Linear Algebra I

Ronald van Luijk, 2014

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

Contents

CHAPTER 1

Vector spaces

Many sets in mathematics come with extra structure. In the set $\mathbb{R}$ of real numbers, for instance, we can add and multiply elements. In linear algebra, we study *vector spaces*, which are sets in which we can *add* and *scale* elements. By proving theorems using only the addition and the scaling, we prove these theorems for all vector spaces at once.

All we require from our scaling factors, or *scalars*, is that they come from a set in which we can add, subtract, and multiply elements, and divide by any nonzero element. Sets with this extra structure are called *fields* (see Appendix B). We will often use the field $\mathbb{R}$ of real numbers in our examples, but by allowing ourselves to work over more general fields, we also cover linear algebra over finite fields, such as the field $\mathbb{F}_2 = \{0, 1\}$ of two elements, which has important applications in computer science and coding theory.

1.1. Examples

We start with some examples of a set with an addition and a scaling, the latter often being referred to as *scalar multiplication*.

EXAMPLE 1.1. 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.

8	1	6
3	5	7
4	9	2

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 magic square, which we denote by λX. If all rows, columns, and diagonals of X add up to s, then those of λX add up to λs. Moreover, if we have two magic squares X and Y, then we can make a new magic square Z by letting the top-left number in Z 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 Z are all $s + t$. Check this, and verify that $5A - B - 3C$ equals the well-known magic square above.

EXAMPLE 1.2. Consider the set $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ of all pairs of real numbers. The pairs can be interpreted as points in the plane, where the two numbers of the pair correspond to the coordinates of the point. We define the sum of two pairs (a, b) and (c, d) in $\mathbb{R}^2$ by adding the first elements of each pair, as well as the second, so

$$(a, b) + (c, d) = (a + c, b + d).$$

We define the scalar multiplication of a pair $(a, b) \in \mathbb{R}^2$ by a factor $\lambda \in \mathbb{R}$ by setting

$$\lambda \cdot (a, b) = (\lambda a, \lambda b).$$

EXAMPLE 1.3. Let $\mathrm{Map}(\mathbb{R}, \mathbb{R})$ be the set 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 $f + g$ that is given by

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

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 $\lambda \cdot f$ that is given by

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

for all $x \in \mathbb{R}$.

REMARK 1.4. Obviously, if f is a function from $\mathbb{R}$ to $\mathbb{R}$ and x is a real number, then $f(x)$ is also a real number. In our notation, we will always be careful to distinguish between the function f and the number $f(x)$. Therefore, 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 1.5. Nothing stops us from taking any set X and the set $\mathrm{Map}(X, \mathbb{R})$ of all functions from X to $\mathbb{R}$ and repeating the construction of addition and scalar multiplication from Example 1.3 on $\mathrm{Map}(X, \mathbb{R})$. We will do this in a yet more general situation in Example 1.13.

EXAMPLE 1.6. A real *polynomial* in the variable x 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 different integral powers x^i multiplied by a real constant a_i; we say that a_i is the coefficient of the *monomial* x^i in f. The *degree* of $f = \sum_{i=0}^{d} a_i x^i$ with $a_d \neq 0$ is d. By definition the degree of 0 equals $-\infty$. Let $\mathbb{R}[x]$ denote the set of all real polynomials in the variable x. We define the addition of polynomials coefficientwise, so that the sum of the polynomials

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

equals

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

The scalar multiplication of f by $\lambda \in \mathbb{R}$ is given by

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

We could also define the product of two polynomials, but that has nothing to do with the vector space structure, for which we only care about the addition and scalar multiplication.

In the examples above, we used the ordinary addition on the set $\mathbb{R}$ of real numbers to define an addition on other sets. When reading an equation as

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

in Example 1.3, one should always make sure to identify which addition the plus-symbols + refer to. In this case, the left + refers to the addition on $\mathrm{Map}(\mathbb{R},\mathbb{R})$, while the right + refers to the ordinary addition on $\mathbb{R}$.

All examples describe an addition on a set V that satisfies all the rules that one would expect from the use of the word sum and the notation $v+w$. For example, one easily checks that in all examples we have

$$u+v=v+u \qquad \text{and} \qquad u+(v+w)=(u+v)+w$$

for all elements u,v,w in V. Also the scalar multiplication acts as its notation suggests. For instance, in all examples we have

$$\lambda\cdot(\mu\cdot v)=(\lambda\mu)\cdot v$$

for all scalars λ,μ and all elements v in V.

We will define vector spaces in Section 1.2 as a set with an addition and a scalar multiplication satisfying these same three rules and five more. The examples above are all vector spaces in which the scaling is done by real numbers.

Exercises

1.1.1. Let A,B,C be the magic squares as in Example 1.1. Prove that all 3×3 magic squares can be written as $\lambda A+\mu B+\nu C$ for some real numbers λ,μ,ν.

***1.1.2.** 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?

1.2. Definition of a vector space

We can now define the general notion of a vector space. Roughly speaking, it is just a set V of elements that we can add and scale. The sum $x+y$ of two elements x and y of V is again an element of V. In other words, the addition is a map $V\times V\to V$ from its domain $V\times V$ of pairs (x,y) with $x,y\in V$, to its codomain V. For the scaling, we use scalars from a field F (see Appendix B), which we often choose to be the field $\mathbb{R}$ of real numbers. The scalar multiple λx of a scalar $\lambda\in F$ and an element $x\in V$ is again an element of V. In other words, the scalar multiplication is a map $F\times V\to V$ from its domain $F\times V$ of pairs (λ,x) with $\lambda\in F$ and $x\in V$, to its codomain V. The addition and scalar multiplication have to satisfy some rules, and the exact definition of a vector space is as follows.

Definition 1.7. Let F be a field. A *vector space* or *linear space* over F, or an *F-vector space*, is a set V with a distinguished zero element $0\in V$, together with two maps $+:V\times V\to V$ ('addition') and $\cdot:F\times V\to V$ ('scalar multiplication'), written, as usual, $(x,y)\mapsto x+y$ and $(\lambda,x)\mapsto\lambda\cdot x$ or λx, respectively, satisfying the following axioms.

(1) For all $x,y\in V$, $\boxed{x+y=y+x}$ (addition is commutative).
(2) For all $x,y,z\in V$, $\boxed{(x+y)+z=x+(y+z)}$ (addition is associative).
(3) For all $x\in V$, $\boxed{x+0=x}$ (adding the zero element does nothing).

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

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

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

(7) For all $\lambda \in \mathbb{R}$ and $x, y \in V$, $\boxed{\lambda(x + y) = \lambda x + \lambda y}$ (distributivity I).

(8) For all $\lambda, \mu \in \mathbb{R}$ and $x \in V$, $\boxed{(\lambda + \mu)x = \lambda x + \mu x}$ (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 1.8.

(1) The first four axioms above exactly state that $(V, 0, +)$ is an (additive) *abelian group*. (If you didn't know what an abelian group is, then this is the definition.)

(2) Instead of writing $(V, 0, +, \cdot)$ (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.

The examples of Section 1.1 are real vector spaces. In the examples below, many will be generalized to general fields. In each case we also specify the zero of the vectorspace. It is crucial to always distinguish this from the zero of the field F, even though both are written as 0.

Example 1.9. The simplest (and perhaps least interesting) example of a vector space over a field F is $V = \{0\}$, with addition given by $0 + 0 = 0$ and scalar multiplication by $\lambda \cdot 0 = 0$ 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 1.10. 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.

Example 1.11. Now we come to a very important example, which is *the* model of a vector space. Let F be a field. We consider $V = F^n$, the set of n-tuples of elements of F, with zero element $0 = (0, 0, \ldots, 0)$. We define addition and scalar multiplication 'component-wise':

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

Of course, we now have to *prove* that our eight axioms are satisfied by our choice of $(V, 0, +, \cdot)$. In this case, this is very easy, since everything reduces to addition and multiplication in the field F. As an example, let us show that the first distributive law (7) and the existence of negatives (4) are satisfied. For the first, take $x, y \in F^n$ and write them as

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

Then we have

$$\begin{aligned}\lambda(x+y) &= \lambda\big((x_1,x_2,\ldots,x_n)+(y_1,y_2,\ldots,y_n)\big)\\ &= \lambda\cdot(x_1+y_1,x_2+y_2,\ldots,x_n+y_n)\\ &= \big(\lambda(x_1+y_1),\lambda(x_2+y_2),\ldots,\lambda(x_n+y_n)\big)\\ &= (\lambda x_1+\lambda y_1,\lambda x_2+\lambda y_2,\ldots,\lambda x_n+\lambda y_n)\\ &= (\lambda x_1,\lambda x_2,\ldots,\lambda x_n)+(\lambda y_1,\lambda y_2,\ldots,\lambda y_n)\\ &= \lambda(x_1,x_2,\ldots,x_n)+\lambda(y_1,y_2,\ldots,y_n)=\lambda x+\lambda y.\end{aligned}$$

This proves the first distributive law (7) for F^n. Note that for the fourth equality, we used the distributive law for the field F. For the existence of negatives (4), take an element $x \in F^n$ and write it as $x = (x_1, x_2, \ldots, x_n)$. For each i with $1 \le i \le n$, we can take the negative $-x_i$ of x_i in the field F and set

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

Then, of course, we have

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

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

Of course, for $n = 2$ and $n = 3$ and $F = \mathbb{R}$, this is more or less what you know as 'vectors' from high school; the case $n = 2$ is also Example 1.2. For $n = 1$, this example reduces to the previous one (if one identifies 1-tuples (x) with elements 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.)

Example 1.12. A special case of Example 1.11 is when $F = \mathbb{R}$. The vector space $\mathbb{R}^n$ is called Euclidean n-space. In Chapter 3 we will consider lengths, angles, reflections, and projections in $\mathbb{R}^n$. 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 $\mathbb{R}^2$, if we take an orthogonal coordinate system in the plane, with 0 at the origin, 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. This way, the vectors, which are pairs of real numbers, 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. By doing so, we can now add points in the plane, as well as in space!
In physics, more precisely in relativity theory, $\mathbb{R}^4$ is often interpreted as space with a fourth coordinate for time.
For $n = 2$ or $n = 3$, we may also interpret vectors as arrows in the plane or space, respectively. In the plane, the arrow from the point $p = (p_1, p_2)$ to the 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 $q - p$! Of course we can do the same for $\mathbb{R}^3$.

For example, take the two 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.
We can now interpret negation, scalar multiples, sums, and differences of vectors geometrically, namely in terms of arrows. Make your own pictures! If a vector v corresponds to a certain arrow, then $-v$ corresponds to any arrow with the same length but opposite direction; more generally, for $\lambda \in \mathbb{R}$ the vector λv corresponds to the arrow obtained by scaling the arrow for v by a factor λ.
If v and w correspond to two arrows that have common tail p, then these two arrows are the sides of a unique parallelogram; the vector $v + w$ corresponds to 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 $v + w$ corresponds to the arrow from the tail of the first to the head of the second. Compare the two constructions in a picture!
For the same v and w, still with common tail and with heads q and r respectively, the difference $v - w$ corresponds to 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 above. Make a picture again!

Example 1.13. This example generalizes Example 1.5. Let F be a field. Let us consider any set X and look at the set $\mathrm{Map}(X, F)$ or F^X of all maps (or functions) from X to F:

$$V = \mathrm{Map}(X, F) = F^X = \{f : X \to F\}\,.$$

We take the zero vector 0 to be the constant zero function that sends each element of X to 0 in $\mathbb{R}$. In order to get a vector space, we have to define addition and scalar multiplication. To define addition, for every pair of functions $f, g : X \to F$, we have to define a new function $f + g : X \to F$. The only reasonable way to do this is as follows ('point-wise'):

$$f + g : X \longrightarrow F\,, \quad x \longmapsto f(x) + g(x)\,,$$

or, in a more condensed form, by writing $(f+g)(x) = f(x)+g(x)$. (Make sure that you understand these notations!) In a similar way, we define scalar multiplication:

$$\lambda f : X \longrightarrow F\,, \quad x \longmapsto \lambda \cdot f(x)\,.$$

We then have to check the axioms in order to convince ourselves that we really get a vector space. Let us do again the first distributive law as an example. We have to check that $\lambda(f + g) = \lambda f + \lambda g$, which means that for all $x \in X$, we want

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

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

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

Note the parallelism of this proof with the one in the previous example. That parallelism goes much further. If we take $X = \{1, 2, \ldots, n\}$, then the set $F^X = \mathrm{Map}(X, F)$ of maps $f : \{1, 2, \ldots, n\} \to F$ can be identified with F^n by letting such a map f correspond to the n-tuple $(f(1), f(2), \ldots, f(n))$. It is not a coincidence that the notations F^X and F^n are chosen so similar! What do we get when X is the empty set?

EXAMPLE 1.14. This example generalizes Example 1.6. A *polynomial* in the variable x over a field 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 different integral powers x^i multiplied by a constant $a_i \in F$; the products $a_i x^i$ are called the *terms* of f and we say that a_i is the coefficient of x^i in f. We let the zero vector 0 be the zero polynomial, for which $a_i = 0$ holds for all i. 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. We define the addition and scalar multiplication of polynomials as in Example 1.6. 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.

WARNING 1.15. 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 $\alpha \in \mathbb{F}_2$.

REMARK 1.16. 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 a field 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, e.g., $\mathbb{F}_2$.

EXAMPLE 1.17. There are other examples that may appear more strange. 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*: $A + B = (A \setminus B) \cup (B \setminus A)$ (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: $0 \cdot A = \emptyset$, $1 \cdot A = A$. These operations turn V into an $\mathbb{F}_2$-vector space.

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 earlier on $\mathbb{F}_2^X$.

Exercises

1.2.1. Compute the sum of the given vectors v and w in $\mathbb{R}^2$ and draw a corresponding picture (cf. Example 1.12).
(1) $v = (-2, 5)$ and $w = (7, 1)$,
(2) $v = 2(-3, 2)$ and $w = (1, 3) + (-2, 4)$,
(3) $v = (-3, 4)$ and $w = (4, 3)$,
(4) $v = (-3, 4)$ and $w = (8, 6)$,
(5) $v = (2, -7)$ and $w = (x, y)$,
(6) $v = w = (a, b)$.

1.2.2. In Example 1.11, 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.

1.2.3. In Example 1.13, 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.

1.2.4. Let $(V, 0, +, \cdot)$ be a real vector space and define $x - y = x + (-y)$, as usual. Which of the vector space axioms are satisfied and which are not (in general), for $(V, 0, -, \cdot)$?

NOTE. You are expected to give proofs for the axioms that hold and to give counterexamples for those that do not hold.

1.2.5. Prove that the set $F[x]$ of polynomials over F, together with addition, scalar multiplication, and the zero as defined in Example 1.14 is a vector space.

1.2.6. Given the field F and the set V in the following cases, together with the described addition and scalar multiplication, as well as the implicit element 0, which cases determine a vector space? If not, 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 1.17.
(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.

1.2.7. 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?

1.2.8. We can generalize Example 1.13 further. Let F be a field and V 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.

1.2.9. 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.

1.2.10. 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.

***1.2.11.** For each of the eight axioms in Definition 1.7, try to find a system $(V, 0, +, \cdot)$ that does not satisfy that axiom, while it does satisfy the other seven.

1.3. 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 1.7. We will show that many such rules follow from the basic eight. The first is a cancellation rule.

LEMMA 1.18. *If three elements x, y, z of a vector space V satisfy $x + z = y + z$, then $x = y$.*

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

$$x = x + 0 = x + (z + z') = (x + z) + z' = (y + z) + z' = y + (z + z') = y + 0 = y,$$

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

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

PROPOSITION 1.19. In a vector space V, there is only one zero element, i.e., if two elements $0' \in V$ and $z \in V$ satisfy $0' + z = z$, then $0' = 0$.

PROOF. Exercise. □

PROPOSITION 1.20. *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, i.e., $x + a = 0$, $x + b = 0$. Then by commutativity we have

$$a + x = x + a = 0 = x + b = b + x,$$

so $a = b$ by Lemma 1.18. □

NOTATION 1.21. Since negatives are unique, given $x \in V$ we may write $-x$ for the unique element that satisfies $x + (-x) = 0$. As usual, we write $x - y$ for $x + (-y)$.

Here are some more harmless facts.

REMARKS 1.22. *Let $(V, 0, +, \cdot)$ be a vector space over a field F.*

(1) *For all* $x \in V$*, we have* $0 \cdot x = 0$*.*
(2) *For all* $x \in V$*, we have* $(-1) \cdot x = -x$*.*
(3) *For all* $\lambda \in F$ *and* $x \in V$ *such that* $\lambda x = 0$*, we have* $\lambda = 0$ *or* $x = 0$*.*
(4) *For all* $\lambda \in F$ *and* $x \in V$*, we have* $-(\lambda x) = \lambda \cdot (-x)$*.*
(5) *For all* $x, y, z \in V$*, we have* $z = x - y$ *if and only if* $x = y + z$*.*

PROOF. Exercise. □

Exercises

1.3.1. Prove Proposition 1.19.

1.3.2. Prove Remarks 1.22.

1.3.3. Is the following statement correct? "Axiom (4) of Definition 1.7 is redundant because we already know by Remarks 1.22(2) that for each vector $x \in V$ the vector $-x = (-1) \cdot x$ is also contained in V."

CHAPTER 2

Subspaces

2.1. Definition and examples

In many applications, we do not want to consider all elements of a given vector space V, rather we only consider 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 2.1. 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 $u_1 + u_2 \in U$.
(3) If $\lambda \in F$ and $u \in U$, then $\lambda u \in U$.

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 2.2. Let $(V, +, \cdot, 0)$ be an F-vector space. If $U \subset V$ is a linear subspace of V, then $(U, +|_{U\times U}, \cdot|_{F\times U}, 0)$ is again an F-vector space.

The notation $+|_{U\times U}$ means that we take the addition map $+ : V \times V$, but *restrict* it to $U \times U$. (Strictly speaking, we also restrict its target set from V to U. However, this is usually suppressed in the notation.)

PROOF OF LEMMA 2.2. 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 ..., $\boxed{\ldots}$' 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 1.20, Notation 1.21, and Remarks 1.22). 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 2.3. Let V be a vector space. Then $\{0\} \subset V$ and V itself are linear subspaces of V.

EXAMPLE 2.4. Consider $V = \mathbb{R}^2$ and, for $a \in \mathbb{R}$, $U_a = \{(x, y) \in \mathbb{R}^2 : x + y = a\}$. When is U_a a linear subspace?

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

$$\begin{aligned}(x_1, y_1), (x_2, y_2) \in U_0 &\Longrightarrow x_1 + y_1 = 0, \quad x_2 + y_2 = 0 \\ &\Longrightarrow (x_1 + x_2) + (y_1 + y_2) = 0 \\ &\Longrightarrow (x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2) \in U_0\end{aligned}$$

and

$$\begin{aligned}\lambda \in \mathbb{R}, (x, y) \in U_0 &\Longrightarrow x + y = 0 \\ &\Longrightarrow \lambda x + \lambda y = \lambda(x + y) = 0 \\ &\Longrightarrow \lambda(x, y) = (\lambda x, \lambda y) \in U_0\,.\end{aligned}$$

We conclude that U_0 is indeed a subspace.

EXAMPLE 2.5. Let F be a field, X any set, and $x \in X$ an element. Consider the subset

$$U_x = \{f \colon X \to F \mid f(x) = 0\}$$

of the vector space F^X. 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.

EXAMPLE 2.6. Consider $V = \mathbb{R}^{\mathbb{R}} = \{f : \mathbb{R} \to \mathbb{R}\}$, 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

$$\mathcal{C}(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} \mid f \text{ is continuous}\}$$

is a linear subspace of V.

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

$$\mathcal{C}^n(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} \mid f \text{ is } n \text{ times differentiable and } f^{(n)} \text{ is continuous}\}$$

is again a linear subspace of V.

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

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

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 V.

To define subspaces of F^n it is convenient to introduce the following notation.

DEFINITION 2.7. Let F be a field. 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 *dot product* of x and y as

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

Note that the dot product $\langle x, y \rangle$ is an element of F.

The dot product is often written in other pieces of literature as $x \cdot y$, which explains its name. Although this notation looks like scalar multiplication, it should always be clear from the context which of the two is mentioned, as one involves two vectors and the other a scalar and a vector. Still, we will always use the notation $\langle x, y \rangle$ to avoid confusion. When the field F equals $\mathbb{R}$ (or a subset of $\mathbb{R}$), then the dot product satisfies the extra property $\langle x, x \rangle \geq 0$ for all $x \in \mathbb{R}^n$; over these fields we also refer to the dot product as the *inner product* (see Section 3). Other pieces of literature may use the two phrases interchangeably over all fields.

EXAMPLE 2.8. Suppose we have $x = (3, 4, -2)$ and $y = (2, -1, 5)$ in $\mathbb{R}^3$. Then we get

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

EXAMPLE 2.9. Suppose we have $x = (1, 0, 1, 1, 0, 1, 0)$ and $y = (0, 1, 1, 1, 0, 0, 1)$ in $\mathbb{F}_2^7$. Then we get

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

The dot product satisfies the following useful properties.

PROPOSITION 2.10. *Let F be a field with an element $\lambda \in F$. 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. The two identities (1) and (3) are an exercise for the reader. We will prove the second identity. Write x and y as

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

Then we have $\lambda x = (\lambda x_1, \lambda x_2, \ldots, \lambda x_n)$, so

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

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). □

Note that from properties (1) and (3) we also find the equality $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$. Properties (2) and (3), together with this last property, mean that the dot product is *bilinear*. Note that from the properties above it also follows that $\langle x, y - z \rangle = \langle x, y \rangle - \langle x, z \rangle$ for all vectors $x, y, z \in F^n$; of course this is also easy to check directly.

EXAMPLE 2.11. Consider $\mathbb{R}^2$ with coordinates x and y. Let $L \subset \mathbb{R}^2$ be the line given by $3x + 5y = 7$. For the vector $a = (3, 5)$ and $v = (x, y)$, we have

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

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

The following example is very similar to Example 2.4. The dot product and Proposition 2.10 allow us to write everything much more efficiently.

EXAMPLE 2.12. 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 $b = 0$.

We can generalize this to F^n for any positive integer n.

DEFINITION 2.13. Let F be a field, $a \in F^n$ a nonzero vector, and $b \in F$ a constant. Then the set

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

is called a *hyperplane*.

EXAMPLE 2.14. Any line in $\mathbb{R}^2$ is a hyperplane, cf. Example 2.12.

EXAMPLE 2.15. Any plane in $\mathbb{R}^3$ is a hyperplane. If we use coordinates x, y, z, then any plane is given by the equation $px + qy + rz = b$ for some constants $p, q, r, b \in \mathbb{R}$ with p, q, r not all 0; equivalently, this plane consists of all points $v = (x, y, z)$ that satisfy $\langle a, v \rangle = b$ with $a = (p, q, r) \neq 0$.

PROPOSITION 2.16. *Let F be a field, $a \in F^n$ a nonzero vector, and $b \in F$ a constant. Then the hyperplane*

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

is a subspace of F^n if and only if $b = 0$.

PROOF. The proof is completely analogous to Example 2.12. See also Exercise 2.1.11. □

DEFINITION 2.17. Let F be a field and $a, v \in F^n$ vectors with v nonzero. Then the subset

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

of F^n is called a *line*.

PROPOSITION 2.18. *Let F be a field and $a, v \in F^n$ vectors with v nonzero. Then the line*

$$L = \{\, a + \lambda v \;:\; \lambda \in F \,\} \subset F^n$$

is a subspace if and only if there exists a scalar $\lambda \in F$ such that $a = \lambda v$.

PROOF. Exercise. □

Exercises

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

2.1.2. Write the following equations for lines in $\mathbb{R}^2$ with coordinates x_1 and x_2 in the form $\langle a, x\rangle = c$, i.e., 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) $L_7 \colon 6x_1 - 2x_2 = 7$

2.1.3. 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 $0 \cdot v = 0$. (What do the zeros in this statement refer to?)

2.1.4. Given an integer $d \geq 0$, let $\mathbb{R}[x]_d$ denote the set of 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.

2.1.5. Let X be a set with elements $x_1, x_2 \in X$, and let F be a field. Is the set

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

a subspace of F^X?

2.1.6. Let X be a set with elements $x_1, x_2 \in X$. Is the set

$$U = \{\, f \in \mathbb{R}^X \;:\; f(x_1) = f(x_2)^2 \,\}$$

a subspace of $\mathbb{R}^X$?

2.1.7. Which of the following are linear subspaces of the vector space $\mathbb{R}^2$? Explain your answers!

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

2.1.8. 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) $U_6 = \{f \in V \;:\; f(0) \geq 0\}$

2.1.9. Prove Proposition 2.10.

2.1.10. Prove Proposition 2.18.

2.1.11. Let F be any field. 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 $b_1 = \ldots = b_t = 0$.

2.1.12. Let X be a set and F a field.

(1) Show that the set $F^{(X)}$ of all functions $f \colon X \to F$ 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 $f\colon X \to V$ that satisfy $f(x) = 0$ for all but finitely many $x \in X$ is a subspace of the vector space V^X (cf. Exercise 1.2.8).

2.1.13. Let X be a set and F a field.

(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 2.1.12.

(2) More generally, let V be a vector space over F. Show that the set of all functions $f\colon X \to V$ with finite image is a subspace of the vector space V^X that contains $V^{(X)}$ of Exercise 2.1.12.

2.2. Intersections

The following result can be used, for example, to show that, with U and $\mathcal{C}(\mathbb{R})$ as in Example 2.6, 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 2.19. 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. □

Note that in general, if U_1 and U_2 are linear subspaces, then $U_1 \cup U_2$ is not (it is if and only if $U_1 \subset U_2$ or $U_2 \subset U_1$ — Exercise!).

EXAMPLE 2.20. 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

2.2.1. 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 $U_2 \subset U_1$.

2.2.2. 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.

2.2.3. 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.

2.3. 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 2.21. Let V be a vector space over a field 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. This implies that every element of the form $\lambda_1 v_1 + \lambda_2 v_2$ is contained in W. So set

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

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

$$\begin{aligned}
&0 = 0 \cdot v_1 + 0 \cdot v_2 \in U,\\
&(\lambda_1 v_1 + \lambda_2 v_2) + (\mu_1 v_1 + \mu_2 v_2) = (\lambda_1 + \mu_1) v_1 + (\lambda_2 + \mu_2) v_2 \in U,\\
&\lambda(\lambda_1 v_1 + \lambda_2 v_2) = (\lambda\lambda_1) v_1 + (\lambda\lambda_2) v_2 \in U.
\end{aligned}$$

(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 generalizes.

DEFINITION 2.22. Let V be an F-vector space, $v_1, v_2, \ldots, v_n \in V$. The *linear combination* (or, more precisely, *F-linear combination*) of $v_1, v_2, \ldots, v_n$ with coefficients $\lambda_1, \lambda_2, \ldots, \lambda_n \in F$ is the element

$$v = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n.$$

If $n = 0$, 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 on S* is a linear combination of *finitely many* elements of S.

DEFINITION 2.23. Let V be a vector space over a field 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 $L_F(S)$. If $v_1, v_2, \ldots, v_n \in V$, we also write $L(v_1, v_2, \ldots, v_n)$ for $L(\{v_1, v_2, \ldots, v_n\})$.

Note that if $S = \{v_1, v_2, \ldots, v_n\}$, then any linear combination on S, a priori not necessarily using all n vectors, is also a linear combination of all $v_1, v_2, \ldots, v_n$, as we can just add coefficients zero for the vectors that were not used. This means that $L(v_1, v_2, \ldots, v_n)$ is the set of all linear combinations of $v_1, v_2, \ldots, v_n$.

PROPOSITION 2.24. *Let V be a vector space, and $v_1, v_2, \ldots, v_n \in V$. Then the set $L(v_1, v_2, \ldots, v_n)$ 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 $U = L(v_1, v_2, \ldots, v_n)$. First of all, $0 \in U$, since $0 = 0v_1 + 0v_2 + \cdots + 0v_n$ (this even works for $n = 0$). To check that U is closed under addition, let $v = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n$ and $w = \mu_1 v_1 + \mu_2 v_2 + \cdots + \mu_n v_n$ be two elements of U. Then

$$\begin{aligned}
v + w &= (\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n) + (\mu_1 v_1 + \mu_2 v_2 + \cdots + \mu_n v_n)\\
&= (\lambda_1 + \mu_1) v_1 + (\lambda_2 + \mu_2) v_2 + \cdots + (\lambda_n + \mu_n) v_n
\end{aligned}$$

is again a linear combination of $v_1, v_2, \ldots, v_n$. Also, for $\lambda \in F$,

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

is a linear combination of $v_1, v_2, \ldots, v_n$. So 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. 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. □

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 $L(S) = V$, we say that S *generates* V, or that S is a *generating set* for V. If V can be generated by a finite set S, then we say that V is *finitely generated*.

