Linear Algebra I

Michael Stoll, 2007
additions by Ronald van Luijk, 2010

Linear Algebra I

Course No. 100 221

Fall 2006

Michael Stoll

With corrections made for version of December 10, 2007
With additions by Ronald van Luijk, 2010

Contents

1. Preliminaries

In the following, we will assume that you are familiar with the basic notions and notations of ('naive') *set theory*. This includes, for example, the following.

- Notation for *sets*: $\{1,2,3\}$, $\{x \in S : \text{some property of } x\}$, the *empty set* $\emptyset$.
- Symbols for the set of *natural numbers:*
$$\mathbb{N} = \{1,2,3,\dots\}, \quad \mathbb{N}_0 = \{0,1,2,3,\dots\},$$
integers: $\mathbb{Z}$, *rational numbers:* $\mathbb{Q}$, *real numbers:* $\mathbb{R}$, and *complex numbers:* $\mathbb{C}$ (which we will introduce shortly).
- Notations $x \in S$ (x is an element of S), $A \cap B$ (intersection of A and B), $A \cup B$ (union of A and B) $A \setminus B = \{x \in A : x \notin B\}$ (set difference of A and B), $A \subset B$ (A is a subset of B; this *includes* the case $A = B$ — my convention), $A \subsetneq B$ (A is a *proper* subset of B).
- *Pairs* (a,b), *triples* (a,b,c) and general *n-tuples* $(a_1,a_2,\dots,a_n)$, and *cartesian products* $A \times B = \{(a,b) : a \in A, b \in B\}$ etc. The product $A \times A \times \dots \times A$ of n copies of A is denoted by A^n.
- The notion of a *map* $f : A \to B$ and properties like *injectivity*, *surjectivity*, *bijectivity* (we will recall these when we need them).
- The notion of an *equivalence relation* on a set S: formally, a relation on S is a subset R of $S \times S$. For simplicity, we write $a \sim b$ for $(a,b) \in R$. Then R is an equivalence relation if it is *reflexive* ($a \sim a$ for all $a \in S$), *symmetric* (if $a \sim b$, then $b \sim a$), and *transitive* (if $a \sim b$ and $b \sim c$, then $a \sim c$). Given an equivalence relation on S, the set S has a natural decomposition into *equivalence classes*, and we can consider the *quotient set* $S/\!\sim$: the set of all equivalence classes. We will recall this in more detail later.

We will also use notation like the following.

$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \dots \cup A_n \quad \text{with} \quad \bigcup_{i=1}^{0} A_i = \emptyset$$

$$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \dots \cap A_n\,; \quad \bigcap_{i=1}^{0} A_i \quad \text{usually does not make sense}$$

$$\bigcup_{A \in \mathcal{S}} A = \bigcup \mathcal{S} \quad \text{is the union of all the sets } A \text{ that are elements of } \mathcal{S}$$

$$\sum_{i=1}^{n} a_i = a_1 + a_2 + \dots + a_n \quad \text{and is zero for } n = 0$$

$$\prod_{i=1}^{n} a_i = a_1 a_2 \cdots a_n \quad \text{and is one for } n = 0$$

2. What Is Linear Algebra?

Linear Algebra is the theory of 'linear structures'. So what is a linear structure? Well, the notion of linearity involves *addition* — you want to be able to form sums, and this addition should behave in the way you expect (it is commutative ($x+y = y+x$) and associative ($(x+y)+z = x+(y+z)$), has a zero element, and every element has a negative) — and *multiplication*, not between the elements of your

structure, but by 'scalars', for example, real numbers. This scalar multiplication should also behave reasonably (you want to have $1 \cdot x = x$ and $\lambda \cdot (\mu \cdot x) = (\lambda\mu) \cdot x$) and be compatible with the addition (in the sense that both distributive laws hold: $(\lambda + \mu)x = \lambda x + \mu x$, $\lambda(x + y) = \lambda x + \lambda y$). We will make this precise in the next section.

A set with a linear structure in the sense of our discussion is called a *linear space* or *vector space*. So Linear Algebra studies these linear spaces and the maps between them that are compatible with the linear structure: *linear maps*. This may sound somewhat abstract, and indeed, it is. However, it is exactly this level of abstraction that makes Linear Algebra an extremely useful tool. The reason for this is that linear structures abound in mathematics, and so Linear Algebra has applications everywhere (see below). It is this method of abstraction that extracts the common features of various situations to create a general theory, which forms the basis of all essential progress in mathematics. It took the mathematicians quite a long time before they came up with our modern clean formulation of the theory.

As already hinted at above, many other structures in mathematics are built on linear spaces, usually by putting some additional structure on them. Here are some examples. Even though the various notions probably do not have much of a meaning to you now, this list should show you that you can expect to use what you will learn in this course quite heavily in your further studies.

- **Geometry.** First of all, there is of course the elementary analytic geometry of the plane and space, which is based on the fact that plane and space can be turned into linear spaces by choosing a point as the origin. Usually one adds some structure that allows to talk about distances and angles. This is based on the 'dot product', which is a special case of an 'inner product', which turns a linear space into a Euclidean vector space. We will discuss this in detail later in the course.

  A more advanced branch of geometry is *Differential Geometry* that studies 'smooth' curves, surfaces and higher-dimensional 'manifolds'. Every point on such a manifold has a 'tangent space', which is a linear space, and there are many other linear spaces associated in a natural way to a manifold (for example, spaces of differential forms).
- **Analysis.** The notion of derivative is based on linearity — the derivative describes the best *linear* approximation to a function in a given point. In the case of functions of one variable, this comes down to the slope of the tangent line to the graph. However, the concept can be generalized to functions in several variables (which are functions defined on a vector space!) and even to functions between manifolds. Also the operations of differentiating and integrating functions are *linear*. Finally, spaces of functions usually carry a structure as a linear space (for example, the space of differentiable functions on the real line). Functional Analysis is concerned with these function spaces; usually one adds some structure by introducing a suitable 'norm', leading to Banach Spaces and Hilbert Spaces.
- **Algebra.** Algebra studies more general algebraic structures (like groups, rings, fields), many of which are based on linear spaces, like for example Lie Algebras. Then there is a whole bunch of so-called homology and cohomology theories involving (co-)homology groups, which are in many cases linear spaces. Such groups occur, for example, in Algebraic Topology, where they can be used to show that two spaces are essentially distinct from

each other, in that one cannot continuously deform one into the other. They also play an important role in Algebraic Geometry.

Another, somewhat different, application (which may be interesting to the EECS, CS and ECE students) is to Coding Theory, where one tries to construct good 'error-correcting codes'. Essentially all the codes that are considered are *linear* codes, which means that the codewords form a vector space (where the scalar multiplication is not by real numbers, but by elements from a 'finite field' like the field $\mathbb{F}_2 = \{0, 1\}$).

To illustrate the versatility of Linear Algebra, let me write down a few *linear equations* from various areas of mathematics. Even though they involve objects of quite different nature, the general theory applies to all of them to provide (for example) general information on the structure of the solution set.

- A system of linear equations in real numbers:
$$x + y = 3\,, \quad x + z = 5\,, \quad y + z = 6\,.$$
- A linear recurrence relation for a sequence $(F_n)_{n\geq 0}$ of real numbers (the Fibonacci numbers provide one solution):
$$F_{n+2} = F_{n+1} + F_n \quad \text{for all } n \geq 0.$$
- A linear ordinary differential equation for a (twice differentiable) function $y : \mathbb{R} \to \mathbb{R}$ (the sine and cosine functions solve it):
$$y'' + y = 0\,.$$
- A linear partial differential equation for a (twice differentiable) function f of time t and space x, y, z (the Heat Equation):
$$\frac{\partial f}{\partial t} = -\frac{\partial^2 f}{\partial x^2} - \frac{\partial^2 f}{\partial y^2} - \frac{\partial^2 f}{\partial z^2}\,.$$

The fact that Linear Algebra applies to all of them accounts for the general feeling that linear equations are 'easy', whereas nonlinear equations are 'difficult'.

3. Vector Spaces

In this section, we will give the complete formal definition of what a (real) vector space or linear space is. 'Real' here refers to the fact that the scalars are real numbers.

3.1. **Definition.** A *real vector space* or *linear space over* $\mathbb{R}$ is a set V with a distinguished zero element $0 \in V$, together with two maps $+ : V \times V \to V$ ('addition') and $\cdot : \mathbb{R} \times V \to V$ ('scalar multiplication') satisfying the following axioms. Note that for $\lambda \cdot x$, we usually write λx.

(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*.

3.2. **Remarks.**

(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.

Before we continue with the general theory, we should look at some examples.

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

3.4. **Example.** The next (still not very interesting) example is $V = \mathbb{R}$, with addition and multiplication the usual ones on real numbers. The axioms above in this case just reduce to well-known rules for computing with real numbers.

3.5. **Example.** Now we come to a very important example, which is *the* model of a vector space. We consider $V = \mathbb{R}^n$, the set of n-tuples of real numbers, 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 known facts about real numbers. As an example, let us show that the first distributive law is satisfied.

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

Of course, for $n = 2$ and $n = 3$, this is more or less what you know as 'vectors' from high school. 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 $\mathbb{R}^0$ has exactly one element, the empty tuple $()$, which we can call 0 here.)

3.6. **Exercise.** Write down proofs for the other seven axioms.

3.7. **Example.** The preceding example can be generalized. Note that we can identify $\mathbb{R}^n$ with the set of maps $f : \{1, 2, \ldots, n\} \to \mathbb{R}$, by letting such a map f correspond to the n-tuple $(f(1), f(2), \ldots, f(n))$. But there is no need to limit ourselves to these specific sets. So let us consider any set X and look at the set of all maps (or functions) from X to $\mathbb{R}$:

$$V = \mathbb{R}^X = \{f : X \to \mathbb{R}\}.$$

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 $f, g : X \to \mathbb{R}$, we have to define a new function $f + g : X \to \mathbb{R}$. The only reasonable way to do this is as follows ('point-wise'):

$$f + g : X \longrightarrow \mathbb{R}, \quad x \longmapsto f(x) + g(x),$$

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:

$$\lambda f : X \longrightarrow \mathbb{R}, \quad x \longmapsto \lambda \cdot f(x).$$

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 have

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

So let $\lambda \in \mathbb{R}$, $f, g : X \to \mathbb{R}$, and $x \in X$.

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

Note the parallelism of this proof with the one in the previous example! What do we get when X is the empty set?

3.8. **Exercise.** Write down proofs for the other seven axioms.

3.9. **Example (Leiden).** 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 $x \in \mathbb{R}^2$ satisfying $\langle a, x\rangle = b$. We wonder when L is a vector space itself, with the same addition and scalar multiplication as in $\mathbb{R}^2$ and some zero $0_L \in L$. When this is the case, the zero vector 0_L of L has to equal $0 \in \mathbb{R}^2$ (why?). This forces $b = 0$.
Conversely, assume $b = 0$. Then for two elements $x, y \in L$ we have $\langle a, x + y\rangle = \langle a, x\rangle + \langle a, y\rangle = 2b = 0$, so $x + y \in L$; this shows that the addition restrict to an addition $+_L : L \times L \to L$ on L. Similarly, one shows that the scalar multiplication restricts to a scalar multiplication $\cdot_L : \mathbb{R} \times L \to L$. The eight axioms of Definition 3.1 then hold trivially for $(L, 0, +_L, \cdot_L)$, so L is a vector space if and only if $b = 0$.

3.10. **Example.** 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 is given by

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

Anybody who can prove that the previous examples are vector spaces, will have no problems showing that $P(\mathbb{R})$ is a vector space as well.

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 rules that hold for the traditional addition and multiplication than just the eight axioms given in Definition 3.1. We will show that many such rules follow from the basic eight. The first is a cancellation rule.

3.11. **Remark.** *If three elements x, y, z of a real 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.

3.12. **Remark.** *In a real 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. □

3.13. **Remark.** *In a real vector space V, there is a unique negative for each element.*

Proof. Let $x \in V$ 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 Remark 3.11. □

3.14. **Notation.** 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.

3.15. **Remarks.** *Let $(V, 0, +, \cdot)$ be a real vector space.*

(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 $\lambda \in \mathbb{R}$ and $x \in V$ such that $\lambda x = 0$, we must have $\lambda = 0$ or $x = 0$.*

Proof. Exercise. □

Exercises

Exercise 3.15.1. *Proof Remark 3.12.*

Exercise 3.15.2. *Proof Remark 3.15.*

Exercise 3.15.3. *Check that $\mathbb{R}^n$, together with the zero and the addition and scalar multiplication as given in Example 3.5 is a vector space.*

Exercise 3.15.4. *Check that for every set X, the set $\mathbb{R}^X$ of functions $X \to \mathbb{R}$, together with the zero function and the addition and scalar multiplication as given in Example 3.7 is a vector space.*

Exercise 3.15.5. *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 (Leiden) 3.15.6. *Let $a_1, \ldots, a_t \in \mathbb{R}^n$ be vectors and $b_1, \ldots, b_t \in \mathbb{R}$ constants. Let $V \subset \mathbb{R}^n$ be the subset*

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

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

Exercise 3.15.7. *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 3.15.8. *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 3.15.9. *Is the following statement correct? "Axiom (4) of Definition 3.1 is redundant because we already know by Remark 3.15(2) that for each vector $x \in V$ the vector $-x = (-1) \cdot x$ is also contained in V."*

Exercise 3.15.10. *For each of the eight axioms in Definition 3.1, try to find a system $(V, 0, +, \cdot)$ that does not satisfy that axiom, while it does satisfy the other seven.*

4. Fields

So far, we have required our scalars to be real numbers. It turns out, however, that we can be quite a bit more general without adding any complications, by replacing the real numbers with other structures with similar properties. These structures are called *fields*.

In the formulation of the axioms below, we will use the shorthands '$\forall x \in X : \ldots$' and '$\exists x \in X : \ldots$' for the phrases 'for all $x \in X$, we have $\ldots$' and 'there exists some $x \in X$ such that $\ldots$'.

4.1. **Definition.** A *field* is a set F, together with two distinguished elements $0, 1 \in F$ with $0 \neq 1$ and two maps $+ : F \times F \to F$ ('addition') and $\cdot : F \times F \to F$ ('multiplication'), written, as usual, $(\lambda, \mu) \mapsto \lambda + \mu$ and $(\lambda, \mu) \mapsto \lambda \cdot \mu$ or $\lambda\mu$, respectively, satisfying the following axioms.

(1) $\forall \lambda, \mu \in F : \lambda + \mu = \mu + \lambda$.
(2) $\forall \lambda, \mu, \nu \in F : (\lambda + \mu) + \nu = \lambda + (\mu + \nu)$.
(3) $\forall \lambda \in F : \lambda + 0 = \lambda$.
(4) $\forall \lambda \in F \; \exists \lambda' \in F : \lambda + \lambda' = 0$.
(5) $\forall \lambda, \mu \in F : \lambda\mu = \mu\lambda$.
(6) $\forall \lambda, \mu, \nu \in F : (\lambda\mu)\nu = \lambda(\mu\nu)$.
(7) $\forall \lambda \in F : 1 \cdot \lambda = \lambda$.
(8) $\forall \lambda \in F \setminus \{0\} \; \exists \lambda'' \in F : \lambda''\lambda = 1$.
(9) $\forall \lambda, \mu, \nu \in F : \lambda(\mu + \nu) = \lambda\mu + \lambda\nu$.

Well-known examples of fields are the field $\mathbb{Q}$ of rational numbers and the field $\mathbb{R}$ of real numbers. We will introduce the field $\mathbb{C}$ of complex numbers shortly. But there are other examples of fields. For example, the smallest possible field just has the required elements 0 and 1, with the only 'interesting' definition being that $1 + 1 = 0$.

4.2. **Exercise.** Check the axioms for this field $\mathbb{F}_2 = \{0, 1\}$.

As before for vector spaces, we have some simple statements that easily follow from the axioms.

4.3. **Remarks.** *Let F be a field.*

(1) *There is only one zero 0 and one unit 1 of F.*
(2) *Likewise, for every $\lambda \in F$, there is only one $\lambda' \in F$ with $\lambda + \lambda' = 0$ as in axiom (4); we denote this element by $-\lambda$.*
(3) *For every $\lambda \in F \setminus \{0\}$, there is only one $\lambda'' \in F$ with $\lambda''\lambda = 1$ as in axiom (8); we denote this element by λ^{-1}.*
(4) *We have $0 \cdot \lambda = 0$ and $(-1)\lambda = -\lambda$ for all $\lambda \in F$. In particular, $(-1)(-1) = 1$ (taking $\lambda = -1$).*
(5) *If $\lambda\mu = 0$ for $\lambda, \mu \in F$, then $\lambda = 0$ or $\mu = 0$ (or both).*

Proof. Exercise. □

4.4. **Remark.** It is perhaps easier to remember the field axioms in the following way.

(1) $(F, 0, +)$ is an (additive) abelian group with zero element 0, called the *additive group* of F.
(2) $(F \setminus \{0\}, 1, \cdot)$ is a (multiplicative) abelian group, called the *multiplicative group* of F. (But note that the multiplication map is defined on all of F, not just on $F \setminus \{0\}$.)
(3) The distributive law holds: $\forall \lambda, \mu, \nu \in F : \lambda(\mu + \nu) = \lambda\mu + \lambda\nu$.

Given this, we can now define the notion of an F-vector space for an arbitrary field F.

4.5. **Definition.** Let F be a field. An *F-vector space* or *linear space over F* 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'), satisfying the vector space axioms given in Def. 3.1 with $\mathbb{R}$ replaced by F.

Note that in order to prove the statements in Remarks 3.11, 3.12, 3.13 and 3.15, we only had to use that $\mathbb{R}$ is a field. Hence these statements are valid for F-vector spaces in general.

The same observation applies to our examples of real vector spaces.

4.6. **Examples.**

(1) $V = \{0\}$, with the unique addition and scalar multiplication maps, is an F-vector space, again called the *zero space* (over F).
(2) F, with the addition and multiplication from its structure as a field, is an F-vector space.
(3) F^n, with addition and scalar multiplication defined component-wise, is an F-vector space.
(4) For any set X, the set F^X of all maps $X \to F$, with addition and scalar multiplication defined point-wise, is an F-vector space.
(5) The set $P(F)$ of all polynomials $a_d x^d + \ldots + a_1 x + a_0$ with coefficients a_i in F, together with addition and scalar multiplication defined coefficient-wise, is an F-vector space.
(6) The set $P_d(F)$ of all polynomials with coefficients in F of degree at most d (with $d \geq 0$ an integer).

4.7. **Warning.** 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 $\alpha \in \mathbb{F}_2$.

4.8. **Example.** 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 that we had defined earlier on $\mathbb{F}_2^X$.

4.9. **The Field of Complex Numbers.** Besides real vector spaces, *complex* vector spaces are very important in many applications. These are linear spaces over the field of complex numbers, which we now introduce.

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 the real vector space $\mathbb{R}^2$, 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. (Later, in the "Introductory Algebra" course, you will learn that there is 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*.

4.10. **Definition.** 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 $|\bar{z}| = |z|$.

4.11. **Remark.**

(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:

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

4.12. **Remark.** Historically, the necessity of introducing complex numbers was realized through the study of *cubic* (and not quadratic) equations. The reason for this is that there is a solution formula for cubic equations that in some cases requires complex numbers in order to express a real solution. See Section 2.7 in Jänich's book [J].

The importance of the field of complex numbers lies in the fact that they provide solutions to *all* polynomial equations. This is the 'Fundamental Theorem of Algebra':

Every non-constant polynomial with complex coefficients has a root in $\mathbb{C}$.

We will have occasion to use it later on. A proof, however, is beyond the scope of this course.

4.13. **Definition.** A *complex vector space* is a linear space over $\mathbb{C}$.

Exercises

Exercise 4.13.1. *Prove the Remarks of 4.3.*

Exercise 4.13.2. *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* $\{0, 1, 2\}$ *with the usual addition and multiplication, followed by taking the remainder after division by* 3.

Exercise 4.13.3. *Show that* $\mathbb{R}$ *is a vector space over* $\mathbb{Q}$*, with the usual addition and scalar multiplication.*

Exercise 4.13.4. *Prove Remark 4.11.*

Exercise 4.13.5. *For every complex number* z *we have* $\mathrm{Re}(z) = \frac{1}{2}(z + \overline{z})$ *and* $\mathrm{Im}(z) = \frac{1}{2i}(z - \overline{z})$.

Exercise 4.13.6. *Given the field* F *and the set* V *in the following cases, together with the described addition and scalar multiplication, as well as the implicit element* 0*, which cases determine a vector space? If not, then which rule is not satisfied?*

(1) *The field* $F = \mathbb{R}$ *and the set* V *of all functions* $[0, 1] \to \mathbb{R}_{>0}$*, together with the usual addition and scalar multiplication.*
(2) *Any field* F *and the set* F^n *of all* n*-tuples of elements in* F*, with coordinatewise addition and scalar multiplication.*
(3) *Example 4.8 in the notes.*
(4) *The field* $F = \mathbb{Q}$ *and the set* $V = \mathbb{R}$ *with the usual addition and multiplication.*

(5) *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.*
(6) *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.*
(7) *Any field* F *together with the subset*

$$\{(x, y, z) \in F^3 \ : \ x + 2y - z = 0\},$$

with coordinatewise addition and scalar multiplication.
(8) *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 4.13.7. *Suppose the set* X *contains exactly* n *elements. Then how many elements does the vector space* $\mathbb{F}_2^X$ *of functions* $X \to \mathbb{F}_2$ *consist of?*

Exercise 4.13.8. *Let* F *be any field. Let* X *be any set and* $x \in X$ *an element. Define*

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

Show that the usual addition and scalar multiplication of F^X *induce an addition and scalar multiplication on* U *that make* U *into a vector space.*

5. Linear Subspaces and Linear Hulls

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.

5.1. **Definition.** 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.

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'.

5.2. **Lemma.** *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. 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 ..., we have ...' and do not involve any existence statements, this is clear, since they hold (by assumption) even for all elements of V. 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 Remarks 3.13, 3.14, and 3.15). This element $u' = -u$ is contained in U by the third property of linear subspaces (take $\lambda = -1 \in F$). □

It is time for some examples.

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

5.4. **Example.** 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:

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

and

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

We conclude that U_0 is indeed a subspace.

5.5. **Example.** Let X be any set and $x \in X$ an element. Consider the subset

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

of the vector space $\mathbb{R}^X$. 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 $\lambda \in \mathbb{R}$ 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.

5.6. **Examples.** 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.

The following result now tells us that $U \cap \mathcal{C}(\mathbb{R})$, the set of all continuous periodic functions, is again a linear subspace.

