Linear Algebra I

October 23, 2011

Ronald van Luijk, 2011

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

Contents

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*. 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. 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 1.2. 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 1.3. 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 1.4. 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 1.2 on $\mathrm{Map}(X, \mathbb{R})$. We will do this in a yet more general situation in Example 1.22.

Example 1.5. A real *polynomial* in the variable x is a formal sum

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

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 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 $P(\mathbb{R})$ denote the set of all real polynomials. We define the addition of polynomials coefficientwise, so that the sum of the polynomials

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

equals

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

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

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

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 1.2, 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 1.4 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 the next section we introduce fields, which can function as sets of scalars.

1.2. **Fields.**

Definition 1.6. A field is a set F, together with two distinguished elements $0, 1 \in F$ with $0 \neq 1$ and four maps

$$\begin{aligned} +\colon F \times F \to F, &\quad (x, y) \mapsto x + y \quad (\text{'addition'}),\\ -\colon F \times F \to F, &\quad (x, y) \mapsto x - y \quad (\text{'subtraction'}),\\ \cdot\colon F \times F \to F, &\quad (x, y) \mapsto x \cdot y \quad (\text{'multiplication'}),\\ /\colon F \times (F \setminus \{0\}) \to F, &\quad (x, y) \mapsto x/y \quad (\text{'division'}), \end{aligned}$$

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 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 1.7. 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 1.8. 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 1.9. 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

$$(a + b\sqrt{2}) \pm (c + d\sqrt{2}) = (a \pm c) + (b \pm d)\sqrt{2},$$
$$(a + b\sqrt{2}) \cdot (c + d\sqrt{2}) = (ac + 2bd) + (ad + bc)\sqrt{2}.$$

To see that for any $x, y \in \mathbb{Q}(\sqrt{2})$ with $y \neq 0$ we also have $x/y \in \mathbb{Q}(\sqrt{2})$, we first note that if c and d are integers with $c^2 = 2d^2$, then $c = d = 0$, as otherwise c^2 would have an even and $2d^2$ an odd number of factors 2. Now for any $x, y \in \mathbb{Q}(\sqrt{2})$ with $y \neq 0$, we can write x/y as

$$\frac{a + b\sqrt{2}}{c + d\sqrt{2}}$$

with integers a, b, c, d, where c and d are not both 0; we find

$$\begin{aligned} \frac{x}{y} = \frac{a + b\sqrt{2}}{c + d\sqrt{2}} &= \frac{(a + b\sqrt{2}) \cdot (c - d\sqrt{2})}{(c + d\sqrt{2}) \cdot (c - d\sqrt{2})} = \frac{(ac - 2bd) + (bc - ad)\sqrt{2}}{c^2 - 2d^2} \\ &= \frac{ac - 2bd}{c^2 - 2d^2} + \frac{bc - ad}{c^2 - 2d^2}\sqrt{2} \in \mathbb{Q}(\sqrt{2}). \end{aligned}$$

We conclude that we also have division by nonzero elements on $\mathbb{Q}(\sqrt{2})$. Since the requirements of Definition 1.6 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 1.6 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 1.10. *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 1.11. The smallest field $\mathbb{F}_2 = \{0, 1\}$ has no more than the two required elements, with the only 'interesting' definition being that $1 + 1 = 0$. One easily checks that all requirements of Definition 1.6 are satisfied.

Warning 1.12. 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.

Exercise 1.2.1. Prove Proposition 1.10.

Exercise 1.2.2. Check that $\mathbb{F}_2$ is a field (see Example 1.11).

Exercise 1.2.3. Which of the following are fields?

(1) The set $\mathbb{N}$ together with the usual addition and multiplication.
(2) The set $\mathbb{Z}$ together with the usual addition and multiplication.
(3) The set $\mathbb{Q}$ together with the usual addition and multiplication.
(4) The set $\mathbb{R}_{\geq 0}$ together with the usual addition and multiplication.
(5) The set $\mathbb{Q}(\sqrt{3}) = \{a + b\sqrt{3} \ : \ a, b \in \mathbb{Q}\}$ together with the usual addition and multiplication.
(6) The set $\mathbb{F}_3 = \{0, 1, 2\}$ with the usual addition and multiplication, followed by taking the remainder after division by 3.

1.3. **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. (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 and multiplication.

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

$$(a + bi) + (c + di) = (a + c) + (b + d)i$$

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 hold. (The theory of rings and fields in later courses provides a rather elegant way of doing this.)

If $z = a + bi$ as above, then we call $\operatorname{Re} z = a$ the *real part* and $\operatorname{Im} z = b$ the *imaginary part* of z.

The least straightforward statement is probably the existence of multiplicative inverses. In this context, it is advantageous to introduce the notion of *conjugate complex number.*

Definition 1.13. If $z = a + bi \in \mathbb{C}$, then the *complex conjugate* of z is $\bar{z} = a - bi$. Note that $z\,\bar{z} = a^2 + b^2 \geq 0$. 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 1.14.

(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* $z \in \mathbb{C} \setminus \{0\}$, *we have* $z^{-1} = |z|^{-2} \cdot \bar{z}$.
(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
$$|z|^{-2}\bar{z} \cdot z = |z|^{-2} \cdot z\bar{z} = |z|^{-2}|z|^2 = 1\,.$$
(3) Exercise.

□

For example:
$$\frac{1}{1+2i} = \frac{1-2i}{(1+2i)(1-2i)} = \frac{1-2i}{1^2+2^2} = \frac{1-2i}{5} = \frac{1}{5} - \frac{2}{5}i\,.$$

Remark 1.15. 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 [**?**].

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.

Exercise 1.3.1. Prove Remark 1.14.

Exercise 1.3.2. For every complex number z we have $\mathrm{Re}(z) = \frac{1}{2}(z + \bar{z})$ and $\mathrm{Im}(z) = \frac{1}{2i}(z - \bar{z})$.

1.4. **Definition of a vector space.** We can now define the general notion of a vector space.

Definition 1.16. 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 1.17.

(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, they will all 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 1.18. 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 1.19. 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 1.20. 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

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

This proves teh 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 $x = (x_1, x_2, \dots, x_n)$. For each i with $1 \le i \le n$, we can take the negative $-x_i$ of x_i in the field F and set

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

Then, of course, we have

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

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 1.1. 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 1.21. A special case of Example 1.20 is when $F = \mathbb{R}$. The vector space $\mathbb{R}^n$ is called Euclidean n-space. In Sections 2.5 and **??** 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 1.22. This examples generalizes Example 1.4. 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 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 $X = \{1, 2, \dots, n\}$, then the set $F^X = \mathrm{Map}(X, F)$ of maps $f : \{1, 2, \dots, n\} \to F$ can be identified with F^n by letting such a map f correspond to the n-tuple $(f(1), f(2), \dots, f(n))$. 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 1.23. This example generalizes Example 1.5. A *polynomial* in the variable x over a field F is a formal sum

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

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 $f = \sum_{i=0}^{d} a_i x^i$ with $a_d \neq 0$ is d. By definition the degree of 0 equals $-\infty$. Let $P(F)$ denote the set of all polynomials over F. We define the addition and scalar multiplication of polynomials as in Example 1.5. Anybody who can prove that the previous examples are vector spaces, will have no problems showing that $P(F)$ is a vector space as well.

Warning 1.24. The polynomials x and x^2 in $P(\mathbb{F}_2)$ 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 $\alpha \in \mathbb{F}_2$.

Example 1.25. 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.

Exercise 1.4.1. Compute the inner product of the given vectors v and w in $\mathbb{R}^2$ and draw a corresponding picture (cf. Example 1.21).

(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) $v = w = (a, b)$.

Exercise 1.4.2. 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.

(1) L_1: $2x_1 + 3x_2 = 0$,
(2) L_2: $x_2 = 3x_1 - 1$,

(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) $L_7\colon 6x_1 - 2x_2 = 7$

Exercise 1.4.3. True or False? If true, explain why. If false, give a counterexample.

(1) If $a, b \in \mathbb{R}^2$ are nonzero vectors and $a \neq b$, then the lines in $\mathbb{R}^2$ given by $\langle a, x\rangle = 0$ and $\langle b, x\rangle = 1$ are not parallel.
(2) If $a, b \in \mathbb{R}^2$ are nonzero vectors and the lines in $\mathbb{R}^2$ given by $\langle a, x\rangle = 0$ and $\langle b, x\rangle = 1$ are parallel, then $a = b$.
(3) Two different hyperplanes in F^n may be given by the same equation.
(4) The intersection of two lines in F^n is either empty or consists of one point.
(5) For each vector $v \in \mathbb{R}^2$ we have $0 \cdot v = 0$. (What do the zeros in this statement refer to?)

Exercise 1.4.4. In Example 1.20, 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.

Exercise 1.4.5. In Example 1.22, 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.

Exercise 1.4.6. 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.

Exercise 1.4.7. Prove that the set $P(F)$ of polynomials over F, together with addition, scalar multiplication, and the zero as defined in Example 1.23 is a vector space.

Exercise 1.4.8. 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 1.25.
(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
$$\{(x, y, z) \in F^3 \ : \ x - z = 1\},$$
with coordinatewise addition and scalar multiplication.

Exercise 1.4.9. 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?

Exercise 1.4.10. We can generalize Example 1.22 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.

Exercise 1.4.11. Let S be the set of all sequences $(a_n)_{n\geq 0}$ of real numbers satisfying the recurrence relation
$$a_{n+2} = a_{n+1} + a_n \quad \text{for all } n \geq 0.$$
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.

Exercise 1.4.12. 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.

***Exercise 1.4.13.** For each of the eight axioms in Definition 1.16, try to find a system $(V, 0, +, \cdot)$ that does not satisfy that axiom, while it does satisfy the other seven.

1.5. **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 1.16. We will show that many such rules follow from the basic eight. The first is a cancellation rule.

Lemma 1.26. *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 1.27. 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 1.28. *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 1.26. □

Notation 1.29. 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 1.30. *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.

Exercise 1.5.1. Proof Proposition 1.27.

Exercise 1.5.2. Proof Remarks 1.30.

Exercise 1.5.3. Is the following statement correct? "Axiom (4) of Definition 1.16 is redundant because we already know by Remarks 1.30(2) that for each vector $x \in V$ the vector $-x = (-1) \cdot x$ is also contained in V."

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 $\dots$, $\boxed{\dots}$' and do not involve any existence statements, this is clear, since they hold (by assumption) even for all elements of V, so certainly for all elements of U. This covers all axioms but axiom (4). For axiom (4), we need that for all $u \in U$ there is an element $u' \in U$ with $u + u' = 0$. In the vector space V there is a unique such an element, namely $u' = -u = (-1)u$ (see Proposition 1.28, Notation 1.29, and Remarks 1.30). This element $u' = -u$ is contained in U by the third property of linear subspaces (take $\lambda = -1 \in F$). $\square$

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: $0 = (0, 0) \in U_a \iff 0 + 0 = a$, 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

$$\langle x, y\rangle = x_1y_1 + x_2y_2 + \ldots + x_ny_n.$$

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 2.5). 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

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

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

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

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

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

which proves the second equality of (2). □

Note that from properties (1) and (2) we also conclude that $\langle x+y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$. 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 vector space if and only if $b = 0$.

We can generalize 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 $px + qy + rz = d$ 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 given by $\langle a, v\rangle = b$ is a subspace if and only if $b = 0$.*

Proof. The proof is completely analogous to Example 2.12. See also Exercise 2.1.8. □

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.

