 we have $\dim R(A) = \dim R(A') = r$ and $\dim \ker(A) = \dim \ker(A') = n - r$, and thus $\dim R(A) + \dim \ker(A) = n$.[1]

Theorem 7.47. *Let V be a vector space containing elements $v_1, \ldots, v_r$. Then the following statements hold.*

(1) *If $v_1, v_2, \ldots, v_r$ are linearly independent, then we have $r \leq \dim V$ with equality if and only if $(v_1, \ldots, v_r)$ is a basis for V.*
(2) *If $v_1, v_2, \ldots, v_r$ generate V, then we have $\dim V \leq r$ with equality if and only if $(v_1, \ldots, v_r)$ is a basis for V.*
(3) *If $r = \dim V$, then $v_1, \ldots, v_r$ are linearly independent if and only if they generate V.*

Proof. For (1), we are done if $\dim V = \infty$, so we assume that $\dim V$ is finite-dimensional, say $\dim V = s$ with a basis $w_1, w_2, \ldots, w_s$ for V. We apply the Basis Extension Theorem to the sequences $v_1, \ldots, v_r$ and $w_1, \ldots, w_s$. As we have

$$V = L(w_1, \ldots, w_s) = L(v_1, \ldots, v_r, w_1, \ldots, w_s),$$

we can extend $v_1, \ldots, v_r$ to a basis of length s. We immediately conclude $r \leq s = \dim V$ and equality holds if and only if $(v_1, \ldots, v_r)$ needs no extension, that is, it is already a basis.

For (2), we apply the Basis Extension Theorem to the empty sequence and the sequence $v_1, \ldots, v_r$. The empty sequence can be extended to a basis by adding suitably chosen elements from $v_1, \ldots, v_r$. As no element occurs doubly in such a basis (or it would not be linearly independent), the basis contains at most r elements, so $\dim V \leq r$.

If the inequality $\dim V \leq r$ is an equality, then each v_i is included in the basis, as otherwise some element would occur doubly. This shows that $v_1, \ldots, v_r$ are linearly independent, so $(v_1, \ldots, v_r)$ is a basis for V. Conversely, if $(v_1, \ldots, v_r)$ is a basis for V, then we have $\dim V = r$. Statement (3) follows from (1) and (2). □

Remark 7.48. Theorem 7.47(2) shows that if V is a finitely generated vector space, then V has a finite basis and a finite dimension.

Note that Theorem 7.47 yields a quite strong existence statement: if V is a vector space of dimension $\dim V = n$, then part (1) of Theorem 7.47 guarantees the *existence* of a nontrivial linear relation among any r elements $v_1, v_2, \ldots, v_r \in V$ whenever $r > n$ without the need to do any computation. This is very useful in many applications. On the other hand, it is quite a different matter to actually *find* such a relation: the proof is non-constructive and we usually need some computational method to exhibit an explicit relation.

---

[1] This argument uses the row echelon form and Proposition 7.22, which relies on Proposition 6.19, which tells us how to compute generators of the kernel. This proof can therefore be considered 'computational', which is the type of proofs we avoid as much as possible in this book. A computation-free proof will be given in Theorem 8.12.

Part (1) of Theorem 7.47 tells us that in a vector space of (finite) dimension n, the length of a linearly independent sequence of vectors is bounded by n. We can use this to show in another way that $\dim F[x] = \infty$ (see Example 7.44).

Example 7.49. Let F be a field. In Example 7.44 we showed that the vector space $F[x]$ of all *polynomials* in the variable x with coefficients in F is infinite-dimensional by showing that it can not be generated by finitely many polynomials. Using Theorem 7.47 we can give a new argument using linear independence. The space $F[x]$ contains the monomials $1, x, x^2, x^3, x^4, \ldots$, which are linearly independent, see Example 7.7. This means that we can find arbitrarily many linearly independent elements in $F[x]$, so $F[x]$ can not have a finite basis by Theorem 7.47(1). We conclude, again, $\dim F[x] = \infty$. Note that since $F[x] = L(\{x^n : n \in \mathbb{Z}_{\geq 0}\})$, we have shown that the collection $(x^n)_{n\in\mathbb{Z}_{\geq 0}}$ is a basis of $F[x]$.

With a little more effort, we can also show that the subspace $P(\mathbb{R})$ of $\mathbb{R}^{\mathbb{R}}$ of real *polynomial functions* does not have a finite basis either. Note that this follows from Example 7.49 if we use the fact that the bijection $\varphi\colon \mathbb{R}[x] \to P(\mathbb{R})$ from Remark 3.36 is an isomorphism, which in turn follows from the fact that φ is linear, as we have seen in Exercise 4.1.9. However, for the fact that φ is injective we used Exercise 11.3.8 from a later chapter, while the following example is independent of that.

Example 7.50. Let us consider again the linear subspace $P(\mathbb{R})$ of *polynomial functions* in $\mathcal{C}(\mathbb{R})$ (the vector space of continuous functions on $\mathbb{R}$), compare Example 3.35.

$$P(\mathbb{R}) = \{f \in \mathcal{C}(\mathbb{R}) : \exists n \in \mathbb{Z}_{\geq 0}\ \exists a_0, \ldots, a_n \in \mathbb{R}\ \forall x \in \mathbb{R} : f(x) = a_n x^n + \cdots + a_1 x + a_0\}$$

Denote as before by f_n the n-th power function: $f_n(x) = x^n$. We claim that the collection $(f_0, f_1, f_2, \ldots) = (f_n)_{n\in\mathbb{Z}_{\geq 0}}$ is linearly independent. Recall that this means that the only way of writing zero (that is, the zero function) as a *finite* linear combination of the f_j is with all coefficients equal to zero. If we let n be the largest number such that f_n occurs in the linear combination, then it is clear that we can write the linear combination as

$$\lambda_0 f_0 + \lambda_1 f_1 + \cdots + \lambda_n f_n = 0\,.$$

We have to show that this is only possible when $\lambda_0 = \lambda_1 = \cdots = \lambda_n = 0$.

Note that our assumption means that

$$\lambda_n x^n + \cdots + \lambda_1 x + \lambda_0 = 0 \qquad \text{for all } x \in \mathbb{R}.$$

There are various ways to proceed from here. For example, we can make use of the fact that a polynomial of degree $n \geq 0$ can have at most n zeros in $\mathbb{R}$. This is the theorem that we used without proof in Remark 3.36. Since there are infinitely many real numbers, the polynomial above has infinitely many zeros, hence it must be the zero polynomial.

Another possibility is to use *induction* on n (which, by the way, is implicit in the proof above: it is used in proving the statement on zeros of polynomials). Let us do this in detail. The *claim* we want to prove is

$$\forall n \in \mathbb{Z}_{\geq 0}\ \forall \lambda_0, \ldots, \lambda_n \in \mathbb{R} : \Big(\big(\forall x \in \mathbb{R} : \lambda_n x^n + \cdots + \lambda_0 = 0\big) \Longrightarrow \lambda_0 = \cdots = \lambda_n = 0\Big).$$

We now have to establish the *induction base*: the claim holds for $n = 0$. This is easy — let $\lambda_0 \in \mathbb{R}$ and assume that for all $x \in \mathbb{R}$, $\lambda_0 = 0$ (the function is constant here: it does not depend on x). Since there are real numbers, this implies $\lambda_0 = 0$.

Next, and this is usually the hard part, we have to do the *induction step*. We assume that the claim holds for a given n (this is the *induction hypothesis*) and deduce that it then also holds for $n+1$. To prove the statement for $n+1$, we have to consider coefficients $\lambda_0, \dots, \lambda_{n+1} \in \mathbb{R}$ such that for all $x \in \mathbb{R}$,

$$f(x) = \lambda_{n+1}x^{n+1} + \lambda_n x^n + \cdots + \lambda_1 x + \lambda_0 = 0\,.$$

Now we want to use the induction hypothesis, so we have to reduce this to a statement involving a polynomial of degree at most n. One way of doing that is to borrow some knowledge from Analysis about differentiation. This tells us that the derivative of f is zero again, and that it is a polynomial function of degree $\leq n$:

$$0 = f'(x) = (n+1)\lambda_{n+1}x^n + n\lambda_n x^{n-1} + \cdots + \lambda_1\,.$$

Now we can apply the induction hypothesis to this polynomial function; it tells us that $(n+1)\lambda_{n+1} = n\lambda_n = \cdots = \lambda_1 = 0$, hence $\lambda_1 = \cdots = \lambda_n = \lambda_{n+1} = 0$. So $f(x) = \lambda_0$ is in fact constant, which finally implies $\lambda_0 = 0$ as well (by our reasoning for the induction base).

This completes the induction step and therefore the whole proof of the fact that the collection $(f_n)_{n\in\mathbb{Z}_{\geq 0}}$ is linearly independent. From Proposition 7.47 we conclude $\dim P(\mathbb{R}) = \infty$.

Note that since $P(\mathbb{R}) = L(\{f_n : n \in \mathbb{Z}_{\geq 0}\})$, we have shown that the collection $(f_n)_{n\in\mathbb{Z}_{\geq 0}}$ is a basis for $P(\mathbb{R})$.

Example 7.51. We have inclusions

$$P(\mathbb{R}) \subset \mathcal{C}^\infty(\mathbb{R}) = \bigcap_{n=0}^{\infty} \mathcal{C}^n(\mathbb{R}) \subset \cdots \subset \mathcal{C}^2(\mathbb{R}) \subset \mathcal{C}^1(\mathbb{R}) \subset \mathcal{C}(\mathbb{R}) \subset \mathbb{R}^{\mathbb{R}}\,.$$

Since $P(\mathbb{R})$ contains arbitrarily long sequences of linearly independent functions, so do all these spaces and therefore they are all infinite-dimensional.

Warning 7.52. In Examples 7.49 and 7.50 we actually found infinite bases for $F[x]$ and $P(\mathbb{R}) \subset \mathbb{R}^{\mathbb{R}}$, but for example for $\mathbb{R}^{\mathbb{R}}$, it is a priori not at all clear that there even exists a collection C of functions in $\mathbb{R}^{\mathbb{R}}$ that is linearly independent and generates the whole vector space $\mathbb{R}^{\mathbb{R}}$. Using Zorn's Lemma, one can indeed show that all vector spaces do have a basis (see Appendix E), but, with the exception of Appendix E, we will not assume this in this book. By definition, the claim $\dim V = \infty$ only means that there is no finite basis, and does not directly state that there would exist an infinite basis.

The following proposition also justifies the word infinite-dimensional for those vector spaces that are not finite-dimensional.

Proposition 7.53. *Let V be a vector space. Then the following statements are equivalent.*

(1) *We have* $\dim V = \infty$.
(2) *The space V is not finitely generated.*
(3) *Every sequence $v_1, \ldots, v_n$ of n linearly independent elements in V can be extended to a sequence $v_1, \ldots v_n, v_{n+1}, \ldots, v_r$ of linearly independent vectors in V of arbitrary length $r \geq n$.*

Proof. The implication (1) $\Rightarrow$ (2) follows from part (2) of Theorem 7.47: if V were finitely generated, it would have finite dimension. For the implication (2) $\Rightarrow$ (3), assume that V is not finitely generated. Let $v_1, \ldots, v_n \in V$ be linearly independent vectors and set $U = L(v_1, \ldots, v_n)$. As these n vectors do not generate V, we have $U \subsetneq V$, so there is an element $v_{n+1} \in V$ with $v_{n+1} \notin U$. By Proposition 7.15, the vectors $v_1, \ldots, v_n, v_{n+1}$ are linearly independent. By induction to r, we can extend $v_1, \ldots, v_n$ to a sequence $v_1, \ldots v_n, v_{n+1}, \ldots, v_r$ of linearly independent vectors in V of arbitrary length $r \geq n$, which proves the implication (2) $\Rightarrow$ (3). For the final implication (3) $\Rightarrow$ (1), we assume that (3) holds. This implies that we can extend the empty sequence to a sequence of r linearly independent vectors in V for every $r > 0$. If the dimension of V were finite, then for $r = \dim V + 1$ we would get a contradiction with part (1) of Theorem 7.47. Hence, we conclude $\dim V = \infty$. $\square$

Exercises

7.3.1. Show that the real polynomials $f_1 = x^2 + 2$, $f_2 = 2x^2 - 3$, and $f_3 = x^3 + x - 1$ are linearly independent and extend them to a basis for the space $\mathbb{R}[x]_4$ of all real polynomials of degree at most 4. In other words, give polynomials $f_4, \ldots, f_t$ for a certain t, such that $(f_1, \ldots, f_t)$ is a basis for $\mathbb{R}[x]_4$.

7.3.2. Redo Exercise 7.1.4 using Theorem 7.47.

7.3.3. Let $V \subset \mathbb{R}^4$ be the hyperplane $V = a^\perp$ with $a = (1, 1, 1, 1)$.
(1) What is the dimension of V?
(2) Show that the vectors $v_1 = (2, -3, -1, 2)$ and $v_2 = (-1, 3, 2, -4)$ are linearly independent and contained in V.
(3) Extend (v_1, v_2) to a basis for V.

7.3.4. Let V be a finite-dimensional vector space and $S \subset V$ a subset that generates V.
(1) Show that there is a finite subset of S that generates V.
(2) Show that there is a finite subset of S of which the elements form a basis of V.

7.3.5. Let V be a vector space. Suppose there is an integer m such that for all linearly independent $v_1, v_2, \ldots, v_r \in V$ we have $r \leq m$. Prove that we have $\dim V \leq m$.

7.3.6. This exercise gives three alternative definitions for the dimension of a vector space. Let V be a vector space.
(1) Show that $\dim V$ equals the supremum (possibly ∞) of the set of all integers r for which there exists a sequence

$$\{0\} = V_0 \subsetneq V_1 \subsetneq V_2 \subsetneq \ldots \subsetneq V_{r-1} \subsetneq V_r = V$$

of subspaces of V, each properly contained in the previous.

(2) Show that $\dim V$ equals the supremum (possibly ∞) of the set of all integers r for which there exists a sequence

$$v_1, v_2, \ldots, v_r$$

of linearly independent elements in V (note that $r = 0$ is contained in this set).

(3) Show that $\dim V$ equals the infimum (possibly ∞) of the set of all integers r for which there exists a sequence

$$v_1, v_2, \ldots, v_r$$

of elements that generate V (the infimum of the empty set is ∞).

The Basis Extension Theorem as stated in Theorem 7.33 uses r linearly independent elements and s extra elements to generate V. The General Basis Extension Theorem E.1 of Appendix E also deals with infinite collections. It is proved using Zorn's Lemma. In the exercises below, we prove some partial generalisations that do not require Zorn's Lemma.

7.3.7. Let V be a vector space and $(v_j)_{j\in J}$ a (not necessarily finite) linearly independent collection of elements in V, labeled by an index set J. Prove the following statements.

(1) Let $(w_1, w_2, \ldots, w_s)$ be a sequence of elements of V such that $(v_j)_{j\in J}$ and $(w_1, w_2, \ldots, w_s)$ together generate V. Let $I \subset \{1, 2, \ldots, s\}$ be the set of indices i for which w_i is not a linear combination of $(v_j)_{j\in J}$ and $(w_1, w_2, \ldots, w_{i-1})$. Then $(v_j)_{j\in J}$ and $(w_i)_{i\in I}$ together form a basis for V.

(2) Let $(w_i)_{i\in\mathbb{Z}_{\geq 1}}$ be an infinite sequence of elements of V such that $(v_j)_{j\in J}$ and $(w_i)_{i\in\mathbb{Z}_{\geq 1}}$ together generate V. Let $I \subset \mathbb{Z}_{\geq 1}$ be the set of indices i for which w_i is not a linear combination of $(v_j)_{j\in J}$ and $(w_1, w_2, \ldots, w_{i-1})$. Then $(v_j)_{j\in J}$ and $(w_i)_{i\in I}$ together form a basis for V.

7.3.8. Let V be a vector space with a basis B.

(1) Let $v \in V$ be nonzero. Show that we can replace some element of B by v to obtain a basis B' of V that contains v.

(2) Let $v_1, v_2, \ldots, v_n \in V$ be linearly independent. Show that we can replace n elements of B by $v_1, \ldots, v_n$ to obtain a basis B' of V that contains $v_1, \ldots, v_n$.

7.4. Dimensions of subspaces

In the following proposition, and thereafter, we use the usual convention that $n < \infty$ for $n \in \mathbb{Z}_{\geq 0}$.

The following result shows that our intuition that dimension is a measure for the 'size' of a vector space is not too far off: larger spaces have larger dimension.

Lemma 7.54. *Let U be a linear subspace of the vector space V. Then we have $\dim U \leq \dim V$. If $\dim V$ is finite, then we have equality if and only if $U = V$.*

Note that in the case that $\dim V$ is finite, the statement also implies the existence of a finite basis of U.

Proof. There is nothing to show if $\dim V = \infty$. So let us assume $\dim V = n$ for some integer n. If $u_1, \ldots, u_r \in U$ are linearly independent, then $r \leq n$ by Theorem 7.47(1). From Proposition 7.53, applied to U, we conclude that the dimension of U is not infinite, say $\dim U = m$. Applying the same argument to a basis $(u_1, \ldots, u_m)$ for U gives $m \leq n$, so $\dim U \leq \dim V$.

To prove the second part, first assume $U \neq V$ and consider a basis B of U. It can be extended to a basis for V by the Basis Extension Theorem 7.33. Since B does not generate V, at least one element has to be added, which implies $\dim U < \dim V$. Conversely, obviously if $U = V$, then we have $\dim U = \dim V$. $\square$

Now we have the following nice formula relating the dimensions of subspaces U_1, U_2 of a vector space V to the dimension of their intersection $U_1 \cap U_2$ and their sum $U_1 + U_2$. We use the convention that $\infty + n = n + \infty = \infty + \infty = \infty$ for $n \in \mathbb{Z}_{\geq 0}$.

Theorem 7.55. *Let U_1 and U_2 be linear subspaces of a vector space V. Then*

$$\dim(U_1 + U_2) + \dim(U_1 \cap U_2) = \dim U_1 + \dim U_2 \,.$$

Proof. First note that the statement is trivially true when U_1 or U_2 is infinite-dimensional, since then both sides are ∞. So we can assume that U_1 and U_2 are both finite-dimensional.

We use the Basis Extension Theorem 7.33 again. Since U_1 is finite-dimensional, we know by Lemma 7.54 that its subspace $U_1 \cap U_2 \subset U_1$ is also finite-dimensional. Let $(v_1, \ldots, v_r)$ be a basis for $U_1 \cap U_2$. Using the Basis Extension Theorem, we can extend it on the one hand to a basis $(v_1, \ldots, v_r, w_1, \ldots, w_s)$ for U_1 and on the other hand to a basis $(v_1, \ldots, v_r, x_1, \ldots, x_t)$ for U_2. We claim that then $(v_1, \ldots, v_r, w_1, \ldots, w_s, x_1, \ldots, x_t)$ is a basis for $U_1 + U_2$. It is clear that these vectors generate $U_1 + U_2$ (since they are obtained by putting generating sets of U_1 and of U_2 together, see Lemma 3.41). So it remains to show that they are linearly independent. Consider a general linear relation

$$\lambda_1 v_1 + \cdots + \lambda_r v_r + \mu_1 w_1 + \cdots + \mu_s w_s + \nu_1 x_1 + \cdots + \nu_t x_t = 0 \,.$$

Then for $z = \nu_1 x_1 + \cdots + \nu_t x_t \in U_2$ we also have

$$z = -\lambda_1 v_1 - \cdots - \lambda_r v_r - \mu_1 w_1 - \cdots - \mu_s w_s \in U_1 \,,$$

so $z \in U_1 \cap U_2$, which implies that

$$z = \alpha_1 v_1 + \cdots + \alpha_r v_r$$

for suitable α_j, since $v_1, \ldots, v_r$ is a basis of $U_1 \cap U_2$. Since z has unique coefficients with respect to the basis $(v_1, \ldots, v_r, x_1, \ldots, x_t)$ for U_2 (see Lemma 7.30), we find $\alpha_i = 0$ for $1 \leq i \leq r$ and $\nu_j = 0$ for $1 \leq j \leq t$. Since z also has unique coefficients with respect to the basis $(v_1, \ldots, v_r, w_1, \ldots, w_s)$ for U_1, we also find $\mu_j = 0$ for $1 \leq j \leq s$ and $\lambda_i = -\alpha_i = 0$ for $1 \leq i \leq r$.

We conclude that $(v_1, \ldots, v_r, w_1, \ldots, w_s, x_1, \ldots, x_t)$ is indeed a linearly independent sequence and therefore a basis for $U_1 + U_2$. So we get $\dim(U_1 + U_2) = r + s + t$, $\dim(U_1 \cap U_2) = r$, $\dim U_1 = r + s$ and $\dim U_2 = r + t$, from which the claim follows. $\square$

Remark 7.56. Note the analogy with the formula

$$\#(X \cup Y) + \#(X \cap Y) = \#X + \#Y$$

for the number of elements in a set. However, there is no analogue of the corresponding formula for three sets:

$$\#(X \cup Y \cup Z) = \#X + \#Y + \#Z - \#(X \cap Y) - \#(X \cap Z) - \#(Y \cap Z) + \#(X \cap Y \cap Z) \,.$$

It is an exercise to find a vector space V and linear subspaces $U_1, U_2, U_3 \subset V$ such that

$$\dim(U_1 + U_2 + U_3) + \dim(U_1 \cap U_2) + \dim(U_1 \cap U_3) + \dim(U_2 \cap U_3) \neq \dim U_1 + \dim U_2 + \dim U_3 + \dim(U_1 \cap U_2 \cap U_3)\,.$$

For given dimensions of U_1 and U_2, we see that if the intersection $U_1 \cap U_2$ is relatively small, then the sum $U_1 + U_2$ is relatively big, and vice versa.

Note that if $U_1 \cap U_2 = \{0\}$, then we simply have $\dim(U_1 + U_2) = \dim U_1 + \dim U_2$ (and conversely). Complementary subspaces (see Definition 3.42) give an especially nice case.

Proposition 7.57. *If U_1 and U_2 are complementary subspaces in a vector space V, then we have*

$$\dim U_1 + \dim U_2 = \dim V.$$

Proof. Follows immediately from Theorem 7.55 and the fact that $U_1 \cap U_2 = \{0\}$ and $U_1 + U_2 = V$. □

Example 7.58. Let $a \in \mathbb{R}^n$ be nonzero and H the hyperplane $H = a^\perp$. By Example 7.43 we have $\dim(L(a)) = 1$. The subspaces $L(a)$ and H are complementary subspaces in F^n by Corollary 3.45, so Proposition 7.57 yields $\dim H = n-1$. In Example 8.21 we will see that the same holds for a hyperplane over any field F.

Example 7.59. Let L and V be a line and a plane in $\mathbb{R}^3$, both containing 0, so that they are subspaces. Then $\dim L = 1$ and $\dim V = 2$. By Theorem 7.55 we have

$$\dim(L \cap V) + \dim(L + V) = 1 + 2 = 3.$$

From $\dim(L + V) \geq \dim V = 2$, we find that there are two possibilities. The first possibility is $\dim(L+V) = 3$ and $\dim(L \cap V) = 0$, which means $L+V = \mathbb{R}^3$ and $L \cap V = \{0\}$. The second is $\dim(L+V) = 2$ and $\dim(L \cap V) = 1 = \dim L$, which implies $L \cap V = L$, so L is contained in V in this case.

We can use the Basis Extension Theorem to show the existence of complementary subspaces in finite-dimensional vector spaces.

Proposition 7.60. *Let V be a finite-dimensional vector space. If $U \subset V$ is a linear subspace, then there is a linear subspace $U' \subset V$ that is complementary to U.*

Proof. The subspace U is finite-dimensional by Proposition 7.54, say with basis $u_1, \ldots, u_m$. By the Basis Extension Theorem 7.33, we can extend this to a basis $u_1, \ldots, u_m, v_1, \ldots, v_n$ of V. Let $U' = L(v_1, \ldots, v_n)$. Then we clearly have $V = U + U'$ (Lemma 3.41). But we also have $U \cap U' = \{0\}$: if $v \in U \cap U'$, then

$$v = \lambda_1 u_1 + \cdots + \lambda_m u_m = \mu_1 v_1 + \cdots + \mu_n v_n\,,$$

for some coefficients $\lambda_1, \ldots, \lambda_m$ and $\mu_1, \ldots, \mu_n$, which gives

$$\lambda_1 u_1 + \cdots + \lambda_m u_m - \mu_1 v_1 - \cdots - \mu_n v_n = v - v = 0.$$

But $u_1, \ldots, u_m, v_1, \ldots, v_n$ are linearly independent, so this relation yields

$$\lambda_1 = \ldots = \lambda_m = \mu_1 = \ldots = \mu_n = 0,$$

and hence $v = 0$. □

Example 7.61. Given $U \subset V$, there usually are many complementary subspaces. For example, consider $V = \mathbb{R}^2$ and $U = \{(x,0) : x \in \mathbb{R}\}$. What are its complementary subspaces U'? We have $\dim V = 2$ and $\dim U = 1$, so we must have $\dim U' = 1$ as well. Let $u' = (x', y')$ be a basis of U'. Then $y' \neq 0$ (otherwise $0 \neq u' \in U \cap U'$). Then we can scale u' by $1/y'$ (replacing u', x', y' by $\frac{1}{y'}u', x'/y', 1$, respectively) to obtain a basis for U' of the form $u' = (x', 1)$, and $U' = L(u')$ then is a complementary subspace for every $x' \in \mathbb{R}$ — note that $U + U' = \mathbb{R}^2$ as every elements (x, y) can be written as $(x, y) = (x - yx', 0) + y(x', 1) \in U + U'$.

Remark 7.62. For any two subspaces U_1 and U_2 of a vector space V, we have $\dim(U_1 + U_2) \leq \dim V$ by Lemma 7.54. If V is finite-dimensional, then together with Theorem 7.55 this implies the inequality

$$\dim(U_1 \cap U_2) \geq \dim U_1 + \dim U_2 - \dim V.$$

Example 7.63. Let $a_1, a_2 \in \mathbb{R}^n$ be nonzero and H_i the hyperplane $H_i = \{a_i\}^\perp$ for $i = 1, 2$. Then $\dim H_i = n - 1$ by Example 7.58, so we have

$$n - 1 = \dim H_1 \geq \dim(H_1 \cap H_2) \geq \dim H_1 + \dim H_2 - \dim \mathbb{R}^n = n - 2.$$

Now there are two cases, namely $\dim(H_1 \cap H_2) = n-2$ and $\dim(H_1 \cap H_2) = n-1$. In the former case we have $\dim(H_1 + H_2) = n$, so $H_1 + H_2 = \mathbb{R}^n$ by Lemma 7.54. In the latter we have $H_1 \cap H_2 = H_1$ and thus $H_1 \subset H_2$; by symmetry we obtain $H_1 = H_2 = H_1 + H_2$. For $\mathbb{R}^3$ we conclude that two *different* planes that both contain 0 intersect in a subspace of dimension 1, that is, a line.

Exercises

7.4.1. (1) Let $U \subset F^n$ be a subspace of dimension $\dim U = 1$. Show that U is a line.
(2) Let $U \subset F^n$ be a subspace of dimension $\dim U = n - 1$. Show that U is a hyperplane. (See Example 8.22 for a clean proof.)

7.4.2. Let $d \geq 1$ be an integer, and for any $r \in \mathbb{R}$, let $U_r \subset \mathbb{R}[x]_d$ be the kernel of the evaluation map $\mathbb{R}[x]_d \to \mathbb{R}$ that sends f to $f(r)$.
(1) Prove $\dim U_r = d$ and give a basis for U_r.
(2) Prove that for $r, s \in \mathbb{R}$ with $r \neq s$, we have $\dim(U_r \cap U_s) = d - 1$ and give a basis for $U_r \cap U_s$.
(3) Prove that $U_r + U_s = \mathbb{R}[x]_d$.

7.4.3. Let U_1, U_2 be subspaces of a finite-dimensional vector space V satisfying $U_1 \cap U_2 = \{0\}$ and $\dim U_1 + \dim U_2 \geq \dim V$. Show that U_1 and U_2 are complementary subspaces.

7.4.4. Find a vector space V and linear subspaces $U_1, U_2, U_3 \subset V$ such that

$$\begin{aligned}\dim(U_1 + U_2 + U_3) + \dim(U_1 \cap U_2) + \dim(U_1 \cap U_3) + \dim(U_2 \cap U_3)\\ \neq \dim U_1 + \dim U_2 + \dim U_3 + \dim(U_1 \cap U_2 \cap U_3)\,.\end{aligned}$$

(See Remark 7.56.)

7.4.5. Let V be a vector space of dimension $\dim V = 10$. Let $U_1 \subset V$ and $U_2 \subset V$ be subspaces of dimensions $\dim U_1 = 6$ and $\dim U_2 = 7$, respectively. Prove that the intersection $U_1 \cap U_2$ is not zero.

7.4.6. Let F be a *finite* field, and consider the F-vector space $P(F) \subset F^F$ of polynomial functions as defined in Appendix D. Show that $\dim_F P(F)$ is finite.
This is in contrast with Example 7.50, which deals with infinite fields. It is also in contrast with Example 7.44, which deals with polynomials instead of polynomial functions (cf Warning D.8).

7.4.7. Let F be a finite field. Show that the map $\varphi\colon F[x] \to F^F$ of Exercise D.2.1 is not injective, cf. Exercise 4.1.9.

[Remark: one can show that if $q = |F|$, then the kernel of φ consists of all polynomials that are a multiple of $x^q - x$.]

CHAPTER 8

Ranks

8.1. The rank of a linear map

There is an important result that relates the dimensions of the kernel, image and domain of a linear map.

Definition 8.1. Let $f : V \to W$ be a linear map. Then we call the dimension of the image of f the *rank* of f: $\mathrm{rk}(f) = \dim \mathrm{im}(f)$.

Lemma 8.2. *Let $f : V \to W$ be a linear map. If f is surjective, then we have $\mathrm{rk}\, f = \dim W$. If W is finite-dimensional, then the converse is true as well.*

Proof. The map f is surjective if and only if the inclusion $\mathrm{im}\, f \subset W$ is an equality, so this follows from Lemma 7.54. $\square$

Theorem 8.3 (Dimension Formula for Linear Maps). *Let $f : V \to W$ be a linear map. Then*

$$\dim \ker(f) + \mathrm{rk}(f) = \dim V\,.$$

Proof. First we consider the case that V is finite-dimensional. By Proposition 7.60, there is a complementary subspace U of $\ker(f)$ in V and we have $\dim \ker f + \dim U = \dim V$ by Proposition 7.57.

Let $f' \colon U \to \mathrm{im}(f)$ be the restriction of f to U. We will show that f' is an isomorphism. Note that $\ker(f') = \ker(f) \cap U = \{0\}$, so f' is injective. To show that f' is also surjective, take $w \in \mathrm{im}(f)$. Then there is $v \in V$ such that $f(v) = w$. We can write $v = u' + u$ with $u' \in \ker(f)$ and $u \in U$ (see Lemma 3.44). Now

$$f'(u) = f(u) = f(v - u') = f(v) - f(u') = w - 0 = w\,,$$

so we have $w \in \mathrm{im}(f')$ as well. This implies that f' is surjective and thus an isomorphism. Since isomorphisms send bases to bases (see Corollary 7.32), we conclude $\dim U = \dim \mathrm{im}(f) = \mathrm{rk}\, f$ and therefore

$$\dim V = \dim \ker f + \dim U = \dim \ker f + \mathrm{rk}\, f.$$

Now consider the case $\dim V = \infty$. If $\mathrm{rk}\, f = \infty$, then we are done, so assume $\mathrm{rk}\, f = n$ for some integer n. Let r be any positive integer. Let $U \subset V$ be any r-dimensional subspace of V, which exists because we can take r linearly independent elements $v_1, \ldots, v_r \in V$ (see Proposition 7.53) and set $U = L(v_1, \ldots, v_r)$. Let $f' \colon U \to \mathrm{im}\, f$ be the linear map given by restricting f to U. Then by the finite-dimensional case, we have

$$\dim \ker f \geq \dim \ker f' = \dim U - \mathrm{rk}\, f' \geq \dim U - \dim \mathrm{im}\, f = r - n,$$

where the two inequalities follow from the inclusion $\ker f' \subset \ker f$ and the inclusion $\operatorname{im} f' \subset \operatorname{im} f$, respectively. Since r was an arbitrary positive integer, we conclude $\dim \ker f = \infty$, which proves the dimension formula for linear maps. $\square$

For a proof working directly with bases, see Chapter 4 in Jänich's book [**J**].

Example 8.4. Let $k \leq n$ be positive integers, and $F[x]_{n-k}$ and $F[x]_n$ the vector spaces of polynomials over F of degree at most $n-k$ and n, respectively. Let $\alpha_1, \alpha_2, \ldots, \alpha_k \in F$ be distinct elements, and set $p = (x-\alpha_1)(x-\alpha_2)\cdots(x-\alpha_k)$. The map $T\colon F[x]_{n-k} \to F[x]_n$ that sends an element f to $f \cdot p$ is linear and clearly injective, so the rank of T equals

$$\operatorname{rk} T = \dim F[x]_{n-k} - \dim \ker T = (n-k+1) - 0 = n-k+1.$$

The $(n-k+1)$-dimensional image of T consists of all polynomials in $F[x]_n$ that are multiples of p.