5.7. **Lemma.** *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!).

5.8. **Example.** 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.

The property we just proved in Lemma 5.7 is very important, since it tells 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.

5.9. **Definition.** 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 $v_1, v_2, \ldots, v_n \in V$, we also write $L(v_1, v_2, \ldots, v_n)$ for $L(\{v_1, v_2, \ldots, v_n\})$.

If $L(S) = V$, we say that S *generates* V, or that S is a *generating set* for V.

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

5.10. **Example.** 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 $L(\emptyset) = \{0\}$.

5.11. **Example (Leiden).** 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 Exercise 5.20.4) that both contain v_1, v_2, v_3. So certainly we have an inclusion $L(v_1, v_2, v_3) \subset H_1 \cap H_2$. The matrix with a_1 and a_2 as two rows is already in row echelon form, so we quickly find the parametrization

$$H_1 \cap H_2 = \{x \in \mathbb{R}^n \ : \ \langle x, a_1 \rangle = \langle x, a_2 \rangle = 0\} = \{\lambda_1 v_1 + \lambda_2 v_2 \ : \ \lambda_1, \lambda_2 \in \mathbb{R}\}.$$

Check this! Since every subspace containing v_1, v_2, v_3 contains at least all vectors $\lambda_1 v_1 + \lambda_2 v_2$ with $\lambda_1, \lambda_2 \in \mathbb{R}$, we see that the set

$$\{\lambda_1 v_1 + \lambda_2 v_2 \ : \ \lambda_1, \lambda_2 \in \mathbb{R}\}$$

is contained in all subspaces that $L(S)$ is the intersection of. Therefore we also have the inclusion

$$\{\lambda_1 v_1 + \lambda_2 v_2 \ : \ \lambda_1, \lambda_2 \in \mathbb{R}\} \subset L(S)$$

and we deduce the equality $L(S) = H_1 \cap H_2$.

5.12. **Remark.** 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 5.9 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 5.11.

5.13. **Example.** 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 5.12. 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.

5.14. **Definition.** 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 subset, then an *(F-)linear combination on S* is a linear combination of *finitely many* elements of S.

5.15. **Proposition.** *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 5.12. 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. □

5.16. **Remark.** 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 5.15 states that our definition is equivalent, so from now on we can use both.

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

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

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

5.18. **Example.** For any field we can consider the power functions $f_n : x \mapsto x^n$ inside the vector space F^F of all functions from F to F. Their linear hull $L(\{f_n : n \in \mathbb{N}_0\}) \subset F^F$ is the linear subspace of *polynomial functions* from F to F, i.e, functions that are of the form

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

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

5.19. **Warning.** In Example 3.10 we defined real *polynomials* in the variable x as formal (or abstract) sums of powers x^i multiplied by a real constant a_i. These are not to be confused with the *polynomial functions* $f\colon \mathbb{R} \to \mathbb{R}$, though the difference is subtle.

As stated in Warning 4.7, though, over some other fields the difference 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.

5.20. **Exercises.**

Exercise 5.20.1. *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 5.20.2. *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 5.20.3. *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 (Leiden) 5.20.4. *Let $a \in \mathbb{R}^n$ be a vector and $b \in R$ a constant. Show that the set*

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

is a subspace of $\mathbb{R}^n$ if and only if $b = 0$.

Exercise (Leiden) 5.20.5. *Let S be any subset of $\mathbb{R}^n$. Define*

$$S^\perp = \{x \in \mathbb{R}^n \mid \forall s \in S \ : \ \langle x, s\rangle = 0\}.$$

(1) *Show that $S^\perp$ is a subspace of $\mathbb{R}^n$.*
(2) *Show that $S^\perp = L(S)^\perp$.*
(3) *Show that for any subset $T \subset S$ we have $S^\perp \subset T^\perp$.*
(4) *Show that for any subset $T \subset \mathbb{R}^n$ we have $S^\perp \cap T^\perp = (S \cup T)^\perp$.*

Exercise 5.20.6. *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 5.12, give short proofs of the following equalities of subspaces.*

(1) *$W = L(v'_1, \ldots, v'_n)$ 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),*
(2) *$W = L(v'_1, \ldots, v'_n)$ 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).*
(3) *$W = L(v'_1, \ldots, v'_n)$ 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).*

Exercise (Leiden) 5.20.7. *Let $v_1, v_2, \ldots, v_r \in \mathbb{R}^n$ be r vectors and consider the $(r \times n)$-matrix M whose rows are these vectors. Recall that the elementary row operations on a matrix were*

(1) *switch two rows,*
(2) *multiply a row by a nonzero scalar,*
(3) *add a multiple of some row to another row.*

Let M' be a matrix obtained from M by applying a sequence of elementary row operations. Show that for the rows $v'_1, v'_2, \ldots, v'_r$ of M' we have $L(v_1, \ldots, v_r) = L(v'_1, \ldots, v'_r)$.

Exercise 5.20.8. *Let U_1 and U_2 be two linear subspaces of a vector space V. Show that $U_1 \cup U_2$ is a linear subspace of V if and only if $U_1 \subset U_2$ or $U_2 \subset U_1$.*

6. Linear Independence and Dimension

This section essentially follows Chapter 3 in Jänich's book [J].

In the context of looking at linear hulls, it is a natural question whether we really need all the given vectors in order to generate their linear hull. Also (maybe in order to reduce waste...), it is interesting to consider *minimal* generating sets. These questions lead to the notions of linear independence and basis.

6.1. **Definition.** Let V be an F-vector space, $v_1, v_2, \ldots, v_n \in V$. We say that $v_1, v_2, \ldots, v_n$ are *linearly independent*, if for all $\lambda_1, \lambda_2, \ldots, \lambda_n \in F$, the equality

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

implies $\lambda_1 = \lambda_2 = \cdots = \lambda_n = 0$. ("The zero vector cannot be written as a nontrivial linear combination of $v_1, \ldots, v_n$.")

In a similar way, we can define linear independence for arbitrary subsets of V. This is equivalent to the following. $S \subset V$ is *linearly independent* if every choice of finitely many (distinct) vectors from S is linearly independent.

As a special case, the empty sequence or empty set of vectors is considered to be linearly independent.

If we want to refer to the field of scalars F, we say that the given vectors are *F-linearly independent* or *linearly independent over F*.

If $v_1, v_2, \ldots, v_n$ (resp., S) are not linearly independent, we say that they are *linearly dependent.*

6.2. **Example.** For an easy example that the field of scalars matters in the context of linear independence, consider $1, i \in \mathbb{C}$, where $\mathbb{C}$ can be considered as a real or as a complex vector space. We then have that 1 and i are $\mathbb{R}$-linearly independent (essentially by definition of $\mathbb{C}$ — $0 = 0 \cdot 1 + 0 \cdot i$, and this representation is unique), whereas they are $\mathbb{C}$-linearly dependent — $i \cdot 1 + (-1) \cdot i = 0$.

6.3. **Example.** The vectors

$$v_1 = (1, 2, 3, 4), \qquad v_2 = (5, 6, 7, 8), \qquad v_3 = (9, 10, 11, 12)$$

in $\mathbb{R}^4$ are linearly dependent, as we have $v_1 - 2v_2 + v_3 = 0$.

6.4. **Example.** Consider the vectors

$$w_1 = (1, 1, 1), \quad w_2 = (1, 2, 4), \qquad w_3 = (1, 3, 9)$$

in $\mathbb{R}^3$ and suppose we have $\lambda_1 w_1 + \lambda_2 w_2 + \lambda_3 w_3 = 0$. Then we have

$$\begin{aligned} \lambda_1 + \lambda_2 + \lambda_3 &= 0, \\ \lambda_1 + 2\lambda_2 + 3\lambda_3 &= 0, \\ \lambda_1 + 4\lambda_2 + 9\lambda_3 &= 0. \end{aligned}$$

These equations imply $\lambda_1 = \lambda_2 = \lambda_3 = 0$, so w_1, w_2, and w_3 are linearly independent.

6.5. **Example.** In $\mathcal{C}(\mathbb{R})$, the functions

$$x \longmapsto 1\,, \quad x \longmapsto \sin x\,, \quad x \longmapsto \cos x\,, \quad x \longmapsto \sin^2 x\,, \quad x \longmapsto \cos^2 x$$

are linearly dependent, since $1 - \sin^2 x - \cos^2 x = 0$ for all $x \in \mathbb{R}$.

On the other hand,

$$x \longmapsto 1\,, \quad x \longmapsto \sin x\,, \quad x \longmapsto \cos x$$

are linearly independent. To see this, assume that $\lambda + \mu \sin x + \nu \cos x = 0$ for all $x \in \mathbb{R}$. Plugging in $x = 0$, we obtain $\lambda + \nu = 0$. For $x = \pi$, we get $\lambda - \nu = 0$, which together imply $\lambda = \nu = 0$. Then taking $x = \pi/2$ shows that $\mu = 0$ as well.

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

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

for some $j \in \{1, 2, \ldots, n\}$. A similar statement holds for subsets $S \subset V$.

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

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

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

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

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

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

Then

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

so the given vectors are linearly dependent. Bearing in mind that S is linearly dependent if and only if some finite subset of S is linearly dependent, the last statement also follows. □

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

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

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

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

Proof. Exercise! □

6.8. **Example.** Take the vectors

$$\begin{aligned} v_1 &= (1, 2, 1, -1, 2, 1, 0),\\ v_2 &= (0, 1, 1, 0, -1, -2, 3),\\ v_3 &= (0, 0, 0, 3, 3, -1, 2),\\ v_4 &= (0, 0, 0, 0, 0, 6, 4) \end{aligned}$$

in $\mathbb{Q}^7$. We consider them in opposite order, so v_4, v_3, v_2, v_1. Then for each vector, the first coordinate that is nonzero (namely the sixth, fourth, second, and first coordinate respectively), is zero for all previous vectors. This implies by Remark 6.7 that no vector is a linear combination of the previous ones, so the vectors are linearly independent.

6.9. **Remark (Leiden).** In the same way as in Example 6.8, one shows that in general the nonzero rows of a matrix in row echelon form over any field F are linearly independent. See Exercise 6.56.3.

6.10. **Example (Leiden).** Suppose M is the matrix

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

which is already in row echelon form with its pivots circled. Call its rows v_1, v_2, v_3, v_4, let $U = L(v_1, v_2, v_3, v_4) \subset \mathbb{R}^7$ be the space generated by the rows, and let $U^\perp$ be the subspace of all vectors orthogonal to U, i.e.,

$$\begin{aligned} U^\perp &= \{x \in \mathbb{R}^7 \ : \ \langle v, x\rangle = 0 \text{ for all } v \in U\} \\ &= \{x \in \mathbb{R}^7 \ : \ \langle v_i, x\rangle = 0 \text{ for } i = 1,2,3,4\} \\ &= \{x \in \mathbb{R}^7 \ : \ \langle v_i, x\rangle = 0 \text{ for } i = 1,2,3\}, \end{aligned}$$

cf. Exercise 5.20.5. Suppose $x = (x_1, x_2, x_3, x_4, x_5, x_6, x_7)$ is contained in $U^\perp$. Then we have seen that whatever the coordinates are that belong to columns without a pivot, the other coordinates are uniquely determined; indeed, if we set $q = x_2$, $r = x_4$, $s = x_6$, and $t = x_7$, then we obtain $x_5 = -s - t$, $x_3 = r + 3s$ and $x_1 = -2q + r + 4s + 5t$, using the equations $\langle v_i, x\rangle = 0$ for $i = 3, 2$, and 1 respectively. This means we can write

$$x = (-2q + r + 4s + 5t, q, r + 3s, r, -s - t, s, t) = qw_1 + rw_2 + sw_3 + tw_4,$$

with

$$\begin{aligned} w_1 &= (-2, 1, 0, 0, \quad 0, 0, 0), \\ w_2 &= (\ \ 1, 0, 1, 1, \quad 0, 0, 0), \\ w_3 &= (\ \ 4, 0, 3, 0, -1, 1, 0), \\ w_4 &= (\ \ 5, 0, 0, 0, -1, 0, 1). \end{aligned}$$

Note that for each of the columns without pivot, there is exactly one w_j with 1 for the corresponding coordinate and 0 for all other coordinates belonging to a column without a pivot. This could also be used to find w_1, w_2, w_3, w_4 directly: choose any column without a pivot and set the corresponding coordinate equal to 1, set all other coordinates corresponding to columns without pivot equal to 0, and compute the remaining coordinates.

We see that for each of the vectors w_1, w_2, w_3, w_4, there is a nonzero coordinate that is zero for the other three vectors, so none of the four vectors is a linear combination of the others. From Remark 6.6 we find that w_1, w_2, w_3, w_4 are linearly independent.

6.11. **Remark (Leiden).** We have seen before how to find generators for $U^\perp$ when U is a subspace of $\mathbb{R}^n$ generated by r vectors $v_1, \ldots, v_r$. First let M be the $(r \times n)$-matrix whose rows are the vectors $v_1, \ldots, v_r$. Then use Gaussian elimination, i.e., apply elementary row operations, to bring M into a row echelon form M'. By Exercise 5.20.7, the rows of M' also generate U. As in Example 6.10, we find a vector in $U^\perp$ for each pivot-less column in M', with 1 for the corresponding coordinate and 0 for all other coordinates that correspond to a

pivot-less column. As in Example 6.10, these vectors generate $U^\perp$ and they are linearly independent by Remark 6.6. See also Lemma 12.12.

6.12. **Example.** Consider the real polynomials

$$f_1 = 1, \quad f_2 = x + 2, \quad f_3 = x^2 - 2x + 3, \quad f_4 = 2x^4 - 2x^2 + 5$$

inside the real vector space $P(\mathbb{R})$ (cf. Example 4.6 and Warning 4.7). The degree of each polynomial is higher than the degree of all the previous ones, so none of the polynomials is a linear combination of the previous ones and we conclude by Remark 6.7 that the polynomials are linearly independent.

6.13. **Example.** Over any field F, the powers $1, x, x^2, x^3, x^4, \ldots$ of x in $P(F)$ (called *monomials* , cf. Example 4.6 and Warning 4.7) are linearly independent by Remark 6.7 because none is a linear combination of the previous ones, as their degree grows.

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

Note that the elements of a basis have a specific order. If we forget about the order, we speak of *unordered bases*. (By definition these are *not* bases!)

6.15. **Definition.** Let V be a vector space. A subset $S \subset V$ is called an *unordered basis* of V if S is linearly independent, and $V = L(S)$.

6.16. **Example.** The most basic example of a basis is the *canonical basis* of F^n. This is $e_1, e_2, \ldots, e_n$, where

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

6.17. **Example.** Let X be a finite set and F a field. For each $x \in X$, we define the function $f_x \colon X \to F$ that sends x to 1 and every other element of X to 0. Then the set $\{f_x \ : \ x \in X\}$ is an unordered basis of the vector space F^X. Compare this to the previous example.

6.18. **Remark (Leiden).** Remark 6.9 can be used to find a basis of a subspace U of F^n generated by $v_1, \ldots, v_r \in F^n$; if we let M be the $r \times n$ matrix whose rows are $v_1, \ldots, v_r$, and M' is a row echelon form of M, then the rows of M' generate the space $U = L(v_1, \ldots, v_r)$ as well (see Exercise 5.20.7), so the nonzero rows of M' form a basis for U by Remark 6.9. See Proposition 12.4.

6.19. **Remark (Leiden).** Remark 6.11 shows in fact that the generators computed in the method described, form a basis for $U^\perp$.

From Remark 6.6 above, we see that a basis (or unordered basis) of V is a *minimal* generating set of V, in the sense that we cannot leave out some element and still have a generating set.

What is special about a basis among generating sets?

6.20. **Lemma.** *If $v_1, v_2, \ldots, v_n$ is a basis of the F-vector space V, then for every $v \in V$, there are* unique *scalars $\lambda_1, \lambda_2, \ldots, \lambda_n \in F$ such that*

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

Proof. The *existence* of $(\lambda_1, \lambda_2, \ldots, \lambda_n) \in F^n$ such that

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

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

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

$$v = \mu_1 v_1 + \mu_2 v_2 + \cdots + \mu_n v_n \,.$$

Taking the difference, we obtain

$$0 = (\lambda_1 - \mu_1) v_1 + (\lambda_2 - \mu_2) v_2 + \cdots + (\lambda_n - \mu_n) v_n \,.$$

Since $v_1, v_2, \ldots, v_n$ are linearly independent, it follows that

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

i.e., $(\lambda_1, \lambda_2, \ldots, \lambda_n) = (\mu_1, \mu_2, \ldots, \mu_n)$. □

The converse of Lemma 6.20 holds as well.

6.21. **Lemma.** *Let V be a vector space over a field F, with a sequence $v_1, v_2, \ldots, v_n$ of elements in V. If for every $v \in V$, there are* unique *scalars $\lambda_1, \lambda_2, \ldots, \lambda_n \in F$ such that*

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

then $v_1, v_2, \ldots, v_n$ is a basis of the vector space V.

Proof. Exercise! □

6.22. **Exercise.** Formulate and prove the corresponding statements for unordered bases.

In a precise sense (to be discussed in detail later), knowing a basis $v_1, v_2, \ldots, v_n$ of a vector space V allows us to express everything in V in terms of the standard vector space F^n.

Since we seem to know "everything" about a vector space as soon as we know a basis, it makes sense to use bases to measure the "size" of vector spaces. In order for this to make sense, we need to know that any two bases of a given vector space have the same size. The key to this (and many other important results) is the following.

6.23. **Basis Extension Theorem.** *Let V be an F-vector space, and let $v_1, \dots, v_r$, $w_1, \dots, w_s \in V$. If $v_1, \dots, v_r$ are linearly independent and V is generated by $v_1, \dots, v_r, w_1, \dots, w_s$, then by adding suitably chosen vectors from $w_1, \dots, w_s$, we can extend $v_1, \dots, v_r$ to a basis of V.*

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

The Basis Extension Theorem implies another important statement.

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

This says that we can trade any vector of our choice in the first basis for a vector in the second basis in such a way as to still have a basis.

We will postpone the proofs and first look at some consequences.

6.25. **Theorem.** *If $v_1, v_2, \dots, v_n$ and $w_1, w_2, \dots, w_m$ are two bases of a vector space V, then $n = m$.*

Proof. Assume, without loss of generality, that $n > m$. By repeatedly applying the Exchange Lemma, we can successively replace v_1, v_2, …, v_n by some w_j and still have a basis. Since there are more v's than w's, the resulting sequence must have repetitions and therefore cannot be linearly independent, contradiction. □

This implies that the following definition makes sense.

6.26. **Definition.** If the vector space V has a basis $v_1, v_2, \dots, v_n$, then $n \geq 0$ is called the *dimension* of V, written $n = \dim V = \dim_F V$.

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

6.28. **Example.** The canonical basis of F^n has length n, so $\dim F^n = n$.

6.29. **Theorem.** *Let V be a vector space, $\dim V = n$, and $v_1, \dots, v_m \in V$ a sequence of vectors. Then the following statements hold.*

(1) *If $m > n$, then $v_1, \dots, v_m$ are linearly dependent.*
(2) *If $m < n$, then $v_1, \dots, v_m$ do not generate V.*
(3) *If $m = n$, then $v_1, \dots, v_m$ are linearly independent if and only if they generate V.*

Proof. Let $w_1, \dots, w_n$ be a basis of V. For (1), assume that $v_1, \dots, v_m$ were linearly *in*dependent. Then by the Basis Extension Theorem (note that

$$V = L(w_1, \dots, w_n) = L(v_1, \dots, v_m, w_1, \dots, w_n) \quad),$$

we could extend $v_1, \dots, v_m$ to a basis of V by adding some vectors from $w_1, \dots, w_n$. Since $m > n$, the resulting basis would be too long, contradiction. We leave the proofs of (2) and (3) as an exercise. □

Note that this theorem is a quite strong existence statement: part (1) guarantees the *existence* of a nontrivial linear relation, i.e., a nontrivial linear combination that is zero, among the given vectors without the need to do any computation. This is very useful in many applications. On the other hand, it is quite a different matter to actually *find* such a relation: the proof is non-constructive (as is usually the case with proofs by contradiction), and we usually need some computational method to exhibit a relation.

6.30. **Remark (Leiden).** If you bring an $m \times n$ matrix M over a field F in row echelon form using elementary row operations (Gaussian elimination), then each row in the row echelon form is a linear combination of the rows of M. By keeping track of the specific linear combinations that make up the rows during the process, we find one linear relation among the rows of M for each row of zeroes in the row echelon form.

6.31. **Example (Leiden).** Consider the matrix

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

whose rows we will call v_1, v_2, and v_3, respectively. After three obvious row operations, its row echelon form is found to be

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

where the rows are v_1, $v_2 - 2v_1$, and $5v_1 - 3v_2 + v_3$ respectively. We have therefore found the linear relation $5v_1 - 3v_2 + v_3 = 0$.

Part (1) of Theorem 6.29 tells us that in a vector space of (finite) dimension n, there is an upper bound (namely, n) for the length of a linearly independent sequence of vectors. We can use this to show that there are vector spaces that do not have dimension n for any integer $n \geq 0$.

6.32. **Example.** The vector space $P(F)$ of all *polynomials* with coefficients in F contains the monomials $1, x, x^2, x^3, x^4, \ldots$, which are linearly independent, see Example 6.13. This means that we can find arbitrarily many linearly independent elements in $P(F)$, so $P(F)$ can not have a finite basis by Theorem 6.29(1).

Note that since $P(F) = L(\{x^n : n \in \mathbb{N}_0\})$, we have shown that $\{x^n : n \in \mathbb{N}_0\}$ is an unordered basis of $P(F)$.

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

6.33. **Example.** Let us consider again the linear subspace of *polynomial functions* in $\mathcal{C}(\mathbb{R})$ (the vector space of continuous functions on $\mathbb{R}$), compare Example 5.17. Let us call this space P:

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

Denote as before by f_n the nth power function: $f_n(x) = x^n$. I claim that $\{f_0, f_1, f_2, \ldots\}$ is linearly independent. Recall that this means that the only way of writing zero (i.e., the zero function) as a *finite* linear combination of the f_j is with all coefficients equal to zero. If we let n be the largest number such that

f_n occurs in the linear combination, then it is clear that we can write the linear combination as

$$\lambda_0 f_0 + \lambda_1 f_1 + \cdots + \lambda_n f_n = 0\,.$$

We have to show that this is only possible when $\lambda_0 = \lambda_1 = \cdots = \lambda_n = 0$.

