Linear Algebra

Notes for Math 115A(H)

Contents

Introduction

This document is meant to be a *free* set of notes supporting a 30-hour course of linear algebra, like the Math115A and Math115AH classes at UCLA. It presupposes some familiarity with solving linear equations (the Gaussian technique, a. k. a. Gauss-Jordan elimination method), some knowledge of vectors and matrices, etc, as provided in an entry-level linear algebra class, like Math33A at UCLA. This text addresses the reader as a student of one such linear algebra class.

All comments in footnote can be safely ignored. They provide general education to students interested in learning more mathematics later on.

We use little mathematical notation. They are recalled in Appendix A.

Some techniques of proof are gathered in Appendix C. In some sense, this course is a gateway to rigorous proofs. Some arguments are therefore spelled out in more detail than necessary, also to build this skill set. The result is that every given section may be a little longer than a 50-minute class, although the rough estimate should be *one section, one class*. In the case of some longer sections, it will hopefully be clear that two 50-minute classes are preferable.

Specifically for Math115A at UCLA, the sections marked with a * (for instance in the table-of-content) are usually done only in the honors class, Math115AH. Some of the sections should probably be taught by the TA, in discussion, *e.g.* the first section on fields.

Students should begin by refreshing their knowledge of vector and matrix calculus (addition, multiplication, and the geometric interpretation in Euclidean plane $\mathbb{R}^2$ and space $\mathbb{R}^3$) and the Gauss-Jordan technique of solving systems of linear equations in several variables, by row-reduction.

Acknowledgements: I am very thankful to many students who reported misprints in earlier versions. Special thanks go to Tom Hong and to Pablo S. Ocal for a very useful list of suggestions.

CHAPTER 1

Vector spaces

1.1. Fields

Linear algebra can be done over the real numbers $\mathbb{R}$, over the complex numbers $\mathbb{C}$, or over other collections of so-called '*scalars*'. Despite what you may think from the easy problems of a first-course linear algebra, you *cannot* do linear algebra with using only the *integers*

$$\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, 4, \dots\}.$$

The reason is that when you solve systems of linear equations, using the Gaussian elimination method, you need to *divide* by non-zero numbers, to create 'pivots', *i.e.* 1's at the beginning of a row. Therefore, the simplest set of 'scalars' for linear algebra are the *rational numbers*, *i.e.* our old friends the *fractions* of integers:

$$\mathbb{Q} = \Big\{ \frac{a}{b} \,\Big|\, a, b \in \mathbb{Z},\ b \neq 0 \Big\}. \tag{1.1.1}$$

Let us formalize the list of rules that those 'scalars' should verify in general.

1.1.2. Definition. A *field* is a triple $(\mathbb{F}, +, \cdot)$, often just written $\mathbb{F}$, consisting of a set $\mathbb{F}$ equipped with two operations $+$ and $\cdot$, called addition and multiplication

$$\begin{tikzcd}
\mathbb{F} \times \mathbb{F} \arrow[r, "+"] & \mathbb{F} & & \mathbb{F} \times \mathbb{F} \arrow[r, "\cdot"] & \mathbb{F} \\
(a, b) \arrow[r, mapsto] & a + b & & (a, b) \arrow[r, mapsto] & a \cdot b
\end{tikzcd}$$

respectively, and satisfying the following ten axioms:

(F1) Associativity of addition: $(a + b) + c = a + (b + c)$ for all $a, b, c \in \mathbb{F}$.
(F2) Zero: There is an element $0 \in \mathbb{F}$ such that $0 + a = a = a + 0$ for all $a \in \mathbb{F}$.
(F3) Commutativity of addition: $a + b = b + a$ for all $a, b \in \mathbb{F}$.
(F4) Opposite: For every $a \in \mathbb{F}$ there exists some $b \in \mathbb{F}$ such that $a + b = 0$.

So far, we only discussed addition. [1] Let us add multiplication to the mix:

(F5) Associativity of multiplication: $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all $a, b, c \in \mathbb{F}$.
(F6) One: There is an element $1 \in \mathbb{F}$ such that $1 \cdot a = a = a \cdot 1$ for all $a \in \mathbb{F}$.
(F7) Commutativity of multiplication: $a \cdot b = b \cdot a$ for all $a, b \in \mathbb{F}$.
(F8) Distributivity: $a \cdot (b + c) = a \cdot b + a \cdot c$ for all $a, b, c \in \mathbb{F}$.
(F9) Non-triviality: $0 \neq 1$.

So far, we have 'usual rules'. [2] The most important axiom of a field is:

(F10) Invertibility: For every $a \in \mathbb{F}$ non-zero, there exists $b \in \mathbb{F}$ such that $a \cdot b = 1$.

[1] A pair $(\mathbb{F}, +)$ satisfying (F1)-(F4) is what one calls an *abelian group*. The group is called abelian because of (F3).

[2] A triple $(\mathbb{F}, +, \cdot)$ satisfying (F1)-(F8) is what one calls a *commutative ring*. The ring is called commutative because of (F7). Axiom (F9) only excludes the trivial ring $A = 0$.

1.1.3. Remark. Let $a \in \mathbb{F}$ be a fixed element of a field $\mathbb{F}$. We know by (F4) that there exists $b \in \mathbb{F}$ such that $a+b=0$. Note that $b+a=0$ as well by commutativity. We claim that this b is *unique*. Indeed let $b, b' \in \mathbb{F}$ be such that $a+b=0$ and $a+b'=0$. Then compute $b+a+b'$. Technically, the latter means either side of

$$(b+a)+b' = b+(a+b')$$

which agree by associativity. The left-hand side is $0+b'=b'$ and the right-hand side is $b+0=b$. So $b'=b$. (This uniqueness holds in any abelian group.) A similar argument shows that the scalar 0 satisfying (F2) is unique.

1.1.4. Notation. The unique b satisfying $a+b=b+a=0$ is called *the opposite* of a and denoted $-a$. (Read "minus a", not "negative a", as $-a$ can be positive, as for $\mathbb{F}=\mathbb{R}$ and $a=-1$; or $\mathbb{F}$ might have no positive and negative, as for $\mathbb{F}=\mathbb{C}$.)

1.1.5. Exercise. Let $a \in \mathbb{F}$ be a non-zero element in a field $\mathbb{F}$. Show that the $b \in \mathbb{F}$ such that $a \cdot b = 1$ is unique.

1.1.6. Notation. Let $a \in \mathbb{F}$ be non-zero in a field. The unique b satisfying $a \cdot b = 1$ is called *the inverse* of a and denoted a^{-1}. (Read "a inverse".)

1.1.7. Exercise. Let $\mathbb{F}$ be a field. Prove the following:

(1) We have $0+0=0$ and $-0=0$. In fact, 0 is the only solution $x \in \mathbb{F}$ of $x+x=x$.
(2) We have $1 \cdot 1 = 1$ and $1^{-1}=1$ and 1 is the only solution $y \in \mathbb{F}$ of $y \cdot y = y$.
(3) For all $a, b \in \mathbb{F}$, we have $-(a+b)=(-a)+(-b)$ and $-(a \cdot b) = (-a) \cdot b = a \cdot (-b)$.
(4) For all $a, b \in \mathbb{F}$, we have $(a \cdot b)^{-1} = (a^{-1}) \cdot (b^{-1})$.
(5) If $a \cdot b = 0$ then either $a=0$ or $b=0$. Hence if $ab=ac$ and $a \neq 0$ then $b=c$.

It is high time to give examples. To prepare for it, the reader should refresh the notion of *equivalence relation* on a set X as recalled in Remark B.1.2.

1.1.8. Example. When counting in a 'normal' way, the 'first' field we encounter is the field of *rational numbers* $\mathbb{Q}$ of (1.1.1). Recall that the notation $\frac{a}{b}$ refers to an equivalence class, where we declare $\frac{a}{b} = \frac{a'}{b'}$ if $ab' = a'b$, for any $a, a', b, b' \in \mathbb{Z}$ with $b \neq 0$, $b' \neq 0$. In particular, for any $c \neq 0$ in $\mathbb{Z}$, we have $\frac{a}{b} = \frac{ac}{bc}$. Recall that we multiply fractions very easily: $\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}$ but that addition is slightly more tricky:

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

where we first put the fractions in the same denominator.