Let $S\colon F[x]_n \to F^k$ be the linear map that sends the polynomial $f \in F[x]_n$ to the sequence $\big(f(\alpha_1), f(\alpha_2), \ldots, f(\alpha_k)\big)$. Then for each $1 \leq i \leq k$, the map S sends the polynomial $p_i = p/(x-\alpha_i)$ to a nonzero multiple of $e_i \in F^k$, so these k images are linearly independent and thus $\operatorname{rk} S = \dim \operatorname{im} S \geq k$. Of course we also have $\dim \operatorname{im} S \leq k$, as $\operatorname{im} S$ is a subspace of F^k. Thus $\operatorname{rk} S = k$ and $\dim \ker S = \dim F[x]_n - \operatorname{rk} S = n+1-k$.

Clearly, the kernel $\ker S$ of S contains the image $\operatorname{im} T$ of T, and as they both have dimension $n-k+1$, we conclude $\ker S = \operatorname{im} T$. This shows that a polynomial f satisfies $f(\alpha_1) = f(\alpha_2) = \ldots = f(\alpha_k) = 0$ if and only if f is a multiple of p.

Corollary 8.5. *Let $f\colon V \to W$ be a linear map between finite-dimensional vector spaces with $\dim V = \dim W$. Then the following statements are equivalent.*

(1) *The map f is injective.*
(2) *The map f is surjective.*
(3) *The map f is an isomorphism.*

Proof. Note that f is injective if and only if $\dim \ker f = 0$ (Lemma 4.7) and f is surjective if and only if $\operatorname{rk}(f) = \dim W = \dim V$ (Lemma 8.2). By Theorem 8.3, these two statements are equivalent. $\square$

Example 8.6. Let $T\colon F[x]_n \to F[x]_n$ be the linear map that sends a polynomial f to $f + f'$, where f' is the derivative of f. Since f' has smaller degree than f, we have $\deg T(f) = \deg(f+f') = \deg f$. This shows that the only polynomial f with $T(f) = 0$, is $f = 0$, so T is injective and therefore, it is surjective. This proves, without explicit computations, that for every polynomial g, there is a polynomial f with $f + f' = g$.

Proposition 8.7. *Suppose $f\colon V \to W$ is a linear map of vector spaces. Then the following statements hold.*

(1) *If f is injective, then $\dim V \leq \dim W$.*
(2) *If f is surjective, then $\dim V \geq \dim W$.*
(3) *If f is an isomorphism, then $\dim V = \dim W$.*

Proof. If f is injective, then $\dim\ker f = 0$, so Theorem 8.3 yields

$$\dim V = \dim \operatorname{im} f \leq \dim W,$$

where the inequality follows from the inclusion $\operatorname{im} f \subset W$. If f is surjective, then $\operatorname{im} f = W$, so Theorem 8.3 yields $\dim V = \dim W + \dim\ker f \geq \dim W$. Implication (3) follows from (1) and (2). It also follows from the fact that isomorphisms send bases to bases (see Corollary 7.32). □

Example 8.8. We conclude, just from the dimensions, that the 3×4 matrix A of Example 5.10 induces a linear map $F^4 \to F^3$ that is not injective.

In Exercise 6.3.5 we could already prove that invertible matrices are square by using Proposition 6.20, which relied on the row echelon form. Instead of those computational arguments, we can now give a nicer proof.

Corollary 8.9. *Every invertible matrix is a square matrix.*

Proof. Suppose an $m \times n$ matrix A over F is invertible. Then the associated map $f_A \colon F^n \to F^m$ is an isomorphism, so we get $m = \dim F^m = \dim F^n = n$ by Proposition 8.7. □

Proposition 8.7(3) shows that if V and W are isomorphic, then $\dim V = \dim W$. The next proposition shows that the converse also holds if V and W are finite-dimensional. Together, these results show that essentially ('up to isomorphism'), there is only one F-vector space of any given dimension n (namely F^n, cf. Proposition 7.29).

Proposition 8.10. *If V and W are finite-dimensional vector spaces over the same field F with $\dim V = \dim W$, then V and W are isomorphic.*

Proof. If we have $\dim W = \dim V = n$, then V has a basis $B = (v_1, \ldots, v_n)$ and W has a basis $C = (w_1, \ldots, w_n)$, so $\varphi_B \colon F^n \to V$ and $\varphi_C \colon F^n \to W$ are isomorphisms by Proposition 7.29 and the composition $\varphi_C \circ \varphi_B^{-1} \colon V \to W$ is an isomorphism. □

In particular, we see that if V is an F-vector space of dimension $\dim V = n$, then V is isomorphic to F^n; indeed, an isomorphism is given by φ_B for any basis B for V. Note, however, that in general there is no *natural* (or *canonical*) isomorphism $V \xrightarrow{\sim} F^n$. The choice of isomorphism is equivalent to the choice of a basis, and there are many bases of V. In particular, we may want to choose different bases for V for different purposes, so it does not make sense to identify V with F^n in a specific way.

Exercises

8.1.1. Is the statement of Corollary 8.5 true without the assumption that V and W be finite-dimensional? If not, then give a counterexample and show where in the proof of Corollary 8.5 finite-dimensionality is used.

8.1.2. Let n be a positive integer and $F[x]_n$ the vector space of polynomials over F of degree at most n. Assume $\alpha_1, \alpha_2, \ldots, \alpha_{n+1} \in F$ are distinct elements. Let $S \colon F[x]_n \to F^{n+1}$ be the function given by

$$S(f) = \big(f(\alpha_1), f(\alpha_2), \ldots, f(\alpha_{n+1})\big)$$

as in Example 8.4 (for $k = n + 1$).

(1) Show that S is indeed a linear map as stated in Example 8.4.
(2) Show that S is surjective (cf. Example 8.4).
(3) Show that S is an isomorphism.
(4) Show that for every $i \in \{1, \ldots, n+1\}$, there is a unique polynomial $f_i \in F[x]_n$ such that $f_i(\alpha_j) = 1$ if $i = j$ and $f_i(\alpha_j) = 0$ if $i \neq j$.
(5) Show that $f_1, f_2, \ldots, f_{n+1}$ form a basis for $F[x]_n$.
(6) The polynomials $f_1, \ldots, f_{n+1}$ are called *Lagrange polynomials*. Give an explicit expression for them in terms of the elements $\alpha_1, \alpha_2, \ldots, \alpha_{n+1}$.

8.1.3. Let n be a positive integer and $T\colon \mathbb{R}[x]_n \to \mathbb{R}[x]_n$ the map that sends f to xf', where f' is the derivative of f. Show that T is a linear map and determine the rank of T.

8.1.4. Let $f\colon U \to V$ and $g\colon V \to W$ be linear maps of vector spaces.
(1) Show that we have $\operatorname{rk}(g \circ f) \leq \operatorname{rk} f$ with equality if (but not only if) g is injective.
(2) Show that we have $\operatorname{rk}(g \circ f) \leq \operatorname{rk} g$ with equality if (but not only if) f is surjective.

8.1.5. This exercise generalises Exercise 8.1.4. Let $f\colon U \to V$ and $g\colon V \to W$ be linear maps of vector spaces.
(1) Show that $\operatorname{rk}(g \circ f) \leq \operatorname{rk} f$ with equality if and only if $\operatorname{rk}(g \circ f) = \infty$ or $\ker g \cap \operatorname{im} f = \{0\}$.
(2) Show that $\operatorname{rk}(g \circ f) \leq \operatorname{rk} g$ with equality if and only if $\operatorname{rk}(g \circ f) = \infty$ or $\ker g + \operatorname{im} f = V$.

8.2. The rank of a matrix

Definition 8.11. Let $A \in \operatorname{Mat}(m \times n, F)$. Then the *rank* $\operatorname{rk} A$ of A is the rank of the associated linear map $f_A\colon F^n \to F^m$.

Recall that for a matrix $A \in \operatorname{Mat}(m \times n, F)$, the image of f_A equals the column space $C(A) \subset F^m$ of A (see Proposition 5.32). Therefore, $\operatorname{rk} A = \dim C(A) \leq m$, with equality if and only if $C(A) = F^m$ (see Lemma 7.54). Since the image $\operatorname{im} f_A = C(A)$ is generated by the n columns of A, we also have $\operatorname{rk} A \leq n$ by part (2) of Theorem 7.47. Hence, we have $\operatorname{rk} A \leq \min\{m, n\}$.

By this definition, the rank of A is the same as the *column rank* of A, that is, the dimension of the column space $C(A) \subset F^m$ of A. We can as well define the *row rank* of A to be the dimension of the row space $R(A) \subset F^n$ of A. Part (3) of the following theorem tells us that these additional notions are not really necessary, as the row rank of any matrix equals the column rank.

Theorem 8.12. *Let $A \in \operatorname{Mat}(m \times n, F)$ be a matrix. Then the following are true.*

(1) *We have* $\dim \ker A + \dim C(A) = n$.
(2) *We have* $\dim \ker A + \dim R(A) = n$.
(3) *We have* $\dim C(A) = \dim R(A)$.

Part (2) was already proved computationally (that is, using a row echelon form and Proposition 7.22, which uses Proposition 6.19) in Example 7.46. We will give several proofs of this important theorem. All except for the second alternative proof include a new computation-free proof of part (2).

Proof. Clearly, any two of the three statements imply the third. Statement (1) is true because it is a restatement of Theorem 8.3, so statements (2) and

(3) are equivalent. After repeatedly deleting from A some row that is a linear combination of the other rows, thus not changing the row space, we obtain an $r \times n$ matrix A' of which the rows are linearly independent. As the row spaces $R(A')$ and $R(A)$ are equal, we have $\ker A' = \ker A$ by Proposition 5.32, and therefore $\dim C(A') = \dim C(A)$ by statement (1). The r rows of A' form a basis of the row space $R(A')$, so we have $r = \dim R(A')$. The column space $C(A')$ is contained in F^r, so we find

$$\dim C(A) = \dim C(A') \leq \dim F^r = r = \dim R(A') = \dim R(A).$$

By symmetry, or applying the same argument to $A^\top$, we also get the opposite inequality $\dim R(A) \leq \dim C(A)$, so statement (3), and thus also (2), follows. □

First alternative proof. Again, any two of the three statements imply the third. Statement (1) is true because it is a restatement of Theorem 8.3, so statements (2) and (3) are equivalent.

Applying elementary row operations to A does not change $\ker A$ and $R(A)$ (see Proposition 6.3), so the truth of statement (2) is invariant under row operations, and therefore so is the truth of statement (3). Since statement (3) is symmetric in the rows and columns, the truth of both statements is also invariant under elementary column operations.

Using row and column operations, we can transform A into a matrix A' of which all entries are zero, except for some ones along the diagonal. For example, we could first use row operations to find the reduced row echelon form of A, then apply some permutation of the columns so that all pivots are along the diagonal, and finally apply column operations to make all non-diagonal entries zero; then A' would have the form of a block matrix

$$A' = \left(\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right).$$

It is clear that the row rank and column rank of A' both equal the number of ones along the diagonal, which proves statement (3) and therefore also (2). □

Second alternative proof. Statement (1) is true because it is a restatement of Theorem 8.3. Statement (2) is proved Example 7.46. Statement (3) follows from (1) and (2). □

Third alternative proof. Assume A' is as in the first proof. We now only give an alternative proof of one step of the first proof, namely that the equality $\ker A' = \ker A$ implies $\dim C(A') = \dim C(A)$.

So assume $\ker A' = \ker A$. Then the linear relations among the columns of A' correspond exactly with the linear relations among the columns of A. This means that for any maximal linearly independent subset of the columns of A (and thus a basis of the column space $C(A)$), the corresponding columns of A' form a maximal linearly independent subset of the columns of A', (and thus a basis of $C(A')$). This yields $\dim C(A') = \dim C(A)$. □

Remark 8.13. Statement (3) of Theorem 8.12 can be stated as $\operatorname{rk} A = \operatorname{rk} A^\top$.

Remark 8.14. By statement (3) of Theorem 8.12, the rank of a matrix A equals the row rank of A, which also equals the number of nonzero rows in a row equivalent matrix A' that is in row echelon form by Proposition 7.22.

Remark 8.15. The first proof, with the argument for the implication

$$\ker A' = \ker A \quad \Rightarrow \quad \dim C(A') = \dim C(A)$$

replaced by the argument in the third alternative proof, gives a proof of statement (3) that does not depend on (1). The second alternative proof contains a direct proof of statement (2). Together they imply (1), which gives an alternative proof of the dimension formula for linear maps between vector spaces F^n and F^m. Since every finite-dimensional vector space over F is isomorphic to F^n for some integer n (Proposition 8.10), we get a new proof of the dimension formula for general finite-dimensional vector spaces from Proposition 4.41.

Remark 8.16. In Proposition 7.22, we found that for an $m \times n$ matrix A in row echelon form with r nonzero rows, the $n-r$ elements w_k of Proposition 6.19 form a basis of the kernel $\ker A$ by showing that they are linearly independent and they generate $\ker A$. Theorem 8.12, statement (2), shows independently that the dimension of the kernel equals $n - r$ (independent as long as we do not use the second alternative proof). Using this and Theorem 7.47, we find that in order to reprove that the w_k form a basis for $\ker A$, it would suffices to show only one of the two: either that they are linearly independent or that they generate $\ker A$.

Example 8.17. Consider the matrix

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

over $\mathbb{R}$. The reduced row echelon form of A is

$$A' = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix},$$

which has two nonzero rows, so we find $\mathrm{rk}(A) = 2$.

Proposition 8.18. *For any $m \times n$ matrix A we have $\ker A = \{0\}$ if and only if $\mathrm{rk}\, A = n$.*

Proof. This follows immediately from Theorem 8.3. □

Remark 8.19. Corollary 7.11 states that n vectors $w_1, w_2, \ldots, w_n \in F^m$ are linearly independent if and only if the $m \times n$ matrix A of which the columns are $w_1, w_2, \ldots, w_n$ has kernel $\ker A = \{0\}$. By Proposition 8.18, this is the case if and only if $\mathrm{rk}\, A = n$. As we have $\mathrm{rk}\, A = \mathrm{rk}\, A^\top$ by Theorem 8.12, we may also check that the rank of $A^\top$, which has the n vectors as rows, equals n (cf. Remark 7.23).

Proposition 8.20. *Let F be a field, n a positive integer, and U a subspace of F^n. Then $\dim U + \dim U^\perp = n$ and $(U^\perp)^\perp = U$.*

Proof. By Lemma 7.54 there is a finite basis $v_1, v_2, \ldots, v_r$ for U. Let A be the $r \times n$ matrix of which the rows are $v_1, v_2, \ldots, v_r$. Then $R(A) = U$ and $\ker A = U^\perp$ by Proposition 5.32. The first equality follows immediately from

Theorem 8.12, statement (2). It implies

$$\dim(U^\perp)^\perp = n - \dim U^\perp = n - (n - \dim U) = \dim U,$$

and since U is contained in $(U^\perp)^\perp$ (Proposition 3.33), we conclude $(U^\perp)^\perp = U$ from Lemma 7.54. □

Example 8.21. Let $a \in F^n$ be nonzero and set $H = a^\perp = L(a)^\perp$. By Example 7.43, we have $\dim L(a) = 1$, so we find $\dim H = n - \dim L(a) = n - 1$.

Example 8.22. Let $U \subset F^n$ be a subspace of dimension $n-1$. Then $U^\perp$ has dimension 1, so there is a nonzero element $a \in U^\perp$ with $U^\perp = L(a)$. Then $U = (U^\perp)^\perp = L(a)^\perp = a^\perp$, so U is a hyperplane (cf. Exercise 7.4.1).

As in Example 8.22, we can think of any element $a \in F^n$ as an equation for the hyperplane $a^\perp$ (see Definition 3.12). Similarly, the elements of a subset $S \subset F^n$ correspond to equations for the subspace $S^\perp$ (See Definition 3.16).

Suppose $v_1, v_2, \ldots, v_m \in F^n$ generate a subspace $U \subset F^n$, and write $V = U^\perp$. Then we have $\{v_1, \ldots, v_m\}^\perp = V$, so $v_1, \ldots, v_m$ correspond to equations for V in this sense. Recall from Remark 5.33 that the space V is equal to the kernel of the $m \times n$ matrix M that has $v_1, v_2, \ldots, v_m$ as rows. After finding a row equivalent matrix M' in row echelon form, we can use Proposition 6.19 to find a set S of generators for $\ker M' = \ker M = V$. This way we go from equations to generators for V.

To go from generators to equations, we switch our point of view to U, for which $v_1, v_2, \ldots, v_m$ are generators. By Proposition 8.20, the set S can be viewed as a set of equations for U, in the sense that $S^\perp = L(S)^\perp = (U^\perp)^\perp = U$.

Example 8.23. Take $U \subset \mathbb{R}^4$ generated by $v_1 = (1,0,1,0)$ and $v_2 = (1,1,1,0)$. The kernel of the 2×4 matrix

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

with v_1 and v_2 is generated by $w_1 = (0,0,0,1)$ and $w_2 = (1,0,-1,0)$. We conclude that w_1 and w_2 correspond to equations for U in the sense that $\{w_1, w_2\}^\perp = U$.

Corollary 8.24. *Let U be a subspace of $\mathbb{R}^n$. Then U and $U^\perp$ are complementary subspaces.*

Proof. Suppose $x \in U \cap U^\perp$, so that we have $\langle x, x \rangle = 0$. Because we work over $\mathbb{R}$, we conclude $x = 0$, so we have $U \cap U^\perp = \{0\}$. From the dimension formula 7.55 and Proposition 8.20 we then find

$$\dim(U + U^\perp) = \dim U + \dim U^\perp - \dim(U \cap U^\perp) = n - 0 = n,$$

so from Lemma 7.54 we conclude $U + U^\perp = \mathbb{R}^n$ and U and $U^\perp$ are complementary spaces. □

For any subset $U \subset \mathbb{R}^n$, we call $U^\perp$ the *orthogonal complement* of U.

Warning 8.25. For some fields F, such as $\mathbb{F}_2$ and $\mathbb{C}$, there exist subspaces $U \subset F^n$ with $U \cap U^\perp \neq \{0\}$, so Corollary 8.24 is not true over general fields.

Exercises

8.2.1. Determine the rank of the matrices in Exercises 5.5.4 and 5.5.5.

8.2.2. Determine the rank of the matrices in Exercise 6.3.4.

8.2.3. Determine the rank of the linear maps and matrices of the exercises of Section 5.4.

8.2.4. Show that for any subset S of F^n, we have $L(S) = (S^\perp)^\perp$ (cf. Proposition 3.33 and Remark 3.34).

8.2.5. For the matrices A in Exercise 5.5.5, compute a basis for $(\ker A)^\perp$ and $(\operatorname{im} A)^\perp$ and determine the dimensions of these spaces.

8.2.6. Let l, m, n be non-negative integers. Suppose that A is an $l \times m$ matrix and B is an $m \times n$ matrix, so that the product AB exists. Prove the following statements (cf. Exercise 8.1.4).

(1) We have $\operatorname{rk} AB \leq \operatorname{rk} A$ with equality if (but not only if) $\operatorname{rk} B = m$.
(2) We have $\operatorname{rk} AB \leq \operatorname{rk} B$ with equality if (but not only if) $\operatorname{rk} A = m$.
(3) We have $\operatorname{rk} AB = \operatorname{rk} A$ if B is invertible.
(4) We have $\operatorname{rk} AB = \operatorname{rk} B$ if A is invertible.

8.3. Computing intersections

Proposition 8.26. *Suppose F is a field and $U_1, U_2 \subset F^n$ are subspaces. Then we have*

$$U_1 \cap U_2 = (U_1^\perp + U_2^\perp)^\perp \qquad \textit{and} \qquad (U_1 \cap U_2)^\perp = U_1^\perp + U_2^\perp.$$

Proof. In Proposition 3.33 we have already seen that $S^\perp \cap T^\perp = (S \cup T)^\perp$ for all subsets $S, T \subset F^n$. For $S = U_1^\perp$ and $T = U_2^\perp$ we obtain

$$U_1 \cap U_2 = (U_1^\perp)^\perp \cap (U_2^\perp)^\perp = (U_1^\perp \cup U_2^\perp)^\perp = \big(L(U_1^\perp \cup U_2^\perp)\big)^\perp = (U_1^\perp + U_2^\perp)^\perp,$$

where the first equality follows from Proposition 8.20, the second and third from Proposition 3.33 (part (4) and (2)), and the last from the definition of sums of subspaces (Definition 3.37). This proves the first identity of the proposition. Applying $(\_)^\perp$ to both sides gives the second identity by Proposition 8.20. □

Proposition 8.26 expresses taking intersections in terms of taking sums and orthogonal subspaces. If we view $U^\perp$ as a set of equations for the subspace U, as we did in the previous section, then Proposition 8.26 follows from the fact that if U_1 and U_2 are subspaces, each given by a set of linear equations, then the union of these sets is a set of equations for the intersection $U_1 \cap U_2$.

This allows us to explicitly compute generators for the intersection $U_1 \cap U_2$ if we know generators for the subspaces U_1 (or $U_1^\perp$) and U_2 (or $U_2^\perp$). Indeed, we already know how to take sums and orthogonal subspaces: if we have generating subsets S_1 and S_2 for two subspaces V_1 and V_2 of F^n, then the union $S_1 \cup S_2$ generates $V_1 + V_2$ by Lemma 3.41, and if $v_1, v_2, \ldots, v_r \in F^n$ generate a subspace $V \subset F^n$, then $V^\perp$ is the kernel of the matrix whose rows are $v_1, v_2, \ldots, v_n$ by Proposition 5.32 and we can compute generators for this kernel with Proposition 6.19.

Example 8.27. Let $U \subset \mathbb{R}^5$ be generated by the elements

$$\begin{aligned} u_1 &= (1, 3, 1, 2, 2), \\ u_2 &= (-1, 2, -2, 3, 2), \\ u_3 &= (3, 2, 0, -1, -4), \end{aligned}$$

and $V \subset \mathbb{R}^5$ by the elements

$$\begin{aligned} v_1 &= (-2, 0, -6, 3, -2), \\ v_2 &= (1, 2, -3, 1, -3), \\ v_3 &= (-1, 0, -3, -2, -1). \end{aligned}$$

To determine generators for the intersection $U \cap V$, we use the identity

$$U \cap V = (U^\perp + V^\perp)^\perp.$$

The subspaces $U^\perp$ and $V^\perp$ equal the kernels of the matrices

$$M = \begin{pmatrix} 1 & 3 & 1 & 2 & 2 \\ -1 & 2 & -2 & 3 & 2 \\ 3 & 2 & 0 & -1 & -4 \end{pmatrix} \qquad \text{and} \qquad N = \begin{pmatrix} -2 & 0 & -6 & 3 & -2 \\ 1 & 2 & -3 & 1 & -3 \\ -1 & 0 & -3 & -2 & -1 \end{pmatrix},$$

respectively, where the rows of M are u_1, u_2, u_3 and those of N are v_1, v_2, v_3. The reduced row echelon forms of M and N are

$$M' = \begin{pmatrix} 1 & 0 & 0 & -1 & -2 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix} \qquad \text{and} \qquad N' = \begin{pmatrix} 1 & 0 & 3 & 0 & 1 \\ 0 & 1 & -3 & 0 & -2 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix},$$

respectively. The dimensions of U and V equal the number of nonzero rows in M and N, respectively, so $\dim U = \dim V = 3$. By Proposition 6.27, the kernels $\ker M' = \ker M = U^\perp$ and $\ker N' = \ker N = V^\perp$ are generated by $\{w_4, w_5\}$ and $\{x_3, x_5\}$ respectively, with

$$w_4 = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \qquad w_5 = \begin{pmatrix} 2 \\ -1 \\ -1 \\ 0 \\ 1 \end{pmatrix}, \qquad x_3 = \begin{pmatrix} -3 \\ 3 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \qquad x_5 = \begin{pmatrix} -1 \\ 2 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Therefore, the subspace $U^\perp + V^\perp$ is generated by w_4, w_5, x_3, x_5, so the subspace $U \cap V = (U^\perp + V^\perp)^\perp$ is the kernel of the matrix

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

which has w_4, w_5, x_3, x_5 as rows. The reduced row echelon form of this matrix is

$$A' = \begin{pmatrix} 1 & 0 & 0 & 2 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

so the kernel $\ker A = \ker A' = U \cap V$ is generated by the vectors (now not written as column vectors)

$$z_4 = (-2, -1, -3, 1, 0) \qquad \text{and} \qquad z_5 = (-1, -1, 0, 0, 1).$$

Note that the row space of the last matrix equals $U^\perp + V^\perp$, so even without computing its kernel explicitly, we find $\dim(U^\perp + V^\perp) = 3$ and thus

$$\dim(U \cap V) = \dim(U^\perp + V^\perp)^\perp = 5 - \dim(U^\perp + V^\perp) = 2$$

by Proposition 8.20. We also conclude

$$\dim(U + V) = \dim U + \dim V - \dim(U \cap V) = 3 + 3 - 2 = 4.$$

Indeed, U and V are both contained in the 4-dimensional hyperplane H with normal $a = (2, -1, -1, 0, 1)$, so $U + V = H$. This is of course easier to verify immediately than through the computation we just did.

There is a different way to compute the intersection of two subspaces, based on the equality

$$U_1 \cap U_2 = (U_1^\perp)^\perp \cap U_2 = \{u \in U_2 \ : \ u \perp U_1^\perp\}.$$

Example 8.28. Let U and V be as in Example 8.27. Just as in Example 8.27, we first determine that $U^\perp = \ker M$ is generated by w_4 and w_5. This shows

$$U \cap V = (U^\perp)^\perp \cap V = \{v \in V \ : \ \langle v, w_4\rangle = \langle v, w_5\rangle = 0\}.$$

Every $v \in V$ can be written as $v = \lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3$ for some $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$. In terms of the λ_i, the equation $\langle v, w_k\rangle = 0$ (for $k = 4, 5$) is equivalent to

$$0 = \langle \lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3, w_k\rangle = \lambda_1\langle v_1, w_k\rangle + \lambda_2\langle v_2, w_k\rangle + \lambda_3\langle v_3, w_k\rangle,$$

so the two equations $\langle v, w_4\rangle = \langle v, w_5\rangle = 0$ are equivalent to $(\lambda_1, \lambda_2, \lambda_3)$ lying in the kernel of the matrix

$$\begin{pmatrix} \langle v_1, w_4\rangle & \langle v_2, w_4\rangle & \langle v_3, w_4\rangle \\ \langle v_1, w_5\rangle & \langle v_2, w_5\rangle & \langle v_3, w_5\rangle \end{pmatrix} = \begin{pmatrix} 1 & 0 & -3 \\ 0 & 0 & 0 \end{pmatrix}.$$

It turns out (as the bottom row is zero) that w_5 is orthogonal to V and this matrix is already in reduced row echelon form. Its kernel is generated by $(0, 1, 0)$ and $(3, 0, 1)$, which correspond to the vectors $0 \cdot v_1 + 1 \cdot v_2 + 0 \cdot v_3 = v_2$ and $3 \cdot v_1 + 0 \cdot v_2 + 1 \cdot v_3 = 3v_1 + v_3$. We conclude that $U \cap V$ is generated by v_2 and $3v_1 + v_3$.

Remark 8.29. The method you choose to compute an intersection $U_1 \cap U_2$ obviously depends on whether you have generators for U_i or equations (that is, generators for $U_i^\perp$), and whether you want generators for the intersection or equations. Also, if U_i requires many generators, then $U_i^\perp$ only needs few, so it is worth considering a method where you can do the bulk of the computation with $U_i^\perp$ instead of U_i. Another point to consider is that the method of Example 8.28 yields generators for $U_1 \cap U_2$ that are given as explicit linear combinations of the generators of U_1 and/or U_2, which in some applications is an advantage. The big advantage of the method of Example 8.27 is that it always yields a minimal number of generators, regardless of whether the number of given generators for U_1 and U_2 is minimal.

Exercises

8.3.1. Compute the intersection $U \cap V$ with U and V as in Example 8.27 with the method of Example 8.28, but with the roles of U and V reversed.

8.3.2. Let $F = \mathbb{F}_2$ be the field of two elements. Let $U \subset F^4$ be the subspace generated by

$$(1, 1, 1, 1), \qquad (1, 1, 0, 0), \qquad \text{and} \qquad (0, 1, 1, 0),$$

and let $V \subset F^4$ be the subspace generated by

$$(1,1,1,0) \qquad \text{and} \qquad (0,1,1,1).$$

Find generators for the intersection $U \cap V$.

8.3.3. Take two subspaces of $\mathbb{R}^6$ generated by four elements and compute generators for the intersection.

8.4. Inverses of matrices

Recall that every invertible matrix is square by Corollary 8.9. Proposition 5.25 shows that a matrix is invertible if and only if it has both a right and a left inverse. The following lemma implies that a *square* matrix A has a left inverse if and only if it has a right inverse, in which case A is invertible and these left and right inverses both equal A^{-1}.

Lemma 8.30. *Let A be an $n \times n$ matrix over F. Then the following statements are equivalent.*

(1) *The matrix A is invertible.*
(2) *The map f_A is injective.*
(3) *The map f_A is surjective.*
(4) *We have $\ker A = \ker f_A = \{0\}$.*
(5) *We have $\operatorname{rk} A = \operatorname{rk} f_A = n$.*
(6) *There exists an $n \times n$ matrix B such that $AB = I_n$.*
(7) *There exists an $n \times n$ matrix B such that $BA = I_n$.*

Moreover, if a matrix B as in (6) or (7) exists, then we have $B = A^{-1}$.

Proof. By definition, the matrix A is invertible when $f_A : F^n \to F^n$ is an isomorphism. Hence, Corollary 8.5 shows that the first three statements are equivalent. Lemmas 4.7 and 8.2 show that statements (2) and (3) are equivalent with (4) and (5), respectively. Clearly, statement (1) implies statements (6) and (7), as we may take $B = A^{-1}$. We finish the proof that all seven statements are equivalent by noting that the implication (6) $\Rightarrow$ (3) and the implication (7) $\Rightarrow$ (2) both follow from Lemma 5.23. Suppose a matrix B as in (6) exists. Then A is invertible by statement (1). From Proposition 5.25 with $C = A^{-1}$ we then conclude $B = A^{-1}$. If a matrix B as in (7) exists, then taking the transpose yields $A^\top B^\top = (BA)^\top = I_n$, which by the previous arguments means $B^\top = (A^\top)^{-1} = (A^{-1})^\top$, so $B = A^{-1}$. $\square$

Remark 8.31. Lemma 8.30 is analogous to the situation for functions. Suppose $f : X \to Y$ is a function between sets X and Y. If f is a bijection, then any left inverse g, that is, a function $g : Y \to X$ with $g \circ f = \mathrm{id}_X$, is the inverse of f; and any right inverse h, that is, a function $h : Y \to X$ with $f \circ h = \mathrm{id}_Y$, is the inverse of f. Moreover, if X and Y are finite sets of the same size, then f is injective if and only if it is surjective.

In this section, we will give a method to check whether a square matrix is invertible, and, if so, to compute the inverse.

Lemma 8.32. *Let A, B, C be matrices satisfying $AB = C$. Let A' be the matrix obtained from A by a sequence of elementary row operations, and let C' be the matrix obtained from C by the same sequence of operations. Then we have $A'B = C'$.*

Proof. By Proposition 6.4, there is an invertible matrix M, depending only on the applied sequence of row operations, such that $A' = MA$ and $C' = MC$. We immediately see $A'B = (MA)B = M(AB) = MC = C'$. Alternatively, the identity $A'B = C'$ also follows easily from the fact that the entries of C are the scalar products of the rows of A and the columns of B, and the fact that the scalar product is linear in its variables. $\square$

Lemma 8.32 states that if we start with a product $AB = C$, written as

$$
\begin{array}{rcl}
 & & \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix} = B \\
A = & \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{l1} & a_{l2} & \cdots & a_{lm} \end{pmatrix} & \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{l1} & c_{l2} & \cdots & c_{ln} \end{pmatrix} = C
\end{array} \tag{8.1}
$$

as in (5.6), and we perform an elementary row operation on the two bottom matrices A and C simultaneously, then we obtain the matrices A' and C' and, together with B, these resulting matrices depict the equality $A'B = C'$.

Given the matrices A and C, one might be interested in finding a matrix B such that $AB = C$, if such B exists. If A is invertible, then such a B does exist, as we have $B = A^{-1}(AB) = A^{-1}C$. If A^{-1} is known, then the matrix B is readily computed by multiplying A^{-1} with C. The following proposition gives a criterion for A being invertible and, if so, for determining $A^{-1}C$ efficiently if the inverse A^{-1} is not yet known.