Note that our assumption means that

$$\lambda_n x^n + \cdots + \lambda_1 x + \lambda_0 = 0 \qquad \text{for all } x \in \mathbb{R}.$$

There are various ways to proceed from here. For example, we can make use of the fact that a polynomial of degree $n \geq 0$ can have at most n zeros in $\mathbb{R}$. Since there are infinitely many real numbers, the polynomial above has infinitely many zeros, hence it must be the zero polynomial (which does not have a well-defined degree).

Another possibility is to use *induction* on n (which, by the way, is implicit in the proof above: it is used in proving the statement on zeros of polynomials). Let us do this in detail. The *claim* we want to prove is

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

We now have to establish the *induction base*: the claim holds for $n = 0$. This is easy — let $\lambda_0 \in \mathbb{R}$ and assume that for all $x \in \mathbb{R}$, $\lambda_0 = 0$ (the function is constant here: it does not depend on x). Since there are real numbers, this implies $\lambda_0 = 0$.

Next, and this is usually the hard part, we have to do the *induction step.* We assume that the claim holds for a given n (this is the *induction hypothesis*) and deduce that it then also holds for $n+1$. To prove the statement for $n+1$, we have to consider coefficients $\lambda_0, \ldots, \lambda_{n+1} \in \mathbb{R}$ such that for all $x \in \mathbb{R}$,

$$f(x) = \lambda_{n+1} x^{n+1} + \lambda_n x^n + \cdots + \lambda_1 x + \lambda_0 = 0\,.$$

Now we want to use the induction hypothesis, so we have to reduce this to a statement involving a polynomial of degree at most n. One way of doing that is to borrow some knowledge from Analysis about differentiation. This tells us that the derivative of f is zero again, and that it is a polynomial function of degree $\leq n$:

$$0 = f'(x) = (n+1)\lambda_{n+1} x^n + n\lambda_n x^{n-1} + \cdots + \lambda_1\,.$$

Now we can apply the induction hypothesis to this polynomial function; it tells us that $(n+1)\lambda_{n+1} = n\lambda_n = \cdots = \lambda_1 = 0$, hence $\lambda_1 = \cdots = \lambda_n = \lambda_{n+1} = 0$. So $f(x) = \lambda_0$ is in fact constant, which finally implies $\lambda_0 = 0$ as well (by our reasoning for the induction base).

This completes the induction step and therefore the whole proof.

Note that since $P = L(\{f_n : n \in \mathbb{N}_0\})$, we have shown that $\{f_n : n \in \mathbb{N}_0\}$ is an unordered basis of P.

So we see that P cannot have a finite basis, since we can find arbitrarily many linearly independent elements. This motivates the following definition.

6.34. **Definition.** If a vector space V does not have a finite basis, then V is said to be *infinite-dimensional*, and we write $\dim V = \infty$.

In particular, we see that $\dim P = \infty$. Since P is a subspace of $\mathbb{R}^{\mathbb{R}}$, by Example we can also find arbitrarily many linearly independent functions in $\mathbb{R}^{\mathbb{R}}$, namely the polynomial functions, so we also have $\dim \mathbb{R}^{\mathbb{R}} = \infty$.

6.35. **Warning.** In Examples and we actually found infinite unordered bases for $P(F)$ and $P \subset \mathbb{R}^{\mathbb{R}}$, but for example for $\mathbb{R}^{\mathbb{R}}$, it is a priori not at all clear that there even exists a subset $S \subset \mathbb{R}^{\mathbb{R}}$ that is linearly independent and generates the whole vector space $\mathbb{R}^{\mathbb{R}}$.

Although in Section 7 we will see that indeed all vector spaces do magically turn out to have some unordered basis, by definition the claim $\dim V = \infty$ only means that there is no finite basis, and does not directly state that there would exist an infinite basis.

The following result shows that our intuition that dimension is a measure for the 'size' of a vector space is not too far off: larger spaces have larger dimension.

6.36. **Lemma.** *Let U be a linear subspace of the vector space V. Then we have $\dim U \leq \dim V$. If $\dim V$ is finite, then we have equality if and only if $U = V$.*

Here we use the usual convention that $n < \infty$ for $n \in \mathbb{N}_0$.

Note that in the case that $\dim V$ is finite, the statement also asserts the existence of a (finite) basis of U.

Proof. There is nothing to show if $\dim V = \infty$. So let us assume that V has a basis $v_1, \ldots, v_n$. If $u_1, \ldots, u_m \in U$ are linearly independent, then $m \leq n$ by Thm. 6.29. Hence there is a sequence $u_1, \ldots, u_m$ of linearly independent vectors in U of maximal length m (and $m \leq n$). We claim that $u_1, \ldots, u_m$ is in fact a basis of U. The first claim follows, since then $\dim U = m \leq n = \dim V$.

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

To prove the second part, first note that $\dim U < \dim V$ implies $U \subsetneq V$ (if $U = V$, a basis of U would also be a basis of V, but it is too short for that by Thm. 6.25). On the other hand, assume $U \subsetneq V$, and consider a basis of U. It can be extended to a basis of V by the Basis Extension Theorem 6.23. Since it does not generate V, at least one element has to be added, which implies $\dim U < \dim V$. □

6.37. **Examples.** Since

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

all these spaces are infinite-dimensional.

Although the vector space of real polynomial functions is infinite dimensional, the following exercise states over finite fields this is not the case.

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

$$P_F = L_F(\{f_0, f_1, f_2, \ldots\})$$

where $f_n : x \mapsto x^n$ (for $x \in F$). Show that $\dim_F P_F$ is finite.

We have seen that the intersection of linear subspaces is again a linear subspace, but the union usually is not, see Example 5.8. 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.

6.39. **Definition.** 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.

6.40. **Lemma.** *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 5.12 (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. □

6.41. **Example.** The union $U = U_1 \cup U_2$ of Example 5.8 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 $U_1 + U_2 = \mathbb{R}^2$.

6.42. **Example.** 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

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

6.43. **Remark.** 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. □

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

6.44. **Theorem.** *Let U_1 and U_2 be linear subspaces of a vector space V. Then*

$$\dim(U_1 + U_2) + \dim(U_1 \cap U_2) = \dim U_1 + \dim U_2 \,.$$

Proof. First note that the statement is trivially true when U_1 or U_2 is infinite-dimensional, since then both sides are ∞. So we can assume that U_1 and U_2 are both finite-dimensional.

For the proof, we use the Basis Extension Theorem 6.23 again. Since $U_1 \cap U_2 \subset U_1$ and U_1 is finite-dimensional, we know by Lemma 6.36 that $U_1 \cap U_2$ is also finite-dimensional. Let $v_1, \dots, v_r$ be a basis of $U_1 \cap U_2$. Using the Basis Extension Theorem, we can extend it on the one hand to a basis $v_1, \dots, v_r, w_1, \dots, w_s$ of U_1 and on the other hand to a basis $v_1, \dots, v_r, z_1, \dots, z_t$ of U_2. I claim that then $v_1, \dots, v_r, w_1, \dots, w_s, z_1, \dots, z_t$ is a basis of $U_1 + U_2$. It is clear that these vectors generate $U_1 + U_2$ (since they are obtained by putting generating sets of U_1 and of U_2 together, see Remark 6.43). So it remains to show that they are linearly independent. Consider a general linear combination

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

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

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

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

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

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

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

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

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

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

6.45. **Remark and Exercise.** Note the analogy with the formula

$$\#(X \cup Y) + \#(X \cap Y) = \#X + \#Y$$

for the number of elements in a set. However, there is no analogue of the corresponding formula for three sets:

$$\#(X \cup Y \cup Z) = \#X + \#Y + \#Z - \#(X \cap Y) - \#(X \cap Z) - \#(Y \cap Z) + \#(X \cap Y \cap Z) \,.$$

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

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

Note that if $U_1 \cap U_2 = \{0\}$, then we simply have $\dim(U_1 + U_2) = \dim U_1 + \dim U_2$ (and conversely). So this is an especially nice case; it motivates the following definition.

6.46. **Definition.** 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 $U_1 + U_2 = V$.

Note that by the above, we then have $\dim U_1 + \dim U_2 = \dim V$.

6.47. **Proposition (Leiden).** Let U be a subspace of $\mathbb{R}^n$. Then $U^\perp$ is complmentary to U.

Because of Proposition 6.47, we call $U^\perp$ the *orthogonal complement* of U in $\mathbb{R}^n$. To prove Proposition 6.47, we use the following two lemmas. The proof of the first uses row echelon forms; in Remark 11.10 we will see a nicer proof that does not depend on that.

6.48. **Lemma (Leiden).** For any subspace $U \subset \mathbb{R}^n$ we have $\dim U + \dim U^\perp = n$.

Proof. Let $v_1, \ldots, v_s$ be generators of U and let M be the $s \times n$ matrix whose rows are $v_1, \ldots, v_s$. Let M' be a matrix in row echelon form associated to M, and let r denote the number of nonzero rows of M'. Then we have $\dim U = r$ by Remark 6.18. Since each nonzero row contains exactly one pivot, there are $n - r$ pivot-less columns in M', so by Remark 6.19 we have $\dim U^\perp = n - r$ and the equality follows. □

6.49. **Lemma (Leiden).** Let U_1 and U_2 be subspaces of $\mathbb{R}^n$ satisfying $U_1 \cap U_2 = \{0\}$ and $\dim U_1 + \dim U_2 \geq n$. Then U_1 and U_2 are complementary subspaces.

Proof. From the Dimension Formula 6.44 we get

$$n \leq \dim U_1 + \dim U_2 = \dim(U_1 \cap U_2) + \dim(U_1 + U_2) = \dim(U_1 + U_2) \leq \dim \mathbb{R}^n = n,$$

so all inequalities are in fact equalities and we have $U_1 + U_2 = \mathbb{R}^n$, as required. □

Proof of Proposition 6.47. Suppose $x \in U \cap U^\perp$. Then we have $\langle x, x \rangle = 0$, so $x = 0$ and thus $U \cap U^\perp = \{0\}$. The statement now follows immediately from Lemmas 6.48 and 6.49. □

6.50. **Proposition (Leiden).** Let $U_1, U_2 \subset \mathbb{R}^n$ be subspaces. Then we have $(U_i^\perp)^\perp = U_i$ and

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

Proof. Exercise. □

6.51. **Remark (Leiden).** For a subspace $U \subset \mathbb{R}^n$, we can think of $U^\perp$ as the subspace of equations for U, as for each $v \in U^\perp$ we have $\langle v, x \rangle = 0$ for all $x \in U$. From this point of view, Proposition 6.50 states that U is itself the subspace of equations for $U^\perp$ and the subspace of equations for the intersection $U_1 \cap U_2$ is the sum of the subspaces of equations for U_1 and U_2. See also the more general statement Prop. 12.15 over arbitrary fields and Example 12.16.

We now know that in $\mathbb{R}^n$ every subspace has a complementary space, namely its orthogonal complement. The first part of the following Lemma states that every subspace of any finite-dimensional vector space V over any field has a complementary space. In fact, the methods of Section 7 will show that the vector space V does not need to be finite dimensional.

6.52. **Lemma.** *Let V be a vector space.*

(1) *If V is finite-dimensional and $U \subset V$ is a linear subspace, then there is a linear subspace $U' \subset V$ that is complementary to U.*
(2) *If U and U' are complementary linear subspaces of V, then for every $v \in V$ there are unique $u \in U$, $u' \in U'$ such that $v = u + u'$.*

Proof.

(1) In this case, U is finite-dimensional, with basis $u_1, \dots, u_m$ (say). By the Basis Extension Theorem 6.23, we can extend this to a basis $u_1, \dots, u_m, v_1, \dots, v_n$ of V. Let $U' = L(v_1, \dots, v_n)$. Then we clearly have $V = U + U'$. But we also have $U \cap U' = \{0\}$: if $v \in U \cap U'$, then
$$v = \lambda_1 u_1 + \dots + \lambda_m u_m = \mu_1 v_1 + \dots + \mu_n v_n \,,$$
but $u_1, \dots, u_m v_1, \dots, v_n$ are linearly independent, so all the λs and μs must be zero, hence $v = 0$.
(2) 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 $u = w$, $u' = w'$.

□

6.53. **Example.** Given $U \subset V$, there usually are many complementary subspaces. For example, consider $V = \mathbb{R}^2$ and $U = \{(x, 0) : x \in \mathbb{R}\}$. What are its complementary subspaces U'? We have $\dim V = 2$ and $\dim U = 1$, so we must have $\dim U' = 1$ as well. Let $u' = (x', y')$ be a basis of U'. Then $y' \neq 0$ (otherwise $0 \neq u' \in U \cap U'$). Then we can scale u' by $1/y'$ (replacing u', x', y' by $\frac{1}{y'}u', x'/y', 1$, respectively) to obtain a basis of the form $u' = (x', 1)$, and $U' = L(u')$ then is a complementary subspace for every $x' \in \mathbb{R}$ — note that $(x, y) = (x - yx', 0) + y(x', 1) \in U + U'$.

Most of the previous results depend on the still unproven Basis Extension Theorem 6.23 and the Exchange Lemma 6.24. We will now prove those.

6.54. **Proof of Basis Extension Theorem.** Here we prove the Basis Extension Theorem. A precise version of the statement is as follows.

Let V be a vector space, and let $v_1, \dots, v_r, w_1, \dots, w_s \in V$ be vectors such that $v_1, \dots, v_r$ are linearly independent and $V = L(v_1, \dots, v_r, w_1, \dots, w_s)$. Then there is $t \in \mathbb{N}_0$ and indices $i_1, \dots, i_t \in \{1, \dots, s\}$ such that $v_1, \dots, v_r, w_{i_1}, \dots, w_{i_t}$ is a basis of V.

Make sure you understand how we have formalized the notion of "suitably chosen vectors from $w_1, \dots, w_s$!"

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

$v_1, \dots, v_r, w_{i_1}, \dots, w_{i_t}, w_j$ would be linearly independent, which would contradict our choice of a linearly independent sequence of maximal length. So w_j must be a linear combination of our vectors, and the theorem is proved.

Here is an alternative proof, using induction on the number s of vectors w_j.

The base case is $s = 0$. In this case, the assumptions tell us that $v_1, \dots, v_r$ are linearly independent and generate V, so we have a basis.

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

6.55. **Proof of Exchange Lemma.** Now we prove the Exchange Lemma 6.24. Recall the statement.

If $v_1, v_2, \dots, v_n$ and $w_1, w_2, \dots, w_m$ are two bases of a vector space V, then for each $i \in \{1, \dots, n\}$ there is some $j \in \{1, \dots, m\}$ such that $v_1, \dots, v_{i-1}, w_j, v_{i+1}, \dots, v_n$ is again a basis of V.

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

6.56. **Exercises.**

Exercise 6.56.1. *Which of the following sequences of vectors in $\mathbb{R}^3$ are linearly independent?*

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

Exercise 6.56.2. *Prove Remark 6.7.*

Exercise (Leiden) 6.56.3. *Prove Remark 6.9, i.e., show that the nonzero rows of a matrix in row echelon form are linearly independent.*

Exercise 6.56.4. *Determine a basis for the subspaces of $\mathbb{R}^n$ generated by*

(1) $v_1 = (1,3), v_2 = (2,1), v_3 = (1,1)$,
(2) $v_1 = (1,3,1), v_2 = (2,1,2), v_3 = (1,1,1)$,
(3) $v_1 = (1,3,1), v_2 = (3,1,3), v_3 = (1,1,1)$,
(4) $v_1 = (1,2,3), v_2 = (4,5,6), v_3 = (7,8,9)$,

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

Exercise 6.56.5. *Are the polynomials* $3, x-1, x^2-3x+2, x^4-3x+13, x^7-x+14$ *linearly independent?*

Exercise 6.56.6. *Are the polynomials* $x^7-2x+1, 5x^2, 2x^4-5x^3, x, x^6-3x$ *linearly independent?*

Exercise 6.56.7. *Are the vectors*

$$\begin{aligned} v_1 &= (1,4,2,3,5),\\ v_2 &= (-1,7,2,3,6),\\ v_3 &= (4,2,3,-3,4),\\ v_4 &= (2,-3,1,4,2),\\ v_5 &= (6,5,3,-2,-4),\\ v_6 &= (1,-7,3,2,5) \end{aligned}$$

in $\mathbb{R}^5$ *linearly independent? (Hint: do not start a huge computation)*

Exercise 6.56.8. *Prove Lemma 6.21.*

Exercise 6.56.9. *Do Exercise 6.22.*

Exercise 6.56.10. *Prove part (2) and (3) of Theorem 6.29.*

Exercise 6.56.11. *In Example 6.17, is it necessary that* X *is finite?*

Exercise 6.56.12. *Do exercise 6.38 from the text.*

Exercise 6.56.13. *Prove Remark 6.43.*

Exercise 6.56.14. *Formulate and prove a version of Remark 6.43 for arbitrary collections* $\{S_i\}_{i\in I}$ *of subsets of* V.

Exercise 6.56.15. *Do exercise 6.45 from the text.*

Exercise (Leiden) 6.56.16. *Use Exercise 5.20.5 and Lemma 6.48 to prove Proposition 6.50.*

Exercise 6.56.17. *Let* V *be a finite-dimensional vector space and* $S \subset V$ *a subset that generates* V. *Show that there is a finite subset of* S *that generates* V.

7. Digression: Infinite-Dimensional Vector Spaces and Zorn's Lemma

We have seen that a vector space V such that the number of linearly independent vectors in V is bounded has a (finite) basis and so has finite dimension. What can we say about the existence of a basis in an *infinite-dimensional* vector space?

We have seen an example of an infinite-dimensional vector space that has a basis: this was the space of polynomial functions on $\mathbb{R}$, with basis given by the monomials $x \mapsto x^n$, $n \in \mathbb{N}_0$.

On the other hand, you would be very hard put to write down a basis for (say) $\mathcal{C}(\mathbb{R})$, or also a basis for $\mathbb{R}$ as a $\mathbb{Q}$-vector space.

In order to prove the existence of a basis and other related results, we would need an 'infinite' version of the Basis Extension Theorem.

7.1. **General Basis Extension Theorem.** *Let V be a vector space, $X, Y \subset V$ two subsets such that X is linearly independent and $V = L(X \cup Y)$. Then there is a subset $Z \subset Y$ such that $X \cup Z$ is a basis of V.*

Now, how can we prove such a statement? Recall that in the proof of the finite version (if we formulate it for sets instead of sequences), we took a *maximal* subset Z of Y such that $X \cup Z$ is linearly independent and showed that $X \cup Z$ is already a basis. This last step will work here in the same way: assume that Z is maximal as above, then for every $y \in Y \setminus Z$, $X \cup Z \cup \{y\}$ is linearly dependent, and so $y \in L(X \cup Z)$. This implies that $V = L(X \cup Y) \subset L(X \cup Z)$, so $X \cup Z$ generates V and is therefore a basis.

However, the key point is the *existence* of a maximal set Z with the required property. Note that if $\mathcal{S}$ is an arbitrary set of subsets of some set, $\mathcal{S}$ need not necessarily have maximal elements. For example, $\mathcal{S}$ could be empty. Or consider the set of all finite subsets of $\mathbb{N}$. So we need some extra condition to ensure the existence of maximal elements. (Of course, when $\mathcal{S}$ is finite (and nonempty), then there is no problem — we can just take a set of maximal size.)

This condition is formulated in terms of *chains.*

7.2. **Definition.** Let X be a set, and let $\mathcal{S}$ be a set of subsets of X. A subset $\mathcal{C} \subset \mathcal{S}$ is called a *chain* if all elements of $\mathcal{C}$ are comparable, i.e., if for all $U, V \in \mathcal{C}$, we have $U \subset V$ or $V \subset U$. (Note that this is trivially true when $\mathcal{C}$ is empty.)

7.3. **Remark.** The notion of 'chain' (as well as Zorn's Lemma below) applies more generally to (partially) ordered sets: a chain then is a subset that is totally ordered.

Now a statement of the kind we need is the following.

7.4. **Zorn's Lemma.** *Let X be a set, and let $\mathcal{S}$ be a collection of subsets of X. If for every chain $\mathcal{C} \subset \mathcal{S}$, there is a set $U \in \mathcal{S}$ such that $Z \subset U$ for all $Z \in \mathcal{C}$, then $\mathcal{S}$ has a maximal element.*

Note that the condition, when applied to the empty chain, ensures that $\mathcal{S} \neq \emptyset$. Also note that there can be more than one maximal element in $\mathcal{S}$.

Let us see how we can apply this result to our situation. The set $\mathcal{S}$ we want to consider is the set of all subsets $Z \subset Y$ such that $X \cup Z$ is linearly independent. We have to verify the assumption on chains. So let $\mathcal{C} \subset \mathcal{S}$ be a chain. We have to exhibit a set $U \in \mathcal{S}$ containing all the elements of $\mathcal{C}$. In such a situation, our first guess is to try $U = \bigcup \mathcal{C}$ (the union of all sets in $\mathcal{C}$); usually it works. In our case, we have to show that this U has the property that $X \cup U$ is linearly independent. Assume it is not. Then there is a *finite* non-trivial linear combination of elements of $X \cup U$ that gives the zero vector. This linear combination will only involve finitely many elements of U, which come from finitely many sets $Z \in \mathcal{C}$. Since $\mathcal{C}$ is a chain, there is a maximal set Z_{max} among these, and our nontrivial linear combination only involves elements from $X \cup Z_{max}$. But Z_{max} is in $\mathcal{S}$, and so $X \cup Z_{max}$ is linearly independent, a contradiction. Therefore our assumption must be false, and $X \cup U$ must be linearly independent.

7.5. **Exercise.** Use Zorn's Lemma to prove that for every subset X of a vector space V such that X contains the zero vector, there is a maximal linear subspace of V contained in X.

7.6. **Discussion.** Based on Zorn's Lemma, we can prove the general Basis Extension Theorem. In particular, this shows that every vector space must have a basis (take $X = \emptyset$ and $Y = V$). However, Zorn's Lemma is an extremely inconstructive result — it does not give us any information on how to find a maximal element. And in fact, nobody has ever been able to 'write down' (or explicitly construct) a $\mathbb{Q}$-basis of $\mathbb{R}$, say. Still, such bases must exist.

The next question then is, how does one prove Zorn's Lemma? It turns out that it is equivalent (given the more 'harmless' axioms of set theory) to the *Axiom of Choice*, which states the following.

Let I be a set, and let $(X_i)_{i\in I}$ be a family of nonempty sets indexed by I. Then there is a 'choice function' $f : I \to \bigcup_{i\in I} X_i$ such that $f(i) \in X_i$ for all $i \in I$.