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

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

LEMMA 2.25. *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 $L(S) \subset U$. □

LEMMA 2.26. *Let V be an F-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 2.25 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 $L(S) = L(T)$. □

LEMMA 2.27. *Let V be an F-vector space and $S \subset V$ a set that generates V. Suppose $T \subset V$ is a subset whose span $L(T)$ contains S. Then T also generates V.*

PROOF. We have $T \subset V = L(S)$, so from Lemma 2.26 we find $L(T) = L(S) = V$, which proves the lemma. □

EXAMPLE 2.28. Note that for any nonzero $v \in F^n$, the subspace $L(v)$ consists of all multiples of v, so $L(v) = \{\lambda v : \lambda \in F\}$ is a line (see Definition 2.17).

EXAMPLE 2.29. 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 $x = x_1 e_1 + x_2 e_2 + x_3 e_3$, so every element in $\mathbb{R}^3$ is a linear combination of e_1, e_2, e_3. We conclude $\mathbb{R}^3 \subset L(e_1, e_2, e_3)$ and therefore $L(e_1, e_2, e_3) = \mathbb{R}^3$, so $\{e_1, e_2, e_3\}$ generates $\mathbb{R}^3$.

DEFINITION 2.30. Let F be a field and n a positive integer. The *standard generators* of F^n are

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

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

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

By Lemma 2.27, if we want to show that a certain set $T \subset F^n$ generates F^n, then it suffices to show that the standard generators of F^n are linear combinations of T.

The following proposition again shows that $L(S)$ is the smallest subspace containing S.

PROPOSITION 2.31. *Let V be an F-vector space and S a subset of V. Then we have*

$$L(S) = \bigcap \{U \subset V : U \text{ linear subspace of } V \quad \text{and} \quad S \subset U\}.$$

Note that the notation in this proposition means the intersection of all elements of the specified set: we intersect all linear subspaces containing S.

PROOF. The space V itself is a subspace containing S, so the set of subspaces of which we take the intersection is non-empty. Let U denote this intersection. By Lemma 2.19, the intersection U is a subspace of V. Since U contains S, we obtain $L(S) \subset U$ from Lemma 2.25. As the subspace $L(S)$ contains S, this subspace $L(S)$ is one of the subspaces that U is the intersection of, so we also have $U \subset L(S)$, which proves the proposition. □

REMARK 2.32. What do we get in the extreme case that $S = \emptyset$? Well, then we have to intersect *all* linear subspaces of V, so Proposition 2.31 reduces to the statement $L(\emptyset) = \{0\}$.

EXAMPLE 2.33. 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 2.16) that both contain v_1, v_2, v_3. So certainly we have an inclusion $L(v_1, v_2, v_3) \subset H_1 \cap H_2$.

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 $x = x_1v_1 + x_2v_2$. We conclude $x \in L(v_1, v_2)$, 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 7.3 we will see how to compute generators of intersections more systematically.

EXAMPLE 2.34. Let us consider again the vector space $\mathcal{C}(\mathbb{R})$ of continuous functions on $\mathbb{R}$. The power functions $f_n : x \mapsto x^n$ $(n = 0, 1, 2, \dots)$ are certainly continuous and defined on $\mathbb{R}$, so they are elements of $\mathcal{C}(\mathbb{R})$. We find that their linear hull $L(\{f_n : n \in \mathbb{N}_0\})$ is the linear subspace of *polynomial functions*, 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 $n \in \mathbb{N}_0$ and $a_0, a_1, \dots, a_n \in \mathbb{R}$.

EXAMPLE 2.35. For any field we can consider the power functions $f_n : x \mapsto x^n$ inside the vector space F^F of all functions from F to F. Their linear hull $L(\{f_n : n \in \mathbb{N}_0\}) \subset F^F$ 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 $n \in \mathbb{N}_0$ and $a_0, a_1, \dots, a_n \in F$. By definition, the power functions f_n generate the subspace of polynomial functions.

WARNING 2.36. In Example 1.6 we defined real *polynomials* in the variable x as formal (or abstract) sums of powers x^i multiplied by a real constant a_i. These are not to be confused with the *polynomial functions* $f\colon \mathbb{R} \to \mathbb{R}$, though the difference is subtle: over a general field, the subspace of polynomial functions is generated by the power functions f_n from Example 2.35, while the space $F[x]$ of polynonomials is generated by the formal powers x^i of a variable x.
As stated in Warning 1.15, though, over some fields the difference between polynomials, as defined in Example 1.14, and polynomial functions, as defined in Example 2.35, 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.

DEFINITION 2.37. Let F be a field and S any subset of F^n. Then we set

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

In Remark 3.9 we will clarify the notation $S^\perp$.

EXAMPLE 2.38. Let F be a field. Then for every element $a \in F^n$, the hyperplane $H_a = \{x \in F^n \ : \ \langle a, x\rangle = 0\}$ equals $\{a\}^\perp$. Moreover, the set $S^\perp$ is the intersection of all hyperplanes H_a with $a \in S$, i.e.,

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

For instance, the intersection $H_1 \cap H_2$ of Example 2.33 can also be written as $\{a_1, a_2\}^\perp$.

PROPOSITION 2.39. *Let F be a field and S any subset of F^n. Then the following statements hold.*

(1) *The set $S^\perp$ is a subspace of F^n.*
(2) *We have $S^\perp = L(S)^\perp$.*
(3) *We have $L(S) \subset (S^\perp)^\perp$.*
(4) *For any subset $T \subset S$ we have $S^\perp \subset T^\perp$.*
(5) *For any subset $T \subset F^n$ we have $S^\perp \cap T^\perp = (S \cup T)^\perp$.*

PROOF. We leave (1), (3), (4), and (5) as an exercise to the reader. To prove (2), note that from $S \subset L(S)$ and (4) 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

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

□

REMARK 2.40. Later we will see that the inclusion $L(S) \subset (S^\perp)^\perp$ of Proposition 2.39 is in fact an equality, so that for every subspace U we have $(U^\perp)^\perp = U$. See Corollary 7.15 and Exercise 7.2.4.

Exercises

2.3.1. Prove Proposition 2.39.

2.3.2. Do the vectors

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

generate $\mathbb{R}^3$?

2.3.3. Do the vectors

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

generate $\mathbb{R}^3$?

2.3.4. 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.

2.3.5. Let F be a field and X 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 2.1.12). 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 $F^{(X)}$.

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

2.3.7. 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?

2.3.8. Let F be a field and $F[x]$ 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, i.e., to $f(a) := \sum_{i=0}^d c_i a^i$, cf. Warning 1.15.

(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.]

2.3.9. Given a vector space V over a field F and vectors $v_1, v_2, \dots, v_n \in V$. Set $W = L(v_1, v_2, \dots, v_n)$. Using Lemma 2.26, give short proofs of the following equalities of subspaces.

(1) $W = L(v_1', \dots, v_n')$ 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) $W = L(v_1', \dots, v_n')$ 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) $W = L(v_1', \dots, v_n')$ 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),

2.4. 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 2.20. 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 2.41. 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 2.42. *If U_1 and U_2 are linear subspaces of the vector space V, then*

$$U_1 + U_2 = \{u_1 + u_2 : u_1 \in U_1, u_2 \in U_2\}\,.$$

If $(U_i)_{i\in I}$ is a family of linear subspaces of V, then

$$\sum_{i\in I} U_i = \Big\{\sum_{j\in J} u_j : J \subset I \text{ finite and } u_j \in U_j \text{ for all } j \in J\Big\}\,.$$

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 2.25 (applied with S equal to the union $U_1 \cup U_2$, resp. $\bigcup_{i\in I} U_i$, which is obviously contained in the right-hand side) to show that the right-hand sides are linear subspaces.

We have $0 = 0 + 0$ (resp., $0 = \sum_{j\in\emptyset} u_j$), so 0 is an element of the right-hand side sets. Closure under scalar multiplication is easy to see:

$$\lambda(u_1 + u_2) = \lambda u_1 + \lambda u_2\,,$$

and we have $\lambda u_1 \in U_1$, $\lambda u_2 \in U_2$, because U_1, U_2 are linear subspaces. Similarly,

$$\lambda \sum_{j\in J} u_j = \sum_{j\in J} \lambda u_j\,,$$

and $\lambda u_j \in U_j$, since U_j is a linear subspace. 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 $v_j = u_j \in U_j$ if $j \in J_1 \setminus J_2$, $v_j = u_j' \in U_j$ if $j \in J_2 \setminus J_1$, and $v_j = u_j + u_j' \in U_j$ if $j \in J_1 \cap J_2$. □

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 union $U_1 \cup U_2$, resp. $\bigcup_{i\in I} U_i$, is contained in the 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 nonnegative integer $r \geq s$ 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 2.43. The union $U = U_1 \cup U_2$ of Example 2.20 contains the vectors $e_1 = (1, 0)$ and $e_2 = (0, 1)$, so the sum $U_1 + U_2 = L(U)$ contains $L(e_1, e_2) = \mathbb{R}^2$ and we conclude $U_1 + U_2 = \mathbb{R}^2$.

Example 2.44. Let $V \subset \mathbb{R}^{\mathbb{R}}$ be the vector space 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 the claim $U_0 + U_1 = V$. 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

$$f_0 = \frac{f(1)}{f(1) - f(0)}(f - f(0)), \qquad f_1 = \frac{f(0)}{f(0) - f(1)}(f - f(1)),$$

while in the case $f(0) = f(1) = c$ we can take f_0 and f_1 given by

$$f_0(x) = c(f(x) + x - c) + (f(x) - c), \qquad f_1(x) = -c(f(x) + x - c - 1).$$

Lemma 2.45. *Suppose V is a vector space containing two subsets S and T. Then the equality $L(S) + L(T) = L(S \cup T)$ holds. In other words, 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.*

Proof. Exercise. □

DEFINITION 2.46. Let V be a vector space. Two linear subspaces $U_1, U_2 \subset V$ are said to be *complementary* if $U_1 \cap U_2 = \{0\}$ and $U_1 + U_2 = V$.

EXAMPLE 2.47. 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 2.48. *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 $u \in U$, $u' \in U'$ 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 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 Proposition 6.64 we will see that this is indeed the case, at least when V is finitely generated. In the next section, we will see an easy special case, namely when U is a subspace of F^n generated by an element $a \in F^n$ satisfying $\langle a, a\rangle \neq 0$. It turns out that in that case the hyperplane $\{a\}^\perp$ is a complementary subspace (see Corollary 3.15).

Exercises

2.4.1. Prove Lemma 2.45.

2.4.2. State and prove a version of Lemma 2.45 for an arbitrary collection of $(S_i)_{i\in I}$ of subsets.

2.4.3. Suppose F is a field and $U_1, U_2 \subset F^n$ subspaces. Show that we have

$$(U_1 + U_2)^\perp = U_1^\perp \cap U_2^\perp.$$

2.4.4. 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.

2.4.5. 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.

2.4.6. Let U_+ and U_- be the subspaces of $\mathbb{R}^\mathbb{R}$ of even and odd functions, respectively (cf. Exercise 2.3.7).

(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.

2.4.7. Are the subspaces U_0 and U_1 of Example 2.44 complementary subspaces?

2.4.8. 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.

CHAPTER 3

Euclidean space: lines and hyperplanes

This chapter, with the exception of Proposition 3.14, Corollary 3.15, and Exercise 3.2.10, deals with *Euclidean n-space* $\mathbb{R}^n$, as well as F^n for fields F that are contained in $\mathbb{R}$, such as the field $\mathbb{Q}$ of rational numbers.

3.1. Angles and orthogonality

As usual, we identify $\mathbb{R}^2$ and $\mathbb{R}^3$ with the plane and three-space through an orthogonal coordinate system, as in Example 1.12. Vectors correspond with points and vectors can be represented by arrows. In the plane and three-space, we have our usual notions of length, angle, and orthogonality. (Two lines are called *orthogonal*, or *perpendicular*, if the angle between them is $\pi/2$, or $90°$.) In this section we will generalize these notions to all $n \geq 0$. 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 3.2, as well as Proposition 3.5.

In $\mathbb{R}$ we can talk about elements being 'positive' or 'negative' and 'smaller' or 'bigger' than other elements. The dot product satisfies an extra property in this situation.

PROPOSITION 3.1. *Suppose F is a field contained in $\mathbb{R}$. Then for any 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 nonnegative, this sum of squares is also nonnegative and it equals 0 if and only if each terms equals 0, so if and only if $x_i = 0$ for all i with $1 \leq i \leq n$. □

Over $\mathbb{R}$ and fields that are contained in $\mathbb{R}$, we will also refer to the dot product as the *standard inner product* or just *inner product*. In other pieces of literature, the dot product may be called the inner product over any field.

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 $\sqrt{\langle x, x\rangle}$. Similarly, in $\mathbb{R}^2$ the length of an arrow representing the vector $x \in \mathbb{R}^2$ equals $\sqrt{\langle x, x\rangle}$. We define, more generally, the length of a vector in $\mathbb{R}^n$ for any integer $n \geq 0$ accordingly.

DEFINITION 3.2. Suppose F is a field contained in $\mathbb{R}$. Then for any element $x \in F^n$ we define the *length* $\|x\|$ of x as $\|x\| = \sqrt{\langle x, x\rangle}$.

Note that by Proposition 3.1, 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 $\mathbb{Q}$.

EXAMPLE 3.3. The length of the vector $(1,-2,2,3)$ in $\mathbb{R}^4$ equals $\sqrt{1+4+4+9} = 3\sqrt{2}$.

LEMMA 3.4. *Suppose F is a field contained in $\mathbb{R}$. Then for all $\lambda \in F$ and $x \in F^n$ we have $\|\lambda x\| = |\lambda| \cdot \|x\|$.*

PROOF. Exercise. □

PROPOSITION 3.5. *Suppose $n = 2$ or $n = 3$. Let v, w be two nonzero elements in $\mathbb{R}^n$ and let α be the angle between the arrow from 0 to v and the arrow from 0 to w. Then we have*

$$\cos\alpha = \frac{\langle v, w\rangle}{\|v\| \cdot \|w\|}. \tag{1}$$

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. 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

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

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

EXAMPLE 3.6. Let the lines l and m in the (x,y)-plane $\mathbb{R}^2$ be given by $y = ax + b$ and $y = cx + d$, respectively. Then their directions are the same as the lines $l' = L((1,a))$ and $m' = L((1,c))$, respectively. By Proposition 3.5, the lines l' and m', and thus l and m, are orthogonal to each other when $0 = \langle (1,a), (1,c)\rangle = 1+ac$, so when $ac = -1$.

Inspired by Proposition 3.5, we define orthogonality for vectors in $\mathbb{R}^n$ for all $n \geq 0$.

DEFINITION 3.7. Suppose F is a field contained in $\mathbb{R}$. Then we say that two vectors $v, w \in F^n$ are *orthogonal*, or *perpendicular* to each other, when $\langle v, w\rangle = 0$. Note that the zero vector is orthogonal to every vector.

WARNING 3.8. Proposition 3.1 implies that the only vector in $\mathbb{R}^n$ that is perpendicular to itself, is 0. Over other fields, however, we may have $\langle v, v\rangle = 0$ for nonzero v. For instance, the vector $v = (1, i) \in \mathbb{C}^2$ satisfies $\langle v, v\rangle = 0$, so in $\mathbb{C}^2$ we have $v \in \{v\}^\perp$. Also the vector $w = (1,1) \in \mathbb{F}_2^2$ satisfies $\langle w, w\rangle = 0$.

REMARK 3.9. If two vectors v and w in $\mathbb{R}^n$ are orthogonal, we sometimes write $v \perp w$. This explains the notation $S^\perp$ (see Definition 2.37) for $S \subset \mathbb{R}^n$, as the set

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

consists exactly of all elements that are orthogonal to all elements of S.

DEFINITION 3.10. Suppose F is a field contained in $\mathbb{R}$ and $a \in F^n$ a nonzero vector and $b \in F$ a constant. Then we say that a is a *normal* of the hyperplane

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

PROPOSITION 3.11. *Suppose F is a field contained in $\mathbb{R}$ and H a hyperplane with a normal a. Then for any $p, q \in H$, the vector $q - p$ is orthogonal to a. If H contains 0, then every $q \in H$ is orthogonal to a.*

PROOF. There is a constant $b \in F$ such that H consists exactly of all $x \in F^n$ with $\langle a, x\rangle = b$. This implies that for $p, q \in H$ we have $\langle a, q-p\rangle = \langle a, q\rangle - \langle a, p\rangle = b - b = 0$, so a is orthogonal to $q - p$. The last statement follows by taking $p = 0$. □

Because of Proposition 3.11, we say that a normal a of a hyperplane is orthogonal to that hyperplane. Beware though, as for hyperplanes not containing 0, it does not mean that a is orthogonal to the elements of H, but to the differences between elements. Draw a picture to clarify this for yourself!

EXAMPLE 3.12. Suppose $H \subset \mathbb{R}^n$ is a hyperplane with normal a, containing the point p. Then there is a constant b such that H consists of all points $x \in \mathbb{R}^n$ with $\langle a, x\rangle = b$. From $p \in H$ we obtain $b = \langle a, p\rangle$.

With Definitions 3.2 and 3.7 we immediately have the following analogon of Pythagoras' Theorem.

PROPOSITION 3.13. *Suppose F is a field contained in $\mathbb{R}$. Then two vectors $v, w \in F^n$ are orthogonal if and only if they satisfy $\|v - w\|^2 = \|v\|^2 + \|w\|^2$.*

PROOF. We have

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

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. □

Exercises

3.1.1. Prove Lemma 3.4.

3.1.2. 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 volume of the parallelepipid spanned by vectors $A, B, C \in \mathbb{R}^3$ equals $|\langle A, B \times C\rangle|$. (Does this equal $|\langle C, A \times B\rangle|$? If so, why?).

3.2. Orthogonal projections, distances, and reflections

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). However, before we can do that, we need to know that the value on the right-hand side of (1) lies in the interval $[-1, 1]$. We will use various results from this section to see that this is indeed the case in Proposition 3.29 of the next section.

The following proposition and its first corollary are the only results of this section that hold for all fields.

PROPOSITION 3.14. *Let F be any field, $n \geq 0$ an integer, and $a \in F^n$ an element with $\langle a, a\rangle \neq 0$. Then 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 $\langle w, w\rangle = \langle v, v\rangle - \frac{\langle a,v\rangle^2}{\langle a,a\rangle}$.*

PROOF. For any $\lambda \in F$, we have $\langle a, v - \lambda a\rangle = \langle a, v\rangle - \lambda\langle a, a\rangle$, so we have $\langle a, v - \lambda a\rangle = 0$ if and only if $\langle a, v\rangle = \lambda\langle a, a\rangle$, so if and only if $\lambda = \frac{\langle a,v\rangle}{\langle a,a\rangle}$. The dot products of λa and $w = v - \lambda a$ with themselves follow from

$$\langle \lambda a, \lambda a\rangle = \lambda^2\langle a, a\rangle$$

and

$$\langle w, w\rangle = \langle w, v - \lambda a\rangle = \langle w, v\rangle - \lambda\langle w, a\rangle = \langle v - \lambda a, v\rangle - 0 = \langle v, v\rangle - \lambda\langle a, v\rangle.$$

□

COROLLARY 3.15. *Let F be any field, $n \geq 0$ an integer, and $a \in F^n$ an element with $\langle a, a\rangle \neq 0$. Then the subspaces $L(a)$ and*

$$H_a = \{a\}^\perp = \{\, x \in F^n \;:\; \langle a, x\rangle = 0 \,\}$$

are complementary subspaces.

PROOF. Proposition 3.14 says that every $v \in F^n$ can be written uniquely as the sum of an element $\lambda a \in L(a)$ and an element w in the hyperplane $H_a = \{a\}^\perp$ given by $\langle a, x\rangle = 0$. By Lemma 2.48, the spaces $L(a)$ and H_a are complementary subspaces. Alternatively, we first only conclude $L(a) + H_a = F^n$ from Proposition 3.14. We also claim $L(a) \cap H_a = \{0\}$. Indeed, for $v = \lambda a \in L(a)$ we have $\langle v, a\rangle = \lambda\langle a, a\rangle$, so $\langle v, a\rangle = 0$ if and only if $\lambda = 0$, which means $v = 0$. □

COROLLARY 3.16. *Suppose F is a field contained in $\mathbb{R}$ and $a \in F^n$ is a vector. Then every element $v \in F^n$ can be written uniquely as a sum $v = v_1 + v_2$ of a multiple v_1 of a and an element v_2 that is orthogonal to a. Moreover, if a is nonzero, then we have $v_1 = \lambda a$ with $\lambda = \langle a, v\rangle \cdot \|a\|^{-2}$.*

PROOF. The statement is just a reformulation of Proposition 3.14 for $F \subset \mathbb{R}$, with $v_1 = \lambda a$ and $v_2 = w$. Indeed, for $a = 0$ the statement is trivial and for $a \neq 0$, we have $\langle a, a\rangle \neq 0$ by Proposition 3.1. □

DEFINITION 3.17. Using the same notation as in Corollary 3.16 and assuming a is nonzero, we call v_1 the *orthogonal projection* of v onto a or the line $L = L(a)$, and we call v_2 the orthogonal projection of v onto the hyperplane $H = \{a\}^\perp = L^\perp$.

LEMMA 3.18. *Suppose F is a field contained in $\mathbb{R}$ and $a, v \in F^n$ elements with a nonzero. Set $L = L(a)$ and $H = \{a\}^\perp = L^\perp$. Let $v_1 \in L$ and $v_2 \in H$ be the orthogonal projections of v on L and H respectively. Then for any $x \in L$ we have $\|v - x\| \geq \|v - v_1\| = \|v_2\|$ and for any $y \in H$ we have $\|v - y\| \geq \|v - v_2\| = \|v_1\|$.*

PROOF. We have $v = v_1 + v_2$. Suppose $x \in L$. we can write $v - x$ as the sum $(v - v_1) + (v_1 - x)$ of two orthogonal vectors, so that by Proposition 3.13 (Pythagoras) we have

$$\|v - x\|^2 = \|v - v_1\|^2 + \|v_1 - x\|^2 \geq \|v - v_1\|^2 = \|v_2\|^2.$$

This proves the first part of the statement. The second part follows similarly by writing $v - y$ as $(v - v_2) + (v_2 - y)$. □

Lemma 3.18 shows that the distance $\|v - x\|$ from v to any point $x \in L$ is at least the distance from v to the orthogonal projection v_1 of v on L. Similarly, the distance from v to any point in H is at least the distance from v to the orthogonal projection of v on H. This shows that the minimum in the following definition exists, at least if the line or hyperplane W contains 0. Of course the same holds when W does not contain 0, as we can translate W and v, so the definition makes sense.

DEFINITION 3.19. Suppose F is a field contained in $\mathbb{R}$. 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, i.e.,

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

One can in fact similarly define the distance from any vector $v \in F^n$ to any translate of any subspace of F^n.

PROPOSITION 3.20. *Suppose F is a field contained in $\mathbb{R}$ and $a, v \in F^n$ are elements with a nonzero. Let v_1 and v_2 be the orthogonal projections of v on $L(a)$ and $\{a\}^\perp$, respectively. Then we have*

$$d(v, \{a\}^\perp) = \|v_1\| = \frac{|\langle a, v\rangle|}{\|a\|} \qquad \text{and} \qquad d(v, L(a)) = \|v_2\| = \sqrt{\|v\|^2 - \tfrac{\langle a,v\rangle^2}{\|a\|^2}}.$$

PROOF. This follows immediately from Proposition 3.14, Corollary 3.16 and Lemma 3.18. Make a picture to support this! □

Note that $L(a)$ and $\{a\}^\perp$ are subspaces, so they contain 0. In order to find the distance to a line or hyperplane that does *not* contain 0, we first apply an appropriate translation to make sure the line or hyperplane *does* contain 0 (cf. Examples 3.23 and 3.24).

EXAMPLE 3.21. Take $a = (1, 1, 1) \in \mathbb{R}^3$. Then the hyperplane $H = \{a\}^\perp$ is the set

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

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

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

so we get $v_1 = 2a = (2, 2, 2)$ and thus $v_2 = v - v_1 = (2, 1, 3) - (2, 2, 2) = (0, -1, 1)$. Indeed, we have $v_2 \in H$. We find that the distance $d(v, L(a))$ from v to $L(a)$ equals $\|v_2\| = \sqrt{2}$ and the distance from v to H equals $d(v, H) = \|v_1\| = 2\sqrt{3}$.

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

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

and

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

Verify this and derive it yourself! Also find the distance from x to L and H in this general setting.

EXAMPLE 3.22. 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 will compute the distance from p to V. The normal $a = (1, -2, 3)$ of V satisfies $\langle a, a\rangle = 14$. We have $V = \{a\}^\perp$, so by Proposition 3.20, 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 3.23. 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 $\mathbb{R}^3$ with normal a. 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

$$W' = \{\, x \in \mathbb{R}^3 \;:\; \langle a, x - (-q)\rangle = 1 \,\} = \{\, x \in \mathbb{R}^3 \;:\; \langle a, x\rangle = 0 \,\} = \{a\}^\perp,$$

which does contain 0 (by construction, of course, because it is the image of $q \in W$ under the translation). Note the minus sign in the derived equation $\langle a, x-(-q)\rangle = 1$ for W' and make sure you understand why it is there! By Proposition 3.20, 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 3.24. 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 $q - p = (1, -1, -1)$. 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 Corollary 3.16. 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 its length is indeed $\frac{1}{3}\sqrt{42}$.

DEFINITION 3.25. Let $a \in \mathbb{R}^n$ be nonzero and set

$$H_a = \{a\}^\perp = \{\, x \in \mathbb{R}^n \;:\; \langle a, x\rangle = 0 \,\}.$$

Then for any $v \in \mathbb{R}^n$, we define the *reflection* of v in H_a to be

$$v' = v - 2\frac{\langle v, a\rangle}{\langle a, a\rangle}a.$$

Note that if we write $v = v_1 + v_2$ with v_1 a multiple of a and $v_2 \in H_a$, as in Corollary 3.16, then we have $v' = v_2 - v_1$; note also that $\langle v', a\rangle = \langle -v_1, a\rangle = -\langle v, a\rangle$, so the

reflection v'' of v' in H_a is v, as we have

$$v'' = v' - 2\frac{\langle v', a\rangle}{\langle a, a\rangle}a = v' + 2\frac{\langle v, a\rangle}{\langle a, a\rangle}a = v.$$

Draw a picture to see why v' is called the reflection of v and compare it with the following proposition.

PROPOSITION 3.26. *Let $a \in \mathbb{R}^n$ be nonzero and set $H_a = \{a\}^\perp$. Let $v \in \mathbb{R}^n$ be any vector and v' the reflection of v in H_a. Then the following statements hold.*

(1) *The vector $v - v'$ is orthogonal to H_a.*
(2) *The distances of v and v' to H_a are the same, i.e., $d(v, H_a) = d(v', H_a)$.*
(3) *If v is not contained in H_a, then v' is the unique point different from v itself that satisfies the two points above.*
(4) *If v is contained in H_a, then $v' = v$.*

PROOF. Exercise. □

EXAMPLE 3.27. Let $L \subset \mathbb{R}^2$ be the line given by $y = -2x$. Then $L = \{a\}^\perp$ for $a = (2, 1)$, i.e., a is a normal of L. The reflection of the point $p = (3, 4)$ in L is

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

Draw a picture to verify!

EXAMPLE 3.28. 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 V does not contain 0, so we first translate everything over $-p$ with $p = (0, 1, 0) \in V$. Then we get $\tilde{q} = q - p = (0, 2, 1)$ and

$$\tilde{V} = \{v - p \;:\; v \in V\} = \{a\}^\perp.$$

The reflection of $\tilde{q}$ in $\tilde{V}$ equals

$$\tilde{q}' = \tilde{q} - 2\frac{\langle \tilde{q}, a\rangle}{\langle a, a\rangle}a = \tilde{q} - 2 \cdot \frac{7}{14} \cdot a = \tilde{q} - a = (1, 0, -2).$$

Finally, to get the reflection q' of q in V, we have to translate back over p, so $q' = \tilde{q}' + p = (1, 1, -2)$.

Exercises

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

3.2.2.
(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.

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

3.2.4. 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 $v = (2, 1, 3)$.

3.2.5. 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 $|f(p) - b|$.

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

3.2.7. 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.

3.2.8. Prove Proposition 3.26.

3.2.9. 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, i.e.,

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

(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.