1.1.9. Exercise. Verify that the above $(\mathbb{Q}, +, \cdot)$ is indeed a field. Show that $\frac{a}{b} \neq 0$ if and only if $a \neq 0$. In that case, prove that $(\frac{a}{b})^{-1} = \frac{b}{a}$.

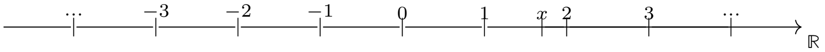
1.1.10. Remark. The more elementary triple $(\mathbb{Z}, +, \cdot)$ is an example of a commutative ring. It fails only (F10) in the list of Definition 1.1.2. Indeed, the element $2 \in \mathbb{Z}$ has no inverse. (The only two elements that admit an inverse are ±1.)

1.1.11. Remark. We shall often use the notation

$$\mathbb{N} = \{0, 1, 2, 3, \ldots\}$$

for the set of *natural numbers*. This set has addition and multiplication as in $\mathbb{Z}$ but it is not a ring since 1 has no opposite in $\mathbb{N}$. This set is nonetheless very useful for enumerations, *e.g.* when discussing sequences $(x_0, x_1, x_2, \ldots) = (x_n)_{n \in \mathbb{N}}$. We are following here the 'European' tradition of starting the count with 0, for people in Europe are born at 0 year old. American mathematicians often prefer the convention $\{1, 2, 3, \ldots\}$. We shall write $\mathbb{N}_{\geq 1}$ for the latter.

1.1.12. Example. Arguably the most common field used in applications is the field $\mathbb{R}$ of *real numbers*. A precise definition of $\mathbb{R}$ is not a banality. In a snapshot, we often picture $\mathbb{R}$ as the *real line*



But what does a real number x represented as above mean? It is between 1 and 2, a little closer to 2, but what else? One can actually build $\mathbb{R}$ from $\mathbb{Q}$ by a process called *completion*. One way to formalize this is as equivalence classes of Cauchy sequences $\underline{x} = (x_n)_{n\in\mathbb{N}}$ of rational numbers $x_n \in \mathbb{Q}$. A sequence is Cauchy if for every $\epsilon > 0$ arbitrary small in $\mathbb{Q}$ there exists a large enough index $n_\epsilon \gg 1$ such that $-\epsilon < x_\ell - x_m < \epsilon$ for all larger indices $\ell, m \geq n_\epsilon$. Two sequences $\underline{x}$ and $\underline{y}$ are called equivalent if $\underline{x} - \underline{y}$ converges to zero in $\mathbb{Q}$, in the usual sense. One can show that the set of equivalence classes is a field, and we call it $\mathbb{R}$. Examples of such sequences are the stationary ones $\underline{x} = (a)_{n\in\mathbb{N}}$ for some fixed $a \in \mathbb{Q}$ (that is, $x_n = a$ for all n). This defines an injection $\mathbb{Q} \hookrightarrow \mathbb{R}$, so we identify the rational numbers as a subfield of the real numbers

$$\mathbb{Q} \subset \mathbb{R}.$$

One can show that this inclusion is far from an equality: There are way more irrational real numbers than rational ones. Indeed, $\mathbb{Q}$ is *countable* (in bijection with $\mathbb{N}$) whereas $\mathbb{R}$ is not.

Alternatively, one can construct $\mathbb{R}$ as the set of so-called ‘Dedekind cuts’ in $\mathbb{Q}$. A ‘cut’ is a pair (A, B) where $\mathbb{Q} = A \cup B$, $A \cap B = \varnothing$, $A, B \neq \varnothing$ and A is closed under taking strictly small numbers (if $a \in A$ and $a' \in \mathbb{Q}$ is such that $a' < a$ then $a' \in A$) whereas B is closed under taking greater or equal numbers (if $b \in B$ and $b' \in \mathbb{Q}$ is such that $b \leq b'$ then $b' \in B$). After the fact, we understand those cuts as follows. Pick $x \in \mathbb{R}$ and define $A_x = \{ a \in \mathbb{Q} \mid a < x \} =]{-\infty}, x[\, \cap \mathbb{Q}$ and $B_x = \{ b \in \mathbb{Q} \mid x \leq b \} = [x, \infty[\, \cap \mathbb{Q}$ the complement (using here the square bracket notation for intervals, not the one with round parentheses). The cut (A_x, B_x) corresponds to $x \in \mathbb{R}$. The point is that conversely, any cut in $\mathbb{Q}$ defines a unique $x \in \mathbb{R}$. This approach is perhaps more elementary than Cauchy sequences and explains how $\mathbb{R}$ ‘fills in all the holes in $\mathbb{Q}$’.

In analysis, $\mathbb{R}$ is useful for the above completeness property and the consequences for Cauchy sequences (they converge) and continuous functions (the Intermediate Value Theorem). There are notoriously useful real numbers like π and e that do not belong to $\mathbb{Q}$.

For linear algebra, one could argue that $\mathbb{R}$ is somewhat imperfect, as it contains some roots, like $\sqrt{2}$ that is not rational, but not all roots. Indeed, $\mathbb{R}$ misses exactly the roots of negative numbers, most famously $\sqrt{-1}$. Creating a new imaginary world in which the latter exists leads us to the next example.

1.1.13. Exercise. Equip $\mathbb{R}^2 = \{ (x, y) \mid x, y \in \mathbb{R} \}$ with the following operations

$$(x, y) + (x', y') = (x + x', y + y') \quad \text{and} \quad (x, y) \cdot (x', y') = (x\,x' - y\,y' \,,\, x\,y' + y\,x').$$

(1) Show that this defines a field, with $0 = (0, 0)$ and $1 = (1, 0)$.
(2) Show that $i := (0, 1)$ satisfies $i^2 = -1$.

(3) Show that every element $z = (x, y)$ in this field can be written as

$$z = (x, 0) + (y, 0) \cdot i$$

and that the subset $\{ (x, 0) \,|\, x \in \mathbb{R} \}$ with the above operations is the same as the field $\mathbb{R}$ with its usual operations.

1.1.14. Example. In view of the above exercise, one usually defines the field of *complex* numbers as

$$\mathbb{C} = \{ x + y\,i \,|\, x, y \in \mathbb{R} \}$$

whose elements are given by pairs of real numbers (x, y), formally written $x + y\,i$ for an 'imaginary' number i, with x the 'real part' and y the 'imaginary part'. The operations are forced by associativity, commutativity and distributivity and one extra rule that breaks our real intuition:

$$i^2 = -1.$$

In particular, multiplication becomes

$$(x + y\,i) \cdot (x' + y'i) = (xx') + (xy' + yx')\,i + (yy')\,i^2 = (xx' - yy') + (xy' + yx')\,i$$

recovering Exercise 1.1.13. The part $y = 0$ identifies with $\mathbb{R}$ so we write

$$\mathbb{R} \subset \mathbb{C}.$$

The difference is of course that $i \notin \mathbb{R}$. (While, as sets, $\mathbb{R}$ and $\mathbb{C}$ have same cardinal.)

There are also amusing little fields. At first, we might think that these other fields just appear because Definition 1.1.2 is imperfect. After all, we gave a list of axioms; every $(\mathbb{F}, +, \cdot)$ satisfying those axioms must now be accepted as a field! Upon further investigation, these fields are interesting in their own right and play a very important role in algebra, notably in number theory.

1.1.15. Example. Let $p \in \mathbb{N}$ be a *prime number*, *i.e.* one that admits exactly two divisors 1 and p. The first primes are $2, 3, 5, 7, 11, 13, \ldots$. Consider the set with p elements written with the following decoration

$$\mathbb{F}_p = \{\bar{0}, \bar{1}, \ldots, \overline{p-1}\} = \{ \bar{i} \,|\, 0 \leq i < p \}.$$

For instance $\mathbb{F}_2 = \{\bar{0}, \bar{1}\}$, $\mathbb{F}_3 = \{\bar{0}, \bar{1}, \bar{2}\}, \ldots, \mathbb{F}_{13} = \{\bar{0}, \bar{1}, \bar{2}, \bar{3}, \bar{4}, \bar{5}, \bar{6}, \bar{7}, \bar{8}, \bar{9}, \bar{10}, \bar{11}, \bar{12}\}$, etc. Let us define operations on $\mathbb{F}_p$ as follows: Add and multiply as in $\mathbb{Z}$ (without the decoration $\bar{\ }$) and if the result is larger than p, reduce modulo p. For instance, in $\mathbb{F}_{13}$ we have $\bar{3} + \bar{4} = \bar{7}$ but $\bar{8} + \bar{9} = \bar{17} = \bar{4}$ since $17 = 13 + 4 \equiv 4$ modulo 13. Still in $\mathbb{F}_{13}$, we have $\bar{6} \cdot \bar{8} = \bar{48} = \bar{9}$ since $48 = 39 + 9 \equiv 9$ modulo 13.