In other words, if all the X_i are nonempty, then the product $\prod_{i\in I} X_i$ of these sets is also nonempty. This looks like a natural property, however it has consequences like the existence of $\mathbb{Q}$-bases of $\mathbb{R}$ which are not so intuitive any more. Also, as it turned out, the Axiom of Choice is *independent* from the other axioms of set theory: it is not implied by them.

For some time, there was some discussion among mathematicians as to whether the use of the Axiom of Choice (and therefore, of Zorn's Lemma) should be allowed or forbidden (because of its inconstructive character). By now, a pragmatic viewpoint has been adapted by almost everybody: use it when you need it. For example, interesting parts of analysis and algebra need the Axiom of Choice, and mathematics would be quite a bit poorer without it.

Finally, a historical remark: Zorn's Lemma was first discovered by Kazimierz Kuratowski in 1922 (and rediscovered by Max Zorn about a dozen years later), so it is not really appropriately named. In fact, when I was a student, one of my professors told us that he talked to Zorn at some occasion, who said that he was not at all happy that the statement was carrying his name...

8. 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.

8.1. **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, $\mathrm{im}(f)$. 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 V under f. Note that when f is bijective, there are two meanings of $f^{-1}(B)$ — one as just defined, and one as $g(B)$ where g is the inverse map f^{-1}. Fortunately, both meanings agree (Exercise), and there is no danger of confusion.

Maps can be *composed:* if $f : X \to Y$ and $g : Y \to Z$, then we can define a map $X \to Z$ that sends $x \in X$ to $g(f(x)) \in Z$. This map is denoted by $g \circ f$ ("g after f") — keep in mind that it is f that is applied first!

Composition of maps is associative: if $f : X \to Y$, $g : Y \to Z$ and $h : Z \to W$, then $(h \circ g) \circ f = h \circ (g \circ f)$. Every set X has a special map, the *identity map* $\mathrm{id}_X : X \to X, x \mapsto x$. It acts as a neutral element under composition: for $f : X \to Y$, we have $f \circ \mathrm{id}_X = f = \mathrm{id}_Y \circ f$. If $f : X \to Y$ is bijective, then its inverse satisfies $f \circ f^{-1} = \mathrm{id}_Y$ and $f^{-1} \circ f = \mathrm{id}_X$.

When talking about several sets and maps between them, we often picture them in a *diagram* like the following.

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ {\scriptstyle g}\downarrow & & \downarrow{\scriptstyle g'} \\ U & \xrightarrow{f'} & V \end{array} \qquad\qquad \begin{array}{ccc} X & & \\ {\scriptstyle f}\downarrow & \searrow^{h} & \\ Y & \xrightarrow{g} & Z \end{array}$$

We call such a diagram *commutative* if all possible ways of going from one set to another lead to the same result. For the left diagram, this means that $g' \circ f = f' \circ g$, for the right diagram, this means that $h = g \circ f$.

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

8.2. **Definition.** 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.

8.3. **Lemma.** *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.

□

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

8.4. **Definition.** Let $f : V \to W$ be a linear map. Then the *kernel* of f is defined to be
$$\ker(f) = \{v \in V : f(v) = 0\}\,.$$

8.5. **Lemma.** *Let $f : V \to W$ be a linear map.*

(1) $\ker(f) \subset V$ *is a linear subspace. More generally, if $U \subset W$ is a linear subspace, then $f^{-1}(U) \subset V$ is again a linear subspace; it contains* $\ker(f)$.
(2) $\mathrm{im}(f) \subset W$ *is a linear subspace. More generally, if $U \subset V$ is a linear subspace, then $f(U) \subset W$ is again a linear subspace; it is contained in* $\mathrm{im}(f)$.
(3) *f is injective if and only if* $\ker(f) = \{0\}$.

Proof.

(1) We have to check the three properties of subspaces for $\ker(f)$. By the previous remark, $f(0) = 0$, so $0 \in \ker(f)$. Now let $v_1, v_2 \in \ker(f)$. Then $f(v_1) = f(v_2) = 0$, so $f(v_1 + v_2) = f(v_1) + f(v_2) = 0 + 0 = 0$, and $v_1 + v_2 \in \ker(f)$. Finally, let λ be a scalar and $v \in \ker(f)$. Then $f(v) = 0$, so $f(\lambda v) = \lambda f(v) = \lambda \cdot 0 = 0$, and $\lambda v \in \ker(f)$.
The more general statement is left as an exercise.

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

The more general statement is proved in the same way.

(3) If f is injective, then there can be only one element of V that is mapped to $0 \in W$, and since we know that $f(0) = 0$, it follows that $\ker(f) = \{0\}$.

Now assume that $\ker(f) = \{0\}$, and let $v_1, v_2 \in V$ such that $f(v_1) = f(v_2)$. Then $f(v_1 - v_2) = f(v_1) - f(v_2) = 0$, so $v_1 - v_2 \in \ker(f)$. By our assumption, this means that $v_1 - v_2 = 0$, hence $v_1 = v_2$.

□

8.6. **Remark.** 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.

8.7. **Examples.**

(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* $\mathrm{pr}_j : F^n \to F$, $(x_1, \ldots, x_n) \mapsto x_j$ 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) 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 $D \circ I_a = \mathrm{id}_P$).

(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!).

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

8.8. **Lemma.** *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

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

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

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

□

Now the next question is, how do we specify a general linear map? It turns out that it suffices to specify the images of the elements of a basis. If our vector spaces are finite-dimensional, this means that only a finite amount of information is necessary (if we consider elements of the field of scalars as units of information).

8.9. **Theorem.** *Let V and W be two F-vector spaces. Let $v_1, \dots, v_n$ be a basis of V, and let $w_1, \dots, w_n \in W$. Then there is a unique linear map $f : V \to W$ such that $f(v_j) = w_j$ for all $j \in \{1, \dots, n\}$.*

More generally, let $B \subset V$ be a basis, and let $\phi : B \to W$ be a map. Then there is a unique linear map $f : V \to W$ such that $f|_B = \phi$ (i.e., $f(b) = \phi(b)$ for all $b \in B$).

Proof. The statement has two parts: *existence* and *uniqueness*. In many cases, it is a good idea to prove uniqueness first, since this usually tells us how to construct the object we are looking for, thus helping with the existence proof. So let us look at uniqueness now.

We show that there is only one way to define a linear map f such that $f(v_j) = w_j$ for all j. Let $v \in V$ be arbitrary. Then v is a linear combination on the basis:

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

If f is to be linear, we must then have

$$f(v) = \lambda_1 f(v_1) + \lambda_2 f(v_2) + \cdots + \lambda_n f(v_n) = \lambda_1 w_1 + \lambda_2 w_2 + \cdots + \lambda_n w_n \,,$$

which fixes $f(v)$. So there is only one possible choice for f.

To show existence, it suffices to prove that f as defined above is indeed linear. Note that f is well-defined, since every $v \in V$ is given by a *unique* linear combination of the v_j, see Lemma 6.20. Let $v, v' \in V$ with

$$\begin{aligned} v &= \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n \\ v' &= \lambda'_1 v_1 + \lambda'_2 v_2 + \cdots + \lambda'_n v_n \,, \quad \text{so} \\ v + v' &= (\lambda_1 + \lambda'_1) v_1 + (\lambda_2 + \lambda'_2) v_2 + \cdots + (\lambda_n + \lambda'_n) v_n \end{aligned}$$

then

$$\begin{aligned} f(v + v') &= (\lambda_1 + \lambda'_1) w_1 + (\lambda_2 + \lambda'_2) w_2 + \cdots + (\lambda_n + \lambda'_n) w_n \\ &= (\lambda_1 w_1 + \lambda_2 w_2 + \cdots + \lambda_n w_n) + (\lambda'_1 w_1 + \lambda'_2 w_2 + \cdots + \lambda'_n w_n) \\ &= f(v) + f(v') \end{aligned}$$

and similarly for $f(\lambda v) = \lambda f(v)$.

The version with basis sets is proved in the same way. □

We can use the images of basis vectors to characterize injective and surjective linear maps.

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

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

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

This leads to an important fact: essentially ('up to isomorphism'), there is only *one* F-vector space of any given finite dimension n.

8.11. **Corollary.** *If V and W are two F-vector spaces of the same finite dimension n, then V and W are isomorphic. In particular, if V is an F-vector space of dimension $n < \infty$, then V is isomorphic to F^n: $V \cong F^n$.*

Proof. Let $v_1, \dots, v_n$ be a basis of V, and let $w_1, \dots, w_n$ be a basis of W. By Thm. 8.9, there exists a linear map $f : V \to W$ such that $f(v_j) = w_j$ for all $j \in \{1, \dots, n\}$. By Prop, 8.10, f is an isomorphism. For the second statement, take $W = F^n$. □

Note, however, that in general there is no *natural* (or *canonical*) isomorphism $V \xrightarrow{\sim} F^n$. The choice of isomorphism is equivalent to the choice of a basis, and there are many bases of V. In particular, we may want to choose different bases of V for different purposes, so it does not make sense to identify V with F^n in a specific way.

There is an important result that relates the dimensions of the kernel, image and domain of a linear map.

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

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

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

Proof. By Lemma 6.52, there is a complementary subspace U of $\ker(f)$ in V. (If $\dim V = \infty$, this is still true by the General Basis Extension Theorem 7.1, based on Zorn's Lemma.) We show that f restricts to an isomorphism between U and $\operatorname{im}(f)$. This implies that $\dim U = \dim \operatorname{im}(f)$. On the other hand, $\dim V = \dim \ker(f) + \dim U$, so the dimension formula follows.

Let $f' : U \to \operatorname{im}(f)$ be the linear map given by restricting f. We note that $\ker(f') = \ker(f) \cap U = \{0\}$, so f' is injective. To show that f' is also surjective, take $w \in \operatorname{im}(f)$. Then there is $v \in V$ such that $f(v) = w$. We can write $v = u' + u$ with $u' \in \ker(f)$ and $u \in U$. Now

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

so $w \in \operatorname{im}(f')$ as well. □

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

As a corollary, we have the following criterion for when an endomorphism is an automorphism.

8.14. **Corollary.** *Let V be a finite-dimensional vector space, and let $f : V \to V$ be a linear map. Then the following statements are equivalent.*

(1) *f is an isomorphism.*
(2) *f is injective.*
(3) *f is surjective.*

Proof. Note that f is injective if and only if $\dim \ker(f) = 0$ and f is surjective if and only if $\operatorname{rk}(f) = \dim V$. By Thm.8.13, these two statements are equivalent. □

9. Quotient Spaces

We have seen in the last section that the kernel of a linear map is a linear subspace. One motivation for introducing quotient spaces is the question, is any given linear subspace the kernel of a linear map?

But before we go into this, we need to review the notion of an equivalence relation.

9.1. **Review of Equivalence Relations.** Recall the following definition. An *equivalence relation* on a set X is a relation $\sim$ on X (formally, we can consider the relation to be a subset $R \subset X \times X$, and we write $x \sim y$ when $(x, y) \in R$) that is

(1) *reflexive:* $x \sim x$ for all $x \in X$;
(2) *symmetric:* for all $x, y \in X$, if $x \sim y$, then $y \sim x$;
(3) *transitive:* for all $x, y, z \in X$, if $x \sim y$ and $y \sim z$, then $x \sim z$.

One extreme example is the relation of equality on X. The other extreme example is when all elements of X are 'equivalent' to each other.

The most important feature of an equivalence relation is that it leads to a partition of X into *equivalence classes:* for every $x \in X$, we consider its equivalence class $C_x = \{y \in X : x \sim y\}$. Note that $x \in C_x$ (by reflexivity), so $C_x \neq \emptyset$. Then $x \sim y$ is equivalent to $C_x = C_y$, and $x \not\sim y$ is equivalent to $C_x \cap C_y = \emptyset$. To see this, let $z \in C_x$, so $x \sim z$, so $z \sim x$, and (since $x \sim y$) therefore $z \sim y$, so $y \sim z$, and $z \in C_y$. The other inclusion follows in a similar way. Conversely, assume that $z \in C_x \cap C_y$. Then $x \sim z$ and $y \sim z$, so (using symmetry and transitivity) $x \sim y$.

So two equivalence classes are either equal or disjoint. We can then consider the *quotient set* $X/\!\sim$, which is the set of all equivalence classes,

$$X/\!\sim \; = \{C_x : x \in X\}\,.$$

Note that we have a natural surjective map $\pi : X \to X/\!\sim$, $x \mapsto C_x$.

We use this construction when we want to 'identify' objects with one another that are possibly distinct, but have a common property.

9.2. **Example.** Let $X = \mathbb{Z}$. I claim that

$$n \sim m \iff n - m \text{ is even}$$

defines an equivalence relation. This is easy to check (Exercise). What are the equivalence classes? Well, $C_0 = \{n \in \mathbb{Z} : n \text{ is even}\}$ is the set of all even integers, and $C_1 = \{n \in \mathbb{Z} : n - 1 \text{ is even}\}$ is the set of all odd integers. Together, they partition $\mathbb{Z}$, and $\mathbb{Z}/\!\sim \; = \{C_0, C_1\}$. The natural map $\mathbb{Z} \to \mathbb{Z}/\!\sim$ maps all even numbers to C_0 and all odd numbers to C_1.

But now on to quotient spaces.

9.3. **Definition and Lemma.** Let V be an F-vector space, $U \subset V$ a linear subspace. For $v, v' \in V$, we set

$$v \equiv v' \bmod U \qquad \iff \qquad v - v' \in U\,.$$

This defines an equivalence relation on V, and the equivalence classes have the form

$$C_v = v + U = \{v + u : u \in U\}\,;$$

these sets $v + U$ are called the *cosets* of U in V. We write

$$V/U = \frac{V}{U} = \{v + U : v \in V\}$$

for the quotient set. We define an addition and scalar multiplication on V/U by

$$(v + U) + (v' + U) = (v + v') + U\,, \qquad \lambda(v + U) = \lambda v + U\,.$$

These operations are well-defined and turn V/U into an F-vector space, the *quotient vector space* of V mod U. The natural map $\pi : V \to V/U$ is linear; it is called the *canonical epimorphism.* We have $\ker(\pi) = U$.

Proof. There is a number of statements that need proof. First we need to show that we have indeed defined an equivalence relation:

(1) $v - v = 0 \in U$, so $v \equiv v \bmod U$;
(2) if $v \equiv v' \bmod U$, then $v - v' \in U$, so $v' - v = -(v - v') \in U$, hence $v' \equiv v \bmod U$;
(3) if $v \equiv v' \bmod U$ and $v' \equiv v'' \bmod U$, then $v' - v \in U$ and $v'' - v' \in U$, so $v'' - v = (v' - v) + (v'' - v') \in U$, hence $v \equiv v'' \bmod U$.

Next we have to show that the equivalence classes have the form $v + U$. So let $v \in V$; then $v \equiv v' \bmod U$ if and only if $u = v - v' \in U$, if and only if $v' = v + u$ for some $u \in U$, if and only if $v' \in v + U$.

Next we have to check that the addition and scalar multiplication on V/U are well-defined. Note that a given coset can (usually) be written as $v + U$ for many different $v \in V$, so we have to check that our definition does not depend on the specific *representatives* chosen. So let $v, v', w, w' \in V$ such that $v + U = w + U$ and $v' + U = w' + U$. We have to show that $(v + v') + U = (w + w') + U$, which is equivalent to $(w + w') - (v + v') \in U$. But this follows easily from $w - v, w' - v' \in U$. So addition is OK. For scalar multiplication, we have to show that $\lambda v + U = \lambda w + U$. But $w - v \in U$ implies $\lambda w - \lambda v \in U$, so this is fine, too.

Then we have to show that V/U with the addition and scalar multiplication we have defined is an F-vector space. This is clear from the definitions and the validity of the vector space axioms for V, if we take $U = 0 + U$ as the zero element and $(-v) + U$ as the additive inverse of $v + U$.

It remains to show that the canonical map $V \to V/U$ is linear and has kernel U. But linearity is again clear from the definitions:

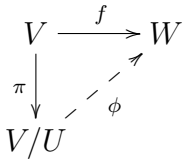
$$\pi(v + v') = (v + v') + U = (v + U) + (v' + U) = \pi(v) + \pi(v') \quad \text{etc.}$$

In fact, the main reason for defining the vector space structure on V/U in the way we have done it is to make π linear! Finally, $\ker(\pi) = \{v \in V : v + U = U\} = U$. □

So we see that indeed every linear subspace of V is the kernel of some linear map $f : V \to W$.

However, the following property of quotient spaces is even more important.

9.4. **Proposition.** *Let $f : V \to W$ be a linear map and $U \subset V$ a linear subspace. If $U \subset \ker(f)$, then there is a unique linear map $\phi : V/U \to W$ such that $f = \phi \circ \pi$, where $\pi : V \to V/U$ is the canonical epimorphism. In other words, there is a unique linear map ϕ that makes the following diagram commutative.*



If $\ker(f) = U$, then ϕ is injective.

Note that this property allows us to construct linear maps with domain V/U.

Proof. It is clear that we must have $\phi(v + U) = f(v)$ (this is what $f = \phi \circ \pi$ means). This already shows uniqueness. For existence, we need to show that ϕ, defined in this way, is well-defined. So let $v, w \in V$ such that $v + U = w + U$. We have to check that $f(v) = f(w)$. But we have $w - v \in U \subset \ker(f)$, so $f(w) - f(v) = f(w - v) = 0$. It is clear from the definitions that ϕ is linear. Finally,

$$\ker(\phi) = \{v + U : f(v) = 0\} = \{U\}$$

if $\ker(f) = U$; this is the zero subspace of V/U, hence ϕ is injective. □

9.5. **Corollary.** *If $f : V \to W$ is a linear map, then $V/\ker(f) \cong \operatorname{im}(f)$.*

Proof. Let $U = \ker(f)$. By Prop. 9.4, there is a linear map $\phi : V/U \to W$ such that $f = \phi \circ \pi$, and ϕ is injective. So ϕ gives rise to an isomorphism $V/U \to \operatorname{im}(\phi) = \operatorname{im}(f)$. □

9.6. **Corollary.** *Let V be a vector space and $U \subset V$ a linear subspace. Then*

$$\dim U + \dim V/U = \dim V\,.$$

Proof. Let $\pi : V \to V/U$ be the canonical epimorphism. By Thm. 8.13, we have $\dim \ker(\pi) + \dim \operatorname{im}(\pi) = \dim V$. But $\ker(\pi) = U$ and $\operatorname{im}(\pi) = V/U$, so the claim follows. □

9.7. **Definition.** Let V be a vector space and $U \subset V$ a linear subspace. Then we call $\dim V/U$ the *codimension* of U in V, written $\operatorname{codim}_V U$.

9.8. **Examples.** Note that $\operatorname{codim}_V U$ can be finite, even though U and V both are infinite-dimensional. A trivial example is $\operatorname{codim}_V V = 0$.

For some less trivial examples, consider again the vector space P of polynomial functions on $\mathbb{R}$. For the subspace C of constant functions, we have $\dim C = 1$ and $\operatorname{codim}_P C = \infty$. For the subspace $Z = \{p \in P : p(0) = 0\}$ of polynomials vanishing at 0, we have $\dim Z = \infty$ and $\operatorname{codim}_P Z = 1$. (Indeed, $Z = \ker(\mathrm{ev}_0)$, so $P/Z \cong \mathbb{R} = \operatorname{im}(\mathrm{ev}_0)$.) Finally, for the subspace

$$E = \{p \in P : \forall x \in \mathbb{R} : p(-x) = p(x)\}$$

of even polynomials, we have $\dim E = \infty$ and $\operatorname{codim}_P E = \infty$.

9.9. **Exercise.** Let $U_1, U_2 \subset V$ be linear subspaces of a (not necessarily finite-dimensional) vector space V. Show that

$$\operatorname{codim}_V(U_1 + U_2) + \operatorname{codim}_V(U_1 \cap U_2) = \operatorname{codim}_V U_1 + \operatorname{codim}_V U_2 \,.$$

For this exercise, the following results may be helpful.

9.10. **Proposition.** *Let $U_1, U_2 \subset V$ be two linear subspaces of the vector space V. Then there is a natural isomorphism*

$$\frac{U_1}{U_1 \cap U_2} \xrightarrow{\sim} \frac{U_1 + U_2}{U_2} \,.$$

Proof. Exercise. □

9.11. **Proposition.** *Let $U \subset V \subset W$ be vector spaces. Then $V/U \subset W/U$ is a linear subspace, and there is a natural isomorphism*

$$\frac{W}{V} \xrightarrow{\sim} \frac{W/U}{V/U} \,.$$

Proof. It is clear that $V/U = \{v + U : v \in V\}$ is a linear subspace of $W/U = \{w + U : w \in W\}$. Consider the composite linear map

$$f : W \longrightarrow W/U \longrightarrow \frac{W/U}{V/U}$$

where both maps involved in the composition are canonical epimorphisms. What is the kernel of f? The kernel of the second map is V/U, so the kernel of f consists of all $w \in W$ such that $w + U \in V/U$. This is equivalent to $w - v \in U$ for some $v \in V$, or $w \in U + V$. Since $U \subset V$, we have $U + V = V$, hence $\ker(f) = V$. The map $\phi : W/V \to (W/U)/(V/U)$ given to us by Prop. 9.4 then is injective and surjective (since f is surjective), hence an isomorphism. □

10. Digression: Finite Fields

Before we embark on studying matrices, I would like to discuss finite fields. First a general notion relating to fields.

10.1. **Lemma and Definition.** Let F be a field with unit element 1_F. As in any abelian group, integral multiples of elements of F are defined:

$$n \cdot \lambda = \begin{cases} \lambda + \lambda + \cdots + \lambda \quad (n \text{ summands}) & \text{if } n > 0, \\ 0 & \text{if } n = 0, \\ (-n) \cdot (-\lambda) & \text{if } n < 0. \end{cases}$$

Consider the set $S = \{n \in \mathbb{N} : n \cdot 1_F = 0\}$. If $S = \emptyset$, we say that F has *characteristic zero*. In this case, all integral multiples of 1_F are distinct. Otherwise, the smallest element p of S is a prime number, and we say that F has *characteristic p*. We write $\operatorname{char} F = 0$ or $\operatorname{char} F = p$.

Proof. First assume that S is empty. Then we have to show that $m \cdot 1_F$ and $n \cdot 1_F$ are distinct when m and n are distinct integers. So assume $m \cdot 1_F = n \cdot 1_F$ and (without loss of generality) $n > m$. Then we have $(n-m) \cdot 1_F = 0$, so $n - m \in S$, a contradiction.