3.2.10. In this exercise, we generalize the notion of reflection to arbitrary fields. Let F be any field, $n \geq 0$ an integer, and $a \in F^n$ an element with $\langle a, a\rangle \neq 0$. Set

$$H_a = \{a\}^\perp = \{\, x \in F^n \;:\; \langle a, x\rangle = 0 \,\}.$$

Then for any $v \in F^n$, we define the *reflection* of v in H_a to be

$$v' = v - 2\frac{\langle v, a\rangle}{\langle a, a\rangle}a.$$

(1) Show that the reflection of v' in H_a equals v.
(2) Suppose that w' is the reflection of a vector $w \in F^n$ and x' is the reflection of the sum $x = v + w$. Show that $x' = v' + w'$. (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 also Examples 4.7.)

3.3. Cauchy-Schwarz

PROPOSITION 3.29 (Cauchy-Schwarz). *Suppose F is a field contained in $\mathbb{R}$ and $n \geq 0$ is an integer. Then for all $v, w \in F^n$ 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. For $v = 0$, we automatically have equality, as well as a nontrivial linear combination that is 0, namely with $\lambda = 1$ and $\mu = 0$. Suppose $v \neq 0$. Let z be the orthogonal projection of w onto $\{v\}^\perp$ (see Definition 3.17, so our vectors v, w, z correspond to a, v, v_2 of Proposition 3.16, respectively). Then by Corollary 3.16 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 nonnegative. 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 a nontrivial linear combination $\lambda v + \mu w = 0$ 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. □

PROPOSITION 3.30 (Triangle inequality). *Suppose F is a field contained in $\mathbb{R}$ and $n \geq 0$ is an integer. Then for all $v, w \in F^n$ we have $\|v + w\| \leq \|v\| + \|w\|$ and equality holds if and only if there are nonnegative scalars $\lambda, \mu \in F$, not both zero, such that $\lambda v = \mu w$.*

PROOF. By the inequality of Cauchy-Schwarz, Proposition 3.29, we have

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

Since all lengths are nonnegative, we may take square roots to find the desired inequality. The investigation of equality is left as an exercise. □

DEFINITION 3.31. Suppose F is a field contained in $\mathbb{R}$ and $n \geq 0$ is an integer. Then for all nonzero $v, w \in F^n$ we define the *angle* between v and w to be the unique real number $\alpha \in [0, \pi]$ that satisfies

$$\cos\alpha = \frac{\langle v, w\rangle}{\|v\| \cdot \|w\|}. \tag{2}$$

Note that the angle α between v and w is well defined, as by Proposition 3.29, the right-hand side of (2) lies between -1 and 1. The angle also corresponds with the usual notion of angle in $\mathbb{R}^2$ and $\mathbb{R}^3$ by Proposition 3.5. Finally, Definitions 3.7 and 3.31 imply that two nonzero vectors v and w in F^n are orthogonal if and only the angle between them is $\pi/2$.

EXAMPLE 3.32. 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 3.33. 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

3.3.1. Take $a = (-1, 2, 1) \in \mathbb{R}^3$ and set $V = \{a\}^\perp \subset \mathbb{R}^3$. Write the element $x = (x_1, x_2, x_3) \in \mathbb{R}^3$ as $x = x' + x''$ with $x \in L(a)$ and $x'' \in V$.

3.3.2. Finish the proof of Proposition 3.30.

3.3.3. Explain why Proposition 3.30 might be called the triangle inequality, which usually refers to $c \leq a + b$ for the sides a, b, c of a triangle. Prove that for all $v, w \in \mathbb{R}^n$ we have $\|v - w\| \leq \|v\| + \|w\|$. When does equality hold?

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

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

3.3.6. 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.

CHAPTER 4

Linear maps

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

We want to single out among all maps between two vector spaces V and W those that are 'compatible with the linear structure.'

DEFINITION 4.1. Let V and W be two F-vector spaces. A map $f : V \to W$ is called an *(F-)linear map* or a *homomorphism* if

(1) for all $v_1, v_2 \in V$, we have $f(v_1 + v_2) = f(v_1) + f(v_2)$,
(2) for all $\lambda \in F$ and all $v \in V$, we have $f(\lambda v) = \lambda f(v)$.

(Note: the first property states that f is a group homomorphism between the additive groups of V and 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 $f : V \to V$ is called an *endomorphism* of V; if f is in addition bijective, then it is called an *automorphism* of V.

Here are some simple properties of linear maps.

LEMMA 4.2. *Let U, V, W be vector spaces over a field F.*

(1) *If $f : V \to W$ is linear, then $f(0) = 0$.*
(2) *If $f : V \to W$ is an isomorphism, then the inverse map f^{-1} is also an isomorphism.*
(3) *If $f : U \to V$ and $g : V \to W$ are linear maps, then $g \circ f : U \to W$ is also linear.*

PROOF.

(1) This follows from either one of the two properties of linear maps:

$$f(0) = f(0 + 0) = f(0) + f(0) \Longrightarrow f(0) = 0$$

or

$$f(0) = f(0 \cdot 0) = 0 \cdot f(0) = 0\,.$$

(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 let $w_1, w_2 \in W$ 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 is checked in a similar way.

(3) Easy.

□

LEMMA 4.3. *Let $f : V \to W$ 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. □

Associated to a linear map there are two important linear subspaces: its kernel and its image.

DEFINITION 4.4. Let $f : V \to W$ be a linear map. Then the *kernel* of f is defined to be

$$\ker(f) = \{v \in V : f(v) = 0\}\,.$$

LEMMA 4.5. *Let $f : V \to W$ be a linear map.*

(1) *$\ker(f) \subset V$ is a linear subspace. More generally, if $U \subset W$ is a linear subspace, then $f^{-1}(U) \subset V$ is again a linear subspace; it contains $\ker(f)$.*
(2) *$\mathrm{im}(f) \subset W$ is a linear subspace. More generally, if $U \subset V$ is a linear subspace, then $f(U) \subset W$ is again a linear subspace; it is contained in $\mathrm{im}(f)$.*
(3) *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 $\lambda v \in \ker(f)$.

The more general statement is left as an exercise.

(2) We check again the subspace properties. We have $f(0) = 0 \in \mathrm{im}(f)$. If $w_1, w_2 \in \mathrm{im}(f)$, then there are $v_1, v_2 \in V$ such that $f(v_1) = w_1$, $f(v_2) = w_2$, hence $w_1 + w_2 = f(v_1 + v_2) \in \mathrm{im}(f)$. If λ is a scalar and $w \in \mathrm{im}(f)$, then there is $v \in V$ such that $f(v) = w$, hence $\lambda w = f(\lambda v) \in \mathrm{im}(f)$.

The more general statement is proved in the same way.

(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 $\ker(f) = \{0\}$.

Now assume that $\ker(f) = \{0\}$, and let $v_1, v_2 \in V$ 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 $v_1 = v_2$.

□

REMARK 4.6. If you want to show that a subset U in a vector space V is a linear subspace, it may be easier to find a linear map $f : V \to W$ such that $U = \ker(f)$ than to check the properties directly.

It is time for some examples.

EXAMPLES 4.7.

(1) Let V be any vector space. Then the unique map $f : V \to \{0\}$ into 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 $\{0\} \subset W$.

(2) For any vector space, 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.

(3) If $V = F^n$, then all the *projection maps* $\pi_j : F^n \to F$, $(x_1, \ldots, x_n) \mapsto x_j$ 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.)

(4) Suppose $V = \mathbb{R}^n$ and $a \in V$ is nonzero. Set $H_a = \{a\}^\perp$. Then the following maps from V to V are linear.

(a) The orthogonal projection $\pi_a \colon \mathbb{R}^n \to \mathbb{R}^n$ onto $L(a)$ given by

$$v \mapsto \frac{\langle v, a\rangle}{\langle a, a\rangle} a$$

(see Definition 3.17). Indeed, linearity follows from the identities $\langle v + w, a\rangle = \langle v, a\rangle + \langle w, a\rangle$ and $\langle \lambda v, a\rangle = \lambda\langle v, a\rangle$. Note that for the $a = e_j$, the j-th standard vector, and the projection map $\pi_j \colon \mathbb{R}^n \to \mathbb{R}$ on the j-th coordinate, we have

$$\pi_{e_j}(v) = \pi_j(v) \cdot e_j.$$

(b) The orthogonal projection $\pi_{a^\perp} \colon \mathbb{R}^n \to \mathbb{R}^n$ onto $H_a = \{a\}^\perp$ given by

$$v \mapsto v - \frac{\langle v, a\rangle}{\langle a, a\rangle} a = v - \pi_a(v)$$

(see Definition 3.17). 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.

(c) The reflection $s_a \colon \mathbb{R}^n \to \mathbb{R}^n$ in the hyperplane $H_a = \{a\}^\perp$ given by

$$v \mapsto v - 2\frac{\langle v, a\rangle}{\langle a, a\rangle} a$$

(see Definition 3.25). The linearity is proven in the same way as for the projection onto H_a. The remark under Definition 3.25 shows that $s_a \circ s_a = \mathrm{id}_V$.

(5) For any two vector spaces V_1, V_2 over the same field 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 1.2.10.

(6) Let P be the vector space of polynomial functions on $\mathbb{R}$. Then the following maps are linear.

(a) Evaluation: given $a \in \mathbb{R}$, the map $\mathrm{ev}_a : P \to \mathbb{R}$, $p \mapsto p(a)$ is linear. The kernel of ev_a consists of all polynomials having a zero at a; the image is all of $\mathbb{R}$.

(b) Differentiation: $D : P \to P$, $p \mapsto p'$ is linear.
The kernel of D consists of the constant polynomials; the image of D is P (see below).

(c) Definite integration: given $a < b$, the map

$$I_{a,b} : P \longrightarrow \mathbb{R}\,, \quad p \longmapsto \int_a^b p(x)\,dx$$

is linear.

(d) Indefinite integration: given $a \in \mathbb{R}$, the map

$$I_a : P \longrightarrow P\,, \quad p \longmapsto \Big(x \mapsto \int_a^x p(t)\,dt\Big)$$

is linear. This map is injective; its image is the kernel of ev_a (see below).

(e) Translation: given $a \in \mathbb{R}$, the map

$$T_a : P \longrightarrow P\,, \quad p \longmapsto \big(x \mapsto p(x+a)\big)$$

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 $p \in P$ we have $(I_{a,b} \circ D)(p) = p(b) - p(a)$ and $(I_a \circ D)(p) = p - p(a)$. 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 $\mathrm{im}(I_a) = \ker(\mathrm{ev}_a)$.

The relation $D \circ I_a = \mathrm{id}_P$ implies that I_a is injective and that D is surjective. Let $C \subset P$ be the subspace of constant polynomials, and let $Z_a \subset P$ 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. D restricts to an isomorphism $Z_a \xrightarrow{\sim} P$, and I_a restricts (on the target side) to an isomorphism $P \xrightarrow{\sim} Z_a$ (Exercise!).

Exercises

4.1.1. Prove Lemma 4.3.

4.1.2. 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) $P \to P$, $f \mapsto f'$, where P is the vector space of real polynomials and f' is the derivative of f,
(6) $P \to \mathbb{R}^2$, $f \mapsto (f(2), f'(0))$.

4.1.3. 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 $L(f(S)) = W$.

4.1.4. 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.
(2) What are the images $\rho((1,0))$ and $\rho((0,1))$?
(3) Show that we have

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

4.1.5. 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.

4.1.6. Let F be a field and $F[x]$ the vector space of polynomials over F.
(1) Given $a \in F$, 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 $f(a) = \sum_{i=0}^{d} c_i a^i$. Cf. Example 4.7(6). Show that ev_a is linear.
(2) Show that the map $\varphi\colon F[x] \to F^F$ of Exercise 2.3.8 is given by

$$f \mapsto (a \mapsto \mathrm{ev}_a(f))$$

and deduce that φ is linear.

4.1.7. 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 $2y - x = 0$.

4.1.8. 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. Characterising linear maps

PROPOSITION 4.8. *Let F be any field and n a nonnegative 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 2.10. □

PROPOSITION 4.9. *Let F be any field and n a nonnegative integer. Suppose $f\colon F^n \to F$ is a linear map. Then there is a unique vector $a \in F^n$ such that for all $x \in F^n$ we have $f(x) = \langle a, x\rangle$.*

PROOF. Suppose there exists such an element a and write $a = (a_1, a_2, \ldots, a_n)$. 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 $a = \big(f(e_1), f(e_2), \ldots, f(e_n)\big)$, so a is completely determined by f and therefore unique, if it exists.

To show there is indeed an a as claimed, we take

$$a = \big(f(e_1), f(e_2), \ldots, f(e_n)\big)$$

(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 $x = (x_1, x_2, \ldots, x_n)$, then we find

$$f(x) = f(x_1 \cdot e_1 + \cdots + x_n \cdot e_n) = x_1 \cdot f(e_1) + \cdots + x_n \cdot f(e_n) = \langle x, a\rangle = \langle a, x\rangle.$$

□

PROPOSITION 4.10. *Let F be a field and W be an F-vector space. Then for every sequence $w_1, w_2, \ldots, w_n$ of n vectors in W, there is a unique linear map $\varphi\colon F^n \to W$ with $\varphi(e_i) = w_i$ for every $i \in \{1, \ldots, n\}$.*

PROOF. Suppose f is a function with $f(e_i) = w_i$ for every $i \in \{1, \ldots, n\}$. Then for $x = (x_1, x_2, \ldots, x_n) \in F^n$ we have

$$f(x) = f(x_1e_1 + \cdots + x_ne_n) = x_1f(e_1) + \cdots + x_1f(e_n) = x_1w_1 + \cdots + x_nw_n,$$

so f is completely determined on all $x \in F^n$ by the vectors $w_1, w_2, \ldots, w_n$ 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

$$\varphi(x) = x_1w_1 + \cdots + x_nw_n$$

(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. □

By construction, the image of the map φ of Proposition 4.10 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 4.11. For any vector space W over a field F, 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 4.10.

Exercises

4.2.1. 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.
(2) Show that

$$U_+ = \{v \in \mathbb{R}^3 \;:\; s(v) = v\} \qquad \text{and} \qquad U_- = \{v \in \mathbb{R}^3 \;:\; s(v) = -v\}$$

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.

4.2.2. This exercise generalizes Exercises 2.4.6 and 4.2.1. Let V be a vector space over a field F and assume $2 \neq 0$ in F, so that and we can divide by 2. 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

$$V_+ = \{\, v \in V \;:\; s(v) = v \,\}, \qquad V_- = \{\, v \in V \;:\; s(v) = -v \,\}.$$

(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 V_+ and V_- are complementary subspaces in V.
(4) For what choice of s does Exercise 2.4.6 become a special case?

4.2.3. Let V be a vector space over a field F. Suppose $f_1, \dots, f_m$ are linear maps from V to F. Show that the combined map

$$V \to F^m, \quad v \mapsto (f_1(x), \dots, f_m(x))$$

is also linear.

4.2.4. Suppose V is a vector space with two complementary subspaces U and U', cf. Definition 2.46. Then for every $v \in V$ there are unique elements $u \in U$ and $u' \in U'$ with $v = u + u'$ by Lemma 2.48; let $\pi_U \colon V \to U$ 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. Show that π_U is a surjective linear map with kernel $\ker \pi_U = U'$ that satisfies $\pi_U \circ \pi_U = \pi_U$. We call π_U the projection of V onto U along U'.

4.2.5. This exercise generalizes Proposition 4.10. Let F be a field and X a (not necessarily finite) set. Consider the subspace $F^{(X)}$ of F^X as in Exercise 2.1.12, and the elements e_x (for $x \in X$) as in Exercise 2.3.5. Let W be a vector space over F containing a collection $C = (w_x)_{x \in X}$ of elements in W. 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 collection C generates W.

4.3. Linear maps form a vector space

One nice property of linear maps is that they are themselves elements of vector spaces.

LEMMA 4.12. *Let V and W be two F-vector spaces. Then the set of all linear maps $V \to W$, with addition and scalar multiplication defined point-wise, forms an F-vector space. It is denoted by* $\operatorname{Hom}(V, W)$.

PROOF. It is easy to check the vector space axioms for the set of *all* maps $V \to W$ (using the point-wise definition of the operations and the fact that W is a vector space). Hence it suffices to show that the linear maps form a linear subspace:

The zero map is a homomorphism. If $f, g : V \to W$ are two linear maps, we have to check that $f + g$ is again linear. So let $v_1, v_2 \in V$; then

$$\begin{aligned}(f+g)(v_1+v_2) &= f(v_1+v_2) + g(v_1+v_2) = f(v_1) + f(v_2) + g(v_1) + g(v_2) \\ &= f(v_1) + g(v_1) + f(v_2) + g(v_2) = (f+g)(v_1) + (f+g)(v_2)\,.\end{aligned}$$

Similarly, if $\lambda \in F$ and $v \in V$, we have

$$(f+g)(\lambda v) = f(\lambda v) + g(\lambda v) = \lambda f(v) + \lambda g(v) = \lambda\big(f(v) + g(v)\big) = \lambda \cdot (f+g)(v)\,.$$

It follows that $f + g$ is linear. Now let $\mu \in F$, and let $f : V \to W$ be linear. We have to check that μf is again linear. So let $v_1, v_2 \in V$; then

$$\begin{aligned}(\mu f)(v_1+v_2) &= \mu f(v_1+v_2) = \mu\big(f(v_1) + f(v_2)\big) \\ &= \mu f(v_1) + \mu f(v_2) = (\mu f)(v_1) + (\mu f)(v_2)\,.\end{aligned}$$

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. □

EXAMPLE 4.13. Suppose $V = \mathbb{R}^n$ and $a \in V$ is nonzero. Set $H_a = \{a\}^\perp$. Let π_a, $\pi_{a^\perp}$, and s_a be the orthogonal projection onto $L(a)$, the orthogonal projection onto $\{a\}^\perp$, and the reflection in H_a, respectively, as in Examples 4.7. Then the linearity of the last two maps follows from the linearity of the first, as we have

$$\pi_{a^\perp} = \mathrm{id}_V - \pi_a, \qquad \text{and} \qquad s_a = \mathrm{id}_V - 2\pi_a.$$

Two isomorphic vector spaces can for all practical purposes be identified. This is illustrated by the following proposition.

PROPOSITION 4.14. *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*

$$\begin{array}{ccc} V & \xrightarrow{f} & W \\ {\scriptstyle\varphi}\downarrow & & \downarrow{\scriptstyle\psi} \\ V' & \xrightarrow{f'} & W' \end{array}$$

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. □

Exercises

4.3.1. Suppose we have a diagram

$$\begin{array}{ccc} V & \xrightarrow{f} & W \\ {\scriptstyle\varphi}\downarrow & & \downarrow{\scriptstyle\psi} \\ V' & \xrightarrow{f'} & W' \end{array}$$

of linear maps that commutes, i.e., 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 4.14.

4.3.2. Let F be a field and n a nonnegative integer. Show that there is an isomorphism

$$F^n \to \operatorname{Hom}(F^n, F)$$

that sends a vector $a \in F^n$ to the linear map $x \mapsto \langle a, x\rangle$.

4.3.3. Let F be field. The dot product on F^n is a map $F^n \times F^n \to F$, satisfying some conditions. In this exercise, we will generalize this to F^X for any set X. Note that if X is finite, then F^X and $F^{(X)}$ as in Exercise 2.1.12 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 generalized dot product satisfies the conditions of Proposition 2.10.
(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 $g \mapsto \langle f, g\rangle$.

4.3.4. Let $f\colon V \to V$ be an endomorphism of a finitely generated vectorspace V. Let $\sigma\colon V \to W$ be a linear map. Suppose that f sends $\ker\sigma$ to itself, i.e., $f(\ker\sigma) \subset \ker\sigma$. 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 $z \in V$.

4.3.5. Let V be a vector space and $\sigma\colon X \to Y$ any map of sets. Define the map
$$\sigma^*\colon V^Y = \mathrm{Map}(Y, V) \to \mathrm{Map}(X, V) = V^X$$
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.

4.3.6.
(1) Suppose $\alpha\colon W \to W'$ is a linear map of vector spaces over a field F. Show that for every vector space V over F there is a linear map
$$\alpha_*\colon \mathrm{Hom}(V, W) \to \mathrm{Hom}(V, W')$$
that sends f to $\alpha \circ f$.
(2) Suppose $\beta\colon V' \to V$ is a linear map of vector spaces over a field F. Show that for every vector space W over F there is a linear map
$$\beta^*\colon \mathrm{Hom}(V, W) \to \mathrm{Hom}(V', W)$$
that sends f to $f \circ \beta$.
(3) Check that in Proposition 4.14 we have
$$f' = (\psi_* \circ (\varphi^{-1})^*)(f) = ((\varphi^{-1})^* \circ \psi_*)(f).$$

4.3.7. Suppose $\alpha, \alpha_1, \alpha_2\colon V' \to V$ and $\beta\colon V'' \to V'$ are linear maps of vector spaces over a field F. Let W be a vector space over F. With the notation of Exercise 4.3.6, show that we have the following.
(1) Show that $(\alpha \circ \beta)^* = \beta^* \circ \alpha^*$.
(2) Show that $(\alpha_1 + \alpha_2)^* = \alpha_1^* + \alpha_2^*$.
(3) Show that $(\lambda\alpha)^* = \lambda \cdot \alpha^*$ for any $\lambda \in F$.

CHAPTER 5

Matrices

By Proposition 4.10, every linear map $\varphi\colon F^n \to F^m$ 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 4.11.

From a different viewpoint, we can interpret $\varphi\colon F^n \to F^m$ as a sequence of m linear maps, one for each coordinate of F^m. More formally, if $\pi_i\colon F^m \to F$ is the projection of F^m onto its i-th coordinate ($1 \leq i \leq m$), then the composition $\pi_i \circ \varphi\colon F^n \to F$, which only remembers the i-th coordinate, is a linear map for each i, and φ equals the combined map

$$F^n \to F^m, \quad x \mapsto \big((\pi_1 \circ \varphi)(x), \ldots, (\pi_m \circ \varphi)(x)\big)$$

(cf. Exercise 4.2.3). Each of the m maps $\pi_i \circ \varphi\colon F^n \to F$ is given by $x \mapsto \langle v_i, x\rangle$ for some $v_i \in F^n$ (see Proposition 4.9), so φ is determined by the m vectors $v_1, \ldots, v_m \in F^n$.

We will see 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!

$$\begin{pmatrix} | & | & & | \\ w_1 & w_2 & \cdots & w_n \\ | & | & & | \end{pmatrix} = \begin{pmatrix} -v_1- \\ -v_2- \\ \vdots \\ -v_m- \end{pmatrix} \tag{3}$$

Such an array is called an $m \times n$ matrix. The linear map φ is completely determined by its associated matrix, say M. In the next section, we will describe how exactly: we will define a product between $m \times n$ matrices and vectors in F^n in such a way that for all $x \in F^n$ we have $\varphi(x) = M \cdot x$ (see Section 5.2). It will make immediately clear why the two matrices in (3) are the same.

5.1. Definition of matrices

Definition 5.1. Let F be a field and m, n nonnegative integers. An *$m \times n$ matrix over F* is an array

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} = (a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}$$

of *entries* or *coefficients* $a_{ij} \in F$.

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 above.

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*.

For any

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \in \mathrm{Mat}(m \times n, F) \qquad \text{and} \qquad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in F^n$$

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 we have written x and Ax vertically.

As announced in the introduction of this chapter, there are (at least) two useful ways to think of the multiplication. If we let

$$v_i = (a_{i1}, a_{i2}, \dots, a_{in})$$

be the i-th row of A, then we can write Ax as

$$Ax = \begin{pmatrix} -v_1- \\ -v_2- \\ \vdots \\ -v_m- \end{pmatrix} \cdot x = \begin{pmatrix} \langle v_1, x \rangle \\ \langle v_2, x \rangle \\ \vdots \\ \langle v_m, x \rangle \end{pmatrix} ,$$

so the entries of Ax are the dot-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, then we can write Ax as

$$Ax = \begin{pmatrix} | & | & & | \\ w_1 & w_2 & \cdots & w_n \\ | & | & & | \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1 w_1 + x_2 w_2 + \ldots x_n w_n,$$

so Ax is the linear combination of the column vectors of A with the entries of x as coefficients. Note that $Ae_j = w_j$.

EXAMPLE 5.2. 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 dot 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.

5.2. Linear maps associated to matrices

DEFINITION 5.3. 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 $x \in F^n$.

The dual description of matrices in the introduction of this chapter corresponds to the following two lemmas. They also imply the equality of the two matrices in (3).

LEMMA 5.4. *Let A be an $m \times n$ matrix over F with rows $v_1, \ldots, v_m \in F^n$. Let $f_A \colon F^n \to F^m$ be the function associated to A. Then we have*

$$f_A(x) = \begin{pmatrix} \langle v_1, x \rangle \\ \langle v_2, x \rangle \\ \vdots \\ \langle v_m, x \rangle \end{pmatrix}$$

for all $x \in F^n$.

PROOF. This follows immediately from first of the two useful ways to think of the matrix multiplication described at the end of the previous section. □

LEMMA 5.5. *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 $1 \leq j \leq n$, and f_A equals φ_C as in Definition 4.11 with $C = (w_1, \ldots, w_n)$.*

PROOF. This follows immediately from second of the two useful ways to think of the matrix multiplication described at the end of the previous section. □

Note that Lemma 5.5 implies that for any $m \times n$ matrix A, the j-th column of A equals $f_A(e_j)$ for any $j \in \{1, \ldots, n\}$. In fact, 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.

LEMMA 5.6. *For any matrix $A \in \mathrm{Mat}(m \times n, F)$, the associated function $f_A \colon F^n \to F^m$ is a linear map.*

PROOF. By Lemma 5.5, the function f_A equals the linear map φ_C of Definition 4.11, where $C = (w_1, \ldots, w_n)$ is the sequence of columns of A. Alternatively, let $v_1, v_2, \ldots, v_m$ denote the row vectors of A. Then for each i, the map $F^n \to F$ that sends x to $\langle v_i, x\rangle$ is linear by Proposition 4.8. By Lemma 5.4, the function f_A is the combination of these m maps as in Exercise 4.2.3, so it is linear as well. □

Clearly, the linear map f_I associated to the matrix $I = I_n$ is the identity map $F^n \to F^n$.

EXAMPLE 5.7. 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 5.8. *Let F be a field and m, n nonnegative 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. Define $w_j = f(e_j)$ for $1 \leq j \leq n$, 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 4.10. 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. □

Lemma 5.6 and Proposition 5.8 together show that there is a bijection

$$\mathrm{Mat}(m \times n, F) \to \mathrm{Hom}(F^n, F^m), \qquad A \mapsto f_A.$$

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 $\operatorname{im} A = \operatorname{im} f_A$.

EXAMPLE 5.9. Let $\rho\colon \mathbb{R}^2 \to \mathbb{R}^2$ be rotation about the origin $(0, 0)$ over an angle θ. From Exercise 4.1.4, 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 5.10. Let $s\colon \mathbb{R}^2 \to \mathbb{R}^2$ 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 5.5, 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 projection of v onto a equals λa with $\lambda = \frac{\langle v,a\rangle}{\langle a,a\rangle}$, so the projection of v onto L is $v - \lambda a$ and the reflection of v in L is $s(v) = v - 2\lambda a$. (Make a picture!) We find

$$s(e_1) = \begin{pmatrix} -\frac{3}{5} \\ \frac{4}{5} \end{pmatrix} \qquad \text{and} \qquad s(e_2) = \begin{pmatrix} \frac{4}{5} \\ \frac{3}{5} \end{pmatrix}$$

(do the calculations yourself), so we get

$$A = \begin{pmatrix} -\frac{3}{5} & \frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{pmatrix}.$$

DEFINITION 5.11. 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.

REMARK 5.12. The column space of a matrix $A \in \mathrm{Mat}(m \times n, F)$ is the same as the image of A, i.e., the image of the linear map f_A.

PROPOSITION 5.13. *Let $A \in \mathrm{Mat}(m \times n, F)$ be a matrix. Then we have*

$$\ker A = (R(A))^\perp \subset F^n.$$

For $F = \mathbb{R}$, the kernel of A consists of all vectors in $\mathbb{R}^n$ that are orthogonal to the row space $R(A)$ of A.

PROOF. Let $v_1, v_2, \ldots, v_m$ 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 $x \in F^n$ (see Lemma 5.4; here we have written $f_A(x)$ normally instead of vertically). Thus, we have $x \in \ker A = \ker f_A$, i.e., $f_A(x) = 0$, 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 2.39(2)). We conclude $\ker A = (R(A))^\perp$, as stated. The last statement is merely a rephrasing of this equality for $F = \mathbb{R}$. □

REMARK 5.14. Let $U \subset F^n$ be a subspace of F^n. We can use Proposition 5.13 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 $i \leq m$.