The formal definition is that $\mathbb{F}_p$ is the set of equivalence classes $\mathbb{Z}/p$ of integers $i \in \mathbb{Z}$ modulo p, that is, under the equivalence relation $i \sim j$ if $i - j$ is divisible by p. This requires a few verifications that we omit.

1.1.16. Exercise. Show that $\mathbb{F}_p$ is a field, with $0_{\mathbb{F}_p} = \bar{0}$ and $1_{\mathbb{F}_p} = \bar{1}$. [Do this at least for $p = 2, 3$.] In $\mathbb{F}_7$, compute the inverse of every non-zero element.

1.1.17. Remark. As the notation $\mathbb{Z}/p$ suggests, one can similarly define $\mathbb{Z}/n$ for any n and obtain a commutative ring. Declare $i \sim j$ if n divides $i - j$ and let $\mathbb{Z}/n = \mathbb{Z}/\sim$ be the set of equivalence classes. If we write $\bar{i} = \{ i + kn \,|\, k \in \mathbb{Z} \}$ for the equivalence class of i then the operations are very easy to describe: $\bar{i} + \bar{j} = \overline{i + j}$ and $\bar{i} \cdot \bar{j} = \overline{i\,j}$. One only needs to check that they make sense.

Such a ring $\mathbb{Z}/n$ will be a field if and only if n is prime. For instance, if $n = 6$, we have $\bar{2} \cdot \bar{3} = \bar{6} = 0$ in $\mathbb{Z}/6$ although $\bar{2} \neq 0$ and $\bar{3} \neq 0$. Compare Exercise 1.1.7 (5).

1.1.18. Remark. Those fields like $\mathbb{F}_p$ feel quite different from the usual $\mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$. The notable difference is that if you add $1_\mathbb{F} = \bar{1}$ with itself p times in $\mathbb{F}_p$ you get zero: $1_\mathbb{F} + 1_\mathbb{F} + \cdots + 1_\mathbb{F} = 0$ whereas this would never happen in $\mathbb{Q}$ and beyond.

Fields where $n \cdot 1_\mathbb{F} = 1_\mathbb{F} + 1_\mathbb{F} + \cdots + 1_\mathbb{F}$ (sum of n terms) is never zero for $n > 0$ are called *characteristic zero fields* ($n = 0$ being the only solution to $n \cdot 1_\mathbb{F} = 0$).

If your field $\mathbb{F}$ has the property that $n \cdot 1_\mathbb{F} = 0$ for some $n > 0$, then you can *prove* that the smallest such n is a prime number p and that all solutions of $n \cdot 1_\mathbb{F} = 0$ are the multiples of p, as is the case in $\mathbb{F}_p$. Indeed, let $p > 0$ be the smallest integer $n > 0$ such that $n \cdot 1_\mathbb{F} = 0$ in $\mathbb{F}$. If $p = a \cdot b$ in $\mathbb{Z}$ with $a, b > 0$ then $(a \cdot 1_\mathbb{F}) \cdot (b \cdot 1_\mathbb{F}) = p \cdot 1_\mathbb{F} = 0$ and $\mathbb{F}$ being a field forces one of $(a \cdot 1_\mathbb{F})$ or $(b \cdot 1_\mathbb{F})$ to be zero in $\mathbb{F}$ by Exercise 1.1.7. As p was the smallest such that $p \cdot 1_\mathbb{F} = 0$ and as $a, b \leq p$, the property $(a \cdot 1_\mathbb{F}) = 0$ would force $a = p$ and therefore $b = 1$, and similarly the property $(b \cdot 1_\mathbb{F}) = 0$ would force $b = p$ and therefore $a = 1$. In short, we have shown that p is prime number: p has only 1 and p as divisors. The argument to show that all other n such that $n \cdot 1_\mathbb{F} = 0$ are multiples of that smallest p is direct from Euclidean division: Write $n = q \cdot p + r$ with $0 \leq r < p$. We have $0 = n \cdot 1_\mathbb{F} = q \cdot p \cdot 1_\mathbb{F} + r \cdot 1_\mathbb{F} = q \cdot 0 + r \cdot 1_\mathbb{F}$. But $r < p$ so it cannot be non-zero otherwise p would not be the smallest. So $r = 0$, that is, p divided n.

In that situation, that is, when $p \cdot 1_\mathbb{F} = 0$ for a prime p, the field $\mathbb{F}$ is called of *characteristic p*.

In summary, every field has a characteristic, a prime $p > 0$ or zero, depending on whether $\left\{ n \in \mathbb{Z} \,\middle|\, n \cdot 1_\mathbb{F} = 0 \right\}$ is $p \cdot \mathbb{Z}$ or $\{0\}$. We write $\mathrm{char}(\mathbb{F}) = p$ or $\mathrm{char}(\mathbb{F}) = 0$ respectively. ([3])

1.2. Vector spaces

For the entire section, $\mathbb{F}$ is a fixed field, called the 'ground field'. The elements of $\mathbb{F}$ are called *scalars*. In first approximation, the reader can think only of one of the three standard fields $\mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$.

The fundamental notion of linear algebra is that of 'vector space'. Its importance lies in the large catalogue of examples of vector spaces that appear throughout algebra and analysis. We give several of the most elementary examples after the definition and many more in the next sections.

1.2.1. Definition. A *vector space* over $\mathbb{F}$ (or $\mathbb{F}$-vector space) is a triple $(V, +, \cdot)$ consisting of a set V equipped with two operations

$$V \times V \xrightarrow{\;+\;} V \qquad\qquad \mathbb{F} \times V \xrightarrow{\;\cdot\;} V$$

$$(x, y) \longmapsto x + y \qquad\qquad (a, x) \longmapsto a \cdot x$$

satisfying Axioms (VS1)-(VS8) below. The first operation (addition) is internal to V. The second operation involves both V and the field $\mathbb{F}$; it is called the *(scalar) action* of $\mathbb{F}$ on V. To be clear, the latter means that we assign an element denoted $a \cdot x$ in V to every choice of $a \in \mathbb{F}$ and $x \in V$. The first four axioms do not involve the action ([4]):

(VS1) Associativity of addition: $(x + y) + z = x + (y + z)$ for all $x, y, z \in V$.

[3] A finite field must have positive characteristic. You will learn later that there are finite fields other than the $\mathbb{F}_p$. And a field of characteristic p does not have to be a finite field.

[4] As in Definition 1.1.2, the first four axioms are saying that $(V, +)$ is an abelian group.

(VS2) Zero vector: There exists $0 \in V$ such that $0 + x = x = x + 0$ for all $x \in V$.
(VS3) Commutativity of addition: $x + y = y + x$ for all $x, y \in V$.
(VS4) Opposite: For every $x \in V$ there exists some $y \in V$ such that $x + y = 0$.

Then the action of $\mathbb{F}$ comes in and should be compatible with addition:

(VS5) Associativity of action: $(a \cdot b) \cdot x = a \cdot (b \cdot x)$ for all $a, b \in \mathbb{F}$ and $x \in V$.
(VS6) Action of one: $1 \cdot x = x$ for all $x \in V$.
(VS7) Distributivity (right): $a \cdot (x + y) = a \cdot x + a \cdot y$ for all $a \in \mathbb{F}$ and $x, y \in V$.
(VS8) Distributivity (left): $(a + b) \cdot x = a \cdot x + b \cdot x$ for all $a, b \in \mathbb{F}$ and $x \in V$.

We often just write V instead of $(V, +, \cdot)$ and we shall soon write $a\,x$ instead of $a \cdot x$. By definition, a *vector* is simply an element of a vector space, *i.e.* any x in V.

1.2.2. Definition. A *real vector space* is a vector space over the field $\mathbb{F} = \mathbb{R}$ of real numbers. A *complex vector space* is one over the field $\mathbb{F} = \mathbb{C}$ of complex numbers.

1.2.3. Remark. As in Remark 1.1.3, given a vector $x \in V$, the $y \in V$ satisfying (VS4) is unique. We call it the opposite of x and write it $-x$.