We also have to show that the smallest element of S is a prime number when S is nonempty. So assume the smallest element n of S is not prime. Then $n = km$ with integers $2 \leq k, m < n$. But then we have

$$0 = n \cdot 1_F = km \cdot 1_F = (k \cdot 1_F)(m \cdot 1_F)\,.$$

Since F is a field, one of the two factors must be zero, so $k \in S$ or $m \in S$. But this contradicts the assumption that n is the smallest element of S. □

10.2. **Corollary.** *If F is a finite field, then* $\operatorname{char} F = p$ *for some prime number p.*

Proof. If $\operatorname{char} F = 0$, then the integral multiples of 1_F are all distinct, so F must be infinite. □

In order to see that finite fields of characteristic p exist, we will construct the smallest of them.

10.3. **Definition and Lemma.** Let p be a prime number. The following defines an equivalence relation on $\mathbb{Z}$:

$$a \equiv b \bmod p \quad \Longleftrightarrow \quad p \text{ divides } a - b.$$

Its equivalence classes have the form

$$a + p\mathbb{Z} = \{a + kp : k \in \mathbb{Z}\}\,.$$

Let $\mathbb{F}_p = \{a + p\mathbb{Z} : a \in \mathbb{Z}\}$ be the quotient set. Then

$$\mathbb{F}_p = \{0 + p\mathbb{Z}, 1 + p\mathbb{Z}, 2 + p\mathbb{Z}, \ldots, (p-1) + p\mathbb{Z}\}\,.$$

The following addition and multiplication on $\mathbb{F}_p$ are well-defined:

$$(a + p\mathbb{Z}) + (b + p\mathbb{Z}) = (a + b) + p\mathbb{Z}\,, \qquad (a + b\mathbb{Z})(b + p\mathbb{Z}) = ab + p\mathbb{Z}\,.$$

$\mathbb{F}_p$ with this addition and multiplication is a field, and $\#\mathbb{F}_p = p$ and $\operatorname{char} \mathbb{F}_p = p$.

Proof. That we have defined an equivalence relation is seen in much the same way as in the case of quotient vector spaces. The same statement applies to the form of the equivalence classes and the proof that addition is well-defined. To see that multiplication is well-defined, assume that

$$a \equiv a' \bmod p \quad \text{and} \quad b \equiv b' \bmod p\,.$$

Then there are $k, l \in \mathbb{Z}$ such that $a' = a + kp$, $b' = b + lp$. Then

$$a'b' = (a + kp)(b + lp) = ab + (kb + al + klp)p\,,$$

so $a'b' + p\mathbb{Z} = ab + p\mathbb{Z}$.

It is then clear that all the field axioms are satisfied (with zero $0 + p\mathbb{Z}$ and one $1 + p\mathbb{Z}$), except possibly the existence of multiplicative inverses. For this, we show first that $(a+p\mathbb{Z})(b+p\mathbb{Z}) = 0+p\mathbb{Z}$ implies that $a+p\mathbb{Z} = 0+p\mathbb{Z}$ or $b+p\mathbb{Z} = 0+p\mathbb{Z}$. Indeed, the vanishing of the product means that p divides ab, so p (as a prime number) must divide a or b.

Now consider $a + p\mathbb{Z} \neq 0 + p\mathbb{Z}$. Then I claim that the map

$$\mathbb{F}_p \longrightarrow \mathbb{F}_p\,, \quad x + p\mathbb{Z} \longmapsto (a + p\mathbb{Z})(x + p\mathbb{Z})$$

is injective. Indeed, if $(a+p\mathbb{Z})(x+p\mathbb{Z}) = (a+p\mathbb{Z})(y+p\mathbb{Z})$, then

$$(a+p\mathbb{Z})\big((x+p\mathbb{Z}) - (y+p\mathbb{Z})\big) = 0 + p\mathbb{Z}\,,$$

hence $x+p\mathbb{Z} = y+p\mathbb{Z}$. Since $\mathbb{F}_p$ is finite, the map must then be surjective as well, so there is $x+p\mathbb{Z} \in \mathbb{F}_p$ such that $(a+p\mathbb{Z})(x+p\mathbb{Z}) = 1+p\mathbb{Z}$.

That $\#\mathbb{F}_p = p$ is clear from the description of the equivalence classes. Finally, $\operatorname{char}\mathbb{F}_p = p$ follows from $n\cdot(1+p\mathbb{Z}) = n+p\mathbb{Z}$. □

10.4. **Theorem.** *Let F be a field of characteristic p. Then F is a vector space over $\mathbb{F}_p$. In particular, if F is finite, then $\#F = p^n$ for some $n \in \mathbb{N}$.*

Proof. We define scalar multiplication $\mathbb{F}_p \times F \to F$ by $(a+p\mathbb{Z})\cdot x = a\cdot x$ (where on the right, we use integral multiples). We have to check that this is well-defined. So let $a' = a+kp$. Then

$$(a'+p\mathbb{Z})\cdot x = a'\cdot x = a\cdot x + kp\cdot x = a\cdot x + (p\cdot 1_F)(k\cdot x) = a\cdot x$$

(since $p\cdot 1_F = 0$). The relevant axioms are then clearly satisfied (they are for integral multiples, as one can prove by induction).

If F is finite, then its dimension over $\mathbb{F}_p$ must be finite, say n. Then every element of F is a unique linear combination with coefficients from $\mathbb{F}_p$ of n basis elements, hence F must have p^n elements. □

So we see that there can be no field with exactly six elements, for example.

We know that for every prime number p, there is a field with p elements. What about *existence* of fields with p^n elements for every n?

10.5. **Theorem.** *If p is a prime number and $n \in \mathbb{N}$, then there exists a field with p^n elements, and all such fields are isomorphic (in a suitable sense).*

Proof. The proof of this result is beyond this course. You should see it in the 'Introductory Algebra' course. □

10.6. **Example.** Let us show that there is a field $\mathbb{F}_4$ with four elements. It will be of characteristic 2. Its elements will be 0, 1, α and $\alpha+1$ (since $\alpha+1$ has to be something and cannot be one of 0, 1 or α, it has to be the fourth element). What is α^2? It cannot be 0 or 1 or α, so we must have

$$\alpha^2 = \alpha + 1\,.$$

This implies $\alpha(\alpha+1) = 1$, which shows that our new two elements have multiplicative inverses. Here are the addition and multiplication tables.

$+$	0	1	α	$\alpha+1$
0	0	1	α	$\alpha+1$
1	1	0	$\alpha+1$	α
α	α	$\alpha+1$	0	1
$\alpha+1$	$\alpha+1$	α	1	0

$\cdot$	0	1	α	$\alpha+1$
0	0	0	0	0
1	0	1	α	$\alpha+1$
α	0	α	$\alpha+1$	1
$\alpha+1$	0	$\alpha+1$	1	α

11. Matrices

Matrices are a convenient way of specifying linear maps from F^n to F^m. Since every finite-dimensional F-vector space is isomorphic to some F^n, they can also be used to describe linear maps between finite-dimensional vector spaces in general. Last, but not least, matrices are a very convenient tool for performing explicit computations with linear maps.

11.1. **Definition.** Recall that by Thm. 8.9, a linear map $f : F^n \to F^m$ is uniquely determined by the images of a basis. Now, F^n has a canonical basis $e_1, \dots, e_n$ (where $e_j = (\delta_{1j}, \dots, \delta_{nj})$ and $\delta_{ij} = 1$ if $i = j$ and 0 otherwise; δ_{ij} is called the *Kronecker symbol*), and so f is uniquely specified by

$$\begin{aligned} f(e_1) &= (a_{11}, a_{21}, \dots, a_{m1}) \in F^m \\ f(e_2) &= (a_{12}, a_{22}, \dots, a_{m2}) \in F^m \\ &\vdots \quad \vdots \qquad\qquad \vdots \\ f(e_n) &= (a_{1n}, a_{2n}, \dots, a_{mn}) \in F^m \,. \end{aligned}$$

We arrange the various coefficients $a_{ij} \in F$ in a rectangular array

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

and call this an *$m \times n$ matrix (with entries in F)*. The a_{ij} are called the *entries* or *coefficients* of A. For $i \in \{1, \dots, m\}$, $(a_{i1}, a_{i2}, \dots, a_{in})$ is a *row* of A, and for $j \in \{1, \dots, n\}$,

$$(a_{1j}, a_{2j}, \dots, a_{mj})^\top := \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}$$

is called a *column* of A. Note that the coefficients of $f(e_j)$ appear in the jth column of A. The set of all $m \times n$ matrices with entries in F is denoted by $\mathrm{Mat}(m \times n, F)$. Note that as a boundary case, $m = 0$ or $n = 0$ (or both) is allowed; in this case $\mathrm{Mat}(m \times n, F)$ has only one element, which is an empty matrix and corresponds to the zero homomorphism.

If $m = n$, we sometimes write $\mathrm{Mat}(n, F)$ for $\mathrm{Mat}(n \times n, F)$. The matrix $I = I_n \in \mathrm{Mat}(n, F)$ that corresponds to the identity map id_{F^n} is called the *identity matrix;* we have

$$I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} = (\delta_{ij})_{1 \le i,j \le n} \,.$$

11.2. **Remark.** By Thm. 8.9, the matrices in $\mathrm{Mat}(m \times n, F)$ correspond bijectively to linear maps in $\mathrm{Hom}(F^n, F^m)$. Therefore, we will usually not distinguish between a matrix A and the linear map $F^n \to F^m$ it describes.

In this context, the elements of F^n (and similarly for F^m) are considered as *column vectors*, and we write the linear map given by the matrix $A = (a_{ij})$, as applied to $x = (x_j) \in F^n$ in the form

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

Note that the result is again a column vector, this time of length m (the length of the columns of A). Note also that Ae_j is the jth column of A, hence A really corresponds (in the sense introduced above) to the linear map we have defined here.

11.3. **Definition.** We know that $\mathrm{Hom}(F^n, F^m)$ has the structure of an F-vector space (see Lemma 8.8). We can 'transport' this structure to $\mathrm{Mat}(m \times n, F)$ using the identification of matrices and linear maps. So for $A, B \in \mathrm{Mat}(m \times n, F)$, we define $A + B$ to be the matrix corresponding to the linear map $x \mapsto Ax + Bx$. It is then 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, for $\lambda \in F$ and $A = (a_{ij}) \in \mathrm{Mat}(m \times n, F)$, we define λA to be the matrix corresponding to the linear map $x \mapsto \lambda \cdot Ax$; we then see easily that $\lambda(a_{ij}) = (\lambda a_{ij})$. With this addition and scalar multiplication, $\mathrm{Mat}(m \times n, F)$ becomes an F-vector space, and it is clear that it is 'the same' as (i.e., isomorphic to) F^{mn} — the only difference is the arrangement of the coefficients in a rectangular fashion instead of in a row or column.

11.4. **Definition.** By Lemma 8.3, the composition of two linear maps is again linear. How is this reflected in terms of matrices?

Let $A \in \mathrm{Mat}(l \times m, F)$ and $B \in \mathrm{Mat}(m \times n, F)$. Then B gives a linear map $F^n \to F^m$, and A gives a linear map $F^m \to F^l$. We define the *product* AB to be the matrix corresponding to the composite linear map $F^n \xrightarrow{B} F^m \xrightarrow{A} F^l$. So AB will be a matrix in $\mathrm{Mat}(l \times n, F)$.

To find out what this means in terms of matrix entries, recall that the kth column of AB gives the image of the basis vector e_k. So the kth column of AB is given by $ABe_k = AB_k$, where $B_k = Be_k$ denotes the kth column of B. The ith entry of this column is then

$$a_{i1}b_{1k} + a_{i2}b_{2k} + \cdots + a_{im}b_{mk} = \sum_{j=1}^{m} a_{ij}b_{jk}\,.$$

As a mnemonic, to compute the entry in row i and column k of AB, we take row i of A: $(a_{i1}, a_{i2}, \ldots, a_{im})$ and column k of B: $(b_{1k}, b_{2k}, \ldots, b_{mk})^\top$, and compute their 'dot product' $\sum_{j=1}^{m} a_{ij}b_{jk}$.

If the linear map corresponding to $A \in \mathrm{Mat}(m \times n, F)$ is an isomorphism, then A is called *invertible*. This implies that $m = n$. The matrix corresponding to the inverse linear map is (obviously) denoted A^{-1}; we then have $AA^{-1} = A^{-1}A = I_n$, and A^{-1} is uniquely determined by this property.

11.5. **Remark.** From the corresponding statements on linear maps, we obtain immediately that matrix multiplication is associative:

$$A(BC) = (AB)C$$

for $A \in \mathrm{Mat}(k \times l, F)$, $B \in \mathrm{Mat}(l \times m, F)$, $C \in \mathrm{Mat}(m \times n, F)$, and is distributive with respect to addition:

$$A(B + C) = AB + AC \quad \text{for } A \in \mathrm{Mat}(l \times m, F),\ B, C \in \mathrm{Mat}(m \times n, F);$$
$$(A + B)C = AC + BC \quad \text{for } A, B \in \mathrm{Mat}(l \times m, F),\ C \in \mathrm{Mat}(m \times n, F).$$

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

$$AB = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = BA\,.$$

The identity matrix acts as a multiplicative identity:

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

If $A, B \in \mathrm{Mat}(n, F)$ are both invertible, then AB is also invertible, and $(AB)^{-1} = B^{-1}A^{-1}$ (note the reversal of the factors!).

11.6. **Definition.** Let $A \in \mathrm{Mat}(m \times n, F)$. Then the *rank* of A, $\mathrm{rk}(A)$, is the rank of A, considered as a linear map $F^n \to F^m$. Note that we have $\mathrm{rk}(A) \leq \min\{m, n\}$, since it is the dimension of a subspace of F^m, generated by n vectors.

By this definition, the rank of A is the same as the *column rank* of A, i.e., the dimension of the linear hull of the columns of A (as a subspace of F^m). We can as well define the *row rank* of A to be the dimension of the linear hull of the rows of A (as a subspace of F^n). The following result tells us that these additional definitions are not really necessary.

11.7. **Proposition.** *Let $A \in \mathrm{Mat}(m \times n, F)$ be a matrix. Then the row and column ranks of A are equal.*

Proof. We first note that the dimension of the linear hull of a sequence of vectors equals the length of a maximal linearly independent subsequence (which is then a basis of the linear hull). If we call a row (column) of A 'redundant' if it is a linear combination of the remaining rows (columns), then the row (column) rank of the matrix is therefore unchanged if we remove a redundant row (column). We want to show that removing a redundant column also does not change the row rank and conversely. So suppose that the jth column is redundant. Now assume that a sequence of rows is linearly dependent after removing the jth column. Since the jth column is a linear combination of the other columns, this dependence relation extends to the jth column, hence the rows are also linearly dependent before removing the jth column. This shows that the row rank does not drop (since linearly independent rows stay linearly independent), and as it clearly cannot increase, it must be unchanged. Similarly, we see that removing a redundant row leaves the column rank unaffected.

We now successively remove redundant rows and columns until this is no longer possible. Let the resulting matrix A' have r rows and s columns. Without loss of generality, we can assume $r \leq s$. The column rank is s (since there are s linearly independent columns), but can at most be r (since the columns have r entries), so $r = s$, and row and column rank of A' are equal. But A has the same row and column ranks as A', hence the row and column ranks of A must also be equal. □

There is another way of expressing this result. To do this, we need to introduce another notion.

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

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

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

11.9. **Remark.** The result of Prop. 11.7 can be stated as $\mathrm{rk}(A) = \mathrm{rk}(A^\top)$.

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 $A \in \mathrm{Mat}(n, F)$ is invertible, this implies that $A^\top$ is also invertible, and $(A^\top)^{-1} = (A^{-1})^\top$.

11.10. **Remark (Leiden).** Let $U \subset \mathbb{R}^n$ be a subspace of $\mathbb{R}^n$ that is generated by the vectors $v_1, v_2, \ldots, v_m$. Let $f\colon \mathbb{R}^n \to \mathbb{R}^m$ be the linear map given by

$$f(x) = \big(\langle v_1, x\rangle, \langle v_2, x\rangle, \ldots, \langle v_m, x\rangle\big).$$

Then the kernel of f equals $U^\perp$, cf. Example 6.10. 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 $i \leq m$. The rank $\mathrm{rk}\, f$ of f equals the column rank of M, which equals its row rank $\dim U$ by Proposition 11.7. From Theorem 8.13 we conclude $\dim U + \dim U^\perp = n$, which we had already seen in Lemma 6.48.

12. Computations with Matrices: 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.'

12.1. **Definition.** 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.

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).

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

12.2. **Remark.** *If A' is obtained from A by a sequence of elementary row operations, then there is an invertible matrix B such that $A' = BA$. Similarly, if A' is obtained from A by a sequence of elementary column operations, then there is an invertible matrix C such that $A' = AC$. In both cases, we have* $\mathrm{rk}(A') = \mathrm{rk}(A)$.

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 $(I_m + \lambda E_{ij})A$.

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, …, 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 $A^\top$.

Finally, the statement on the ranks follows from the fact that invertible matrices represent isomorphisms, or also from the simple observation that elementary row (column) operations preserve the row (column) rank, together with Prop. 11.7. □

The following algorithm is the key to most computations with matrices.

12.3. **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 in order to transform it into a matrix A' in* row echelon form. *This means that A' has the following shape.*

$$A' = \begin{pmatrix}
0\cdots0 & 1 & *\cdots* & * & *\cdots* & * & *\cdots* \\
0\cdots0 & 0 & 0\cdots0 & 1 & *\cdots* & * & *\cdots* \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0 & 1 & *\cdots* \\
0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0
\end{pmatrix}$$

So there are $0 \le r \le m$ and $1 \le j_1 < j_2 < \cdots < j_r \le n$ such that if $A' = (a'_{ij})$, then $a'_{ij} = 0$ if $i > r$ or if $i \le r$ and $j < j_i$, and $a'_{ij_i} = 1$ for $1 \le i \le r$.

1. *Set $A' = A$, $r = 0$ and $j_0 = 0$.*
2. *(At this point, $a'_{ij} = 0$ if $i > r$ and $j \le j_r$ or if $1 \le i \le r$ and $1 \le j < j_i$. Also, $a'_{ij_i} = 1$ for $1 \le i \le r$.)*
   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 $r < i \le m$. 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 $i = r+1, \dots, m$, 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. □

12.4. **Proposition.** *The value of the number r at the end of the Row Echelon Form Algorithm is the rank of A. More precisely, the r nonzero rows form a basis of the row space of A.*

Proof. It is clear that the first r rows of A' are linearly independent. Since all remaining rows are zero, the (row) rank of A' is r. But elementary row operations do not change the rank, so $\mathrm{rk}(A) = \mathrm{rk}(A') = r$. Since elementary row operations do not even change the row space (exercise), the second claim also follows. □

12.5. **Example.** 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 and find its rank!

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, and $\mathrm{rk}(A) = 1$. 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

$$A' = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

This is in row echelon form, and we find $\mathrm{rk}(A) = 2$.

12.6. **Proposition.** *If $A \in \mathrm{Mat}(n, F)$ is invertible, then we can transform it into the identity matrix I_n by elementary row operations. The same operations, applied to I_n in the same order, produce the inverse A^{-1}.*

Proof. If A is invertible, then its rank is n. So the row echelon form our algorithm produces looks like this:

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

(A' is an *upper triangular matrix.*)

By adding suitable multiples of the ith row to the rows above it, we can clear all the entries away from the diagonal and get I_n.

By Remark 12.2, $I_n = BA$, where the left multiplication by B corresponds to the row operations performed on A. This implies that $A^{-1} = B = BI_n$, hence we obtain A^{-1} by performing the same row operations on I_n. □

12.7. **Example.** Let us see how to invert the following matrix (where we assume $\mathrm{char}(F) \neq 2$).

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix}$$

It is convenient to perform the row operations on A and on I in parallel:

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

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

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

So

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

12.8. **Remark.** This inversion procedure will also tell us whether the matrix 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 matrix is not invertible. This corresponds to a gap ($j_{i+1} \geq j_i + 2$) in the sequence $j_1, \ldots, j_r$, which implies that $r < n$.

12.9. **Corollary.** *If $A \in \mathrm{Mat}(n, F)$ 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 Remark 12.2.)*

Proof. By Prop. 12.6, A^{-1} can be transformed into I_n by a sequence of elementary row operations, and $A = (A^{-1})^{-1}$ then equals the matrix B such that left multiplication by B effects the row operations. A look at the proof of Remark 12.2 shows that B is a product of matrices of the required form. □

The next application of the Row Echelon Form Algorithm is to compute a basis for the kernel of a matrix (considered as a linear map $F^n \to F^m$).

12.10. **Definition.** 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 $i \neq k$. This means that the entries above the leading 1's in the nonzero rows are zero as well:

$$A = \begin{pmatrix} 0\cdots0 & 1 & *\cdots* & 0 & *\cdots* & 0 & *\cdots* \\ 0\cdots0 & 0 & 0\cdots0 & 1 & *\cdots* & 0 & *\cdots* \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0 & 1 & *\cdots* \\ 0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0 \end{pmatrix}$$

It is clear that every matrix can be transformed into reduced row echelon form by a sequence of elementary row operations — we only have to change Step 5 of the algorithm to

5. For each $i = 1, \ldots, r-1, r+1, \ldots, m$, add $-a'_{ij_r}$ times the rth row of A' to the ith row of A'.

12.11. **Remark.** *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 $A' = BA$ with $B \in \mathrm{Mat}(m, F)$ invertible, then $A = A'$.*

In other words, if we declare two $m \times n$ matrices to be equivalent if one can be obtained from the other by row operations, then the matrices in reduced row echelon form give a complete system of representatives of the equivalence classes.

Proof. Exercise. □

12.12. **Lemma.** *If $A = (a_{ij}) \in \mathrm{Mat}(m \times n, F)$ is in reduced row echelon form, then $\dim \ker(A) = n - r$, and a basis of $\ker(A)$ is given by*

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

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

Proof. It is clear that the given vectors are linearly independent, since v_k is the only vector in the list whose kth entry is nonzero. We know that $\dim\ker(A) = n - \mathrm{rk}(A) = n - r$. So we only have to check that $Av_k = 0$ for all k. For this, note that $Ae_k = (a_{1k}, a_{2k}, \ldots, a_{mk})^\top$ is the kth column of A, that $Ae_{j_i} = e'_i$, where $e'_1, \ldots, e'_m$ is the canonical basis of F^m, and that $a_{ik} = 0$ if $i > r$ or $j_i > k$. So

$$Av_k = Ae_k - \sum_{\substack{1\le i\le r\\ j_i<k}} a_{ik}e'_i = \sum_{i=1}^{m} a_{ik}e'_i - \sum_{i=1}^{m} a_{ik}e'_i = 0\,.$$

□

12.13. **Lemma.** *If $A \in \mathrm{Mat}(m\times n, F)$ is a matrix and A' is obtained from A by a sequence of elementary row operations, then $\ker(A) = \ker(A')$.*

Proof. By Remark 12.2, $A' = BA$ with an invertible matrix B. We then have

$$x \in \ker(A) \iff Ax = 0 \iff BAx = 0 \iff A'x = 0 \iff x \in \ker(A')\,.$$

□

We can therefore compute (a basis of) the kernel of A by first bringing it into reduced row echelon form; then we read off the basis as described in Lemma 12.12.