Exercises

5.2.1. Prove Lemmas 5.4 and 5.5.

5.2.2. Prove Remark 5.12.

5.2.3. For the given matrix A and the vector x, determine Ax.

(1)

$$A = \begin{pmatrix} -2 & -3 & 1 \\ 1 & 1 & -2 \\ 0 & 1 & 1 \end{pmatrix} \qquad \text{and} \qquad x = \begin{pmatrix} -3 \\ -4 \\ 2 \end{pmatrix},$$

(2)

$$A = \begin{pmatrix} 1 & -3 & 2 \\ -2 & -4 & 2 \end{pmatrix} \qquad \text{and} \qquad x = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix},$$

(3)
$$A = \begin{pmatrix} 4 & 3 \\ 3 & -2 \\ -3 & -1 \\ -1 & 1 \end{pmatrix} \qquad \text{and} \qquad x = \begin{pmatrix} -2 \\ 3 \end{pmatrix}.$$

5.2.4. For each of the linear maps $f\colon F^n \to F^m$ of the exercises of Section 4.1, give a matrix M such that f is given by
$$x \mapsto Mx.$$

5.2.5. 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 $v_1, \ldots, v_n$ such that f is also given by
$$f\big((x_1, x_2, \ldots, x_n)\big) = x_1v_1 + \cdots + x_nv_n.$$

5.2.6. Determine the matrix M for which $f_M\colon \mathbb{R}^3 \to \mathbb{R}^3$ is reflection in the plane given by $x + 2y - z = 0$.

5.2.7. 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 $w_3 = (-2, 4)$.

5.3. Addition and multiplication of matrices

We know that $\mathrm{Hom}(F^n, F^m)$ has the structure of an F-vector space (see Lemma 4.12). We can 'transport' this structure to $\mathrm{Mat}(m \times n, F)$ using the identification of matrices and linear maps.

Definition 5.15. 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 $(a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij})$, i.e., that addition of matrices is done coefficient-wise. Similarly, we see easily that $\lambda(a_{ij}) = (\lambda a_{ij})$. 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 (i.e., isomorphic to) F^{mn} — the only difference is the arrangement of the coefficients in an array instead of in a sequence.

By Lemma 4.2, the composition of two linear maps is again linear. How is this reflected in terms of matrices?

Definition 5.16. 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 $f_A \circ f_B\colon F^n \xrightarrow{B} F^m \xrightarrow{A} F^l$. So AB will be a matrix in $\mathrm{Mat}(l \times n, F)$.

Remark 5.17. 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 5.16, the product AB satisfies $f_{AB} = f_A \circ f_B$, so we have

$$(AB)x = f_{AB}(x) = f_A(f_B(x)) = A(Bx) \tag{4}$$

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 (4) holds in particular for $x = e_k$, the k-th standard vector. 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 dot 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 dot product of the i-th row of A next to this entry and the k-th column of B above the entry.

$$\begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix} = B \tag{5}$$

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{l1} & a_{l2} & \cdots & a_{lm} \end{pmatrix} \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{l1} & c_{l2} & \cdots & c_{ln} \end{pmatrix} = AB$$

EXAMPLE 5.18. 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 diagonally with respect to each other.

$$\begin{array}{ccc} & & \begin{pmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \\ 14 & 16 & 18 \\ 20 & 22 & 24 \end{pmatrix} \\ \begin{pmatrix} 1 & 3 & 5 & 7 \\ 9 & 11 & 13 & 15 \end{pmatrix} & \begin{pmatrix} . & 268 & . \\ . & . & . \end{pmatrix} & \end{array}$$

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 dot 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

$$AB = \begin{pmatrix} 236 & 268 & 300 \\ 588 & 684 & 780 \end{pmatrix}.$$

PROPOSITION 5.19. *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 5.8. □

PROPOSITION 5.20. *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 5.21. If the linear map f_A corresponding to $A \in \mathrm{Mat}(m \times n, F)$ is an isomorphism, then A is called *invertible.*

The matrix corresponding to the inverse linear map is (obviously) denoted A^{-1}, so that $f_{A^{-1}} = f_A^{-1}$; if A is an $m \times n$ matrix, then we have $AA^{-1} = I_m$ and $A^{-1}A = I_n$, and A^{-1} is uniquely determined by this property. We will see in Corollary 6.56 that if $A \in \mathrm{Mat}(m \times n, F)$ is invertible, then $m = n$, so A is in fact a square matrix.

PROPOSITION 5.22. *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 $CA = I_n$.*

PROOF. Exercise. □

PROPOSITION 5.23. *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. Exercise. □

REMARK 5.24. The identity matrix acts as a multiplicative identity:

$$I_m A = A = AI_n \quad \text{for } A \in \mathrm{Mat}(m \times n, F).$$

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 5.3.10.

DEFINITION 5.25. Let $A = (a_{ij}) \in \mathrm{Mat}(m \times n, F)$ be a matrix. The *transpose* of A is the matrix

$$A^\top = (a_{ji})_{1 \leq i \leq n, 1 \leq j \leq m} \in \mathrm{Mat}(n \times m, F)\,.$$

EXAMPLE 5.26. For

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{pmatrix}$$

we have

$$A^\top = \begin{pmatrix} 1 & 5 & 9 \\ 2 & 6 & 10 \\ 3 & 7 & 11 \\ 4 & 8 & 12 \end{pmatrix}\,.$$

PROPOSITION 5.27. *Let F be a field, and l, m, n nonnegative 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 $v_1, \dots, v_l$ be the rows of A and $w_1, \dots, w_n$ 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 $w_1, \dots, w_n$ and the columns of $A^\top$ are $v_1, \dots, v_l$, 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 5.3.10. The fourth statement follows from the third. □

REMARK 5.28. We have expressed the product AB of matrices A and B in terms of the dot products of the rows of A and the columns of B. Conversely, we can interpret the dot 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 $1 \times n$ matrices (implicitly using that F^n and $\mathrm{Mat}(1 \times n, F)$ are isomorphic). Then the transpose $b^\top$ is an $n \times 1$ matrix and the matrix product

$$a \cdot b^\top = \begin{pmatrix} a_1 & a_2 & \dots & a_n \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = (a_1 b_1 + \dots + a_n b_n)$$

is the 1×1 matrix whose single entry equals the dot product $\langle a, b \rangle$.

REMARK 5.29. The product Ax of a matrix $A \in \mathrm{Mat}(m \times n, F)$ and a vector $x \in F^n$ can be interpreted as a product between matrices as well. If we think of x as a $1 \times n$ matrix, then $x^\top$ is an $n \times 1$ matrix and the product Ax corresponds to the matrix product $A \cdot x^\top$.

Exercises

5.3.1. Prove Proposition 5.23. If matrices A and B have a product AB that is invertible, does this imply that A and B are invertible? Cf. Exercise 7.4.4.

5.3.2. Prove Proposition 5.20.

5.3.3. Let $\rho\colon \mathbb{R}^2 \to \mathbb{R}^2$ be rotation around 0 over an angle α, cf. Exercise 4.1.4 and Example 5.9. Show that the matrix

$$A = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$$

satisfies $\rho(v) = Av$ for all $v \in \mathbb{R}^2$. Show that for all $\alpha, \beta \in \mathbb{R}$ we have

$$\cos(\alpha + \beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta,$$
$$\sin(\alpha + \beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta.$$

5.3.4. For which $i, j \in \{1, \dots, 5\}$ does the product of 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.

5.3.5. 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 the previous exercise.

(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?

5.3.6. Take the linear maps f and g of Exercise 5.2.7 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.

5.3.7. This exercise proves Proposition 5.22. Let A be an $m \times n$ matrix over a field F.
(1) Show that if there exists a matrix B such that $AB = I_m$, then f_A is surjective.
(2) Show that if there exists a matrix C such that $CA = I_n$, then f_A is injective.
(3) Show that if there exist matrices B and C such that $AB = I_m$ and $CA = I_n$, then f_A is an isomorphism and $B = C$.
(4) Show that if f_A is an isomorphism, then there exist matrices B and C such that $AB = I_m$ and $CA = I_n$.

5.3.8. Let F be a field and m, n nonnegative integers. Show that there exists an isomorphism

$$\mathrm{Mat}(m \times n, F) \to \mathrm{Hom}(F^n, F^m)$$

that sends A to f_A. (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.)

5.3.9. Let F be a field and m, n nonnegative integers. Some of the previous two sections can be summarized by the following diagram.

$$\begin{array}{ccccc} (F^n)^m & \longrightarrow & \mathrm{Mat}(m \times n, F) & \longleftarrow & (F^m)^n \\ & \searrow & \downarrow & \swarrow & \\ & & \mathrm{Hom}(F^n, F^m) & & \end{array}$$

Describe a natural isomorphism for each arrow, making the diagram commutative.

5.3.10. Let F be a field and m, n nonnegative 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 Proposition 4.9 and Exercise 4.3.2). 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 4.3.5), and the linear map $f^\top = \varphi_n^{-1} \circ f^* \circ \varphi_m \colon F^m \to F^n$.

$$\begin{array}{ccc} \mathrm{Hom}(F^m, F) & \xrightarrow{f^*} & \mathrm{Hom}(F^n, F) \\ {\scriptstyle \varphi_m}\uparrow & & \uparrow{\scriptstyle \varphi_n} \\ F^m & \xrightarrow[f^\top]{} & F^n \end{array}$$

Let A be an $m \times n$ matrix with rows $v_1, \ldots, v_m$, and let $f_A \colon F^n \to F^m$ be the associated linear map. Let $j \in \{1, \ldots, m\}$.
(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, i.e., $f_A^\top = f_{A^\top}$.
(5) Use Exercise 4.3.7 to prove Proposition 5.27.

5.3.11. (infinite matrices) An $m \times n$ matrix over a field F can be viewed as 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 matrix 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 4.3.3.
(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 $N \in \mathrm{Mat}(W \times X)$ 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.

5.4. Elementary row and column operations

Matrices are very suitable for doing computations. The main tool for that are the so-called 'elementary row and column operations.'

Definition 5.30. 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 5.31. *If M and M' are row equivalent matrices, then we have*

$$R(M) = R(M') \qquad \text{and} \qquad \ker M = \ker M'.$$

PROOF. Exercise. □

PROPOSITION 5.32. *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, …, B_r be the elementary matrices corresponding to the row operations we have performed on A to obtain A', then

$$A' = B_r\Big(B_{r-1} \cdots \big(B_2(B_1 A)\big) \cdots \Big) = (B_r B_{r-1} \cdots B_2 B_1)A\,,$$

and $B = B_r B_{r-1} \cdots B_2 B_1$ is invertible as 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 5.33. *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 is 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*

$$\begin{array}{ccc} F^n & \xrightarrow{f_A} & F^m \\ {\scriptstyle\varphi}\uparrow & & \downarrow{\scriptstyle\psi} \\ F^n & \xrightarrow[f_{A'}]{} & F^m \end{array}$$

is commutative.

PROOF. Exercise. □

COROLLARY 5.34. *Let M and M' be row equivalent matrices. Then f_M is injective if and only if $f_{M'}$ is injective and f_M is surjective if and only if $f_{M'}$ is surjective.*

PROOF. By Proposition 5.33 there is an isomorphism ψ with $f_{M'} = \psi \circ f_M$. Indeed, the composition is surjective or injective if and only if f_M is, cf. Proposition 4.14. □

Exercises

5.4.1. Let $v_1, v_2, \ldots, v_m \in \mathbb{R}^n$ be m vectors and consider the $m \times n$ matrix M whose rows are these vectors. Let M' be a matrix that is row equivalent to M. Use Exercise 2.3.9 to show that for the rows $v'_1, v'_2, \ldots, v'_m$ of M' we have $L(v_1, \ldots, v_m) = L(v'_1, \ldots, v'_m)$.

5.4.2. Prove Proposition 5.31.

5.4.3. Show that column equivalent matrices have the same column space, cf. Proposition 5.31.

5.4.4. 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.

$$A_0 = \begin{pmatrix} 2 & 5 & 4 & -3 & 1 \\ 1 & 3 & -2 & 2 & 1 \\ 0 & 4 & -1 & 0 & 3 \\ -1 & 2 & 2 & 3 & 1 \end{pmatrix} \qquad A_1 = \begin{pmatrix} 1 & 3 & -2 & 2 & 1 \\ 2 & 5 & 4 & -3 & 1 \\ 0 & 4 & -1 & 0 & 3 \\ -1 & 2 & 2 & 3 & 1 \end{pmatrix}$$

$$A_2 = \begin{pmatrix} 1 & 3 & -2 & 2 & 1 \\ 0 & -1 & 8 & -7 & -1 \\ 0 & 4 & -1 & 0 & 3 \\ 0 & 5 & 0 & 5 & 2 \end{pmatrix} \qquad A_3 = \begin{pmatrix} 1 & 3 & -2 & 2 & 1 \\ 0 & -1 & 8 & -7 & -1 \\ 0 & 0 & 31 & -28 & -1 \\ 0 & 0 & 40 & -30 & -3 \end{pmatrix}$$

$$A_4 = \begin{pmatrix} 1 & 3 & -2 & 2 & 1 \\ 0 & -1 & 8 & -7 & -1 \\ 0 & 0 & 31 & -28 & -1 \\ 0 & 0 & 9 & -2 & -2 \end{pmatrix} \qquad A_5 = \begin{pmatrix} 1 & 3 & -2 & 2 & 1 \\ 0 & -1 & 8 & -7 & -1 \\ 0 & 0 & 4 & -22 & 5 \\ 0 & 0 & 9 & -2 & -2 \end{pmatrix}$$

$$A_6 = \begin{pmatrix} 1 & 3 & -2 & 2 & 1 \\ 0 & -1 & 8 & -7 & -1 \\ 0 & 0 & 4 & -22 & 5 \\ 0 & 0 & 1 & 42 & -12 \end{pmatrix} \qquad A_7 = \begin{pmatrix} 1 & 3 & -2 & 2 & 1 \\ 0 & -1 & 8 & -7 & -1 \\ 0 & 0 & 1 & 42 & -12 \\ 0 & 0 & 4 & -22 & 5 \end{pmatrix}$$

$$A_8 = \begin{pmatrix} 1 & 3 & -2 & 2 & 1 \\ 0 & -1 & 8 & -7 & -1 \\ 0 & 0 & 1 & 42 & -12 \\ 0 & 0 & 0 & -190 & 53 \end{pmatrix} \qquad A_9 = \begin{pmatrix} 1 & 3 & -2 & 2 & 1 \\ 0 & 1 & -8 & 7 & 1 \\ 0 & 0 & 1 & 42 & -12 \\ 0 & 0 & 0 & 190 & -53 \end{pmatrix}$$

5.4.5. Show that row operations commute with column operations. In other words, if M is a matrix and M' is the matrix obtained from M by first applying a certain row operation and then a certain column operation, then applying the two operations in the opposite order to M yields the same matrix M'.

5.4.6. Prove Proposition 5.33.

5.4.7. Is Corollary 5.34 also true for column equivalent matrices M and M'? What about matrices M and M' that can be obtained from each other by a sequence of row *or* column operations?

5.5. 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 5.31 we may replace A by any row equivalent matrix.

Example 5.35. 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 5.31. 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 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 5.35 is said to be in *row echelon form*. This form made it very easy to solve the equations (in terms of the coefficients of $x \in \mathbb{R}^4$) that describe the fact that $x \in \ker A'$. In this section we will define the row echelon form, and we explain how to find a matrix in row echelon form that is row equivalent to some given matrix. In the next section we will see in full generality how to obtain generators for the kernel from the row echelon form.

Definition 5.36. 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.

Example 5.37. The matrix A_9 of Exercise 5.4.4 is in row echelon form. The following matrices are all in row echelon form as well, with the last one describing

the most general shape with all pivots equal to 1.

$$\begin{pmatrix} 1 & 4 & -2 & 4 & 3 \\ 0 & 2 & 7 & 2 & 5 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & 4 & -2 & 4 \\ 0 & 5 & 7 & 2 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\begin{array}{c} 1 \\ 2 \\ \vdots \\ r \\ r+1 \\ \vdots \\ m \end{array} \begin{pmatrix} 0\cdots0 & 1 & *\cdots* & * & *\cdots* & * & *\cdots* \\ 0\cdots0 & 0 & 0\cdots0 & 1 & *\cdots* & * & *\cdots* \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0 & 1 & *\cdots* \\ 0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0 \end{pmatrix} \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \end{array}$$
$$\begin{array}{ccccccc} & j_1 & & j_2 & \cdots & j_r & \end{array}$$

To make the matrix A 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 if $A = (a_{ij})$, then $a_{ij} = 0$ if $i > r$ or if $i \leq r$ and $j < j_i$, and $a_{ij_i} = 1$ for $1 \leq i \leq r$.

Every matrix can be brought into row echelon form by a sequence of elementary row operations. The following procedure describes precisely how to do this. 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 5.38 (The 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 $j_0 = 0$.*
2. *(At this point, $a'_{ij} = 0$ if $i > r$ and $j \leq j_r$ or if $1 \leq i \leq r$ 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 mth 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 rth and the ith row of A' if $r \neq i$. Note that $j_r > j_{r-1}$.*
4. *Multiply the rth row of A' by $(a'_{rj_r})^{-1}$.*
5. *For each $i = r+1, \ldots, m$, add $-a'_{ij_r}$ times the rth row of A' to the ith 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 correct. It is trivially satisfied when we reach step 2 for the first time. We now assume it is OK 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 $a'_{ij} = 0$ for $i > r$ and $j \leq j_r$ and for $i = r$ 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. □

Example 5.39. Consider the following 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 have to distinguish two cases. If $3 = 0$ in F, then

$$A' = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

is already in row echelon form. Otherwise, $-3 \neq 0$, so 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 5.40. In Example 5.35, the matrix A' is a matrix in row echelon form that is row equivalent to A.

Remark 5.41. The row space of A in Example 5.39 is spanned by its three rows. Assume $3 \neq 0$. By Proposition 5.31, 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 6.46 and Proposition 7.9).

Example 5.42 (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.

$$\begin{pmatrix} 22 & 5 \\ 9 & 2 \end{pmatrix} \rightsquigarrow \begin{matrix} R_1 - 2R_2 \\ R_2 \end{matrix} \begin{pmatrix} 4 & 1 \\ 9 & 2 \end{pmatrix} \rightsquigarrow \begin{matrix} R_1 \\ R_2 - 2R_1 \end{matrix} \begin{pmatrix} 4 & 1 \\ 1 & 0 \end{pmatrix}$$
$$\rightsquigarrow \begin{matrix} R_2 \\ R_1 \end{matrix} \begin{pmatrix} 1 & 0 \\ 4 & 1 \end{pmatrix} \rightsquigarrow \begin{matrix} R_1 \\ R_2 - 4R_1 \end{matrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

We see that the 2×2 identity matrix is also a row echelon form for the original matrix.

Note that in Example 5.42 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. Note that if in some step you add a multiple of a row, say R_i, to another row, say R_j, then row R_i should appear unchanged as one of the rows in the new matrix.

We give one more example, where we avoid denominators all the way, except for the last step.

Example 5.43.

$$\begin{pmatrix} 3 & 5 & 2 & 2 \\ 1 & 3 & -4 & 3 \\ 2 & -2 & 5 & -1 \\ -1 & 3 & 1 & -3 \end{pmatrix} \rightsquigarrow \begin{matrix} R_2 \\ R_1 \\ R_3 \\ R_4 \end{matrix} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 3 & 5 & 2 & 2 \\ 2 & -2 & 5 & -1 \\ -1 & 3 & 1 & -3 \end{pmatrix}$$
$$\rightsquigarrow \begin{matrix} R_1 \\ R_2 - 3R_1 \\ R_3 - 2R_1 \\ R_4 + R_1 \end{matrix} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 0 & -4 & 14 & -7 \\ 0 & -8 & 13 & -7 \\ 0 & 6 & -3 & 0 \end{pmatrix} \rightsquigarrow \begin{matrix} R_1 \\ R_2 \\ R_3 \\ R_4 + R_2 \end{matrix} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 0 & -4 & 14 & -7 \\ 0 & -8 & 13 & -7 \\ 0 & 2 & 11 & -7 \end{pmatrix}$$
$$\rightsquigarrow \begin{matrix} R_1 \\ R_4 \\ R_3 \\ R_2 \end{matrix} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 0 & 2 & 11 & -7 \\ 0 & -8 & 13 & -7 \\ 0 & -4 & 14 & -7 \end{pmatrix} \rightsquigarrow \begin{matrix} R_1 \\ R_2 \\ R_3 + 4R_2 \\ R_4 + 2R_2 \end{matrix} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 0 & 2 & 11 & -7 \\ 0 & 0 & 57 & -35 \\ 0 & 0 & 36 & -21 \end{pmatrix}$$
$$\rightsquigarrow \begin{matrix} R_1 \\ R_2 \\ R_3 - R_4 \\ R_4 \end{matrix} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 0 & 2 & 11 & -7 \\ 0 & 0 & 21 & -14 \\ 0 & 0 & 36 & -21 \end{pmatrix} \rightsquigarrow \begin{matrix} R_1 \\ R_2 \\ R_3 \\ R_4 - R_3 \end{matrix} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 0 & 2 & 11 & -7 \\ 0 & 0 & 21 & -14 \\ 0 & 0 & 15 & -7 \end{pmatrix}$$
$$\rightsquigarrow \begin{matrix} R_1 \\ R_2 \\ R_3 - R_4 \\ R_4 \end{matrix} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 0 & 2 & 11 & -7 \\ 0 & 0 & 6 & -7 \\ 0 & 0 & 15 & -7 \end{pmatrix} \rightsquigarrow \begin{matrix} R_1 \\ R_2 \\ R_3 \\ R_4 - 2R_3 \end{matrix} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 0 & 2 & 11 & -7 \\ 0 & 0 & 6 & -7 \\ 0 & 0 & 3 & 7 \end{pmatrix}$$
$$\rightsquigarrow \begin{matrix} R_1 \\ R_2 \\ R_4 \\ R_3 \end{matrix} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 0 & 2 & 11 & -7 \\ 0 & 0 & 3 & 7 \\ 0 & 0 & 6 & -7 \end{pmatrix} \rightsquigarrow \begin{matrix} R_1 \\ R_2 \\ R_3 \\ R_4 - 2R_3 \end{matrix} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 0 & 2 & 11 & -7 \\ 0 & 0 & 3 & 7 \\ 0 & 0 & 0 & -21 \end{pmatrix}$$

$$\rightsquigarrow \quad \begin{matrix} R_1 \\ \frac{1}{2}R_2 \\ \frac{1}{3}R_3 \\ -\frac{1}{21}R_4 \end{matrix} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 0 & 1 & \frac{11}{2} & -\frac{7}{2} \\ 0 & 0 & 1 & \frac{7}{3} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

While the row echelon form of a matrix is not unique, the reduced row echelon form below is (see Corollary 5.47).

DEFINITION 5.44. 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:

$$A = \begin{pmatrix} 0\cdots 0 & 1 & *\cdots * & 0 & *\cdots * & 0 & *\cdots * \\ 0\cdots 0 & 0 & 0\cdots 0 & 1 & *\cdots * & 0 & *\cdots * \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0\cdots 0 & 0 & 0\cdots 0 & 0 & 0\cdots 0 & 1 & *\cdots * \\ 0\cdots 0 & 0 & 0\cdots 0 & 0 & 0\cdots 0 & 0 & 0\cdots 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0\cdots 0 & 0 & 0\cdots 0 & 0 & 0\cdots 0 & 0 & 0\cdots 0 \end{pmatrix}$$

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 rth row of A' to the ith row of A'.

PROPOSITION 5.45. *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 5.46. *The following statements about two matrices $A, A' \in \mathrm{Mat}(m \times n, F)$ 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 5.31, 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 5.31 we have $R(B) = R(A)$ and $R(B') = R(A')$, so we conclude $R(B) = R(B')$. Therefore, by Proposition 5.45, 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 (2) $\Rightarrow$ (3) follows from the fact that if $B = B'$ is row equivalent to both A and A', then A and A' are row equivalent. □

COROLLARY 5.47. *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 5.46 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 5.48. It follows from Corollary 5.47 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 7.2.

5.6. 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 5.31, 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 5.49. Suppose M is the matrix (over $\mathbb{R}$)

$$\begin{pmatrix} \textcircled{1} & 2 & -1 & 0 & 2 & 1 & -3 \\ 0 & 0 & \textcircled{1} & -1 & 2 & -1 & 2 \\ 0 & 0 & 0 & 0 & \textcircled{1} & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

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

$$\begin{aligned} x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{pmatrix} &= \begin{pmatrix} -2q + r + 4s + 5t \\ q \\ r + 3s \\ r \\ -s - t \\ s \\ t \end{pmatrix} = q \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + r \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + s \begin{pmatrix} 4 \\ 0 \\ 3 \\ 0 \\ -1 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 5 \\ 0 \\ 0 \\ 0 \\ -1 \\ 0 \\ 1 \end{pmatrix} \\ &= qw_2 + rw_4 + sw_6 + tw_7, \end{aligned}$$

where

$$w_2 = \begin{pmatrix} \textcircled{-2} \\ 1 \\ \textcircled{0} \\ 0 \\ \textcircled{0} \\ 0 \\ 0 \end{pmatrix}, \qquad w_4 = \begin{pmatrix} \textcircled{1} \\ 0 \\ \textcircled{1} \\ 1 \\ \textcircled{0} \\ 0 \\ 0 \end{pmatrix}, \qquad w_6 = \begin{pmatrix} \textcircled{4} \\ 0 \\ \textcircled{3} \\ 0 \\ \textcircled{-1} \\ 1 \\ 0 \end{pmatrix}, \qquad w_7 = \begin{pmatrix} \textcircled{5} \\ 0 \\ \textcircled{0} \\ 0 \\ \textcircled{-1} \\ 0 \\ 1 \end{pmatrix}.$$

This shows that the kernel $\ker M$ is generated by w_2, w_4, w_6, w_7, i.e., 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, i.e.,

$$w = \begin{pmatrix} * \\ 0 \\ * \\ 0 \\ * \\ 1 \\ 0 \end{pmatrix}.$$

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 theorem says that we can find generators for the kernel of any matrix in row echelon form in the same manner.

PROPOSITION 5.50. *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 numbers 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 $l \neq k$ and $l \notin \{j_1, j_2, \ldots, j_r\}$.*

Furthermore, 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 5.49 and is left to the reader. □

The computation of generators of the kernel of a matrix A is easier when A is in *reduced* row echelon form. A reduced row echelon form for the matrix M of Example 5.50, for instance, is

$$\begin{pmatrix} \textcircled{1} & 2 & 0 & -1 & 0 & -4 & -5 \\ 0 & 0 & \textcircled{1} & -1 & 0 & -3 & 0 \\ 0 & 0 & 0 & 0 & \textcircled{1} & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The circled entries of w_6 of Example 5.49 are exactly the negatives of the elements $-4, -3, 1$ in the nonzero rows and the sixth column. The same holds for the other generators w_2, w_4, and w_7. In terms of Proposition 5.50, 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 5.50 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.

PROPOSITION 5.51. *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 5.50. □

Proposition 5.51 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.

We can now also check efficiently whether the map associated to a matrix is injective.

PROPOSITION 5.52. *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 5.34, 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 4.5, the map f_A is injective if and only if the kernel $\ker f_A = \ker A$ is zero, which, according to Proposition 5.50, 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. □

Proposition 5.45 and Corollaries 5.46 and 5.47 state that if A is an $m \times n$ matrix and A' is the associated reduced row echelon form, then the nonzero rows of A' are uniquely determined by the row space $R(A)$ of A. The following proposition shows how the columns of A determine which of the columns of A' contain pivots.

PROPOSITION 5.53. *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 $v_k = \lambda_1 v_1 + \cdots + \lambda_{k-1} v_{k-1}$, i.e., 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 5.31, 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 $A = A'$ and without loss of generality, we may and will also assume that $A = A' = (a_{ij}) \in \mathrm{Mat}(m \times n, F)$ is in *reduced* row echelon form.

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, let r denote the number of nonzero rows of A and let the columns with pivot be numbered $j_1, j_2, \ldots, j_r$. If the k-th column does not contain a pivot, then by Proposition 5.51 the element

$$w_k = e_k - \sum_{\substack{1 \le i \le r \\ j_i < k}} a_{ik} e_{j_i}$$

is contained in the kernel, so we have $Aw_k = 0$, i.e.,

$$v_k = \sum_{\substack{1 \le i \le r \\ j_i < k}} a_{ik} v_{j_i},$$

and we conclude that v_k is indeed a linear combination of $v_1, v_2, \ldots, v_{k-1}$. □

Exercises

5.6.1. Prove Proposition 5.50.

5.6.2. Determine the "reduced row echelon form" for the following matrices over $\mathbb{C}$ and give generators for their kernels.

$$\begin{pmatrix} 2+i & 1 & 1+i \\ 2 & 1-3i & 3-5i \end{pmatrix} \qquad \begin{pmatrix} 3 & 0 & 3 \\ 2 & 3 & 0 \\ 3 & 3 & 1 \end{pmatrix}$$

$$\begin{pmatrix} -1 & 0 & 0 & 1 & 2 \\ 2 & 1 & -1 & 0 & 2 \\ 0 & 0 & 0 & -1 & 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 2 & 2 & -2 \\ 2 & 3 & 1 & 0 \\ -2 & 0 & 2 & 1 \end{pmatrix}$$

CHAPTER 6

Linear independence and dimension

6.1. 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 6.1. Let V be an F-vector space, $v_1, v_2, \ldots, v_n \in V$. 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 $(v_i)_{i \in I}$. Note that such a collection has more structure than a set, as for each index i, we know which element of the collection belongs to that index i. In other words, we know which is the i-th element. Also, elements may occur multiple times, so for $i, j \in I$ with $i \neq j$, we may have $v_i = v_j$. Such a collection is also called a labeled set, where the index i is called the label of the element v_i.