1.2.4. Remark. Let us not mix up the two zeros discussed so far. If necessary, the zero of the field $\mathbb{F}$, from (F2), can be denoted $0_{\mathbb{F}}$ whereas the zero of the vector space V, from (VS2), can be denoted 0_V. They are related in the obvious way:

1.2.5. Exercise. Let V be a vector space over $\mathbb{F}$. Show the following:

(1) We have $0_{\mathbb{F}} \cdot x = 0_V$ for all $x \in V$.
(2) We have $(-1) \cdot x = -x$ for all $x \in V$. (In fact, Axiom (VS4) is superfluous.)
(3) If $a \cdot x = 0_V$ for $x \in V$ and $a \in \mathbb{F}$ then either $a = 0_{\mathbb{F}}$ or $x = 0_V$. Similarly, if $a \in \mathbb{F}$ is non-zero and $x, y \in V$ satisfy $a \cdot x = a \cdot y$ then $x = y$.

Let us discuss examples. The smallest vector space is:

1.2.6. Example. Let $V = \{0\}$ be the set with only one element. Then V is a vector space with the only possible addition and multiplication ($0 + 0 = 0$ and $a \cdot 0 = 0$ for all $a \in \mathbb{F}$). Axioms (VS1)-(VS8) amount to check $0 = 0$ many times.

1.2.7. Exercise. Why can't we take any non-empty set V, pick some element $0 \in V$ and define the operations as $x + y = 0$ and $a \cdot x = 0$ for all $x, y \in V$ and all $a \in \mathbb{F}$? Which of the axioms would hold and which ones would fail?

1.2.8. Example. The reader pondering the resemblance between the axioms for a field (Definition 1.1.2) and for a vector space (Definition 1.2.1) might be helped, or confused further, by the case where we take $V = \mathbb{F}$ and use as action the multiplication of $\mathbb{F}$. Then Axioms (VS1)-(VS8) are actually a subset of Axioms (F1)-(F10).

It might be easier to see the above example as the case $n = 1$ of the following more general class of examples. Indeed, $\mathbb{F} = \mathbb{F}^1$ is the *line* over $\mathbb{F}$ and we can speak of n-space $\mathbb{F}^n$ for any $n \geq 1$.

1.2.9. Example. The classical example from introductory linear algebra is n-space $V = \mathbb{F}^n$ for some $n \geq 1$. (You might have seen this only for $\mathbb{F} = \mathbb{R}$, or for $n = 2$ and 3, *i.e.* for Euclidean plane and space.) The set V is that of n-tuples

$$\mathbb{F}^n = \big\{\, \underline{x} = (x_1, x_2, \ldots, x_n) \,\big|\, x_1, \ldots, x_n \in \mathbb{F} \,\big\}.$$

Addition and scalar action are defined *componentwise*, meaning:

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

$$(\cdot) \qquad a\cdot(x_1,x_2,\dots,x_n) = (a\cdot x_1\,,\, a\cdot x_2,\dots,a\cdot x_n)$$

for all vectors $(x_1,x_2,\dots,x_n),(y_1,y_2,\dots,y_n)\in\mathbb{F}^n$ and all scalars $a\in\mathbb{F}$. For the action, the 'inner' expressions $a\cdot x_i$ on the right-hand side refer to multiplication in the field $\mathbb{F}$. So the action of $\mathbb{F}$ on $V=\mathbb{F}^n$ relies on the multiplication in $\mathbb{F}$.

1.2.10. Notation. It is sometimes more convenient to write the elements of $\mathbb{F}^n$ in columns $\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, in particular when we do matrix multiplication. This writing artifice has no deep meaning but should not be used to create (or hide) confusion.

1.2.11. Exercise. Verify that the above $(\mathbb{F}^n,+,\cdot)$ is indeed an $\mathbb{F}$-vector space.

1.2.12. Exercise. Refresh your geometric understanding: Pick two vectors in $\mathbb{R}^2$, draw them on the plane and add them. Do the same exercise with scalar action.

We can generalize the above. We refer to Appendix A for notation.

1.2.13. Notation. Let I be a set (of 'indices'). Let

$$\mathbb{F}^I = \big\{\, f\colon I\to\mathbb{F} \,\big|\, f \text{ is a function}\,\big\}$$

be the set of *all* functions $f\colon I\to\mathbb{F}$. We shall make $\mathbb{F}^I$ into a vector space over $\mathbb{F}$. Let us emphasize: *One* vector in this (possible very big) vector space $\mathbb{F}^I$ will be a *function* from the set I to our field $\mathbb{F}$. We add functions and multiply them by scalars ('constants') in the usual way:

(+) For every two functions $f,g\in\mathbb{F}^I$ we define the function $(f+g)\in\mathbb{F}^I$, that is, $f+g\colon I\to\mathbb{F}$, by the formula $(f+g)(i)=f(i)+g(i)$ for all $i\in I$:

$$\begin{aligned} (f+g)\colon \qquad I &\longrightarrow \qquad \mathbb{F} \\ i &\longmapsto \quad f(i)+g(i). \end{aligned}$$

$(\cdot)$ For every function $f\in\mathbb{F}^I$ and scalar $a\in\mathbb{F}$, we define the function $(a\cdot f)\in\mathbb{F}^I$, that is, $a\cdot f\colon I\to\mathbb{F}$, by the formula $(a\cdot f)(i)=a\cdot f(i)$ for all $i\in I$:

$$\begin{aligned} (a\cdot f)\colon \qquad I &\longrightarrow \qquad \mathbb{F} \\ i &\longmapsto \quad a\cdot f(i). \end{aligned}$$

In both cases, the explicit formulas involve the operations $+$ and $\cdot$ in the field $\mathbb{F}$.

1.2.14. Exercise. Verify that $\mathbb{F}^I$ is an $\mathbb{F}$-vector space with the above operations.

There are two elementary prototypes of the above construction $\mathbb{F}^I$.

1.2.15. Example. First, for $n\geq 1$ and the indexing set $I=\{1,2,\dots,n\}$, the vector space $\mathbb{F}^I$ is nothing but our old friend $\mathbb{F}^n$. Indeed, to give a function $f\colon\{1,2,\dots,n\}\to\mathbb{F}$ is simply to give the n-tuple $(f(1),f(2),\dots,f(n))$. [In fact, if you try to define 'tuples' carefully, this is *the* definition!] The operations $+$ and $\cdot$ on functions amounts exactly to the componentwise operations of Example 1.2.9.

1.2.16. Example. Secondly, if $I=\mathbb{N}$ then the vector space $\mathbb{F}^I$ is that of sequences in $\mathbb{F}$. Again, the data of a function $f\colon\mathbb{N}=\{0,1,2,\ldots\}\to\mathbb{F}$ consists exactly of the data of a sequence of scalars $f(0),f(1),f(2),\ldots$ that we usually denote $x_0,x_1,x_2,\ldots$ or $(x_i)_{i\in\mathbb{N}}$ setting $x_i=f(i)$ for every index i. Some sources will have special notation for the $\mathbb{F}$-vector space of sequences in $\mathbb{F}$ but we simply write $\mathbb{F}^{\mathbb{N}}$.

1.2.17. Example. Here is a mildly silly variation on the above theme. Let $p, q \geq 1$ and take $I = \big\{ (i,j) \in \mathbb{N} \times \mathbb{N} \,\big|\, 1 \leq i \leq p\,,\ 1 \leq j \leq q \big\}$ the set of pairs of indices (i,j) that we use to describe the entries of a $(p \times q)$-matrix. Then $\mathbb{F}^I$ is precisely the set

$$\mathrm{M}_{p\times q}(\mathbb{F}) = \Big\{ \begin{pmatrix} a_{1,1} & \cdots & a_{1,q} \\ \vdots & \ddots & \vdots \\ a_{p,1} & \cdots & a_{p,q} \end{pmatrix} \,\Big|\, a_{i,j} \in \mathbb{F},\ \text{for all } 1 \leq i \leq p\,,\ 1 \leq j \leq q \Big\}$$

of $p \times q$ matrices with entries in $\mathbb{F}$. As before, we just need to write the data of the function $I \to \mathbb{F}$ as a table with p rows and q columns and $a_{i,j} = f(i,j)$ in the i-th row and j-th column. For instance, for $p = 2$ and $q = 3$ then

$$\mathbb{F}^I = \Big\{ \begin{pmatrix} f(1,1) & f(1,2) & f(1,3) \\ f(2,1) & f(2,2) & f(2,3) \end{pmatrix} \,\Big|\, f\colon I \to \mathbb{F} \Big\}.$$

This example is just a rearrangement of $\mathbb{F}^n$ from Example 1.2.9 with $n = pq$. Again, it gives the usual addition and scalar multiplication of matrices.