12.14. **Example.** Let us compute the kernel of the 'telephone matrix'

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

We have seen earlier that we can transform it into the row echelon form (for $\mathrm{char}(F) \neq 3$)

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

From this, we obtain the reduced row echelon form

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

Hence the kernel has dimension 1 and is generated by $(1, -2, 1)^\top$.

If $\mathrm{char}(F) = 3$, the reduced row echelon form is

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

Here, the kernel has dimension 2; a basis is given by $(1, 1, 0)^\top$, $(0, 0, 1)^\top$.

As a final application, we show how we can write a given linear subspace of F^n (given as the linear hull of some vectors) as the kernel of a suitable $m \times n$ matrix.

12.15. **Proposition.** *Let $V = L\big(v_1, \dots, v_k\big) \subset F^n$ be a linear subspace. Write $v_i = (a_{ij})_{1 \le j \le n}$, and let $A = (a_{ij})$ be the $k \times n$ matrix whose rows are given by the coefficients of the v_i. If $U = L\big(w_1, \dots, w_m\big)$ is the kernel of A, then V is the kernel of the $m \times n$ matrix B whose rows are given by the coefficients of the w_j.*

Proof. Let $l = \dim V$; then $\dim U = n - l$. In terms of matrices, we have the relation $AB^\top = 0$, which implies $BA^\top = 0$, hence the v_i, which are the columns of $A^\top$, are in the kernel of B. We also know that $\mathrm{rk}(A) = l$, $\mathrm{rk}(B) = n - l$, therefore $\dim \ker(B) = n - \mathrm{rk}(B) = l = \dim V$. Since, as we have seen, $V \subset \ker(B)$, this implies $V = \ker B$. □

Of course, we use the Row Echelon Form Algorithm in order to compute (a basis of) the kernel of the matrix A.

12.16. **Example.** Let us use Prop. 12.15 to find the intersection of two linear subspaces. Consider

$$U = L\big((1,1,1),(1,2,3)\big) \quad \text{and} \quad V = L\big((1,0,0),(1,-1,1)\big)$$

as linear subspaces of $\mathbb{R}^3$. We want to find (a basis of) $U \cap V$.

To do this, we first write U and V as kernels of suitable matrices. To get that for U, we apply the Row Echelon Form Algorithm to the following matrix.

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

The kernel of this matrix is therefore generated by $(1, -2, 1)^\top$, and U is the kernel of $\begin{pmatrix} 1 & -2 & 1 \end{pmatrix}$.

For V, we proceed as follows.

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & -1 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \end{pmatrix}$$

So V is the kernel of the matrix $\begin{pmatrix} 0 & 1 & 1 \end{pmatrix}$.

Then $U \cap V$ will be the kernel of the following matrix, which we compute using the Row Echelon Form Algorithm again.

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

We see that $U \cap V = L\big((-3,-1,1)\big)$.

12.17. **Remark.** Let $A \in \mathrm{Mat}(k \times n, F)$ and $B \in \mathrm{Mat}(m \times n, F)$. Performing row operations, we can find a basis for the row space of a given matrix, and we can find a basis for the kernel of a given matrix. If we stack the matrix A on top of B to form a matrix $C \in \mathrm{Mat}((k+m) \times n, F)$, then the row space of C will be the sum of the row spaces of A and of B, and the kernel of C will be the intersection of the kernels of A and of B. Since we can switch between representing a linear subspace of F^n as a row space (i.e., as the linear hull of given vectors) or as the kernel of a matrix (i.e., of a linear map $F^n \to F^m$ for some m), we can use our Row Echelon Form Algorithm in order to compute (bases of) sums and intersections of linear subspaces.

13. Linear Equations

One of the very useful applications of Linear Algebra is to linear equations.

13.1. **Definition.** Let $f : V \to W$ be a linear map between two F-vector spaces. The equation

$$f(x) = 0\,,$$

to be solved for $x \in V$, is called a *homogeneous linear equation.* If $V = F^n$ and $W = F^m$ (with $m > 1$), we also speak of a *homogeneous system of linear equations.* (Since the equation consists of m separate equations in F, coming from the coordinates of F^m.)

If $b \in W \setminus \{0\}$, the equation

$$f(x) = b$$

(again to be solved for $x \in V$) is called an *inhomogeneous linear equation,* or in the case $V = F^n$, $W = F^m$, an *inhomogeneous system of linear equations.* The equation or system of equations is called *consistent,* if it has a solution, i.e., if $b \in \mathrm{im}(f)$.

With the theory we have built so far, the following result is essentially trivial.

13.2. **Theorem.** *Let $f : V \to W$ be a linear map between two F-vector spaces.*

(1) *The solution set of the homogeneous linear equation $f(x) = 0$ forms a linear subspace U of V.*
(2) *Let $b \in W \setminus \{0\}$. If the inhomogeneous linear equation $f(x) = b$ is consistent, and $a \in V$ is a solution, then the set of all solutions is the coset $a + U$.*

Proof.

(1) The solution set U is exactly the kernel of f, which is a linear subspace of V by Lemma 8.5.
(2) Let x be any solution. Then $f(x - a) = f(x) - f(a) = b - b = 0$, so $x - a \in U$, and $x \in a + U$. Conversely, if $x \in a + U$, then $f(x) = f(a) = b$.

□

13.3. **Example.** Consider the *wave equation*

$$\frac{\partial^2 f}{\partial t^2} = c^2 \frac{\partial^2 f}{\partial x^2}$$

for $f \in \mathcal{C}^2(\mathbb{R} \times [0, \pi])$, with boundary conditions $f(t, 0) = f(t, \pi) = 0$ and initial conditions $f(0, x) = f_0(x)$ and $\frac{\partial f}{\partial t}(0, x) = 0$. If we ignore the first initial condition for a moment, we can consider this as a homogeneous linear equation, where we let

$$V = \{f \in \mathcal{C}^2(\mathbb{R}\times[0,\pi]) : \forall t \in \mathbb{R} :\ f(t,0) = f(t,\pi) = 0,\ \forall x \in]0,\pi[:\ \tfrac{\partial f}{\partial t}(0,x) = 0\}$$

and $W = \mathcal{C}(\mathbb{R} \times [0, \pi])$, and the linear map $V \to W$ is the *wave operator*

$$w : f \longmapsto \frac{\partial^2 f}{\partial t^2} - c^2 \frac{\partial^2 f}{\partial x^2}\,.$$

We can find fairly easily a bunch of solutions using the trick of 'separating the variables' — we look for solutions of the form $f(t, x) = g(t)h(x)$. This leads to an equation

$$\frac{1}{c^2}\frac{g''(t)}{g(t)} = \frac{h''(x)}{h(x)}\,,$$

and the common value of both sides must be constant. The boundary conditions then force $h(x) = \sin kx$ (up to scaling) for some $k \geq 1$, and then $g(t) = \cos kct$ (again up to scaling). Since we know that the solution set is a linear subspace, we see that all linear combinations

$$f(t,x) = \sum_{k=1}^{n} a_k \cos kct \sin kx$$

are solutions. Such a solution has

$$f(0,x) = \sum_{k=1}^{n} a_k \sin kx\,,$$

so if f_0 is of this form, we have found a (or the) solution to the original problem. Otherwise, we have to use some input from Analysis, which tells us that we can approximate f_0 by linear combinations as above and that the corresponding solutions will approximate the solution we are looking for.

Let us now look at the more familiar case where $V = F^n$ and $W = F^m$, so that we have a system of m linear equations in n variables. This is most conveniently written in matrix notation as $Ax = 0$ in the homogeneous case and $Ax = b$ in the inhomogeneous case, where $x \in F^n$ and $0 \in F^m$ or $b \in F^m$ are considered as column vectors.

13.4. **Algorithm.** To solve a homogeneous system of linear equations $Ax = 0$, use elementary row operations to bring A into reduced row echelon form; then read off a basis of the kernel (which is the solution space) according to Lemma 12.12.

13.5. **Algorithm.** To solve an inhomogeneous system of linear equations $Ax = b$, let $A' = (A|b)$ denote the extended matrix of the system (the matrix A with b attached as an $(n+1)$st column). Use elementary row operations to bring A' into reduced row echelon form. The system is consistent if and only if $n+1$ is not one of the j_k, i.e., the last column does not contain the leading '1' of a row. In this case, the first n coordinates of $-v_{n+1}$ (in the notation of Lemma 12.12) give a solution of the system. A basis of the solution space of the corresponding homogeneous system can be read off from the first n columns of the reduced row echelon form of A'.

Note that the last column does not contain a leading '1' of a row if and only if the rank of the first n columns equals the rank of all $n+1$ columns, i.e., if and only if $\mathrm{rk}(A) = \mathrm{rk}(A')$. The latter is equivalent to saying that b is in the linear hull of the columns of A, which is the image of A as a linear map. The statement on how to find a solution is then easily verified.

13.6. **Example.** Consider the following system of linear equations:

$$\begin{array}{rcrcrcrcl} x & + & y & + & z & + & w & = & 0 \\ x & + & 2y & + & 3z & + & 4w & = & 2 \\ x & + & 3y & + & 5z & + & 7w & = & 4 \end{array}$$

We will solve it according to the procedure outlined above. The extended matrix is

$$A' = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 2 & 3 & 4 & 2 \\ 1 & 3 & 5 & 7 & 4 \end{pmatrix}.$$

We transform it into reduced row echelon form:

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 2 & 3 & 4 & 2 \\ 1 & 3 & 5 & 7 & 4 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 2 & 3 & 2 \\ 0 & 2 & 4 & 6 & 4 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & -1 & -2 & -2 \\ 0 & 1 & 2 & 3 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Since the last column does not contain the leading 1 of a row, the system is consistent, and a solution is given by $(x, y, z, w) = (-2, 2, 0, 0)$. The kernel of the non-extended matrix has basis $(1, -2, 1, 0)$, $(2, -3, 0, 1)$. So all solutions are given by

$$(x, y, z, w) = (-2 + r + 2s, 2 - 2r - 3s, r, s)\,,$$

where r and s are arbitrary.

14. Matrices and Linear Maps

So far, we have considered matrices as representing linear maps between F^n and F^m. But on the other hand, we have seen earlier (see Cor. 8.11) that any n-dimensional F-vector space is isomorphic to F^n, the isomorphism coming from the choice of a basis. This implies that we can use matrices to represent linear maps between arbitrary finite-dimensional vector spaces. One important thing to keep in mind here is that this representation will *depend on the bases chosen* for the two vector spaces — it does not make sense to say that A is "the matrix of f", one has to say that A is the matrix of f with respect to the chosen bases.

14.1. **Definition.** If V is an F-vector space with basis $v_1, \ldots, v_n$, then the isomorphism

$$\Phi_{(v_1,\ldots,v_n)} : F^n \longrightarrow V\,, \quad (\lambda_1, \ldots, \lambda_n) \longmapsto \lambda_1 v_1 + \ldots \lambda_n v_n$$

is called the *canonical basis isomorphism* with respect to the basis $v_1, \ldots, v_n$.

14.2. **Definition.** Let V and W be F-vector spaces with bases $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$, respectively. Let $f : V \to W$ be a linear map. Then the matrix $A \in \mathrm{Mat}(m \times n, F)$ that is defined by the commutative diagram

$$\begin{array}{ccc} V & \xrightarrow{f} & W \\ {\scriptstyle \Phi_{(v_1,\ldots,v_n)}}\uparrow{\scriptstyle \cong} & & {\scriptstyle \cong}\uparrow{\scriptstyle \Phi_{(w_1,\ldots,w_m)}} \\ F^n & \xrightarrow{A} & F^m \end{array}$$

is called the *matrix associated to* f relative to the chosen bases. In terms of maps, we then have

$$\Phi_{(w_1,\ldots,w_m)} \circ A \circ \Phi^{-1}_{(v_1,\ldots,v_n)} = f \quad \text{and} \quad \Phi^{-1}_{(w_1,\ldots,w_m)} \circ f \circ \Phi_{(v_1,\ldots,v_n)} = A\,.$$

Now what happens when we change to a different basis? Let $v_1, \ldots, v_n$ and $v'_1, \ldots, v'_n$ be two bases of the F-vector space V. For simplicity, write $\Phi = \Phi_{(v_1,\ldots,v_n)}$ and $\Phi' = \Phi_{(v'_1,\ldots,v'_n)}$. Then we have a diagram as follows.

$$\begin{array}{ccc} & V & \\ {\scriptstyle \Phi'}\nearrow{\scriptstyle \cong} & & {\scriptstyle \cong}\nwarrow{\scriptstyle \Phi} \\ F^n & \xrightarrow{P} & F^n \end{array}$$

14.3. **Definition.** The matrix P defined by the diagram above is called the *basis change matrix* associated to changing the basis from $v_1, \dots, v_n$ to $v'_1, \dots, v'_n$. Since $\Phi^{-1} \circ \Phi'$ is an isomorphism, $P \in \mathrm{Mat}(n, F)$ is invertible.

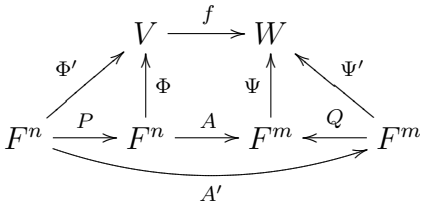
Conversely, given an invertible matrix $P \in \mathrm{Mat}(n, F)$ and the basis $v_1, \dots, v_n$, we can define $\Phi' = \Phi \circ P$ and hence the new basis $v'_1 = \Phi'(e_1), \dots, v'_n = \Phi'(e_n)$ (where, as usual, $e_1, \dots, e_n$ is the canonical basis of F^n).

14.4. **Lemma.** *Let V and W be F-vector spaces, let $v_1, \dots, v_n$ and $v'_1, \dots, v'_n$ be bases of V, and let $w_1, \dots, w_m$ and $w'_1, \dots, w'_m$ be bases of W. Let $f : V \to W$ be a linear map, and let A be the matrix associated to f relative to the bases $v_1, \dots, v_n$ and $w_1, \dots, w_m$, and let A' be the matrix associated to f relative to the bases $v'_1, \dots, v'_n$ and $w'_1, \dots, w'_m$. Then*

$$A' = Q^{-1}AP$$

where P is the basis change matrix associated to changing the basis of V from $v_1, \dots, v_n$ to $v'_1, \dots, v'_n$, and Q is the basis change matrix associated to changing the basis of W from $w_1, \dots, w_m$ to $w'_1, \dots, w'_m$.

Proof. Write $\Phi = \Phi_{(v_1,\dots,v_n)}$, $\Phi' = \Phi_{(v'_1,\dots,v'_n)}$, $\Psi = \Phi_{(w_1,\dots,w_m)}$ and $\Psi' = \Phi_{(w'_1,\dots,w'_m)}$. We have a commutative diagram



from which the statement can be read off. □

14.5. **Corollary.** *If $f : V \to W$ is a linear map between finite-dimensional F-vector spaces and $A \in \mathrm{Mat}(m \times n, F)$ is the matrix associated to f relative to some choice of bases of V and W, then the set of all matrices associated to f relative to any choice of bases is*

$$\{QAP : P \in \mathrm{Mat}(n, F), Q \in \mathrm{Mat}(m, F), P \text{ and } Q \text{ invertible}\}.$$

Proof. By Lemma 14.4, every matrix associated to f is in the given set. Conversely, given invertible matrices P and Q, we can change the bases of V and W in such a way that P and Q^{-1} are the corresponding basis change matrices. Then (by Lemma 14.4 again) QAP is the matrix associated to f relative to the new bases. □

If we choose bases that are well-adapted to the linear map, then we will obtain a very nice matrix. This is used in the following result.

14.6. **Corollary.** *Let $A \in \mathrm{Mat}(m \times n, F)$. Then there are invertible matrices $P \in \mathrm{Mat}(n, F)$ and $Q \in \mathrm{Mat}(m, F)$ such that*

$$QAP = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix} = \left(\begin{array}{c|c} I_r & 0_{r\times(n-r)} \\ \hline 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{array}\right),$$

where $r = \mathrm{rk}(f)$.

Proof. Let $V = F^n$, $W = F^m$, and let $f : V \to W$ be the linear map given by A. Let $v_1, \ldots, v_n$ be a basis of V such that $v_{r+1}, \ldots, v_n$ is a basis of $\ker(f)$. Then $w_1 = f(v_1), \ldots, w_r = f(v_r)$ are linearly independent in W, and we can extend to a basis $w_1, \ldots, w_m$. We then have

$$f(v_i) = \begin{cases} w_i & \text{if } 1 \le i \le r \\ 0 & \text{if } r+1 \le i \le n. \end{cases}$$

So the matrix A' associated to f relative to these bases has the required form. Let P and Q^{-1} be the basis change matrices associated to changing the bases of V and W from the canonical bases to the new ones. By Lemma 14.4, we then have $A' = QAP$ (since A is the matrix associated to f relative to the canonical bases). □

14.7. **Remarks.**

(1) If we say that two matrices $A, A' \in \mathrm{Mat}(m \times n, F)$ are *equivalent* if there are invertible matrices $P \in \mathrm{Mat}(n, F)$ and $Q \in \mathrm{Mat}(m, F)$ such that $A' = QAP$ (exercise: this really defines an equivalence relation), then Cor. 14.6 tells us that A and A' are equivalent if and only if $\mathrm{rk}(A) = \mathrm{rk}(A')$. To see this, first note that if A and A' are equivalent, they must have the same rank (since the rank does not change under multiplication by invertible matrices). Then Cor. 14.6 tells us that if A has rank r, it is equivalent to the matrix given there, so any two matrices of rank r are equivalent to the same matrix.

(2) Recall that by Remark 12.2, row operations on a matrix A correspond to multiplication on the left by an invertible matrix, and column operations on A correspond to multiplication on the right by an invertible matrix. Interpreting A as the matrix associated to a linear map relative to some bases, we see that row operations correspond to changing the basis of the target space (containing the columns) of A, whereas column operations correspond to changing the basis of the domain space (containing the rows) of A. The result of Cor. 14.6 then also means that any matrix A can be transformed into the given simple form by elementary row and column operations. The advantage of this approach is that by keeping track of the operations, we can also determine the matrices P and Q explicitly, much in the same way as when inverting a matrix.

14.8. **Endomorphisms.** If we consider *endomorphims* $f : V \to V$, then there is only one basis to choose. If A is the matrix associated to f relative to one basis, and P is the basis change matrix associated to changing that basis to another one, then the matrix associated to f relative to the new basis will be $A' = P^{-1}AP$, see Lemma 14.4. Matrices $A, A' \in \mathrm{Mat}(n, F)$ such that there is an invertible matrix $P \in \mathrm{Mat}(n, F)$ with $A' = P^{-1}AP$ are said to be *similar*. This defines again an equivalence relation (exercise).

We have seen that it is easy to classify matrices with respect to equivalence: the equivalence class is determined by the rank. In contrast to this, the classification of matrices with respect to similarity is much more complicated. For example, the 'multiplication by λ' endomorphism (for $\lambda \in F$) has matrix λI_n regardless of the basis, and so λI_n and μI_n are not similar if $\lambda \neq \mu$.

14.9. **Example.** As another example, consider the matrices

$$M_{\lambda,t} = \begin{pmatrix} \lambda & t \\ 0 & \lambda \end{pmatrix} .$$

The corresponding endomorphism $f_{\lambda,t}$ has $\ker(f_{\lambda,t} - \mu \,\mathrm{id}) = 0$ if $\lambda \neq \mu$, and has nontrivial kernel otherwise. This shows that $M_{\lambda,t}$ and $M_{\mu,u}$ can be similar only when $\lambda = \mu$. Since $\dim \ker(f_{\lambda,t} - \lambda \,\mathrm{id})$ is 1 if $t \neq 0$ and 2 if $t = 0$, $M_{\lambda,0}$ and $M_{\lambda,1}$ are not similar. On the other hand, $M_{\lambda,t}$ is similar to $M_{\lambda,1}$ if $t \neq 0$, since

$$\begin{pmatrix} \lambda & t \\ 0 & \lambda \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & t^{-1} \end{pmatrix} \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & t \end{pmatrix} .$$

This example gives you a first glimpse of the classification theorem, the 'Jordan Normal Form Theorem', which will be a topic later.

14.10. **The Trace.** For purposes of classification, it is useful to have *invariants*, i.e., functions that are constant on the equivalence classes. In the case of equivalence of matrices, the rank is an invariant, and in this case, it gives the complete classification. The rank is (of course) still an invariant with respect to similarity, but as the example above shows, it is by no means sufficient to separate the classes. Here is another invariant.

14.11. **Definition.** For $A = (a_{ij}) \in \mathrm{Mat}(n, F)$, we define the *trace* of A to be

$$\mathrm{Tr}(A) = a_{11} + a_{22} + \cdots + a_{nn} .$$

14.12. **Lemma.** *If $A \in \mathrm{Mat}(m \times n, F)$ and $B \in \mathrm{Mat}(n \times m, F)$, so that both products AB and BA are defined, then*

$$\mathrm{Tr}(AB) = \mathrm{Tr}(BA) .$$

Proof. The (i, i)-entry of AB is $\sum_{j=1}^{n} a_{ij} b_{ji}$. The (j, j)-entry of BA is $\sum_{i=1}^{m} b_{ji} a_{ij}$. So we get

$$\mathrm{Tr}(AB) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ji} = \sum_{j=1}^{n} \sum_{i=1}^{m} b_{ji} a_{ij} = \mathrm{Tr}(BA) .$$

□

14.13. **Corollary.** *Let $A, A' \in \mathrm{Mat}(n, F)$ be similar. Then $\mathrm{Tr}(A) = \mathrm{Tr}(A')$.*

Proof. There is an invertible matrix $P \in \mathrm{Mat}(n, F)$ such that $A' = P^{-1}AP$. It follows that

$$\mathrm{Tr}(A') = \mathrm{Tr}(P^{-1} \cdot AP) = \mathrm{Tr}(AP \cdot P^{-1}) = \mathrm{Tr}(A)\,.$$

□

This allows us to make the following definition.

14.14. **Definition.** Let V be a finite-dimensional F-vector space and $f : V \to V$ an endomorphism of V. We define the *trace* of f, $\mathrm{Tr}(f)$, to be the trace of any matrix associated to f relative to some basis of V.