DEFINITION 6.2. 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, i.e., 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 6.1 and 6.2 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 6.3. 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 6.4. 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 $v_1 = -\frac{t}{s}v_2$ and $t \neq 0$ implies $v_2 = -\frac{s}{t}v_1$.

EXAMPLE 6.5. 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 6.6. 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 6.7. Let F be a field and $V = F[x]$ be the vector space of all polynomials in the variable x over F (see Example 1.14). 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 6.8. 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 6.9. 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 $\lambda_1 = \lambda_2 = \lambda_3 = 0$, so w_1, w_2, and w_3 are linearly independent.

Recall from Definition 4.11 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 6.10. *Suppose W is a vector space over the field F and $C = (w_1, w_2, \ldots, w_n)$ a sequence of n vectors in W. Then the elements $w_1, w_2, \ldots, w_n$ are linearly independent if and only if* $\ker \varphi_C = \{0\}$.

PROOF. The kernel of φ_C consists of all the n-tuples $(\lambda_1, \ldots, \lambda_n)$ with $\lambda_1 w_1 + \cdots + \lambda_n w_n = 0$, so indeed, we have $\ker \varphi_C = \{0\}$ if and only if the elements $w_1, w_2, \ldots, w_n$ are linearly independent. □

In fact, the proof shows that the nontrivial linear relations on $w_1, \ldots, w_n$ correspond exactly with the nonzero elements of the kernel of φ_C. A statement similar to Proposition 6.10 holds for arbitrary collections (exercise). For $W = F^m$, we have the following corollary.

COROLLARY 6.11. *Let F be a field and m a nonnegative integer. Then any vectors $w_1, w_2, \ldots, w_n \in F^m$ are linearly independent if and only if the $m \times n$ matrix that has $w_1, w_2, \ldots, w_n$ 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 5.5 and Proposition 5.8, so this follows from Proposition 6.10. □

EXAMPLE 6.12. Let $w_1, w_2, w_3 \in \mathbb{R}^3$ be as in Example 6.9. 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 $-w_2 + 3w_2 - 3w_3 + w_4 = 0$. 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 5.52.

LEMMA 6.13. *Let $f\colon V \to W$ be a linear map of vector spaces. Then any vectors $v_1, v_2, \ldots, v_n \in V$ are linearly independent if their images $f(v_1), f(v_2), \ldots, f(v_n)$ 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 6.10, 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. □

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 any linearly independent vectors $v_1, v_2, \ldots, v_n \in V$. 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. □

From the finite case, it follows immediately that Lemma 6.13 holds for arbitrary collections as well (exercise).

EXAMPLE 6.14. 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 6.13, 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 6.13, 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 6.11, 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

$$M \cdot \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} = 0.$$

The equality $\ker M = \{0\}$ shows $\lambda_1 = \lambda_2 = \lambda_3 = 0$, and we conclude again that f_1, f_2, and f_3 are linearly independent.

PROPOSITION 6.15. *Let V be a vector space, $v_1, v_2, \ldots, v_n \in V$. Then $v_1, v_2, \ldots, v_n$ are linearly dependent if and only if one of the v_j is a linear combination of the others, i.e., if and only if*

$$L(v_1, v_2, \ldots, v_n) = L(v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_n)$$

for some $j \in \{1, 2, \dots, n\}$. A similar statement holds for any collection $(v_i)_{i\in I}$ of vectors in V.

PROOF. Let us first assume that $v_1, v_2, \dots, v_n$ are linearly dependent. Then there are scalars $\lambda_1, \lambda_2, \dots, \lambda_n$, not all zero, such that

$$\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n = 0\,.$$

Let j be such that $\lambda_j \neq 0$. Then

$$v_j = -\lambda_j^{-1}(\lambda_1 v_1 + \cdots + \lambda_{j-1} v_{j-1} + \lambda_{j+1} v_{j+1} + \cdots + \lambda_n v_n)\,.$$

Conversely, assume that v_j is a linear combination of the other vectors:

$$v_j = \lambda_1 v_1 + \cdots + \lambda_{j-1} v_{j-1} + \lambda_{j+1} v_{j+1} + \cdots + \lambda_n v_n\,.$$

Then

$$\lambda_1 v_1 + \cdots + \lambda_{j-1} v_{j-1} - v_j + \lambda_{j+1} v_{j+1} + \cdots + \lambda_n v_n = 0\,,$$

so the given vectors are linearly dependent. Given 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, the last statement also follows. □

If we take the order of the vectors into consideration, we can make the following stronger statement.

PROPOSITION 6.16. *Let V be a vector space, $v_1, v_2, \dots, v_n \in V$. Then the elements $v_1, v_2, \dots, v_n$ are linearly dependent if and only if one of the v_j is a linear combination of the previous ones, i.e., if and only if*

$$v_j \in L(v_1, \dots, v_{j-1})$$

for some $j \in \{1, 2, \dots, n\}$. A similar statement holds for infinite sequences of vectors in V.

PROOF. Exercise. □

EXAMPLE 6.17. 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 1.6 and Warning 1.15). 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 6.16 that the polynomials are linearly independent.

EXAMPLE 6.18. 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 6.16.

PROPOSITION 6.19. *Let $v_1, v_2, \dots, v_r$ be the nonzero rows of a matrix in row echelon form. Then $v_1, v_2, \dots, v_r$ are linearly independent.*

PROOF. The proof is completely analogous to Example 6.18 and is left to the reader. □

PROPOSITION 6.20. *Let A be an $m \times n$ matrix in row echelon form. Let r be the number of nonzero rows in A. Then the $n - r$ elements w_k (for all $1 \leq k \leq n$ for which the k-th column contains no pivot) of Proposition 5.50 (or Proposition 5.51 if A is in reduced row echelon form) are linearly independent.*

PROOF. 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 6.15, these $n - r$ elements are linearly independent. □

Exercises

6.1.1. 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) $\big((1,3,2),(1,1,1),(-1,3,1)\big)$.

6.1.2. Are the polynomials $3, x-1, x^2-3x+2, x^4-3x+13, x^7-x+14$ linearly independent?

6.1.3. Are the polynomials $x^7-2x+1, 5x^2, 2x^4-5x^3, x, x^6-3x$ linearly independent?

6.1.4. 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)

6.1.5. Prove Proposition 6.19.

6.1.6.
(1) Prove Proposition 6.16.
(2) Phrase and prove a version of Proposition 6.16 for collections of vectors indexed by $\mathbb{Z}_{\geq 0}$, i.e., for infinite sequences $v_0, v_1, v_2, \ldots$.
(3) Phrase and prove a version of Proposition 6.16 any collection of vectors indexed by a totally ordered set I.

6.1.7. 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 4.2.5). Show that the collection $(w_i)_{i\in I}$ is linearly independent if and only if φ is injective.

6.1.8. State and prove a generalization of Proposition 6.10 for arbitrary collections of vectors, cf. Exercise 4.2.5.

6.1.9. State and prove a generalization of Lemma 6.13 for arbitrary collections of vectors.

6.2. Bases

DEFINITION 6.21. Let V be a vector space. A sequence $(v_1, v_2, \ldots, v_n)$ of elements of V is called a *basis* of V if $v_1, v_2, \ldots, v_n$ are linearly independent, and $V =$

$L(v_1, v_2, \ldots, v_n)$. We also say that the elements $v_1, v_2, \ldots, v_n$ *form* a basis for V. More generally, a basis is a collection $(v_i)_{i \in I}$ of vectors in V that is linearly independent and generates 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 collections, a basis $(v_i)_{i \in I}$ has more structure than just a set. For each index $i \in I$, we know which element of the basis belongs to that index i. In other words, we know which is the i-th element. See also the remark between Definitions 6.1 and 6.2.

EXAMPLE 6.22. 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

$$\begin{aligned} e_1 &= (1, 0, 0, \ldots, 0, 0) \\ e_2 &= (0, 1, 0, \ldots, 0, 0) \\ &\vdots \qquad\qquad \vdots \\ e_n &= (0, 0, 0, \ldots, 0, 1)\,. \end{aligned}$$

EXAMPLE 6.23. 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. For infinite sets X, see Exercise 6.2.5.

EXAMPLE 6.24 (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 $R(A)$, as they generate the row space by definition and they are linearly independent by Proposition 6.19. The $n - r$ elements w_k (for all $1 \leq k \leq n$ for which the k-th column contains no pivot) of Proposition 5.50 (or Proposition 5.51 if A is in reduced row echelon form) form a basis of the kernel of A, as they generate the kernel by Proposition 5.50 or 5.51 and they are linearly independent by Proposition 6.20.

REMARK 6.25 (Basis of U and $U^\perp$ using rows). We can use Example 6.24 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 5.31, the nonzero rows of A' form a basis for $R(A') = R(A) = U$ by Example 6.24. Moreover, the subspace $U^\perp$ equals $\ker A = \ker A'$ by Propositions 5.13 and 5.31, so Example 6.24 also gives a basis for $U^\perp$.

Remark 6.25 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 6.28 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 6.26. *Suppose V is a vector space and $v_1, v_2, \ldots, v_n \in V$. 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. Exercise. □

EXAMPLE 6.27. 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 5.53, the columns with a pivot, i.e., the first, second, fourth, and sixth, are exactly the columns that are not a linear combination of the previous columns of A. From Lemma 6.26 we conclude that these four columns form a basis for the column space $C(A)$ of A.

We can combine Proposition 5.53 and Lemma 6.26 to make a method to determine a basis for the column space of a matrix.

PROPOSITION 6.28 (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 5.53, 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 Lemma 6.26, this implies that this collection $(w_i)_{i \in I}$ is a basis for the space $L(w_1, \ldots, w_n) = C(A)$. □

REMARK 6.29 (Basis of U using columns). We can use Proposition 6.28 to determine a basis of a subspace U of F^n generated by elements $w_1, w_2, \ldots, w_m$. First we let B denote the $n \times m$ matrix of which the columns are $w_1, w_2, \ldots, w_m$. Note that $B = A^\top$ for A as in Remark 6.25. Then we apply a sequence of elementary row operations to B to obtain a matrix B' 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 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 6.30. 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 5.51).

Recall from Definition 4.11, as in the previous section, that for any sequence $C = (w_1, \dots, w_n)$ 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 $(a_1, \dots, a_n) \in F^n$ to $a_1 w_1 + \cdots + a_n w_n$.

PROPOSITION 6.31. *Suppose W is a vector space over the field F and $C = (w_1, w_2, \dots, w_n)$ 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 $w_1, \dots, w_n$ are linearly independent by Proposition 6.10. The map φ_C is surjective if and only if $w_1, \dots, w_n$ generate W (see the remark below Proposition 4.10). The statement follows. □

A statement similar to Proposition 6.31 holds for arbitrary collections (exercise).

From Proposition 6.15 above, we see that a basis of V is a *minimal* generating set of V, be it with an ordering or labels, in the sense that we cannot leave out some element and still have a generating set.

What is special about a basis among generating sets?

LEMMA 6.32. *Suppose V is an F-vector space. Then a sequence $(v_1, v_2, \dots, v_n)$ of elements in V is a basis for V if and only if for every $v \in V$, there are* unique *scalars $\lambda_1, \lambda_2, \dots, \lambda_n \in F$ such that*

$$v = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n \,.$$

PROOF. Set $C = (v_1, v_2, \dots, v_n)$. Then by Proposition 6.31, 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 $\lambda_1, \lambda_2, \dots, \lambda_n \in F$ 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 $(v_1, v_2, \dots, v_n)$ is a basis for V. The *existence* of $(\lambda_1, \lambda_2, \dots, \lambda_n) \in F^n$ such that

$$v = \lambda_1 v_1 + \lambda_2 + \cdots + \lambda_n v_n$$

follows from the fact that $v_1, v_2, \dots, v_n$ generate V.

To show *uniqueness,* assume that $(\mu_1, \mu_2, \dots, \mu_n) \in F^n$ 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

$$\lambda_1 - \mu_1 = \lambda_2 - \mu_2 = \cdots = \lambda_n - \mu_n = 0\,,$$

i.e., $(\lambda_1, \lambda_2, \ldots, \lambda_n) = (\mu_1, \mu_2, \ldots, \mu_n)$. The converse is left as an exercise. □

LEMMA 6.33. *Let $f\colon V \to W$ be an isomorphism of vector spaces and $v_1, v_2, \ldots, v_n$ elements of V. Then the elements $v_1, v_2, \ldots, v_n$ 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.*

PROOF. Set $C = (v_1, v_2, \ldots, v_n)$. By Proposition 6.31, the elements $v_1, v_2, \ldots, v_n$ form a basis for V if and only if φ_C is an isomorphism. The composition $f \circ \varphi_C\colon F^n \to W$ sends e_j to $f(v_j)$, so the elements $f(v_1), f(v_2), \ldots, f(v_n)$ form a basis for W if and only if $f \circ \varphi_C$ is an isomorphism. The lemma now follows from the fact that φ_C is an isomorphism if and only if the composition $f \circ \varphi_C$ is. □

ALTERNATIVE PROOF. Suppose $v_1, v_2, \ldots, v_n$ form a basis for V. Then the elements $v_1, \ldots, v_n$ are linearly independent and since f is injective, the linear independence of $f(v_1), \ldots, f(v_n)$ follows from Lemma 6.13. Because $v_1, \ldots, v_n$ generate V, we also have

$$L\left(f(v_1), \ldots, f(v_n)\right) = f\left(L(v_1, \ldots, v_n)\right) = f(V) = W$$

by Lemma 4.3, so $f(v_1), \ldots, f(v_n)$ generate W, so they form a basis. The converse statement follows by applying the same argument to f^{-1}. □

PROPOSITION 6.34. *Let V and W be vector spaces, $f : V \to W$ a linear map, and let $v_1, \ldots, v_n$ be a basis of V. Then*

1. *f is injective if and only if $f(v_1), \ldots, f(v_n)$ are linearly independent,*
2. *f is surjective if and only if $L(f(v_1), \ldots, f(v_n)) = W$, and*
3. *f is an isomorphism if and only if $f(v_1), \ldots, f(v_n)$ is a basis of W.*

PROOF. The proof of the first two statements is an exercise; the third follows from the first two. □

Lemmas 6.32 and 6.33 and Proposition 6.34 also hold for arbitrary collections (exercise).

Exercises

6.2.1. 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. $v_1 = (1, 2, 3, 4), v_2 = (4, 3, 2, 1), v_3 = (1, -1, 1, -1)$,

6.2.2. Redo Exercise 6.1.4.

6.2.3. Finish the alternative proof of Lemma 6.32.

6.2.4. For each of the matrices of Exercise 5.6.2, select some columns that form a basis for the column space of that matrix.

6.2.5. This exercise generalizes Example 6.23. 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 $F^{(X)}$.

6.2.6. State and prove a generalization of Proposition 6.31 for arbitrary collections of vectors, cf. Exercise 4.2.5.

6.2.7. State and prove an analog of Lemma 6.32 for arbitrary collections $(v_i)_{i\in I}$ of vectors in V.

6.2.8. Use Proposition 4.10 to prove the following generalization of Proposition 4.10 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 $w_1, w_2, \dots, w_n$ of vectors in W there is a unique linear map $f \colon V \to W$ such that $f(v_j) = w_j$ for all $j \in \{1, \dots, n\}$." 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).

6.2.9. Prove Lemma 6.26. Is the same statement true for infinite sequences $v_1, v_2, v_3, \dots$? 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 6.13 and 6.33 and Proposition 6.34. We will not use/assume in these statements that every vector space has a basis, cf. Warning 6.52, which is why it is included as explicit hypothesis whenever needed.

6.2.10. State and prove an analog of Lemma 6.33 for arbitrary collections $(v_i)_{i\in I}$ of vectors in V.

6.2.11. Prove Proposition 6.34. Also state and prove an analog of Proposition 6.34 for an arbitrary collection $(v_i)_{i\in I}$ of vectors as a basis for V (follows from the next three exercises).

6.2.12. 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 nonnegative 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 also 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.

6.2.13. 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 generate 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 $v_1, v_2, \ldots, v_n$ of elements in V that generate V, the images $f(v_1), f(v_2), \ldots, f(v_n)$ generate W.

6.2.14. 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.

6.3. The basis extension theorem and dimension

Proposition 6.31 says that if $v_1, v_2, \ldots, v_n$ form a basis for a vector space V, then V is isomorphic to the standard vector space F^n, so that 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 6.35 (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 $V = L(v_1, \ldots, v_r, w_1, \ldots, w_s)$. Then there is $t \in \mathbb{N}_0$ and indices $i_1, \ldots, i_t \in \{1, \ldots, s\}$ such that $(v_1, \ldots, v_r, w_{i_1}, \ldots, w_{i_t})$ is a basis of V.*

The Basis Extension Theorem says that when $v_1, \ldots, v_r$ and $w_1, \ldots, w_s$ are as given, then by adding suitably chosen vectors from $w_1, \ldots, w_s$, we can extend $v_1, \ldots, v_r$ to a basis of V. Make sure you understand how we have formalized the notion of "suitably chosen vectors from $w_1, \ldots, w_s$!"

Note that this is an *existence theorem* — what it says is that if we have a bunch of vectors that is 'too small' (linearly independent, but not necessarily generating) and a larger bunch of vectors that is 'too large' (generating but not necessarily linearly independent), then there is a basis 'in between'. Proposition 6.38 tells us how to actually find such a basis, i.e., how to select the w_j that we have to add, in the case V is a subspace of F^n.

PROOF OF THEOREM 6.35. The idea of the proof is simply to add vectors from the w_j's as long as this is possible while keeping the sequence linearly independent. When no further lengthening is possible, we should have a basis. So we are looking for a maximal linearly independent sequence $v_1, \ldots, v_r, w_{i_1}, \ldots, w_{i_t}$. Note that there cannot be repetitions among the $w_{i_1}, \ldots, w_{i_t}$ if this sequence is to be linearly independent. Therefore $t \leq s$, and there must be such a sequence of maximal length. We have to show that it generates V. It suffices to show that $w_j \in L(v_1, \ldots, v_r, w_{i_1}, \ldots, w_{i_t})$ for all $j \in \{1, \ldots, s\}$. This is clear if $j = i_k$ for some $k \in \{1, \ldots, t\}$. Otherwise, assume that w_j is *not* a linear combination of $v_1, \ldots, v_r, w_{i_1}, \ldots, w_{i_t}$. Then $v_1, \ldots, v_r, w_{i_1}, \ldots, w_{i_t}, w_j$ would be linearly independent, which would contradict our choice of a linearly independent sequence of maximal length. So w_j must be a linear combination of our vectors, and the theorem is proved. □

ALTERNATIVE PROOF. Here is an alternative proof, using induction on the number s of vectors w_j. The base case is $s = 0$. In this case, the assumptions tell us that $v_1, \ldots, v_r$ are linearly independent and generate V, so we have a basis. For the induction step, we assume the statement of the theorem is true for

$w_1, \dots, w_s$ (and any choice of linearly independent vectors $v_1, \dots, v_r$), and we have to prove it for $w_1, \dots, w_s, w_{s+1}$. First assume that $L(v_1, \dots, v_r, w_1, \dots, w_s) = V$. Then the induction hypothesis immediately gives the result. So we assume now that $L(v_1, \dots, v_r, w_1, \dots, w_s) \subsetneq V$. Then w_{s+1} is not contained in the subspace $L(v_1, \dots, v_r, w_1, \dots, w_s)$, so w_{s+1} is not a linear combination of $v_1, \dots, v_r$, hence $v_1, \dots, v_r, w_{s+1}$ are linearly independent. Now we can apply the induction hypothesis again (to $v_1, \dots, v_r, w_{s+1}$ and $w_1, \dots, w_s$); it tells us that we can extend $v_1, \dots, v_r, w_{s+1}$ to a basis by adding suitable vectors from $w_1, \dots, w_s$. This gives us what we want. □

Example 6.36. 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 and x^2 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 non-trivial 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 also 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 6.37. 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 6.13) and bases to bases (Lemma 6.33). 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 6.28 and Remark 6.29 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 6.28 and Remark 6.29, the column space $C(A)$ has a basis consisting

of the corresponding columns of B. We conclude that (v_1, v_2, v_3, e_2) is a basis of $C(A) = \mathbb{R}^4$ and 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 5.53.

The idea of the example above can be used in general to extend some linearly independent vectors in a subspace V of F^n to a basis of V. The following proposition makes this precise.

PROPOSITION 6.38 (Explicit Basis Extension Theorem). *Let $V \subset F^n$ be a subspace containing elements $v_1, \ldots, v_r, w_1, \ldots, w_s \in V$ such that $v_1, \ldots, v_r$ are linearly independent and $V = L(v_1, \ldots, v_r, w_1, \ldots, w_s)$. Let A be the $n \times (r + s)$ matrix with columns $v_1, \ldots, v_r, w_1, \ldots, w_s$, let A' be the associated reduced row echelon form, and I the set of all indices i with $1 \leq i \leq s$ 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. The vectors $v_1, \ldots, v_r$ are linearly independent, so none is a linear combination of the others, so the first r columns of A' contain a pivot by Proposition 5.53. This means that the elements $v_1, v_2, \ldots, v_r$ and $(w_i)_{i \in I}$ correspond exactly to the columns of A' that contain a pivot. By Proposition 6.28, these elements form a basis for the column space $C(A)$ of A, which equals V by construction. □

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 6.39 (Exchange Lemma). *If $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$ are two bases of a vector space V, then for each $i \in \{1, 2, \ldots, n\}$ there is some $j \in \{1, 2, \ldots, m\}$ such that $v_1, \ldots, v_{i-1}, w_j, v_{i+1}, \ldots, v_n$ is again a basis of V.*

PROOF. Fix $i \in \{1, \ldots, n\}$. Since $v_1, \ldots, v_n$ are linearly independent, v_i cannot be a linear combination of the remaining v's. So $U = L(v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n) \subsetneq V$. This implies that there is some $j \in \{1, \ldots, m\}$ such that $w_j \notin U$ (if all $w_j \in U$, then $V \subset U$). This in turn implies that $v_1, \ldots, v_{i-1}, w_j, v_{i+1}, \ldots, v_n$ is linearly independent. If it is not a basis of V, then by the Basis Extension Theorem, $v_1, \ldots, v_{i-1}, w_j, v_{i+1}, \ldots, v_n, v_i$ must be a basis (we apply the Basis Extension Theorem to the linearly independent vectors $v_1, \ldots, v_{i-1}, w_j, v_{i+1}, \ldots, v_n$ and the additional vector v_i; together they generate V). However, the vectors in this latter sequence are not linearly independent, since w_j is a linear combination of $v_1, \ldots, v_n$. So $v_1, \ldots, v_{i-1}, w_j, v_{i+1}, \ldots, v_n$ must already be a basis of V. □

THEOREM 6.40. *If $v_1, v_2, \ldots, v_n$ and $w_1, w_2, \ldots, w_m$ 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, …, 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. □

This implies that the following definition makes sense.

DEFINITION 6.41. If a vector space V over a field F has a basis $(v_1, v_2, \ldots, v_n)$, 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 does not have a finite basis, then we write $\dim V = \infty$ and we say that V is *infinite-dimensional*.

EXAMPLE 6.42. The empty sequence is a basis of the zero space, so $\dim\{0\} = 0$.

EXAMPLE 6.43. The canonical basis of F^n has length n, so $\dim F^n = n$.

EXAMPLE 6.44. 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 $f_1, f_2, \ldots, f_r$ 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 6.45. Let F be a field and $d \geq 0$ an integer. Then the vector space $F[x]_d$ of all polynomials of degree at most d has a basis $(1, x, x^2, \ldots, x^d)$, so $\dim F[x]_d = d + 1$.

THEOREM 6.46. *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, i.e., 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 6.47. Theorem 6.46(2) shows that if V is a finitely generated vector space, then V has a finite basis and a finite dimension.

Note that Theorem 6.46 yields a quite strong existence statement: if V is a vector space of dimension $\dim V = n$ containing a sequence $C = (v_1, v_2, \ldots, v_r)$ of r elements in V, then the nontrivial linear relations among $v_1, v_2, \ldots, v_r$ correspond to

the nonzero elements in the kernel of $\varphi_C\colon F^r \to V$ (see remark below Proposition 6.10), and part (1) guarantees the *existence* of such a nontrivial linear relation 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.

Part (1) of Theorem 6.46 tells us that in a vector space of (finite) dimension n, there is an upper bound (namely, n) for the length of a linearly independent sequence of vectors. We can use this to show in another way that $\dim F[x] = \infty$ (see Example 6.44).

EXAMPLE 6.48. Let F be a field. The vector space $F[x]$ of all *polynomials* in the variable x with coefficients in F contains the monomials $1, x, x^2, x^3, x^4, \ldots$, which are linearly independent, see Example 6.7. 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 6.46(1). We conclude, again, $\dim F[x] = \infty$. Note that since $F[x] = L(\{x^n : n \in \mathbb{N}_0\})$, we have shown that the collection $(x^n)_{n\in\mathbb{N}_0}$ is a basis of $F[x]$.

With a little more effort, we can also show that the subspace of $\mathbb{R}^{\mathbb{R}}$ of real polynomial *functions* does not have a finite basis either.

EXAMPLE 6.49. Let us consider again the linear subspace of *polynomial functions* in $\mathcal{C}(\mathbb{R})$ (the vector space of continuous functions on $\mathbb{R}$), compare Example 2.34. Let us call this space P:

$$P = \{f \in \mathcal{C}(\mathbb{R}) : \exists n \in \mathbb{N}_0\ \exists a_0, \ldots, a_n \in \mathbb{R}\ \forall x \in \mathbb{R} : f(x) = a_n x^n + \cdots + a_1 x + a_0\}$$

Denote as before by f_n the nth power function: $f_n(x) = x^n$. I claim that the collection $(f_0, f_1, f_2, \ldots) = (f_n)_{n\in\mathbb{N}_0}$ is linearly independent. Recall that this means that the only way of writing zero (i.e., 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 $\mathbb{R}$. 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

$$\forall n \in \mathbb{N}_0\ \forall \lambda_0, \ldots, \lambda_n \in \mathbb{R} : \Big(\big(\forall x \in \mathbb{R} : \lambda_n x^n + \cdots + \lambda_0 = 0\big) \Longrightarrow \lambda_0 = \cdots = \lambda_n = 0\Big).$$

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 $(f_n)_{n\in\mathbb{N}_0}$ is linearly independent. From Proposition 6.46 we conclude $\dim P = \infty$.

Note that since $P = L(\{f_n : n \in \mathbb{N}_0\})$, we have shown that the collection $(f_n)_{n\in\mathbb{N}_0}$ is a basis for P.

Example 6.50. We have inclusions

$$P \subset \mathcal{C}^\infty(\mathbb{R}) = \bigcap_{n=0}^{\infty} \mathcal{C}^n(\mathbb{R}) \subset \cdots \subset \mathcal{C}^2(\mathbb{R}) \subset \mathcal{C}^1(\mathbb{R}) \subset \mathcal{C}(\mathbb{R}) \subset \mathbb{R}^{\mathbb{R}}\,.$$

Since P contains arbitrarily long sequences of linearly independent functions, so do all these spaces and therefore they are all infinite-dimensional.

Warning 6.51. Although the vector space of real polynomial functions is infinite-dimensional, over finite fields this is not the case (see Exercise 6.4.2).

Warning 6.52. In Examples 6.48 and 6.49 we actually found infinite bases for $F[x]$ and $P \subset \mathbb{R}^{\mathbb{R}}$, 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 something called "Zorn's Lemma," one can show that all vector spaces do magically turn out to have some basis, but 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 6.53. *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 $v_1, \dots, v_n$ of n linearly independent elements in V can be extended to a sequence $v_1, \dots v_n, v_{n+1}, \dots, v_r$ of linearly independent vectors in V of arbitrary length $r \geq n$.*

Proof. Exercise. □

In the following proposition, and thereafter, we use the usual convention that $n < \infty$ for $n \in \mathbb{N}_0$.