So we have already lots of examples, inside linear algebra. But there are more examples of vector spaces of the form $\mathbb{F}^I$ in analysis, where one uses such spaces of functions for I even bigger than the above $\mathbb{N}$.

1.2.18. Example. Let us take the ground field to be $\mathbb{F} = \mathbb{R}$ the real numbers. If we take $I = \mathbb{R}$ then $\mathbb{F}^I = \mathbb{R}^{\mathbb{R}}$ is the set of *all* functions $f\colon \mathbb{R} \to \mathbb{R}$. Alternatively, if we take $I = [a,b]$ an interval, with $a < b$, then our space $\mathbb{F}^I = \mathbb{R}^{[a,b]}$ becomes the space of functions $f\colon [a,b] \to \mathbb{R}$. The addition and scalar multiplication of Notation 1.2.13 are the ‘usual’ addition of functions and multiplication of a function by a scalar (a constant function).

Of course, in complex analysis, one would replace $\mathbb{R}$ by $\mathbb{C}$ and construct similar spaces $\mathbb{C}^X$ for many possible X, like a piece of the complex plane for instance. Such variations are limitless.

In both cases, one is often interested in continuous functions, or differentiable functions, holomorphic ones, etc. We return to this in the next section.

1.2.19. Remark. Saying that the function set $\mathbb{F}^I$ is a vector space does not keep track of another operation, namely the usual multiplication of functions, given by $(f \cdot g)(i) = f(i) \cdot g(i)$ for all $i \in I$. In fact, $\mathbb{F}^I$ is what one calls an $\mathbb{F}$-*algebra*: both an $\mathbb{F}$-vector space and a ring, in a compatible way.

In light of the previous remark, we do not need our functions f to land in the ground field $\mathbb{F}$. In fact any vector space would suffice:

1.2.20. Exercise. Let V be an $\mathbb{F}$-vector space and I be a set. Define an ‘obvious’ structure of $\mathbb{F}$-vector space on V^I, the set of functions $f\colon I \to V$ from I to V.

Let us give one more class of examples, coming from algebra.

1.2.21. Definition. Recall that a *polynomial* in a variable X, with coefficients in a field $\mathbb{F}$, consists of an expression of the form

$$P(X) = a_0 + a_1 X + a_2 X^2 + \cdots + a_n X^n = \sum_{i=0}^{n} a_i X^i \qquad (1.2.22)$$

where $n \in \mathbb{N}$ and $a_0, a_1, \ldots, a_n \in \mathbb{F}$. (Note that X^0 is omitted, that is, we read $X^0 = 1$.) We do not think of $P(X)$ as a function of X but as a way to record the $n+1$ scalars $a_0, \dots, a_n$. Note that the number n is not fixed, nor capped in advance. To be precise, we identify the following two expressions for any $m > n$

$$a_0 + a_1 X + \cdots + a_n X^n = a_0 + a_1 X + \cdots + a_n X^n + \mathbf{0}\, X^{n+1} + \cdots + \mathbf{0}\, X^m$$

that is, we ignore the top terms as long as their coefficient is zero. If $a_n \neq 0$, we say that the polynomial P as in (1.2.22) has *degree* n and write $\deg(P) = n$. The set of all polynomials, for all possible degrees, is denoted

$$\mathbb{F}[X] = \big\{\, a_0 + a_1 X + a_2 X^2 + \cdots + a_n X^n \,\big|\, n \in \mathbb{N} \text{ and } a_0, a_1, \dots, a_n \in \mathbb{F} \,\big\}.$$

1.2.23. Exercise. Show that the usual addition of polynomials and the obvious $\mathbb{F}$-action given by $b \cdot \big(\sum_{i=0}^{n} a_i X^i\big) = \sum_{i=0}^{n} (b \cdot a_i) X^i$ define a structure of $\mathbb{F}$-vector space on $\mathbb{F}[X]$.

1.2.24. Remark. We emphasize that the variable X plays no real role in the above vector space structure. It is simply a way to 'distinguish' a_i from a_j for $i \neq j$. The mathematical reason for the variable X comes in the multiplication of polynomials, where we use the obvious distributivity and commutativity rules together with $X^i \cdot X^j = X^{i+j}$, resulting in the not entirely obvious multiplication

$$\big(\sum_{i=0}^{m} a_i X^i\big) \cdot \big(\sum_{j=0}^{n} b_j X^j\big) = \sum_{k=0}^{m+n} \big(\sum_{i+j=k} a_i \cdot b_j \big)\; X^k.$$

But we do not use multiplication of polynomials for the moment.

There are also abstract way of producing new vector spaces out of old ones.

1.2.25. Exercise. Let V_1 and V_2 be two vector spaces over the same field $\mathbb{F}$. Denote by $V_1 \oplus V_2$ the cartesian product $V_1 \times V_2$

$$V_1 \times V_2 = \big\{\, (x_1, x_2) \,\big|\, x_1 \in V_1,\ x_2 \in V_2 \,\big\}$$

with the componentwise addition $(x_1, x_2) + (y_1, y_2) = (x_1 + y_1\, ,\, x_2 + y_2)$ and scalar multiplication $a \cdot (x_1, x_2) = (a \cdot x_1\, ,\, a \cdot x_2)$. Show that this is still a vector space over $\mathbb{F}$. It is called the *(external) direct sum* of V_1 and V_2 and denoted

$$V_1 \oplus V_2.$$

1.2.26. Remark. This operation can be repeated (and is associative), to define the direct sum

$$V_1 \oplus V_2 \oplus \cdots \oplus V_n$$

of any number of vector spaces $V_1, \dots, V_n$ over the same field $\mathbb{F}$. It consists of n-tuples $(x_1, \dots, x_n)$ where each x_i is a vector in V_i, with componentwise operations.

1.3. Subspaces

As before, $\mathbb{F}$ is a fixed ground field.

In the previous section, we saw many examples of vector spaces. However, most of them look essentially like $\mathbb{F}^I$ for a set I. One big strength of the concept of vector space is the ability to handle new vector spaces that appear *inside* a given one, even if the given one appear 'easy', like $\mathbb{F}^I$. For instance, if you think of the usual Euclidean space $V = \mathbb{R}^3$ then there are many planes in it. The 'obvious' ones containing two of the three axes, $\mathbb{R} \times \mathbb{R} \times \{0\}$, $\mathbb{R} \times \{0\} \times \mathbb{R}$ and $\{0\} \times \mathbb{R} \times \mathbb{R}$, which are essentially just $\mathbb{R}^2$ in disguise. But there are many more, in many directions, like the solutions (x_1, x_2, x_3) of $ax_1 + bx_2 + cx_3 = 0$ for $a, b, c \neq 0$. Such a plane in $\mathbb{R}^3$ is also a vector space in its own right, even if it is not $\mathbb{R}^2$ in a naive way.

1.3.1. Definition. Let V be a vector space over $\mathbb{F}$. A *subspace* is a subset $W \subseteq V$ satisfying the following three axioms:

(SS1) $0 \in W$.

(SS2) $W + W \subseteq W$, meaning that whenever two vectors $x, y \in W$ belong to our subset, their sum $x + y$ (in V) still belongs to our subset: $x + y \in W$.

(SS3) $\mathbb{F} \cdot W \subseteq W$, meaning that whenever a vector $x \in W$ belongs to our subset, every scalar multiple $a \cdot x$ (in V) still belongs to our subset: $a \cdot x \in W$ for all $a \in \mathbb{F}$.

1.3.2. Remark. Before giving examples, let us make one comment. Axiom (SS1) simply says that W cannot be taken empty. Indeed, in the presence of (SS3) if $x \in W$ then $0 \cdot x \in W$, which gives (SS1). However, to do that, we need at least some $x \in W$. The subset $\varnothing \subset V$ would satisfy (SS2) and (SS3) but not (SS1).

There are always two trivial subspaces, whose verification is very easy:

1.3.3. Example. For any vector space V, the subset $\{0\}$ is a subspace.

1.3.4. Example. For any vector space V, the subset V itself is a subspace.

1.3.5. Exercise. In $V = \mathbb{F}^1$, there are only the above two subspaces.

1.3.6. Example. In your introductory linear algebra, you surely saw, and perhaps proved, the following two facts (that we shall prove again later):

(1) In the real vector space $\mathbb{R}^2$, there are three types of subspaces: $\{0\}$, lines $L \subset \mathbb{R}^2$ through the origin, and $\mathbb{R}^2$ itself.