Exercise 2.1.1. Given an integer $d \geq 0$, let $P_d(\mathbb{R})$ denote the set of polynomials of degree at most d. Show that the addition of two polynomials $f, g \in P_d(\mathbb{R})$ satisfies $f + g \in P_d(\mathbb{R})$. Show also that any scalar multiple of a polynomial $f \in P_d(\mathbb{R})$ is contained in $P_d(\mathbb{R})$. Prove that $P_d(\mathbb{R})$ is a vector space.

Exercise 2.1.2. 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?

Exercise 2.1.3. 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 $\mathbb{R}^X$?

Exercise 2.1.4. 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) $U_3 = \{(x, y) \in \mathbb{R}^2 : xy = 0\}$.

Exercise 2.1.5. 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) $U_6 = \{f \in V \ : \ f(0) \geq 0\}$

Exercise 2.1.6. Prove Proposition 2.10.

Exercise 2.1.7. Prove Proposition 2.18.

Exercise 2.1.8. 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 $b_1 = \ldots = b_t = 0$.

Exercise 2.1.9.

(1) Let X be a set and F a field. 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 X be a set, F a field, and V 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 1.4.10).

2.2. **Intersections.** The following result now tells us 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, $U_1, U_2 \subset V$ 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.

Exercise 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 $U_2 \subset U_1$.

Exercise 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.

Exercise 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.** The property we proved in Lemma 2.19 is very important, since it will tell us that there is always a *smallest* linear subspace of V that contains a given subset S of V. This means that there is a linear subspace U of V such that $S \subset U$ and such that U is contained in every other linear subspace of V that contains S.

Definition 2.21. Let V be a vector space, $S \subset V$ a subset. The *linear hull* or *linear span* of S, or the linear subspace *generated by* S is

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

(This notation means the intersection of all elements of the specified set: we intersect all linear subspaces containing S. Note that V itself is such a subspace, so this set of subspaces is non-empty, so by the preceding result, $L(S)$ really *is* a linear subspace.)

If we want to indicate the field F of scalars, we write $L_F(S)$. If $v_1, v_2, \dots, v_n \in V$, we also write $L(v_1, v_2, \dots, v_n)$ for $L(\{v_1, v_2, \dots, v_n\})$.

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 Span(S) or $\langle S \rangle$ (which in LaTeX is written `$\langle S \rangle$` and *not* `$<S>$`!).

Example 2.22. What do we get in the extreme case that $S = \emptyset$? Well, then we have to intersect *all* linear subspaces of V, so we get $L(\emptyset) = \{0\}$.

Lemma 2.23. *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. By definition, U is one of the subspaces that $L(S)$ is the intersection of. The claim follows immediately. □

Definition 2.21 above has some advantages and disadvantages. Its main advantage is that it is very elegant. Its main disadvantage is that it is rather abstract and non-constructive. To remedy this, we show that in general we can build the linear hull in a constructive way "from below" instead of abstractly "from above." This generalizes the idea of Example 2.31.

Example 2.24. Let us look at another specific case first. Given a vector space V over a field F, and vectors $v_1, v_2 \in V$, how can we describe $L(v_1, v_2)$?

According to the definition of linear subspaces, we must be able to add and multiply by scalars in $L(v_1, v_2)$; also $v_1, v_2 \in L(v_1, v_2)$. This implies that every element of the form $\lambda_1 v_1 + \lambda_2 v_2$ must be in $L(v_1, v_2)$. So set

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

(where F is the field of scalars); then $U \subset L(v_1, v_2)$. On the other hand, U is itself a linear subspace:

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

(Exercise: which of the vector space axioms have we used where?)

Therefore, U is a linear subspace containing v_1 and v_2, and hence $L(v_1, v_2) \subset U$ by Remark 2.23. We conclude that

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

This observation generalizes.

Definition 2.25. 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.

Proposition 2.26. *Let V be a vector space, $v_1, v_2, \ldots, v_n \in V$. Then the set of all linear combinations of $v_1, v_2, \ldots, v_n$ is a linear subspace of V; it equals the linear hull $L(v_1, v_2, \ldots, v_n)$.*

More generally, let $S \subset V$ be a subset. Then the set of all linear combinations on S is a linear subspace of V, equal to $L(S)$.

Proof. Let U be the set of all linear combinations of $v_1, v_2, \ldots, v_n$. We have to check that U is a linear subspace of V. 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. We have $v_1, v_2, \ldots, v_n \in U$, since

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

so $L(v_1, v_2, \ldots, v_n) \subset U$ by Remark 2.23. On the other hand, it is clear that any linear subspace containing $v_1, v_2, \ldots, v_n$ has to contain all linear combinations of these vectors. Hence U is contained in all the subspaces that $L(v_1, v_2, \ldots, v_n)$ is the intersection of, so $U \subset L(v_1, v_2, \ldots, v_n)$. Therefore

$$L(v_1, v_2, \ldots, v_n) = U.$$

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. □

Remark 2.27. In many books the linear hull $L(S)$ of a subset $S \subset V$ is in fact *defined* to be the set of all linear combinations on S. Proposition 2.3 states that our definition is equivalent, so from now on we can use both.

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 $x = x_1e_1 + x_2e_2 + x_3e_3$, so every element in $\mathbb{R}^3$ is a linear combination of e_1, e_2, e_3. We conclude $\mathbb{R}^3 \subset L(e_1, e_2, e_3)$ and therefore $L(e_1, e_2, e_3) = \mathbb{R}^3$, so $\{e_1, e_2, e_3\}$ generates $\mathbb{R}^3$.

Example 2.30. Let F be a field and n a positive integer. Set

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

with e_i the vector in F^n whose i-th entry equals 1 while all other entries equal 0. Then for every vector $x = (x_1, x_2, \ldots, x_n) \in F^n$ we have $x = x_1e_1+x_2e_2+\ldots+x_ne_n$, so as in the previous example we find that $\{e_1, e_2, \ldots, e_n\}$ generates F^n. These generators are called the *standard generators* of F^n.

Example 2.31. Take $V = \mathbb{R}^4$ and consider $S = \{v_1, v_2, v_3\}$ with

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

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

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

are subspaces (see Proposition 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 **??** we will see how to compute generators of intersections more systematically.

Example 2.32. Let us consider again the vector space $\mathcal{C}(\mathbb{R})$ of continuous functions on $\mathbb{R}$. The power functions $f_n : x \mapsto x^n$ $(n = 0, 1, 2, \ldots)$ 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

$$x \longmapsto a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$$

with $n \in \mathbb{N}_0$ and $a_0, a_1, \ldots, a_n \in \mathbb{R}$.

Example 2.33. 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 $a_0, a_1, \ldots, a_n \in F$. By definition, the power functions f_n generate the subspace of polynomial functions.

Warning 2.34. In Example 1.5 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* $f : \mathbb{R} \to \mathbb{R}$, though the difference is subtle.
Over a general field, the subspace of polynomial functions is generated by the power functions f_n from Example 2.33, while the space $P(F)$ of polynonomials is generated by the formal powers x^i of a variable x.
As stated in Warning 1.24, though, over some other fields the difference between polynomials, as defined in Example 1.23, and polynomial functions, as defined in Example 2.33, 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 2.35. 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 2.55 we will clarify the notation $S^\perp$.

Example 2.36. Let F be a field. Then for every element $a \in F^n$, the hyperplane H_a given by $\langle a, x\rangle = 0$ is $\{a\}^\perp$. 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 2.31 can also be written as $\{a_1, a_2\}^\perp$.

Proposition 2.37. *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 $t = \lambda_1 s_1 + \ldots + \lambda_n s_n$, which implies

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

□

Remark 2.38. Later we will see that the inclusion $L(S) \subset (S^\perp)^\perp$ of Proposition 2.37 is in fact an equality, so that for every subspace U we have $(U^\perp)^\perp = U$. A computational proof is based on Proposition 5.30, cf. Exercise 5.3.6. A more conceptual proof uses the notion of *dimension* and is given in Section 7.1.

Exercises.

Exercise 2.3.1. Prove Proposition 2.37.

Exercise 2.3.2. Do the vectors

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

generate $\mathbb{R}^3$?

Exercise 2.3.3. Do the vectors

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

generate $\mathbb{R}^3$?

Exercise 2.3.4. Let $U \subset \mathbb{R}^4$ be the subspaces 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.

Exercise 2.3.5. Let F be a field and X a set. Consider the subspace $F^{(X)}$ of F^X consisting of all functions $f : X \to F$ that satisfy $f(x) = 0$ for all but finitely many $x \in X$ (cf. Exercise 2.1.9). For every $x \in X$ we define the function $e_x : X \to F$ 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 $F^{(X)}$.

Exercise 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?

Exercise 2.3.7. We say that a function $f : \mathbb{R} \to \mathbb{R}$ 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 al even functions a linear subspace?
(2) Is the subset of $\mathbb{R}^\mathbb{R}$ consisting of al odd functions a linear subspace?

Exercise 2.3.8. 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 Remark 2.23, give short proofs of the following equalities of subspaces.

(1) $W = L(v'_1, \dots, v'_n)$ 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) $W = L(v'_1, \dots, v'_n)$ 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).
(3) $W = L(v'_1, \dots, v'_n)$ 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).

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, the following definition is natural.

Definition 2.39. 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

$$\sum_{i \in I} U_i = L\Big(\bigcup_{i \in I} U_i\Big)\,.$$

As before in our discussion of linear hulls, we want a more explicit description of these sums.

Lemma 2.40. *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 Remark 2.23 (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 $v = \lambda_1 w_1 + \ldots + \lambda_s w_s$ 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 $u_1 = \lambda_1 w_1 + \ldots + \lambda_r w_r \in U_1$ and $u_2 = \lambda_{r+1} w_{r+1} + \ldots + \lambda_s w_s \in U_2$ we have $v = u_1 + u_2$, as required.

Suppose we have $v = \lambda_1 w_1 + \ldots + \lambda_s w_s$ 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 2.41. 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 2.42. 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 2.43. *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 2.44. 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 2.45. 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 2.46. *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 Section 6 we will see that this is indeed the case. 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 2.61).

Exercises.

Exercise 2.4.1. Prove Lemma 2.43.

Exercise 2.4.2. Suppose F is a field and $U_1, U_2 \subset F^n$ subspaces. Show that we have

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

Exercise 2.4.3. Suppose V is a vector space with a subspace $U \subset V$. Suppose that $U_1, U_2 \subset V$ subspaces of V that are contained in U. Show that the sum $U_1 + U_2$ is also contained in U.

Exercise 2.4.4. 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.

Exercise 2.4.5. 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.

Exercise 2.4.6. Are the subspaces U_0 and U_1 of Example 2.42 complementary subspaces?

Exercise 2.4.7. 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.

2.5. **Euclidean space.** This section, with the exception of Proposition 2.60, 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. 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 1.21. 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 2.48, as well as Proposition 2.51.

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 2.47. *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 $\langle x, x\rangle = x_1^2 + x_2^2 + \ldots + x_n^2$. 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 2.48. 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 2.47, 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 2.49. The length of the vector $(1, -2, 2, 3)$ in $\mathbb{R}^4$ equals $\sqrt{1 + 4 + 4 + 9} = 3\sqrt{2}$.

Lemma 2.50. *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 2.51. *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 2.52. 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 2.51, 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 2.51, we define orthogonality for vectors in $\mathbb{R}^n$ for all $n \geq 0$.

Definition 2.53. 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 2.54. Proposition 2.47 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 2.55. 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 2.35) 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 2.56. 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 2.57. *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 2.57, 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 the elements of H are orthogonal to a, but the differences between elements. Draw a picture to clarify this for yourself!

Example 2.58. 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 2.48 and 2.53 we immediately have the following analogon of Pythagoras' Theorem.

Proposition 2.59. *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. □

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 see that this is the case in Proposition 2.69. First we state some auxiliary results.