Proposition 8.33. *A matrix $A \in \mathrm{Mat}(n, F)$ is invertible if and only if its reduced row echelon form is the identity matrix I_n. Suppose I_n is obtained from A by a sequence of elementary row operations. Then A^{-1} is obtained from I_n by the same sequence of operations. More generally, for any matrix C with n rows, the matrix $A^{-1}C$ is obtained from C by the same sequence of operations.*

Proof. If A is invertible, then f_A is injective, and by Proposition 6.20 we conclude that any row echelon form of A has n nonzero rows, so every row has a pivot and all pivots are on the diagonal; it follows that the *reduced* row echelon form is the identity matrix. Conversely, suppose that the reduced row echelon form of A is the identity matrix I_n. Then by Proposition 6.4 there is an invertible matrix B, such that $I_n = BA$, so A is invertible by Lemma 8.30. Applying Lemma 8.32 to the products $A \cdot A^{-1} = I_n$ and $A \cdot (A^{-1}C) = C$ and the sequence of elementary row operations that transform A into I_n, yields the last two statements. $\square$

Here is a visual interpretation of Proposition 8.33. If we write $X = A^{-1}C$ for A and C as in Proposition 8.33, then we can depict the equality $AX = C$ as in (8.1) by

$$\begin{array}{cc} & \boxed{X} \\ \boxed{A} & \boxed{C} \end{array}.$$

Applying elementary row operations to the combined matrix $\boxed{A}\boxed{C}$ yields a combined matrix $\boxed{A'}\boxed{C'}$ of matrices A' and C' that satisfy $A'X = C'$ by Lemma 8.32, depicted as follows.

$$\begin{array}{cc} & \boxed{X} \\ \boxed{A} & \boxed{C} \end{array} \quad \rightsquigarrow \quad \begin{array}{cc} & \boxed{X} \\ \boxed{A'} & \boxed{C'} \end{array}$$

In particular, if we obtain $A' = I$, then we have $C' = A'X = IX = X$.

$$\begin{array}{cc} & \boxed{X} \\ \boxed{A} & \boxed{C} \end{array} \quad \rightsquigarrow \quad \begin{array}{cc} & \boxed{X} \\ \boxed{I} & \boxed{X} \end{array}$$

Therefore, if a priori we do not yet know the matrix $X = A^{-1}C$, then we can find X by writing down the combined matrix $\boxed{A}\boxed{C}$ and applying row operations until the left part of the combined matrix equals I. The right part then automatically equals $X = A^{-1}C$.

Example 8.34. Let us see how to invert the following real matrix

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix}.$$

We perform the row operations on A and on I in parallel, as above.

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 4 & 0 & 1 & 0 \\ 1 & 3 & 9 & 0 & 0 & 1 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 3 & -1 & 1 & 0 \\ 0 & 2 & 8 & -1 & 0 & 1 \end{array}\right)$$

$$\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 0 & -2 & 2 & -1 & 0 \\ 0 & 1 & 3 & -1 & 1 & 0 \\ 0 & 0 & 2 & 1 & -2 & 1 \end{array}\right)$$

$$\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 3 & -3 & 1 \\ 0 & 1 & 0 & -\frac{5}{2} & 4 & -\frac{3}{2} \\ 0 & 0 & 1 & \frac{1}{2} & -1 & \frac{1}{2} \end{array}\right)$$

So

$$A^{-1} = \begin{pmatrix} 3 & -3 & 1 \\ -\frac{5}{2} & 4 & -\frac{3}{2} \\ \frac{1}{2} & -1 & \frac{1}{2} \end{pmatrix}.$$

Remark 8.35. This inversion procedure will also tell us whether a matrix A is invertible or not. Namely, if at some point in the computation of the row echelon form, the lower part of the next column has no non-zero entries, then the reduced row echelon form of A is not the identity, so the matrix is not invertible by Proposition 8.33.

Corollary 8.36. *If $A \in \mathrm{Mat}(m, F)$ is invertible, then A can be written as a product of matrices $L_i(\lambda)$ (for $\lambda \neq 0$) and $M_{ij}(\lambda)$ (for $i \neq j$) and N_{ij} of Section 6.1.*

Proof. By Proposition 8.33, the matrix A can be transformed into I_m by a sequence of elementary row operations. Let r be the number of operations. The i-th operation can also be obtained by multiplication from the left by an elementary matrix B_i, which is of the form $L_i(\lambda)$ (for $\lambda \neq 0$) or $M_{ij}(\lambda)$ (for $i \neq j$) or N_{ij}. We obtain $I_m = BA$ with $B = B_r B_{r-1} \cdots B_1$. Cf. the proof of Proposition 6.4. $\square$

Example 8.37. Let A be the matrix of Example 8.34 and $b \in F^3$ the vector

$$b = \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}.$$

Using the inverse A^{-1}, it is easy to find an element $x \in F^3$ with $Ax = b$, namely

$$x = A^{-1}(Ax) = A^{-1}b = \begin{pmatrix} 3 & -3 & 1 \\ -\frac{5}{2} & 4 & -\frac{3}{2} \\ \frac{1}{2} & -1 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} -8 \\ 9 \\ -2 \end{pmatrix}.$$

If we had not know A^{-1} yet, then we can apply Lemma 8.32 directly to the product $Ax = b$ and the sequence of row operations that transforms A into I_3, so that we need not compute A^{-1} first. We put A and b in an *extended matrix*

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & -1 \\ 1 & 2 & 4 & 2 \\ 1 & 3 & 9 & 1 \end{array}\right)$$

and transform the left part to I_3:

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & -1 \\ 1 & 2 & 4 & 2 \\ 1 & 3 & 9 & 1 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccc|c} 1 & 1 & 1 & -1 \\ 0 & 1 & 3 & 3 \\ 0 & 2 & 8 & 2 \end{array}\right)$$

$$\rightsquigarrow \left(\begin{array}{ccc|c} 1 & 0 & -2 & -4 \\ 0 & 1 & 3 & 3 \\ 0 & 0 & 2 & -4 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccc|c} 1 & 0 & 0 & -8 \\ 0 & 1 & 0 & 9 \\ 0 & 0 & 1 & -2 \end{array}\right),$$

so

$$x = \begin{pmatrix} -8 \\ 9 \\ -2 \end{pmatrix}.$$

Exercises

8.4.1. Determine the inverses of the following matrices

$$\begin{pmatrix} -3 & -1 \\ -2 & -1 \end{pmatrix}, \quad \begin{pmatrix} -1 & -2 & -1 \\ 1 & 3 & 1 \\ 1 & -2 & 0 \end{pmatrix}, \quad \begin{pmatrix} -1 & 2 & -2 \\ 0 & -1 & 0 \\ 1 & -2 & 3 \end{pmatrix}, \quad \begin{pmatrix} 0 & -1 & 0 & 1 \\ 3 & -2 & -2 & 1 \\ -1 & -2 & -2 & 0 \\ 0 & 0 & -1 & -1 \end{pmatrix}.$$

8.4.2. Are the matrices

$$\begin{pmatrix} 1 & 2 \\ -2 & 4 \end{pmatrix}, \quad \begin{pmatrix} -2 & 1 & -2 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix}$$

invertible?

8.4.3. Determine the inverse of those matrices (over $\mathbb{R}$) that are invertible.

$$\begin{pmatrix} 0 & -2 & -1 \\ -1 & 1 & 0 \\ -2 & -2 & 1 \end{pmatrix} \qquad \begin{pmatrix} -1 & 1 & -2 & 2 \\ -2 & 1 & 1 & -1 \\ 2 & -1 & 1 & 0 \\ 0 & 1 & 2 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 2 & -1 & 1 \\ -2 & -1 & -2 & 0 \\ 1 & 0 & -1 & 2 \\ 2 & 2 & 0 & 2 \end{pmatrix} \qquad \begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & -1 \\ 1 & 0 & 0 \end{pmatrix}.$$

8.4.4. Suppose the product AB of square matrices $A, B \in \mathrm{Mat}(n, F)$ is invertible. Prove that A and B are also invertible. Cf. Exercise 5.5.1.

8.4.5. Suppose M, N are $n \times n$ matrices with $MN = I_n$. Prove that then also $NM = I_n$.

8.4.6. Write the following matrices as a product of elementary matrices (see Section 6.1), if possible:

$$\begin{pmatrix} 1 & -1 & 0 \\ -1 & -2 & -1 \\ 2 & 2 & 1 \end{pmatrix} \quad \begin{pmatrix} -1 & 0 & -2 \\ -1 & -1 & -2 \\ 2 & 3 & 3 \end{pmatrix} \quad \begin{pmatrix} 2 & 3 & -2 \\ 3 & 2 & 2 \\ 0 & -1 & 2 \end{pmatrix}$$

8.5. Solving linear equations

As mentioned in the beginning of Chapter 6, the system

$$\begin{cases} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ & & & & & & & \vdots & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{cases}$$

of linear equations over F can be written as $Ax = b$ with

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \in \mathrm{Mat}(m \times n, F) \qquad \text{and} \qquad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \in F^m$$

and the vector

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

of unknowns. The solution set is the inverse image $f_A^{-1}(b)$, where $f_A \colon F^n \to F^m$ is the usual map that sends $x \in F^n$ to $Ax \in F^m$.

If $b = 0$, then the system is homogeneous and the solution set equals $\ker A$, for which we have seen in Chapter 6 how to find generators.

If $b \neq 0$, then the system is inhomogeneous. by Theorem 4.33 it suffices to do two things to solve the system: the first is to find a single solution, and if this exists, the second is to compute $\ker A$. Below we describe an algorithm to do both at once. For completeness, we also summarise an algorithm for the homogeneous case.

Algorithm for a homogeneous system. To solve a homogeneous system of linear equations $Ax = 0$, use elementary row operations to bring the matrix A

into (reduced) row echelon form; then read off a basis of the kernel (which is the solution space) according to Proposition 6.19 (or Proposition 6.27 for the reduced row echelon form).

Algorithm for an inhomogeneous system. To solve an inhomogeneous system of linear equations $Ax = b$, we do the same as in Example 8.37 (though this time we do not assume A is invertible). Let $A^\circ = (A|b)$ denote the extended matrix of the system (the matrix A with b attached as an $(n+1)$-st column). Use elementary row operations to bring A° into reduced row echelon form. The system is consistent if and only if b is a linear combination of the columns of A, so if and only if the last column of A° does not contain a pivot (see Proposition 6.21). In this case, the first n coordinates of $-w_{n+1}$ (in the notation of Proposition 6.27) give a solution of the system, but such a solution can also easily be found by solving the equations corresponding to the nonzero rows of the row echelon form from the bottom up. A basis of the solution space of the corresponding *homogeneous* system (needed to find the complete solution set with Theorem 4.33) can be read off from the first n columns of the reduced row echelon form of A°, as these form the reduced row echelon form of A.

To see that this algorithm is correct, we depict the system, as in Section 8.4, as

$$\begin{array}{rc} & \boxed{x} \\ \boxed{\quad A \quad} & \boxed{b} \end{array} .$$

Applying elementary row operations to the combined matrix $A^\circ = \boxed{\quad A \quad}\boxed{b}$ yields a combined matrix $\boxed{\quad A' \quad}\boxed{b'}$, for which the solution set to the equation $A'x = b'$ is the same as the solution set to the original equation $Ax = b$ by Lemma 8.32. Note that the last column of the row echelon form of A° does not contain a pivot if and only if the rank of the first n columns equals the rank of all $n + 1$ columns, that is, if and only if $\operatorname{rk}(A) = \operatorname{rk}(A^\circ)$. The latter is equivalent to saying that b is in the span of the columns of A, which is the image of the linear map f_A. The statement on how to find a solution is then easily verified.

The process of solving a system of linear equations by first bringing the corresponding matrix of coefficients into (reduced) row echelon form through elementary row operations, and then solving the system using the rows of the row echelon form from bottom to top is called *Gaussian elimination.* Indeed, when we add a multiple of a row with a pivot in the j-th column to the i-th row in order to make the (i, j)-entry zero, we are essentially eliminating the j-th variable from the equation corresponding to the i-th row.

Remark 8.38. Of course, if A is an invertible $n \times n$ matrix over F, then for any $b \in F^n$, the solution to the equation $Ax = b$ is just $x = A^{-1}b$ (cf. Remark 4.36).

Example 8.39. Consider the following system of linear equations:

$$\begin{array}{ccccccccc} x & + & y & + & z & + & w & = & 0 \\ x & + & 2y & + & 3z & + & 4w & = & 2 \\ x & + & 3y & + & 5z & + & 7w & = & 4 \end{array}$$

We will solve it according to the procedure outlined above. The extended matrix is

$$A^\circ = \left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 0 \\ 1 & 2 & 3 & 4 & 2 \\ 1 & 3 & 5 & 7 & 4 \end{array}\right).$$

We transform it into reduced row echelon form.

$$\left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 0 \\ 1 & 2 & 3 & 4 & 2 \\ 1 & 3 & 5 & 7 & 4 \end{array}\right) \longrightarrow \left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 2 & 3 & 2 \\ 0 & 2 & 4 & 6 & 4 \end{array}\right) \longrightarrow \left(\begin{array}{cccc|c} 1 & 0 & -1 & -2 & -2 \\ 0 & 1 & 2 & 3 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right) = A^{\circ\prime}$$

Since the last column of $A^{\circ\prime}$ does not contain the leading 1 of a row, the system is consistent. To find a solution, we find the element $w_{n+1} \in F^5$ in the kernel of $A^{\circ\prime}$, as in Proposition 6.19. It has a 1 as last coordinate and a 0 for all other coordinates that correspond to columns without a pivot. Hence, we have $w_{n+1} = (*, *, 0, 0, 1)$. Using the nonzero rows of $A^{\circ\prime}$, we determine the two remaining unknown coordinates, and we find $w_{n+1} = (2, -2, 0, 0, 1)$. Proposition 6.27 would have given this directly.

Hence, following our algorithm, we let $a \in F^4$ be the vector of the first 4 coordinates of $-w_{n+1}$, so a solution is given by $a = (x, y, z, w) = (-2, 2, 0, 0)$. It is easy to check that this is indeed a solution. Alternatively, if we write $A^{\circ\prime}$ as $(A'|b')$, then we could also find a by taking the coordinates corresponding to columns of A' without pivots to be 0 (so $a = (*, *, 0, 0)$), and solving for the remaining coordinates using the equation $A'a = b'$, working from the bottom nonzero row to the top.

The kernel of the non-extended matrix has basis (u, v) with $u = (1, -2, 1, 0)$ and $v = (2, -3, 0, 1)$. So all solutions are given by

$$(x, y, z, w) = a + ru + sv = (-2 + r + 2s, 2 - 2r - 3s, r, s)\,,$$

for some r and s.

Exercises

8.5.1. For each of the following systems of linear equations over $\mathbb{R}$, find a matrix A and a vector b, such that the system is equivalent with the equation $Ax = b$ in x. Then describe the full solution set.

$$\left\{\begin{array}{rrrcr} 2x_1+ & 3x_2+ & -2x_3 & = & 0 \\ 3x_1+ & 2x_2+ & 2x_3 & = & 0 \\ & -x_2+ & 2x_3 & = & 0 \end{array}\right.$$

$$\left\{\begin{array}{rrrcr} 2x_1+ & 3x_2+ & -2x_3 & = & 1 \\ 3x_1+ & 2x_2+ & 2x_3 & = & -1 \\ & -x_2+ & 2x_3 & = & -1 \end{array}\right.$$

$$\left\{\begin{array}{rrrcr} 2x_1+ & 3x_2+ & -2x_3 & = & 1 \\ 3x_1+ & 2x_2+ & 2x_3 & = & 1 \\ & -x_2+ & 2x_3 & = & 1 \end{array}\right.$$

$$\left\{\begin{array}{rrrrcr} 3x_1+ & x_2+ & 2x_3+ & -2x_4 & = & 1 \\ 2x_1+ & -x_2+ & 2x_3 & & = & 2 \\ x_1+ & & x_3 & & = & 3 \\ -2x_1+ & -x_2+ & -x_3+ & x_4 & = & 4 \end{array}\right.$$

8.5.2. The formula for trinitrotoluene (TNT) is $C_7H_5N_3O_6$. If it explodes, then the products of that reaction are N_2, H_2O, CO en C. Determine the balanced equation:

$$a{\cdot}C_7H_5N_3O_6 \to b{\cdot}N_2 + c{\cdot}H_2O + d{\cdot}CO + e{\cdot}C.$$

8.5.3. Consider the points $a = (1, 2, 1)$ and $b = (2, 1, -1)$ and $s = (-1, 4, 5)$ in $\mathbb{R}^3$. Set $\lambda = \frac{1}{3}$ and $\mu = \frac{2}{3}$.

(1) Verify that s lies in the plane $V = L(a, b)$.

(2) Find $p \in L(a)$ and $q \in L(b)$ such that $s = \lambda p + \mu q$.

CHAPTER 9

Linear maps and matrices

9.1. The matrix associated to a linear map

Proposition 8.10 shows that any finite-dimensional vector space V over a field F is isomorphic with F^n with $n = \dim V$. For any basis B for V, there is an isomorphism $\varphi_B \colon F^n \to V$ (Proposition 7.29). As we have seen in Proposition 4.41, this means that for all practical purposes, we can identify V and F^n, though we should keep in mind that the identification depends on the choice of a basis B. If we identify a second finite-dimensional vector space W over F with F^m for $m = \dim W$ (based on a choice of basis for W), then any linear map $f \colon V \to W$ corresponds with a linear map $F^n \to F^m$, which is given by some matrix. The following definition makes this precise.

Definition 9.1. Let F be a field and V, W finite-dimensional vector spaces over F with bases B and C, respectively, and dimensions $n = \dim V$ and $m = \dim W$. Then for every linear map $f \colon V \to W$, *the matrix associated to f with respect to the bases B and C*, denoted $[f]^B_C$, is the unique $m \times n$ matrix whose associated function as in Proposition 5.11 is the linear map $(\varphi_C^{-1} \circ f \circ \varphi_B) \colon F^n \to F^m$.

In the case $V = W$ and $B = C$, we also refer to $[f]^B_B$ as the matrix associated to f with respect to B.

If we write $M = [f]^B_C$, then by definition we have

$$f_M = \varphi_C^{-1} \circ f \circ \varphi_B.$$

If we identify the matrix M with the map $f_M \colon F^n \to F^m$ it defines, then we have the following commutative diagram.

(9.1)
$$\begin{tikzcd} V \arrow[r, "f"] & W \\ F^n \arrow[u, "\cong"', "\varphi_B"] \arrow[r, "{[f]^B_C}"'] & F^m \arrow[u, "\cong"', "\varphi_C"] \end{tikzcd}$$

Note that the map $\varphi_C^{-1} \circ f \circ \varphi_B \colon F^n \to F^m$ is nothing but the composition of (1) the identification of F^n with V, (2) the map $f \colon V \to W$, and (3) the identification of W with F^m. In other words, if we identify V with F^n and W with F^m, through the choice of bases B and C for V and W, respectively, then the map $f \colon V \to W$ corresponds with the map $F^n \to F^m$ given by the $m \times n$ matrix $[f]^B_C$.

Example 9.2. For the standard bases E_n and E_m for F^n and F^m, the maps φ_{E_n} and φ_{E_m} are the identity maps on F^n and F^m, respectively. Hence, for any linear map $f \colon F^n \to F^m$, the matrix $[f]^{E_n}_{E_m}$ is the matrix associated to f as in Proposition 5.11. This implies that for any $m \times n$ matrix A over F, and its associated linear map $f_A\colon F^n \to F^m$, we have

$$[f_A]^{E_n}_{E_m} = A.$$

Example 9.3. Let V be a finite-dimensional vector space over F with basis B and dimension n. Then the matrix $[\mathrm{id}_V]^B_B$ is the matrix whose associated function $F^n \to F^n$ equals $\varphi_B^{-1} \circ \mathrm{id}_V \circ \varphi_B = \mathrm{id}_{F^n}$, so $[\mathrm{id}_V]^B_B = I_n$.

Example 9.4. Let $\mathbb{R}[x]_3$ be the vector space of real polynomials of degree at most 3 with basis $B = (1, x, x^2, x^3)$. Let $D\colon \mathbb{R}[x]_3 \to \mathbb{R}[x]_3$ denote the map that sends $g \in \mathbb{R}[x]_3$ to its derivative g'.

$$\begin{array}{c} a_0 + a_1x + a_2x^2 + a_3x^3 \\ \big\uparrow \\ (a_0, a_1, a_2, a_3) \end{array} \qquad \begin{array}{ccc} \mathbb{R}[x]_3 & \xrightarrow{D} & \mathbb{R}[x]_3 \\ \cong\Big\uparrow \varphi_B & & \cong\Big\uparrow \varphi_B \\ \mathbb{R}^4 & \xrightarrow[{[D]^B_B}]{} & \mathbb{R}^4 \end{array}$$

Consider the composition $\varphi_B^{-1} \circ D \circ \varphi_B$. The map φ_B sends a quadruple (a_0, a_1, a_2, a_3) to the polynomial $g = a_0 + a_1x + a_2x^2 + a_3x^3$, of which the derivative $D(g) = g'$ equals $a_1 + 2a_2x + 3a_3x^2$, which in turn is identified through φ_B^{-1} with the quadruple $(a_1, 2a_2, 3a_3, 0)$. This means that the map associated to the matrix $[D]^B_B$ sends

$$(a_0, a_1, a_2, a_3) \qquad \text{to} \qquad (a_1, 2a_2, 3a_3, 0),$$

so the matrix equals

$$[D]^B_B = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Example 9.5. Let F be a field with k elements $\alpha_1, \alpha_2, \ldots, \alpha_k \in F$ and let n be a positive integer. Let $T\colon F[x]_n \to F^k$ be the linear map that sends a polynomial $g \in F[x]_n$ to the vector $\big(g(\alpha_1), \ldots, g(\alpha_k)\big)$. We determine the matrix associated to T with respect to the basis $B = (1, x, x^2, \ldots, x^n)$ for $F[x]_n$ and the standard basis E for F^k. Note that $\varphi_E\colon F^k \to F^k$ is the identity. Therefore, the composition $\varphi_E^{-1} \circ T \circ \varphi_B$ sends the j-th standard basis vector e_j to

$$\varphi_E^{-1}(T(\varphi_B(e_j))) = T(x^{j-1}) = (\alpha_1^{j-1}, \alpha_2^{j-1}, \ldots, \alpha_k^{j-1}).$$

By definition of the matrix $[T]^B_E$, this vector also equals $[T]^B_E \cdot e_j$, that is, the j-th column of $[T]^B_E$, cf. Lemma 5.9. Hence, we find

$$[T]^B_E = \begin{pmatrix} 1 & \alpha_1 & \alpha_1^2 & \cdots & \alpha_1^n \\ 1 & \alpha_2 & \alpha_2^2 & \cdots & \alpha_2^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_k & \alpha_k^2 & \cdots & \alpha_k^n \end{pmatrix}.$$

Such a matrix is called a *Vandermonde matrix.*

Definition 9.6. If V is a vector space over a field F of dimension n with basis $B = (v_1, \ldots, v_n)$, then we say that the n-tuple $a = (a_1, \ldots, a_n) \in F^n$ is the *sequence of coefficients* of the vector $v = \varphi_B(a) = a_1v_1 + \cdots + a_nv_n$ with respect to B, and we write $v_B = a = \varphi_B^{-1}(v)$.

Lemma 9.7. *Let $f\colon V \to W$ be a linear map, $B = (v_1, v_2, \ldots, v_n)$ a basis for V, and C a basis for W. Then for any $1 \leq j \leq n$, the j-th column of the $m \times n$ matrix $[f]_C^B$ is the sequence $f(v_j)_C$ of coefficients of $f(v_j)$ with respect to C.*

$$[f]_C^B = \begin{pmatrix} | & | & & | \\ f(v_1)_C & f(v_2)_C & \cdots & f(v_n)_C \\ | & | & & | \end{pmatrix}$$

Proof. As for any matrix, the j-th column of the matrix $[f]_C^B$ equals the image of the j-th standard basis vector e_j under the map associated to the matrix. By definition of $[f]_C^B$, this is equal to $(\varphi_C^{-1} \circ f \circ \varphi_B)(e_j) = \varphi_C^{-1}(f(v_j)) = f(v_j)_C$. $\square$

Example 9.8. Indeed, in Example 9.5, the columns are as described in Lemma 9.7. Also in Example 9.4, the j-th element in the basis B is x^{j-1}, and the j-th column of $[D]_B^B$ is the sequence of coefficients of $D(x^{j-1}) = (j-1)x^{j-2}$ with respect to the basis $B = (1, x, x^2, x^3)$.

Remark 9.9. If we identify $[f]_C^B$ with the linear map that it induces, then the commuting diagram (9.1) can also be expressed as $\varphi_C^{-1} \circ f = [f]_C^B \circ \varphi_B^{-1}$, that is, for each $v \in V$ we have

$$f(v)_C = [f]_C^B \cdot v_B.$$

In words: the sequence of coefficients of $f(v)$ with respect to C equals the product of the matrix $[f]_C^B$ with the sequence of coefficients of v with respect to B.

Example 9.10. The sequence $B = \big((x-1)^3, (x-1)^2, x-1, 1\big)$ is a basis for $F[x]_3$. Let C denote the usual basis $(1, x, x^2, x^3)$. Then the matrix associated to the identity map $\mathrm{id}\colon F[x]_3 \to F[x]_3$ with respect to the bases B and C is

$$[\mathrm{id}]_C^B = \begin{pmatrix} -1 & 1 & -1 & 1 \\ 3 & -2 & 1 & 0 \\ -3 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

This can be found directly from Lemma 9.7 (the j-th column contains the sequence of coefficients of $(x-1)^{4-j}$ with respect to C), but the identity

$$\begin{aligned} &a_1(x-1)^3 + a_2(x-1)^2 + a_3(x-1) + a_4 \\ &\qquad = (-a_1 + a_2 - a_3 + a_4) + (3a_1 - 2a_2 + a_3)x + (-3a_1 + a_2)x^2 + a_1x^3 \end{aligned}$$

also shows that $[\mathrm{id}]_C^B$ sends the quadruple (a_1, a_2, a_3, a_4) to

$$\big(-a_1 + a_2 - a_3 + a_4,\ \ 3a_1 - 2a_2 + a_3,\ \ -3a_1 + a_2,\ \ a_1\big).$$

Example 9.11. Let $V \subset \mathbb{R}^3$ be the plane spanned by $v_1 = (1, 2, 1)$ and $v_2 = (1, 1, 0)$. Then the vector $v_3 = (1, -1, 1)$ is a normal to V. Let B be the basis (v_1, v_2, v_3) of $\mathbb{R}^3$, and let $s = s_V\colon \mathbb{R}^3 \to \mathbb{R}^3$ denote the reflection in V. Note that $s(v_i) = v_i$ for $i = 1, 2$, and $s(v_3) = -v_3$. This means that the matrix associated to s with respect to B is easy to find; we have

$$[s]_B^B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Indeed, for any triple $a = (a_1, a_2, a_3) \in \mathbb{R}^3$ we have $[s]_B^B \cdot a = (a_1, a_2, -a_3)$, which corresponds to the fact that by linearity of s we have

$$s(\varphi_B(a)) = s(a_1v_1 + a_2v_2 + a_3v_3) = a_1v_1 + a_2v_2 - a_3v_3 = \varphi_B\left([s]_B^B \cdot a\right).$$

Example 9.12. Let $B = (v_1, v_2, v_3)$ be the basis for $\mathbb{R}^3$ as in Example 9.11, and let E be the standard basis for $\mathbb{R}^3$. Then $\varphi_E\colon \mathbb{R}^3 \to \mathbb{R}^3$ is the identity, which reflects the fact that the sequence of coefficients of a vector $v \in \mathbb{R}^3$ with respect to E is the vector v itself. Therefore, the columns of the matrix $[\mathrm{id}]_E^B$ are v_1, v_2, v_3 and we have

$$[\mathrm{id}]_E^B = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & -1 \\ 1 & 0 & 1 \end{pmatrix}.$$

Again, we can check for consistency by verifying that for $a = (a_1, a_2, a_3)$ we have

$$\mathrm{id}\left(\varphi_B(a)\right) = a_1v_1 + a_2v_2 + a_3v_3 = \begin{pmatrix} a_1 + a_2 + a_3 \\ 2a_1 + a_2 - a_3 \\ a_1 + a_3 \end{pmatrix} = \varphi_E\big([\mathrm{id}]_E^B \cdot a\big).$$

Example 9.13. Let E be the standard basis of F^n. Then every vector $v \in F^n$ is its own sequence of coefficients with respect to E, that is, $v_E = v$. This makes it easy to determine the matrix $[\mathrm{id}_{F^n}]_E^B$ for any basis B of F^n. Indeed, let $B = (w_1, w_2, \ldots, w_n)$ be a basis for F^n and let M be the $n \times n$ matrix whose columns are $w_1, w_2, \ldots, w_n$. Then we have $M = [\mathrm{id}_{F^n}]_E^B$ by Lemma 9.7.

As mentioned before, if we use the bases B and C of the vector spaces V and W to identify V and W with F^n and F^m, respectively, then every linear map $f\colon V \to W$ corresponds to a linear map $F^n \to F^m$, which in turn corresponds to an $m \times n$ matrix over F by Proposition 5.11. Exercise 9.1.5 makes this more precise by showing that this correspondence induces an isomorphism between $\mathrm{Hom}(V, W)$ and $\mathrm{Mat}(m \times n, F)$. In particular, we see that $\dim \mathrm{Hom}(V, W) = mn$.

Exercises

9.1.1. Let $\rho\colon \mathbb{R}^2 \to \mathbb{R}^2$ be the rotation around 0 over an angle of 90 degrees. Let $E = (e_1, e_2)$ be the standard basis of $\mathbb{R}^2$, and let $C = (v_1, v_2)$ be the basis with $v_1 = (1, 1)$ and $v_2 = (2, 1)$.

(1) Compute $\varphi_C^{-1}(\rho(\varphi_E(e_i)))$ for $i = 1$ and $i = 2$.

(2) Determine the matrix associated to $\varphi_C^{-1} \circ \rho \circ \varphi_E$ (as in Section 5.2, so with respect to the standard basis).

(3) Verify that your answer to the previous part equals the matrix $[\rho]_C^E$ as described in Lemma 9.7.

9.1.2. Let $T\colon \mathbb{R}[x]_4 \to \mathbb{R}[x]_4$ be the linear map given by $T(f) = 3f + (x - 2)f''$. Determine the matrix $[T]_B^B$ of T with respect to the basis $B = (1, x, x^2, x^3, x^4)$.

9.1.3. Let F be a field containing k distinct elements $\alpha_1, \alpha_2, \ldots, \alpha_k \in F$. Show that the square Vandermonde matrix

$$\begin{pmatrix} 1 & \alpha_1 & \alpha_1^2 & \cdots & \alpha_1^{k-1} \\ 1 & \alpha_2 & \alpha_2^2 & \cdots & \alpha_2^{k-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_k & \alpha_k^2 & \cdots & \alpha_k^{k-1} \end{pmatrix}.$$

is invertible, cf. Exercise 8.1.2 and Example 9.5.

9.1.4. Let V_1 be the vector space of 2×2 matrices over $\mathbb{R}$ and V_2 the vector space of 3×2 matrices over $\mathbb{R}$ with bases

$$B = \left(\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right)$$

and

$$C = \left(\begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \right),$$

respectively. Let $T\colon V_1 \to V_2$ be the linear map given by

$$T(M) = \begin{pmatrix} 3 & 7 \\ -1 & 5 \\ 8 & 2 \end{pmatrix} \cdot M.$$

Determine $[T]^B_C$.

9.1.5. Let B and C be bases for the F-vector spaces V and W of dimensions n and m, respectively. Show that the map

$$\mathrm{Hom}(V, W) \to \mathrm{Mat}(m \times n, F), \quad f \mapsto [f]^B_C$$

is an isomorphism (cf. Exercises 4.5.7 and 5.5.11).

9.2. The matrix associated to the composition of linear maps

Suppose U, V, W are finite-dimensional vector spaces of dimensions $\dim U = p$, $\dim V = n$, and $\dim W = m$, and with bases A, B, C respectively. Then for any linear maps $g\colon U \to V$ and $f\colon V \to W$, we get associated matrices $[g]^A_B$ and $[f]^B_C$. The two commuative diagrams as in (9.1) can be combined into one.

(9.2)
$$\begin{tikzcd} U \arrow[r, "g"] & V \arrow[r, "f"] & W \\ F^p \arrow[u, "\cong"', "\varphi_A"] \arrow[r, "{[g]^A_B}"'] & F^n \arrow[u, "\cong"', "\varphi_B"] \arrow[r, "{[f]^B_C}"'] & F^m \arrow[u, "\cong"', "\varphi_C"] \end{tikzcd}$$

Proposition 9.14. *With the notation as above, we have $[f \circ g]^A_C = [f]^B_C \cdot [g]^A_B$.*

Proof. The commutative diagram above simplifies to the following diagram.

$$\begin{tikzcd} U \arrow[r, "f \circ g"] & W \\ F^p \arrow[u, "\cong"', "\varphi_A"] \arrow[r, "{[f]^B_C \cdot [g]^A_B}"'] & F^m \arrow[u, "\cong"', "\varphi_C"] \end{tikzcd}$$

In other words, identifying matrices with the maps they induce, we obtain from the identities

$$[f]^B_C = \varphi_C^{-1} \circ f \circ \varphi_B \qquad \text{and} \qquad [g]^A_B = \varphi_B^{-1} \circ g \circ \varphi_A,$$

that

$$[f]^B_C \cdot [g]^A_B = \varphi_C^{-1} \circ (f \circ g) \circ \varphi_A = [f \circ g]^A_C,$$

which proves the statement. ☐

Alternative proof. Suppose $u \in U$ is any element. We apply Remark 9.9 twice. By first multiplying the matrix $[g]^A_B$ with the sequence u_A of coefficients of u

with respect to A, we obtain the sequence $(g(u))_B$ of coefficients of $g(u)$ with respect to B; multiplying the matrix $[f]^B_C$ with that vector yields the sequence $(f(g(u)))_C$ of coefficients of $f(g(u))$ with respect to C. In other words, we have
$$\big(f(g(u))\big)_C = [f]^B_C \cdot \big(g(u)\big)_B = [f]^B_C \cdot [g]^A_B \cdot u_A.$$
Similarly, we have
$$\big(f(g(u))\big)_C = \big((f \circ g)(u)\big)_C = [f \circ g]^A_C \cdot u_A.$$
This holds for all $u \in U$, in particular for the j-th element of the basis A, for which we have $u_A = e_j \in F^p$, so we find
$$[f]^B_C \cdot [g]^A_B \cdot e_j = [f \circ g]^A_C \cdot e_j$$
for all j. This shows that the two matrices $[f]^B_C \cdot [g]^A_B$ and $[f \circ g]^A_C$ have the same columns, so they are equal. □

Note that the order of f and g in the product $[f]^B_C \cdot [g]^A_B$ of matrices, and in the composition $f \circ g$, is opposite of the order in which they appear in diagram (9.2).