(2) In the real vector space $\mathbb{R}^3$, there are four types of subspaces: $\{0\}$, lines through the origin, planes through the origin, and $\mathbb{R}^3$ itself.

1.3.7. Exercise. Let $A \in \mathrm{M}_{p\times q}(\mathbb{F})$ be a matrix. Think of $\mathbb{F}^q$ as column vectors. Show that $\mathrm{Ker}(A) = \left\{ \underline{x} \in \mathbb{F}^q \,\middle|\, A \cdot \underline{x} = 0 \right\}$ is a subspace of $\mathbb{F}^q$.

1.3.8. Remark. The above says that the solutions of a homogeneous (meaning whose constant terms are zero) system of p equations in q variables $x_1, \ldots, x_q$

$$\left\{\begin{array}{ccccccccc} a_{11}\,x_1 & + & a_{12}\,x_2 & + & \cdots & + & a_{1q}\,x_q & = & 0 \\ a_{21}\,x_1 & + & a_{22}\,x_2 & + & \cdots & + & a_{2q}\,x_q & = & 0 \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ a_{p1}\,x_1 & + & a_{p2}\,x_2 & + & \cdots & + & a_{pq}\,x_q & = & 0 \end{array}\right.$$

defines a subspace of $\mathbb{F}^q$, no matter how the matrix of coefficients $A = (a_{ij})_{i,j} \in \mathrm{M}_{p\times q}(\mathbb{F})$ is chosen.

There are many examples of such 'linear conditions', sometimes less explicit.

1.3.9. Exercise. Let $n \geq 1$ and consider the $\mathbb{F}$-vector space $V = \mathrm{M}_{n\times n}(\mathbb{F})$ of square matrices of size n in $\mathbb{F}$, as in Example 1.2.17. Prove that the following are subspaces or give an explicit counter-example to one of the Axioms (SS1)-(SS3).

(1) The subset of *symmetric* matrices $\left\{ A \in \mathrm{M}_{n\times n}(\mathbb{F}) \,\middle|\, A^t = A \right\}$ where A^t denotes the transpose $(A^t)_{ij} = A_{ji}$.

(2) The *antisymmetric* (or skew-symmetric) matrices $\left\{ A \in \mathrm{M}_{n\times n}(\mathbb{F}) \,\middle|\, A^t = -A \right\}$.

(3) The *upper-triangular* matrices $\{\, A \in \mathrm{M}_{n\times n}(\mathbb{F}) \,|\, A_{ij} = 0 \text{ for all } i > j \,\}$.

(4) The *diagonal* matrices $\{\, A \in \mathrm{M}_{n\times n}(\mathbb{F}) \,|\, A_{ij} = 0 \text{ for all } i \neq j \,\}$.

(5) The *strictly lower-triangular* matrices $\{\, A \in \mathrm{M}_{n\times n}(\mathbb{F}) \,|\, A_{ij} = 0 \text{ for all } i \leq j \,\}$.

(6) The *orthogonal* matrices $\{\, A \in \mathrm{M}_{n\times n}(\mathbb{F}) \,|\, A^t \cdot A = I_n \,\}$, where I_n is the *identity* $(n \times n)$-matrix $I_n = (\delta_{ij})_{ij}$ using the Kronecker symbol of Notation A.1.1.

The following observation is simple but important:

1.3.10. Proposition. *Let $W \subseteq V$ be a subspace of a vector space over $\mathbb{F}$. Then W is itself a vector space over the same field $\mathbb{F}$, with the operations inherited from V.*

Proof. By (SS2) and (SS3) we can define $x + y$ and $a \cdot x$ for all $x, y \in W$ and $a \in \mathbb{F}$ the way they are defined in V. The point being that the output still belongs to W. So, for instance, the sum is not an operation $W \times W \to V$ but really $W \times W \to W$. Axioms (VS1)-(VS8) are easy to prove: For each of them, some equality must be verified in W. Each such equality makes sense because W is a subspace. And it holds true because it does in V. Only Axiom (VS2) and (VS4) require the *existence* of some vector in W, respectively the existence of zero, which is guaranteed by (SS1), and the existence of the opposite that follows from $-x = (-1) \cdot x$ and (SS3). Details are left as a good exercise for beginners. □

To see examples of subsets that are not subspaces, we can of course use the ones we learned in Euclidean plane.

1.3.11. Example. The subset $W = \{\, (x, y) \in \mathbb{R}^2 \,|\, x, y \geq 0 \,\}$ is not a subspace of real $\mathbb{R}^2$. [Draw a picture.] Indeed, it satisfies (SS1) and (SS2) but not (SS3) for negative scalars: $(-1) \cdot (1, 1) = (-1, -1)$ is out of W although $(1, 1) \in W$.

1.3.12. Example. The subset $W = \{\, (x, y) \in \mathbb{R}^2 \,|\, x \cdot y = 0 \,\}$ is not a subspace of real $\mathbb{R}^2$. [Draw a picture.] Indeed, it satisfies (SS1) and (SS3) but not (SS2). For instance $(1, 0) + (0, 1) = (1, 1)$ is out of W although $(1, 0)$ and $(0, 1)$ belong to W.

1.3.13. Exercise. Fix $d \in \mathbb{N}$ and consider $V = \mathbb{F}[X]$ the $\mathbb{F}$-vector space of polynomials in one variable X. Show that the subset $W = \{\, P \in \mathbb{F}[X] \,|\, \deg(f) \leq d \,\}$ of polynomials of degree at most d (including $P = 0$) is a subspace of V. Show that the polynomials of degree exactly d do *not* form a subspace.

1.3.14. Example. Let $a < b$ in $\mathbb{R}$ and $V = \mathbb{R}^{[a,b]} = \{f \colon [a, b] \to \mathbb{R}\}$ the real vector space of functions from the interval $[a, b]$ to $\mathbb{R}$. Then $W = \{\, f \in V \,|\, f \text{ is contnuous} \,\}$ is a subspace of V. Indeed, the zero function is continuous, the sum of continuous functions remains continuous and multiplying a continuous function by a constant gives another continuous function.

A very important example of (sub)space is the following:

1.3.15. Example. Let I be a set and consider the subset of $\mathbb{F}^I$ (Notation 1.2.13) of those functions $f \colon I \to \mathbb{F}$ that are zero almost everywhere, *i.e.* except for finitely many 'indices' $i \in I$. Formally, we let

$$\mathbb{F}^{(I)} = \{\, f \in \mathbb{F}^I \,|\, \text{the set } \{\, i \in I \,|\, f(i) \neq 0 \,\} \text{ is finite} \,\}.$$

For instance, if I is finite, there is no condition. In formula, $\mathbb{F}^{(I)} = \mathbb{F}^I$ when I is finite. When I is infinite, the two are different, as for instance non-zero constant functions belong to $\mathbb{F}^I$ but not to $\mathbb{F}^{(I)}$.

1.3.16. Exercise. Verify that $\mathbb{F}^{(I)}$ is a subspace of $\mathbb{F}^I$.

1.3.17. Definition. The above $\mathbb{F}$-vector space $\mathbb{F}^{(I)}$ is called the *free vector space* on the set I (or on the 'basis' I).

1.3.18. Example. For instance, for $I = \mathbb{N}$, the free vector space $\mathbb{F}^{(\mathbb{N})}$ consists of sequences $\underline{a} = (a_i)_{i\in\mathbb{N}}$ of scalars such that there exists n (depending on $\underline{a}$) such that $a_i = 0$ for all $i > 0$. Let me repeat: The n can vary for different vectors a. Giving such a vector $\underline{a} = (a_0, a_1, \ldots, a_n, 0, 0, \ldots)$ is *the same information* as giving a polynomial $a_0 + a_1 X + \cdots + a_n X^n$. As an exercise, the reader should compare the addition and $\mathbb{F}$-actions on $\mathbb{F}^{(\mathbb{N})}$ and $\mathbb{F}[X]$ under this identification. (We shall make more precise sense of such 'identifications' in Chapter 2.)

1.3.19. Exercise (Operations on subspaces)**.** Let $\{W_i\}_{i\in I}$ be a collection of subspaces of an $\mathbb{F}$-vector space V. Consider their intersection, and their union.