Note that $\mathrm{Tr}(f)$ is well-defined, since all matrices associated to f have the same trace according to Cor. 14.13.

In the next section, we will introduce another invariant, which is even more important than the trace: the determinant.

14.15. **Remark.** To finish off this section, let us remark that, having chosen bases of the F-vector spaces V and W of dimensions n and m, respectively, we obtain an isomorphism

$$\mathrm{Hom}(V, W) \xrightarrow{\cong} \mathrm{Mat}(m \times n, F)\,, \quad f \longmapsto A\,,$$

where A is the matrix associated to f relative to the chosen bases. In particular, we see that $\dim \mathrm{Hom}(V, W) = mn$.

15. Determinants

Let V be a real vector space of dimension n. The determinant will be a number associated to an endomorphism f of V that tells us how f scales 'oriented volume' in V. So we have to think a little bit about functions that define 'oriented volume'.

We will only consider *parallelotopes;* these are the bodies spanned by n vectors in V:

$$P(v_1, \ldots, v_n) = \{\lambda_1 v_1 + \cdots + \lambda_n v_n : \lambda_1, \ldots, \lambda_n \in [0, 1]\}$$

If $V = \mathbb{R}^n$, then $P(v_1, \ldots, v_n)$ is the image of the 'unit cube' $P(e_1, \ldots, e_n)$ under the linear map that sends the canonical basis vectors $e_1, \ldots, e_n$ to $v_1, \ldots, v_n$.

Now let $D : V^n \to \mathbb{R}$ be a function that is supposed to measure oriented volume of parallelotopes — $D(v_1, \ldots, v_n)$ gives the volume of $P(v_1, \ldots, v_n)$. What properties should such a function D satisfy?

One property should certainly be that the volume vanishes when the parallelotope is of lower dimension, i.e., when its spanning vectors are linearly dependent. It will be sufficient to only consider the special case when two of the vectors are equal:

$$D(v_1, \ldots, v_n) = 0 \quad \text{if } v_i = v_j \text{ for some } 1 \leq i < j \leq n.$$

Also, volume should scale corresponding to scaling of the vectors:

$$D(v_1, \ldots, v_{i-1}, \lambda v_i, v_{i+1}, \ldots, v_n) = \lambda D(v_1, \ldots, v_n)\,.$$

Finally, volumes are additive in the following sense:

$$D(v_1, \ldots, v_{i-1}, v_i + v_i', v_{i+1}, \ldots, v_n) \\ = D(v_1, \ldots, v_{i-1}, v_i, v_{i+1}, \ldots, v_n) + D(v_1, \ldots, v_{i-1}, v_i', v_{i+1}, \ldots, v_n)\,.$$

The last two properties can be stated simply by saying that D is linear in each argument separately. Such a function is said to be *multilinear*. A multilinear function satisfying the first property is said to be *alternating*. So the functions we are looking for are alternating multilinear functions from V^n to $\mathbb{R}$.

Note that it makes sense to talk about alternating multilinear functions over any field F, not just over $\mathbb{R}$ (even though we cannot talk about volumes any more). So we will from now on allow arbitrary fields again.

15.1. **Definition.** Let V be an n-dimensional F-vector space. An alternating multilinear function $D : V^n \to F$ is called a *determinantal function* on V.

How many determinantal functions are there? First, it is pretty clear that the set of all determinantal functions on V forms an F-vector space. So the question we should ask is, what is the dimension of this vector space?

Before we state the relevant theorem, let us first prove a few simple properties of determinantal functions.

15.2. **Lemma.** *Let V be an n-dimensional F-vector space, and let $D : V^n \to F$ be a determinantal function on V.*

(1) *If $v_1, \ldots, v_n \in V$ are linearly dependent, then $D(v_1, \ldots, v_n) = 0$.*
(2) *If we add a scalar multiple of v_i to v_j, where $i \neq j$, then $D(v_1, \ldots, v_n)$ is unchanged.*
(3) *If we interchange two of the vectors $v_1, \ldots, v_n \in V$, then $D(v_1, \ldots, v_n)$ changes sign.*

Proof.

(1) If $v_1, \ldots, v_n$ are linearly dependent, then one of them, say v_i, will be a linear combination of the others, say

$$v_i = \sum_{j \neq i} \lambda_j v_j\,.$$

This implies

$$\begin{aligned} D(v_1, \ldots, v_i, \ldots, v_n) &= D(v_1, \ldots, \sum_{j \neq i} \lambda_j v_j, \ldots, v_n) \\ &= \sum_{j \neq i} \lambda_j D(v_1, \ldots, v_j, \ldots, v_j, \ldots, v_n) \\ &= \sum_{j \neq i} \lambda_j \cdot 0 = 0\,. \end{aligned}$$

(2) Say, we replace v_j by $v_j + \lambda v_i$. Assuming that $i < j$, we have

$$D(v_1, \ldots, v_i, \ldots, v_j + \lambda v_i, \ldots, v_n) \\ = D(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) + \lambda D(v_1, \ldots, v_i, \ldots, v_i, \ldots, v_n) \\ = D(v_1, \ldots, v_n) + \lambda \cdot 0 = D(v_1, \ldots, v_n)\,.$$

(3) To interchange v_i and v_j (with $i < j$), we proceed as follows.

$$\begin{aligned} D(v_1, \dots, v_i, \dots, v_j, \dots, v_n) &= D(v_1, \dots, v_i, \dots, v_j + v_i, \dots, v_n) \\ &= D(v_1, \dots, v_i - (v_j + v_i), \dots, v_j + v_i, \dots, v_n) \\ &= D(v_1, \dots, -v_j, \dots, (v_j + v_i) + (-v_j), \dots, v_n) \\ &= -D(v_1, \dots, v_j, \dots, v_i, \dots, v_n)\,. \end{aligned}$$

Alternatively, we can use that (omitting all except the ith and jth arguments)

$$\begin{aligned} 0 &= D(v_i + v_j, v_i + v_j) \\ &= D(v_i, v_i) + D(v_i, v_j) + D(v_j, v_i) + D(v_j, v_j) \\ &= D(v_i, v_j) + D(v_j, v_i)\,. \end{aligned}$$

□

15.3. **Theorem.** *Let V be an n-dimensional F-vector space, with basis $v_1, \dots, v_n$, and let $\lambda \in F$. Then there is a unique determinantal function $D : V^n \to F$ such that $D(v_1, \dots, v_n) = \lambda$. In particular, the determinantal functions on V form a one-dimensional F-vector space.*

Proof. As usual, we have to prove existence and uniqueness, and we will start with uniqueness. Let $w_1, \dots, w_n$ be vectors in V, and let A be the matrix whose columns are given by the coefficients of the w_j when written as linear combinations of the v_i. If the w_j are linearly dependent, then $D(w_1, \dots, w_n) = 0$. Otherwise, the matrix A is invertible, and we can transform it into the identity matrix by elementary column operations. The multilinearity of D and Lemma 15.2 tell us how the value of D changes in the process: we see that

$$D(w_1, \dots, w_n) = (-1)^k \delta^{-1} D(v_1, \dots, v_n) = (-1)^k \delta^{-1} \lambda\,,$$

where k is the number of times we have swapped two columns and δ is the product of all the scaling factors we have used when scaling a column. Note that the identity matrix corresponds to $v_1, \dots, v_n$. This shows that there is at most one choice for $D(w_1, \dots, w_n)$.

We cannot use the observation made in the uniqueness proof easily to show existence (we would have to show that $(-1)^k \delta^{-1}$ does not depend on the sequence of elementary column operations we have performed in order to obtain I_n). Instead, we use induction on the dimension n of V.

As the base case, we consider $n = 0$. Then $V = \{0\}$, the basis is empty, and V^0 has just one element (which coincides with the empty basis). So the function that sends this element to λ is trivially a determinantal function with the required property. (If you suffer from *horror vacui,* i.e. you are afraid of the empty set, you can consider $n = 1$. Then $V = L(v_1)$, and the required function is given by sending $\mu v_1 \in V^1 = V$ to $\mu\lambda$.)

For the induction step, we assume $n \geq 1$ and let $W = L(v_2, \dots, v_n)$. By the induction hypothesis, there is a determinantal function D' on W that takes the value λ on $(v_2, \dots, v_n)$. Any element $w_j \in V$ can be written uniquely as $w_j = \mu_j v_1 + w_j'$ with $w_j' \in W$. We now set

$$D(w_1, \dots, w_n) = \sum_{j=1}^{n} (-1)^{j-1} \mu_j D'(w_1', \dots, w_{j-1}', w_{j+1}', \dots, w_n')$$

and have to check that D is a determinantal function on V. We first verify that D is linear in w_k. This follows from the observation that each term in the sum is linear in $w_k = \mu_k v_1 + w'_k$ — the term with $j = k$ only depends on w_k through μ_k, and the other terms only depend on w_k through w'_k, which is linear in w_k; also D' is linear in each of its arguments. Next assume that $w_k = w_l$ for $k < l$. Then $w'_k = w'_l$, and so in all terms that have $j \notin \{k, l\}$, the value of D' is zero. The remaining terms are, writing $w_k = w_l = \mu v_1 + w'$,

$$(-1)^{k-1}\mu D'(w'_1, \ldots, w'_{k-1}, w'_{k+1}, \ldots, w'_{l-1}, w', w'_{l+1}, \ldots, w'_n) \\ + (-1)^{l-1}\mu D'(w'_1, \ldots, w'_{k-1}, w', w'_{k+1}, \ldots, w'_{l-1}, w'_{l+1}, \ldots, w'_n)\,.$$

These terms cancel since we have to swap adjacent arguments $(l-k-1)$ times to go from one value of D' to the other, which results in a sign of $(-1)^{l-k-1}$.

Finally, we have $D(v_1, v_2, \ldots, v_n) = 1 \cdot D'(v_2, \ldots, v_n) = \lambda$. □

We can now make use of this fact in order to define determinants of matrices and of endomorphisms.

15.4. **Definition.** Let $n \geq 0$. The *determinant* on $\mathrm{Mat}(n, F)$ is the unique determinantal function on the columns of the $n \times n$ matrices that takes the value 1 on the identity matrix I_n. If $A \in \mathrm{Mat}(n, F)$, then then its value $\det(A)$ on A is called the *determinant* of A.

If $A = (a_{ij})$ is written as an $n \times n$ array of entries, we also write

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

15.5. **Remarks.**

(1) Note that $\det(A) \neq 0$ is equivalent to "A invertible".

(2) The uniqueness proof gives us a procedure to compute determinants: we perform elementary column operations on A, keeping track of the scalings and swappings, until we get a zero column (then $\det(A) = 0$), or we reach the identity matrix.

15.6. **Example.** We compute a determinant by elementary column operations. Note that we can avoid divisions (and hence fractions) by choosing the operations cleverly.

$$\begin{vmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \\ 3 & 4 & 2 & 1 \\ 4 & 3 & 1 & 2 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 2 & -3 & -2 & -5 \\ 3 & -2 & -7 & -11 \\ 4 & -5 & -11 & -14 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & -2 & -5 \\ 3 & 12 & -7 & -11 \\ 4 & 17 & -11 & -14 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -21 & 12 & 17 & 49 \\ -30 & 17 & 23 & 71 \end{vmatrix}$$

$$= \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -21 & 12 & 17 & -2 \\ -30 & 17 & 23 & 2 \end{vmatrix} = 2 \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -21 & 12 & 1 & 17 \\ -30 & 17 & -1 & 23 \end{vmatrix} = 2 \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -51 & 29 & -1 & 40 \end{vmatrix}$$

$$= 2 \cdot 40 \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} = 80$$

15.7. **Lemma (Expansion by Rows).** *Let $A \in \mathrm{Mat}(n, F)$ with $n \geq 1$. For $1 \leq i, j \leq n$, denote by A_{ij} the matrix obtained from A by deleting the ith row and jth column. Then for all $1 \leq i \leq n$, we have*

$$\det(A) = \sum_{j=1}^{n} (-1)^{j-i} a_{ij} \det(A_{ij}) \,.$$

This is called the *expansion of the determinant by the ith row.*

Proof. As in the proof of Thm. 15.3 above, we check that the right hand side is a determinantal function that takes the value 1 on I_n. □

15.8. **Examples.** For $n = 0$, we find that the empty determinant is 1. For $n = 1$ and $A = (a)$, we have $\det(A) = a$. For $n = 2$ and $n = 3$, we obtain by an application of the lemma the following formulas.

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + cdh - afh - bdi - ceg$$

We will see later how this generalizes to larger n.

15.9. **Definition.** Let V be an F-vector space of dimension n, and let $f : V \to V$ be an endomorphism. We fix a nontrivial determinantal function $D : V^n \to F$. Then

$$D' : V^n \longrightarrow F \,, \quad (v_1, \ldots, v_n) \longmapsto D\big(f(v_1), \ldots, f(v_n)\big)$$

is again a determinantal function, hence there is $\lambda \in F$ such that $D' = \lambda D$ (since D is a basis of the space of determinantal functions on V). We set $\det(f) = \lambda$ and call $\det(f)$ the *determinant* of f.

Note that this is well-defined, since any other choice of D differs from the one we have made by a non-zero scalar, by which D' is also multiplied, and so it drops out when computing λ.

15.10. **Theorem.** *Let V be an F-vector space of dimension n, and let $f, g : V \to V$ be two endomorphisms. Then $\det(f \circ g) = \det(f) \det(g)$.*

Proof. Let D be a fixed nontrivial determinantal function on V, and let D' be "D after f" and let D'' be "D after g" as in the definition above. Then

$$\begin{aligned} D\big(f \circ g(v_1), \ldots, f \circ g(v_2)\big) &= D'\big(g(v_1), \ldots, g(v_n)\big) \\ &= \det(f) D\big(g(v_1), \ldots, g(v_n)\big) \\ &= \det(f) D''(v_1, \ldots, v_n) \\ &= \det(f) \det(g) D(v_1, \ldots, v_n) \,. \end{aligned}$$

The definition then implies that $\det(f \circ g) = \det(f) \det(g)$. □

15.11. **Lemma.** *Let V be an F-vector space of dimension n, let $f : V \to V$ be an endomorphism. We fix a basis $v_1, \dots, v_n$ of V and let A be the matrix associated to f relative to this basis. Then* $\det(A) = \det(f)$.

Proof. Let $D : V^n \to F$ be the determinantal function that takes the value 1 on the fixed basis. The columns of A contain the coefficients of $f(v_1), \dots, f(v_n)$ with respect to this basis, hence

$$\det(A) = D\big(f(v_1), \dots, f(v_n)\big) = \det(f) D(v_1, \dots, v_n) = \det(f)\,.$$

□

15.12. **Theorem (Multiplicativity of the Determinant).**
Let $A, B \in \mathrm{Mat}(n, F)$. Then $\det(AB) = \det(A)\det(B)$.

Proof. Consider $V = F^n$ and let (for sake of clarity) $f_A, f_B : V \to V$ be the endomorphisms given by A and B. By Lemma 15.11 and Thm. 15.10, we have

$$\det(AB) = \det(f_A \circ f_B) = \det(f_A)\det(f_B) = \det(A)\det(B)\,.$$

□

15.13. **Theorem.** *Let $A \in \mathrm{Mat}(n, F)$. Then* $\det(A^\top) = \det(A)$.

Proof. We show that $A \mapsto \det(A^\top)$ is a determinantal function of the columns of A. First, we have

$$\det(A) = 0 \iff \mathrm{rk}(A) < n \iff \mathrm{rk}(A^\top) < n \iff \det(A^\top) = 0\,,$$

so our function is alternating. Second, we have to show that $\det(A^\top)$ is linear in each of the columns of A. This is obviously equivalent to saying that $\det(A)$ is linear in each of the *rows* of A. To check that this is the case for the ith row, we expand $\det(A)$ by the ith row according to Lemma 15.7. For $A = (a_{ij})$,

$$\det(A) = \sum_{j=1}^{n} (-1)^{j-i} a_{ij} \det(A_{ij})\,.$$

Now in A_{ij} the ith row of A has been removed, so $\det(A_{ij})$ does not depend on the ith row of A; linearity is then clear from the formula. Finally, we have $\det(I_n^\top) = \det(I_n) = 1$, so $\det(A^\top)$ must coincide with $\det(A)$ because of the uniqueness of determinantal functions. □

15.14. **Corollary (Expansion by Columns).** *We can also expand determinants by columns. Let $n \geq 1$ and $A = (a_{ij}) \in \mathrm{Mat}(n, F)$; we use the notation A_{ij} as before. Then for $1 \leq j \leq n$,*

$$\det(A) = \sum_{i=1}^{n} (-1)^{j-i} a_{ij} \det(A_{ij})\,.$$

Proof. We have

$$\begin{aligned}\det(A) = \det(A^\top) &= \sum_{j=1}^{n}(-1)^{j-i}a_{ji}\det\big((A^\top)_{ij}\big)\\ &= \sum_{j=1}^{n}(-1)^{j-i}a_{ji}\det\big((A_{ji})^\top\big) = \sum_{j=1}^{n}(-1)^{j-i}a_{ji}\det(A_{ji})\\ &= \sum_{i=1}^{n}(-1)^{j-i}a_{ij}\det(A_{ij})\,.\end{aligned}$$

□

15.15. **Example.** A matrix $A \in \mathrm{Mat}(n,F)$ is said to be *orthogonal* if $AA^\top = I_n$. What can we deduce about $\det(A)$? Well,

$$1 = \det(I_n) = \det(AA^\top) = \det(A)\det(A^\top) = \det(A)^2\,,$$

so $\det(A) = \pm 1$.

15.16. **Definition.** Let $A \in \mathrm{Mat}(n,F)$ with $n \geq 1$. Then the *adjugate* matrix of A (sometimes called the *adjoint* matrix, but this has also other meanings) is the matrix $\tilde{A} \in \mathrm{Mat}(n,F)$ whose (i,j)-entry is $(-1)^{j-i}\det(A_{ji})$. Here A_{ij} is, as before, the matrix obtained from A by removing the ith row and jth column. Note the reversal of indices — $\tilde{A}_{ij} = (-1)^{j-i}\det(A_{ji})$ and not $\det(A_{ij})$!

15.17. **Proposition.** *Let $A \in \mathrm{Mat}(n,F)$ with $n \geq 1$. Then*

$$A\tilde{A} = \tilde{A}A = \det(A)I_n\,.$$

In particular, if A is invertible, then $\det(A) \neq 0$, and

$$A^{-1} = \det(A)^{-1}\tilde{A}\,.$$

Proof. The (i,k)-entry of $A\tilde{A}$ is

$$\sum_{j=1}^{n} a_{ij}(-1)^{k-j}\det(A_{kj})\,.$$

If $i = k$, then this is just the formula that expands $\det(A)$ by the ith row, so $A\tilde{A}$ has diagonal entries equal to $\det(A)$. If $i \neq k$, then the result is unchanged if we modify the kth row of A (since A_{kj} does not involve the kth row of A). So we get the same result as for the matrix $A' = (a'_{ij})$ which we obtain from A by replacing the kth row by the ith row. We find that

$$0 = \det(A') = \sum_{j=1}^{n}(-1)^{k-j}a'_{kj}\det(A'_{kj}) = \sum_{j=1}^{n}(-1)^{k-j}a_{ij}\det(A_{kj})\,.$$

This shows that the off-diagonal entries of $A\tilde{A}$ vanish. The assertion on $\tilde{A}A$ is proved in the same way (or by applying what we have just proved to $A^\top$). □

16. The Determinant and the Symmetric Group

If $A = (a_{ij}) \in \mathrm{Mat}(n, F)$ and we recursively expand $\det(A)$ along the first row (say), then we end up with an expression that is a sum of products of n matrix entries with a plus or minus sign. For example:

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}$$

Since in the process of recursively expanding, the row and column of the entry we multiply with are removed, each product in the final expression has the form

$$a_{1,\sigma(1)}a_{2,\sigma(2)} \cdots a_{n,\sigma(n)}$$

where $\sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ is a bijective map, a so-called *permutation* of $\{1, 2, \ldots, n\}$. All these permutations together form a group, the symmetric group.

16.1. **Definition.** A *group* is a set G, together with a map $m : G \times G \to G$, usually written $m(x, y) = x \cdot y$ or xy, such that the following axioms are satisfied.

(1) For all $x, y, z \in G$, we have $(xy)z = x(yz)$ (associativity).
(2) There is $e \in G$ such that for all $x \in G$, we have $ex = xe = x$ (identity).
(3) For all $x \in G$, there is $x^{-1} \in G$ such that $xx^{-1} = x^{-1}x = e$ (inverse).

As usual, the identity element and the inverse of x are uniquely determined.

16.2. **Examples.**

(1) If F is a field, then we have the *additive group of* F $(F, +)$ and the *multiplicative group* $(F \setminus \{0\}, \cdot)$ *of* F. If V is a vector space, we have the additive group $(V, +)$. All these groups are *commutative* or *abelian*: they satisfy $xy = yx$ (or $x + y = y + x$ in additive notation).
(2) If X is a set, then the set of all bijective maps $f : X \to X$ forms a group, where the 'multiplication' is given by composition of maps. The identity element is given by id_X, the inverse by the inverse map. This group is called the *symmetric group* of X and denoted $\mathcal{S}(X)$. If $X = \{1, 2, \ldots, n\}$, we also write $\mathcal{S}_n$. If X has more than two elements, this group is *not* abelian (Exercise!). Note that $\#\mathcal{S}_n = n!$.
(3) Let F be a field, $n \geq 0$. The set of all invertible matrices in $\mathrm{Mat}(n, F)$ forms a group under matrix multiplication. This group is called the *general linear group* and denoted by $\mathrm{GL}_n(F)$.

16.3. **The 'Leibniz Formula'.** From the considerations above, we know that there is a formula, for $A = (a_{ij}) \in \mathrm{Mat}(n, F)$:

$$\det(A) = \sum_{\sigma \in \mathcal{S}_n} \varepsilon(\sigma)\, a_{1,\sigma(1)}a_{2,\sigma(2)} \cdots a_{n,\sigma(n)}$$

with a certain map $\varepsilon : \mathcal{S}_n \to \{\pm 1\}$. We call $\varepsilon(\sigma)$ the *sign* of the permutation σ.