Corollary 9.15. *With the notation as above, if f is an isomorphism, then we have $[f^{-1}]^C_B = \big([f]^B_C\big)^{-1}$.*

Proof. If f is an isomorphism, then $m = n$, and $[f]^B_C$ is a square matrix. Apply Proposition 9.14 with $g = f^{-1}$ and $A = C$ to find
$$[f]^B_C \cdot [f^{-1}]^C_B = [\mathrm{id}]^C_C = I_m.$$
The statement follows. □

Example 9.16. Let B and E be the bases for $\mathbb{R}^3$ as in Examples 9.11 and 9.12. Then
$$[\mathrm{id}]^E_B = \big([\mathrm{id}]^B_E\big)^{-1} = \begin{pmatrix} -\frac{1}{3} & \frac{1}{3} & \frac{2}{3} \\ 1 & 0 & -1 \\ \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$
Since the sequence of coefficients of any vector $v \in \mathbb{R}^3$ with respect to E is equal to itself, we have
$$v_B = (\mathrm{id}(v))_B = [\mathrm{id}]^E_B \cdot v_E = [\mathrm{id}]^E_B \cdot v,$$
so the sequence v_B of coefficients of a vector $v \in \mathbb{R}^3$ with respect to B equals $[\mathrm{id}]^E_B \cdot v$. Indeed, the sequence of coefficients with respect to B of the j-th standard vector is the j-th column of $[\mathrm{id}]^E_B$, as we have
$$e_1 = -\tfrac{1}{3}v_1 + v_2 + \tfrac{1}{3}v_3, \qquad e_2 = \tfrac{1}{3}v_1 - \tfrac{1}{3}v_3, \qquad e_3 = \tfrac{2}{3}v_1 - v_2 + \tfrac{1}{3}v_3.$$

Example 9.17. Let $d\colon \mathbb{R}[x]_3 \to \mathbb{R}^4$ be the linear map that sends a polynomial $f \in \mathbb{R}[x]_3$ to
$$(f(2) + f'(2), f(3) + f'(3), f(4) + f'(4), f(5) + f'(5)),$$
where f' is the derivative of f. Then d is the composition of the map
$$d_1\colon \mathbb{R}[x]_3 \to \mathbb{R}[x]_3$$
that sends f to $f + f'$ and the map
$$d_2\colon \mathbb{R}[x]_3 \to \mathbb{R}^4$$

that sends g to $(g(2), g(3), g(4), g(5))$. With respect to the basis $B = (1, x, x^2, x^3)$ for $\mathbb{R}[x]_3$ and the standard basis E for $\mathbb{R}^4$, we get

$$[d]_E^B = [d_2]_E^B \cdot [d_1]_B^B = \begin{pmatrix} 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \\ 1 & 4 & 16 & 64 \\ 1 & 5 & 25 & 125 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 3 & 8 & 20 \\ 1 & 4 & 15 & 54 \\ 1 & 5 & 24 & 112 \\ 1 & 6 & 35 & 200 \end{pmatrix},$$

cf. Examples 9.4 and 9.5.

Exercises

9.2.1. Let $B = (v_1, v_2, v_3, v_4)$ be a basis for a vector space V over $\mathbb{R}$. Show that $B' = (v_1', v_2', v_3', v_4')$ with

$$\begin{aligned} v_1' &= v_1, \\ v_2' &= v_1 + 2v_2, \\ v_3' &= v_1 + 2v_2 + 3v_3, \\ v_4' &= v_1 + 2v_2 + 3v_3 + 4v_4 \end{aligned}$$

is also a basis for V.

(1) Determine the matrices $M = [\mathrm{id}]_B^{B'}$ and $N = [\mathrm{id}]_{B'}^B$.

(2) Explain that for $x = (x_1, x_2, x_3, x_4) \in \mathbb{R}^4$, the vector Mx is the sequence of coefficients with respect to B of the vector $v = x_1 v_1' + x_2 v_2' + x_3 v_3' + x_4 v_4'$.

(3) Explain that for $x = (x_1, x_2, x_3, x_4) \in \mathbb{R}^4$, the vector Nx is the sequence of coefficients with respect to B' of the vector $v = x_1 v_1 + x_2 v_2 + x_3 v_3 + x_4 v_4$.

9.2.2. Let $E = (e_1, e_2, e_3)$ be the standard basis for $\mathbb{R}^3$ and $C = (v_1, v_2, v_3)$ a basis with

$$v_1 = (-1, -2, 0), \quad v_2 = (-2, 1, 3), \quad v_3 = (1, -1, -2).$$

Determine the matrices $[\mathrm{id}]_E^C$ and $[\mathrm{id}]_C^E$.

9.3. Changing bases

Proposition 9.18. *Let $f\colon V \to W$ be a linear map of finite-dimensional vector spaces. Suppose B and B' are bases for V and C and C' are bases for W. Then we have*

$$[f]_{C'}^{B'} = [\mathrm{id}]_{C'}^{C} \cdot [f]_C^B \cdot [\mathrm{id}]_B^{B'}. \tag{9.3}$$

Proof. This follows immediately from Proposition 9.14. □

The following commuting diagram corresponds to the identity (9.3) of Proposition 9.18.

$$\begin{tikzcd}
V \arrow[r, "\mathrm{id}"] \arrow[rrr, bend left, "f"] & V \arrow[r, "f"] & W \arrow[r, "\mathrm{id}"] & W \\
F^n \arrow[u, "\varphi_{B'}"] \arrow[r, "{[\mathrm{id}]_B^{B'}}"'] \arrow[rrr, bend right, "{[f]_{C'}^{B'}}"'] & F^n \arrow[u, "\varphi_B"] \arrow[r, "{[f]_C^B}"'] & F^m \arrow[u, "\varphi_C"] \arrow[r, "{[\mathrm{id}]_{C'}^C}"'] & F^m \arrow[u, "\varphi_{C'}"]
\end{tikzcd}$$

In the spirit of the alternative proof of Proposition 9.14, we can explain the identity (9.3) as follows. Take a vector $v \in V$. By first multiplying the sequence $v_{B'}$ of coefficients of v with respect to B' with the matrix $[\mathrm{id}]_B^{B'}$, we obtain the sequence v_B of coefficients of v with respect to B; multiplying that vector with the matrix $[f]_C^B$ yields the sequence $(f(v))_C$ of coefficients of $f(v)$ with respect to C. Finally, multiplying this last vector with the matrix $[\mathrm{id}]_{C'}^C$ gives the sequence $(f(v))_{C'}$ of coefficients of $f(v)$ with respect to C'. This sequence could also have been obtained directly by multiplying $[f]_{C'}^{B'}$ with the vector $v_{B'}$. In other words, we have

$$[f]_{C'}^{B'} \cdot v_{B'} = \big(f(v)\big)_{C'} = \big([\mathrm{id}]_{C'}^C \cdot [f]_C^B \cdot [\mathrm{id}]_B^{B'}\big) \cdot v_{B'}$$

for all $v \in V$, in particular for the j-th element of the basis B', for which we have $v_{B'} = e_j \in F^n$. So we find

$$[f]_{C'}^{B'} \cdot e_j = \big([\mathrm{id}]_{C'}^C \cdot [f]_C^B \cdot [\mathrm{id}]_B^{B'}\big) \cdot e_j$$

for all j. This shows that the two matrices $[f]_{C'}^{B'}$ and $[\mathrm{id}]_{C'}^C \cdot [f]_C^B \cdot [\mathrm{id}]_B^{B'}$ have the same columns, so they are equal.

Note again that the order of the matrices in the right-hand side of (9.3) is opposite of the order in which they appear in this diagram. Because of Proposition 9.18, the matrices $[\mathrm{id}]_B^{B'}$ and $[\mathrm{id}]_{C'}^C$ associated as in Proposition 5.11 to the linear maps $\varphi_B^{-1} \circ \varphi_{B'} \colon F^n \to F^n$ and $\varphi_{C'}^{-1} \circ \varphi_C \colon F^m \to F^m$, respectively, are often called *basis change matrices*. The latter, for example, satisfies $[\mathrm{id}]_{C'}^C \cdot w_C = w_{C'}$ for all $w \in W$, so multiplying $[\mathrm{id}]_{C'}^C$ with the sequence w_C of coefficients of a vector w with respect to C gives the sequence $w_{C'}$ of coefficients of w with respect to C'.

Exercises

9.3.1. Let E_2 and E_3 be the standard bases of $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively. Let $T\colon \mathbb{R}^2 \to \mathbb{R}^3$ be the map given by

$$T\big((x,y)\big) = (3x+2y, x-y, -x+2y).$$

(1) Determine the matrix $[T]_{E_3}^{E_2}$.
(2) Determine the matrix $[T]_C^B$ for the basis $B = \big((1,2),(-1,1)\big)$ of $\mathbb{R}^2$ and the basis $C = (v_1, v_2, v_3)$ of $\mathbb{R}^3$ with the vectors

$$v_1 = (-1,-2,0), \quad v_2 = (-2,1,3), \quad v_3 = (1,-1,-2)$$

as in Exercise 9.2.2.

9.3.2. Let $V \subset \mathbb{R}^3$ be the subspace spanned by v_1 and v_3 as in Exercise 9.3.1. Then $B = (v_1, v_3)$ is a basis for V. Let $T\colon V \to \mathbb{R}^3$ be the inclusion map. Let E be the standard basis for $\mathbb{R}^3$. Let C be the basis for $\mathbb{R}^3$ as in Exercise 9.3.1.
(1) Determine the matrices $[T]_E^B$ and $[T]_C^B$ directly.
(2) Verify the equality that should hold between one of the matrices $[T]_E^B$ and $[T]_C^B$ on the one hand and the product of the other with $[\mathrm{id}]_E^C$ on the other hand.

9.3.3. Let B be a basis for F^n and let E be the standard basis.
(1) Show that we have $[\varphi_B]_B^E = I_n$.
(2) Show that for $M = [\mathrm{id}_{F^n}]_E^B$ we have $f_M = \varphi_B$.
(3) Show that we have $[\varphi_B]_E^E = [\mathrm{id}_{F^n}]_E^B$.

9.3.4. Let B and C be the standard bases of $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively. Let $T\colon \mathbb{R}^2 \to \mathbb{R}^3$ be the linear map given by

$$T\big((x,y)\big) = (2x-3y, x+y, 3x+y).$$

(1) Determine the matrix $[T]_C^B$.

(2) Determine the matrix $[T]^{B'}_{C'}$ for the basis $B' = \big((3,4),(1,-2)\big)$ for $\mathbb{R}^2$ and the basis $C' = (v_1, v_2, v_3)$ for $\mathbb{R}^3$ with

$$v_1 = (1,1,1), \qquad v_2 = (1,2,3), \qquad v_3 = (1,4,9).$$

(3) Verify that for the vector $v \in \mathbb{R}^2$ with $v_{B'} = (1,1)$ (that is, $v = \varphi_{B'}((1,1))$), we indeed have

$$[T]^{B'}_{C'} \cdot v_{B'} = \big(T(v)\big)_{C'}.$$

(4) Repeat this verification for $v_{B'} = (1,0)$ and $v_{B'} = (0,1)$.

9.4. Endomorphisms

In the special case of Proposition 9.18 that we have $V = W$, we can take $B = C$ and $B' = C'$ to obtain the following.

Proposition 9.19. *Let $f \colon V \to V$ be an endomorphism of a finite-dimensional vector space V with bases B and B'. Then we have*

$$[f]^{B'}_{B'} = [\mathrm{id}]^{B}_{B'} \cdot [f]^{B}_{B} \cdot [\mathrm{id}]^{B'}_{B} = [\mathrm{id}]^{B}_{B'} \cdot [f]^{B}_{B} \cdot \big([\mathrm{id}]^{B}_{B'}\big)^{-1}.$$

Proof. This follows immediately from Proposition 9.18 and Corollary 9.15. $\square$

Example 9.20. Let $B = (v_1, v_2, v_3)$ be the basis for $\mathbb{R}^3$ as in Examples 9.11, 9.12, and 9.16. As in Example 9.11, let s denote the reflection in the plane V spanned by v_1 and v_2. Then with the matrices of those examples, we find that the matrix associated to s with respect to the standard basis E is

$$\begin{aligned}[s]^{E}_{E} &= [\mathrm{id}]^{B}_{E} \cdot [s]^{B}_{B} \cdot [\mathrm{id}]^{E}_{B} = [\mathrm{id}]^{B}_{E} \cdot [s]^{B}_{B} \cdot \big([\mathrm{id}]^{B}_{E}\big)^{-1} \\ &= \begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & -1 \\ 1 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} -\frac{1}{3} & \frac{1}{3} & \frac{2}{3} \\ 1 & 0 & -1 \\ \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ -\frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{pmatrix}.\end{aligned}$$

Example 9.21. Let $B = (v_1, v_2, v_3)$ be the basis for $\mathbb{R}^3$ as in Example 9.20 and let $\pi \colon \mathbb{R}^3 \to \mathbb{R}^3$ be the orthogonal projection onto the plane V spanned by v_1 and v_2. Then we have $\pi(v_i) = v_i$ for $i = 1, 2$, and $\pi(v_3) = 0$, as v_3 is a normal to V. Therefore, we find

$$[\pi]^{B}_{B} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and as in Example 9.20, we find the matrix $[\pi]^{E}_{E}$ with Proposition 9.18:

$$\begin{aligned}[\pi]^{E}_{E} &= [\mathrm{id}]^{B}_{E} \cdot [\pi]^{B}_{B} \cdot [\mathrm{id}]^{E}_{B} = [\mathrm{id}]^{B}_{E} \cdot [\pi]^{B}_{B} \cdot \big([\mathrm{id}]^{B}_{E}\big)^{-1} \\ &= \begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & -1 \\ 1 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} -\frac{1}{3} & \frac{1}{3} & \frac{2}{3} \\ 1 & 0 & -1 \\ \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} & \frac{2}{3} \end{pmatrix}.\end{aligned}$$

Exercises

9.4.1. Let B be the basis $\big(1, 1+x, 1+x+x^2, 1+x+x^2+x^3\big)$ for $\mathbb{R}[x]_3$. Let $T \colon \mathbb{R}[x]_3 \to \mathbb{R}[x]_3$ be the linear map given by $T(f) = f'$.

(1) Determine the matrix $[T]^{B}_{B}$ directly.

(2) Determine the matrix $[T]^{B}_{B}$ by first determining the matrix $[T]^{C}_{C}$ for the basis $C = (1, x, x^2, x^3)$, and then using a basis change matrix.

9.4.2. Let $L \subset \mathbb{R}^2$ be the line given by $y = 2x$. Let $\pi \colon \mathbb{R}^2 \to \mathbb{R}^2$ be the orthogonal projection of $\mathbb{R}^2$ on L.

(1) Determine $[\pi]_B^B$, where B is the standard basis.
(2) Determine v_1 and v_2 such that (v_1) is a basis for L and (v_2) is a basis for $L^\perp$. Set $C = (v_1, v_2)$. Determine $[\pi]_C^C$.
(3) Determine $[\pi]_B^B$ again, this time using $[\pi]_C^C$ and a basis change matrix.

9.4.3. Let $V \subset \mathbb{R}^3$ be the plane given by $x + 3y - 2z = 0$. Let $\pi \colon \mathbb{R}^3 \to \mathbb{R}^3$ be the orthogonal projection of $\mathbb{R}^3$ on V. Let B be the standard basis for $\mathbb{R}^3$.

(1) Determine $[\pi]_B^B$ directly.
(2) Determine $[\pi]_B^B$ via $[\pi]_C^C$, where $C = (v_1, v_2, v_3)$ is a basis consisting of a basis (v_1, v_2) for V and a basis (v_3) for $V^\perp$.

9.5. Similar matrices and the trace

Definition 9.22. We say that two $n \times n$ matrices M and M' are *similar* if there is an invertible $n \times n$ matrix Q such that $M' = QMQ^{-1}$.

The notion of similarity defines an equivalence relation on $\mathrm{Mat}(n, F)$ (see Exercise 9.5.2). Proposition 9.19 shows that any two matrices associated to the same endomorphism of V, but with respect to different bases, are similar. The converse, namely that any two similar $n \times n$ matrices are associated to the same endomorphism with respect to two appropriately chosen bases, will be proved in the next section (see Proposition 9.29).

The next section also touches on the classification of matrices with respect to similarity, which is complicated. For purposes of classification, it is useful to have *invariants*, that is, functions that are constant on the equivalence classes.

The rank is an invariant with respect to similarity, that is, any two similar matrices have the same rank (see Exercise 9.5.3). Here is another invariant (shown to be invariant in Corollary 9.25).

Definition 9.23. For $A = (a_{ij}) \in \mathrm{Mat}(n, F)$, we define the *trace* of A to be
$$\mathrm{Tr}(A) = a_{11} + a_{22} + \cdots + a_{nn}\,.$$

Lemma 9.24. *If $A \in \mathrm{Mat}(m \times n, F)$ and $B \in \mathrm{Mat}(n \times m, F)$, then*
$$\mathrm{Tr}(AB) = \mathrm{Tr}(BA)\,.$$

Proof. The (i,i)-entry of AB is $\sum_{j=1}^n a_{ij}b_{ji}$. The (j,j)-entry of BA is $\sum_{i=1}^m b_{ji}a_{ij}$. So we get
$$\mathrm{Tr}(AB) = \sum_{i=1}^m \sum_{j=1}^n a_{ij}b_{ji} = \sum_{j=1}^n \sum_{i=1}^m b_{ji}a_{ij} = \mathrm{Tr}(BA)\,.$$
□

Corollary 9.25. *Let $A, A' \in \mathrm{Mat}(n, F)$ be similar. Then* $\mathrm{Tr}(A) = \mathrm{Tr}(A')$.

Proof. There is an invertible matrix $Q \in \mathrm{Mat}(n, F)$ such that $A' = QAQ^{-1}$. It follows from Lemma 9.24 that
$$\mathrm{Tr}(A') = \mathrm{Tr}(QA \cdot Q^{-1}) = \mathrm{Tr}(Q^{-1} \cdot QA) = \mathrm{Tr}(A)\,.$$
□

This allows us to make the following definition.

Definition 9.26. Let V be a finite-dimensional F-vector space and $f : V \to V$ an endomorphism of V. We define the *trace* $\mathrm{Tr}(f)$ of f to be the trace of any matrix associated to f relative to some basis of V.

Note that $\mathrm{Tr}(f)$ is well-defined, since all matrices associated to f are similar by Proposition 9.19 and therefore have the same trace according to Corollary 9.25.

In the next chapter, we will introduce another invariant, which is even more important than the trace: the determinant.

Exercises

9.5.1. Determine the trace of the following three matrices.

$$M_1 = \begin{pmatrix} 1 & 2 & 2 & 1 \\ 4 & -3 & 5 & 2 \\ -2 & 1 & 5 & 11 \\ 3 & 2 & 7 & -13 \end{pmatrix}$$

$$M_2 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 4 & 9 & 16 \\ 1 & 8 & 27 & 64 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 2 & 2 & 1 \\ 4 & -3 & 5 & 2 \\ -2 & 1 & 5 & 11 \\ 3 & 2 & 7 & -13 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 4 & 9 & 16 \\ 1 & 8 & 27 & 64 \end{pmatrix}$$

$$M_3 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 5 & 7 \\ 1 & 25 & 49 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 5 & 6 \\ 0 & 2 & 7 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 2 & 2 \\ 4 & -3 & 5 \\ -2 & 1 & 5 \end{pmatrix} \begin{pmatrix} 1 & 5 & 6 \\ 0 & 2 & 7 \\ 0 & 0 & 3 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 5 & 7 \\ 1 & 25 & 49 \end{pmatrix}$$

9.5.2. Show that the notion of similarity defines an equivalence relation on the space $\mathrm{Mat}(n, F)$ of $n \times n$ matrices, as claimed.

9.5.3. Show that any two similar matrices have the same rank.

9.6. Classifying matrices

9.6.1. Similar matrices. Proposition 9.19 shows that any two matrices associated to the same endomorphism of V, but with respect to different bases, are similar. Conversely, Proposition 9.29 implies that for any two similar $n \times n$ matrices M_1, M_2 over F, there are an endomorphism f of F^n and two bases B_1 and B_2 for F^n such that $M_i = [f]^{B_i}_{B_i}$ for $i \in \{1, 2\}$. The proof of Proposition 9.29 uses the following lemma.

Lemma 9.27. *Suppose V is an n-dimensional vector space over F with basis B. Then for every invertible $n \times n$ matrix P, there is a basis B' for V such that $[\mathrm{id}_V]^{B'}_B = P$.*

Proof. Set $w_j = \varphi_B(P \cdot e_j)$ for all $1 \le j \le n$, and set $B' = (w_1, w_2, \ldots, w_n)$. Then we have $\varphi_{B'} = \varphi_B \circ f_P$. The map φ_B is an isomorphism by Proposition 7.29 and f_P is an isomorphism because P is invertible, so their composition $\varphi_{B'}$ is invertible as well and B' is a basis by Proposition 7.29. From $f_P = \varphi_B^{-1} \circ \mathrm{id}_V \circ \varphi_{B'}$, we conclude $P = [\mathrm{id}_V]^{B'}_B$. □

Example 9.28. For $V = F^n$ and $B = E$, the standard basis for F^n, we can make this much more concrete. Let P be an every invertible $n \times n$ matrix, and let $w_1, \ldots, w_n \in F^n$ be the columns of P. Then the sequence $B' = (w_1, \ldots, w_n)$ is linearly independent, so it is a basis for F^n. We have $P = [\mathrm{id}]_E^{B'}$ by Example 9.13. See Exercise 9.6.1 for the case that B is any basis of F^n.

Proposition 9.29. *Let M and M' be two similar $n \times n$ matrices over F. Then there exists a basis B of F^n such that for $M' = [f_M]_B^B$.*

Proof. Since M' and M are similar, there is an invertible $n \times n$ matrix P such that $M' = P^{-1}MP$. By Lemma 9.27 there is a basis B such that $[\mathrm{id}_{F^n}]_E^B = P$, where E is the standard basis for F^n. Then we have

$$M' = P^{-1}MP = \left([\mathrm{id}_{F^n}]_E^B\right)^{-1} \cdot [f_M]_E^E \cdot [\mathrm{id}_{F^n}]_E^B = [\mathrm{id}_{F^n}]_B^E \cdot [f_M]_E^E \cdot [\mathrm{id}_{F^n}]_E^B = [f_M]_B^B.$$

□

The classification of matrices in $\mathrm{Mat}(n, F)$ with respect to similarity is complex. What is still easy, is that the 'multiplication by λ' endomorphism (for $\lambda \in F$) has matrix λI_n regardless of the basis, and so λI_n and μI_n are not similar if $\lambda \neq \mu$.

Before we give a more complex example, we state the following lemma.

Lemma 9.30. *Suppose that the $n \times n$ matrices M and N over F are similar. Then for every scalar $\lambda \in F$, the matrices $M - \lambda I_n$ and $N - \lambda I_n$ are similar as well.*

Proof. Suppose M and N are similar. Then there is an invertible matrix Q such that $N = QMQ^{-1}$. Then the identity

$$Q(M - \lambda I_n)Q^{-1} = QMQ^{-1} - Q(\lambda I_n)Q^{-1} = N - \lambda Q I_n Q^{-1} = N - \lambda I_n$$

shows that $M - \lambda I_n$ and $N - \lambda I_n$ are similar as well. □

Example 9.31. Consider the real matrices

$$M_{\lambda,t} = \begin{pmatrix} \lambda & t \\ 0 & \lambda \end{pmatrix}$$

with trace $\mathrm{Tr}(M_{\lambda,t}) = 2\lambda$. Since any two similar matrices have the same trace, we find that $M_{\lambda,t}$ and $M_{\mu,u}$ can be similar only when $\lambda = \mu$. We have

$$\dim \ker(M_{\lambda,t} - \lambda I_2) = \begin{cases} 1 & \text{if } t \neq 0, \\ 2 & \text{if } t = 0. \end{cases}$$

By Lemma 9.30 and the fact that similar matrices have kernels of the same dimension, we conclude that $M_{\lambda,0}$ and $M_{\lambda,1}$ are not similar. On the other hand, $M_{\lambda,t}$ is similar to $M_{\lambda,1}$ if $t \neq 0$, since

$$\begin{pmatrix} \lambda & t \\ 0 & \lambda \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & t^{-1} \end{pmatrix} \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & t \end{pmatrix}.$$

This example gives us a first glimpse of the classification theorem, the 'Jordan Normal Form Theorem'.

9.6.2. Equivalent matrices.

Proposition 9.32. *If $f : V \to W$ is a linear map between finite-dimensional F-vector spaces and $M \in \mathrm{Mat}(m \times n, F)$ is the matrix associated to f relative to some choice of bases of V and W, then the set of all matrices associated to f relative to any choice of bases is*

$$\{QMP \;:\; P \in \mathrm{Mat}(n, F),\ Q \in \mathrm{Mat}(m, F),\ P \text{ and } Q \text{ invertible}\}.$$

Proof. By Proposition 9.18, every matrix associated to f is in the given set. Conversely, let B and C be the original bases for V and W, so that $M = [f]^B_C$. Given invertible matrices P and Q, we can find bases B' and C' for V and W, respectively, such that $P = [\mathrm{id}]^{B'}_B$ and $Q^{-1} = [\mathrm{id}]^{C'}_C$ by Lemma 9.27. Then (by Proposition 9.18 again) we have $QMP = [f]^{B'}_{C'}$. □

Definition 9.33. We say that two matrices $M, M' \in \mathrm{Mat}(m \times n, F)$ are *equivalent* if there are invertible matrices $P \in \mathrm{Mat}(n, F)$ and $Q \in \mathrm{Mat}(m, F)$ such that $M' = QMP$.

This notion does indeed define an equivalence relation on $\mathrm{Mat}(m \times n, F)$ (see Exercise 9.6.2). It is weaker than the notion of similarity in the sense that any two similar square matrices are equivalent, while two equivalent matrices need not necessarily be similar.

Proposition 9.18 shows that any two matrices associated to the same linear map $f\colon V \to W$, but with respect to different bases, are equivalent. Proposition 9.32 shows that the converse holds as well.

If we choose bases that are well-adapted to the linear map, then we will obtain a very nice matrix. This is used in the following result.

Corollary 9.34. *Let $M \in \mathrm{Mat}(m \times n, F)$. Then there are invertible matrices $P \in \mathrm{Mat}(n, F)$ and $Q \in \mathrm{Mat}(m, F)$ such that*

$$QMP = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix} = \left(\begin{array}{c|c} I_r & 0_{r\times(n-r)} \\ \hline 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{array}\right),$$

where $r = \mathrm{rk}(M)$.

Proof. Let $V = F^n$, $W = F^m$, and let $f = f_M : V \to W$ be the linear map given by M. By the Basis Extension Theorem 7.33 we can choose a basis $B = (v_1, \dots, v_n)$ for V such that $v_{r+1}, \dots, v_n$ is a basis of $\ker(f)$. We have $\dim\ker(f) = n - r$, so $\mathrm{rk}\, f = r$ by Theorem 8.3. Since $f(v_i) = 0$ for $r < i \le n$, the r elements $w_1 = f(v_1), \dots, w_r = f(v_r)$ generate the image $\mathrm{im}\, f$, which has dimension r, so these r elements are linearly independent by Theorem 7.47.

Hence, we can extend them to a basis $C = (w_1, \dots, w_m)$ for W. We then have

$$f(v_i) = \begin{cases} w_i & \text{if } 1 \le i \le r \\ 0 & \text{if } r+1 \le i \le n. \end{cases}$$

So the matrix $M' = [f]^B_C$ associated to f with respect to B and C has the required form. Set $P = [\mathrm{id}]^B_{E_n}$ and $Q = [\mathrm{id}]^{E_m}_C$, where E_n and E_m are the standard bases of F^n and F^m, respectively. Then by Proposition 9.18, we have

$$M' = [f]^B_C = [\mathrm{id}]^{E_m}_C \cdot [f]^{E_n}_{E_m} \cdot [\mathrm{id}]^B_{E_n} = QMP,$$

as M is the matrix associated to $f = f_M$ relative to the standard bases E_n and E_m. □

Corollary 9.34 implies the following, which shows that it is easy to classify matrices up to equivalence.

Corollary 9.35. *Two $m \times n$ matrices M and M' are equivalent if and only if they have the same rank.*

Proof. First note that if M and M' are equivalent, they must have the same rank, since the rank does not change under multiplication by invertible matrices (see Exercise 9.6.3). For the converse, suppose M, M' are $m \times n$ matrices of rank r. Then Corollary 9.34 tells us that M and M' are both equivalent to the matrix given there, and hence equivalent to each other. □

Remark 9.36. Recall that by Proposition 6.4, row operations on a matrix M correspond to multiplication on the left by an invertible matrix, and column operations on M correspond to multiplication on the right by an invertible matrix. Conversely, Corollary 8.36 shows that any invertible matrix is the product of elementary matrices, each corresponding with an elementary row operation (if multiplied by from the left) or column operation (if multiplied by from the right). This has two interesting implications.

(1) Corollary 9.34 implies that any matrix M can be transformed into the given simple form by elementary row and column operations. The advantage of this approach is that by keeping track of the operations, we can also determine the matrices P and Q explicitly, much in the same way as when inverting a matrix, cf. the first alternative proof of Theorem 8.12.
(2) Interpreting M as the matrix $[f]^B_C$ associated to a linear map $f\colon V \to W$ relative to some bases B and C for V and W, respectively, we see from Proposition 9.18 that row operations on M correspond to changing the basis of the target space W, whereas column operations correspond to changing the basis of the domain space V.

Exercises

9.6.1. This exercise generalises Example 9.28 and makes Lemma 9.27 concrete for $V = F^n$. Let $B = (w_1, \dots, w_n)$ be a basis for F^n and let Q be the matrix whose columns are $w_1, \dots, w_n$. Let P be any invertible $n \times n$ matrix and let $v_1, v_2, \dots, v_n$ be the columns of the matrix QP. Show that $B' = (v_1, \dots, v_n)$ is a basis for F^n and we have $P = [\mathrm{id}_{F^n}]^{B'}_B$.

9.6.2. Show that the notion of equivalent matrices defines an equivalence relation on the space $\mathrm{Mat}(m \times n, F)$ of $m \times n$ matrices, as claimed.

9.6.3. Show that any two equivalent matrices have the same rank.

CHAPTER 10

Determinants

We will define the *determinant* $\det f$ of any endomorphism $f\colon V \to V$ of a finite-dimensional vector space V over a field F. The most important properties of the determinant include the fact that f is an isomorphism if and only if $\det f \neq 0$, and the fact that it is multiplicative, that is, $\det(f \circ g) = (\det f) \cdot (\det g)$.

10.1. Determinants of matrices

We start with the case $V = F^n$, so that $f\colon V \to V$ is given by some matrix. In the case $F = \mathbb{R}$, the determinant of $f\colon \mathbb{R}^n \to \mathbb{R}^n$ will turn out to correspond with the factor by which f scales 'oriented volumes' (see Remark 10.14). So we have to think a little bit about functions that define 'oriented volume'.

We will only consider *parallelotopes;* these are the bodies spanned by n vectors $v_1, \ldots, v_n \in \mathbb{R}^n$:

$$P(v_1, \ldots, v_n) = \{\lambda_1 v_1 + \cdots + \lambda_n v_n : \lambda_1, \ldots, \lambda_n \in [0, 1]\}.$$

The parallelotope $P(v_1, \ldots, v_n)$ is the image of the 'unit cube' $P(e_1, \ldots, e_n)$ under the linear map that sends the standard basis vectors $e_1, \ldots, e_n$ to $v_1, \ldots, v_n$; this map is $\varphi_C\colon \mathbb{R}^n \to \mathbb{R}^n$ for the sequence $C = (v_1, \ldots, v_n)$, and it is given by the matrix that has $v_1, \ldots, v_n$ as columns.

Now let us assume $D\colon \operatorname{Mat}(n, \mathbb{R}) \to \mathbb{R}$ is a function that measures oriented volumes in the sense that for any $n \times n$ matrix A, the absolute value $|D(A)|$ can be interpreted as a volume of the image of the 'unit cube' $P(e_1, \ldots, e_n)$ under f_A, that is, a volume of the parallelotope $P(v_1, \ldots, v_n)$, where $v_1, \ldots, v_n$ are the columns of A.