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

Proposition 2.60. *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 2.61. *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 2.60 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 2.46, the spaces $L(a)$ and H_a are complementary subspaces. Alternatively, we first only conclude $L(a) + H_a = F^n$ from Proposition 2.60. 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 2.62. *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 $\lambda = \langle a, v\rangle \cdot \|a\|^{-2}$ and the lengths of v_1 and v_2 are given by*

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

Proof. The statement is just a reformulation of Proposition 2.60 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 2.47. □

Definition 2.63. Using the same notation as in Corollary 2.62, 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 $H = \{a\}^\perp = L^\perp$. We define the *distance* $d(v, L)$ from v to L by $d(v, L) = \|v_2\|$ and the distance $d(v, H)$ from v to H by $d(v, H) = \|v_1\|$. In section **??** we will define the orthogonal projection onto (and distances to) any subspace of $\mathbb{R}^n$.

Remark 2.64. Suppose F is a field contained in $\mathbb{R}$ and a is a nonzero element in F^n. Set $L = L(a)$ and $H = \{a\}^\perp = L^\perp$ as in Definition 2.63. Let $v_1 \in L$ and $v_2 \in H$ be the orthogonal projections of v on L and H respectively, so that $v = v_1 + v_2$. Then for any $x \in L$, we can write $v - x$ as the sum $(v_1 - x) + v_2$ of two orthogonal vectors, so that by Proposition 2.59 (Pythagoras) we have

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

We conclude $\|v - x\| \geq \|v_2\| = d(v, L)$, so the distance $d(v, L)$ is the minimal distance from v to any point on L. Similarly, the distance $d(v, H)$ is the minimal distance from v to any point on H. Make a picture to support these arguments!

Example 2.65. 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 2.66. 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 Definition 2.63, 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 2.67. 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 noral 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 Definition 2.63, 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 2.68. 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 2.62. 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 $\frac{1}{3}\sqrt{42}$.

Proposition 2.69 (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 2.63, so our vectors v, w, z correspond to a, v, v_2 of Proposition 2.62, respectively). Then by Corollary 2.62 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 2.70 (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 2.69, 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 2.71. 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 2.69, 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 2.51. Finally, Definitions 2.53 and 2.71 imply that two nonzero vectors v and w in F^n are orthogonal if and only the angle between them is $\pi/2$.

Example 2.72. 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 2.73. 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.

Exercise 2.5.1. Prove Lemma 2.50.

Exercise 2.5.2. 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 $x'' \in V$.

Exercise 2.5.3. Finish the proof of Proposition 2.70.

Exercise 2.5.4. Explain why Proposition 2.70 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 $\|v - w\| \leq \|v\| + \|w\|$.

Exercise 2.5.5. Prove the cosine rule in $\mathbb{R}^n$.

Exercise 2.5.6. Show that two vectors $v, w \in \mathbb{R}^n$ have the same length if and only $v - w$ and $v + w$ are orthogonal.

Exercise 2.5.7. Prove that the diagonals of a parallelogram are orthogonal to each other if and only if all sides have the same length.

Exercise 2.5.8. Compute the distance from the point $(1, 1, 1, 1) \in \mathbb{R}^4$ to the line $L(a)$ with $a = (1, 2, 3, 4)$.

Exercise 2.5.9. 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 $v = (2, 1, 3)$.

Exercise 2.5.10. 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.

Exercise 2.5.11. Determine the angle between the vectors $(1, -1, 2)$ and $(-2, 1, 1)$ in $\mathbb{R}^3$.

Exercise 2.5.12. 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.

3. 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. But before we give a definition, we have to review what a map or function is and their basic properties.

3.1. **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.

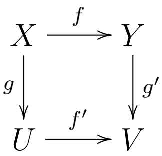
If $f : X \to Y$ is a map, then we call $\{f(x) : x \in X\} \subset Y$ the *image* of f, $\operatorname{im}(f)$. The map f is called *injective* or *one-to-one (1–1)* if no two elements of X are mapped to the same element of Y. More formally, if $x, x' \in X$ and $f(x) = f(x')$, then $x = x'$. The map f is called *surjective* or *onto* 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 $f^{-1}(y) = x \iff f(x) = y$.

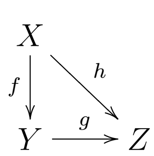
A map $f : X \to Y$ induces maps from subsets of X to subsets of Y and conversely, which are denoted by f and f^{-1} again (so you have to be careful to check the 'datatype' of the argument). Namely, if $A \subset X$, we set $f(A) = \{f(x) : x \in A\}$ (for example, the image of f is then $f(X)$), and 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.





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 $h = g \circ f$.

3.2. **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 3.1. 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.)

An injective homomorphism is called a *monomorphism*, a surjective homomorphism is called an *epimorphism*, and 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.

Lemma 3.2. *Here are some simple properties of linear maps.*

(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 3.3. *Let $f : V \to W$ be a linear map of F-vector spaces.*

(1) *For all $v_1, v_2, \ldots, v_n \in V$ and $\lambda_1, \lambda_2, \ldots, \lambda_n \in F$ we have*

$$f(\lambda_1 v_1 + \ldots + \lambda_n v_n) = \lambda_1 f(v_1) + \ldots + \lambda_n f(v_n).$$

(2) *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 3.4. 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 3.5. *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) $\operatorname{im}(f) \subset W$ *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* $\operatorname{im}(f)$.

(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 $f(0) = 0 \in \operatorname{im}(f)$. If $w_1, w_2 \in \operatorname{im}(f)$, then there are $v_1, v_2 \in V$ such that $f(v_1) = w_1$, $f(v_2) = w_2$, hence $w_1 + w_2 = f(v_1 + v_2) \in \operatorname{im}(f)$. If λ is a scalar and $w \in \operatorname{im}(f)$, then there is $v \in V$ such that $f(v) = w$, hence $\lambda w = f(\lambda v) \in \operatorname{im}(f)$.

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 3.6. 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 3.7.

(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. 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) 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 1.4.12.

(5) 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 (since $D \circ I_a = \mathrm{id}_P$, cf. (d) 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 $I_{a,b} \circ D = \mathrm{ev}_b - \mathrm{ev}_a$ and $I_a \circ D = \mathrm{id}_P - \mathrm{ev}_a$. 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!).

Proposition 3.8. *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 3.9. *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 $1 \leq i \leq n$ we have

$$f(e_i) = \langle a, e_i\rangle = a_1 \cdot 0 + \ldots + a_{i-1} \cdot 0 + a_i \cdot 1 + a_{i+1} \cdot 0 + \ldots + a_n \cdot 0 = a_i.$$

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

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

□

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

Lemma 3.10. *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* $\mathrm{Hom}(V, W)$.

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

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

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

$$(\mu f)(v_1 + v_2) = \mu f(v_1 + v_2) = \mu\big(f(v_1) + f(v_2)\big)$$
$$= \mu f(v_1) + \mu f(v_2) = (\mu f)(v_1) + (\mu f)(v_2)\,.$$

Finally, let $\lambda \in F$ and $v \in V$. Then

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

□

Proposition 3.11. *Let F be a field and n a nonnegative integer. Let W be an F-vector space containing n (not necessarily different) vectors $w_1, w_2, \ldots, w_n$. Then there is a unique linear map $f\colon F^n \to W$ with $f(e_i) = w_i$ 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

$$f(x) = f(x_1 e_1 + \ldots + x_n e_n) = x_1 f(e_1) + \ldots + x_1 f(e_n) = x_1 w_1 + \ldots + x_n w_n,$$

so f is completely determined on all $x \in F^n$ by the vectors $w_1, w_2, \ldots, w_n$ and therefore f is unique, if it exists.

To show there is indeed an f as claimed, we define a function $f\colon F^n \to W$ by

$$f(x) = x_1 w_1 + \ldots + x_n w_n$$

(we have no choice by the above). One easily checks that f is linear. (Do this!) For i with $1 \le i \le n$, we have

$$f(e_i) = 0 \cdot w_1 + \ldots + 0 \cdot w_{i-1} + 1 \cdot w_i + 0 \cdot w_{i+1} + \ldots + 0 \cdot w_n = w_i,$$

so f indeed satisfies the requirements. □

Exercises.

Exercise 3.2.1. Prove Lemma 3.3.

Exercise 3.2.2. 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) $P(\mathbb{C}) \to P(\mathbb{C})$, $f \mapsto f'$, where $P(\mathbb{C})$ is the vector space of polynomials over $\mathbb{C}$ and f' teh derivative of f,
(6) $P(\mathbb{R}) \to \mathbb{R}^2$, $f \mapsto (f(2), f'(0))$.

Exercise 3.2.3. 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 $L(f(S)) = W$.

Exercise 3.2.4. 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
$$\rho((x,y)) = (x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta).$$

Exercise 3.2.5. 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 formule for s.

Exercise 3.2.6. 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 $2y - x = 0$.

Exercise 3.2.7. Give an explicit expression for the linear map $s\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by reflecting in the line $y = 3x$.

Exercise 3.2.8. Let F be a field and n a nonnegative integer. Show that there is an isomorphism
$$F^n \to \mathrm{Hom}(F^n, F)$$
that sends a vector $a \in F^n$ to te linear map $x \mapsto \langle a, x\rangle$.

Exercise 3.2.9. Suppose V is a vector space with two complementary subspaces U and U', cf. Definition 2.44. Then for every $v \in V$ there are unique elements $u \in U$ and $u' \in U'$ with $v = u + u'$ by Lemma 2.46; 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 $\ker \pi_U = U'$. We call π_U the projection of V onto U along U'.

4. Matrices

4.1. Definition of matrices.

Definition 4.1. Let F be a field and m, n nonnegative integers. An *$m \times n$ matrix over F* is an array
$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} = (a_{ij})_{1\le i\le m, 1\le j\le n}$$
of *entries* or *coefficients* $a_{ij} \in F$.

For $i \in \{1, \ldots, m\}$, the vector $(a_{i1}, a_{i2}, \ldots, a_{in})$ is a *row* of A, and for $j \in \{1, \ldots, n\}$, the vector
$$(a_{1j}, a_{2j}, \ldots, a_{mj})^\top := \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}$$
is called a *column* of A. The set of all $m \times n$ matrices with entries in F is denoted by $\mathrm{Mat}(m \times n, F)$ or $\mathrm{Mat}_{m,n}(F)$. 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 $\mathrm{Mat}(n, F)$ or $\mathrm{Mat}_n(F)$ 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

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

we define the product Ax as

$$Ax = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix}.$$

Example 4.2. We have

$$\begin{pmatrix} 3 & 2 & 1 \\ -1 & 2 & 7 \\ -3 & 5 & -2 \end{pmatrix} \begin{pmatrix} 2 \\ -2 \\ -1 \end{pmatrix} = \begin{pmatrix} 3 \cdot 2 + 2 \cdot (-2) + 1 \cdot (-1) \\ (-1) \cdot 2 + 2 \cdot (-2) + 7 \cdot (-1) \\ (-3) \cdot 2 + 5 \cdot (-2) + (-2) \cdot (-1) \end{pmatrix} = \begin{pmatrix} 1 \\ -13 \\ -14 \end{pmatrix}.$$

There are (at least) two useful ways to think of the multiplication. If we let

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

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 $Ae_j = w_j$.

4.2. **Linear maps associated to matrices.**

Definition 4.3. To any matrix $A \in \mathrm{Mat}_{m,n}(F)$ we associate the function $f_A\colon F^n \to F^m$ given by
$$f_A(x) = Ax$$
for all $x \in F^n$.

Remark 4.4. Note that $f_A(e_j)$ equals the j-th column of A for any $j \in \{1, \dots, n\}$.