(1) Show that $\cap_{i\in I} W_i = \{\, v \in V \,|\, v \in W_i \text{ for all } i \in I \,\}$ is still a subspace of V.

(2) Show that $\cup_{i\in I} W_i = \{\, v \in V \,|\, v \in W_i \text{ for some } i \in I \,\}$ is not a subspace of V in general. Give a counterexample already for two subspaces.

(3) Suppose that we have only finitely many indices $I = \{1, 2, \ldots, n\}$. Show that

$$W_1 + W_2 + \cdots + W_n = \sum_{i=1}^{n} W_i := \{\, \sum_{i\in I} w_i \,|\, w_i \in W_i \,\}$$

is a subspace of V, which is the smallest subspace containing W_i for all $i \in I$.

(4) Describe $W_1 + W_2$ in an example where $W_1 \cup W_2$ is not a subspace; see (2).

(5) Construct $\sum_{i\in I} W_i$ satisfying the conclusion of (3) when I is infinite.

1.4. Linear combinations and span

The following basic notion will accompany us throughout the course.

1.4.1. Definition. Let V be an $\mathbb{F}$-vector space. Let $x_1, \ldots, x_n \in V$ be n vectors (repetition and $n = 0$ allowed). A *linear combination* of $x_1, \ldots, x_n$ is any vector in V of the form

$$a_1\, x_1 + \cdots + a_n\, x_n = \sum_{i=1}^{n} \; a_i\, x_i$$

for any choice of n scalars $a_1, \ldots, a_n \in \mathbb{F}$. [The right-hand side is just a short of the left-hand side.] Let me repeat: To say that a vector $v \in V$ *is* a linear combination of some given $x_1, \ldots, x_n$ means that you can find $a_1, \ldots, a_n \in \mathbb{F}$ such that when you compute $a_1\, x_1 + \cdots + a_n\, x_n$ you find v. By convention, when $n = 0$, the empty sum means 0. ([5])

1.4.2. Exercise. Let $x_1 = (1, 2, 3)$ and $x_2 = (4, 5, 6)$ in $\mathbb{R}^3$. Show explicitly that $v = (7, 8, 9)$ is a linear combination of x_1 and x_2.

1.4.3. Proposition. *Let $W \subseteq V$ be a subset of an $\mathbb{F}$-vector space V. The following are equivalent:*

(i) *W is a subspace of V (Definition 1.3.1).*

(ii) *W is closed under forming linear combination: For every $n \geq 0$ and for every $x_1, \ldots, x_n \in W$ and every $a_1, \ldots, a_n \in \mathbb{F}$ we have $a_1\, x_1 + \cdots + a_n\, x_n \in W$.*

[5] Suppose nobody gives you money. How much do you receive? So indeed $\sum_{\varnothing} = 0$.

PROOF. Suppose (i) and let us prove (ii). The case $n = 0$ is (SS1). For $n \geq 1$, given $x_1, \dots, x_n \in W$ and $a_1, \dots, a_n \in \mathbb{F}$, we have $a_i x_i \in W$ for all i by (SS3). Since W is closed under finite sums, by (SS2) and induction on n (see Exercise C.3.2), we conclude that $a_1 x_1 + \cdots + a_n x_n \in W$.

The converse (ii)⇒(i) is easy: Axioms (SS1)-(SS3) are special cases of (ii). □

We can generalize to an infinite collection of vectors G, but with some care since infinite sums $\sum_{g \in G} a_g \cdot g$ are not allowed.

1.4.4. Definition. Let $G \subseteq V$ be a set of vectors in a fixed $\mathbb{F}$-vector space V. We say that $v \in V$ is a *linear combination* of vectors in G if there exist $n \geq 0$ and $x_1, \dots, x_n \in G$ and $a_1, \dots, a_n \in \mathbb{F}$ such that $a_1 \, x_1 + \cdots + a_n \, x_n = v$. We denote by

$$\begin{aligned} \operatorname{Span}(G) &= \left\{ v \in V \,\middle|\, v \text{ is a linear combination of vectors in } G \right\} \\ &= \left\{ a_1 \, x_1 + \cdots + a_n \, x_n \,\middle|\, n \in \mathbb{N},\ a_1, \dots, a_n \in \mathbb{F} \text{ and } x_1, \dots, x_n \in G \right\} \end{aligned}$$

the subset of all linear combinations of vectors in G. We call it the subset *spanned* by G or *generated* by G. By convention $\operatorname{Span}(\varnothing) = \{0\}$, following the rule that an empty sum is zero.

1.4.5. Remark. It is clear that if $G \subseteq G'$ then by definition $\operatorname{Span}(G) \subseteq \operatorname{Span}(G')$.

1.4.6. Exercise. Give an example of $\mathbb{F}$-vector space $V$and two (finite) subsets G and G' of V such that $\operatorname{Span}(G) \subseteq \operatorname{Span}(G')$ and yet G is not contained in G'. Even give an example where G has strictly more elements that G'.

1.4.7. Proposition. *For any $G \subseteq V$ the subset* $\operatorname{Span}(G)$ *is a subspace of V. More precisely,* $\operatorname{Span}(G)$ *is the smallest subspace of V that contains G: If $W \subseteq V$ is a sub*space *that contains a set G then W also contains its span:* $\operatorname{Span}(G) \subseteq W$.

PROOF. For (SS1), we have $0 \in \operatorname{Span}(G)$ by convention (case $n = 0$). The subset $\operatorname{Span}(G)$ clearly satisfies (SS2) since the sum of two linear combinations

$$a_1 \, x_1 + \cdots + a_n \, x_n \ + \ b_1 \, y_1 + \cdots + b_m \, y_m$$

is just another (longer) linear combination of $n + m$ vectors $x_1, \dots, x_n, y_1, \dots, y_m$. Similarly, $\operatorname{Span}(G)$ satisfies (SS3) since a scalar multiple of a linear combination

$$b \cdot \sum_{i=1}^{n} a_i \, x_i = \sum_{i=1}^{n} (b \cdot a_i) \, x_i$$

is just another linear combination of the same n vectors $x_1, \dots, x_n$. And clearly $G \subseteq \operatorname{Span}(G)$ since $v = 1 \cdot v$, for any $v \in G$, that is, for $n = 1$ and $a_1 = 1$ and $x_1 = v$. In summary, $\operatorname{Span}(G)$ is indeed a subspace of V that contains G.

Let now W be any subspace of V containing G and let us show that it contains $\operatorname{Span}(G)$. We need to show that any linear combination of vectors $x_1, \dots, x_n$ in G still belongs to W. This is immediate from Proposition 1.4.3. □

1.4.8. Exercise. With notation as in Proposition 1.4.7, show that $\operatorname{Span}(G)$ is the intersection of all subspaces W of V that contain G:

$$\operatorname{Span}(G) = \bigcap_{\substack{W \text{ subspace of } V \\ \text{such that } G \subseteq W}} W. \tag{1.4.9}$$

Proposition 1.4.7 gives us an abundance of subspaces: Pick any subset $G \subseteq V$ at random and look at what it spans. There is possibly a massive 'waste' in this operation, as one easily realizes. For instance, $\mathrm{Span}(\mathbb{R}^3) = \mathbb{R}^3$ but $\mathrm{Span}(\{e_1, e_2, e_3\}) = \mathbb{R}^3$ as well for the usual canonical basis vectors $e_1 = (1, 0, 0), e_2 = (0, 1, 0), e_3 = (0, 0, 1)$. So different G's can span the same subspace, some much smaller than others. In particular, adding to G a vector that is already spanned by G will not make the spanned subspace any bigger. And the converse is true:

1.4.10. Proposition. *Let V be an $\mathbb{F}$-vector space. Let G' and G'' be two subsets of V. Then the following are equivalent.*

(i) *G' and $G' \cup G''$ span the same subspace:* $\mathrm{Span}(G') = \mathrm{Span}(G' \cup G'')$.

(ii) *Every vector v in G'' is already a linear combination of elements of G', that is, $G'' \subseteq \mathrm{Span}(G')$.*

Proof. Suppose (i). Then $G'' \subseteq G' \cup G'' \subseteq \mathrm{Span}(G' \cup G'') = \mathrm{Span}(G')$ by (i).