Note that if the $n \times n$ matrices A and A' have the same columns, but in a different order, then their associated parallelotopes are the same, so we have $|D(A)| = |D(A')|$. Even though the parallelotope $P(v_1, \ldots, v_n)$ does not determine the order of the vectors $v_1, \ldots, v_n$, by abuse of language we say that the oriented volume of the parallelotope $P(v_1, \ldots, v_n)$ is $D(A)$, where A is the matrix with columns $v_1, \ldots, v_n$ in the same order as in the notation $P(v_1, \ldots, v_n)$.

Example 10.1. For $n = 1$, the absolute value is a 1-dimensional volume, also known as length: a number $v \in \mathbb{R}$ has a length $|v|$. As oriented volume we can take the identity: the oriented volume of v is v itself, which is (obviously) negative when v is negative, while its absolute value is the usual length.

With a generalisation to arbitrary fields in mind, what properties should such a function

$$D\colon \operatorname{Mat}(n, F) \to F$$

satisfy?

For notational convenience, for any $m \times n$ matrix A over F, any integer $1 \leq j \leq n$, and any vector $x \in F^m$, we denote by $r_j(A, x)$ the matrix obtained by replacing the

j-th column of A by x; similarly, for integers $1 \le j, k \le n$ and vectors $x, y \in F^m$, we denote by $r_{jk}(A, x, y)$ the matrix obtained by replacing the j-th and k-th column of A by x and y, respectively.

The oriented volume should scale corresponding to scaling of the vectors, that is,

$$D(r_j(A, \lambda x)) = \lambda D(r_j(A, x)) \text{ for all } 1 \le j \le n \text{ and all } x \in F^m . \tag{10.1}$$

Also, volumes should be additive in the following sense:

$$D(r_j(A, x + y)) = D(r_j(A, x)) + D(r_j(A, y)) \text{ for all } 1 \le j \le n \text{ and all } x \in F^m . \tag{10.2}$$

We will now give a motivation why we want D to satisfy these properties over $\mathbb{R}$. If the $n-1$ columns $v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_n$ of A other than the j-th column, span an $(n-1)$-dimensional parallelotope $B = P(v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_n)$ inside a hyperplane H with normal a, and this so-called base B has $(n-1)$-dimensional volume b, then the volume $|D(A)|$ of $P(v_1, \ldots, v_n)$ equals b times the absolute value of the oriented height of $P(v_1, \ldots, v_n)$ with respect to this base; this oriented height is the oriented length of the projection of the j-th column onto a, which is indeed additive in the j-th column. This is depicted in Figure 10.1 for $\mathbb{R}^3$. In the first picture we see the base B and in the second and third pictures we see two parallelotopes with base B and third vector x and y, respectively. The fourth picture has these two parallelotopes stacked on top of each other and the final picture shows a parallelotope with base B and third vector $x + y$. One way to think about these parallelotopes in $\mathbb{R}^3$ is as stacks of sheets of paper, each sheet having the shape as the base B. We start with two skew stacks, put them on top of each other, and straighten them to one (still skew) stack, keeping the top and bottom sheet in place. Of course, the total volume of paper does not change in the process. Clearly, any decent person would immediately make all the stacks straight and vertical; a stack of papers with third vector z then becomes a stack with third vector $\pi_a(z)$. The three straight (but skew) stacks in Figure 10.1 then become stacks with third vector equal to $\pi_a(x), \pi_a(y)$, and $\pi_a(x+y)$, respectively. Since π_a is linear, we see again that the volumes of the first two stacks add up to the volume of the big stack.

Over any field, the two properties (10.1) and (10.2) can be stated simply by saying that D is linear in each column separately, when the other $n-1$ columns are held constant. That is, for each $n \times n$ matrix A and each $1 \le j \le n$, the function $F^n \to F,\ x \mapsto D(r_j(A, x))$ is linear. Such a function $D\colon \operatorname{Mat}(n, F) \to F$ is said to be *multilinear* as a function in the columns.

Still inspired by the case $F = \mathbb{R}$, another property of D should certainly be that the n-dimensional volume $D(A)$ vanishes when the parallelotope spanned by the columns of A is of lower dimension, that is, when the columns are linearly dependent. Together with multilinearity, it suffices to only require the special case when two of the columns are equal (see Lemma 10.3(1)), that is,

$$D(r_{ij}(A, x, x)) = 0 \text{ for all } 1 \le i, j \le n \text{ with } i \ne j \text{ and all } x \in F^n. \tag{10.3}$$

A function $D\colon \operatorname{Mat}(n, F) \to F$ that is multilinear in the columns and that satisfies this third property (10.3) is said to be *alternating.* So these are the functions we are looking for. Note that it makes sense over any field F to talk about functions $\operatorname{Mat}(n, F) \to F$ that are multilinear and alternating in the columns.

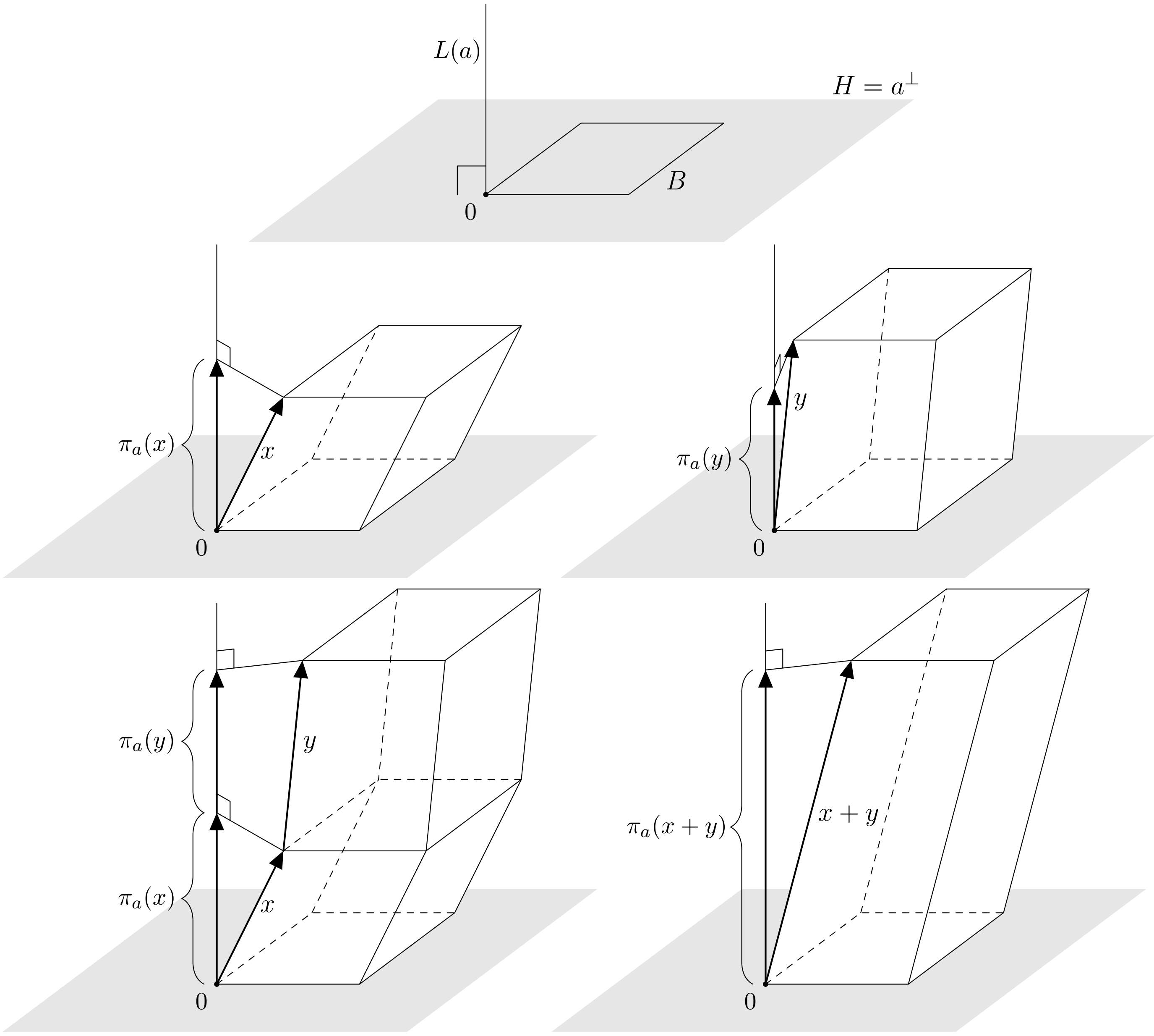


FIGURE 10.1. For a fixed base, the volume is additive in the final vector

Definition 10.2. Let F be a field and let n be a positive integer. A function $\mathrm{Mat}(n, F) \to F$ is called a *determinantal function* if it is multilinear and alternating as function in the columns.

How many determinantal functions are there? First, it is pretty clear that the set of all determinantal functions on V forms an F-vector space. So the question we should ask is, what is the dimension of this vector space?

Before we state the relevant theorem, let us first prove a few simple properties of determinantal functions.

Lemma 10.3. *Let F be a field, n a positive integer, and $A \in \mathrm{Mat}(n, F)$. Let $D : \mathrm{Mat}(n, F) \to F$ be a determinantal function.*

(1) *If A is not invertible, then $D(A) = 0$.*
(2) *If we add a scalar multiple of the i-th column of a matrix A to the j-th column, where $i \neq j$, then $D(A)$ is unchanged, that is,*
$$D(r_{ij}(A, x, y)) = D(r_{ij}(A, x, y + \lambda x)).$$
(3) *If we interchange two of the columns, then $D(A)$ changes sign, that is, for $i \neq j$ we have*
$$D(r_{ij}(A, x, y)) = -D(r_{ij}(A, y, x)).$$

Proof. For (1), assume that $A \in \mathrm{Mat}(n, F)$ is not invertible. Then its columns $v_1, v_2, \ldots, v_n$ are linearly dependent, so one of them, say v_j, is a linear combination of the others, say
$$v_j = \sum_{i \neq j} \lambda_i v_i.$$
By linearity of D in the j-th column, this implies
$$D(A) = D(r_j(A, v_j)) = D\Big(r_j\Big(A, \sum_{i \neq j} \lambda_i v_i\Big)\Big) = \sum_{i \neq j} \lambda_i D(r_j(A, v_i)) = \sum_{i \neq j} \lambda_i \cdot 0 = 0,$$
where the second-to-last equality follows from the fact that for $i \neq j$, the matrix $r_j(A, v_i)$ has two identical columns, namely the i-th and the j-th.

We now prove (2). By linearity of D in the j-th column and the fact that D is alternating, we have
$$D(r_{ij}(A, x, y + \lambda x)) = D(r_{ij}(A, x, y)) + \lambda D(r_{ij}(A, x, x)) = D(r_{ij}(A, x, y)).$$
Finally, for (3), suppose we have $x, y \in F^n$. Then we obtain
$$\begin{aligned} 0 = D(r_{ij}(A, x + y, x + y)) =& D(r_{ij}(A, x, x)) + D(r_{ij}(A, x, y)) + D(r_{ij}(A, y, x)) \\ &+ D(r_{ij}(A, y, y)) = D(r_{ij}(A, x, y)) + D(r_{ij}(A, y, x)), \end{aligned}$$
so $D(r_{ij}(A, x, y)) = -D(r_{ij}(A, y, x))$. □

Proposition 10.4. *For any field F, non-negative integer n, and element $\lambda \in F$, there is at most one determinantal function $D\colon \mathrm{Mat}(n, F) \to F$ with $D(I_n) = \lambda$.*

Proof. Suppose $D\colon \mathrm{Mat}(n, F) \to F$ is a determinantal function with $D(I_n) = \lambda$. Lemma 10.3(1) gives $D(A) = 0$ if A is not invertible. Otherwise, the matrix A is invertible, and we can transform it into the identity matrix I_n by elementary column operations. The multilinearity of D and Lemma 10.3 tell us how the value of D changes in the process: we see that
$$D(A) = (-1)^k \delta^{-1} D(I_n) = (-1)^k \delta^{-1} \lambda\,,$$
where k is the number of times we have swapped two columns and δ is the product of all the scaling factors we have used when scaling a column. This shows that D is uniquely determined, as $D(A)$ is determined for any matrix A. □

We cannot use the observation made in the proof of Proposition 10.4 easily to show the existence of a determinantal function on F^n, as we would have to show that $(-1)^k \delta^{-1}$ does not depend on the sequence of elementary column operations we have performed in order to obtain I_n. Instead, we define an explicit function and show that it is determinantal.

Definition 10.5. We define the functions
$$d_n\colon \operatorname{Mat}(n,F) \to F$$
(for $n \geq 0$) inductively. We set $d_0(I_0) = 1$ for the unique 0×0 matrix I_0. For $n > 0$ we choose an index $1 \leq i \leq n$ and set
$$d_n(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \cdot d_{n-1}(A_{ij}), \tag{10.4}$$
where a_{ij} is the entry in the i-th row and j-th column of A and A_{ij} is the submatrix of A obtained by deleting the i-th row and the j-th column from A.

Note that we have $d_1\big((\lambda)\big) = \lambda$, which could also have been used as the base case in the inductive definition of the functions d_n.

A priori, the function d_n might depend on the choice of the index i, and the analogous choices made to define $d_1, \ldots, d_{n-1}$). The following proposition says that d_n does not depend on these choices.

Proposition 10.6. *For any integer $n \geq 0$, the function $d_n\colon \operatorname{Mat}(n,F) \to F$ is a determinantal function with $d_n(I_n) = 1$ that is independent of the choice of i in Definition 10.5.*

Proof. We use induction on n. For $n = 0$ the statement is trivial. (If you suffer from *horror vacui,* that is, you are afraid of the empty set, you can consider $n = 1$; then $d_1\colon \operatorname{Mat}(1,F) \to F$ sends the 1×1 matrix (λ) to λ.) For the induction step, we assume $n \geq 1$ and let i be the corresponding choice from Definition 10.5.

We first show that d_n is linear in each of its columns. Indeed, note that the function $F^n \to F^{n-1}$ that deletes the i-th coordinate is linear. By the induction hypothesis, this implies that for $1 \leq j, k \leq n$, the function $\operatorname{Mat}(n,F) \to F$ that sends A to $d_{n-1}(A_{ij})$ is linear as a function in the k-th column of A for $k \neq j$ and constant for $k = j$; the function $A \mapsto a_{ij}$ is the opposite, constant as a function in the k-th column of A for $k \neq j$ and linear for $k = j$. So the j-th term in the right-hand side of (10.4) is linear in all columns. Therefore, so is the sum d_n.

To see that d_n is alternating, we will show that for any $n \times n$ matrix A of which the k-th and l-th column are the same for some $k < l$, we have $d_n(A) = 0$. Let A be such a matrix. Then for $1 \leq j \leq n$ with $j \neq k, l$, the submatrix A_{ij} also has two identical columns, so $d_{n-1}(A_{ij}) = 0$ by the induction hypothesis. We conclude
$$d_n(A) = (-1)^{i+k} c \cdot d_{n-1}(A_{ik}) + (-1)^{i+l} c \cdot d_{n-1}(A_{il})$$
with $c = a_{ik} = a_{il}$. The matrices A_{ik} and A_{il} have the same columns, but in a different order: the matrix A_{ik} can be obtained from A_{il} by shifting the k-th column $l-k-1$ positions to the right, or, equivalently, swapping this column with its right neighbor $l-k-1$ times. Since d_{n-1} is an alternating multilinear function in the columns, we find $d_{n-1}(A_{ik}) = (-1)^{l-k-1} d_{n-1}(A_{il})$ by Lemma 10.3(3). This means that the two terms for $j = k$ and $j = l$ cancel and we have $d_n(A) = 0$.

We conclude that d_n is indeed a determinantal function. It is easy to check that $d_n(I_n) = 1$. From Proposition 10.4, we conclude that these two properties already determine d_n uniquely, so it is independent of the choice of i, which finishes the proof. $\square$

Corollary 10.7. *The determinantal functions* $\mathrm{Mat}(n, F) \to F$ *form an* F*-vector space of dimension* 1*.*

Proof. From Proposition 10.4, it follows that the dimension is at most 1, while Proposition 10.6 implies it is at least 1. $\square$

Definition 10.8. For any field F and any non-negative integer n, we let

$$\det\colon \mathrm{Mat}(n, F) \to F$$

be the unique determinantal function with $\det(I_n) = 1$; for $A \in \mathrm{Mat}(n, F)$, we call $\det(A)$ the *determinant* of the matrix A.

Note that the field F and the dimension n are not explicit in the notation det; by Proposition 10.6, we have $\det = d_n$. If $A = (a_{ij})$ is written as an $n \times n$ array of entries, we also write

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

and by (10.4) we have

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}) \tag{10.5}$$

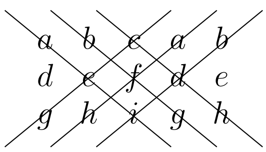
for all $1 \leq i \leq n$; this is called the *expansion of the determinant along the i-th row.*

Example 10.9. For 2×2 matrices and 3×3 matrices, we find

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

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + cdh - afh - bdi - ceg.$$

A mnemonic to remember the formula for 3×3 matrices is to repeat the first two columns after the third, and to add the products of the entries along the diagonals in one direction and subtract the products of the entries along the diagonals in the other direction.



Note that for $n > 3$ this does **not** directly generalise to $n \times n$ matrices! We will see in Exercise 10.1.5 that for $n > 1$ there is a way to write the determinant of an $n \times n$ matrix as a sum of $n!$ terms.

Example 10.10. If one of the rows of a square matrix contains many zeros, then it is useful to expand the determinant along that row. If we expand the following determinant along the second row, then we get

$$\begin{vmatrix} 1 & -1 & 2 & 1 \\ 1 & 0 & 2 & 0 \\ 2 & 1 & 2 & 1 \\ 3 & -1 & 1 & 0 \end{vmatrix} = -1 \cdot \begin{vmatrix} -1 & 2 & 1 \\ 1 & 2 & 1 \\ -1 & 1 & 0 \end{vmatrix} - 2 \begin{vmatrix} 1 & -1 & 1 \\ 2 & 1 & 1 \\ 3 & -1 & 0 \end{vmatrix} = -1 \cdot 2 - 2 \cdot (-7) = 12.$$

Example 10.11. Using induction, it is easy to show that the determinant of a diagonal matrix

$$\begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

equals the product $\prod_{i=1}^n \lambda_i$ of the diagonal elements. The same holds for upper triangular matrices, which are matrices of which all entries below the diagonal are zero. See Exercise 10.1.2.

The proof of Proposition 10.4 gives us a second procedure to compute determinants: we perform elementary column operations on A, keeping track of the scalings and swappings, until we get a zero column (then $\det(A) = 0$), or we reach the identity matrix.

Example 10.12. We compute a determinant by elementary column operations. Note that we can avoid divisions (and hence fractions) by choosing the operations cleverly, cf. Example 6.15.

$$\begin{vmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \\ 3 & 4 & 2 & 1 \\ 4 & 3 & 1 & 2 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 2 & -3 & -2 & -5 \\ 3 & -2 & -7 & -11 \\ 4 & -5 & -11 & -14 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & -2 & -5 \\ 3 & 12 & -7 & -11 \\ 4 & 17 & -11 & -14 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -21 & 12 & 17 & 49 \\ -30 & 17 & 23 & 71 \end{vmatrix}$$

$$= \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -21 & 12 & 17 & -2 \\ -30 & 17 & 23 & 2 \end{vmatrix} = 2 \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -21 & 12 & 1 & 17 \\ -30 & 17 & -1 & 23 \end{vmatrix} = 2 \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -51 & 29 & -1 & 40 \end{vmatrix}$$

$$= 2 \cdot 40 \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} = 80$$

Exercises

10.1.1. Determine the determinants of the following matrices, both by expansion along a row and by using elementary column operations.

$$\begin{pmatrix} -1 & -2 \\ -3 & -2 \end{pmatrix} \qquad \begin{pmatrix} -2 & -3 & 2 \\ 0 & 1 & 2 \\ -3 & -3 & 0 \end{pmatrix} \qquad \begin{pmatrix} 2 & -2 & -2 \\ 1 & 3 & -1 \\ 2 & -2 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & -2 & -2 & -1 \\ 1 & -1 & -1 & 2 \\ -2 & -2 & 0 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix} \qquad \begin{pmatrix} -3 & 2 & 1 & 2 \\ -1 & -1 & -3 & 1 \\ 3 & -2 & -3 & -2 \\ 3 & -2 & -1 & -1 \end{pmatrix}$$

10.1.2. An *upper triangular* matrix is a square matrix of which all entries below the main diagonal are 0. Show that the determinant of an upper triangular matrix is equal to the product of its diagonal entries. The same is true for *lower triangular* matrices.

10.1.3. For each $x, y \in \mathbb{R}$, determine the determinant of the matrix

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

In Exercise 10.2.6 this will be generalised to Vandermonde matrices of arbitrary size.

10.1.4. Let M_n denote the $n \times n$ matrix over $\mathbb{R}$ of which the entry in the i-th row and the j-th column equals 1 if $|i-j| \leq 1$ and 0 otherwise. For example,

$$M_6 = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}.$$

(1) Compute the determinant of M_n for $2 \leq n \leq 5$.
(2) Give (with proof) a general formula in terms of n for the determinant of M_n.

10.1.5. Let $n \geq 1$ be an integer, and let S_n be the set of all permutations of the set $\{1, 2, \ldots, n\}$. (A permutation of a set is a bijection from that set to itself.) For any such permutation $\sigma \colon \{1, \ldots, n\} \to \{1, \ldots, n\}$, we define the sign $\varepsilon(\sigma)$ of σ by $\varepsilon(\sigma) = (-1)^{m(\sigma)}$ with

$$m(\sigma) = \#\{\ (i,j)\ :\ 1 \leq i < j \leq n \text{ and } \sigma(i) > \sigma(j)\ \}.$$

(1) Suppose $\sigma \in S_n$ is a permutation, and $k, l \in \{1, 2, \ldots, n\}$ two different elements. Let σ' be the permutation obtained by composing σ with the permutation that just switches k and l, so

$$\sigma(i) = \begin{cases} \sigma(i) & \text{if } \sigma(i) \neq k, l, \\ k & \text{if } \sigma(i) = l, \\ l & \text{if } \sigma(i) = k. \end{cases}$$

Show that $\varepsilon(\sigma') = \varepsilon(\sigma)$.

(2) Show that for any $n \times n$ matrix $A = \big(a_{i,j}\big)_{i,j}$ we have

$$\det A = \sum_{\sigma \in S_n} \varepsilon(\sigma) \cdot a_{1,\sigma(1)} a_{2,\sigma(2)} \cdots a_{n,\sigma(n)}.$$

10.2. Some properties of the determinant

Proposition 10.13. *For any $n \times n$ matrices A and B, we have*

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

Proof. Let A be an $n \times n$ matrix. Consider the functions $D_1, D_2 \colon \operatorname{Mat}(n, F) \to F$, given by

$$D_1(M) = (\det A) \cdot (\det M),$$
$$D_2(M) = \det(AM).$$

Then D_1 is a multiple of det, so D_1 is a determinantal function and it satisfies $D_1(I_n) = \det A$. Note that in Section 5.5, just under (5.5), we have seen that the j-th column of AM equals (A times the j-th column of M). This implies $A \cdot r_j(M, x) = r_j(AM, Ax)$, from which it is easily seen that the function D_2 is linear in each column of M. It is also alternating, because if M has two identical columns, then so does AM and so $\det(AM) = 0$. We conclude that D_2 is a determinantal function satisfying $D_2(I_n) = \det A$ as well. By Proposition 10.4 we conclude $D_1 = D_2$ and in particular $D_1(B) = D_2(B)$, that is, $\det(AB) = (\det A) \cdot (\det B)$. □

Remark 10.14. We look back at our earlier motivation for the determinant: oriented volumes. For two real $n \times n$ matrices A and B, we can interpret $\det B$ as the oriented volume of the parallelotope P spanned by the columns of B, and $\det(AB)$ as the oriented volume of the image $f_A(P)$ of P under the map f_A, namely the parallelotope spanned by the columns of AB. Then Proposition 10.13 states that the oriented volume of $f_A(P)$ is $(\det A)$ times the oriented volume of P. Hence, instead of viewing $\det A$ as the volume of the one parallelotope spanned by the columns of A, that is, the image of the unit cube, we can view $\det A$ as the factor by which the endomorphism f_A scales the volumes of all polytopes.

Corollary 10.15. *If A is an invertible matrix, then $\det A \neq 0$ and $\det(A^{-1}) = (\det A)^{-1}$.*

Proof. Let n be the number of rows (and thus also the number of columns) of A. By Proposition 10.13, we have

$$(\det(A^{-1})) \cdot (\det A) = \det(A^{-1}A) = \det(I_n) = 1,$$

from which the statement follows. □

Theorem 10.16. *A square matrix A is invertible if and only if $\det A \neq 0$.*

Proof. If A is not invertible, then $\det A = 0$ by Lemma 10.3, and if A is invertible, then $\det A \neq 0$ by Corollary 10.15. □

Theorem 10.17. *Let $A \in \operatorname{Mat}(n, F)$. Then $\det(A^\top) = \det(A)$.*

Proof. We show that $A \mapsto \det(A^\top)$ is a determinantal function. First, if A has two columns that are the same, then we have $\operatorname{rk} A < n$, so we also have $\operatorname{rk}(A^\top) < n$ by Theorem 8.12(3) (see Remark 8.13) and therefore $\det(A^\top) = 0$ by Lemma 10.3(1).

This implies that our function is alternating. Second, we have to show that $\det(A^\top)$ is linear in each of the columns of A. This is obviously equivalent to saying that $\det(A)$ is linear in each of the *rows* of A. To check that this is the case for the i-th row, we expand $\det(A)$ along the i-th row according to (10.5). For $A = (a_{ij})$, we have

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}) .$$

Now in A_{ij} the i-th row of A has been removed, so $\det(A_{ij})$ does not depend on the i-th row of A; linearity is then clear from the formula. Finally, we have $\det(I_n^\top) = \det(I_n) = 1$, so $\det(A^\top)$ must coincide with $\det(A)$ because of the uniqueness of determinantal functions (see Proposition 10.4). □

Corollary 10.18 (Expansion along Columns). *We can also expand determinants along columns. Let $n \geq 1$ and $A = (a_{ij}) \in \mathrm{Mat}(n, F)$; we use the notation A_{ij} as before. Then for $1 \leq j \leq n$,*

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}) .$$

Proof. We expand the determinant of $A^\top$ along the j-th row as in (10.5), with the roles of i and j switched. The elements in the j-th row of $A^\top$ are $a_{1j}, a_{2j}, \ldots, a_{ij}$, so we get

$$\begin{aligned} \det(A) = \det(A^\top) &= \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det\big((A^\top)_{ji}\big) \\ &= \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det\big((A_{ij})^\top\big) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}) . \end{aligned}$$

□

Remark 10.19. Just as Lemma 10.3 tells us how the determinant of a matrix behaves under elementary column operations, we conclude from Theorem 10.17 that it behaves similarly under elementary row operations.

Example 10.20. A matrix $A \in \mathrm{Mat}(n, F)$ is said to be *orthogonal* if $AA^\top = I_n$. What can we deduce about $\det(A)$? Well,

$$1 = \det(I_n) = \det(AA^\top) = \det(A)\det(A^\top) = \det(A)^2 ,$$

so $\det(A) = \pm 1$.

Exercises

10.2.1. Determine (again) the determinants of the matrices of Exercise 10.1.1, this time using elementary row operations or expansion along a column.

10.2.2. Let A, B be two $n \times n$ matrices. True or not true?
(1) $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$.
(2) $\mathrm{Tr}(AB) = (\mathrm{Tr}\, A)(\mathrm{Tr}\, B)$.
(3) $\mathrm{Tr}(A + B) = \mathrm{Tr}\, A + \mathrm{Tr}\, B$.
(4) $\det(AB) = \det BA$.
(5) $\det(AB) = (\det A)(\det B)$.
(6) $\det(A + B) = \det A + \det B$.

(7) $\det A \neq 0$ if and only if A is invertible.

10.2.3. Let M be a block matrix

$$M = \left(\begin{array}{c|c} A & B \\ \hline 0 & C \end{array}\right)$$

over a field F with A and C square matrices, say $A \in \mathrm{Mat}(m, F)$ and $C \in \mathrm{Mat}(n, F)$, and $B \in \mathrm{Mat}(m \times n, F)$ and where 0 denotes the zero matrix in $\mathrm{Mat}(n \times m, F)$. Show that $\det M = (\det A) \cdot (\det C)$.

10.2.4. Show that for any block matrix

$$A = \left(\begin{array}{c|c|c|c} A_{11} & A_{12} & \cdots & A_{1t} \\ \hline 0 & A_{22} & \cdots & A_{2t} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \cdots & A_{tt} \end{array}\right)$$

with square blocks and zeros below the diagonal blocks, we have

$$\det A = (\det A_{11})(\det A_{22}) \cdots (\det A_{tt}).$$

10.2.5. Let M_n denote the $n \times n$ matrix over $\mathbb{R}$ with zeros on the diagonal and ones for every entry off the diagonal.

(1) Compute the determinant of M_n for $2 \leq n \leq 5$.
(2) Guess a general formula in terms of n for the determinant of M_n.
(3) Can you prove your guess?

10.2.6. Let F be a field containing k distinct elements $\alpha_1, \alpha_2, \ldots, \alpha_k \in F$. By Exercise 9.1.3, the square Vandermonde matrix

$$\begin{pmatrix} 1 & \alpha_1 & \alpha_1^2 & \cdots & \alpha_1^{k-1} \\ 1 & \alpha_2 & \alpha_2^2 & \cdots & \alpha_2^{k-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_k & \alpha_k^2 & \cdots & \alpha_k^{k-1} \end{pmatrix}.$$

is invertible, so it has nonzero determinant. Prove that the determinant equals

$$\prod_{i<j} (\alpha_j - \alpha_i).$$

10.3. Cramer's rule

Definition 10.21. Let $A \in \mathrm{Mat}(n, F)$ with $n \geq 1$. Then the *adjugate* matrix of A (sometimes called the *adjoint* matrix, but this also has other meanings) is the matrix $\tilde{A} \in \mathrm{Mat}(n, F)$ whose (i, j)-entry $\tilde{a}_{ij}$ is $(-1)^{i+j} \det(A_{ji})$. Here A_{ij} is, as before, the matrix obtained from A by removing the i-th row and j-th column. Note the reversal of indices: $\tilde{a}_{ij} = \pm \det(A_{ji})$ and not $\pm \det(A_{ij})$!

Proposition 10.22 (Cramer's rule). *Let $A \in \mathrm{Mat}(n, F)$ with $n \geq 1$. Then*

$$A\tilde{A} = \tilde{A}A = \det(A) I_n \,.$$

If A is invertible, then $\det(A) \neq 0$, and

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

Proof. We denote the (i, j)-th entry of A by a_{ij}. The (i, k)-th entry of $A\tilde{A}$ is

$$\sum_{j=1}^{n} a_{ij} (-1)^{j+k} \det(A_{kj}) \,.$$

Let $A' = (a'_{ij})$ be the matrix that we obtain from A by replacing the k-th row by the i-th row. Expanding the determinant of A' by the k-th row, we find

$$\det(A') = \sum_{j=1}^{n} (-1)^{k+j} a'_{kj} \det(A'_{kj}) = \sum_{j=1}^{n} (-1)^{j+k} a_{ij} \det(A_{kj}),$$

which equals the (i,k)-th entry of $A\tilde{A}$ mentioned above. The claimed identity $A\tilde{A} = \det(A) I_n$ now follows from the fact that for $i = k$ we have $A' = A$, so $\det A' = \det A$, while for $i \neq k$, we have $\det A' = 0$, as the i-th and k-th row of A' are equal.

The assertion on $\tilde{A}A$ is proved in the same way (or by applying what we have just proved to $A^\top$). The final claim of the proposition follows immediately. $\square$

Example 10.23. The inverse of a 2×2 matrix

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

with determinant $ad - bc \neq 0$ is

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

Exercises

10.3.1. Let $A \in \mathrm{Mat}(n, f)$ be invertible, with $n \geq 1$. Let $b \in F^n$ be a vector. Then the equation $Ax = b$ has a unique solution $s \in F^n$. Show that if we write $s = (s_1, s_2, \ldots, s_n)$ for this solution, then for all indices $1 \leq i \leq n$ we have

$$s_i = \frac{\det A_i(b)}{\det A},$$

where $A_i(b)$ denotes the $n \times n$ matrix obtained from A by replacing the i-th column by b.

10.4. Determinants of endomorphisms

Proposition 10.24. *Two similar $n \times n$ matrices have the same determinant.*

Proof. Let A and A' be similar $n \times n$ matrices. Then there is an invertible $n \times n$ matrix P such that $A' = PAP^{-1}$. Then

$$\det A' = \det(PAP^{-1}) = (\det P)(\det A)(\det P^{-1}) = (\det P)(\det A)(\det P)^{-1} = \det A$$

by Proposition 10.13. $\square$

Corollary 10.25. *Let $f\colon V \to V$ be an endomorphism of a finite-dimensional vector space V with two bases B and B'. Then we have $\det[f]^{B'}_{B'} = \det[f]^{B}_{B}$.*

Proof. For $P = [\mathrm{id}]^{B}_{B'}$ we have $P^{-1} = [\mathrm{id}]^{B'}_{B}$ and $[f]^{B'}_{B'} = P \cdot [f]^{B}_{B} \cdot P^{-1}$ by Proposition 9.19, so $[f]^{B}_{B}$ and $[f]^{B'}_{B'}$ are similar matrices. They have the same determinant by Proposition 10.24. $\square$

Corollary 10.25 shows that the following definition makes sense.