Example 4.5. Let $A \in \mathrm{Mat}_{3,4}(\mathbb{R})$ be the matrix
$$\begin{pmatrix} 3 & 2 & 0 & -1 \\ 1 & -2 & 5 & -3 \\ 0 & 1 & 4 & 7 \end{pmatrix}.$$

Then the linear map f_A sends
$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \in \mathbb{R}^4 \qquad \text{to} \qquad \begin{pmatrix} 3x_1 & +2x_2 & & -x_4 \\ x_1 & -2x_2 & +5x_3 & -3x_4 \\ & x_2 & +4x_3 & +7x_4 \end{pmatrix} \in \mathbb{R}^3.$$

Indeed, one checks that the images of the standard vectors e_1, e_2, e_3, and e_4 are the columns of A.

Lemma 4.6. *For any matrix $A \in \mathrm{Mat}_{m,n}(F)$, the associated function $f_A\colon F^n \to F^m$ is a linear map.*

Proof. This can be checked straight from the definition, but it is easier to use the two ways to think of the product Ax just described. We will use the first way. Let $v_1, v_2, \dots, v_m$ denote the row vectors of A. Then for any $x, y \in F^n$ we have
$$f_A(x+y) = A(x+y) = \begin{pmatrix} \langle v_1, x+y\rangle \\ \langle v_2, x+y\rangle \\ \vdots \\ \langle v_m, x+y\rangle \end{pmatrix} = \begin{pmatrix} \langle v_1, x\rangle + \langle v_1, y\rangle \\ \langle v_2, x\rangle + \langle v_2, y\rangle \\ \vdots \\ \langle v_m, x\rangle + \langle v_m, y\rangle \end{pmatrix}$$
$$= \begin{pmatrix} \langle v_1, x\rangle \\ \langle v_2, x\rangle \\ \vdots \\ \langle v_m, x\rangle \end{pmatrix} + \begin{pmatrix} \langle v_1, y\rangle \\ \langle v_2, y\rangle \\ \vdots \\ \langle v_m, y\rangle \end{pmatrix} = Ax + Ay = f_A(x) + f_A(y).$$

Similarly, one easily checks that for any $\lambda \in F$ we have $f_A(\lambda x) = \lambda f_A(x)$, so f_A is indeed linear. □

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

Proposition 4.7. *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 $A \in \mathrm{Mat}_{m,n}(F)$ with $f = f_A$.*

Proof. We use the first view point. For any i with $1 \le i \le m$, the composition of f with the projection $\pi_i\colon F^m \to F$ (see Examples 3.7) is the linear map $\pi_i \circ f\colon F^n \to$

F that sends any $x \in F^n$ to the i-th entry of $f(x)$. By Lemma 3.9 there is a unique vector $v_i \in F^n$ such that $(\pi_i \circ f) = \langle v_i, x\rangle$ for all $x \in F^n$. Then we have

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

so $f = f_A$ for the matrix A whose rows are $v_1, v_2, \ldots, v_m$. The uniqueness of A follows from the uniqueness of v_i for all i. □

Alternative proof. We now use the second view point. Suppose $A \in \mathrm{Mat}_{m,n}(F)$ satisfies $f = f_A$. Then by Remark 4.4 the j-th column of A equals $f_A(e_j) = f(e_j)$, so A is completely determined by f and therefore, A is unique, if it exists.

To show there is indeed an A as claimed, we set $w_j = f(e_j)$ for $1 \le j \le n$ and let A be the matrix whose columns are $w_1, w_2, \ldots, w_n$ (we have no choice by the above). Then for any $x = (x_1, \ldots, x_n)$ we have

$$\begin{aligned} f(x) &= f(x_1e_1 + \ldots + x_ne_n) = x_1f(e_1) + \ldots + x_nf(e_n) \\ &= x_1w_1 + \ldots + x_nw_n = Ax = f_A(x), \end{aligned}$$

which implies $f = f_A$. □

Because of Proposition 4.7, 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 4.8. Let $\rho\colon \mathbb{R}^2 \to \mathbb{R}^2$ be rotation about the origin $(0,0)$ over an angle θ. From Exercise 3.2.4, 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 4.9. 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 Remark 4.4, 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

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

We know that $\mathrm{Hom}(F^n, F^m)$ has the structure of an F-vector space (see Lemma 3.10). We can 'transport' this structure to $\mathrm{Mat}(m \times n, F)$ using the identification of matrices and linear maps.

Definition 4.10. For $A, B \in \text{Mat}(m \times n, F)$, 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, $\text{Mat}(m \times n, F)$ 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 a rectangular fashion instead of in a row or column.

By Lemma 3.2, the composition of two linear maps is again linear. How is this reflected in terms of matrices?

Definition 4.11. Let $A \in \text{Mat}(l \times m, F)$ and $B \in \text{Mat}(m \times n, F)$. 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 $\text{Mat}(l \times n, F)$.

In other words, the product AB satisfies $f_{AB} = f_A \circ f_B$, so we have

$$(AB)x = f_{AB}(x) = f_A(f_B(x)) = A(Bx) \tag{3}$$

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 (3) 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

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{l1} & a_{l2} & \cdots & a_{lm} \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix}$$

we get

$$v_i = (a_{i1}, a_{i2}, \ldots, a_{im}) \quad \text{and} \quad w_k = \begin{pmatrix} b_{1k} \\ b_{2k} \\ \vdots \\ b_{mk} \end{pmatrix},$$

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.

(4)
$$\begin{array}{cc} & \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix} = B \\ A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{l1} & a_{l2} & \cdots & a_{lm} \end{pmatrix} & \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{l1} & c_{l2} & \cdots & c_{ln} \end{pmatrix} = AB \end{array}$$

Example 4.12. 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.

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

The product AB is a matrix with as many rows as A and as many columns as B, so it is a 2×3 matrix. The $(1,2)$-th entry of AB, for instance, is the 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 4.13. *The matrix multiplication is associative: for* $A \in \mathrm{Mat}(k \times l, F)$*,* $B \in \mathrm{Mat}(l \times m, F)$*,* $C \in \mathrm{Mat}(m \times n, F)$*, 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 $(f_A \circ f_B) \circ f_C$, and 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 4.7. $\square$

Proposition 4.14. *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. □

However, matrix multiplication is *not* commutative in general — BA need not even be defined even though AB is — and $AB = 0$ (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 4.15. 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 $f_{A^{-1}} = f_A^{-1}$; we then have $AA^{-1} = A^{-1}A = I_n$, and A^{-1} is uniquely determined by this property.

Proposition 4.16. *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 4.17. If $A \in \mathrm{Mat}(m \times n, F)$ is invertible, then $m = n$, as we will see computationally in the next chapter and more insightfully in the chapter after. This means that the matrices A and B in Proposition 4.16 are in fact square matrices of the same size.

Remark 4.18. The identity matrix acts as a multiplicative identity:

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

Definition 4.19. Let $A = (a_{ij}) \in \mathrm{Mat}(m \times n, F)$ be a matrix. The *transpose* of A is the matrix

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

(So we get $A^\top$ from A by a 'reflection on the main diagonal.')

Example 4.20. For

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

we have

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

As simple properties of transposition, we have that

$$(A+B)^\top = A^\top + B^\top, \quad (\lambda A)^\top = \lambda A^\top, \quad (AB)^\top = B^\top A^\top$$

(note the reversal of factors!) — this is an exercise. If A is invertible, this implies that $A^\top$ is also invertible, and $(A^\top)^{-1} = (A^{-1})^\top$.

Remark 4.21. 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 $\mathrm{Mat}_{1,n}(F)$ 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 4.22. The product Ax of a matrix $A \in \mathrm{Mat}_{m,n}(F)$ 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 $A \cdot x^\top$. 4.1.

Definition 4.23. The row space $R(A)$ of an $m \times n$ matrix $A \in \mathrm{Mat}_{m,n}(F)$ 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 4.24. The column space of a matrix $A \in \mathrm{Mat}_{m,n}(F)$ is the same as the image of A, i.e., the image of the linear map f_A.

Proposition 4.25. *Let $A \in \mathrm{Mat}_{m,n}(F)$ 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 $v_1, v_2, \dots, v_m$ be the rows of A. Then $R(A) = L(v_1, \dots, v_m)$. The map $f_A \colon F^n \to F^m$ is then given by $f_A(x) = (\langle v_1, x\rangle, \dots, \langle v_m, x \rangle)$ for all $x \in F^n$. 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

$$\{v_1, \dots, v_m\}^\perp = L(v_1, \dots, v_m)^\perp = (R(A))^\perp.$$

We conclude $\ker A = (R(A))^\perp$, as stated. The last statement is merely a rephrasing of this equality for $F = \mathbb{R}$. □

Remark 4.26. Let $U \subset F^n$ be a subspace of F^n. We can use Proposition 4.25 to reinterpret $U^\perp$. Let U be generated by the vectors $v_1, v_2, \dots, v_m$. 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.

Exercise 4.2.1. Prove Lemma 4.6 using the column vectors of A.

Exercise 4.2.2. Prove Proposition 4.16. If matrices A and B have a product AB that is invertible, does this imply that A and B are invertible? Cf. Exercise 5.3.4.

Exercise 4.2.3. Prove Remark 4.24.

Exercise 4.2.4. For the given matrix A and the vector x, determine Ax.

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

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

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

Exercise 4.2.5. Prove Proposition 4.14.

Exercise 4.2.6. For each of the linear maps $f\colon F^n \to F^m$ of the exercises of Section 3.2, give a matrix M such that f is given by
$$x \mapsto Mx.$$

Exercise 4.2.7. 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
$$f\big((x_1, x_2, \ldots, x_n)\big) = x_1 v_1 + \ldots + x_n v_n.$$

Exercise 4.2.8. Determine the matrix M for which $f_M\colon \mathbb{R}^3 \to \mathbb{R}^3$ is reflection in the plane given by $x + 2y - z = 0$.

Exercise 4.2.9. For which $i, j \in \{1, \ldots, 5\}$ 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.

Exercise 4.2.10. For each $i \in \{1, , \dots, 5\}$, we definer 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?

Exercise 4.2.11. 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) $j\colon \mathbb{R}^2 \to \mathbb{R}^3$, $v \mapsto (\langle v, w_1\rangle, \langle v, w_2\rangle, \langle v, w_3\rangle)$, met $w_1 = (1, -1)$, $w_2 = (2, 3)$ en $w_3 = (-2, 4)$.

Exercise 4.2.12. Take the linear maps f and g of the previous exercise 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) en verify that this product does indeed correspond with the composition of the linear maps.

Exercise 4.2.13. Another way to define when a matrix $A \in \mathrm{Mat}(m \times n, F)$ is invertible, is to require that there exist matrices B and C such that $AB = I_m$ and $CA = I_n$. Show that this implies $B = C$. Show also that the existence of B implies that f_A is surjective, while the existence of C implies that f_A is injective, so that f_A is indeed an isomorphism. Later we will see that this implies $m = n$.

Exercise 4.2.14. Let F be a field and m, n nonnegative integers. Show that there exists an isomorphism

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

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. Matrix manipulations

In this chapter we will see how to do some explicit computations with matrices, such as inverting them (if possible), determining their kernels, and using them to compute intersections of subspaces. Most computations are based on a so-called row echelon form of a matrix, which we will introduce in Section 5.2. This same notion can also be used to prove surprisingly many theoretical results, such as the fact that invertible matrices are square, and for any subspace $U \subset F^n$, the inclusion $U \subset (U^\perp)^\perp$ is in fact an equality. Several of these results, however, can be proved a lot more easily and conceptually after we introduce the notion of *dimension* in Chapter 6.