It remains to determine the map ε. Let us introduce the *permutation matrix* $P(\sigma) = (\delta_{\sigma(i),j})$; this is the matrix whose entries at positions $(i, \sigma(i))$ are 1, and the other entries are 0. By specializing the formula above, we find that

$$\varepsilon(\sigma) = \det\big(P(\sigma)\big)\,.$$

We also have that

$$\varepsilon(\sigma\tau) = \det\big(P(\sigma\tau)\big) = \det\big(P(\tau)P(\sigma)\big) = \det\big(P(\tau)\big)\det\big(P(\sigma)\big) = \varepsilon(\sigma)\varepsilon(\tau)$$

so ε is a *group homomorphism*, and $\varepsilon(\sigma) = -1$ when σ is a *transposition*, i.e., a permutation that exchanges two elements and leaves the others fixed. Since every permutation can be written as a product of such transpositions (even transpositions of neighboring elements), this determines ε uniquely: write σ as a product of, say, k transpositions, then $\varepsilon(\sigma) = (-1)^k$. We can also give an 'explicit' formula.

16.4. **Proposition.** *Let $n \geq 0$, $\sigma \in \mathcal{S}_n$. Then*

$$\varepsilon(\sigma) = \prod_{1\leq i<j\leq n} \frac{\sigma(j) - \sigma(i)}{j - i}\,.$$

Note that for each pair $i < j$, we will either have $j - i$ or $i - j$ as a factor in the numerator (since σ permutes the 2-element subsets of $\{1, 2, \ldots, n\}$). Therefore the right hand side is ± 1. Since the factor $(\sigma(j) - \sigma(i))/(j - i)$ is negative if and only if $\sigma(j) < \sigma(i)$, the right hand side is $(-1)^m$, where m counts the number of pairs $i < j$ such that $\sigma(i) > \sigma(j)$.

Proof. Denote the right hand side by $\varepsilon'(\sigma)$. We first show that ε' is a group homomorphism.

$$\begin{aligned}\varepsilon'(\sigma\tau) &= \prod_{i<j} \frac{\sigma(\tau(j)) - \sigma(\tau(i))}{j - i} = \prod_{i<j} \frac{\sigma(\tau(j)) - \sigma(\tau(i))}{\tau(j) - \tau(i)} \, \frac{\tau(j) - \tau(i)}{j - i} \\ &= \prod_{i<j} \frac{\sigma(j) - \sigma(i)}{j - i} \prod_{i<j} \frac{\tau(j) - \tau(i)}{j - i} = \varepsilon'(\sigma)\varepsilon'(\tau)\end{aligned}$$

(Note that $\{\tau(i), \tau(j)\}$ runs exactly through the two-element subsets of $\{1, 2, \ldots, n\}$, and the ordering within the subset does not influence the value of the fraction.)

Next we show that $\varepsilon'(\tau) = -1$ when τ is a transposition. So assume that τ interchanges k and l, where $k < l$. Then $\varepsilon'(\tau) = (-1)^m$, where m is the number of pairs $i < j$ such that $\tau(j) > \tau(i)$. This is the case when (i) $i = k$ and $k < j < l$, or (ii) $k < i < l$ and $j = l$, or (iii) $i = k$ and $j = l$. Therefore $m = 2(l - k - 1) + 1$ is odd, and $\varepsilon'(\tau) = -1$.

As discussed above, these two properties determine ε' uniquely, hence $\varepsilon = \varepsilon'$. □

16.5. **Remark.** As a concluding remark, let us recall the following equivalences, which apply to an endomorphism f of a finite-dimensional vector space V.

$$f \text{ is an automorphism } \iff \ker(f) = \{0\} \iff \det(f) \neq 0\,.$$

17. Eigenvalues and Eigenvectors

We continue with the study of endomorphisms $f : V \to V$. Such an endomorphism is particularly easy to understand if it is just multiplication by a scalar: $f = \lambda \operatorname{id}_V$ for some $\lambda \in F$. Of course, these are very special (and also somewhat boring) endomorphisms. So we look for linear subspaces of V on which f behaves in this way.

17.1. **Definition.** Let V be an F-vector space, and let $f : V \to V$ be an endomorphism. Let $\lambda \in F$. If there exists a vector $0 \neq v \in V$ such that $f(v) = \lambda v$, then λ is called an *eigenvalue* of f, and v is called an *eigenvector* of f for the eigenvalue λ. In any case, the linear subspace $E_\lambda(f) = \{v \in V : f(v) = \lambda v\}$ is called the *λ-eigenspace of f*. (So that λ is an eigenvalue of f if and only if $E_\lambda(f) \neq \{0\}$.)

17.2. **Examples.**

(1) Let $V = \mathbb{R}^2$ and consider $f(x, y) = (y, x)$. Then 1 and -1 are eigenvalues of f, and $E_1(f) = \{(x, x) : x \in \mathbb{R}\}$, $E_{-1}(f) = \{(x, -x) : x \in \mathbb{R}\}$. The eigenvectors $(1, 1)$ and $(1, -1)$ form a basis of V, and the matrix of f relative to that basis is
$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

(2) Let $V = \mathcal{C}^\infty(\mathbb{R})$ be the space of infinitely differentiable functions on $\mathbb{R}$. Consider the endomorphism $D : f \mapsto f''$. Then every $\lambda \in \mathbb{R}$ is an eigenvalue, and all eigenspaces are of dimension two:
$$E_\lambda(D) = \begin{cases} L(x \mapsto 1, x \mapsto x) & \text{if } \lambda = 0 \\ L(x \mapsto e^{\mu x}, x \mapsto e^{-\mu x}) & \text{if } \lambda = \mu^2 > 0 \\ L(x \mapsto \sin \mu x, x \mapsto \cos \mu x) & \text{if } \lambda = -\mu^2 < 0 \end{cases}$$

Since matrices can be identified with linear maps, it makes sense to speak about eigenvalues and eigenvectors of a square matrix $A \in \operatorname{Mat}(n, F)$.

17.3. **The Characteristic Polynomial.** How can we find the eigenvalues (and eigenvectors) of a given endomorphism f, when V is finite-dimensional?

Well, we have
$$\begin{aligned} \lambda \text{ is an eigenvalue of } f &\iff \text{there is } 0 \neq v \in V \text{ with } f(v) = \lambda v \\ &\iff \text{there is } 0 \neq v \in V \text{ with } (f - \lambda \operatorname{id}_V)(v) = 0 \\ &\iff \ker(f - \lambda \operatorname{id}_V) \neq \{0\} \\ &\iff \det(f - \lambda \operatorname{id}_V) = 0 \end{aligned}$$

It is slightly more convenient to consider $\det(\lambda \operatorname{id}_V - f)$ (which of course vanishes if and only if $\det(f - \lambda \operatorname{id}_V) = 0$). If $A = (a_{ij}) \in \operatorname{Mat}(n, F)$ is a matrix associated to f relative to some basis of V, then
$$\det(\lambda \operatorname{id}_V - f) = \det(\lambda I_n - A) = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{vmatrix}.$$

Expanding the determinant, we find that

$$\det(\lambda I_n - A) = \lambda^n - \mathrm{Tr}(A)\lambda^{n-1} + \cdots + (-1)^n \det(A)\,.$$

This polynomial of degree n in λ is called the *characteristic polynomial* of A (or of f). We will denote it by $P_A(\lambda)$ (or $P_f(\lambda)$). By the discussion above, the eigenvalues of A (or of f) are exactly the roots (in F) of this polynomial.

17.4. **Examples.**

(1) Let us come back to the earlier example $f : (x, y) \mapsto (y, x)$ on $\mathbb{R}^2$. With respect to the canonical basis, the matrix is

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \text{so the characteristic polynomial is} \quad \begin{vmatrix} \lambda & -1 \\ -1 & \lambda \end{vmatrix} = \lambda^2 - 1$$

and has the two roots 1 and -1.

(2) Let us consider the matrix

$$A = \begin{pmatrix} 5 & 2 & -6 \\ -1 & 0 & 1 \\ 3 & 1 & -4 \end{pmatrix}.$$

What are its eigenvalues and eigenspaces? We compute the characteristic polynomial:

$$\begin{vmatrix} \lambda - 5 & -2 & 6 \\ 1 & \lambda & -1 \\ -3 & -1 & \lambda + 4 \end{vmatrix} = (\lambda - 5)\big(\lambda(\lambda + 4) - 1\big) + 2\big((\lambda + 4) - 3\big) + 6\big(-1 + 3\lambda\big)$$
$$= \lambda^3 - \lambda^2 - \lambda + 1 = (\lambda - 1)^2(\lambda + 1)\,.$$

The roots are 1 and -1; these are therefore the eigenvalues. To find (bases of) the eigenspaces, note that $E_\lambda(A) = \ker(A - \lambda I_3)$. For $\lambda = 1$, we have

$$A - I_3 = \begin{pmatrix} 4 & 2 & -6 \\ -1 & -1 & 1 \\ 3 & 1 & -5 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

(by elementary row operations), so $E_1(A)$ is generated by $(2, -1, 1)^\top$. For $\lambda = -1$, we obtain

$$A + I_3 = \begin{pmatrix} 6 & 2 & -6 \\ -1 & 1 & 1 \\ 3 & 1 & -3 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and $E_{-1}(A)$ is generated by $(1, 0, 1)^\top$.

17.5. **Diagonalizable Matrices and Endomorphisms.** If the canonical basis of F^n consists of eigenvectors of A, then A is *diagonal:*

$$A = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues corresponding to the basis vectors. More generally, if F^n has a basis consisting of eigenvectors of A, then changing from the canonical basis to that basis will make A diagonal: $P^{-1}AP$ is diagonal, where P is invertible. In this case, A is called *diagonalizable*. Similarly, if f is an endomorphism of a finite-dimensional F-vector space V, then f is *diagonalizable* if

V has a basis consisting of eigenvectors of f. The matrix associated to f relative to this basis is then diagonal.

The big question is now: when is a matrix or endomorphism diagonalizable?

This is certainly not always true. For example, in our second example above, we only found two linearly independent eigenvectors in F^3, and so there cannot be a basis of eigenvectors. Another kind of example is $f : (x, y) \mapsto (-y, x)$ on $\mathbb{R}^2$. The characteristic polynomial comes out as $\lambda^2 + 1$ and does not have roots in $\mathbb{R}$, so there are no eigenvalues and therefore no eigenvectors. (If we take $\mathbb{C}$ instead as the field of scalars, then we do have two roots $\pm i$, and f becomes diagonalizable.)

17.6. **Definition.** Let V be a finite-dimensional F-vector space, $f : V \to V$ an endomorphism and $\lambda \in F$. Then $\dim E_\lambda(f)$ is called the *geometric multiplicity* of the eigenvalue λ of f. (So the geometric multiplicity is positive if and only if λ is indeed an eigenvalue.)

17.7. **Lemma.** *Let V be an F-vector space and $f : V \to V$ an endomorphism. Let $\lambda_1, \dots, \lambda_m \in F$ be distinct, and for $i = 1, \dots, m$, let $v_i \in E_{\lambda_i}(f)$. If*

$$v_1 + v_2 + \cdots + v_m = 0\,,$$

then $v_i = 0$ for all i.

Proof. By induction on m. The case $m = 0$ (or $m = 1$) is trivial. So assume the claim is true for m, and consider the case with $m + 1$ eigenvalues. We apply the endomorphism $f - \lambda_{m+1} \mathrm{id}_V$ to the equation

$$v_1 + v_2 + \cdots + v_m + v_{m+1} = 0$$

and obtain (note $(f - \lambda_{m+1} \mathrm{id}_V)(v_{m+1}) = 0$)

$$(\lambda_1 - \lambda_{m+1})v_1 + (\lambda_2 - \lambda_{m+1})v_2 + \cdots + (\lambda_m - \lambda_{m+1})v_m = 0\,.$$

By induction, we find that $(\lambda_i - \lambda_{m+1})v_i = 0$ for all $1 \le i \le m$. Since $\lambda_i \ne \lambda_{m+1}$, this implies $v_i = 0$ for $1 \le i \le m$. But then we must also have $v_{m+1} = 0$. □

17.8. **Corollary.** *In the situation above, the union of bases of distinct eigenspaces of f is linearly independent.*

Proof. Consider a linear combination on the union of such bases that gives the zero vector. By the preceding lemma, each part of this linear combination that comes from one of the eigenspaces is already zero. Since the vectors involved there form a basis of this eigenspace, they are linearly independent, hence all the coefficients vanish. □

17.9. **Example.** We can use this to show once again that the power functions $f_n : x \mapsto x^n$ for $n \in \mathbb{N}_0$ are linearly independent as elements of the space P of polynomial functions on $\mathbb{R}$ (say). Namely, consider the endomorphism $D : P \to P$, $f \mapsto (x \mapsto xf'(x))$. Then $D(f_n) = nf_n$, so the f_n are eigenvectors of D for eigenvalues that are pairwise distinct, hence they must be linearly independent.

17.10. **Corollary.** *Let V be a finite-dimensional F-vector space and $f : V \to V$ an endomorphism. Then f is diagonalizable if and only if*

$$\sum_{\lambda \in F} \dim E_\lambda(f) = \dim V\,.$$

Proof. By Cor. 17.8, we always have "$\leq$". If f is diagonalizable, then there is a basis consisting of eigenvectors, and so we must have equality. Conversely, if we have equality, then the union of bases of the eigenspaces will be a basis of V, which consists of eigenvectors of f. □

17.11. **Proposition.** *Let V be an n-dimensional F-vector space and $f : V \to V$ an endomorphism. If $P_f(\lambda)$ has n distinct roots in F, then f is diagonalizable.*

Proof. In this case, there are n distinct eigenvalues $\lambda_1, \ldots, \lambda_n$. Therefore, $E_{\lambda_i}(f)$ is nontrivial for $1 \leq i \leq n$, which means that $\dim E_{\lambda_i}(f) \geq 1$. So

$$\dim V = n \leq \sum_{i=1}^{n} \dim E_{\lambda_i}(f) \leq \dim V\,,$$

and we must have equality. The result then follows by the previous corollary. □

The converse of this statement is false in general, as the identity endomorphism id_V shows (for $\dim V \geq 2$).

However, some statement in the converse direction is true. In order to state it, we need some preparations.

17.12. **Definition.** Let F be a field. The *polynomial ring* in x over F, $F[x]$, is an F-vector space with basis $1 = x^0, x = x^1, x^2, \ldots, x^n, \ldots$, on which a multiplication $F[x] \times F[x] \to F[x]$ is defined by the following two properties: (i) it is F-bilinear, and (ii) $x^m \cdot x^n = x^{m+n}$.

Written out, this means that

$$\begin{aligned}(a_n x^n + \cdots + a_1 x + a_0)(b_m x^m + \cdots + b_1 x + b_0) \\ = a_n b_m x^{n+m} + (a_n b_{m-1} + a_{n-1} b_m) x^{n+m-1} + \ldots \\ + \Big(\sum_{i+j=k} a_i b_j \Big) x^k + \cdots + (a_1 b_0 + a_0 b_1) x + a_0 b_0\,.\end{aligned}$$

It can then be checked that $F[x]$ is a *commutative ring with unit*, i.e., it satisfies the axioms of a field with the exception of the existence of multiplicative inverses.

If $p(x) = a_n x^n + \cdots + a_1 x + a_0 \in F[x]$ and $a_n \neq 0$, then p is said to have *degree* n; we write $\deg(p) = n$. In this case a_n is called the *leading coefficient* of $p(x)$; if $a_n = 1$, $p(x)$ is said to be *monic*.

For example, if V is an n-dimensional vector space and $f : V \to V$ is an endomorphism, then the characteristic polynomial $P_f(x)$ of f is monic of degree n.

17.13. **Theorem.** *Let $p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 \in F[x]$ be a monic polynomial, and let $\alpha \in F$. If $p(\alpha) = 0$, then there is a polynomial $q(x) = x^{n-1} + b_{n-2}x^{n-2} + \cdots + b_0$ such that $p(x) = (x - \alpha)q(x)$.*

Proof. If $\alpha = 0$, this is certainly true, since then $0 = p(0) = a_0$, and visibly $p(x) = xq(x)$. In general, we replace x by $x + \alpha$. Then the polynomial $\tilde{p}(x) = p(x + \alpha)$ is again monic of degree n, and $\tilde{p}(0) = p(\alpha) = 0$, so $\tilde{p}(x) = x\tilde{q}(x)$ with a monic polynomial $\tilde{q}$ of degree $n - 1$. Then

$$p(x) = \tilde{p}(x - \alpha) = (x - \alpha)\tilde{q}(x - \alpha) = (x - \alpha)q(x)\,,$$

where $q(x) = \tilde{q}(x - \alpha)$ is monic of degree $n - 1$. □

17.14. **Corollary and Definition.** *Let $p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 \in F[x]$ and $\alpha \in F$. Then there is a largest $m \in \mathbb{N}_0$ such that $p(x) = (x - \alpha)^m q(x)$ with a polynomial $q(x)$; we then have $q(\alpha) \neq 0$.*

This number m is called the *multiplicity* of the root α of p; we have $m > 0$ if and only if $p(\alpha) = 0$.

Proof. Write $p(x) = (x-\alpha)^m q(x)$ with m as large as possible. (Note that $\deg(p) = m + \deg(q)$, so $m \leq n$.) Then we must have $q(\alpha) \neq 0$, since otherwise we could write $q(x) = (x - \alpha)r(x)$, so $p(x) = (x - \alpha)^{m+1}r(x)$, contradicting our choice of m. □

Now we can make another definition.

17.15. **Definition.** Let V be a finite-dimensional F-vector space, $f : V \to V$ an endomorphism. Then the multiplicity of $\lambda \in F$ as a root of the characteristic polynomial $P_f(x)$ is called the *algebraic multiplicity* of the eigenvalue λ of f.

Note that the following statements are then equivalent.

(1) λ is an eigenvalue of f;
(2) the geometric multiplicity of λ is ≥ 1;
(3) the algebraic multiplicity of λ is ≥ 1.

We also know that the sum of the geometric multiplicities of all eigenvalues as well as the sum of the algebraic multiplicities of all eigenvalues are bounded by $\dim V$.

There is one further important relation between the multiplicities.

17.16. **Theorem.** *Let V be a finite-dimensional F-vector space, $f : V \to V$ an endomorphism, and $\lambda \in F$. Then the geometric multiplicity of λ as an eigenvalue of f is not larger than its algebraic multiplicity.*

Proof. We can choose a basis $v_1, \ldots, v_k, v_{k+1}, \ldots, v_n$ of V such that $v_1, \ldots, v_k$ form a basis of the eigenspace $E_\lambda(f)$; then k is the geometric multiplicity. The matrix associated to f relative to this basis then has the form

$$A = \begin{pmatrix} \lambda & 0 & \ldots & 0 & * & \ldots & * \\ 0 & \lambda & \ldots & 0 & * & \ldots & * \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & \lambda & * & \ldots & * \\ 0 & 0 & \ldots & 0 & * & \ldots & * \\ \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & * & \ldots & * \end{pmatrix} = \left(\begin{array}{c|c} \lambda I_k & B \\ \hline 0 & C \end{array}\right).$$

We then have

$$P_f(x) = \det(xI_n - A) = \det\left(\begin{array}{c|c} (x-\lambda)I_k & -B \\ \hline 0 & xI_{n-k} - C \end{array}\right)$$
$$= \det\big((x-\lambda)I_k\big)\det(xI_{n-k} - C) = (x-\lambda)^k q(x)$$

where $q(x)$ is some monic polynomial of degree $n-k$. We see that λ has multiplicity at least k as a root of $P_f(x)$. □

17.17. **Lemma.** *Let $f : V \to V$ be an endomorphism of an n-dimensional F-vector space V, and let P_f be its characteristic polynomial. Then the sum of the algebraic multiplicities of the eigenvalues of f is at most n; it is equal to n if and only if $P_f(x)$ is a product of linear factors $x - \lambda$ (with $\lambda \in F$).*

Proof. By Thm. 17.13, if λ is a root of P_f, we can write $P_f(x) = (x-\lambda)q(x)$ with a monic polynomial q of degree $n-1$. Continuing in this way, we can write

$$P_f(x) = (x-\lambda_1)^{m_1} \cdots (x-\lambda_k)^{m_k} q(x)$$

with a monic polynomial q that does not have roots in F and distinct elements $\lambda_1, \dots, \lambda_k \in F$. If $\mu \in F$, then

$$P_f(\mu) = (\mu-\lambda_1)^{m_1} \cdots (\mu-\lambda_k)^{m_k} q(\mu)\,,$$

so if $P_f(\mu) = 0$, then $\mu \in \{\lambda_1, \dots, \lambda_k\}$ (since $q(\mu) \neq 0$). Therefore the eigenvalues are exactly $\lambda_1, \dots, \lambda_k$, with multiplicities $m_1, \dots, m_k$, and

$$m_1 + m_2 + \cdots + m_k \leq m_1 + m_2 + \cdots + m_k + \deg(q) = n\,.$$

We have equality if and only if $\deg(q) = 0$, i.e., if and only if $q(x) = 1$; then

$$P_f(\mu) = (\mu-\lambda_1)^{m_1} \cdots (\mu-\lambda_k)^{m_k}$$

is a product of linear factors. □

17.18. **Corollary.** *Let V be a finite-dimensional F-vector space and $f : V \to V$ an endomorphism. Then f is diagonalizable if and only if*

(1) *$P_f(x)$ is a product of linear factors, and*
(2) *for each $\lambda \in F$, its geometric and algebraic multiplicities as an eigenvalue of f agree.*

Proof. By Cor. 17.10, f is diagonalizable if and only if the sum of the geometric multiplicities of all eigenvalues equals $n = \dim V$. By Thm. 17.16, this implies that the sum of the algebraic multiplicities is at least n; however it cannot be larger than n, so it equals n as well. This already shows that geometric and algebraic multiplicities agree. By Lemma 17.17, we also see that $P_f(x)$ is a product of linear factors.

Conversely, if we can write $P_f(x)$ as a product of linear factors, this means that the sum of the algebraic multiplicities is n. If the geometric multiplicities equal the algebraic ones, their sum must also be n, hence f is diagonalizable. □

17.19. **Remark.** If F is an *algebraically closed* field, for example $F = \mathbb{C}$, then condition (1) in the corollary is automatically satisfied (by definition!). However, condition (2) can still fail. It is then an interesting question to see how close we can get to a diagonal matrix in this case. This is what the *Jordan Normal Form Theorem* is about, which will be a topic for the second semester.

References

[BR1] T.S. Blyth and E.F. Robertson: *Basic Linear Algebra.* Springer Undergraduate Mathematics Series, 2002.

[BR2] T.S. Blyth and E.F. Robertson: *Further Linear Algebra.* Springer Undergraduate Mathematics Series, 2002.

[J] K. Jänich: *Linear Algebra.* Springer Undergraduate Texts in Mathematics, 1994

[KM] A. Kostrykin and Y. Manin: *Linear Algebra and Geometry.* Gordan and Breach, 1988.

Index of notation

Index