PROPOSITION 6.54. *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. For (1), suppose f is injective. Suppose V is finite-dimensional, say $\dim V = n$. If $\dim W = \infty$, then we are done, so assume $\dim W = n$. If $v_1, \ldots, v_s \in V$ are linearly independent, then so are $f(v_1), \ldots, f(v_s)$ by Lemma 6.13, and by Proposition 6.46 we obtain $s \leq n$. We conclude that V contains no sequences of more than n linearly independent vectors. By Proposition 6.53 this implies that V is not infinite-dimensional, say $\dim V = m$. Repeating the argument for $s = m$ with a basis $(v_1, \ldots, v_m)$, we conclude $m \leq n$.

For (2), suppose f is surjective. If $\dim V = \infty$, then we are done, so assume V is finite-dimensional, say $\dim V = n$, and let $(v_1, \ldots, v_n)$ be a basis for V. Then $f(v_1), \ldots, f(v_n)$ generate W by Proposition 6.34, so $\dim W \leq n = \dim V$ by Proposition 6.46.

Implication (3) follows from (1) and (2). □

EXAMPLE 6.55. We conclude, just from the dimensions, that the 3×4 matrix A of Example 5.7 induces a linear map $F^4 \to F^3$ that is not injective.

COROLLARY 6.56. *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 $m = \dim F^m = \dim F^n = n$. □

The next proposition shows that the converse of Proposition 6.54(3) also holds. 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 6.31).

PROPOSITION 6.57. *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 6.31 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 of V for different purposes, so it does not make sense to identify V with F^n in a specific way.

Exercises

6.3.1. 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 $\mathbb{R}[x]_4$.

6.3.2. 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.

6.3.3. Let V be a finite-dimensional vector space and $S \subset V$ a subset that generates V. Show that there is a finite subset of S that generates V.

6.3.4. Prove Proposition 6.53.

6.3.5. This exercise gives two alternative definitions for the dimension of a matrix. 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.

6.4. Dimensions of subspaces

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 6.58. *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 asserts the existence of a finite basis of U.

PROOF. There is nothing to show if $\dim V = \infty$. So let us assume that V has a basis $v_1, \ldots, v_n$. If $u_1, \ldots, u_m \in U$ are linearly independent, then $m \leq n$ by Theorem 6.46(1). Hence there is a sequence $u_1, \ldots, u_m$ of linearly independent vectors in U of maximal length m (and $m \leq n$). We claim that $u_1, \ldots, u_m$ is in fact a basis of U. The first claim then follows, since then $\dim U = m \leq n = \dim V$.

We have to show that $u_1, \ldots, u_m$ generate U. So assume that there is $u \in U$ that is not a linear combination of the u_j. Then $u_1, \ldots, u_m, u$ are linearly independent, which contradicts our choice of $u_1, \ldots, u_m$ as a *maximal* linearly independent sequence in U. So there is no such u, hence $U = L(u_1, \ldots, u_m)$.

To prove the second part, first assume $\dim U < \dim V$. Then by Theorem 6.40, no basis of U would also be a basis of V, so $U \neq V$. Conversely, assume $U \neq V$ and consider a basis of U. It can be extended to a basis for V by the Basis Extension Theorem 6.35. Since it does not generate V, at least one element has to be added, which implies $\dim U < \dim V$. □

Now we have the following nice formula relating the dimensions of U_1, U_2, $U_1 \cap U_2$ and $U_1 + U_2$. We use the convention that $\infty + n = n + \infty = \infty + \infty = \infty$ for $n \in \mathbb{N}_0$.

THEOREM 6.59. *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.

For the proof, we use the Basis Extension Theorem 6.35 again. Since $U_1 \cap U_2 \subset U_1$ and U_1 is finite-dimensional, we know by Lemma 6.58 that $U_1 \cap U_2$ is also finite-dimensional. Let $v_1, \dots, v_r$ be a basis of $U_1 \cap U_2$. Using the Basis Extension Theorem, we can extend it on the one hand to a basis $v_1, \dots, v_r, w_1, \dots, w_s$ of U_1 and on the other hand to a basis $v_1, \dots, v_r, z_1, \dots, z_t$ of U_2. I claim that then $v_1, \dots, v_r, w_1, \dots, w_s, z_1, \dots, z_t$ is a basis of $U_1 + U_2$. 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 2.45). So it remains to show that they are linearly independent. Consider a general linear combination

$$\lambda_1 v_1 + \dots + \lambda_r v_r + \mu_1 w_1 + \dots + \mu_s w_s + \nu_1 z_1 + \dots + \nu_t z_t = 0 \,.$$

Then $z = \nu_1 z_1 + \dots + \nu_t z_t \in U_2$, but also

$$z = -\lambda_1 v_1 - \dots - \lambda_r v_r - \mu_1 w_1 - \dots - \mu_s w_s \in U_1 \,,$$

so $z \in U_1 \cap U_2$, which implies that

$$z = \alpha_1 v_1 + \dots + \alpha_r v_r$$

for suitable α_j, since $v_1, \dots, v_r$ is a basis of $U_1 \cap U_2$. Then we have

$$0 = z - z = \alpha_1 v_1 + \dots + \alpha_r v_r - \nu_1 z_1 - \dots - \nu_t z_t.$$

But $v_1, \dots, v_r, z_1, \dots, z_t$ are linearly independent (being a basis of U_2), so this is only possible if $\alpha_1 = \dots = \alpha_r = \nu_1 = \dots = \nu_t = 0$. This then implies that $z = 0$, so

$$0 = -z = \lambda_1 v_1 + \dots + \lambda_r v_r + \mu_1 w_1 + \dots + \mu_s w_s \,,$$

and since $v_1, \dots, v_r, w_1, \dots, w_s$ are linearly independent (being a basis of U_1), we get $\lambda_1 = \dots = \lambda_r = \mu_1 = \dots = \mu_t = 0$ as well. So we have $\dim(U_1 + U_2) = r + s + t$, $\dim(U_1 \cap U_2) = r$, $\dim U_1 = r + s$ and $\dim U_2 = r + t$, from which the claim follows. □

REMARK 6.60. 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

$$\begin{aligned} \dim(U_1 + U_2 + U_3) + \dim(U_1 \cap U_2) + \dim(U_1 \cap U_3) + \dim(U_2 \cap U_3) \\ \neq \dim U_1 + \dim U_2 + \dim U_3 + \dim(U_1 \cap U_2 \cap U_3) \,. \end{aligned}$$

EXAMPLE 6.61. 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 6.59 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. Either $\dim(L+V) = 3$ and $\dim(L \cap V) = 0$, which means $L + V = \mathbb{R}^3$ and $L \cap V = \{0\}$, or $\dim(L+V) = 2$ and $\dim(L \cap V) = 1$, which implies $L \cap V = L$, so L is contained in V.

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 2.46) give an especially nice case.

PROPOSITION 6.62. *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 6.59. □

EXAMPLE 6.63. Let $a \in \mathbb{R}^n$ be nonzero and H the hyperplane $H = \{a\}^\perp$. Then $\dim H = n - 1$ by Corollary 3.15.

We can use the Basis Extension Theorem to show the existence of complementary subspaces in finite-dimensional vector spaces.

PROPOSITION 6.64. *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. In this case, U is finite-dimensional by Proposition 6.58, with basis $u_1, \ldots, u_m$ (say). By the Basis Extension Theorem 6.35, 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 2.45). But we also have $U \cap U' = \{0\}$: if $v \in U \cap U'$, then

$$v = \lambda_1 u_1 + \cdots + \lambda_m u_m = \mu_1 v_1 + \cdots + \mu_n v_n \,,$$

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 all the λs and μs must be zero, hence $v = 0$. □

EXAMPLE 6.65. 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 of the form $u' = (x', 1)$, and $U' = L(u')$ then is a complementary subspace for every $x' \in \mathbb{R}$ — note that $(x, y) = (x - yx', 0) + y(x', 1) \in U + U'$.

REMARK 6.66. 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 6.58. This means that Theorem 6.59 implies the inequality

$$\dim(U_1 \cap U_2) \geq \dim U_1 + \dim U_2 - \dim V.$$

EXAMPLE 6.67. 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 6.63, so we have

$$n - 1 \geq \dim H_1 \geq \dim(H_1 \cap H_2) \geq \dim H_1 + \dim H_2 - \dim \mathbb{R}^n = n - 2.$$

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 6.58. 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, i.e., a line.

Exercises

6.4.1. Do the exercise in Remark 6.60.

6.4.2. Let F be a *finite* field, and consider the F-vector space V of functions from F to F (so $V = F^F$ in our earlier notation). Consider again the linear subspace of polynomial functions:

$$P(F) = L_F(\{f_0, f_1, f_2, \dots\})$$

where $f_n : x \mapsto x^n$ (for $x \in F$). Show that $\dim_F P(F)$ is finite. (Warning: do not confuse the space $P(F)$ of polynomial **functions** with the space $F[x]$ of polynomials, which has infinite dimension, cf. Warning 2.36 and Examples 2.35, 6.48, 6.49, and 6.50.)

6.4.3. Let F be a finite field. Show that the map $\varphi\colon F[x] \to F^F$ of Exercise 2.3.8 is not injective, cf. Exercise 4.1.6.

[Remark: one can show that if $q = |F|$, then the kernel of φ consists of all polynomials that are a multiple of $x^q - x$.]

6.4.4. 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 $U_r + U_s = \mathbb{R}[x]_d$.

6.4.5. 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.

6.4.6. Find a vector space V and linear subspaces $U_1, U_2, U_3 \subset V$ such that

$$\begin{aligned}\dim(U_1 + U_2 + U_3) + \dim(U_1 \cap U_2) + \dim(U_1 \cap U_3) + \dim(U_2 \cap U_3) \\ \neq \dim U_1 + \dim U_2 + \dim U_3 + \dim(U_1 \cap U_2 \cap U_3)\,.\end{aligned}$$

CHAPTER 7

Ranks

7.1. 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 7.1. Let $f : V \to W$ be a linear map. Then we call the dimension of the image of f the *rank* of f: $\operatorname{rk}(f) = \dim \operatorname{im}(f)$.

THEOREM 7.2 (Dimension Formula for Linear Maps). *Let $f : V \to W$ be a linear map. Then*

$$\dim \ker(f) + \operatorname{rk}(f) = \dim V\,.$$

PROOF. First we consider the case that V is finite-dimensional. By Proposition 6.64, there is a complementary subspace U of $\ker(f)$ in V and we have $\dim \ker f + \dim U = \dim V$ by Proposition 6.62.

Let $f' \colon U \to \operatorname{im}(f)$ be the linear map given by restricting 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 $w \in \operatorname{im}(f)$. 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 2.48). Now

$$f'(u) = f(u) = f(v - u') = f(v) - f(u') = w - 0 = w\,,$$

so $w \in \operatorname{im}(f')$ as well. This implies that f' is surjective and thus an isomorphism. We conclude $\dim U = \dim \operatorname{im}(f) = \operatorname{rk} f$ and therefore

$$\dim V = \dim \ker f + \dim U = \dim \ker f + \operatorname{rk} f.$$

Now consider the case $\dim V = \infty$. If $\operatorname{rk} f = \infty$, then we are done, so assume $\operatorname{rk} f = n$ 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 6.53) and set $U = L(v_1, \ldots, v_r)$. Let $f' \colon U \to \operatorname{im} f$ be the linear map given by restricting f to U. Then by the finite-dimensional case, we have

$$\dim \ker f \geq \dim \ker f' = \dim U - \operatorname{rk} f' \geq \dim U - \dim \operatorname{im} f = r - n,$$

where the two inequalities follow from the inclusions $\ker f' \subset \ker f$ and $\operatorname{im} f' \subset \operatorname{im} f$, respectively. Since r was an arbitrary positive integer, we conclude $\dim \ker f = \infty$, which proves the dimension formula for linear maps. □

For a proof working directly with bases, see Chapter 4 in Jänich's book [**J**].

EXAMPLE 7.3. 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 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 map that sends $f \in F[x]_n$ to $\big(f(\alpha_1), f(\alpha_2), \ldots, f(\alpha_k)\big)$. Then for each $1 \le i \le k$, 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 $\operatorname{rk} S = \dim \operatorname{im} S \ge k$. Of course we also have $\dim \operatorname{im} S \le k$, 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 7.4. *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 $\dim \ker f = 0$ and f is surjective if and only if $\operatorname{rk}(f) = \dim W = \dim V$. By Theorem 7.2, these two statements are equivalent. □

EXAMPLE 7.5. 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 $f + f' = g$.

Exercises

7.1.1. Is the statement of Corollary 7.4 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 7.4 finite-dimensionality is used.

7.1.2. Let n be a positive integer and $F[x]_n$ the vector space of polynomials over F of degree at most n. Assume $\alpha_0, \alpha_1, \ldots, \alpha_n \in F$ are distinct elements. Let $T\colon F[x]_n \to F^{n+1}$ be the function given by

$$T(f) = \big(f(\alpha_0), f(\alpha_1), \ldots, f(\alpha_n)\big).$$

(1) Show that T is a linear map.
(2) Show that T is surjective (cf. Example 7.3).
(3) Show that T is an isomorphism.
(4) Show that for every $i \in \{0, \ldots, n\}$, 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 $i \ne j$.
(5) Show that $f_0, f_1, \ldots, f_n$ form a basis for $F[x]_n$.
(6) The polynomials $f_0, \ldots, f_n$ are called *Lagrange polynomials*. Give an explicit expression for them in terms of the elements $\alpha_0, \alpha_1, \ldots, \alpha_n$.

7.1.3. Let n be a positive integer and $T\colon F[x]_n \to F[x]_n$ 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.

7.2. The rank of a matrix

DEFINITION 7.6. Let $A \in \mathrm{Mat}(m \times n, F)$. 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 $A \in \mathrm{Mat}(m \times n, F)$, the image of f_A equals the column space $C(A) \subset F^m$ of A (see Remark 5.12). Therefore, we have $\operatorname{rk}(A) \leq \min\{m, n\}$, as the rank $\operatorname{rk} A$ is the dimension of a subspace of F^m, generated by n vectors.

By this definition, the rank of A is the same as the *column rank* of A, i.e., 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 definitions are not really necessary, as the row rank of any matrix equals the column rank.

THEOREM 7.7. *Let $A \in \mathrm{Mat}(m \times n, F)$ 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* $\dim C(A) = \dim R(A)$.

We will give several proofs of this important theorem.

PROOF. Clearly, any two of the three statements imply the third. Statement (1) is true because it is a restatement of Theorem 7.2, 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 5.13, 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 7.2, so statements (2) and (3) are equivalent.

Applying elementary row operations to A does not change $\ker A$ and $R(A)$ (see Proposition 5.31), 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. Again, any two of the three statements imply the third. Statement (1) is true because it is a restatement of Theorem 7.2, so statements (2) and (3) are equivalent and it suffices to prove statement (2). By Proposition 5.31, the subspaces $\ker A$ and $R(A)$ do not change under elementary row operations, so we may assume that A is in reduced row echelon form. Let r be the number of nonzero rows of A. Then from Example 6.24 we find $\dim R(A) = r$ and $\dim \ker A = n - r$, so $\dim R(A) + \dim \ker A = n$. □

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 7.8. Statement (3) of Theorem 7.7 can be stated as $\operatorname{rk} A = \operatorname{rk} A^\top$.

Remark 7.9. By statement (3) of Theorem 7.7, 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 Example 6.24.

Remark 7.10. The first proof, with the argument for the implication

$$\ker A' = \ker A \;\Rightarrow\; \dim C(A') = \dim C(A)$$

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 6.57), the dimension formula for general finite-dimensional vector spaces follows (again) from Proposition 4.14.

Remark 7.11. In Example 6.24, we found that for an $m \times n$ matrix in row echelon form with r nonzero rows, the $n - r$ elements w_k of Proposition 5.50 form a basis of the kernel of the matrix, as they generate the kernel (Proposition 5.50) and they are linearly independent (Proposition 6.20). Theorem 7.7, statement (2), shows independently that the dimension of the kernel equals $n - r$. Therefore, by Theorem 6.46, in order to show that the w_k form a basis, it suffices in hind sight to show only one of the two: either that they are linearly independent or that they generate the kernel.

Example 7.12. Consider the matrix

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

over $\mathbb{R}$. The reduce 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 $\mathrm{rk}(A) = 2$.

COROLLARY 7.13. *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 7.7. □

REMARK 7.14. Corollary 6.11 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 Corollary 7.13, this is the case if and only if $\mathrm{rk}\, A = n$. As we have $\mathrm{rk}\, A = \mathrm{rk}\, A^\top$, we may also put the n vectors as rows in a matrix and check that the rank of this matrix (namely $A^\top$) equals n. We could have also concluded this from Remark 6.25.

COROLLARY 7.15. *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 6.58 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 5.13. The first equality follows immediately from Theorem 7.7, 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 2.39), we conclude $(U^\perp)^\perp = U$ from Lemma 6.58. □

COROLLARY 7.16. *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 6.59 we then find

$$\dim(U + U^\perp) = \dim U + \dim U^\perp - \dim(U \cap U^\perp) = n - 0 = n,$$

so from Lemma 6.58 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 7.17. 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 7.16 is not true over general fields.

Exercises

7.2.1. Determine the rank of the matrices in Exercises 5.3.3 and 5.3.4.

7.2.2. Determine the rank of the matrices in Exercise 5.6.2.

7.2.3. Determine the rank of the linear maps and matrices of the exercises of Section 5.2.

7.2.4. Show that for any subset S of F^n, we have $L(S) = (S^\perp)^\perp$.

7.3. Computing intersections

PROPOSITION 7.18. *Suppose F is a field and $U_1, U_2 \subset F^n$ are subspaces. Then we have*

$$U_1 \cap U_2 = (U_1^\perp + U_2^\perp)^\perp \qquad \text{and} \qquad (U_1 \cap U_2)^\perp = U_1^\perp + U_2^\perp.$$

PROOF. Exercise, cf. Exercise 2.4.3. □

Proposition 7.18 expresses taking intersections in terms of taking sums and orthogonal subspaces. 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 2.45, 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 5.13 and we can compute generators for this kernel with Proposition 5.50.

EXAMPLE 7.19. 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 5.51, 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

$$\begin{pmatrix} 1 & -1 & 0 & 1 & 0 \\ 2 & -1 & -1 & 0 & 1 \\ -3 & 3 & 1 & 0 & 0 \\ -1 & 2 & 0 & 0 & 1 \end{pmatrix},$$

which has w_4, w_5, x_3, x_5 as rows. The reduced row echelon form of this matrix is

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

so the kernel $U \cap V$ 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 $\dim(U \cap V) = \dim(U^\perp + V^\perp)^\perp = 5 - 3 = 2$. 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 hyperplane H with normal $a = (2, -1, -1, 0, 1)$, which has dimension 4, 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 7.20. Let U and V be as in Example 7.19. Just as in Example 7.19, 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 7.21. 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 (i.e., 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 7.20 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 7.19 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

7.3.1. Prove Proposition 7.18.

7.3.2. Compute the intersection $U \cap V$ with U and V as in Example 7.19 with the method of Example 7.20, but with the roles of U and V reversed.

7.3.3. 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 $U \cap V$.

7.3.4. Take two subspaces of $\mathbb{R}^6$ generated by four elements and compute generators for the intersection.

7.4. Inverses of matrices

Recall that every invertible matrix is square by Corollary 6.56.

LEMMA 7.22. *An $n \times n$ matrix A is invertible if and only if* $\ker A = \{0\}$ *and if and only if* $\operatorname{rk} A = n$.

PROOF. By Corollary 7.4, a square matrix A is invertible if and only if f_A is injective, i.e., $\ker A = \{0\}$, and if and only if f_A is surjective, i.e., $\operatorname{rk} A = n$. □

In this section, we will give a method to check whether a square matrix is invertible, and, if so, to compute the inverse.

LEMMA 7.23. *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 5.32, 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, this also follows easily from the fact that the entries of C are the dot products of the rows of A and the columns of B, and the fact that the dot product is linear in its variables. □

Lemma 7.23 states that if we start with a product $AB = C$, written as

$$\begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix} = B \tag{6}$$

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{l1} & a_{l2} & \cdots & a_{lm} \end{pmatrix} \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{l1} & c_{l2} & \cdots & c_{ln} \end{pmatrix} = C$$

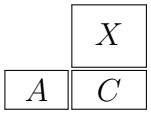
as in (5), and we perform an elementary row operation on the two bottom matrices simultaneously, then we obtain the matrices A' and C' and, together with B, these resulting matrices depict the equality $A'B = C'$.

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 we have $B = A^{-1}(AB) = A^{-1}C$. If A^{-1} is known, then this 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 7.24. *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.*

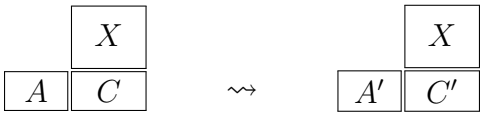
PROOF. If A is invertible, then f_A is injective, and by Proposition 5.52 we conclude that any row echelon form of A has n nonzero rows, so every row has a pivot and all pivots are on the diagonal; it follows that the *reduced* row echelon form is the identity matrix. Conversely, suppose that the reduced row echelon form of A is the identity matrix I_n. Then by Proposition 5.32 there is an invertible matrix M, such that $I_n = MA$, so $A = M^{-1}$ is invertible. Applying Lemma 7.23 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. □

Here is a visual interpretation of Proposition 7.24. If we write $X = A^{-1}C$ for A and C as in Proposition 7.24, then we can depict the equality $AX = C$ as in (6) by



.

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 7.23, depicted as follows.



In particular, if we obtain $A' = I$, then we have $C' = A'X = IX = X$.

$$\begin{array}{|c|c|}\hline & X \\ \hline A & C \\ \hline \end{array} \quad \rightsquigarrow \quad \begin{array}{|c|c|}\hline & X \\ \hline I & X \\ \hline \end{array}$$

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 7.25. Let us see how to invert the following matrix

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix},$$

where we assume $2 \neq 0$, so that we can divide by 2.

We perform the row operations on A and on I in parallel, as above.

$$\begin{aligned} \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 4 & 0 & 1 & 0 \\ 1 & 3 & 9 & 0 & 0 & 1 \end{array}\right) &\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 3 & -1 & 1 & 0 \\ 0 & 2 & 8 & -1 & 0 & 1 \end{array}\right) \\ &\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 0 & -2 & 2 & -1 & 0 \\ 0 & 1 & 3 & -1 & 1 & 0 \\ 0 & 0 & 2 & 1 & -2 & 1 \end{array}\right) \\ &\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 3 & -3 & 1 \\ 0 & 1 & 0 & -\frac{5}{2} & 4 & -\frac{3}{2} \\ 0 & 0 & 1 & \frac{1}{2} & -1 & \frac{1}{2} \end{array}\right) \end{aligned}$$

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 7.26. 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.

COROLLARY 7.27. *If $A \in \mathrm{Mat}(m, F)$ is invertible, then A can be written as a product of matrices $M_i(\lambda)$ ($\lambda \neq 0$) and $I_n + \lambda E_{ij}$ ($i \neq j$) of Section 5.4.*

PROOF. Exercise. □

EXAMPLE 7.28. Let A be the matrix of Example 7.25 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 7.23 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

7.4.1. Determine the inverses of the following matrices

$$\begin{pmatrix} -3 & -1 \\ -2 & -1 \end{pmatrix}, \quad \begin{pmatrix} -1 & -2 & -1 \\ 1 & 3 & 1 \\ 1 & -2 & 0 \end{pmatrix}, \quad \begin{pmatrix} -1 & 2 & -2 \\ 0 & -1 & 0 \\ 1 & -2 & 3 \end{pmatrix}, \quad \begin{pmatrix} 0 & -1 & 0 & 1 \\ 3 & -2 & -2 & 1 \\ -1 & -2 & -2 & 0 \\ 0 & 0 & -1 & -1 \end{pmatrix}.$$

7.4.2. Are the matrices

$$\begin{pmatrix} 1 & 2 \\ -2 & 4 \end{pmatrix}, \qquad \begin{pmatrix} -2 & 1 & -2 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix}$$

invertible?

7.4.3. Determine the inverse of those matrices (over $\mathbb{R}$) that are invertible.

$$\begin{pmatrix} 0 & -2 & -1 \\ -1 & 1 & 0 \\ -2 & -2 & 1 \end{pmatrix} \qquad \begin{pmatrix} -1 & 1 & -2 & 2 \\ -2 & 1 & 1 & -1 \\ 2 & -1 & 1 & 0 \\ 0 & 1 & 2 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 2 & -1 & 1 \\ -2 & -1 & -2 & 0 \\ 1 & 0 & -1 & 2 \\ 2 & 2 & 0 & 2 \end{pmatrix} \qquad \begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & -1 \\ 1 & 0 & 0 \end{pmatrix}.$$

7.4.4. Suppose the product AB of matrices $A, B \in \mathrm{Mat}(n, F)$ is invertible. Prove that A and B are also invertible. Cf. 5.3.1.

7.4.5. Suppose M, N are $n \times n$ matrices with $MN = I_n$. Prove that then also $NM = I_n$.

7.4.6. Let F be a field and m a positive integer. Recall the elementary matrices from Section 5.4.

(1) Show that multiplication by an elementary matrix (from the left) corresponds to applying an elementary row operation.

(2) Conclude that if A and A' are row equivalent, then there is an invertible matrix B such that $A' = BA$ (see Proposition 5.32).

(3) Prove that a matrix A is invertible if and only if A can be written as the product of elementary matrices.
(4) Prove Corollary 7.27.
(5) Write the following matrices as a product of elementary matrices, if possible:

$$\begin{pmatrix} 1 & -1 & 0 \\ -1 & -2 & -1 \\ 2 & 2 & 1 \end{pmatrix} \quad \begin{pmatrix} -1 & 0 & -2 \\ -1 & -1 & -2 \\ 2 & 3 & 3 \end{pmatrix} \quad \begin{pmatrix} 2 & 3 & -2 \\ 3 & 2 & 2 \\ 0 & -1 & 2 \end{pmatrix}$$

CHAPTER 8

Linear maps and matrices

8.1. The matrix associated to a linear map

Proposition 6.57 shows that any finite-dimensional vector space V over a field F is isomorphic with F^n for $n = \dim V$. For any basis B for V, there is an isomorphism $\varphi_B \colon F^n \to V$. As we have seen before in Proposition 4.14, 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 8.1. 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 $[f]_C^B$, is the unique $m \times n$ matrix whose associated function 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 identify the matrix $[f]_C^B$ with the map $F^n \to F^m$ it defines, then we have the following commutative diagram.

(7)
$$\begin{array}{ccc} V & \xrightarrow{f} & W \\ {\scriptstyle\cong}\uparrow{\scriptstyle\varphi_B} & & {\scriptstyle\cong}\uparrow{\scriptstyle\varphi_C} \\ F^n & \xrightarrow[[f]_C^B]{} & F^m \end{array}$$

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 $[f]_C^B$.

Example 8.2. Let A be an $m \times n$ matrix over F, and let $f_A \colon F^n \to F^m$ be the associated linear map. Then with respect to the standard bases E_n and E_m for F^n and F^m, respectively, we have

$$[f_A]_{E_m}^{E_n} = A,$$

which follows from the fact that φ_{E_n} and φ_{E_m} are the identity maps on F^n and F^m, respectively.

Example 8.3. 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'.

$$\begin{array}{ccc} a_0 + a_1x + a_2x^2 + a_3x^3 & & \mathbb{R}[x]_3 \xrightarrow{D} \mathbb{R}[x]_3 \\ \uparrow & & \cong\uparrow\varphi_B \qquad \cong\uparrow\varphi_B \\ (a_0, a_1, a_2, a_3) & & \mathbb{R}^4 \xrightarrow[[D]_B^B]{} \mathbb{R}^4 \end{array}$$

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 8.4. 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 $[T]_E^B$, this vector also equals $[T]_E^B \cdot e_j$, i.e., the j-th column of $[T]_E^B$, cf. Lemma 5.5. Hence, we find

$$[T]_E^B = \begin{pmatrix} 1 & \alpha_1 & \alpha_1^2 & \cdots & \alpha_1^n \\ 1 & \alpha_2 & \alpha_2^2 & \cdots & \alpha_2^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_k & \alpha_k^2 & \cdots & \alpha_k^n \end{pmatrix}.$$

Such a matrix is called a *Vandermonde matrix*.

DEFINITION 8.5. 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 8.6. *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 8.7. Indeed, in Example 8.4, the columns are as described in Lemma 8.6. Also in Example 8.3, 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 8.8. If we identify $[f]_C^B$ with the linear map that it induces, then the commuting diagram (7) can also be expressed as $\varphi_C^{-1} \circ f = [f]_C^B \circ \varphi_B^{-1}$, i.e., 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 $[f]_C^B$ and the sequence of coefficients of v with respect to B.