It is useful to see both how much theory can be derived from the computational approach, and how much easier the conceptual approach makes life. We therefore include the computational proofs of several theoretical results in this chapter, and will revisit them in Section 7.1.

This also means that the reader can, if wanted, skip the proofs of several results at first reading, as they will be redone in Section 7.1. This includes the proofs of 5.13, 5.17, 5.20, 5.22, 5.23, 5.24, 5.27, 5.28, 5.29, 5.30 of Sections 5.2 and 5.3.

5.1. **Elementary row and column operations.** Matrices are not only a convenient means to specify linear maps, they are also very suitable for doing computations. The main tool for that are the so-called 'elementary row and column operations.'

Definition 5.1. 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). Also note that row equivalence is indeed an equivalence, as all elementary row operations are invertible by an elementary row operation of the same type.

We define *elementary column operations* and *column equivalence* in a similar way, replacing the word 'row' by 'column' each time it appears.

Proposition 5.2. *Suppose A and A' are $m \times n$ matrices. If A and A' are row equivalent, then there is an invertible $m \times m$ matrix B, only depending 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 $A \in \mathrm{Mat}(m \times n, F)$. We denote by $E_{ij} \in \mathrm{Mat}(m, F)$ the matrix whose only non-zero entry is at position (i, j) and has value 1. (So $E_{ij} = (\delta_{ik}\delta_{jl})_{1 \le k,l \le m}$.) Also, we set $M_i(\lambda) = I_m + (\lambda - 1)E_{ii}$; this is a matrix whose non-zero entries are all on the diagonal, and have the value 1 except the entry at position (i, i), which has value λ.

Then it is easily checked that multiplying the ith row of A by λ amounts to replacing A by $M_i(\lambda)A$, and that adding λ times the jth row of A to the ith row of A amounts to replacing A by $(I_m + \lambda E_{ij})A$. As noted before, switching two rows can be achieved by a sequence of operations of the first two types, so it amounts to performing a sequence of corresponding replacements of the matrix A.

Now we have that $M_i(\lambda)$ and $I_m + \lambda E_{ij}$ (for $i \neq j$) are invertible, with inverses $M_i(\lambda^{-1})$ and $I_m - \lambda E_{ij}$, respectively. (We can undo the row operations by row operations of the same kind.) Let $B_1, B_2, \dots, B_r$ be the matrices corresponding to the row operations we have performed on A to obtain A', then

$$A' = B_r\Big(B_{r-1} \cdots \big(B_2(B_1A)\big) \cdots\Big) = (B_rB_{r-1} \cdots B_2B_1)A\,,$$

and $B = B_rB_{r-1} \cdots B_2B_1$ is invertible as 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 $A^\top$. $\square$

The following proposition shows that the elementary row operations do not change the row space and the kernel of a matrix.

Proposition 5.3. *If M and M' are row equivalent matrices, then we have*

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

Proof. Exercise. □

Proposition 5.4. *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 5.2 there is an invertible matrix B such that $M' = BM$. Since f_B is an isomorphism, the composition

$$f_{M'} = f_B \circ f_M$$

is surjective or injective if and only if f_M is. □

Exercises.

Exercise 5.1.1. Given a vector space V over a field F and vectors $v_1, v_2, \ldots, v_n \in V$. Set $W = L(v_1, v_2, \ldots, v_n)$. Using Remark 2.23, give short proofs of the following equalities of subspaces.

(1) $W = L(v'_1, \ldots, v'_n)$ where for some fixed j and some nonzero scalar $\lambda \in F$ we have $v'_i = v_i$ for $i \neq j$ and $v'_j = \lambda v_j$ (the j-th vector is scaled by a nonzero factor λ).
(2) $W = L(v'_1, \ldots, v'_n)$ where for some fixed j, k with $j \neq k$ and some scalar $\lambda \in F$ we have $v'_i = v_i$ for $i \neq k$ and $v'_k = v_k + \lambda v_j$ (a scalar multiple of v_j is added to v_k).
(3) $W = L(v'_1, \ldots, v'_n)$ where for some fixed j and k we set $v'_i = v_i$ for $i \neq j, k$ and $v'_j = v_k$ and $v'_k = v_j$ (the elements v_j and v_k are switched),

Exercise 5.1.2. 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. Show that for the rows $v'_1, v'_2, \ldots, v'_m$ of M' we have $L(v_1, \ldots, v_m) = L(v'_1, \ldots, v'_m)$.

Exercise 5.1.3. Prove Proposition 5.3.

Exercise 5.1.4. In the following sequence of matrices, each is obtained from the previous by one or two elementary row operations. Find, for each $1 \le i \le 9$, 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.

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

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

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

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

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

Exercise 5.1.5. 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'.

Exercise 5.1.6. Suppose $A \in \mathrm{Mat}_{m,n}(F)$ is a matrix. Let A' be a matrix obtained from A by applying a sequence of elementary row and column operations. Then there are isomorphisms $\varphi\colon F^n \to F^n$ and $\psi\colon F^m \to F^m$ with

$$f_{A'} = \psi \circ f_A \circ \varphi.$$

5.2. **Row Echelon Form.** 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 equal to 1 and (except for the top row) farther to the right than the pivot in the row above.

Example 5.5. The matrix A_9 of Exercise 5.1.4 is almost in row echelon form, except for scaling the bottom row by a factor $\frac{1}{190}$. The following matrices are all in row echelon form, with the last one describing the most general shape.

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

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

To make the most general shape more precise, note that there are $0 \le r \le m$ and $1 \le j_1 < j_2 < \cdots < j_r \le n$ 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. This algorithm is the key to most computations with matrices.

Proposition 5.6 (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 5.7. 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 $\mathrm{char}(F) = 3$, 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.

Remark 5.8. The row space of A in Example 5.7 is spanned by its three rows. Assume $\mathrm{char}\, F \neq 3$, so $3 \neq 0$. By Proposition 5.3, the row spaces of A and A' are the same, so this space is also spanned by the two nonzero rows of A'. We will see in the next chapter that the space can not be generated by fewer elements. More generally, the number of nonzero rows in a matrix in row echelon form is the minimal number of vectors needed to span its row space.

Example 5.9 (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 5.9 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 only introduce the denominators in the very last step, when we scale the rows so that all pivots are 1.

Example 5.10.

$$\begin{pmatrix} 3 & 5 & 2 & 2 \\ 1 & 3 & -4 & 3 \\ 2 & -2 & 5 & -1 \\ -1 & 3 & 1 & -3 \end{pmatrix} \rightsquigarrow \begin{matrix} R_2 \\ R_1 \\ R_3 \\ R_4 \end{matrix} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 3 & 5 & 2 & 2 \\ 2 & -2 & 5 & -1 \\ -1 & 3 & 1 & -3 \end{pmatrix}$$

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

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

$$\rightsquigarrow \begin{matrix} R_1 \\ R_2 \\ R_3 - R_4 \\ R_4 \end{matrix} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 0 & 2 & 11 & -7 \\ 0 & 0 & 21 & -14 \\ 0 & 0 & 36 & -21 \end{pmatrix} \rightsquigarrow \begin{matrix} R_1 \\ R_2 \\ R_3 \\ R_4 - R_3 \end{matrix} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 0 & 2 & 11 & -7 \\ 0 & 0 & 21 & -14 \\ 0 & 0 & 15 & -7 \end{pmatrix}$$

$$\rightsquigarrow \begin{matrix} R_1 \\ R_2 \\ R_3 - R_4 \\ R_4 \end{matrix} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 0 & 2 & 11 & -7 \\ 0 & 0 & 6 & -7 \\ 0 & 0 & 15 & -7 \end{pmatrix} \rightsquigarrow \begin{matrix} R_1 \\ R_2 \\ R_3 \\ R_4 - 2R_3 \end{matrix} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 0 & 2 & 11 & -7 \\ 0 & 0 & 6 & -7 \\ 0 & 0 & 3 & 7 \end{pmatrix}$$

$$\rightsquigarrow \begin{matrix} R_1 \\ R_2 \\ R_4 \\ R_3 \end{matrix} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 0 & 2 & 11 & -7 \\ 0 & 0 & 3 & 7 \\ 0 & 0 & 6 & -7 \end{pmatrix} \rightsquigarrow \begin{matrix} R_1 \\ R_2 \\ R_3 \\ R_4 - 2R_3 \end{matrix} \begin{pmatrix} 1 & 3 & -4 & 3 \\ 0 & 2 & 11 & -7 \\ 0 & 0 & 3 & 7 \\ 0 & 0 & 0 & -21 \end{pmatrix}$$

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

While the row echelon form of a matrix is not unique, the reduced row echelon form below is.

Definition 5.11. 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 $a_{ij_k} = 0$ for all $1 \leq k \leq r$ and $i \neq k$. This means that the entries above the leading 1's, the pivots, in the nonzero rows 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 5.12. *Suppose that $A \in \mathrm{Mat}_{m,n}(F)$ 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

$$\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_r v_r$$

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 5.13. *Suppose that $A \in \mathrm{Mat}_{m,n}(F)$ is a matrix in reduced row echelon form with r nonzero rows. Then the row space $R(A) \subset F^n$ can not be generated by fewer than r elements.*

Proof. Suppose $v_1, v_2, \ldots, v_s \in F^n$ generate the row space $R(A)$. Let $A' \in \mathrm{Mat}_{s,n}(F)$ be the matrix of which the rows are $v_1, v_2, \ldots, v_s$, and let $A'' \in \mathrm{Mat}_{s,n}(F)$ be a matrix in reduced row echelon form that is row equivalent to A'. Then the row spaces $R(A)$, $R(A')$, and $R(A'')$ are all the same by Proposition 5.3, so the r nonzero rows of A are the same as those of A'' by Proposition 5.12. Since A'' has s rows in total, this implies $s \geq r$. □

Corollary 5.14. *The reduced row echelon form is unique in the sense that if $A, A' \in \mathrm{Mat}(m \times n, F)$ are both in reduced row echelon form, and A and A' are row equivalent, then $A = A'$.*

Proof. For any two matrices A and A' that can be obtained from each other by a sequence of elementary row operations, the row spaces $R(A)$ and $R(A')$ are the same by Proposition 5.3. Therefore, by Proposition 5.12, the nonzero rows of A and A' coincide as well, and as the matrices have the same size, they also have the same number of zero rows. □

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 5.15. It follows from Corollary 5.14 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. By Corollary 5.13 it is the minimal number of generators for the row space $R(A)$ of A. We will see later that this number r equals the so-called *rank* of the matrix A.

Whether or not a linear map $F^n \to F^m$ is surjective can be read off from a row echelon form of the matrix that determines the map.

Proposition 5.16. *Let $A \in \mathrm{Mat}_{m,n}(F)$ 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 surjective if and only if A' has no zero rows.*

Proof. By Proposition 5.4, the map f_A is surjective if and only if $f_{A'}$ is surjective, so it suffices to do the case $A = A'$. Similarly, since A' has the same number of zero rows as the *reduced* row echelon form associated to A', we may also assume that $A = A'$ is in reduced row echelon form.

If A has a zero bottom row, then obviously f_A is not surjective, as the last entry of each element in the image equals 0, so for instance the standard vector e_m is not in the image. Suppose conversely that A has no zero row. Then each row contains a pivot, say in the columns $j_1, j_2, \ldots, j_m$. This implies $e_i = f_A(e_{j_i}) \in \operatorname{im} f_A$ for each $1 \leq i \leq m$, which implies $F^m = L(e_1, \ldots, e_m) \subset \operatorname{im} f_A$, so f_A is indeed surjective. □

Corollary 5.17. *Suppose $f\colon F^n \to F^m$ is a surjective linear map. Then we have $n \geq m$.*

Proof. Exercise. □

In the next section we will see that injectivity can also be read off from a row echelon form.