Definition 10.26. Let $f: V \to V$ be an endomorphism of a finite-dimensional vector space V with basis B. Then we define the *determinant* of f, written $\det f$, to be the determinant $\det[f]_B^B$ of the matrix associated to f with respect to B.

Example 10.27. If V is a finite-dimensional vector space, then for the identity $\mathrm{id}_V: V \to V$ we have $\det \mathrm{id}_V = 1$.

Example 10.28. By Example 9.2, we of course have $\det f_A = \det[f_A]_E^E = \det A$ for any square matrix A.

Example 10.29. Let $V \subset \mathbb{R}^3$ be a plane and $s: \mathbb{R}^3 \to \mathbb{R}^3$ the reflection in V, cf. Examples 9.11 and 9.20. To compute the determinant of s, we may choose any basis. Take a basis (v_1, v_2) for V and a normal v_3 of V. Then $B = (v_1, v_2, v_3)$ is a basis for $\mathbb{R}^3$ (why?), and as in Example 9.11, we find

$$[s]_B^B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

We conclude $\det s = \det([s]_B^B) = -1$. Note that this is consistent with the fact that the reflection s preserves volumes and changes the orientation of the volumes.

Proposition 10.30. *For any finite-dimensional vector space V and any two endomorphisms $f, g: V \to V$, we have* $\det(f \circ g) = (\det f) \cdot (\det g)$.

Proof. Choose a basis B for V. From Propositions 9.14 and 10.13 we find

$$\det(f \circ g) = \det([f \circ g]_B^B) = \det([f]_B^B \cdot [g]_B^B) = (\det[f]_B^B)(\det[g]_B^B) = (\det f)(\det g).$$

□

Proposition 10.30 implies that if $f: V \to V$ is an automorphism of a finite-dimensional vector space V, then $\det f \neq 0$ and $\det(f^{-1}) = (\det f)^{-1}$. Indeed, apply the proposition with $g = f^{-1}$. The following proposition shows that the converse holds as well.

Proposition 10.31. *Let $f: V \to V$ be an endomorphism of a finite-dimensional vector space V. Then f is an isomorphism if and only if* $\det f \neq 0$.

Proof. Choose a basis for B and set $n = \dim V$. By Proposition 4.41 and Definition 5.22, the map f is an isomorphism if and only if the matrix $[f]_B^B$ is invertible. By Theorem 10.16, this is the case if and only if the determinant $\det f = \det([f]_B^B)$ is nonzero. □

Exercises

10.4.1. Determine the determinant of the following linear maps.

(1) $f: \mathbb{R}^3 \to \mathbb{R}^3, (x, y, z) \mapsto (2x + z, y - 3z, -x + 2y + 3z)$,
(2) the rotation $\mathbb{R}^2 \to \mathbb{R}^2$ about 0 over an angle φ,
(3) the orthogonal projection $\mathbb{R}^3 \to \mathbb{R}^3$ of $\mathbb{R}^3$ onto the plane given by the equation $x - 2y + z = 0$,
(4) the map $\mathbb{R}[x]_3 \to \mathbb{R}[x]_3$ given by $f \mapsto xf'$ with f' the derivative of f,

10.4.2. Let $\varphi \colon V \to W$ be an isomorphism of finite-dimensional vector spaces, and $f \colon V \to V$ an endomorphism of V. Show that $f' = \varphi \circ f \circ \varphi^{-1}$ is an endomorphism of W satisfying $\det f' = \det f$.

10.4.3. Let $f \colon V \to V$ be an endomorphism of a finite-dimensional vectorspace V. Let $\sigma \colon V \to W$ be a linear map. Suppose that $f(\ker \sigma) \subset \ker \sigma$. Let f' be the restriction of f to $\ker \sigma$ and let f'' be the endomorphism of $\operatorname{im} \sigma$ induced by f (see Exercise 4.5.1). Show that $\det f = (\det f') \cdot (\det f'')$.

[Hint: use Exercise 10.2.3.]

10.4.4. For every positive integer n, let M_n denote the matrix over $\mathbb{R}$ of Exercise 10.2.5.

(1) Show that for every $2 \leq i \leq n$, the element $v_i = e_i - e_{i-1}$ satisfies $M_n v_i = -v_i$.
(2) Show that $v_1 = (1, 1, \ldots, 1)$ satisfies $M_n(v_1) = (n-1)v_1$.
(3) Show that $B = (v_1, v_2, \ldots, v_n)$ is a basis for $\mathbb{R}^n$.
(4) Set $f = f_{M_n} \colon \mathbb{R}^n \to \mathbb{R}^n$. Show that $[f]_B^B$ is a diagonal matrix with its $(1,1)$-entry equal to $n-1$ and the other diagonal entries equal to -1.
(5) Show that $\det M_n = (-1)^{n-1}(n-1)$.

10.5. Linear equations with parameters

The determinant is very useful in studying systems of linear equations with parameters.

Example 10.32. For any $c \in \mathbb{R}$ we set

$$A_c = \begin{pmatrix} 1 & -1 & c \\ 1 & 1 & -2 \\ -1 & c & 2 \end{pmatrix} \qquad \text{and} \qquad b = \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}.$$

For each $c \in \mathbb{R}$, we want to know whether the linear equation $A_c \cdot x = b$ has no solutions, exactly one solution, or more than one solution. We first compute the determinant by expanding it along the first column.

$$\det A_c = \begin{vmatrix} 1 & -2 \\ c & 2 \end{vmatrix} - \begin{vmatrix} -1 & c \\ c & 2 \end{vmatrix} - \begin{vmatrix} -1 & c \\ 1 & -2 \end{vmatrix} = (2+2c)-(-2-c^2)-(2-c) = (c+1)(c+2).$$

We see that for $c \neq -2, -1$, the determinant $\det A_c$ is nonzero, so the matrix A_c is invertible and there is exactly one x with $A_c \cdot x = b$. For $c = -1$, the extended matrix is

$$\left(\begin{array}{ccc|c} 1 & -1 & -1 & 2 \\ 1 & 1 & -2 & 1 \\ -1 & -1 & 2 & -1 \end{array}\right)$$

with reduced row echelon form

$$\left(\begin{array}{ccc|c} 1 & 0 & -\frac{3}{2} & \frac{3}{2} \\ 0 & 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 0 & 0 \end{array}\right).$$

It follows immediately that $a = (\frac{3}{2}, -\frac{1}{2}, 0)$ satisfies $A_{-1} \cdot a = b$. The kernel of A_{-1} is generated by $z = (3, 1, 2)$, so the complete solution set is $\{ \, a + rz \; : \; r \in \mathbb{R} \, \}$. Finally, for $c = -2$, the extended matrix is

$$\left(\begin{array}{ccc|c} 1 & -1 & -2 & 2 \\ 1 & 1 & -2 & 1 \\ -1 & -2 & 2 & -1 \end{array}\right)$$

with reduced row echelon form

$$\left(\begin{array}{ccc|c} 1 & 0 & -2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right).$$

Here, the last column does contain a pivot, so there is no solution.

Exercises

10.5.1. For any real numbers $a, b \in \mathbb{R}$, we define the matrix C_a and the vector v_b by

$$C_a = \begin{pmatrix} a & a & 2 \\ 1 & 0 & a \\ -2 & -3 & 1 \end{pmatrix} \qquad \text{and} \qquad v_b = \begin{pmatrix} 2 \\ 1 \\ b \end{pmatrix}.$$

(1) For each $a \in \mathbb{R}$, determine the rank of the matrix C_a.
(2) Is C_a invertible for $a = 2$? If no, explain why not; if yes, give the inverse.
(3) For which pairs (a, b) does the equation $C_a x = v_b$ have more than one solution $x \in \mathbb{R}^3$?
(4) Describe the complete solution set for the pair of part (3) with the smallest value of a.

CHAPTER 11

Eigenvalues and Eigenvectors

In Example 10.29 we saw that for a reflection $s\colon \mathbb{R}^3 \to \mathbb{R}^3$ in a plane $V \subset \mathbb{R}^3$, there is a special basis B such that the associated matrix $[s]_B^B$ with respect to B is a diagonal matrix. It allowed us to compute the determinant very easily as the product of the diagonal entries, but it also makes other computations easier. The k-th power of a diagonal matrix D, for instance, is just the diagonal matrix of which the diagonal entries are the k-th power of the corresponding entries of D. In this chapter we will investigate these special bases consisting of so-called eigenvectors.

11.1. Eigenvalues and eigenvectors

Definition 11.1. Let $f\colon V \to V$ be an endomorphism of a vector space V. For any $\lambda \in F$, we define the λ-*eigenspace* of f by $E_\lambda(f) = \{\, v \in V \ : \ f(v) = \lambda v \,\}$; we say that λ is an *eigenvalue* of f if $E_\lambda(f)$ contains a nonzero vector, and we call such a vector an *eigenvector* for the eigenvalue λ. The *spectrum* $\Omega(f)$ of f is the set of eigenvalues of f.

Example 11.2. Let $V = \mathbb{R}^2$ and consider the map $f\colon V \to V$ given by $f(x, y) = (y, x)$. Then 1 and -1 are eigenvalues of f, and we have

$$E_1(f) = \{(x, x) : x \in \mathbb{R}\},$$
$$E_{-1}(f) = \{(x, -x) : x \in \mathbb{R}\}.$$

The eigenvectors $(1, 1)$ and $(1, -1)$ form a basis B of V, and the matrix of f relative to that basis is

$$[f]_B^B = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Example 11.3. Let $V = \mathcal{C}^\infty(\mathbb{R})$ be the space of infinitely differentiable functions on $\mathbb{R}$. Consider the endomorphism $D : f \mapsto f''$. Then every $\lambda \in \mathbb{R}$ is an eigenvalue, and all eigenspaces are of dimension two:

$$E_\lambda(D) = \begin{cases} L(x \mapsto 1, x \mapsto x) & \text{if } \lambda = 0 \\ L(x \mapsto e^{\mu x}, x \mapsto e^{-\mu x}) & \text{if } \lambda = \mu^2 > 0 \\ L(x \mapsto \sin \mu x, x \mapsto \cos \mu x) & \text{if } \lambda = -\mu^2 < 0 \end{cases}$$

Example 11.4. Let $s\colon \mathbb{R}^3 \to \mathbb{R}^3$ be the reflection in a plane $W \subset \mathbb{R}^3$. Then 1 is an eigenvalue with eigenspace $E_1(s) = W$, and -1 is an eigenvalue with eigenspace $E_{-1}(s) = W^\perp$.

If $\pi\colon \mathbb{R}^3 \to \mathbb{R}^3$ is the orthogonal projection onto W, then 1 is an eigenvalue with eigenspace $E_1(\pi) = W$, and 0 is an eigenvalue with eigenspace $E_0(\pi) = W^\perp$.

Since any matrices $A \in \mathrm{Mat}(n, F)$ can be identified with the associated linear map $f_A \colon F^n \to F^n$, it makes sense to speak about eigenvalues, eigenvectors, and eigenspaces of a square matrix.

Proposition 11.5. *Let $f \colon V \to V$ be an endomorphism of a vector space V. Suppose $v \in V$ is an eigenvector of f for eigenvalue λ. Then for every positive integer k, the vector v is an eigenvector for eigenvalue λ^k of the endomorphism*

$$f^k = \underbrace{f \circ f \circ \cdots \circ f}_{k} \colon V \to V.$$

If f is an automorphism, then v is an eigenvector of f^{-1} with eigenvalue λ^{-1}.

Proof. Exercise. □

Proposition 11.6. *Let $f \colon V \to V$ be an endomorphism of a vector space V over a field F. Then for any $\lambda \in F$, we have*

$$E_\lambda(f) = \ker(f - \lambda \cdot \mathrm{id}_V).$$

Proof. This follows immediately from the fact that for every $v \in V$ we have $f(v) = \lambda v$ if and only if $(f - \lambda \cdot \mathrm{id}_V)(v) = 0$. □

It follows that eigenspaces are indeed linear subspaces, as kernels are. We also conclude that the scalar λ is an eigenvalue of an endomorphism f if and only if $\ker(f - \lambda \cdot \mathrm{id}_V) \neq \{0\}$. If V is finite-dimensional, then we can use the determinant to find out whether this is the case.

Proposition 11.7. *Let $f \colon V \to V$ be an endomorphism of a finite-dimensional vector space V over a field F with an element $\lambda \in F$. Then λ is an eigenvalue of f if and only if $\det(f - \lambda \cdot \mathrm{id}_V) = 0$.*

Proof. Proposition 11.6 gives that λ is an eigenvalue of f if and only if we have $\ker(f - \lambda \cdot \mathrm{id}_V) \neq \{0\}$, so if and only if $f - \lambda \cdot \mathrm{id}_V$ is not injective, which is equivalent by Corollary 8.5 to the fact that $f - \lambda \cdot \mathrm{id}_V$ is not an isomorphism. By Proposition 10.31 this is the case if and only if $\det(f - \lambda \cdot \mathrm{id}_V) = 0$. □

Exercises

11.1.1. Prove Proposition 11.5.

11.1.2. Let $V = \mathcal{C}^\infty(\mathbb{R})$ be the space of infinitely differentiable functions on $\mathbb{R}$. Consider the endomorphism $D : f \mapsto f'$. Show that every $\lambda \in \mathbb{R}$ is an eigenvalue of D. Cf. Example 11.3 and Proposition 11.5.

11.1.3. For each matrix A of the following real matrices, find a basis for the eigenspace $E_\lambda(A)$ of each eigenvalue λ.

$$\begin{pmatrix} 5 & -4 \\ 8 & -7 \end{pmatrix} \qquad \begin{pmatrix} 3 & 2 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -3 \end{pmatrix}$$

11.1.4. Let V and W be vector spaces over F. Suppose $f \colon V \to V$ is an endomorphism of V and let $\varphi \colon V \to W$ be an isomorphism. Define the endomorphism g of W as $g = \varphi \circ f \circ \varphi^{-1}$. Show that for every scalar $\lambda \in F$, the map φ restricts to an isomorphism $E_\lambda(f) \to E_\lambda(g)$ of eigenspaces.

$$\begin{tikzcd} W \arrow[r, "g"] & W \\ V \arrow[u, "\cong"', "\varphi"] \arrow[r, "f"'] & V \arrow[u, "\cong"', "\varphi"] \end{tikzcd}$$

11.2. The characteristic polynomial

How do we find all eigenvalues (and eigenvectors) of an endomorphism? Of course, we can not just try all elements of F. If we want to find all eigenvalues of an endomorphism $f \colon V \to V$ of a *finite-dimensional* vector space V, then we can use the characteristic polynomial of f, defined as follows.

Definition 11.8. Let A be an $n \times n$ matrix over F. The *characteristic polynomial* $P_A \in F[t]$ of A is a polynomial over F (see Example 2.13) in the variable t, defined by

$$P_A(t) = \det(t \cdot I_n - A).$$

Remark 11.9. This entire book, including the definition of determinants, we have worked over a field F. So what do we mean by the determinant of $t \cdot I_n - A$, which involves a variable t?

One answer we could give is that we temporarily work over the field $F(t)$ of rational functions in t, which are quotients of polynomials in t (of course with nonzero denominator). Rational functions are added and multiplied as you would expect, so if $f_1, f_2, g_1, g_2 \in F[t]$ are polynomials with g_1, g_2 nonzero, then

$$\frac{f_1}{g_1} \cdot \frac{f_2}{g_2} = \frac{f_1 f_2}{g_1 g_2} \qquad \text{and} \qquad \frac{f_1}{g_1} + \frac{f_2}{g_2} = \frac{f_1 g_2 + f_2 g_1}{g_1 g_2}.$$

We leave it as an exercise to show that this does indeed make the set $F(t)$ into a field (with the same 0 and 1 as in F), over which we can do everything we have done so far. The fact that the determinant of $t \cdot I_n - A$ is not just a rational function, but in fact a polynomial, follows from the formula for the determinant given in Definition 10.5: we never divide to create denominators.

However, this answer would not satisfy the readers who have assumed throughout the book that our base field F has been (contained in) $\mathbb{R}$ or $\mathbb{C}$ (see the introduction of Chapter 2). If all of a sudden we are expected to apply the theory over $\mathbb{R}(t)$, then that assumption no longer holds. For those readers we refer again to the formula for the determinant given in Definition 10.5. If we write $A = (a_{ij})_{ij} \in \mathrm{Mat}(n, F)$, then we can apply this formula to the matrix

$$t \cdot I_n - A = \begin{pmatrix} t - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & t - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & t - a_{nn} \end{pmatrix}$$

to define the characteristic polynomial P_A of A. A very careful reader would still protest, as we now do not know whether the polynomial we obtain from the formula in Definition 10.5 depends on the choice of the index i that was made in that formula. After all, a priori, the proof that the determinant does not depend on that choice when we work over a field may have made crucial use of the fact that we were working over a field. However, we claim that the choice stil does not matter. Indeed, suppose that for two different choices, we got two polynomials, say P_A and P'_A. For each scalar $\lambda \in F$, the evaluations $P_A(\lambda)$ and $P'_A(\lambda)$ both equal $\det(\lambda \cdot I_n - A)$, which of course is well defined as the matrix $\lambda \cdot I_n - A$ is defined over the field F. Hence, every $\lambda \in F$ is a zero of the difference $P_A - P'_A$. Because every nonzero polynomial has only finitely many zeroes (cf. Example 7.50) and every subfield of $\mathbb{C}$ has infinitely many elements, we conclude that $P_A - P'_A = 0$, so $P_A = P'_A$, and the characteristic polynomial is indeed independent of the choice of the index i in Definition 10.5.

This second answer can in fact be applied for any field, except for the final step, where we need infinitely many scalars $\lambda \in F$. For finite fields, this can be resolved by taking an infinite field extension F' of F (for example, the field of rational functions over F).

Example 11.10. The characteristic polynomial of the matrix

$$A = \begin{pmatrix} 2 & 3 \\ -1 & 4 \end{pmatrix}$$

is

$$\det(t \cdot I_2 - A) = \det \begin{pmatrix} t-2 & -3 \\ 1 & t-4 \end{pmatrix} = (t-2)(t-4) - (-3) \cdot 1 = t^2 - 6t + 11.$$

Proposition 11.11. *Let n a non-negative integer and $A, Q \in \mathrm{Mat}(n, F)$ matrices with Q invertible. Set $A' = QAQ^{-1}$. Then the characteristic polynomials P_A and $P_{A'}$ are equal.*

Proof. We have

$$t \cdot I_n - A' = t \cdot QI_nQ^{-1} - QAQ^{-1} = Q(t \cdot I_n - A)Q^{-1}.$$

Taking the determinant of both sides gives

$$P_{A'}(t) = \det\big(Q(t \cdot I_n - A)Q^{-1}\big) = \det(Q) \cdot P_A(t) \cdot \det(Q^{-1}).$$

From $\det(Q)\det(Q^{-1}) = 1$, we find $P_{A'}(t) = P_A(t)$. Note that we used the multiplicativity of the determinant, which we proved in Proposition 10.13, but strictly speaking only over fields. This means that our proof here is not quite complete yet. As in Remark 11.9, we make it complete by either working over the field $F(t)$ of rational functions over F, or by noting that every element of F and of any field extension of F is a zero of the difference $P_{A'} - P_A$, which also implies $P_{A'} = P_A$. $\square$

Proposition 11.11 implies that the following definition makes sense.

Definition 11.12. Let $f\colon V \to V$ be an endomorphism of a finite-dimensional vector space V. Then we define the *characteristic polynomial* $P_f \in F[t]$ of f to be the characteristic polynomial of $[f]_B^B$ for any basis B of V.

By Propositions 9.19 and 11.11, the characteristic polynomial P_f is well defined, as it does not depend on the choice of the basis B.

Example 11.13. If V is an n-dimensional vector space, then for the identity $\mathrm{id}_V\colon V \to V$ we have $P_{\mathrm{id}_V} = (t-1)^n$.

Example 11.14. By Example 9.2, we of course have $P_{f_A} = P_A$ for any square matrix A.

Remark 11.15. Let V be an n-dimensional vector space with basis B and let $f\colon V \to V$ be an endomorphism. Set $A = [f]^B_B$. If we write $A = (a_{ij})_{ij}$, then

$$P_f(t) = P_A(t) = \det(t \cdot I_n - A) = \begin{vmatrix} t - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & t - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & t - a_{nn} \end{vmatrix}.$$

Expanding the determinant, we find (Exercise 11.2.1)

$$\begin{aligned} P_f(t) = P_A(t) &= t^n - \operatorname{Tr}(A)t^{n-1} + \cdots + (-1)^n \det(A) \\ &= t^n - \operatorname{Tr}(f)t^{n-1} + \cdots + (-1)^n \det(f). \end{aligned}$$

Proposition 11.16. *Let V be a finite-dimensional vector space over F and let $f\colon V \to V$ be an endomorphism. Then an element $\lambda \in F$ is an eigenvalue of f if and only if λ is a root of the characteristic polynomial P_f, that is, $P_f(\lambda) = 0$.*

Proof. Set $n = \dim V$ and let B be a basis for V. Set $A = [f]^B_B$. We have

$$P_f(\lambda) = P_A(\lambda) = \det(\lambda \cdot I_n - A) = \det(\lambda \cdot \mathrm{id}_V - f) = (-1)^n \cdot \det(f - \lambda \cdot \mathrm{id}_V),$$

so $P_f(\lambda) = 0$ if and only if $\det(f - \lambda \cdot \mathrm{id}_V) = 0$. The statement therefore follows immediately from Proposition 11.7. $\square$

Example 11.17. Let us come back to the earlier example $f : (x, y) \mapsto (y, x)$ on $\mathbb{R}^2$ of Example 11.2. With respect to the canonical basis E, the associated matrix is

$$[f]^E_E = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

so the characteristic polynomial is

$$P_f(t) = \begin{vmatrix} t & -1 \\ -1 & t \end{vmatrix} = t^2 - 1$$

and the eigenvalues are the two roots 1 and -1.

Example 11.18. Let us consider the real matrix

$$A = \begin{pmatrix} 5 & 2 & -6 \\ -1 & 0 & 1 \\ 3 & 1 & -4 \end{pmatrix}.$$

What are its eigenvalues and eigenspaces? We compute the characteristic polynomial:

$$\begin{aligned} P_A(t) &= \begin{vmatrix} t-5 & -2 & 6 \\ 1 & t & -1 \\ -3 & -1 & t+4 \end{vmatrix} \\ &= (t-5)\big(t(t+4) - 1\big) + 2\big((t+4) - 3\big) + 6\big(-1 + 3t\big) \\ &= t^3 - t^2 - t + 1 = (t-1)^2(t+1). \end{aligned}$$

The roots are 1 and -1; these are therefore the eigenvalues. To find (bases of) the eigenspaces, note that $E_\lambda(A) = \ker(A - \lambda I_3)$. For $\lambda = 1$, we have

$$A - I_3 = \begin{pmatrix} 4 & 2 & -6 \\ -1 & -1 & 1 \\ 3 & 1 & -5 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

(by elementary row operations), so $E_1(A) = \ker(A - I_3)$ is generated by $(2, -1, 1)$. For $\lambda = -1$, we obtain

$$A + I_3 = \begin{pmatrix} 6 & 2 & -6 \\ -1 & 1 & 1 \\ 3 & 1 & -3 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and so $E_{-1}(A) = \ker(A + I_3)$ is generated by $(1, 0, 1)$.

Exercises

11.2.1. Let A be an $n \times n$ matrix. Show that we have

$$P_A(t) = t^n - \operatorname{Tr}(A)t^{n-1} + \cdots + (-1)^n \det(A),$$

that is, the coefficients of t^{n-1} equals $-\operatorname{Tr}(A)$ and the constant coefficient equals $(-1)^n \det(A)$.

11.2.2. What is the characteristic polynomial of the reflection $s\colon \mathbb{R}^3 \to \mathbb{R}^3$ in some plane $V \subset \mathbb{R}^3$?

11.2.3. For each matrix A of the following real matrices, find a basis for the eigenspace $E_\lambda(A)$ of each eigenvalue λ.

$$\begin{pmatrix} -6 & -4 \\ 8 & 6 \end{pmatrix} \qquad \begin{pmatrix} 1 & 1 \\ -4 & 5 \end{pmatrix}$$

$$\begin{pmatrix} 7 & 0 & 8 \\ 0 & 3 & 0 \\ -4 & 0 & -5 \end{pmatrix} \qquad \begin{pmatrix} 0 & -1 & 0 \\ 4 & 4 & 0 \\ 2 & 1 & 2 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 1 & 0 & 0 \\ -2 & 0 & 0 & 0 \\ -2 & -2 & 1 & 0 \\ -9 & -9 & 0 & -3 \end{pmatrix} \qquad \begin{pmatrix} 2 & -1 & 0 & 3 \\ 0 & 1 & 0 & 2 \\ -2 & 1 & 1 & -6 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

11.2.4. Let $\varphi\colon V \to W$ be an isomorphism of finite-dimensional vector spaces, and $f\colon V \to V$ an endomorphism of V. Show that $f' = \varphi \circ f \circ \varphi^{-1}$ is an endomorphism of W satisfying $P_{f'} = P_f$, cf. Exercise 10.4.2.

11.2.5. Let A be a square matrix and $A^\top$ its transpose.

(1) Show that the characteristic polynomials P_A and $P_{A^\top}$ are equal.
(2) Show that A and $A^\top$ have the same eigenvalues.

11.2.6. Let F be a field and $a_0, a_1, \ldots, a_{d-1} \in F$. Show that there is a matrix $A \in \operatorname{Mat}(d, F)$ with $P_A = t^d + a_{d-1}t^{d-1} + \ldots + a_1 t + a_0$.

11.2.7. Let $f\colon V \to V$ be an endomorphism of a finite-dimensional vectorspace V. Let $\sigma\colon V \to W$ be a linear map. Suppose that $f(\ker \sigma) \subset \ker \sigma$. Let f' be the restriction of f to $\ker \sigma$ and let f'' be the endomorphism of $\operatorname{im} \sigma$ induced by f (see Exercises 4.5.1 and 10.4.3). Show that $P_f = P_{f'} \cdot P_{f''}$.

11.2.8. (1) Let A and B be two $n \times n$ matrices. Show that if A or B is invertible, then the characteristic polynomials P_{AB} and P_{BA} are equal.
*(2) Show that in fact the same is true if both A and B are not invertible.
*(3) Suppose A is an $m \times n$ matrix and B an $n \times m$ matrix. Show that we have $t^n P_{AB}(t) = t^m P_{BA}(t)$.

11.3. Diagonalization

Definition 11.19. Let $f: V \to V$ be an endomorphism of a finite-dimensional vector space V. Then f is *diagonalizable* if there exists a basis B for V such that the matrix $[f]_B^B$ associated to f with respect to B is diagonal. A matrix A is *diagonalizable* if the associated linear map f_A is diagonalizable.

Recall from Proposition 9.19 that for any two bases B and C for V we have

$$[f]_B^B = P^{-1} \cdot [f]_C^C \cdot P$$

with $P = [\mathrm{id}]_C^B$. In particular, for $V = F^n$, $C = E$, and $f = f_A$ for some matrix A, we have $[f]_C^C = [f_A]_E^E = A$, and we find that A is diagonalizable if and only if there is an invertible matrix P such that $P^{-1}AP$ is diagonal (see Lemma 9.27). We also conclude that, in general, the endomorphism f is diagonalizable if and only if the matrix $[f]_C^C$ is diagonalizable for some (and thus every) basis C for V.

Proposition 11.20. *Let $f: V \to V$ be an endomorphism of a finite-dimensional vector space V with basis $B = (v_1, \ldots, v_n)$. Then $[f]_B^B$ is a diagonal matrix with diagonal entries $\lambda_1, \ldots, \lambda_n$ if and only if for all $1 \leq j \leq n$, the vector v_j is an eigenvector of f for eigenvalue λ_j.*

Proof. The j-th column of $[f]_B^B$ is the sequence $(f(v_j))_B$ of coefficients of $f(v_j)$ with respect to B by Lemma 9.7. Hence, the matrix $[f]_B^B$ is diagonal with diagonal entries $\lambda_1, \ldots, \lambda_n$ if and only if, for each j, the j-th column $(f(v_j))_B$ equals $\lambda_j e_j$, which happens if and only if, for each j, we have $f(v_j) = \lambda_j v_j$, that is, v_j is an eigenvector of f for eigenvalue λ_j. □

It follows that $f: V \to V$ is diagonalizable if and only if there exists a basis for V consisting of eigenvectors of f.

The big question is now: when is a matrix or endomorphism diagonalizable?

This is certainly not always the case. In Example 11.18, for instance, we found only two linearly independent eigenvectors in $\mathbb{R}^3$, and so there cannot be a basis of eigenvectors. Another example is $f : (x, y) \mapsto (-y, x)$ on $\mathbb{R}^2$. The characteristic polynomial equals $t^2 + 1$ and does not have roots in $\mathbb{R}$, so there are no eigenvalues and therefore no eigenvectors. (If we take $\mathbb{C}$ instead as the field of scalars, then we do have two roots $\pm i$, and f becomes diagonalizable.)

Lemma 11.21. *Let V be an F-vector space and $f : V \to V$ an endomorphism. Let $\lambda_1, \ldots, \lambda_m \in F$ be distinct, and for $i = 1, \ldots, m$, let $v_i \in E_{\lambda_i}(f)$. If*

$$v_1 + v_2 + \cdots + v_m = 0\,,$$

then $v_i = 0$ for all i.

Proof. We use induction on m. The case $m = 0$ (or $m = 1$) is trivial. So assume the claim is true for m, and consider the case with $m+1$ eigenvalues. We apply the endomorphism $f - \lambda_{m+1}\,\mathrm{id}_V$ to the equation

$$v_1 + v_2 + \cdots + v_m + v_{m+1} = 0$$

and obtain (note $(f - \lambda_{m+1}\,\mathrm{id}_V)(v_{m+1}) = 0$)

$$(\lambda_1 - \lambda_{m+1})v_1 + (\lambda_2 - \lambda_{m+1})v_2 + \cdots + (\lambda_m - \lambda_{m+1})v_m = 0\,.$$

By induction, we find that $(\lambda_i - \lambda_{m+1})v_i = 0$ for all $1 \le i \le m$. Since $\lambda_i \neq \lambda_{m+1}$, this implies $v_i = 0$ for $1 \le i \le m$. But then we must also have $v_{m+1} = 0$. $\square$

Alternative proof. Set $C = (v_1, \ldots, v_m)$. Set $v = v_1 + v_2 + \cdots + v_m = 0$. Then for every integer k, we have $f^k(v) = 0$, where $f^k = f \circ f \circ \cdots \circ f$ is the composition of k copies of f; this gives

$$0 = f^k(v) = \lambda_1^k v_1 + \cdots + \lambda_m^k v_m,$$

so the vector $a_k = (\lambda_1^k, \ldots, \lambda_m^k)$ is contained in the kernel of the linear map $\varphi_C \colon F^m \to V$ that sends e_j to v_j. By Example 9.5 and Exercise 9.1.3, the Vandermonde matrix with columns $a_0, a_1, \ldots, a_{m-1}$ is invertible, so these columns span F^m. We conclude $\ker \varphi_C = F^m$, so φ_C is the zero map and $v_j = 0$ for all j. $\square$

Example 11.22. We can use this to show once again that the power functions $f_n : x \mapsto x^n$ for $n \in \mathbb{Z}_{\geq 0}$ are linearly independent as elements of the space P of polynomial functions on $\mathbb{R}$. Namely, consider the endomorphism $D : P \to P$, $f \mapsto (x \mapsto xf'(x))$. Then $D(f_n) = nf_n$, so the f_n are eigenvectors of D for eigenvalues that are pairwise distinct, hence they must be linearly independent by Lemma 11.21.

Corollary 11.23. *Let V be an F-vector space and $f : V \to V$ an endomorphism. Let $\lambda_1, \ldots, \lambda_m \in F$ be distinct, and for each $1 \le i \le m$, let B_i be a basis for $E_{\lambda_i}(f)$. Then the concatenation of $B_1, B_2, \ldots, B_m$ is a sequence of linearly independent vectors.*

Proof. Let v be a linear combination of the elements in $B_1, B_2, \ldots, B_m$. Then v can be written as $v = v_1 + v_2 + \cdots + v_m$ with v_i the part of the linear combination that uses elements in B_i, so $v_i \in E_{\lambda_i}(f)$. Suppose $v = 0$. Then by Lemma 11.21, we have $v_i = 0$ for all i. Since the elements of B_i are linearly independent, all the coefficients in the linear combination that gives v_i vanish. We conclude that all coefficients in the original linear combination that gives v vanish, so indeed, the concatenation of $B_1, B_2, \ldots, B_m$ is a sequence of linearly independent vectors. $\square$

Corollary 11.24. *Let V be a finite-dimensional F-vector space and $f : V \to V$ an endomorphism. Then we have*

$$\sum_{\lambda \in F} \dim E_\lambda(f) \le \dim V$$

and equality holds if and only if f is diagonalizable.