EXAMPLE 8.9. 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 $\mathrm{id}\colon F[x]_3 \to F[x]_3$ 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 8.6 (the j-th column contains the sequence of coefficients of $(x-1)^{4-j}$ with respect to C), but the identity

$$\begin{aligned} a_1(x-1)^3 + a_2(x-1)^2 + a_3(x-1) + a_4 \\ = (-a_1 + a_2 - a_3 + a_4) + (3a_1 - 2a_2 + a_3)x + (-3a_1 + a_2)x^2 + a_3x^3 \end{aligned}$$

also shows that $[\mathrm{id}]_C^B$ sends the quadruple (a_1, a_2, a_3, a_4) to

$$\big(-a_1 + a_2 - a_3 + a_4,\ \ 3a_1 - 2a_2 + a_3,\ \ -3a_1 + a_2,\ \ a_3\big).$$

EXAMPLE 8.10. 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 $s\colon \mathbb{R}^3 \to \mathbb{R}^3$ 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 8.11. Let $B = (v_1, v_2, v_3)$ be the basis for $\mathbb{R}^3$ as in Example 8.10, 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

$$\mathrm{id}\big(\varphi_B(a)\big) = a_1v_1 + a_2v_2 + a_3v_3 = \begin{pmatrix} a_1 + a_2 + a_3 \\ 2a_1 + a_2 - a_3 \\ a_1 + a_3 \end{pmatrix} = \varphi_E\big([\mathrm{id}]_E^B \cdot a\big).$$

Exercises

8.1.1. 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 $B = (1, x, x^2, x^3, x^4)$.

8.1.2. 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 7.1.2 and Example 8.4.

8.1.3. 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

$$B = \left(\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right)$$

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

$$T(M) = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \cdot M.$$

Determine $[T]_C^B$.

8.2. 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 $[g]_B^A$ and $[f]_C^B$. The two commuative diagrams as in (7) can be combined into one.

(8)
$$\begin{array}{ccccc} U & \xrightarrow{\ g\ } & V & \xrightarrow{\ f\ } & W \\ {\scriptstyle\cong}\uparrow{\scriptstyle\varphi_A} & & {\scriptstyle\cong}\uparrow{\scriptstyle\varphi_B} & & {\scriptstyle\cong}\uparrow{\scriptstyle\varphi_C} \\ F^p & \xrightarrow[[g]_B^A]{} & F^n & \xrightarrow[[f]_C^B]{} & F^m \end{array}$$

Proposition 8.12. *With the notation as above, we have* $[f \circ g]_C^A = [f]_C^B \cdot [g]_B^A$.

PROOF. The commutative diagram above simplifies to the following diagram.

$$\begin{array}{ccc} U & \xrightarrow{f\circ g} & W \\ \cong\uparrow{\scriptstyle \varphi_A} & & \cong\uparrow{\scriptstyle \varphi_C} \\ F^p & \xrightarrow[[f]_C^B\cdot[g]_B^A]{} & F^m \end{array}$$

In other words, identifying matrices with the maps they induce, we obtain from the identities

$$[f]_C^B = \varphi_C^{-1} \circ f \circ \varphi_B \qquad \text{and} \qquad [g]_B^A = \varphi_B^{-1} \circ g \circ \varphi_A,$$

that

$$[f]_C^B \cdot [g]_B^A = \varphi_C^{-1} \circ (f \circ g) \circ \varphi_A = [f \circ g]_C^A,$$

which proves the statement. □

ALTERNATIVE PROOF. By first multiplying the sequence of coefficients with respect to A of a vector $u \in U$ with the matrix $[g]_B^A$, we obtain the sequence of coefficients of $g(u)$ with respect to B; multiplying that vector with the matrix $[f]_C^B$ yields the sequence of coefficients of $f(g(u))$ with respect to C. In other words, we have

$$\big(f(g(u))\big)_C = [f]_C^B \cdot \big(g(u)\big)_B = [f]_C^B \cdot [g]_B^A \cdot u_A.$$

Similarly, we have

$$\big(f(g(u))\big)_C = \big((f \circ g)(u)\big)_C = [f \circ g]_C^A \cdot u_A.$$

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

$$[f]_C^B \cdot [g]_B^A \cdot e_j = [f \circ g]_C^A \cdot e_j$$

for all j. This shows that the two matrices $[f]_C^B \cdot [g]_B^A$ and $[f \circ g]_C^A$ have the same columns, so they are equal. □

Note that the order of f and g in the product $[f]_C^B \cdot [g]_B^A$ of matrices, and in the composition $f \circ g$, is opposite of the order in which they appear in diagram (8).

COROLLARY 8.13. *With the notation as above, if f is an isomorphism, then we have $[f^{-1}]_B^C = \big([f]_C^B\big)^{-1}$.*

PROOF. If f is an isomorphism, then $m = n$, and $[f]_C^B$ is a square matrix. Apply Proposition 8.12 with $g = f^{-1}$ and $A = C$ to find

$$[f]_C^B \cdot [f^{-1}]_B^C = [\mathrm{id}]_C^C = I_m.$$

The statement follows. □

EXAMPLE 8.14. Let B and E be the bases for $\mathbb{R}^3$ as in Example 8.11. Then

$$[\mathrm{id}]_B^E = \big([\mathrm{id}]_E^B\big)^{-1} = \begin{pmatrix} -\frac{1}{3} & \frac{1}{3} & \frac{2}{3} \\ 1 & 0 & -1 \\ \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

Since the sequence of coefficients of any vector $v \in \mathbb{R}^3$ with respect to E is equal to itself, this shows

$$v_B = (\mathrm{id}(v))_B = [\mathrm{id}]_B^E \cdot v_E = [\mathrm{id}]_B^E \cdot v,$$

so the sequence of coefficients of a vector v with respect to B equals $[\mathrm{id}]_B^E \cdot v$. Indeed, the sequence of coefficients with respect to B of the j-th standard vector is the j-th column of $[\mathrm{id}]_B^E$, 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 8.15. 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 8.3 and 8.4.

Exercises

8.2.1. 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 $v = x_1v_1' + x_2v_2' + x_3v_3' + x_4v_4'$.
(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 $v = x_1v_1 + x_2v_2 + x_3v_3 + x_4v_4$.

8.2.2. Let $E = (e_1, e_2, e_3)$ be the standard basis for $\mathbb{R}^3$ and $B = (v_1, v_2, v_3)$ a basis with

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

Determine the matrices $[\mathrm{id}]_E^B$ and $[\mathrm{id}]_B^E$.

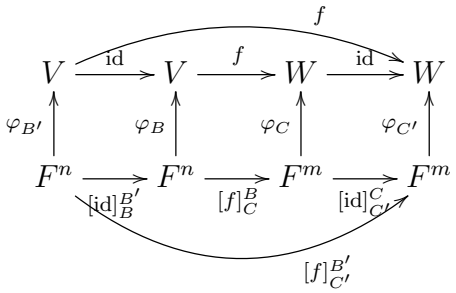
8.3. Changing bases

PROPOSITION 8.16. *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*

$$[f]_{C'}^{B'} = [\mathrm{id}]_{C'}^{C} \cdot [f]_C^B \cdot [\mathrm{id}]_B^{B'}. \tag{9}$$

PROOF. This follows immediately from Proposition 8.12. □

The following commuting diagram corresponds to the identity (9) of Proposition 8.16.



In the spirit of the alternative proof of Proposition 8.12, we can explain the identity (9) 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 $[f]^{B}_{C}$ 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 $[f]^{B'}_{C'}$ with the vector $v_{B'}$. In other words, we have

$$[f]^{B'}_{C'} \cdot u_{B'} = \big(f(v)\big)_{C'} = \big([\mathrm{id}]^{C}_{C'} \cdot [f]^{B}_{C} \cdot [\mathrm{id}]^{B'}_{B}\big) \cdot u_{B'}$$

for all $u \in U$, in particular for the j-th element of the basis B', for which we have $u_{B'} = e_j \in F^n$. So we find

$$[f]^{B'}_{C'} \cdot e_j = \big([\mathrm{id}]^{C}_{C'} \cdot [f]^{B}_{C} \cdot [\mathrm{id}]^{B'}_{B}\big) \cdot e_j$$

for all j. This shows that the two matrices $[f]^{B'}_{C'}$ and $[\mathrm{id}]^{C}_{C'} \cdot [f]^{B}_{C} \cdot [\mathrm{id}]^{B'}_{B}$ have the same columns, so they are equal.

Note again that the order of the matrices in the right-hand side of (9) is opposite of the order in which they appear in this diagram. Because of Proposition 8.16, the matrices

$$[\mathrm{id}]^{B'}_{B} = \varphi_B^{-1} \circ \varphi_{B'} \qquad \text{and} \qquad [\mathrm{id}]^{C}_{C'} = \varphi_{C'}^{-1} \circ \varphi_C$$

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'.

LEMMA 8.17. *Suppose V is an n-dimensional vector space over F with basis B. Then for every invertible matrix P, there is a basis B' for V such that $[\mathrm{id}]^{B'}_{B} = P$.*

PROOF. Set $w_j = \varphi_B(P \cdot e_j)$ and $B' = (w_1, w_2, \ldots, w_n)$. Then we have $\varphi_{B'} = \varphi_B \circ P$, so $\varphi_{B'}$ is invertible and B' is a basis. From $P = \varphi_B^{-1} \circ \mathrm{id} \circ \varphi_{B'}$, we conclude $P = [\mathrm{id}]^{B'}_{B}$. □

PROPOSITION 8.18. *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 8.16, 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 8.17. Then (by Proposition 8.16 again) we have $QMP = [f]_{C'}^{B'}$. □

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 8.19. *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 $f : V \to W$ be the linear map given by M. Let $B = (v_1, \ldots, v_n)$ be a basis of V such that $v_{r+1}, \ldots, v_n$ is a basis of $\ker(f)$. Then $w_1 = f(v_1), \ldots, w_r = f(v_r)$ are linearly independent in W, and we can extend to a basis $C = (w_1, \ldots, w_m)$. 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 $M' = [f]_C^B$ associated to f with respect to B and C has the required form. Set $P = [\mathrm{id}]_{E_n}^B$ and $Q = [\mathrm{id}]_C^{E_m}$, where E_n and E_m are the standard bases of F^n and F^m, respectively. Then by Proposition 8.16, we have

$$M' = [f]_C^B = [\mathrm{id}]_C^{E_m} \cdot [f]_{E_m}^{E_n} \cdot [\mathrm{id}]_{E_n}^B = QMP,$$

as M is the matrix associated to f relative to the standard bases E_n and E_m. □

REMARK 8.20.

(1) If 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 $M' = QMP$ (exercise: this really defines an equivalence relation), then Corollary 8.19 tells us that M and M' are equivalent if and only if $\mathrm{rk}(M) = \mathrm{rk}(M')$. To see this, 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). Then Corollary 8.19 tells us that if M has rank r, it is equivalent to the matrix given there, so any two matrices of rank r are equivalent to the same matrix.

(2) Recall that by Proposition 5.32, 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. Interpreting M as the matrix associated to a linear map relative to some bases, we see that row operations correspond to changing the basis of the target space (containing the columns) of M, whereas column operations correspond to changing the basis of the domain space (containing the rows) of M. The result of Corollary 8.19 then also means 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 7.7.

Exercises

8.3.1. 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

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

as in Exercise 8.2.2.

8.3.2. Let $V \subset \mathbb{R}^3$ be the subspace spanned by v_1 and v_3 as in Exercise 8.3.1. 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 8.3.1.
(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 the other with $[\mathrm{id}]_E^C$ on the other hand.

8.3.3. 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 $[T]_{C'}^{B'}$ 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) Show that for the vector $v \in \mathbb{R}^2$ with $v_{B'} = (1,1)$ (i.e., $v = \varphi_{B'}((1,1))$), we indeed have

$$[T]_{C'}^{B'} \cdot v_{B'} = \big(T(v)\big)_{C'}.$$

(4) Repeat this verification for $v_{B'} = (1,0)$ and $v_{B'} = (0,1)$.

8.4. Endomorphisms

In the special case of Proposition 8.16 that we have $V = W$, we can take $B = C$ and $B' = C'$ to obtain the following.

PROPOSITION 8.21. *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 8.16 and Corollary 8.13. □

EXAMPLE 8.22. Let $B = (v_1, v_2, v_3)$ be the basis for $\mathbb{R}^3$ as in Examples 8.10, 8.11, and 8.14. As in Example 8.10, 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

$$\begin{aligned}[s]_E^E &= [\mathrm{id}]_E^B \cdot [s]_B^B \cdot [\mathrm{id}]_B^E = [\mathrm{id}]_E^B \cdot [s]_B^B \cdot \left([\mathrm{id}]_E^B\right)^{-1} \\ &= \begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & -1 \\ 1 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} -\frac{1}{3} & \frac{1}{3} & \frac{2}{3} \\ 1 & 0 & -1 \\ \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ -\frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{pmatrix}.\end{aligned}$$

EXAMPLE 8.23. Let $B = (v_1, v_2, v_3)$ be the basis for $\mathbb{R}^3$ as in Example 8.22 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 8.22, we find the matrix $[\pi]_E^E$ with Proposition 8.16:

$$\begin{aligned}[\pi]_E^E &= [\mathrm{id}]_E^B \cdot [\pi]_B^B \cdot [\mathrm{id}]_B^E = [\mathrm{id}]_E^B \cdot [\pi]_B^B \cdot \left([\mathrm{id}]_E^B\right)^{-1} \\ &= \begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & -1 \\ 1 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} -\frac{1}{3} & \frac{1}{3} & \frac{2}{3} \\ 1 & 0 & -1 \\ \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} & \frac{2}{3} \end{pmatrix}.\end{aligned}$$

DEFINITION 8.24. 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 $\mathrm{Mat}(n, F)$ (exercise). Proposition 8.21 shows that any two matrices associated to the same endomorphism of V, but possibly with respect to different bases are similar. Conversely, any two similar $n \times n$ matrices over F describe the same endomorphism of F^n with respect to some carefully chosen bases (exercise).

We have seen that it is easy to classify matrices in $\mathrm{Mat}(m \times n, F)$ with respect to equivalence: the equivalence class is determined by the rank. In contrast to this, the classification of matrices in $\mathrm{Mat}(n, f)$ with respect to similarity is much more complicated. For example, 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 $\lambda \neq \mu$.

EXAMPLE 8.25. As another example, consider the matrices

$$M_{\lambda,t} = \begin{pmatrix} \lambda & t \\ 0 & \lambda \end{pmatrix}.$$

The corresponding endomorphism $f_{\lambda,t} = f_{M_{\lambda,t}}$ has $\ker(f_{\lambda,t} - \mu\,\mathrm{id}) = 0$ if $\lambda \neq \mu$, and has nontrivial kernel otherwise. This shows that $M_{\lambda,t}$ and $M_{\mu,u}$ can be similar only when $\lambda = \mu$. Since $\dim \ker(f_{\lambda,t} - \lambda \cdot \mathrm{id})$ is 1 if $t \neq 0$ and 2 if $t = 0$, $M_{\lambda,0}$ 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 you a first glimpse of the classification theorem, the 'Jordan Normal Form Theorem'.

For purposes of classification, it is useful to have *invariants*, i.e., functions that are constant on the equivalence classes. In the case of equivalence of matrices, the rank is an invariant, and in this case, it gives the complete classification. The rank is (of course) still an invariant with respect to similarity, but as the example above shows, it is by no means sufficient to separate the classes. Here is another invariant.

DEFINITION 8.26. 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 8.27. *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 8.28. *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 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 8.29. 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 have the same trace according to Corollary 8.28.

In the next chapter, we will introduce another invariant, which is even more important than the trace: the determinant.

REMARK 8.30. To finish off this section, let us remark that, having chosen bases B and C of the F-vector spaces V and W of dimensions n and m, respectively, we obtain an isomorphism

$$\mathrm{Hom}(V, W) \xrightarrow{\cong} \mathrm{Mat}(m \times n, F)\,, \quad f \longmapsto [f]_C^B\,.$$

In particular, we see that $\dim \mathrm{Hom}(V, W) = mn$.

Exercises

8.4.1. 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.

8.4.2. 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 $[T]_C^C$.
(3) Determine $[T]_B^B$ again, this time using $[T]_C^C$ and a basis change matrix.

8.4.3. 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 $V^\perp$.

8.4.4. Determine the trace of the following three matrices.

$$M_1 = \begin{pmatrix} 1 & 2 & 2 & 1 \\ 4 & -3 & 5 & 2 \\ -2 & 1 & 5 & 11 \\ 3 & 2 & 7 & -13 \end{pmatrix}$$

$$M_2 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 4 & 9 & 16 \\ 1 & 8 & 27 & 64 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 2 & 2 & 1 \\ 4 & -3 & 5 & 2 \\ -2 & 1 & 5 & 11 \\ 3 & 2 & 7 & -13 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 4 & 9 & 16 \\ 1 & 8 & 27 & 64 \end{pmatrix}$$

$$M_3 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 5 & 7 \\ 1 & 25 & 49 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 5 & 6 \\ 0 & 2 & 7 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 2 & 2 \\ 4 & -3 & 5 \\ -2 & 1 & 5 \end{pmatrix} \begin{pmatrix} 1 & 5 & 6 \\ 0 & 2 & 7 \\ 0 & 0 & 3 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 5 & 7 \\ 1 & 25 & 49 \end{pmatrix}$$

8.4.5. Show that the notion of similarity defines an equivalence relation on the space $\mathrm{Mat}(n, F)$ of $n \times n$ matrices, as claimed.

8.4.6. Let M and M' be two similar $n \times n$ matrices over F. Show that there exists a basis B of F^n such that for $M' = [f_M]_B^B$.

8.4.7. Prove Remark 8.30.

CHAPTER 9

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, i.e., $\det(f \circ g) = (\det f) \cdot (\det g)$.

9.1. 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 9.13). 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 given by the matrix that has $v_1, \ldots, v_n$ as columns.

Now let $D\colon \mathrm{Mat}(n, \mathbb{R}) \to \mathbb{R}$ be a function that is supposed to measure oriented volume of n-dimensional parallelotopes — $D(A)$ gives the volume of the image of the 'unit cube' $P(e_1, \ldots, e_n)$ under f_A, i.e., the volume of $P(v_1, \ldots, v_n)$, where $v_1, \ldots, v_n$ are the columns of A. What properties should such a function D 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 $1 \leq j, k \leq n$ 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 volume should scale corresponding to scaling of the vectors, i.e.,

$$D(r_j(A, \lambda x)) = \lambda D(r_j(A, x))\,. \tag{10}$$

Also, volumes should be additive in the following sense:

$$D(r_j(A, x + y)) = D(r_j(A, x)) + D(r_j(A, y))\,. \tag{11}$$

This corresponds to the fact that 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 height of $P(v_1, \ldots, v_n)$ with respect to this base; this height is the oriented length of the projection of the j-th column onto a, which is indeed additive in the j-th column. (Draw a picture supporting this argument for $n = 2$ and $n = 3$!)

These two properties (10) and (11) can be stated simply by saying that D is linear in each column separately, when the other $n-1$ columns are held constant, i.e., for each $n \times n$ matrix A and each $1 \leq j \leq n$, the function $F^n \to F,\ x \mapsto D(r_j(A,x))$ is linear. Such a function $\mathrm{Mat}(n,\mathbb{R}) \to \mathbb{R}$ is said to be *multilinear* as function in the columns.

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, i.e., 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 9.2(1)), i.e.,

$$D(r_{ij}(A,x,x)) = 0 \text{ if } 1 \leq i,j \leq n \text{ and } i \neq j.$$

A function $\mathrm{Mat}(n,\mathbb{R}) \to \mathbb{R}$ that is multilinear in the columns and that satisfies this third property is said to be *alternating*. So these are the functions we are looking for. Note that it makes sense to talk about functions $\mathrm{Mat}(n,F) \to F$ that are multilinear and alternating in the columns for any field F.

Definition 9.1. Let F be a field and n 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 9.2. *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, i.e.,*
$$D(r_{ij}(A,x,y)) = D(r_{ij}(A,x,y+\lambda x)).$$
(3) *If we interchange two of the columns, then $D(A)$ changes sign, i.e., for $i \neq j$ we have*
$$D(r_{ij}(A,x,y)) = -D(r_{ij}(A,y,x)).$$

Proof.

(1) If $A \in \mathrm{Mat}(n,F)$ 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.

(2) By linearity of D in the j-th column and the fact that D is alternating, we have

$$D(r_{ij}(A, x, y + \lambda x)) = D(r_{ij}(A, x, y)) + \lambda D(r_{ij}(A, x, x)) = D(r_{ij}(A, x, y)).$$

(3) For any $x, y \in F^n$, we have

$$\begin{aligned} 0 = D(r_{ij}(A, x + y, x + y)) =& D(r_{ij}(A, x, x)) + D(r_{ij}(A, x, y)) + D(r_{ij}(A, y, x)) \\ &+ D(r_{ij}(A, y, y)) = D(r_{ij}(A, x, y)) + D(r_{ij}(A, y, x)), \end{aligned}$$

so $D(r_{ij}(A, x, y)) = -D(r_{ij}(A, y, x))$.

□

PROPOSITION 9.3. *For any field F, nonnegative 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 9.2(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 9.2 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 9.3 easily to show existence of a determinantal function on F^n (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 9.4. We define the functions

$$d_n\colon \mathrm{Mat}(n, F) \to F$$

(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

$$d_n(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \cdot d_{n-1}(A_{ij}), \tag{12}$$

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.

PROPOSITION 9.5. *For any integer $n \geq 0$, the function $d_n\colon \mathrm{Mat}(n, F) \to F$ is a determinantal function with $d_n(I_n) = 1$ that is independent of the choice of i in Definition 9.4.*

PROOF. We use induction on n. For $n = 0$ the statement is trivial. (If you suffer from *horror vacui,* i.e., you are afraid of the empty set, you can consider $n = 1$; then $d_1\colon \mathrm{Mat}(1, F) \to F$ 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 9.4.

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 $\mathrm{Mat}(n, F) \to F$ that sends A to $d_{n-1}(A_{ij})$ is linear as a function in the k-th column of A for $j \neq k$ and constant for $j = k$; The function $A \mapsto a_{ij}$ is constant as a function in the k-th column of A for $j \neq k$ and linear for $j = k$, so the j-th term in the right-hand side of (12) 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 9.2(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 9.3, we conclude that d_n is uniquely determined, so it is independent of the choice of i, which finishes the proof. □

COROLLARY 9.6. *The determinantal functions* $\mathrm{Mat}(n, F) \to F$ *form a* 1*-dimensional* F*-vector space.*

PROOF. From Proposition 9.3, it follows that the dimension is at most 1, while Proposition 9.5 implies it is at least 1. □

DEFINITION 9.7. For any field F and any nonnegative integer n, we let

$$\det\colon \mathrm{Mat}(n, F) \to F$$

be the unique determinantal function with $\det(I_n) = 1$; for any matrix $A \in \mathrm{Mat}(n, F)$, we call $\det(A)$ the *determinant* of A.

Note that the field F and the dimension n are not explicit in the notation det; by Proposition 9.5, 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 (12) we have

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}) \tag{13}$$

for all $1 \leq i \leq n$; this is called the *expansion of the determinant by the* i*-th row.*

EXAMPLE 9.8. For 2×2 matrices and 3×3 matrices, we find

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

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + cdh - afh - bdi - ceg.$$

EXAMPLE 9.9. If one of the rows of a square matrix contains many zeros, then it is useful to expand the determinant by that row. If we expand the following determinant by the second row, then we get

$$\begin{vmatrix} 1 & -1 & 2 & 1 \\ 1 & 0 & 2 & 0 \\ 2 & 1 & 2 & 1 \\ 3 & -1 & 1 & 0 \end{vmatrix} = - \begin{vmatrix} -1 & 2 & 1 \\ 1 & 2 & 1 \\ -1 & 1 & 0 \end{vmatrix} - 2 \begin{vmatrix} 1 & -1 & 1 \\ 2 & 1 & 1 \\ 3 & -1 & 0 \end{vmatrix} = -1 \cdot 2 - 2 \cdot (-7) = 12.$$

EXAMPLE 9.10. 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.

The proof of Proposition 9.3 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 9.11. We compute a determinant by elementary column operations. Note that we can avoid divisions (and hence fractions) by choosing the operations cleverly, cf. Example 5.42.

$$\begin{vmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \\ 3 & 4 & 2 & 1 \\ 4 & 3 & 1 & 2 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 2 & -3 & -2 & -5 \\ 3 & -2 & -7 & -11 \\ 4 & -5 & -11 & -14 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & -2 & -5 \\ 3 & 12 & -7 & -11 \\ 4 & 17 & -11 & -14 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -21 & 12 & 17 & 49 \\ -30 & 17 & 23 & 71 \end{vmatrix}$$

$$= \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -21 & 12 & 17 & -2 \\ -30 & 17 & 23 & 2 \end{vmatrix} = 2 \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -21 & 12 & 1 & 17 \\ -30 & 17 & -1 & 23 \end{vmatrix} = 2 \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -51 & 29 & -1 & 40 \end{vmatrix}$$

$$= 2 \cdot 40 \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} = 80$$

PROPOSITION 9.12. *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 $D_1(I_n) = \det A$. The function D_2 is easily seen to be 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 9.3 we conclude $D_1 = D_2$ and in particular $D_1(B) = D_2(B)$, i.e., $\det(AB) = (\det A) \cdot (\det B)$. □

REMARK 9.13. 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 9.12 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, i.e., 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 9.14. *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 9.12, we have

$$(\det(A^{-1})) \cdot (\det A) = \det(A^{-1}A) = \det(I_n) = 1,$$

from which the statement follows. □

REMARK 9.15. A square matrix A is invertible if and only if $\det A \neq 0$, because if A is not invertible, then $\det A = 0$ by Lemma 9.2, and if A is invertible, then $\det A \neq 0$ by Corollary 9.14.

THEOREM 9.16. *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, we have

$$\det(A) = 0 \iff \operatorname{rk}(A) < n \iff \operatorname{rk}(A^\top) < n \iff \det(A^\top) = 0\,,$$

so 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 ith row, we expand $\det(A)$ by the ith row according to (13). 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 ith row of A has been removed, so $\det(A_{ij})$ does not depend on the ith 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. □

COROLLARY 9.17 (Expansion by Columns). *We can also expand determinants by 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 $A^\top$ by the j-th row as in (13), but 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 9.18. Just as Lemma 9.2 tells us how the determinant of a matrix behaves under elementary column operations, we conclude from Theorem 9.16 that it behaves similarly under elementary row operations.

EXAMPLE 9.19. 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 $\det(A) = \pm 1$.

DEFINITION 9.20. 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 is $(-1)^{i+j} \det(A_{ji})$. Here A_{ij} is, as before, the matrix obtained from A by removing the ith row and jth column. Note the reversal of indices — $\tilde{A}_{ij} = (-1)^{i+j} \det(A_{ji})$ and not $\det(A_{ij})$!

PROPOSITION 9.21 (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. 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 above. The proposition 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 $A^\top$). □

EXAMPLE 9.22. 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

9.1.1. Determine the determinants of the following matrices, both by expansion by a row or column, or using elementary row and/or column operations.

$$\begin{pmatrix} -1 & -2 \\ -3 & -2 \end{pmatrix} \quad \begin{pmatrix} -2 & -3 & 2 \\ 0 & 1 & 2 \\ -3 & -3 & 0 \end{pmatrix} \quad \begin{pmatrix} 2 & -2 & -2 \\ 1 & 3 & -1 \\ 2 & -2 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & -2 & -2 & -1 \\ 1 & -1 & -1 & 2 \\ -2 & -2 & 0 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix} \quad \begin{pmatrix} -3 & 2 & 1 & 2 \\ -1 & -1 & -3 & 1 \\ 3 & -2 & -3 & -2 \\ 3 & -2 & -1 & -1 \end{pmatrix}$$

9.1.2. 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. Prove the same for *lower triangular* matrices.

9.1.3. Let A, B be two $n \times n$ matrices. True or not true?

(1) $\operatorname{Tr} AB = \operatorname{Tr} BA$.
(2) $\operatorname{Tr} AB = (\operatorname{Tr} A)(\operatorname{Tr} B)$.
(3) $\operatorname{Tr} A + B = \operatorname{Tr} A + \operatorname{Tr} B$.
(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.

9.1.4. 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.

9.1.5. 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 $A \in \operatorname{Mat}(m, F)$ and $C \in \operatorname{Mat}(n, F)$, and $B \in \operatorname{Mat}(m \times n, F)$ and where 0 denotes the zero matrix in $\operatorname{Mat}(n \times m, F)$. Show that $\det M = (\det A) \cdot (\det C)$.

9.1.6. 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 zeros below the diagonal blocks, we have

$$\det A = (\det A_{11})(\det A_{22}) \cdots (\det A_{tt}).$$

9.1.7. 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?