Exercises.

Exercise 5.2.1. Write down some matrices, bring them in row echelon form and determine whether the associated linear map is surjective or not.

Exercise 5.2.2. Prove Corollary 5.17.

5.3. **Generators for the kernel.** If we want to compute generators for the kernel of a matrix $A \in \mathrm{Mat}_{m,n}(F)$, then, according to Proposition 5.3, we may replace A by any row equivalent matrix. In particular, it suffices to understand how to determine generators for the kernel of matrices in row echelon form. We start with an example.

Example 5.18. 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 $x = (x_1, x_2, x_3, x_4, x_5, x_6, x_7)^\top$ is contained in

$$\ker M = R(M)^\perp = \{x \in \mathbb{R}^7 \ : \ \langle v_i, x\rangle = 0 \text{ for } i = 1, 2, 3\}.$$

Then the coordinates x_1, x_3, x_5, which belong to the columns with a pivot, are uniquely determined by the coordinates x_2, x_4, x_6, x_7, which belong to the columns without a pivot. Indeed, starting with the lowest nonzero row, the equation $\langle v_3, x\rangle = 0$ gives $x_5 + x_6 + x_7 = 0$, so

$$x_5 = -x_6 - x_7.$$

The equation $\langle v_2, x\rangle = 0$ then gives $x_3 - x_4 + 2x_5 - x_6 + 2x_7$, so

$$x_3 = x_4 - 2(-x_6 - x_7) + x_6 - 2x_7 = x_4 + 3x_6.$$

Finally, the equation $\langle v_1, x\rangle = 0$ gives

$$x_1 = -2x_2 + (x_4 + 3x_6) - 2(-x_6 - x_7) - x_6 + 3x_7 = -2x_2 + x_4 + 4x_6 + 5x_7.$$

Moreover, any choice for the values x_2, x_4, x_6, x_7, with these corresponding values for x_1, x_3, x_5, does indeed give an element of the kernel $\ker M$, as the equations $\langle v_i, x\rangle = 0$ for $1 \le i \le 3$ are automatically satisfied. With $q = x_2$, $r = x_4$, $s = x_6$, and $t = x_7$, we may write

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

$$= qw_2 + rw_4 + sw_6 + tw_7,$$

where

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

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 5.19. *Let $A \in \mathrm{Mat}_{m,n}(F)$ 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 $1 \le k \le n$ 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 $1 \le l \le n$ with $l \ne k$ and $l \notin \{j_1, j_2, \ldots, j_r\}$.*

Furthermore, the $n-r$ vectors w_k (for $1 \le k \le n$ with $k \notin \{j_1, j_2, \ldots, j_r\}$) generate the kernel $\ker A$.

Proof. The proof is completely analogous to Example 5.18 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 5.19, 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 5.18 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 5.19, with $A = (a_{ij})_{i,j}$ in reduced row echelon form: if $1 \le k \le n$ and $k \notin \{j_1, j_2, \ldots, j_r\}$, then the l-th entry of w_k is given by Proposition 5.19 for $l \notin \{j_1, j_2, \ldots, j_r\}$, while the j_i-th entry of w_k is $-a_{ik}$ for $1 \le i \le r$; this yields $w_k = e_k - \sum_{i=1}^{r} a_{ik} e_{j_i}$. This is summarized in the next proposition.

Proposition 5.20. *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*

$$w_k = e_k - \sum_{\substack{1 \le i \le r \\ j_i < k}} a_{ik} e_{j_i}, \qquad \text{for } k \in \{1, \ldots, n\} \setminus \{j_1, \ldots, j_r\},$$

where $e_1, \ldots, e_n$ is the canonical basis of F^n.

Proof. We leave it as an exercise to show that this follows from Proposition 5.19. □

In Proposition 5.16 we saw how to check whether the map associated to a matrix is surjective. We can now also check whether such a map is injective.

Proposition 5.21. *Let $A \in \mathrm{Mat}_{m,n}(F)$ 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.*

Proof. By Proposition 5.4, 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 3.5, the map f_A is injective if and only if the kernel $\ker f_A = \ker A$ is zero, which, according to Proposition 5.19, happens if and only if each of the n columns of A has a pivot, so if and only if there are exactly n nonzero rows. □

Corollary 5.22. *Suppose $f \colon F^n \to F^m$ is an injective linear map. Then we have $m \geq n$.*

Proof. By Proposition 4.7, the map f coincides with f_A for some matrix $A \in \mathrm{Mat}_{m,n}(F)$. By Proposition 5.21, any row equivalent matrix in row echelon form has n nonzero rows. Since there are m rows in total, we find $n \leq m$. □

We can now show that every invertible matrix is a square matrix, as announced in Remark 4.17.

Corollary 5.23. *The following statements hold.*

(1) *If $f \colon F^n \to F^m$ is an isomorphism, then we have $m = n$.*
(2) *Every invertible matrix is a square matrix.*

Proof. The first statement follows immediately from Corollaries 5.17 and 5.22. The second statement follows immediately from the first. □

Corollary 5.24. *Suppose $f \colon F^n \to F^n$ is a linear map. Then f is injective if and only if it is surjective.*

Proof. By Proposition 4.7, the map f coincides with f_A for some matrix $A \in \mathrm{Mat}_n(F)$. Suppose A' is a row equivalent matrix in row echelon form. Then A' has n rows in total, so it has n nonzero rows if and only if it has no zero rows. By Propositions 5.16 and 5.21, this implies that f is injective if and only if it is surjective. □

If we put the vectors w_2, w_4, w_6, w_7 of Example 5.18 as columns in a matrix, then we obtain

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

If we reflect this matrix in the indicated diagonal (which corresponds to switching rows with column and reversing their orders), then we obtain

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

which is in reduced row echelon form. It has three columns without pivots and the corresponding generators for the kernel of this matrix, as obtained by Proposition 5.19 or 5.20, are the reversed nonzero rows of the reduced row echelon form for the matrix M. (Check this!)

Proposition 5.27 says that this happens in general.

Lemma 5.25. *Let F be a field, let $\sigma\colon \{1,2,\dots,n\} \to \{1,2,\dots,n\}$ be a bijection, and $\sigma^*\colon F^n \to F^n$ the map given by*

$$(x_1, x_2, \dots, x_n) \mapsto (x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)}).$$

Then the map σ^* is an isomorphism and it preserves dot products, i.e., for all $x, y \in F^n$ we have $\langle \sigma^*(x), \sigma^*(y)\rangle = \langle x, y\rangle$.

Proof. Exercise. □

Lemma 5.26. *Suppose that $f\colon F^n \to F^n$ is an isomorphism that preserves dot products, i.e., for all $x, y \in F^n$ we have $\langle f(x), f(y)\rangle = \langle x, y\rangle$. Then for every $U \subset F^n$ we have $f(U^\perp) = (f(U))^\perp$.*

Proof. Suppose $z \in F^n$ and set $x = f^{-1}(z)$. Then for all $u \in U$ we have

$$\langle z, f(u)\rangle = \langle f(x), f(u)\rangle = \langle x, u\rangle,$$

so we have $\langle z, f(u)\rangle = 0$ for all $u \in U$ if and only if $\langle x, u\rangle = 0$ for all $u \in U$. This means $z \in (f(U))^\perp$ if and only if $x \in U^\perp$. Because f is bijective, the latter is the case if and only if we have $z = f(x) \in f(U^\perp)$. We conclude $f(U^\perp) = (f(U))^\perp$. □

Proposition 5.27. *Let $A \in \mathrm{Mat}_{m,n}(F)$ be a matrix in reduced row echelon form and let $v_1, v_2, \dots, v_r$ be its nonzero rows. Let $j_1 < j_2 < \dots < j_r$ denote the numbers of the columns with a pivot and let $w_k \in \ker A$ be as in Proposition 5.20 for k with $1 \le k \le n$ and $k \notin \{j_1, j_2, \dots, j_r\}$. Let*

$$\sigma\colon \{1,2,\dots,n\} \to \{1,2,\dots,n\}$$

be the bijection that reverses the order, so $\sigma(j) = n+1-j$, and set $v_i' = \sigma^*(v_i)$ for $1 \le i \le r$ and $w_k' = \sigma^*(w_k)$ for $1 \le k \le n$ with $k \notin \{j_1, j_2, \dots, j_r\}$. Let B denote the $(n-r) \times n$ matrix of which the rows are the w_k' in decreasing order of k. Then

1. *the matrix B is in reduced row echelon form,*
2. *the generators of $\ker B$, as obtained in Proposition 5.20, are $v_r', \dots, v_1'$,*
3. *we have $\sigma^*(\ker A) = R(B)$, and*
4. *we have $\sigma^*(R(A)) = \ker B$.*

Proof. We leave the first two statements as an exercise to the reader. The third statement follows from the fact that the w_k, with $1 \le k \le n$ and $k \notin \{j_1, j_2, \dots, j_r\}$, generate $\ker A$, so the rows $w_k' = \sigma^*(w_k)$ of B generate $\sigma^*(\ker A)$ by Lemma 3.3. The fourth statement follows immediately from the second, as we have

$$\ker B = L(v_1', \dots, v_r') = L(\sigma^*(v_1), \dots, \sigma^*(v_r)) = \sigma^*(L(v_1, \dots, v_r)) = \sigma^*(R(A)),$$

again by Lemma 3.3. □

Pictorially, we see the following. Let A be a matrix in reduced row echelon form, say

$$\begin{array}{c}\left(\begin{array}{ccccccccccccccc}
0 & \cdots & 0 & 1 & a_{1,j_1+1} & \cdots & a_{1,j_2-1} & 0 & a_{1,j_2+1} & \cdots & a_{1,j_r-1} & 0 & a_{1,j_r+1} & \cdots & a_{1,n} \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & a_{2,j_2+1} & \cdots & a_{2,j_r-1} & 0 & a_{2,j_r+1} & \cdots & a_{2,n} \\
\vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & a_{r,j_r+1} & \cdots & a_{r,n} \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
\vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0
\end{array}\right), \\
\begin{array}{ccccccccccccccc}
 & & & j_1 & & & & j_2 & & \cdots & & j_r & & &
\end{array}\end{array}$$

and let $\tilde{A}$ be the matrix of which the columns are the generators of $\ker A$ as in Proposition 5.20. Then the matrix $\tilde{A}$ equals

$$\begin{matrix} \\ \\ \\ j_1 \\ \\ \\ \\ j_2 \\ \\ \vdots \\ \\ j_r \\ \\ \\ \\ \end{matrix}
\begin{pmatrix}
1 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
0 & \cdots & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & \cdots & 0 & -a_{1,j_1+1} & \cdots & -a_{1,j_2-1} & -a_{1,j_2+1} & \cdots & -a_{1,j_r-1} & -a_{1,j_r+1} & \cdots & -a_{1,n} \\
0 & \cdots & 0 & 1 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
0 & \cdots & 0 & 0 & \cdots & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & \cdots & 0 & 0 & \cdots & 0 & -a_{2,j_2+1} & \cdots & -a_{2,j_r-1} & -a_{2,j_r+1} & \cdots & -a_{2,n} \\
0 & \cdots & 0 & 0 & \cdots & 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & & \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\
0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \\
0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & -a_{r,j_r+1} & \cdots & -a_{r,n} \\
0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 1 & \cdots & 0 \\
\vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 1
\end{pmatrix}.$$

Check that indeed we have $A\tilde{A} = 0$! Proposition 5.27 states that the matrix B, obtained from reflecting $\tilde{A}$ in the indicated diagonal, is in reduced row echelon form; moreover, if $\tilde{B}$ is the matrix associated to B as $\tilde{A}$ is associated to A, i.e., the columns of $\tilde{B}$ are the generators of $\ker B$ as in Proposition 5.20, then $\tilde{B}$ equals the reflection of the nonzero rows of A in the indicated diagonal.