Proof. Since every polynomial of degree d has at most d zeros, it follows from Proposition 11.16 that f has only finitely many eigenvalues, say $\lambda_1, \lambda_2, \ldots, \lambda_m$. For each $1 \le i \le m$, let B_i be a basis of the eigenspace $E_{\lambda_i}(f)$. By Corollary 11.23, the concatenation of $B_1, B_2, \ldots, B_m$ is a sequence of linearly independent vectors of length $\sum_{\lambda \in F} \dim E_\lambda(f)$. It follows from Theorem 7.47(1) that this is at most $\dim V$, which proves the inequality. If f is diagonalizable, then there is a basis consisting of eigenvectors, and so we must have equality. Conversely, if we have equality, then the union of bases of the eigenspaces will be a basis of V that consists of eigenvectors of f. $\square$

The following proposition gives sufficient, though *not* necessary, conditions for an endomorphism to be diagonalizable.

Proposition 11.25. *Let V be an n-dimensional F-vector space and $f : V \to V$ an endomorphism. If $P_f(t)$ has n distinct roots in F, then f is diagonalizable.*

Proof. In this case, there are n distinct eigenvalues $\lambda_1, \dots, \lambda_n$. Therefore, $E_{\lambda_i}(f)$ is nontrivial for $1 \le i \le n$, which means that $\dim E_{\lambda_i}(f) \ge 1$. So

$$\dim V = n \le \sum_{i=1}^{n} \dim E_{\lambda_i}(f) \le \dim V\,,$$

and we must have equality. The result then follows by the previous corollary. □

The converse of this statement is false in general, as the identity endomorphism id_V shows (for $\dim V \ge 2$).

Example 11.26. Consider the real matrix

$$A = \begin{pmatrix} -5 & 6 & 6 \\ 0 & 1 & 0 \\ -3 & 3 & 4 \end{pmatrix}.$$

We want to know if A is diagonalizable and, if so, find an invertible 3×3 matrix P such that $P^{-1}AP$ is diagonal. This means we want to know whether there exists a basis of eigenvectors. We first compute the characteristic polynomial to determine the eigenvalues. We expand along the second row to get

$$P_A(t) = \begin{vmatrix} t+5 & -6 & -6 \\ 0 & t-1 & 0 \\ 3 & -3 & t-4 \end{vmatrix} = (t-1) \cdot \big((t+5)(t-4) + 18\big) = (t-1)^2(t+2).$$

This shows that the eigenvalues are $\lambda_1 = 1$ and $\lambda_2 = -2$. To find the eigenspaces $E_\lambda(A) = \ker(A - \lambda I_3)$, we apply elementary row operations to $A - \lambda I_3$ to obtain the reduced row echelon form. We get

$$A - I_3 = \begin{pmatrix} -6 & 6 & 6 \\ 0 & 0 & 0 \\ -3 & 3 & 3 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & -1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and

$$A + 2I_3 = \begin{pmatrix} -3 & 6 & 6 \\ 0 & 3 & 0 \\ -3 & 3 & 6 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

We conclude that the eigenspace $E_1(A) = \ker(A - I_3)$ has a basis (v_1, v_2) and the eigenspace $E_{-2}(A) = \ker(A + 2I_3)$ has a basis (v_3) with

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \qquad v_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \qquad v_3 = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}.$$

The vectors v_1, v_2, v_3 are linearly independent by Corollary 11.23, so they form a basis $B = (v_1, v_2, v_3)$ for $\mathbb{R}^3$ of eigenvectors of A, which already shows that A

is diagonalizable. The corresponding eigenvalues are 1, 1, -2, respectively, so we get

$$[f_A]_B^B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}$$

by Proposition 11.20. Furthermore, if we set $D = [f_A]_B^B$ and (see Example 9.13)

$$P = [\mathrm{id}]_E^B = \begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix} = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix},$$

then we find

$$D = [f_A]_B^B = [\mathrm{id}]_B^E \cdot [f_A]_E^E \cdot [\mathrm{id}]_E^B = P^{-1}AP.$$

Remark 11.27. Let A be an $n \times n$ matrix over a field F. Assume that, analogously to Example 11.26, there is a basis $B = (v_1, \ldots, v_n)$ for F^n consisting of eigenvectors of A, corresponding to eigenvalues $\lambda_1, \ldots, \lambda_n$, respectively. Set

$$D = [f_A]_B^B = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \quad \text{and} \quad P = [\mathrm{id}]_E^B = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix}.$$

Then again we have

$$D = [f_A]_B^B = [\mathrm{id}]_B^E \cdot [f_A]_E^E \cdot [\mathrm{id}]_E^B = P^{-1}AP.$$

We can verify the equivalent identity $PD = AP$ also differently. Note that for each $1 \le j \le n$, we have $A \cdot v_j = \lambda_j v_j$. This implies

$$AP = \begin{pmatrix} | & | & & | \\ Av_1 & Av_2 & \cdots & Av_n \\ | & | & & | \end{pmatrix} = \begin{pmatrix} | & | & & | \\ \lambda_1 v_1 & \lambda_2 v_2 & \cdots & \lambda_n v_n \\ | & | & & | \end{pmatrix} = PD.$$

Example 11.28. Let F be a field, n a positive integer, and let $D\colon F[x]_n \to F[x]_n$ be the linear map that sends a polynomial $f \in F[x]_n$ to its derivative f'. Note that D^{n+1} is the zero map, so the only eigenvalue of D^{n+1} is 0. It follows from Proposition 11.5 that D can have no other eigenvalue than 0. The corresponding eigenspace $E_0(D) = \ker D$ consists of only the constant polynomials. This implies that there is no basis of eigenvectors, so D is not diagonalizable.

Example 11.29. Let $a, b \in \mathbb{R}^n$ be two nonzero vectors with $\langle a, b \rangle = 0$. Let $T\colon \mathbb{R}^n \to \mathbb{R}^n$ be the map defined by $T(x) = \langle x, a \rangle \cdot b$. Then $T^2 = T \circ T$ is the zero map, so as in the previous example, the map T has no eigenvalue other than 0. The eigenspace $E_0(T) = \ker T$ is the hyperplane $a^\perp$, which is a proper subspace of $\mathbb{R}^n$, so there is no basis of eigenvectors and T is not diagonalizable.

Note that for any $n \times n$ matrices D, P, with P invertible, and $A = PDP^{-1}$, and any positive integer k, we find

$$A^k = (PDP^{-1})^k = \underbrace{(PDP^{-1})(PDP^{-1}) \cdots (PDP^{-1})}_{k} = PD^kP^{-1}.$$

In fact, if D is invertible, then the identity $A^k = PD^kP^{-1}$ holds for every integer k, also if k is negative (Exercise 11.3.1). If D is a diagonal matrix with diagonal entries $\lambda_1, \ldots, \lambda_n$, and $k \ge 0$, then D^k is a diagonal matrix with diagonal entries $\lambda_1^k, \ldots, \lambda_n^k$. This gives an efficient way to compute A^k if A is diagonalizable.

Example 11.30. Take the matrix A as in Example 11.26. We found $A = PDP^{-1}$ with

$$D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} \qquad \text{and} \qquad P = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}.$$

We conclude that for any integer k, we have

$$\begin{aligned} A^k = PD^kP^{-1} &= \begin{pmatrix} 1 & 1 & 2 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & (-2)^k \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ -1 & 1 & 2 \\ 1 & -1 & -1 \end{pmatrix} \\ &= \begin{pmatrix} 2(-2)^k - 1 & (-2)^{k+1} + 2 & (-2)^{k+1} + 2 \\ 0 & 1 & 0 \\ (-2)^k - 1 & 1 - (-2)^k & 2 - (-2)^k \end{pmatrix}. \end{aligned}$$

Proposition 11.25 only gives sufficient conditions for an endomorphism to be diagonalizable. Before we give necessary and sufficient conditions for a matrix (or an endomorphism of a finite-dimensional vector space) to be diagonalizable, we will do some preparations.

Definition 11.31. Let V be a finite-dimensional F-vector space, $f : V \to V$ an endomorphism and $\lambda \in F$. Then $\dim E_\lambda(f)$ is called the *geometric multiplicity* of the eigenvalue λ of f. (So the geometric multiplicity is positive if and only if λ is indeed an eigenvalue.)

Recall that if F is a field, then the *degree* of a nonzero polynomial $p = \sum_{i=0}^{d} a_i t^i \in F[t]$ with $a_d \neq 0$ is d; the coefficient a_d is called the *leading coefficient* of p. A *monic polynomial* is a nonzero polynomial with leading coefficient equal to 1.

For example, if V is an n-dimensional vector space and $f : V \to V$ is an endomorphism, then the characteristic polynomial P_f of f is monic of degree n.

Proposition 11.32. *Let $p \in F[t]$ be a monic polynomial and $\alpha \in F$. Then there is a largest $m \in \mathbb{Z}_{\geq 0}$ such that $p = (t-\alpha)^m q$ for some polynomial $q \in F[t]$; we then have $q(\alpha) \neq 0$.*

Proof. If we have $p = (t-\alpha)^m q$ for some $m \geq 0$ and $q \in F[t]$, then from $\deg(p) = m + \deg(q)$ we obtain $m \leq d$. Hence m is bounded and we can write $p = (t-\alpha)^m q$ with m as large as possible. Then we must have $q(\alpha) \neq 0$, since otherwise we could write $q = (t-\alpha)r$ for some $r \in F[t]$ by Example 8.4 (for $k = 1$), which would yield $p = (t-\alpha)^{m+1} r$, contradicting our choice of m. □

Definition 11.33. Given p and m as in the corollary above, the number m is called the *multiplicity* of the root α of p; we have $m > 0$ if and only if $p(\alpha) = 0$.

Now we can make another definition.

Definition 11.34. Let V be a finite-dimensional F-vector space and $f : V \to V$ an endomorphism. Then the multiplicity of $\lambda \in F$ as a root of the characteristic polynomial P_f is called the *algebraic multiplicity* of the eigenvalue λ of f.

Note that the following statements are then equivalent.

(1) λ is an eigenvalue of f;

(2) the geometric multiplicity of λ is ≥ 1;
(3) the algebraic multiplicity of λ is ≥ 1.

We also know that the sum of the geometric multiplicities of all eigenvalues is bounded by $\dim V$. The following result shows that the same holds for the sum of the algebraic multiplicities of all eigenvalues.

Lemma 11.35. *Let $f : V \to V$ be an endomorphism of an n-dimensional vector space V over F, and let P_f be its characteristic polynomial. Then the sum of the algebraic multiplicities of the eigenvalues of f is at most n; it is equal to n if and only if P_f is a product of linear factors $t - \lambda$ (with $\lambda \in F$).*

Proof. By Example 8.4 (for $k = 1$), if λ is a root of P_f, we can write $P_f = (t-\lambda)q$ for a monic polynomial q of degree $n - 1$. Continuing in this way, we can write

$$P_f = (t - \lambda_1)^{m_1} \cdots (t - \lambda_k)^{m_k} q$$

for a monic polynomial q that does not have roots in F and distinct elements $\lambda_1, \ldots, \lambda_k \in F$. If $\mu \in F$, then

$$P_f(\mu) = (\mu - \lambda_1)^{m_1} \cdots (\mu - \lambda_k)^{m_k} q(\mu)\,,$$

so if $P_f(\mu) = 0$, then $\mu \in \{\lambda_1, \ldots, \lambda_k\}$ (since $q(\mu) \neq 0$). Therefore the eigenvalues are exactly $\lambda_1, \ldots, \lambda_k$, with algebraic multiplicities $m_1, \ldots, m_k$, and

$$m_1 + m_2 + \cdots + m_k \leq m_1 + m_2 + \cdots + m_k + \deg(q) = n\,.$$

We have equality if and only if $\deg(q) = 0$, that is, $q = 1$; this holds if and only if

$$P_f = (t - \lambda_1)^{m_1} \cdots (t - \lambda_k)^{m_k}$$

is a product of linear factors. □

There is one further important relation between the multiplicities.

Theorem 11.36. *Let V be a finite-dimensional F-vector space, $f : V \to V$ an endomorphism, and $\lambda \in F$. Then the geometric multiplicity of λ as an eigenvalue of f is not larger than its algebraic multiplicity.*

Proof. We can choose a basis $v_1, \ldots, v_k, v_{k+1}, \ldots, v_n$ of V such that $v_1, \ldots, v_k$ form a basis of the eigenspace $E_\lambda(f)$; then k is the geometric multiplicity. The matrix associated to f relative to this basis then has the form

$$A = \begin{pmatrix} \lambda & 0 & \ldots & 0 & * & \ldots & * \\ 0 & \lambda & \ldots & 0 & * & \ldots & * \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & \lambda & * & \ldots & * \\ 0 & 0 & \ldots & 0 & * & \ldots & * \\ \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & * & \ldots & * \end{pmatrix} = \left(\begin{array}{c|c} \lambda I_k & B \\ \hline 0 & C \end{array}\right).$$

We then have

$$P_f = \det(t \cdot I_n - A) = \det \left(\begin{array}{c|c} (t-\lambda)\cdot I_k & -B \\ \hline 0 & t \cdot I_{n-k} - C \end{array}\right).$$

Expanding this determinant along the first column shows, by induction, that

$$P_f = (t-\lambda)^k \cdot \det(t \cdot I_{n-k} - C) = (t-\lambda)^k \cdot P_C(t),$$

which could alternatively also be obtained by Exercise 10.2.3. We see that λ has multiplicity at least k as a root of P_f, which proves the theorem. □

Corollary 11.37. *Let V be a finite-dimensional F-vector space and $f : V \to V$ an endomorphism. Then f is diagonalizable if and only if*

(1) *P_f is a product of linear factors, and*
(2) *for each $\lambda \in F$, its geometric and algebraic multiplicities as an eigenvalue of f agree.*

Proof. By Corollary 11.24, the map f is diagonalizable if and only if the sum of the geometric multiplicities of all eigenvalues equals $n = \dim V$. By Theorem 11.36, this implies that the sum of the algebraic multiplicities is at least n; however it cannot be larger than n, so it equals n as well. This already shows that geometric and algebraic multiplicities agree. By Lemma 11.35, we also see that P_f is a product of linear factors.

Conversely, if we can write P_f as a product of linear factors, this means that the sum of the algebraic multiplicities is n. If the geometric multiplicities equal the algebraic ones, their sum must also be n, hence f is diagonalizable. □

Remark 11.38. If F is an *algebraically closed* field, for example $F = \mathbb{C}$, then condition (1) in the corollary is automatically satisfied (by definition!). However, condition (2) can still fail. It is then an interesting question to see how close we can get to a diagonal matrix in this case. This is what the *Jordan Normal Form Theorem* is about, which will be a topic in Linear Algebra II (cf. Example 9.31).

Example 11.39. We will check whether the matrix

$$A = \begin{pmatrix} -3 & 1 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 5 \end{pmatrix}$$

is diagonalizable. The characteristic polynomial of A is $P_A = (t+3)^2(t-5)$, so the eigenvalues of A are -3 and 5 with algebraic multiplicities 2 and 1, respectively. Theorem 11.36 shows that the geometric multiplicity of 5 is 1 as well, so it suffices to check whether the geometric multiplicity of -3 is 2. One easily checks that the eigenspace $E_{-3}(A) = \ker(A + 3I_3)$ is generated by $(1,0,0)$, so the geometric multiplicity of -3 is 1, which does not equal its algebraic multiplicity, so A is not diagonalizable.

Exercises

11.3.1. Show that for any integer k, and any invertible $n \times n$ matrices D, P, we have $(PDP^{-1})^k = PD^kP^{-1}$.

11.3.2. Determine whether the following real matrices are diagonalizable. If not, explain why. If so, then determine an invertible matrix P and a diagonal matrix D, such that the matrix equals PDP^{-1}; also give a closed expression as in Example 11.30 for the k-th power of the matrix, where k is an arbitrary integer (in the case it is diagonalizable).

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 6 & -2 \\ 6 & -1 \end{pmatrix}, \qquad \begin{pmatrix} 3 & -1 & -1 \\ 4 & -2 & -4 \\ -2 & 2 & 4 \end{pmatrix}.$$

11.3.3. For each matrix A of the real matrices in Exercises 11.1.3 and 11.2.3, determine whether A is diagonalizable, and, if it is, determine a diagonal matrix D and an invertible matrix P, such that $A = PDP^{-1}$.

11.3.4. Consider the matrix

$$M = \begin{pmatrix} 4 & 6 & 2 \\ 0 & -3 & 0 \\ -4 & -12 & -2 \end{pmatrix}.$$

(1) Determine an invertible matrix P and a diagonal matrix D such that $M = PDP^{-1}$.
(2) Determine M^k for all positive integers k.

11.3.5. Determine M^k for the following matrices M and all integers k.

$$\begin{pmatrix} 7 & -10 \\ 5 & -8 \end{pmatrix} \qquad \begin{pmatrix} -2 & 3 & -7 \\ 0 & -4 & 6 \\ 0 & -3 & 5 \end{pmatrix}$$

11.3.6. Define the sequence $F_0, F_1, F_2, F_3, \ldots$ of Fibonacci numbers by $F_0 = 0$, $F_1 = 1$, and

$$F_n = F_{n-1} + F_{n-2}$$

for all $n \geq 2$.

(1) Show that for the matrix

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

and the vectors

$$x_n = \begin{pmatrix} F_n \\ F_{n+1} \end{pmatrix},$$

we have $Ax_n = x_{n+1}$ for all $n \geq 0$.
(2) Find constants $\alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathbb{R}$ such that for every $n \geq 0$ we have

$$F_n = \alpha_1 \beta_1^n + \alpha_2 \beta_2^n.$$

11.3.7. Show that a polynomial of degree n over a field F has at most n roots in F.

11.3.8. Let F be an infinite field, that is, $|F| = \infty$, and consider the map $\varphi\colon F[x] \to F^F$ of Exercise D.2.1, cf. Exercises 4.1.9, 7.4.6, and 7.4.7.

(1) Show that φ is injective.
(2) Show that φ induces an isomorphism from $F[x]$ to the subspace $P(F)$ of F^F consisting of *polynomial functions*.
(3) Show that $\dim P(F) = \infty$.

11.3.9. Determine for each of the following matrices M whether they are diagonalizable over F for $F = \mathbb{R}$ and $F = \mathbb{C}$. If so, then give an invertible matrix P and a diagonal matrix D such that $M = PDP^{-1}$.

$$\begin{pmatrix} 2 & 1 \\ -5 & -2 \end{pmatrix} \qquad \begin{pmatrix} 2 & -3 & -2 \\ 0 & 1 & 0 \\ 4 & -2 & -2 \end{pmatrix}.$$

11.3.10. The same as the previous exercise for

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix} \qquad \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

11.3.11. For which angle θ is the rotation $\mathbb{R}^2 \to \mathbb{R}^2$ about 0 over θ diagonalizable over $\mathbb{R}$?

11.3.12. Let $n \geq 2$ be an integer. Let M_n be as in Exercise 10.2.5 and set $N_n = M_n + I_n$, so that N_n is an $n \times n$ matrix with all entries equal to 1.

(1) Show $\operatorname{rk} N_n = 1$ and $\dim \ker N_n = n - 1$.

(2) Show that the eigenvalues of N_n are 0 and n.
(3) Show that N_n is diagonalizable.
(4) Show that the characteristic polynomial of N_n equals $t^n - nt^{n-1}$.
(5) Show $\det M_n = (-1)^{n-1}(n-1)$.

APPENDIX A

Review of maps

Let X, Y and Z be sets.

A *map* or *function* $f : X \to Y$ is a 'black box' that for any given $x \in X$ gives us back some $f(x) \in Y$ that only depends on x. More formally, we can define functions by identifying f with its *graph*

$$\Gamma_f = \{(x, f(x)) : x \in X\} \subset X \times Y .$$

In these terms, a function or map from X to Y is a subset $f \subset X \times Y$ such that for every $x \in X$ there is a unique $y \in Y$ such that $(x, y) \in f$; we then write $f(x) = y$. It is important to keep in mind that the data of a function include the *domain* X and *target* (or *codomain*) Y.

If the domain of f equals $X_1 \times X_2 \times \cdots \times X_n$ for some sets $X_1, X_2, \ldots, X_n$, then we often write $f(x_1, x_2, \ldots, x_n)$ instead of $f((x_1, x_2, \ldots, x_n))$.

If $f \colon X \to Y$ is a map, then we call $\{f(x) : x \in X\} \subset Y$ the *image* of f, $\mathrm{im}(f)$. The map f is called *injective* if no two elements of X are mapped to the same element of Y. More formally, if $x, x' \in X$ and $f(x) = f(x')$, then $x = x'$. The map f is called *surjective* if its image is all of Y. Equivalently, for all $y \in Y$ there is some $x \in X$ such that $f(x) = y$. The map f is called *bijective* if it is both injective and surjective. In this case, there is a unique *inverse map* $g \colon Y \to X$ such that $g(y) = x \iff f(x) = y$ for all $x \in X$ and $y \in Y$; this is equivalent to the statement that both $f \circ g = \mathrm{id}_Y$ and $g \circ f = \mathrm{id}_X$ hold, and in this case we write f^{-1} for the function g.

We denote the set of all functions from X to Y by $\mathrm{Map}(X, Y)$ or Y^X. The following proposition shows how the latter notation relates to more familiar notation.

Proposition A.1. *Let $n \geq 0$ be an integer, and Y a set. Set $I = \{1, 2, \ldots, n\}$. The map*

$$T \colon Y^I \to Y^n$$

that sends a function f to the sequence $(f(1), f(2), \ldots, f(n))$ is a bijection. Its inverse sends a sequence $(y_1, y_2, \ldots, y_n)$ to the function given by $i \mapsto y_i$.

Proof. One easily checks that the two maps are each other's inverses. □

Exercise A.0.1 will give more justification for the notation Y^X.

A map $f : X \to Y$ induces maps from subsets of X to subsets of Y and conversely, which are denoted by f and f^{-1} again (so you have to be careful to check the 'datatype' of the argument). Namely, if $A \subset X$, we set $f(A) = \{f(x) : x \in A\}$ (for example, the image of f is then $f(X)$), and for a subset $B \subset Y$, we set $f^{-1}(B) = \{x \in X : f(x) \in B\}$; this is called the *preimage* of B under f. Note that when f is bijective, there are two meanings of $f^{-1}(B)$ — one as just defined, and one as $g(B)$ where g is the inverse map f^{-1}. Fortunately, both meanings agree (exercise), and there is no danger of confusion.

Maps can be *composed:* if $f : X \to Y$ and $g : Y \to Z$, then we can define a map $X \to Z$ that sends $x \in X$ to $g(f(x)) \in Z$. This map is denoted by $g \circ f$ ("g after f") — keep in mind that it is f that is applied first!

Composition of maps is associative: if $f : X \to Y$, $g : Y \to Z$ and $h : Z \to W$, then $(h \circ g) \circ f = h \circ (g \circ f)$. Every set X has a special map, the *identity map* $\mathrm{id}_X : X \to X, x \mapsto x$. It acts as a neutral element under composition: for $f : X \to Y$, we have $f \circ \mathrm{id}_X = f = \mathrm{id}_Y \circ f$. If $f : X \to Y$ is bijective, then its inverse satisfies $f \circ f^{-1} = \mathrm{id}_Y$ and $f^{-1} \circ f = \mathrm{id}_X$.

If $f\colon X \to Y$ and $g\colon Y \to X$ are maps that satisfy $f \circ g = \mathrm{id}_Y$, then f is surjective and g is injective. If, moreover, $h\colon Y \to X$ is a map that satisfies $h \circ f = \mathrm{id}_X$, then f is also injective, and therefore bijective; we then have $g = h = f^{-1}$. We can apply this in particular to the case where f is invertible and g or h equals f^{-1}. We then obtain that if f is invertible, then any right or left inverse equals f^{-1}.

If $f\colon X \to X$ is a map from a set X to itself, then we often write f^2 instead of $f \circ f$. More generally, for every positive integer n we set

$$f^n = \underbrace{f \circ f \circ \cdots \circ f}_{n}.$$

When talking about several sets and maps between them, we often picture them in a *diagram* like the following.

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "g"'] & Y \arrow[d, "g'"] \\ U \arrow[r, "f'"] & V \end{tikzcd} \qquad \begin{tikzcd} X \arrow[d, "f"'] \arrow[dr, "h"] & \\ Y \arrow[r, "g"] & Z \end{tikzcd}$$

We call such a diagram *commutative* if all possible ways of going from one set to another lead to the same result. For the left diagram, this means that $g' \circ f = f' \circ g$, for the right diagram, this means that $h = g \circ f$.

Exercises

A.0.1. Let W, X, X_1, X_2, Y be sets.

(1) Show that there is a natural bijection
$$(Y^X)^W \to Y^{W \times X}.$$

(2) Denote the disjont union of X_1 and X_2 by $X_1 \coprod X_2$. Show that there is a natural bijection
$$Y^{X_1 \coprod X_2} \to Y^{X_1} \times Y^{X_2}.$$

(3) Assume that X and Y are finite. Show that we have
$$\#\left(Y^X\right) = (\#Y)^{(\#X)}.$$

APPENDIX B

Fields

B.1. Definition of fields

Definition B.1. A field is a set F, together with two distinguished elements $0, 1 \in F$ with $0 \neq 1$ and four maps

$$\begin{aligned} +\colon F \times F &\to F, \quad (x, y) \mapsto x + y \quad (\text{‘addition’}), \\ -\colon F \times F &\to F, \quad (x, y) \mapsto x - y \quad (\text{‘subtraction’}), \\ \cdot\colon F \times F &\to F, \quad (x, y) \mapsto x \cdot y \quad (\text{‘multiplication’}), \\ /\colon F \times (F \setminus \{0\}) &\to F, \quad (x, y) \mapsto x/y \quad (\text{‘division’}), \end{aligned}$$

such that, for all $x, y, z \in F$, the addition and multiplication satisfy

$$\begin{aligned} x + y &= y + x, & x + (y + z) &= (x + y) + z, & x + 0 &= x, \\ x \cdot y &= y \cdot x, & x \cdot (y \cdot z) &= (x \cdot y) \cdot z, & x \cdot 1 &= x, \end{aligned}$$
$$x \cdot (y + z) = (x \cdot y) + (x \cdot z),$$

the subtraction is related to the addition through

$$x + y = z \Leftrightarrow x = z - y,$$

and, if $y \neq 0$, the division is related to the multiplication through

$$x \cdot y = z \Leftrightarrow x = z/y.$$

Example B.2. The set $\mathbb{R}$ of real numbers, together with its 0 and 1 and the ordinary addition, subtraction, multiplication, and division, obviously form a field.

Example B.3. Also the field $\mathbb{Q}$ of rational numbers, together with its 0 and 1 and the ordinary addition, subtraction, multiplication, and division, form a field.

Example B.4. Consider the subset

$$\mathbb{Q}(\sqrt{2}) = \{\, a + b\sqrt{2} \;:\; a, b \in \mathbb{Q} \,\}$$

of $\mathbb{R}$, which contains 0 and 1. The ordinary addition, subtraction, and multiplication of $\mathbb{R}$ clearly give addition, subtraction, and multiplication on $\mathbb{Q}(\sqrt{2})$, as we have

$$\begin{aligned} (a + b\sqrt{2}) \pm (c + d\sqrt{2}) &= (a \pm c) + (b \pm d)\sqrt{2}, \\ (a + b\sqrt{2}) \cdot (c + d\sqrt{2}) &= (ac + 2bd) + (ad + bc)\sqrt{2}. \end{aligned}$$

To see that for any $x, y \in \mathbb{Q}(\sqrt{2})$ with $y \neq 0$ we also have $x/y \in \mathbb{Q}(\sqrt{2})$, we first note that if c and d are integers with $c^2 = 2d^2$, then $c = d = 0$, as otherwise c^2 would have an even and $2d^2$ an odd number of factors 2. Now for any $x, y \in \mathbb{Q}(\sqrt{2})$ with $y \neq 0$, we can write x/y as

$$\frac{a + b\sqrt{2}}{c + d\sqrt{2}}$$

with integers a, b, c, d, where c and d are not both 0; we find

$$\begin{aligned}\frac{x}{y} = \frac{a + b\sqrt{2}}{c + d\sqrt{2}} &= \frac{(a + b\sqrt{2}) \cdot (c - d\sqrt{2})}{(c + d\sqrt{2}) \cdot (c - d\sqrt{2})} = \frac{(ac - 2bd) + (bc - ad)\sqrt{2}}{c^2 - 2d^2} \\ &= \frac{ac - 2bd}{c^2 - 2d^2} + \frac{bc - ad}{c^2 - 2d^2}\sqrt{2} \in \mathbb{Q}(\sqrt{2}).\end{aligned}$$

We conclude that we also have division by nonzero elements on $\mathbb{Q}(\sqrt{2})$. Since the requirements of Definition B.1 are fulfilled for all real numbers, they are certainly fulfilled for all elements in $\mathbb{Q}(\sqrt{2})$ and we conclude that $\mathbb{Q}(\sqrt{2})$ is a field.

In any field with elements x and y, we write $-x$ for $0 - x$ and y^{-1} for $1/y$ if y is nonzero; we also often write xy for $x \cdot y$. The rules of Definition B.1 require that many of the properties of the ordinary addition, subtraction, multiplication, and division hold in any field. The following proposition shows that automatically many other properties hold as well.

Proposition B.5. *Suppose F is a field with elements $x, y, z \in F$.*

(1) *We have $x + z = y + z$ if and only if $x = y$.*
(2) *If z is nonzero, then $xz = yz$ if and only if $x = y$.*
(3) *If $x + z = z$, then $x = 0$.*
(4) *If $xz = z$ and $z \neq 0$, then $x = 1$.*
(5) *We have $0 \cdot x = 0$ and $(-1) \cdot x = -x$ and $(-1) \cdot (-1) = 1$.*
(6) *If $xy = 0$, then $x = 0$ or $y = 0$.*

Proof. Exercise. □

Example B.6. The smallest field $\mathbb{F}_2 = \{0, 1\}$ has no more than the two required elements, with the only 'interesting' definitions being that $1 + 1 = 0$ and $0 - 1 = -1 = 1$. One easily checks that all requirements of Definition B.1 are satisfied.

Warning B.7. Many properties of sums and products that you are used to from the real numbers hold for general fields. There is one important exception: in general there is no ordering and it makes no sense to call an element positive or negative, or bigger than an other element. The fact that this is possible for $\mathbb{R}$ and for fields contained in $\mathbb{R}$, means that these fields have more structure than general fields. In Chapter 1 this extra structure is used to our advantage.

Exercises

B.1.1. Prove Proposition B.5.

B.1.2. Check that $\mathbb{F}_2$ is a field (see Example B.6).

B.1.3. Which of the following are fields?

(1) The set $\mathbb{N}$ together with the usual addition, multiplication, subtraction, division, 0, and 1.
(2) The set $\mathbb{Z}$ together with the usual operations, and the usual 0 and 1.

(3) The set $\mathbb{Q}$ together with the usual operations, and the usual 0 and 1.
(4) The set $\mathbb{R}_{\geq 0}$ together with the usual operations, and the usual 0 and 1.
(5) The set $\mathbb{Q}(\sqrt{3}) = \{a+b\sqrt{3} \ : \ a, b \in \mathbb{Q}\}$ together with the usual operations, and the usual 0 and 1.

B.1.4. Suppose F is a field. Show that the 0, 1, the subtraction, and the division are completely determined by the addition and the multiplication and the fact that F is a field. In other words, once you know the addition and multiplication on a set F, there is no choice anymore for the elements 0 and 1, and the subtraction and division, if you want to make F into a field.

B.1.5. Consider the set $\mathbb{F}_3 = \{0, 1, 2\}$ with the usual addition, subtraction, and multiplication, but where each is followed by taking the remainder after division by 3. Is there a division that makes $\mathbb{F}_3$ into a field?

B.2. The field of complex numbers.

The first motivation for the introduction of complex numbers is a shortcoming of the real numbers: while positive real numbers have real square roots, negative real numbers do not. Since it is frequently desirable to be able to work with solutions to equations like $x^2 + 1 = 0$, we introduce a new number, called i, that has the property $i^2 = -1$. The set $\mathbb{C}$ of *complex numbers* then consists of all expressions $a + bi$, where a and b are real numbers. If $z = a + bi$, then we call $\operatorname{Re} z = a$ the *real part* and $\operatorname{Im} z = b$ the *imaginary part* of z. (More formally, one considers pairs of real numbers (a, b) and so identifies $\mathbb{C}$ with $\mathbb{R}^2$ as sets.) In order to turn $\mathbb{C}$ into a field, we have to define addition, multiplication, subtraction, and division.

If we want the multiplication to be compatible with the scalar multiplication on $\mathbb{R}^2$, then (bearing in mind the field axioms) there is no choice: we have to set

$$(a + bi) \pm (c + di) = (a \pm c) + (b \pm d)i$$

and

$$(a + bi)(c + di) = ac + adi + bci + bdi^2 = (ac - bd) + (ad + bc)i$$

(remember $i^2 = -1$). It is then an easy, but tedious, matter to show that the axioms of Definition B.1 regarding the addition, subtraction, and multiplication hold. (The theory of rings and fields in later courses provides a rather elegant way of doing this.)

We still need to show there is also a division, or, equivalently, we need to show the existence of multiplicative inverses. In this context, it is advantageous to introduce the notion of *conjugate complex number*.