9.2. Determinants of endomorphisms

DEFINITION 9.23. Let $f\colon V \to V$ 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.

The fact that the choice of basis B is not reflected in the notation $\det f$ is justified by the following proposition.

PROPOSITION 9.24. *The determinant* $\det f$ *of an endomorphism* $f\colon V \to V$ *of a finite-dimensional vector space with basis* B *is independent of the basis* B.

PROOF. Let B' be a second basis for V. Then with $P = [\mathrm{id}]_{B'}^B$ we have $P^{-1} = [\mathrm{id}]_B^{B'}$ and $[f]_{B'}^{B'} = P \cdot [f]_B^B \cdot P^{-1}$, so

$$\det[f]_{B'}^{B'} = (\det P) \cdot (\det[f]_B^B) \cdot (\det P^{-1}) = (\det P) \cdot (\det[f]_B^B) \cdot (\det P)^{-1} = \det[f]_B^B.$$

This shows that the determinant $\det f = \det([f]_B^B)$ is indeed independent of the choice of basis B. □

EXAMPLE 9.25. For the identity $\mathrm{id}_V\colon V \to V$ we have $\det \mathrm{id}_V = 1$.

EXAMPLE 9.26. By Example 8.2, we of course have $\det f_A = \det[f_A]_E^E = \det A$ for any square matrix A.

EXAMPLE 9.27. Let $V \subset \mathbb{R}^3$ be a plane and $s\colon \mathbb{R}^3 \to \mathbb{R}^3$ the reflection in V, cf. Examples 8.10 and 8.22. 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 8.10, 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 9.28. *For any two endomorphisms* $f, g\colon V \to V$ *of a finite-dimensional vector space* V*, we have* $\det(f \circ g) = (\det f) \cdot (\det g)$.

PROOF. Choose a basis B for V. Then from Proposition 9.12 we find

$$\det(f \circ g) = \det([f \circ g]_B^B) = \det([f]_B^B \cdot [g]_B^B) = (\det[f]_B^B)(\det[g]_B^B) = (\det f)(\det g).$$

□

Note that by Corollary 7.4, an endomorphism $f\colon V \to V$ of a finite-dimensional vector space V is an isomorphism if and only if it is injective, so if and only if $\ker f \neq \{0\}$. By the following proposition, this happens if and only if $\det f \neq 0$.

PROPOSITION 9.29. *Let $f\colon V \to V$ 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 4.14, the map f is an isomorphism if and only if the matrix $[f]_B^B$ is invertible. By Remark 9.15, this is the case if and only if $\det([f]_B^B) \neq 0$. □

Exercises

9.2.1. Determine the determinant of the following linear maps.
(1) $f\colon \mathbb{R}^3 \to \mathbb{R}^3, (x, y, z) \mapsto (2x + z, y - 3z, -x + 2y + 3z)$,
(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 $x - 2y + z = 0$,
(4) the map $\mathbb{R}[x]_3 \to \mathbb{R}[x]_3$ given by $f \mapsto xf'$ with f' the derivative of f,

9.2.2. 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 $\det f' = \det f$.

9.2.3. 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 4.3.4). Show that $\det f = (\det f') \cdot (\det f'')$.
[Hint: use Exercise 9.1.5.]

9.3. Linear equations

The system

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \qquad\qquad\qquad\qquad\qquad\quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

of m linear equations in n variables $x_1, \ldots, x_n$ over a field F can also be written as $Ax = b$ with

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \in \operatorname{Mat}(m \times n, F)$$

and

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \qquad \text{and} \qquad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \in F^m.$$

In terms of the linear map $f_A\colon F^n \to F^m$, the solution set equals

$$\{\ x \in F^n\ :\ Ax = b\ \} = f_A^{-1}(b).$$

Thus solving systems of linear equations comes down to determining inverse images under linear maps.

DEFINITION 9.30. 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, i.e., if $b \in \mathrm{im}(f)$.

With the theory we have built so far, the following result is essentially trivial.

THEOREM 9.31. *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*

$$f^{-1}(b) = \{\ a + z\ :\ z \in \ker f\ \}. \tag{14}$$

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 (14). 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 9.32. 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.

Let us now look at the more familiar case where $V = F^n$ and $W = F^m$, so that we have a system of m linear equations in n variables. This is most conveniently written in matrix notation as $Ax = 0$ in the homogeneous case and $Ax = b$ in the inhomogeneous case.

Algorithm for homogeneous equations. 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 5.51.

Algorithm for inhomogeneous equations. To solve an inhomogeneous system of linear equations $Ax = b$, we do the same as in Example 7.28 (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 the last column does not contain a pivot. In this case, the first n coordinates of $-w_{n+1}$ (in the notation of Proposition 5.51) give a solution of the system, but such a solution can also be easily 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 9.31) 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 7.4, as

$$\begin{array}{|c|c|}\hline \multicolumn{1}{|c|}{} & x \\ \hline A & b \\ \hline \end{array}\,.$$

Applying elementary row operations to the combined matrix $A^\circ = \begin{array}{|c|c|}\hline A & b \\ \hline \end{array}$ yields a combined matrix $\begin{array}{|c|c|}\hline A' & b' \\ \hline \end{array}$, for which the solution set to the equation $A'x = b'$ is the same as the solution set to the original equation $Ax = b$. 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, i.e., if and only if $\mathrm{rk}(A) = \mathrm{rk}(A^\circ)$. The latter is equivalent to saying that b is in the span of the columns of A, which is the image of A as a linear map. The statement on how to find a solution is then easily verified.

REMARK 9.33. Suppose $f\colon V \to W$ is a linear map of which you already know it is an isomorphism and you know its inverse $g = f^{-1}$ explicitly. Then for any $b \in W$, the solution to the linear equation $f(x) = b$ is of course just $x = g(b)$.

Similarly, 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 $x = A^{-1}b$.

EXAMPLE 9.34. 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

$$A' = \left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 0 \\ 1 & 2 & 3 & 4 & 2 \\ 1 & 3 & 5 & 7 & 4 \end{array}\right).$$

We transform it into reduced row echelon form.

$$\left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 0 \\ 1 & 2 & 3 & 4 & 2 \\ 1 & 3 & 5 & 7 & 4 \end{array}\right) \longrightarrow \left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 2 & 3 & 2 \\ 0 & 2 & 4 & 6 & 4 \end{array}\right) \longrightarrow \left(\begin{array}{cccc|c} 1 & 0 & -1 & -2 & -2 \\ 0 & 1 & 2 & 3 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

Since the last column does not contain the leading 1 of a row, the system is consistent, and a solution is given by $a = (x, y, z, w) = (-2, 2, 0, 0)$. The kernel of the non-extended matrix has basis (z_1, z_2) with $z_1 = (1, -2, 1, 0)$ and $z_2 = (2, -3, 0, 1)$. So all solutions are given by

$$(x, y, z, w) = a + rz_1 + sz_2 = (-2 + r + 2s, 2 - 2r - 3s, r, s),$$

for some r and s.

EXAMPLE 9.35. 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 by 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 $z = (2, 1, 1)$, 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

9.3.1. For each of the following systems of linear equations, 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.

$$\begin{cases} 2x_1+ & 3x_2+ & -2x_3 & = & 0 \\ 3x_1+ & 2x_2+ & 2x_3 & = & 0 \\ & -x_2+ & 2x_3 & = & 0 \end{cases}$$

$$\begin{cases} 2x_1+ & 3x_2+ & -2x_3 & = & 1 \\ 3x_1+ & 2x_2+ & 2x_3 & = & -1 \\ & -x_2+ & 2x_3 & = & -1 \end{cases}$$

$$\begin{cases} 2x_1+ & 3x_2+ & -2x_3 & = & 1 \\ 3x_1+ & 2x_2+ & 2x_3 & = & 1 \\ & -x_2+ & 2x_3 & = & 1 \end{cases}$$

$$\begin{cases} 3x_1+ & x_2+ & 2x_3+ & -2x_4 & = & 1 \\ 2x_1+ & -x_2+ & 2x_3 & & = & 2 \\ x_1+ & & x_3 & & = & 3 \\ -2x_1+ & -x_2+ & -x_3+ & x_4 & = & 4 \end{cases}$$

9.3.2. 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 10

Eigenvalues and Eigenvectors

In Example 9.27 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.

10.1. Eigenvalues and eigenvectors

Definition 10.1. Let $f\colon V \to V$ be an endomorphism of a vector space V. For any $\lambda \in F$, we say that λ is an *eigenvalue* of f if there exists a nonzero vector $v \in V$ with $f(v) = \lambda v$; we call such a vector an *eigenvector* for the eigenvalue λ, and the subspace $E_\lambda(f) = \{\, v \in V \;:\; f(v) = \lambda v\}$ is called the λ-*eigenspace* of f. The *spectrum* $\Omega(f)$ of f is the set of eigenvalues of f.

Note that $\lambda \in F$ is an eigenvalue of f if and only if $E_\lambda(f) \neq \{0\}$.

Example 10.2. 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 of V, and the matrix of f relative to that basis is

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

Example 10.3. 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 10.4. Let $s\colon \mathbb{R}^3 \to \mathbb{R}^3$ be the reflection in a plane $V \subset \mathbb{R}^3$. Then 1 is an eigenvalue with eigenspace $E_1(s) = V$, and -1 is an eigenvalue with eigenspace $E_{-1}(s) = V^\perp$.

If $\pi\colon \mathbb{R}^3 \to \mathbb{R}^3$ is the orthogonal projection onto V, then 1 is an eigenvalue with eigenspace $E_1(\pi) = V$, and 0 is an eigenvalue with eigenspace $E_0(\pi) = V^\perp$.

Since matrices can be identified with linear maps, it makes sense to speak about eigenvalues, eigenvectors, and eigenspaces of a square matrix $A \in \mathrm{Mat}(n, F)$.

PROPOSITION 10.5. *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*

$$f^k = \underbrace{f \circ f \circ \cdots \circ f}_{k}\colon V \to V.$$

PROOF. Exercise. □

PROPOSITION 10.6. *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 $(f - \lambda \cdot \mathrm{id}_V)(v) = 0$. □

So 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 10.7. *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 10.6 gives that λ is an eigenvalue of f if and only if $\ker(f - \lambda \cdot \mathrm{id}_V) \neq \{0\}$, so if and only if $f - \lambda \cdot \mathrm{id}_V$ is not injective, which is equivalent by Corollary 7.4 to the fact that $f - \lambda \cdot \mathrm{id}_V$ is not an isomorphism. By Proposition 9.29 this is the case if and only if $\det(f - \lambda \cdot \mathrm{id}_V) = 0$. □

Exercises

10.1.1. Prove Proposition 10.5.

10.1.2. 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 10.3 and Proposition 10.5.

10.2. 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 10.8. Let $f\colon V \to V$ be an endomorphism of a finite-dimensional vector space V over a field F. Then the characteristic polynomial $P_f \in F[t]$ of f is a polynomial over F (see Example 1.14) in the variable t, defined by

$$P_f(t) = \det(t \cdot \mathrm{id}_V - f).$$

The characteristic polynomial $P_A(t)$ of an $n \times n$ matrix A over F is defined by

$$P_A(t) = \det(t \cdot I_n - A).$$

PROPOSITION 10.9. *Suppose V is a finite-dimensional vector space with basis B and $f\colon V \to V$ an endomorphism. Set $A = [f]_B^B$. Then we have $P_f(t) = P_A(t)$.*

PROOF. Set $n = \dim V$. Then we have

$$P_f(t) = \det(t \cdot \mathrm{id}_V - f) = \det([t \cdot \mathrm{id}_V - f]_B^B) = \det(t \cdot I_n - [f]_B^B) = P_A(t).$$

□

Lemma 10.9 shows that the separate definitions of the characteristic polynomial for endomorphisms and matrices will not cause confusion. Moreover, it shows that if we want to compute the characteristic polynomial of an endomorphism, then we can use the associated matrix with respect to any basis. Applying this to f_A for any square matrix A yields the following result, which can also be proved directly.

PROPOSITION 10.10. *Let F be a field, n a nonnegative integer and $A, P \in \mathrm{Mat}(n, F)$ matrices with P invertible. Set $A' = P^{-1}AP$. Then the characteristic polynomials P_A and $P_{A'}$ are equal.*

PROOF. Exercise. □

If we write $A = (a_{ij})_{ij} \in \mathrm{Mat}(n, F)$, then

$$P_f(t) = \det(t \cdot I_n - A) = \begin{vmatrix} t - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & t - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & t - a_{nn} \end{vmatrix}.$$

Expanding the determinant, we find (exercise)

$$\begin{aligned} P_f(t) = P_A(t) &= t^n - \mathrm{Tr}(A)t^{n-1} + \cdots + (-1)^n \det(A) \\ &= t^n - \mathrm{Tr}(f)t^{n-1} + \cdots + (-1)^n \det(f)\,. \end{aligned}$$

PROPOSITION 10.11. *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 λ is a root of the characteristic polynomial P_f, i.e., $P_f(\lambda) = 0$.*

PROOF. Set $n = \dim V$. We have $P_f(\lambda) = \det(\lambda \cdot \mathrm{id}_V - f) = (-1)^n \cdot \det(f - \lambda \cdot \mathrm{id}_V)$, so $P_f(\lambda) = 0$ if and only if $\det(f - \lambda \cdot \mathrm{id}_V) = 0$. The statement therefore follows immediately from Proposition 10.7. □

EXAMPLE 10.12. Let us come back to the earlier example $f : (x, y) \mapsto (y, x)$ on $\mathbb{R}^2$ of Example 10.2. 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 10.13. 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

10.2.1. Let A be an $n \times n$ matrix. Show that we have

$$P_A(t) = t^n - \operatorname{Tr}(A)t^{n-1} + \cdots + (-1)^n \det(A),$$

i.e., the coefficients of t^{n-1} equals $-\operatorname{Tr}(A)$ and the constant coefficient equals $(-1)^n \det(A)$.

10.2.2. What is the characteristic polynomial of the reflection $s\colon \mathbb{R}^3 \to \mathbb{R}^3$ in some plane $V \subset \mathbb{R}^3$?

10.2.3. For each matrix A of the following real matrices, find a basis for the eigenspace $E_\lambda(A)$ of each eigenvalue λ.

$$\begin{pmatrix} 5 & -4 \\ 8 & -7 \end{pmatrix} \qquad \begin{pmatrix} -6 & -4 \\ 8 & 6 \end{pmatrix} \qquad \begin{pmatrix} 1 & 1 \\ -4 & 5 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 2 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -3 \end{pmatrix} \qquad \begin{pmatrix} 7 & 0 & 8 \\ 0 & 3 & 0 \\ -4 & 0 & -5 \end{pmatrix} \qquad \begin{pmatrix} 0 & -1 & 0 \\ 4 & 4 & 0 \\ 2 & 1 & 2 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 1 & 0 & 0 \\ -2 & 0 & 0 & 0 \\ -2 & -2 & 1 & 0 \\ -9 & -9 & 0 & -3 \end{pmatrix} \qquad \begin{pmatrix} 2 & -1 & 0 & 3 \\ 0 & 1 & 0 & 2 \\ -2 & 1 & 1 & -6 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

10.2.4. 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 9.2.2.

10.2.5. 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 $P_A = t^d + a_{d-1}t^{d-1} + \ldots + a_1 t + a_0$.

10.2.6. 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 4.3.4 and 9.2.3). Show that $P_f = P_{f'} \cdot P_{f''}$.

10.3. Diagonalization

DEFINITION 10.14. Let $f\colon V \to V$ 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 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 8.17). 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 10.15. *Let $f\colon V \to V$ be an endomorphism of a finite-dimensional vector space V with basis $B = (v_1, \dots, v_n)$. Then $[f]_B^B$ is a diagonal matrix with diagonal entries $\lambda_1, \dots, \lambda_n$ if and only if for all $1 \le j \le n$, 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. The matrix $[f]_B^B$ is diagonal with diagonal entries $\lambda_1, \dots, \lambda_n$ 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 $f(v_j) = \lambda_j v_j$, i.e., v_j is an eigenvector of f for eigenvalue λ_j. □

It follows that $f\colon V \to V$ 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 10.13, for instance, we only found 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 10.16. *Let V be an F-vector space and $f : V \to V$ an endomorphism. Let $\lambda_1, \dots, \lambda_m \in F$ be distinct, and for $i = 1, \dots, m$, 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 $\lambda_i \ne \lambda_{m+1}$, this implies $v_i = 0$ for $1 \le i \le m$. But then we must also have $v_{m+1} = 0$. □

ALTERNATIVE PROOF. 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 $\rho\colon F^m \to V$ that sends e_j to v_j. By Example 8.4 and Exercise 8.1.2, the Vandermonde matrix with columns $a_0, a_1, \ldots, a_{m-1}$ is invertible, so these columns span F^m. We conclude $\ker \rho = F^m$, so ρ is the zero map and $v_j = 0$ for all j. □

COROLLARY 10.17. *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 $1 \leq j \leq m$, 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 on 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 10.16, 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. □

EXAMPLE 10.18. We can use this to show once again that the power functions $f_n : x \mapsto x^n$ for $n \in \mathbb{N}_0$ 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.

COROLLARY 10.19. *Let V be a finite-dimensional F-vector space and $f : V \to V$ an endomorphism. Then we have*

$$\sum_{\lambda \in F} \dim E_\lambda(f) \leq \dim V$$

and equality holds if and only if f is diagonalizable.

PROOF. The inequality follows from Theorem 6.46 and Corollary 10.17. 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, which consists of eigenvectors of f. □

PROPOSITION 10.20. *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 $\lambda_1, \ldots, \lambda_n$. Therefore, $E_{\lambda_i}(f)$ is nontrivial for $1 \leq i \leq n$, which means that $\dim E_{\lambda_i}(f) \geq 1$. So

$$\dim V = n \leq \sum_{i=1}^{n} \dim E_{\lambda_i}(f) \leq \dim V\,,$$

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 $\dim V \geq 2$).

EXAMPLE 10.21. 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 by 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 $E_1(A) = \ker(A - I_3)$ has a basis (v_1, v_2) and $E_{-2}(A) = \ker(A + 2I_3)$ 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 10.17, 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 10.15. Furthermore, if we set $D = [f_A]_B^B$ and

$$P = [\mathrm{id}]_E^B = \begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix} = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix},$$

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 10.22. Let A be an $n \times n$ matrix over a field F. Assume that, analogously to Example 10.21, 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

$$AP = \begin{pmatrix} | & | & & | \\ \lambda_1 v_1 & \lambda_2 v_2 & \cdots & \lambda_n v_n \\ | & | & & | \end{pmatrix} = PD.$$

EXAMPLE 10.23. 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 10.5 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 10.24. 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 $\{b\}^\perp$, which is a proper subspace of $\mathbb{R}^n$, so there is no basis of eigenvectors and T is not diagonalizable.

Proposition 10.20 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 10.25. 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 and p is called *monic* if $a_d = 1$.

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.

THEOREM 10.26. Let $p = t^d + a_{d-1}t^{d-1} + \cdots + a_1 t + a_0 \in F[t]$ be a monic polynomial, and let $\alpha \in F$. If $p(\alpha) = 0$, then there is a polynomial $q = t^{d-1} + b_{d-2}t^{d-2} + \cdots + b_0$ such that $p = (t - \alpha)q$.

PROOF. If $\alpha = 0$, this is certainly true, since then $0 = p(0) = a_0$, and visibly $p = tq$ for $q = \sum_{i=1}^d a_i t^{i-1}$. In general, we replace t by $t + \alpha$. Then the polynomial $\tilde{p} = p(t + \alpha)$ is again monic of degree d, and $\tilde{p}(0) = p(\alpha) = 0$, so $\tilde{p} = t\tilde{q}$ for some monic polynomial $\tilde{q}$ of degree $d - 1$. Then

$$p = \tilde{p}(t - \alpha) = (t - \alpha)\tilde{q}(t - \alpha) = (t - \alpha)q\,,$$

where $q = \tilde{q}(t - \alpha)$ is monic of degree $d - 1$. □

COROLLARY 10.27. *Let $p = t^d + a_{d-1}t^{d-1} + \cdots + a_1 t + a_0 \in F[t]$ and $\alpha \in F$. Then there is a largest $m \in \mathbb{N}_0$ such that $p = (t - \alpha)^m q$ for some polynomial $q \in F[t]$; we then have $q(\alpha) \neq 0$.*

PROOF. Write $p = (t-\alpha)^m q$ with m as large as possible. (Note that $\deg(p) = m+\deg(q)$, so $m \le d$.) Then we must have $q(\alpha) \neq 0$, since otherwise we could write $q = (t-\alpha)r$ for some $r \in F[t]$, which would yield $p = (t-\alpha)^{m+1} r$, contradicting our choice of m. □

DEFINITION 10.28. 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 10.29. 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 10.30. *Let $f : V \to V$ be an endomorphism of an n-dimensional F-vector space V, 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 Thm. 10.26, if λ is a root of P_f, we can write $P_f = (t-\lambda)q$ with a monic polynomial q of degree $n-1$. Continuing in this way, we can write

$$P_f = (t-\lambda_1)^{m_1} \cdots (t-\lambda_k)^{m_k} q$$

with a monic polynomial q that does not have roots in F and distinct elements $\lambda_1, \dots, \lambda_k \in F$. 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 $\mu \in \{\lambda_1, \dots, \lambda_k\}$ (since $q(\mu) \neq 0$). Therefore the eigenvalues are exactly $\lambda_1, \dots, \lambda_k$, with algebraic multiplicities $m_1, \dots, m_k$, and

$$m_1 + m_2 + \cdots + m_k \le m_1 + m_2 + \cdots + m_k + \deg(q) = n\,.$$

We have equality if and only if $\deg(q) = 0$, i.e., if and only if $q = 1$; then

$$P_f = (t-\lambda_1)^{m_1} \cdots (t-\lambda_k)^{m_k}$$

is a product of linear factors. □

There is one further important relation between the multiplicities.

THEOREM 10.31. *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 $v_1, \dots, v_k, v_{k+1}, \dots, v_n$ of V such that $v_1, \dots, v_k$ 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

$$A = \begin{pmatrix} \lambda & 0 & \dots & 0 & * & \dots & * \\ 0 & \lambda & \dots & 0 & * & \dots & * \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \lambda & * & \dots & * \\ 0 & 0 & \dots & 0 & * & \dots & * \\ \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & * & \dots & * \end{pmatrix} = \left(\begin{array}{c|c} \lambda I_k & B \\ \hline 0 & C \end{array}\right).$$

We then have

$$\begin{aligned} P_f &= \det(t \cdot I_n - A) = \det\left(\begin{array}{c|c} (t-\lambda)\cdot I_k & -B \\ \hline 0 & t \cdot I_{n-k} - C \end{array}\right) \\ &= \det\big((t-\lambda)\cdot I_k\big) \cdot \det(t \cdot I_{n-k} - C) = (t-\lambda)^k \cdot P_C(t), \end{aligned}$$

by Exercise 9.1.5. We see that λ has multiplicity at least k as a root of P_f. □

COROLLARY 10.32. *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 10.19, the map f is diagonalizable if and only if the sum of the geometric multiplicities of all eigenvalues equals $n = \dim V$. By Theorem 10.31, 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 10.30, 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 10.33. 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.

EXAMPLE 10.34. 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. Lemma 10.30 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.

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). If D is a diagonal matrix with diagonal entries $\lambda_1, \ldots, \lambda_n$, and $k \geq 0$, 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 10.35. Take the matrix A as in Example 10.21. 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

$$\begin{aligned} A^k = PD^kP^{-1} &= \begin{pmatrix} 1 & 1 & 2 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & (-2)^k \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ -1 & 1 & 2 \\ 1 & -1 & -1 \end{pmatrix} \\ &= \begin{pmatrix} 2(-2)^k - 1 & (-2)^{k+1} + 2 & (-2)^{k+1} + 2 \\ 0 & 1 & 0 \\ (-2)^k - 1 & 1 - (-2)^k & 2 - (-2)^k \end{pmatrix}. \end{aligned}$$

Exercises

10.3.1. Show that for any integer k, and any invertible $n \times n$ matrices D, P, we have $(PDP^{-1})^k = PD^kP^{-1}$.

10.3.2. 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 10.35 for the k-th power of the matrix, where k is an arbitrary integer.

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 6 & -2 \\ 6 & -1 \end{pmatrix}, \qquad \begin{pmatrix} 3 & -1 & -1 \\ 4 & -2 & -4 \\ -2 & 2 & 4 \end{pmatrix}.$$

10.3.3. For each matrix A of the real matrices in Exercise 10.2.3, determine whether A is diagonalizable, and, if it is, determine a diagonal matrix D and an invertible matrix P, such that $A = PDP^{-1}$.

10.3.4. 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.

10.3.5. Determine M^k for the following matrices M and all integers k.

$$\begin{pmatrix} 7 & -10 \\ 5 & -8 \end{pmatrix} \qquad \begin{pmatrix} -2 & 3 & -7 \\ 0 & -4 & 6 \\ 0 & -3 & 5 \end{pmatrix}$$

10.3.6. Show that a polynomial of degree n over a field F has at most n roots in F.

10.3.7. Let F be an infinite field, i.e., $|F| = \infty$, and consider the map $\varphi\colon F[x] \to F^F$ of Exercise 2.3.8, cf. Exercises 4.1.6 and 6.4.3.

(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 $\dim P(F) = \infty$.

10.3.8. 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 $M = PDP^{-1}$.

$$\begin{pmatrix} 2 & 1 \\ -5 & -2 \end{pmatrix} \qquad \begin{pmatrix} 2 & -3 & -2 \\ 0 & 1 & 0 \\ 4 & -2 & -2 \end{pmatrix}.$$

10.3.9. The same as the previous exercise for

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix} \qquad \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

10.3.10. For which angle θ is the rotation $\mathbb{R}^2 \to \mathbb{R}^2$ about 0 over θ diagonalizable over $\mathbb{R}$?

10.3.11. Let M_n be as in Exercise 9.1.7 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

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 $f : X \to Y$ 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 an *inverse map* f^{-1} such that $f^{-1}(y) = x \iff f(x) = y$.

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 $f^{-1} \circ f = \mathrm{id}_X$.

When talking about several sets and maps between them, we often picture them in a *diagram* like the following.

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ {\scriptstyle g}\downarrow & & \downarrow{\scriptstyle g'} \\ U & \xrightarrow{f'} & V \end{array} \qquad\qquad \begin{array}{ccc} X & & \\ {\scriptstyle f}\downarrow & \searrow^{h} & \\ Y & \xrightarrow{g} & Z \end{array}$$

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 $h = g \circ f$.

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

$$\begin{aligned} +&: F \times F \to F, \quad (x, y) \mapsto x + y \quad (\text{'addition'}),\\ -&: F \times F \to F, \quad (x, y) \mapsto x - y \quad (\text{'subtraction'}),\\ \cdot&: F \times F \to F, \quad (x, y) \mapsto x \cdot y \quad (\text{'multiplication'}),\\ /&: F \times (F \setminus \{0\}) \to F, \quad (x, y) \mapsto x/y \quad (\text{'division'}), \end{aligned}$$

of which the addition and multiplication satisfy

$$\begin{array}{ccc} x + y = y + x, & x + (y + z) = (x + y) + z, & x + 0 = x,\\ x \cdot y = y \cdot x, & x \cdot (y \cdot z) = (x \cdot y) \cdot z, & x \cdot 1 = x,\\ & x \cdot (y + z) = (x \cdot y) + (x \cdot z) & \end{array}$$

for all $x, y, z \in F$, while the subtraction and division are related to the addition and multiplication through

$$x + y = z \Leftrightarrow x = z - y$$

for all $x, y, z \in F$ and

$$x \cdot y = z \Leftrightarrow x = z/y$$

for all $x, y, z \in F$ with $y \neq 0$.

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

$$\frac{x}{y} = \frac{a+b\sqrt{2}}{c+d\sqrt{2}} = \frac{(a+b\sqrt{2})\cdot(c-d\sqrt{2})}{(c+d\sqrt{2})\cdot(c-d\sqrt{2})} = \frac{(ac-2bd)+(bc-ad)\sqrt{2}}{c^2-2d^2}$$
$$= \frac{ac-2bd}{c^2-2d^2} + \frac{bc-ad}{c^2-2d^2}\sqrt{2} \in \mathbb{Q}(\sqrt{2}).$$

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) *Then $x+z = y+z$ 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 $0-1 = 1$. One easily checks that all requirements of Definition B.1 are satisfied.

WARNING B.7. Many properties of sums 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. We will see later that this extra structure can be 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 $\mathbb{C}$.

We will have occasion to use it later on. A proof, however, 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

$$\mathrm{Re}(z) = \tfrac{1}{2}(z+\bar{z}) \qquad \text{and} \qquad \mathrm{Im}(z) = \tfrac{1}{2i}(z-\bar{z}).$$

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