Corollary 5.28. *Let $A \in \mathrm{Mat}_{m,n}(F)$ be a matrix in row echelon form with r nonzero rows. Then the kernel $\ker A$ can be generated by $n - r$ elements and not by fewer than $n - r$.*

Proof. By Proposition 5.3, we may replace A by its reduced row echelon form without loss of generality. We will use the notation of Proposition 5.27. Since B is in reduced row echelon form, the row space $R(B) = \sigma^*(\ker A)$ can not be generated by fewer than $n - r$ elements, according to Proposition 5.13. Since $\sigma^* \colon F^n \to F^n$ maps any generators of $\ker A$ to generators of $\sigma^*(\ker A)$, the subspace $\ker A$ can not be generated by fewer than $n - r$ elements either. $\square$

Proposition 5.29. *Let $A \in \mathrm{Mat}_{m,n}(F)$ be a matrix. Then in F^n we have $R(A) = (\ker A)^\perp$.*

Proof. By Proposition 5.3, we may assume without loss of generality that A is in reduced row echelon form. We will use the notation of Proposition 5.27. From the last two statements of that proposition and the equality $\ker B = (R(B))^\perp$ given by Proposition 4.25, we conclude $\sigma^*(R(A)) = (\sigma^*(\ker A))^\perp = \sigma^*((\ker A)^\perp)$, where the last equality follows from Lemma 5.26. Applying the inverse of σ^* to both sides gives the result. □

Corollary 5.30. *Let $U \subset F^n$ be a finitely generated subspace of F^n. Then we have $(U^\perp)^\perp = U$. Moreover, if U can be generated by r elements and not by fewer, then $U^\perp$ can be generated by $n - r$ elements and not by fewer.*

Proof. Let $v_1, v_2, \dots, v_m \in F^n$ be generators of U, and let $A \in \mathrm{Mat}_{m,n}(F)$ be the matrix of which the rows are $v_1, v_2, \dots, v_m$. Then we have $R(A) = U$ and $\ker A = U^\perp$, so by Proposition 5.29 we find

$$U = R(A) = (\ker A)^\perp = (U^\perp)^\perp.$$

Let A' be a matrix in row echelon form that is row equivalent to A. Then the number of nonzero rows of A equals r by Corollary 5.13. By Corollary 5.28, the subspace $U^\perp = \ker A$ can be generated by $n - r$ elements and not by fewer. □

Note that from Proposition 2.37 we already knew the trivial inclusion $U \subset (U^\perp)^\perp$. We will see a non-computational proof of equality in Section 7.1.

Remark 5.31. The hypothesis in Corollary 5.30 that U be finitely generated is not necessary. We will see later in Proposition **??** that any subspace of a finitely generated vector space V is itself finitely generated. As F^n is finitely generated, it follows that any subspace of F^n is finitely generated as well. An elegant proof for $V = F^n$ is given in Exercise 5.3.8.

We will use Corollary 5.30 in Section 5.4 to compute intersections inside F^n.

Exercises.

Exercise 5.3.1. Prove Proposition 5.19.

Exercise 5.3.2. Prove Lemma 5.25.

Exercise 5.3.3. Determine the "reduced row echelon form" for the following matrices over $\mathbb{C}$ and give generators for their kernels.

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

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

Exercise 5.3.4. Suppose the product AB of matrices $A, B \in \mathrm{Mat}_n(F)$ is invertible. Prove that A and B are also invertible. Cf. 5.3.4.

Exercise 5.3.5. Suppose M, N are $n \times n$ matrices with $MN = I_n$. Prove that then also $NM = I_n$.

Exercise 5.3.6. Let F be a field and S any subset of F^n. Show that we have $L(S) = (S^\perp)^\perp$, cf. Proposition 2.37.

Exercise 5.3.7. Let F be a field and define a partial ordering $\prec$ on F by $0 \prec 1$ and $1 \prec x$ for all $x \notin \{0,1\}$. Let $A \in \mathrm{Mat}_{m,n}(F)$ be a matrix in reduced row echelon form with r nonzero rows. Show that the following statements hold and use them to prove Proposition 5.27.

(1) Equip F^n with the induced lexicographic ordering. If $v_1, v_2, \dots, v_r$ are the nonzero rows of A, then v_r is the unique smallest nonzero element of $R(A)$. Moreover, for each $1 \leq i \leq r$, the vector v_i is the unique smallest element of $R(A)$ that is not contained in $L(v_{i+1}, \dots, v_r)$.
(2) Equip F^n with the induced anti-lexicographic ordering. Let $J = \{j_1, \dots, j_r\}$ be the set of the numbers of the columns with pivot and for $1 \leq k \leq n$ with $k \notin J$, let w_k be as in Proposition 5.20. Then for each $1 \leq k \leq n$ with $k \notin J$, the element w_k is the unique smallest element of $\ker A$ that is not contained in

$$L(\,\{w_l \;:\; 1 \leq l < k,\; l \notin J\}\,).$$

Exercise 5.3.8. Let F be a field and $U \subset F^n$ a subspace. This exercise proves that U is finitely generated. Take the same partial ordering $\prec$ on F as in Exercise 5.3.7, and equip F^n with the induced lexicographic ordering. We construct a sequence $z_1, z_2, \dots$ of elements in U as follows. For any $k = 1, 2, \dots$, we let z_k be a minimal vector in $U \setminus L(z_1, \dots, z_{k-1})$, if this set is nonempty; otherwise, we stop and the sequence is finite.

(1) Show that every nonempty subset $S \subset F^n$ has minimal elements, so that we can indeed choose z_k as described.
(2) Show that z_k starts with fewer zeros than z_{k-1}.
(3) Show that the sequence is finite of length at most n.
(4) Conclude that U is finitely generated.
(5) Show that each z_k is the *unique* minimal vector in $U \setminus L(z_1, \dots, z_{k-1})$.

5.4. **Computing intersections.**

Proposition 5.32. *Suppose F is a field and $U_1, U_2 \subset F^n$ are finitely generated 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. □

Later we will see that it is not necessary to assume the the subspaces U_1 and U_2 are finitely generated, cf. 5.31.

Proposition 5.32 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 2.43, and if $v_1, v_2, \dots, v_r \in F^n$ generate a subspace $V \subset F^n$, then $V^\perp$ is the kernel of the matrix whose rows are $v_1, v_2, \dots, v_n$ by Proposition 4.25 and we can compute generators for this kernel with Proposition 5.19.

Example 5.33. 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 kernel 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. By Proposition 5.20, 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)

$$z_4 = (-2,-1,-3,1,0) \qquad \text{and} \qquad z_5 = (-1,-1,0,0,1).$$

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 5.34. Let U and V be as in Example 5.33. Just as in Example 5.33, 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 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 $3v_1 + v_3$.

Yet a third method to compute the intersection is based on the following proposition. Recall from Examples 3.7 that 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 $V_1 \times V_2 \to V_2$ are linear.

Proposition 5.35. *Let $f_1\colon V_1 \to W$ and $f_2\colon V_2 \to W$ be two linear maps. Let $f_1 \times f_2\colon V_1 \times V_2 \to W$ denote the linear map given by*

$$V_1 \times V_2 \ni (v_1, v_2) \mapsto f_1(v_1) + f_2(v_2).$$

For $i = 1, 2$, let $U_i \subset V_i$ be the projection of $\ker(f_1 \times f_2)$ onto V_i. Then we have

$$f_1(U_1) = f_2(U_2) = \operatorname{im}(f_1) \cap \operatorname{im}(f_2).$$

Proof. Suppose $z \in f_1(U_1)$. Then there are elements $v_1 \in V_1$ and $v_2 \in V_2$ with $(v_1, v_2) \in \ker(f_1 \times f_2)$, so $v_1 \in U_1$, and $f_1(v_1) = z$. Then we have $f_1(v_1) + f_2(v_2) = 0$, so $z = f(v_1) = f_2(-v_2) \in \operatorname{im}(f_1) \cap \operatorname{im}(f_2)$ and thus $f_1(U_1) \subset \operatorname{im}(f_1) \cap \operatorname{im}(f_2)$. Conversely, suppose $z \in \operatorname{im}(f_1) \cap \operatorname{im}(f_2)$. Then there are $v_1 \in V_1$ and $v_2 \in V_2$ with $z = f_1(v_1) = f_2(v_2)$. This gives $(f_1 \times f_2)\big((v_1, -v_2)\big) = f_1(v_1) + f_2(-v_2) = z - z = 0$, so $(v_1, -v_2) \in \ker(f_1 \times f_2)$. This implies $v_1 \in U_1$ and hence $z \in f_1(U_1)$. We conclude $f_1(U_1) = \operatorname{im}(f_1) \cap \operatorname{im}(f_2)$. The equality $f_2(U_2) = \operatorname{im}(f_1) \cap \operatorname{im}(f_2)$ follows by symmetry. □

Example 5.36. With $u_1, u_2, u_3, v_1, v_2, v_3$ as in Example 5.33, let $f, g\colon \mathbb{R}^3 \to \mathbb{R}^5$ be the linear maps given by

$$\begin{aligned} f\big((\lambda_1, \lambda_2, \lambda_3)\big) &= \lambda_1 u_1 + \lambda_2 u_2 + \lambda_3 u_3, \\ g\big((\lambda_1, \lambda_2, \lambda_3)\big) &= \lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3. \end{aligned}$$

Then we have $\operatorname{im} f = U$ and $\operatorname{im} g = V$. Identifying, $\mathbb{R}^3 \times \mathbb{R}^3$ with $\mathbb{R}^6$, the map $f \times g\colon \mathbb{R}^6 \to \mathbb{R}^5$ is given by the matrix

$$\left(\begin{array}{ccc|ccc} 1 & -1 & 3 & -2 & 1 & -1 \\ 3 & 2 & 2 & 0 & 2 & 0 \\ 1 & -2 & 0 & -6 & -3 & -3 \\ 2 & 3 & -1 & 3 & 1 & -2 \\ 2 & 2 & -4 & -2 & -3 & -1 \end{array}\right),$$

where the line separates the two factors in $\mathbb{R}^3 \times \mathbb{R}^3$. The reduced row echelon form of this matrix, with some of its rows scaled to get rid of denominators, equals

$$\left(\begin{array}{ccc|ccc} 11 & 0 & 0 & 0 & -8 & -70 \\ 0 & 22 & 0 & 0 & 25 & 161 \\ 0 & 0 & 22 & 0 & 21 & 49 \\ \hdashline 0 & 0 & 0 & 1 & 0 & -3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Therefore, the kernel of $f \times g$ is generated by

$$z_1 = (16, -25, -21 \mid 0, 22, 0) \qquad \text{and} \qquad z_2 = (140, -161, -49 \mid 66, 0, 22),$$

which correspond to the relations

$$\begin{array}{rcrl} 16u_1 - 25u_2 - 21u_3 & + & 22v_2 & = 0 \\ 140u_1 - 161u_2 - 49u_3 & + & 66v_1 + 22v_3 & = 0 \end{array}$$

The projections of z_1 and z_2 onto the first factor in $\mathbb{R}^3 \times \mathbb{R}^3$ are $(16, -25, -21)$ and $(140, -161, -49)$ and the projections on the second factor are $(0, 22, 0)$ and $(66, 0, 22)$. We find the two elements

$$f((16, -25, -21)) = 16u_1 - 25u_2 - 21u_3 = -22v_2 = -g((0, 22, 0)),$$
$$f((140, -161, -49)) = 140u_1 - 161u_2 - 49u_3 = -(66v_1 + 22v_3) = -g((66, 0, 22)),$$

and these elements generate $\operatorname{im} f \cap \operatorname{im} g = U \cap V$ by Proposition 5.35.