Definition B.8. If $z = a + bi \in \mathbb{C}$, then the *complex conjugate* of z is $\bar{z} = a - bi$. Note that $z\,\bar{z} = a^2 + b^2$ is real and satisfies $z\bar{z} \geq 0$. We set $|z| = \sqrt{z\bar{z}}$; this is called the *absolute value* or *modulus* of z. It is clear that $|z| = 0$ only for $z = 0$; otherwise $|z| > 0$. We obviously have $\bar{\bar{z}} = z$ and $|\bar{z}| = |z|$.

Proposition B.9.

(1) *For all $w, z \in \mathbb{C}$, we have $\overline{w + z} = \bar{w} + \bar{z}$ and $\overline{wz} = \bar{w}\,\bar{z}$.*
(2) *For all $z \in \mathbb{C} \setminus \{0\}$, the element $z' = |z|^{-2} \cdot \bar{z}$ satisfies $z' \cdot z = 1$.*
(3) *For all $w, z \in \mathbb{C}$, we have $|wz| = |w| \cdot |z|$.*

Proof. (1) Exercise.

(2) First of all, $|z| \neq 0$, so the expression makes sense. Now note that
$$z' \cdot z = |z|^{-2}\bar{z} \cdot z = |z|^{-2} \cdot z\bar{z} = |z|^{-2}|z|^2 = 1\,.$$

(3) Exercise.

□

By Proposition B.9(2), the division on $\mathbb{C}$ has to satisfy $1/z = |z|^{-2} \cdot \bar{z}$, and therefore
$$\frac{y}{z} = y \cdot \frac{1}{z} = \frac{y\bar{z}}{|z|^2}$$
for all $y, z \in \mathbb{C}$ with $z \neq 0$. For example:
$$\frac{1}{1+2i} = \frac{1-2i}{(1+2i)(1-2i)} = \frac{1-2i}{1^2+2^2} = \frac{1-2i}{5} = \frac{1}{5} - \frac{2}{5}i\,.$$
In general, we get
$$\frac{a+bi}{c+di} = \frac{(a+bi)(c-di)}{(c+di)(c-di)} = \frac{ac+bd}{c^2+d^2} + \frac{bc-ad}{c^2+d^2} \cdot i,$$
for $a, b, c, d \in \mathbb{R}$ with c and d not both 0.

Remark B.10. Historically, the necessity of introducing complex numbers was realized through the study of *cubic* (and not quadratic) equations. The reason for this is that there is a solution formula for cubic equations that in some cases requires complex numbers in order to express a real solution. See Section 2.7 in Jänich's book [**J**].

The importance of the field of complex numbers lies in the fact that they provide solutions to *all* polynomial equations. This is the 'Fundamental Theorem of Algebra':

Every non-constant polynomial with complex coefficients has a root in $\mathbb{C}$.

Unfortunately, a proof is beyond the scope of this course.

Exercises

B.2.1. Prove Remark B.9.

B.2.2. Show that for every complex number z we have
$$\mathrm{Re}(z) = \tfrac{1}{2}(z + \overline{z}) \qquad \text{and} \qquad \mathrm{Im}(z) = \tfrac{1}{2i}(z - \overline{z}).$$

APPENDIX C

Labeled collections

Let X be a set and $n \geq 0$ an integer. An *n-tuple* is an ordered sequence of n objects. The Cartesian product X^n consists of all *n-tuples* or *sequences* of n elements in X. A sequence

$$(x_1, x_2, \ldots, x_n) \in X^n$$

could also be written as

$$(x_i)_{i=1}^n = (x_i)_{1 \leq i \leq n} = (x_i)_{i \in \{1,\ldots,n\}}.$$

This sequence consists of one element in X for each $i \in \{1, \ldots, n\}$, so it can be identified with the function $\{1, \ldots, n\} \to X$ that sends i to x_i. Indeed, this was made precise in Proposition A.1. Recall that for any set I, we denote the set of all maps from I to X by $\mathrm{Map}(I, X)$ or X^I. Proposition A.1 states that for $I = \{1, 2, \ldots, n\}$, the map

$$T\colon X^I \to X^n$$

that sends a function f to the sequence $(f(1), f(2), \ldots, f(n))$ is a bijection.

We can think of an n-tuple as a collection of elements labeled by the set $\{1, 2, \ldots, n\}$. For each $i \in \{1, 2, \ldots, n\}$ we have chosen an element in X, namely x_i. Motivated by this viewpoint and Proposition A.1, we give the following definition.

Definition C.1. Let I be a set. A *(labeled) collection*, labeled by I, of elements in X is a map $f\colon I \to X$. We also write the collection as $(x_i)_{i \in I}$ with $x_i = f(i)$.

As for the sequences that we started with, a collection consists of one element x_i in X for each $i \in I$. The elements of I are called the *indices* or *labels*.

Example C.2. For $I = \{1, 2, \ldots, n\}$ we recover n-tuples: finite sequences of length n.

Example C.3. For $I = \mathbb{Z}_{\geq 0}$ we obtain infinite sequences $(x_0, x_1, x_2, x_3, \ldots)$. Cf. Remark 2.12.

Example C.4. For $I = \mathbb{Z}$ we obtain doubly infinite sequences

$$\ldots, x_{-2}, x_{-1}, x_0, x_1, x_2, x_3, \ldots.$$

Example C.5. Let $V = \mathrm{Map}(\mathbb{R}, \mathbb{R})$ be the vector space of all functions from $\mathbb{R}$ to $\mathbb{R}$. Let $\mathcal{I}$ denote the set of all closed intervals $[a, b] \subset \mathbb{R}$. For each interval $I \in \mathcal{I}$, we let h_I denote the function given by

$$h_I(x) = \begin{cases} 1 & \text{if } x \in I, \\ 0 & \text{if } x \notin I. \end{cases}$$

This yields a collection $\big(h_I\big)_{I \in \mathcal{I}}$ of functions from $\mathbb{R}$ to $\mathbb{R}$, labeled by the set $\mathcal{I}$ of all bounded closed intervals (See Exercise 7.1.7).

Suppose $(x_i)_{i \in I}$ is a collection of elements in X, corresponding to the function $f \colon I \to X$. Then its set of elements is

$$\{x_i \,:\, i \in I\} = \operatorname{im} f \subset X.$$

As opposed to its set of elements, the collection $(x_i)_{i \in I}$ may have repetitions, that is, we may have $x_i = x_j$ for $i, j \in I$ with $i \neq j$.

Conversely, we can associate to any subset $S \subset X$ a natural collection that is labeled by the set S itself. The map $S \to X$ that this collection corresponds to is the inclusion and it can be written as $(s)_{s \in S}$.

In other pieces of literature, (labeled) collections are sometimes called *labeled sets*. Given that they are *not* sets (they may contain some elements more than once), we will refrain from using this terminology.

APPENDIX D

Polynomials

Many books that define *polynomials* state something along the lines of them being

> finite sums of terms, each term being a product of a coefficient and a monomial, where monomials are products of nonnegative integral powers of the variables.

But without making precise what a *variable* is, or which addition this sum refers to, this is not a satisfying definition. Authors that recognise this may attempt to make it sound better by calling the variables *indeterminates*, which are supposed to be just symbols or placeholders in some sense, and by calling the sum a *formal sum*, which does not make it more clear at all, unless perhaps one defines this notion of formal sum in terms of the more advanced notion of *external direct sums.*

In an attempt to comfort the reader, such books give some examples, such as

$$x^2+1, \qquad x^7-13x+4, \qquad x^2y^2z^2-xy^2+2x^2y+17z^4,$$

where the first two are polynomials in the one variable x, and the last one is a polynomial in the three variables x, y, z. If F is a field, then the obvious addition and scalar multiplication make the set of polynomials with coefficients in F a vector space over F. There is also a multiplication of polynomials, which works just as you expect, for example

$$(x-y)(x^4+x^3y+x^2y^2+xy^3+y^4)=x^5-y^5.$$

While these examples may indeed seem like polynomials are simple enough objects, we do want to give a better definition. For simplicity, we restrict ourselves to polynomials in one variable. Each can be identified with the sequence of (all infinitely many of) its coefficients, so $x^7-13x+4$ is identified with the sequence

$$(4,-13,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,\ldots).$$

In the next section, we will define polynomials in terms of such sequences.

D.1. Polynomials in one variable

We assume that the reader has read Chapter 2 at least up to Remark 2.12.

Let F be a field, and let F^∞ be the vector space of all infinite sequences of elements in F as in Example 2.6. As in that example, we will use indices starting at 0, so the component with index n of the vector $(a_n)_{n\geq 0}=(a_0,a_1,a_2,\ldots)$ is a_n. Besides the addition and scalar multiplication that come with the vector space F^∞, we also define a multiplication between two vectors: for $f=(f_0,f_1,\ldots)$ and $g=(g_0,g_1,\ldots)$ in F^∞ we let the product $f\cdot g\in F^\infty$ be the vector whose component with index $n\geq 0$, is

$$\sum_{k=0}^{n} f_k g_{n-k}. \tag{D.1}$$

For the first terms, we obtain

$$f \cdot g = \big(f_0 g_0, f_0 g_1 + f_1 g_0, f_0 g_2 + f_1 g_1 + f_2 g_0, \dots\big).$$

Note that we can also write (D.1) as

$$\sum_{\substack{0 \le k, \ell \le n \\ k+\ell=n}} f_k g_\ell.$$

This shows immediately that for f, g as above, we have

$$f \cdot g = g \cdot f; \tag{D.2}$$

it also makes it easy to verify that, for $h = (h_0, h_1, \dots) \in F^\infty$, we have

$$(f \cdot g) \cdot h = f \cdot (g \cdot h), \tag{D.3}$$

as the component with index n of both products equals

$$\sum_{\substack{0 \le k, \ell, m \le n \\ k+\ell+m=n}} f_k g_\ell h_m.$$

So we don't need to write parentheses in products of more than two vectors. One also quickly checks that for $f, g, h \in F^\infty$ we have

$$f \cdot (g + h) = f \cdot g + f \cdot h, \tag{D.4}$$

and for all $\lambda \in F$ we have

$$(\lambda f) \cdot g = \lambda (f \cdot g). \tag{D.5}$$

Furthermore, for all integers $n \ge 0$, let $e_n \in F^\infty$ be the sequence consisting of only zeroes, except for a 1 at index n. Then for all integers $m, n \ge 0$ we have

$$e_m \cdot e_n = e_{m+n}, \tag{D.6}$$

and for all $f \in F^\infty$ we have

$$e_0 \cdot f = f. \tag{D.7}$$

We denote the element e_1 by x. By (D.6) we have $x^n = e_n$ for all $n \ge 1$, and we set $x^0 = e_0$. By abuse of notation, we will identify F with its image under the map

$$\iota \colon F \to F^\infty, \quad \lambda \mapsto \lambda e_0 = (\lambda, 0, 0, 0, 0, 0, , \dots).$$

This means that for any $\lambda \in F$, we may denote λe_0 by λ, which for $\lambda = 1$ yields the usual identity $x^0 = 1$. Any critical reader should now worry, as for any $\lambda \in F$ and $f \in F^\infty$, the notation $\lambda \cdot f$ might refer to the scalar multiplication of the scalar λ and the vector f, but it might also refer to the new multiplication of the vectors λe_0 and f. Fortunately, the two coincide by (D.5) and (D.7), so this will not lead to any confusion.

Definition D.1. A *monomial* is a power of x with a nonnegative integral exponent. A *polynomial* over F is a finite sum of terms, where each term is the product of a scalar in F and a monomial. We let $F[x] \subset F^\infty$ be the set of all polynomials.

A *real polynomial* is a polynomial over the field $\mathbb{R}$. A *complex polynomial* is a polynomial over the field $\mathbb{C}$.

Note that for coefficients $a_0, a_1, a_2, \dots, a_d \in F$ we have

$$\sum_{n=0}^{d} a_n x^n = (a_0, a_1, a_2, \dots, a_d, 0, 0, 0, 0, 0, \dots). \tag{D.8}$$

Proposition D.2. *An element of F^∞ is a polynomial if and only if all but finitely many of its coefficients are zero. The set $F[x]$ is a vector space over F.*

Proof. Every polynomial can be written as the left-hand side of (D.8), possibly by adding terms with coefficient zero. For every vector $a = (a_n)_{n\geq 0} \in F^\infty$ of which all but finitely many coefficients are zero, there is a d such that for all $n > d$ we have $a_n = 0$, that is, a can be written as the right-hand side of (D.8). Therefore, both directions of the first statement follow from (D.8). The second statement is a good exercise for any reader who has just read Chapter 2, or at least Section 2.1, and preferably some examples from Section 2.2. The theory of Section 3.1 may be used to significantly reduce the amount of work. □

By Proposition D.2, the following notions are well defined.

Definition D.3. Let $f = (f_n)_{n\geq 0}$ be a polynomial. The *constant coefficient* of f is f_0. If f is nonzero, then the degree $\deg(f)$ of f is the largest index n for which f_n is nonzero; the *main coefficient* of f is then f_d with $d = \deg(f)$. The degree of the zero polynomial is $-\infty$.

From (D.8) it follows immediately that for every nonzero polynomial $f \in F[x]$ there is a unique integer $d \geq 0$ and there are unique coefficients $a_0, a_1, \ldots, a_d \in F$ with $a_d \neq 0$ such that $f = \sum_{n=0}^{d} a_n x^n$; this integer d is the degree of f and the coefficients $a_0, a_1, \ldots, a_d$ are just the first $d+1$ coefficients of the vector $f \in F^\infty$. Also, a polynomial $f = \sum_{n=0}^{d} a_n x^n \in F[x]$ is zero if and only if for all n with $0 \leq n \leq d$ we have $a_n = 0$.

Warning D.4. (††) The polynomials x and x^2 in $\mathbb{F}_2[x]$ are different; one has degree 1 and the other degree 2. However, by substituting elements of $\mathbb{F}_2$ for x, the two polynomials induce the same function $\mathbb{F}_2 \to \mathbb{F}_2$ as we have $\alpha = \alpha^2$ for all $\alpha \in \mathbb{F}_2$.

Because for some fields we can not use analysis, we define the derivative of a polynomial directly in terms of its coefficients.

Definition D.5. Let $f = \sum_{n=0}^{d} a_n x^n$ be a polynomial. Then the *derivative* of f is the polynomial $\sum_{n=1}^{d} n a_n x^{n-1}$.

Remark D.6. In terms of Chapter 7, the collection $(x^n)_{n\geq 0} = (1, x, x^2, x^3, \ldots)$ is a basis for $F[x]$. Indeed, if we set $I = \mathbb{Z}_{\geq 0}$, then under the isomorphism

$$\chi\colon F^I \to F^\infty$$

that sends $\varphi\colon I \to F$ to $(\varphi(0), \varphi(1), \varphi(2), \ldots)$ (cf. Remark 2.12 and Example C.3), the subspace $F[x] \subset F^\infty$ corresponds to $F^{(I)}$, and the collection $(x^n)_{n\geq 0} = (e_n)_{n\in I}$ corresponds to the basis for $F^{(I)}$ given in part (2) of Exercise 7.2.5.

Under the same isomorphism χ, the new multiplication on F^∞ corresponds with a new multiplication on F^I, which can be expressed, analogously to (D.1), by defining the product of $\varphi \in F^I$ and $\psi \in F^I$ to be the function $\varphi \cdot \psi\colon I \to F$

given by

$$(\varphi \cdot \psi)(n) = \sum_{k=0}^{n} \varphi(k)\psi(n-k).$$

Remark D.7. The properties (D.2), (D.3), (D.4), (D.7), with the first four axioms for a vector space, imply that F^∞, together with its addition and multiplication, carries the structure of what is called a *commutative ring*, with unit element e_0. In fact, together with the embedding $\iota\colon F \to F^\infty$ it is a so-called *F-algebra.*

Exercises

D.1.1. Define a polynomial ring in two variables.

D.2. Polynomial functions

We assume the reader has read up to Section 3.4. We generalise Example 3.35 to arbitrary fields.

Let F be a field. We consider the power functions $p_n\colon x \mapsto x^n$ inside the vector space F^F of all functions from F to F. Their linear hull $L(\{p_n : n \in \mathbb{Z}_{\geq 0}\}) \subset F^F$ is the linear subspace of *polynomial functions* from F to F, i.e, functions that are of the form

$$x \longmapsto a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

with $n \in \mathbb{Z}_{\geq 0}$ and $a_0, a_1, \ldots, a_n \in F$. By definition, the power functions p_n generate the subspace of polynomial functions, which we denote by $P(F)$.

Warning D.8. In Example 2.13 we defined *polynomials* as formal sums of powers x^i, each multiplied by a scalar, and in the previous section we gave a more precise definition. These are not to be confused with the *polynomial functions* $f\colon F \to F$, even though over subfields of $\mathbb{C}$ there will not be much confusion (see Remark 3.36). In fact, the same is true over infinite fields (cf. Exercise 11.3.8).

(††) As stated in Warning D.4, though, over finite fields the difference between polynomials, as defined in Example 2.13, and polynomial functions, as defined in Example 3.35, is clear, as there may be many more polynomials than polynomial functions. For instance, the polynomial $x^2 + x$ and the zero polynomial 0, both with coefficients in the field $\mathbb{F}_2$, are different **polynomials**: the first has degree 2, the second degree $-\infty$. However, the **polynomial function** $\mathbb{F}_2 \to \mathbb{F}_2$ that sends x to $x^2 + x$ is the same as the zero function.

Warning D.9. †† Although the vector space of *real polynomial functions* is infinite-dimensional, over finite fields this is not the case (see Exercise 7.4.6). The vector space $F[x]$ of *polynomials*, however, is infinite-dimensional for any field F (see Example 7.44 and Warning D.8).

Exercises

D.2.1. We expand on Remark 3.36 over a general field F. Let $F[x]$ be the vector space of polynomials over F. Consider the map $\varphi\colon F[x] \to F^F$ that sends a *polynomial* $f = \sum_{i=0}^{d} c_i x^i$ to the *function* that sends an element $a \in F$ to the evaluation of f at a, that is, to $f(a) := \sum_{i=0}^{d} c_i a^i$, cf. Warning D.4.

(1) Show that the image of φ is exactly the subspace of F^F consisting of *polynomial functions*.
(2) (††) Is φ injective for $F = \mathbb{F}_2$?
(3) Is there a field F for which φ is injective?

[Remark: By abuse of notation, the function $\varphi(f)$ is often also denoted by f.]

D.3. Polynomials in more variables

Let F be a field. Let S be a set. We call the elements of S *variables*. A *monomial* in S is a function $m\colon S \to \mathbb{Z}_{\geq 0}$ that is zero at all but finitely many elements of S. We write the evaluation of a monomial m at a variable v with square brackets; we call the image $m[v]$ the *exponent* of v in m.

Let $M(S)$ denote the set of all monomials in S. By abuse of notation, we identify S with its image under the map $S \to M(S)$ that sends each $v \in S$ to its associated indicator function; that is, if v denotes an element in S, then v also denotes the monomial in which the variable v has exponent 1 and every other variable has exponent 0.

The addition on $\mathbb{Z}_{\geq 0}$ induces an addition on $M(S)$, which we will write multiplicatively; we also write 1 for the trivial monomial that sends every $w \in S$ to 0. In particular, for every $v \in S$ and every non-negative integer n, the monomial in which v has exponent n and every other variable $w \in S \setminus \{v\}$ has exponent 0, is written as v^n.

It is easy to verify that if $m \in M(S)$ is a monomial, then we have

$$m = \prod_{v \in S} v^{m[v]},$$

where this a priori potentially infinite product should be taken only over all finitely many $v \in S$ with $m[v] \neq 0$; the factors that we leave out are all equal to 1.

A *polynomial* in S over F is a function $f\colon M(S) \to F$ that is zero at all but finitely many monomials. We write the evaluation of a polynomial f at a monomial m with an index; we call the image f_m the *coefficient* of m in f. Hence we can write f as a labelled collection

$$f = \big(f_m\big)_{m \in M(S)}.$$

Let $F[S]$ denote the set of all polynomials in S over F; this is exactly the vector space $F^{(M(S))}$ (see Exercise 3.1.9). By abuse of notation, we identify $M(S)$ with its image under the map $M(S) \to F[S]$ that sends a monomial m to its associated indicator function; that is, if m denotes a monomial, then m also denotes the polynomial in which the coefficient of the monomial m is 1 and the coefficient of every other monomial is 0.

In this way, the monomials form a basis for the vector space $F[S]$ (see Exercise 7.2.5), and we have

$$f = \sum_{m \in M(S)} f_m \cdot m.$$

In other words, the coefficients of the monomials in f are also the coefficients of f with respect to the basis formed by the monomials. A polynomial f is zero if and only if for all monomials m the coefficient of m in f is zero.

We define the *total degree* of a monomial m by

$$\deg(m) = \sum_{v \in S} m[v],$$

which is well defined as this sum has only finitely many nonzero terms. We define the total degree of a nonzero polynomial f to be the maximum of the total degrees of all monomials that have a nonzero coefficient in f, and the total degree of the zero polynomial to be $-\infty$.

We define a multiplication on the vector space $F[S]$ as follows. For two polynomials $f, g \in F[S]$ we set

$$f \cdot g = \left(\sum_{\substack{m_1, m_2 \in M(S) \\ m_1 m_2 = m}} f_{m_1} g_{m_2} \right)_{m \in M(S)}$$

Note that this multiplication restricts to the natural multiplication on monomials. Also, since the multiplication on monomials is commutative, it follows that this multiplication on polynomials is commutative as well: for two polynomials $f, g \in F[S]$ we have $f \cdot g = g \cdot f$.

The multiplication on monomials is also associative, so the multiplication on polynomials is associative as well, as for three polynomials $f, g, h \in F[S]$ we have

$$(f \cdot g) \cdot h = \left(\sum_{\substack{m_1, m_2, m_3 \in M(S) \\ m_1 m_2 m_3 = m}} f_{m_1} g_{m_2} h_{m_3} \right)_{m \in M(S)} = f \cdot (g \cdot h).$$

It is also easy to check that for $f, g, h \in F[S]$ we have $f \cdot (g + h) = f \cdot g + f \cdot h$, that is, the multiplication is distributive over the addition, and for all $\lambda \in F$ we have $(\lambda f) \cdot g = \lambda (f \cdot g)$. Together with commutativity this implies that the multiplication is *bilinear*: it is linear in both arguments. In fact, bilinearity, together with the fact that the multiplication on polynomials restricts to the natural multiplication on monomials, completely determines the multiplication on polynomials.

By abuse of notation, we identify F with its image under the linear map $\iota \colon F \to F[S]$ that sends the scalar λ to $\lambda \cdot 1$, where 1 stands for the trivial monomial, viewed as polynomial. We call the elements in the image of ι *constant polynomials*. Since the polynomial 1 acts as a unit for the multiplication on polynomials, we find that, for all $\lambda \in F$, multiplication by the constant polynomial $\lambda \cdot 1$ coincides with scalar multiplication by λ, so identifying λ with its associated constant polynomial will not cause any confusion.

Remark D.10. The properties above imply that $F[S]$, together with its addition and multiplication, and the special polynomials 0 and 1, carries the structure of a so-called *commutative ring*. In fact, the map ι gives this ring $F[S]$ the structure of a so-called F-*algebra*.

We now define the evaluation of a polynomial at a collection labelled by S. Let $a = (a_v)_{v \in S}$ be a collection of elements in F labelled by S. Then for every monomial $m = \prod_{v \in S} v^{m[v]}$ we define

$$m(a) = \prod_{v \in S} a_v^{m[v]}.$$

For any polynomial $f = \sum_{m \in M(S)} f_m \cdot m$ we define

$$f(a) = \sum_{m \in M(S)} f_m \cdot m(a).$$

Remark D.11. In fact, for any commutative F-algebra A, we can similarly define the evaluation of a polynomial $f \in F[S]$ at a collection of elements in A labelled by S. Recall that we have identified S with a subset of the commutative F-algebra $F[S]$ in the sense that each variable can be viewed as a polynomial. If we label the set $S \subset$ by itself, then we obtain $f(S) = f$.

Remark D.12. We can now explain why we chose non-standard notations for the evaluation of a monomial m at a variable v, yielding the exponent of v in m, and for the evaluation of a polynomial f at a monomial m, yielding the coefficient of m in f. Suppose $S = \{x\}$, and we have a monomial $m = x^2$ and variable $v = x$. Then

(1) we have $m[v] = 2$, the exponent of v in m;
(2) we have $m_v = 0$, the coefficient of the monomial $v = x$ in the polynomial $m = x^2$;
(3) we have $m(v) = x^2$, the evaluation of the polynomial/monomial $m = x^2$ at the element $x \in F[S]$.

Remark D.13. We now compare polynomials in more variables with polynomials in one variable. If S consists of one element, say $S = \{x\}$, then $M(S)$ consists of the non-negative integral powers of x, that is, $M(S) = \{1, x, x^2, x^3, \ldots\}$. There is a natural bijection with $\mathbb{Z}_{\geq 0}$, with $n \in \mathbb{Z}_{\geq 0}$ corresponding to x^n. Through this bijection, we can identify a polynomial $f \in F[S]$, which is a collection of elements in F labelled by $M(S)$, with a collection of elements in F labelled by $\mathbb{Z}_{\geq 0}$. Such a collection corresponds naturally with an element of F^∞ (cf. Remark 2.12 and Example C.3), which actually lies in the subspace $F[x] \subset F^\infty$.

Exercises

D.3.1. Define the derivative of a polynomial with respect to one of its variables.

APPENDIX E

Infinite-dimensional vector spaces and Zorn's Lemma

We have seen that a finitely generated vector space V has a finite basis and so has finite dimension. What can we say about the existence of a basis in an *infinite-dimensional* vector space?

We have seen examples of an infinite-dimensional vector space that have a basis. For example, the space $F[t]$ of polynomials over the field F (Example 7.49), or the space of polynomial functions on $\mathbb{R}$ (Example 7.50), with basis given by the monomials $x \mapsto x^n$, for $n \in \mathbb{Z}_{\geq 0}$.

On the other hand, you would be very hard put to write down a basis for $\mathcal{C}(\mathbb{R})$, or a basis for $\mathbb{R}$ as a $\mathbb{Q}$-vector space.

In order to prove the existence of a basis and other related results, we would need an 'infinite' version of the Basis Extension Theorem.

Theorem E.1 (General Basis Extension Theorem). *Let V be a vector space, J an index set, and $(v_j)_{j\in J}$ a collection of elements that generates V. Suppose that for a subset $I \subset J$, the subcollection $(v_i)_{i\in I}$ is linearly independent. Then there is a subset $I' \subset J$ with $I \subset I'$ such that the collection $(v_i)_{i\in I'}$ is a basis for V.*

Now, how can we prove such a statement? One idea, which also works for the finite-dimensional case, would be to choose a maximal subset $I' \subset J$ containing I for which the collection $(v_i)_{i\in I'}$ is linearly independent, and then show that this collection also generates V and is therefore a basis.

This last step will work fine: assume that I' is maximal as above, then for every $j \in J \setminus I'$, the collection $(v_i)_{i\in I''}$ with $I'' = I' \cup \{j\}$ is linearly dependent, and so $v_j \in L\big((v_i)_{i\in I'}\big)$. This implies that

$$V = L\big((v_i)_{i\in J}\big) \subset L\big((v_i)_{i\in I'}\big),$$

so $(v_i)_{i\in I'}$ generates V and is therefore a basis.

However, the key point is the *existence* of a maximal set I' with the required property. Note that if $\mathcal{S}$ is an arbitrary set of subsets of some set, $\mathcal{S}$ need not necessarily have maximal elements. For example, $\mathcal{S}$ could be empty. Or consider the set of all finite subsets of $\mathbb{Z}$. So we need some extra condition to ensure the existence of maximal elements. (Of course, when $\mathcal{S}$ is finite (and nonempty), then there is no problem — we can just take a set of maximal size.)

This condition is formulated in terms of *chains*.

Definition E.2. Let X be a set, and let $\mathcal{S}$ be a set of subsets of X. A subset $\mathcal{C} \subset \mathcal{S}$ is called a *chain* if all elements of $\mathcal{C}$ are comparable, i.e., if for all $U, V \in \mathcal{C}$, we have $U \subset V$ or $V \subset U$. (Note that this is trivially true when $\mathcal{C}$ is empty.)

The notion of 'chain' (as well as Zorn's Lemma below) applies more generally to (partially) ordered sets: a chain then is a subset that is totally ordered.

Now a statement of the kind we need is the following.

Zorn's Lemma. *Let X be a set, and let $\mathcal{S}$ be a collection of subsets of X. If for every chain $\mathcal{C} \subset \mathcal{S}$, there is a set $U \in \mathcal{S}$ such that $Z \subset U$ for all $Z \in \mathcal{C}$, then $\mathcal{S}$ has a maximal element.*

Note that the condition, when applied to the empty chain, ensures that $\mathcal{S} \neq \emptyset$. Also note that there can be more than one maximal element in $\mathcal{S}$.

Let us see how we can apply this result to our situation. The set $\mathcal{S}$ we want to consider is the set of all subsets $I' \subset J$ containing I such that $(v_i)_{i \in I'}$ is linearly independent. We have to verify the assumption on chains. So let $\mathcal{C} \subset \mathcal{S}$ be a chain. We have to exhibit a set $U \in \mathcal{S}$ containing all the elements of $\mathcal{C}$. In such a situation, our first guess is to try $U = \bigcup \mathcal{C}$ (the union of all sets in $\mathcal{C}$); usually it works. In our case, we have to show that this $U \subset J$ has the property that $(v_i)_{i \in U}$ is linearly independent. Assume it is not. Then there is a *finite* non-trivial linear combination of elements of $(v_i)_{i \in U}$ that gives the zero vector. This linear combination will only involve finitely many elements of U, which come from finitely many sets $I' \in \mathcal{C}$. Since $\mathcal{C}$ is a chain, there is a maximal set I^* among these finitely many, and our nontrivial linear combination only involves elements from $(v_i)_{i \in I^*}$. But I^* is in $\mathcal{S}$, and so $(v_i)_{i \in I^*}$ is linearly independent, a contradiction. Therefore our assumption must be false, and $(v_i)_{i \in U}$ must be linearly independent.

Hence, Zorn's Lemma implies that our set $\mathcal{S}$ contains a maximal element, which by the discussion before Definition E.2 implies the general Basis Extension Theorem. In particular, this shows that every vector space must have a basis (take $I = \emptyset$ and $J = V$ and $(v_j)_{j \in J} = (v)_{v \in V}$). However, Zorn's Lemma is an extremely inconstructive result; it does not give us any information on how to find a maximal element. And in fact, nobody has ever been able to 'write down' (or explicitly construct) a $\mathbb{Q}$-basis of $\mathbb{R}$, say. Still, such bases must exist.

The next question then is, how does one prove Zorn's Lemma? It turns out that it is equivalent (given the more 'harmless' axioms of set theory) to the *Axiom of Choice*, which states the following.

Let I be a set, and let $(X_i)_{i \in I}$ be a collection of nonempty sets indexed by I. Then there is a 'choice function' $f : I \to \bigcup_{i \in I} X_i$ such that $f(i) \in X_i$ for all $i \in I$.

In other words, if all the X_i are nonempty, then the product $\prod_{i \in I} X_i$ of these sets is also nonempty. This looks like a natural property, however it has consequences like the existence of $\mathbb{Q}$-bases of $\mathbb{R}$ that are not so intuitive any more. Also, as it turned out, the Axiom of Choice is *independent* from the other axioms of set theory: it is not implied by them.

For some time, there was some discussion among mathematicians as to whether the use of the Axiom of Choice (and therefore, of Zorn's Lemma) should be allowed or forbidden (because of its inconstructive character). By now, a pragmatic viewpoint has been adapted by almost everybody: use it when you need it. For example, interesting parts of analysis and algebra need the Axiom of Choice, and mathematics would be quite a bit poorer without it.

Finally, a historical remark: Zorn's Lemma was first discovered by Kazimierz Kuratowski in 1922 (and rediscovered by Max Zorn about a dozen years later), so it is not really appropriately named. In fact, when Michael Stoll was a student, one of his professors told them that he talked to Zorn at some occasion, who said that he was not at all happy that the statement was carrying his name...

Exercises

E.0.1. Use Zorn's Lemma to prove that for every subset X of a vector space V such that X contains the zero vector, there is a maximal linear subspace of V contained in X.

E.0.2. Show that Zorn's Lemma implies that every linear subspace U of a vector space V has a complementary subspace in V.

E.0.3. Suppose V is a vector space with subspaces U and U' satisfying $U \cap U' = \{0\}$. Show that Zorn's Lemma implies that we can extend U' to a complementary space of U, that is, there exists a subspace $W \subset V$ containing U' that is a complementary subspace of U in V.

Bibliography

[BR1] T.S. Blyth and E.F. Robertson: *Basic Linear Algebra.* Springer Undergraduate Mathematics Series, 2002.

[BR2] T.S. Blyth and E.F. Robertson: *Further Linear Algebra.* Springer Undergraduate Mathematics Series, 2002.

[J] K. Jänich: *Linear Algebra.* Springer Undergraduate Texts in Mathematics, 1994.

[KM] A. Kostrykin and Y. Manin: *Linear Algebra and Geometry.* Gordan and Breach, 1988.

[S] M. Stoll: *Linear Algebra I.* 2007.

Index of notation

Index