Note that in Example 5.36, when we determine generators for the kernel of the 5×6 matrix in reduced row echelon form using Proposition 5.20, the only part of the matrix that matters for the last three coordinates of the generators (corresponding to the second factor $\mathbb{R}^3$), is the part right of the vertical line and below the dashed horizontal line. This implies that the projection of the kernel $\ker(f \times g)$ onto the second factor of $\mathbb{R}^3 \times \mathbb{R}^3$ is exactly the kernel of the 2×3 matrix

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

It is a coincidence that this matrix is the same as the one in Example 5.36, but it is not a coincidence that the kernels of these 2×3 matrices coincide. This observation can be turned into an efficient variation of the method of Example 5.36, described in the following proposition.

Proposition 5.37. *Let $U, V \subset F^n$ be subspaces generated by $u_1, u_2, \ldots, u_s$, and $v_1, v_2, \ldots, v_t$ respectively. Let M be the $n \times (s+t)$ matrix of which the columns are $u_1, u_2, \ldots, u_s, v_1, v_2, \ldots, v_t$. Let M' be a row equivalent matrix in row echelon form. Let r be the number of nonzero rows in the left-most $n \times s$ submatrix of M', let N be the lower-right $(n-r) \times t$ submatrix of M', and $K \subset F^t$ the kernel of N. Then $U \cap V = g(K)$, where $g \colon F^t \to F^n$ is given by $g((\lambda_1, \ldots, \lambda_t)) = \lambda_1 v_1 + \ldots + \lambda_t v_t$.*

Proof. One proof uses the same arguments as above. We leave it as an exercise to the reader to phrase them in full generality. Exercise 5.4.8 gives a different proof. □

Example 5.38. Let $U \subset \mathbb{R}^4$ be generated by

$$u_1 = (-1, 3, 3, -3), \qquad u_2 = (-1, 1, -3, 1), \qquad \text{and} \qquad u_3 = (-2, 3, -3, 0),$$

and $V \subset \mathbb{R}^4$ by

$$v_1 = (-1, 3, -1, -3) \qquad \text{and} \qquad v_2 = (-3, 4, 2, 1).$$

We will find generators for the intersection $U \cap V$, using Proposition 5.37. The matrix with u_1, u_2, u_3, v_1, v_2 as columns is

$$M = \left(\begin{array}{ccc|cc} -1 & -1 & -2 & -1 & -3 \\ 3 & 1 & 3 & 3 & 4 \\ 3 & -3 & -3 & -1 & 2 \\ -3 & 1 & 0 & -3 & 1 \end{array}\right),$$

which has row echelon form (with some rows scaled)

$$M' = \left(\begin{array}{ccc|cc} 2 & 0 & 1 & 0 & 5 \\ 0 & 2 & 3 & 0 & 5 \\ \hdashline 0 & 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

The part of the matrix left of the vertical line has $r = 2$ nonzero rows. The kernel of the lower-right 2×2 matrix is generated by $(2, 1)$, so $U \cap V$ is generated by $z = 2v_1 + v_2$. Indeed, the kernel of M' contains the element $(-\frac{5}{2}, -\frac{5}{2}, 0, 2, 1)$, so $-\frac{5}{2}u_1 - \frac{5}{2}u_2 + 2v_1 + v_2 = 0$ and we have $2z = 5u_1 + 5u_2$, which shows $z \in U$.

Remark 5.39. 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 methods of Examples 5.34 and 5.36 and Proposition 5.37 yield 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 5.33 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.

Exercise 5.4.1. Prove Proposition 5.32.

Exercise 5.4.2. Compute the intersection $U \cap V$ with U and V as in Example 5.33 with the method of Example 5.34, but with the roles of U and V reversed. Same for the methods of Example 5.36 and Proposition 5.37.

Exercise 5.4.3. Compute the intersection $U \cap V$ with U and V as in Example 5.38 with all the methods, and compare the amount of work involved with each method. Perform each method again with the roles of U and V reversed.

Exercise 5.4.4. 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 $U \cap V$.

Exercise 5.4.5. Take two subspaces of $\mathbb{R}^6$ generated by four elements and compute generators for the intersection.

Exercise 5.4.6. Prove Proposition 5.37 along the lines of the arguments mentioned above that proposition. For a different proof, see Exercise 5.4.8.

Exercise 5.4.7. Suppose $U \subset F^n$ is a subspace that can be generated by r elements and not by fewer. Suppose $V \subset F^n$ is a subspace that can be generated by $n-r$ elements and not by fewer. Show that the intersection $U \cap V$ equals $\{0\}$ if and only if the sum $U+V$ equals F^n. [Hint: consider the $n \times n$ matrix whose first r columns generate U and whose last $n-r$ columns generate V.]

Exercise 5.4.8. * In this exercise, we will prove Proposition 5.37 more conceptually.

(1) Let $f_1 \colon V_1 \to W$ and $f_2 \colon V_2 \to W$ be linear maps. Suppose $U \subset W$ is a subspace that is complementary to $\operatorname{im} f_1$. Let $\pi \colon W \to U$ be the projection of W onto U along $\operatorname{im} f_1$ as in Exercise 3.2.9, and set $\tilde{f}_2 = \pi \circ f_2 \colon V_2 \to U$. Show that we have $f_2(\ker \tilde{f}_2) = \operatorname{im}(f_1) \cap \operatorname{im}(f_2)$.
(2) Now assume $V_1 = F^s$ and $V_2 = F^t$ and $W = F^n$. Identify $F^s \times F^t$ with F^{s+t} and let M be the matrix associated to $f_1 \times f_2 \colon F^{s+t} \to F^n$. Let M' be a row equivalent matrix in row echelon form and let r denote the number of nonzero rows in the left $n \times s$ submatrix of M'. By Exercise 5.1.6 there is an isomorphism $\psi \colon F^n \to F^n$ such that $f_{M'} = \psi \circ (f_1 \times f_2)$. Show that ψ induces an isomorphism from $\operatorname{im} f_1$ to the subspace of F^n generated by the first r standard vectors $e_1, e_2, \ldots, e_r$.
(3) Let U' be the subspace of F^n generated by the last $n-r$ standard vectors $e_{r+1}, \ldots, e_n$ and set $U = \psi^{-1}(U')$. Show that $\operatorname{im} f_1$ and U are complementary subspaces in F^n.
(4) Show that $\tilde{f}_2$ and $\psi \circ \tilde{f}_2$ have the same kernel.
(5) Show that the map $\psi \circ \tilde{f}_2 \colon F^t \to U' \cong F^{n-r}$ is given by the lower-right $(n-r) \times t$ submatrix of M'.
(6) Prove Proposition 5.37.

5.5. **Inverses.** Recall that every invertible matrix is square by Corollary 5.23. In this section, we will give a method to check whether a square matrix is invertible, and, if so, to compute the inverse.

Lemma 5.40. *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 5.2, 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 C, and the fact that the dot product is linear in its variables. □

Lemma 5.40 states that if we start with a product $AB = C$, written as

$$(5) \qquad \begin{array}{ccc} & \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix} & = B \\ A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{l1} & a_{l2} & \cdots & a_{lm} \end{pmatrix} & \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{l1} & c_{l2} & \cdots & c_{ln} \end{pmatrix} & = C \end{array}$$

as in (4), 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 $A'B = C'$.

Proposition 5.41. *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 surjective, and by Proposition 5.16 or 5.21 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 5.2 there is an invertible matrix M, such that $I_n = MA$, so $A = M^{-1}$ is invertible. Applying Lemma 5.40 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 5.41. If we write $X = A^{-1}C$ for A and C as in Proposition 5.41, then we can depict the equality $AX = C$ as in (5) by

$$\begin{array}{|c|c|} \hline & X \\ \hline A & C \\ \hline \end{array} \; .$$

Applying elementary row operations to the combined matrix $\boxed{A}\boxed{C}$ yields a combined matrix $\boxed{A'}\boxed{C'}$ of matrices A' and C' that satisfy $A'X = C'$ by Lemma 5.40, depicted as follows.

$$\begin{array}{|c|c|} \hline & X \\ \hline A & C \\ \hline \end{array} \quad \rightsquigarrow \quad \begin{array}{|c|c|} \hline & X \\ \hline A' & C' \\ \hline \end{array}$$

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 $X = A^{-1}C$, 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 5.42. 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 $\text{char}(F) \neq 2$, so that $2 \neq 0$ and we can divide by 2.

We perform the row operations on A and on I in parallel, as above.

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 4 & 0 & 1 & 0 \\ 1 & 3 & 9 & 0 & 0 & 1 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 3 & -1 & 1 & 0 \\ 0 & 2 & 8 & -1 & 0 & 1 \end{array}\right)$$

$$\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 0 & -2 & 2 & -1 & 0 \\ 0 & 1 & 3 & -1 & 1 & 0 \\ 0 & 0 & 2 & 1 & -2 & 1 \end{array}\right)$$

$$\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 3 & -3 & 1 \\ 0 & 1 & 0 & -\frac{5}{2} & 4 & -\frac{3}{2} \\ 0 & 0 & 1 & \frac{1}{2} & -1 & \frac{1}{2} \end{array}\right)$$

So

$$A^{-1} = \begin{pmatrix} 3 & -3 & 1 \\ -\frac{5}{2} & 4 & -\frac{3}{2} \\ \frac{1}{2} & -1 & \frac{1}{2} \end{pmatrix}.$$

Remark 5.43. 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 5.44. *If $A \in \text{Mat}(n, F)$ is invertible, then A can be written as a product of matrices $M_i(\lambda)$ $(\lambda \neq 0)$ and $I_n + \lambda E_{ij}$ $(i \neq j)$. (Notation as in the proof of Proposition 5.2.)*

Proof. Exercise. □

Example 5.45. Let A be the matrix of Example 5.42 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 5.40 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.

Exercise 5.5.1. Determine the inverses of the following matrices

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

Exercise 5.5.2. Are the matrices

$$\begin{pmatrix} 1 & 2 \\ -2 & 4 \end{pmatrix}, \quad \begin{pmatrix} -2 & 1 & -2 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix}$$

invertible?

Exercise 5.5.3. Determine the inverse of those matrices (over $\mathbb{R}$) that are invertible.

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

$$\begin{pmatrix} 0 & 2 & -1 & 1 \\ -2 & -1 & -2 & 0 \\ 1 & 0 & -1 & 2 \\ 2 & 2 & 0 & 2 \end{pmatrix} \quad \begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & -1 \\ 1 & 0 & 0 \end{pmatrix}.$$

Exercise 5.5.4. 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, as in the proof of Proposition 5.2. For $1 \leq i, j \leq m$ with $i \neq j$ and $\lambda \in F$, we set

$$\begin{aligned} L_{ij}(\lambda) &= I_m + \lambda E_{ij} \\ M_i(\lambda) &= I_m + (\lambda - 1)E_{ii} \\ N_{ij} &= I_m + E_{ij} + E_{ji} - E_{ii} - E_{jj} \end{aligned}$$

We call these matrices *elementary matrices.*

(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 5.2).
(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 5.44.
(5) Write the following matrices as a product of elementary matrices, if possible:

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

6. Linear independence and dimension

Exercises.

Exercises for computational alternatives.

7. Rank of a linear map

7.1. Matrices revisited.

Exercises.