Suppose (ii). So $\mathrm{Span}(G')$ contains G'', and it always contains G', hence it contains the set $G = G' \cup G''$. By Proposition 1.4.7, applied to the subspace $W = \mathrm{Span}(G')$ and the set G, we see that $\mathrm{Span}(G')$ must contain the span of G, which reads $\mathrm{Span}(G' \cup G'') \subseteq \mathrm{Span}(G')$. As the other inclusion $\mathrm{Span}(G') \subseteq \mathrm{Span}(G' \cup G'')$ is obvious (see Remark 1.4.5), we get equality (i). □

1.4.11. Example. In the real vector space $V = \mathbb{R}^3$, we have by Exercise 1.4.2 that $\mathrm{Span}\big((1, 2, 3), (4, 5, 6)\big) = \mathrm{Span}\big((1, 2, 3), (4, 5, 6), (7, 8, 9)\big)$.

Using Exercise 1.3.19, we can also treat the above as follows:

1.4.12. Exercise. Let $G', G'' \subseteq V$ be subsets. Prove that $\mathrm{Span}(G' \cup G'') = \mathrm{Span}(G') + \mathrm{Span}(G'')$ where the right-hand side uses the notation $W' + W''$ for the two sub*spaces* $W' = \mathrm{Span}(G')$ and $W'' = \mathrm{Span}(G'')$.

It is important to reverse the order of appearance: Sometimes we know a subspace W and we look for a set G such that $\mathrm{Span}(G) = W$. As subspaces are vector spaces (Proposition 1.3.10), it suffices to understand this notion for V itself.

1.4.13. Definition. Let V be an $\mathbb{F}$-vector space. We say that $G \subseteq V$ is a *set of generators* of V if $\mathrm{Span}(G) = V$. Unpacking the meaning of Span(...), this means that for every vector $v \in V$ we can find some $n \in \mathbb{N}$, some $x_1, \ldots, x_n \in G$ and some $a_1, \ldots, a_n \in \mathbb{F}$ such that $a_1 x_1 + \cdots + a_n x_n = v$.

1.4.14. Remark. Given $W \subseteq V$ a subspace, a set of generators of W is a subset $G \subseteq W$ such that $\mathrm{Span}(G) = W$. Nothing mysterious: Just pay attention that the generators $g \in G$ cannot be taken outside W.

1.4.15. Remark. It is customary to speak of "a generator g" to mean some $g \in G$ when G is a *given* set of generators. Note however that the phrase "a generator g" does not make sense on its own and should not be used to mean that there exists some generating set G containing g. Indeed, we always have $\mathrm{Span}(V) = V$ and $g \in V$, so every g would be a generator. If one says that g alone is a generator of V, without specifying a G, we mean that $V = \mathrm{Span}(\{g\})$ is a line (or zero).

Here is a silly observation but it is better to make it now rather than in the middle of a more complicated proof.

1.4.16. Lemma. *Let G be a subset of an $\mathbb{F}$-vector space V. Let $v \in \mathrm{Span}(G)$. Then there exists a number $m \geq 0$ and m vectors $y_1, \ldots, y_m \in G$ all distinct (meaning $y_i \neq y_j$ whenever $i \neq j$) and m scalars $b_1, \ldots, b_m \in \mathbb{F}$ all non-zero such that $v = b_1 y_1 + \cdots + b_m y_m$.*

Proof. This is almost the definition of $\mathrm{Span}(G)$ from Definition 1.4.4, which says that we can write v as the linear combination $v = a_1 x_1 + \cdots + a_n x_n$ for $x_1, \ldots, x_n \in G$ and $a_1, \ldots, a_n \in \mathbb{F}$. We just need to arrange the x_i to be distinct and the a_i to be non-zero. For instance, $4x + 0y + 5z - x$ is simply $3x + 5z$. Making sure that all a_i are non-zero is trivial: We can remove the term $0 \cdot x_i = 0$ from the linear combination. Similarly, if $x_i = x_j$ for $i \neq j$ we can regroup two terms $a_i x_i + a_j x_j = (a_i + a_j) x_i$ into one. By induction on the number of terms, we get the statement. Note that we might have removed *all* terms in this process (only when $v = 0$). But that is alright: that is why we allow $m = 0$ in the statement. □

1.5. Linear dependence and independence

Again, the field $\mathbb{F}$ is fixed for the entire section and so is the $\mathbb{F}$-vector space V.

We saw in the previous section that $\mathrm{Span}(G) = \mathrm{Span}(G')$ can happen for very different G and G'. We want to express what it means for G to be optimal in that respect. This leads to the conceptual definition of linear dependence. We will immediately rephrase this definition in the more traditional form, that readers are used to. We temporarily stop using the letter G that reminded us of 'generators' as the subsets we discuss now do not have to generate the whole space.

1.5.1. Definition. Let $L \subseteq V$ be a subset of an $\mathbb{F}$-vector space. We say that L is *linearly dependent* if there exists a strictly smaller $L' \subsetneq L$ such that $\mathrm{Span}(L) = \mathrm{Span}(L')$. In other words, L spans whatever it spans ($\mathrm{Span}(L)$) in a suboptimal way: We can do as well with a strictly smaller L' contained in L.

We say that L is *linearly independent* when it is not linearly dependent. This means that L does *not* admit a proper subset $L' \subsetneq L$ with $\mathrm{Span}(L') = \mathrm{Span}(L)$.

1.5.2. Example. The empty subset $L = \varnothing$ is always linearly independent as it has no proper subset anyway. Note that $\mathrm{Span}(\varnothing) = \{0\}$ (by our convention).

1.5.3. Example. We also have $\mathrm{Span}(\{0\}) = \{0\}$ and therefore $L = \{0\}$ is linearly dependent in any V: The smaller $L' = \varnothing$ spans the same subspace! In fact, any L that contains 0 is linearly dependent as $\mathrm{Span}(L) = \mathrm{Span}(L \smallsetminus \{0\})$.

1.5.4. Notation. Recall that $X \smallsetminus Y$ means "the set of elements of X that are not in Y". We shall use this simple notation several times below.

1.5.5. Example. Continuing our running Example 1.4.11, the three vectors $x_1 = (1, 2, 3)$, $x_2 = (4, 5, 6)$, $x_3 = (7, 8, 9)$ form a linearly dependent set $L = \{x_1, x_2, x_3\}$ in $V = \mathbb{R}^3$ since $\mathrm{Span}(L) = \mathrm{Span}(L')$ for $L' = \{x_1, x_2\}$.

1.5.6. Example. Let us do a small variation to illustrate one point. Let x_1, x_2, x_3 be as above and $x_4 = (1, 0, 0)$ for instance. Then $\mathrm{Span}(x_1, x_2, x_3, x_4) = V$ and $x_1, \ldots, x_4$ are linearly dependent. Indeed, $V = \mathrm{Span}(x_1, x_2, x_4) = \mathrm{Span}(x_1, x_3, x_4) = \mathrm{Span}(x_2, x_3, x_4)$ *but not* $\mathrm{Span}(x_1, x_2, x_3)$ that is a plane. So L linearly dependent does not mean that you can take *any* $L' \subsetneq L$ and span the same subspace. Beware of quantifiers: "there exists" and "for all"!

Here is the reformulation when expanding the meaning of Span(...).

1.5.7. Proposition. *Let $L \subseteq V$ be a subset of an $\mathbb{F}$-vector space. The following are equivalent:*

(i) *The set L is linearly dependent in V.*
(ii) *There exists a number $n \geq 1$ and n distinct vectors $x_1, \ldots, x_n \in L$ and n non-zero scalars $a_1, \ldots, a_n \in \mathbb{F}$ such that $a_1x_1 + \cdots + a_nx_n = 0$.*

PROOF. (i)⇒(ii): Let $L' \subsetneq L$ be a proper subset such that $\operatorname{Span}(L) = \operatorname{Span}(L')$. Since the complement $L \smallsetminus L'$ is non-empty, we can pick $v \in L \smallsetminus L'$. Then v belongs to $L \subseteq \operatorname{Span}(L) = \operatorname{Span}(L')$. By Lemma 1.4.16 for $v \in \operatorname{Span}(L')$, we can write

$$v = b_1y_1 + \cdots + b_my_m$$

for $m \geq 0$ and m vectors $y_1, \ldots, y_m$ in L' all distinct and m scalars $b_1, \ldots, b_m \in \mathbb{F}$ all non-zero. (We might have $m = 0$ but that is fine.) Rewrite the above as

$$1 \cdot v + (-b_1)y_1 + \cdots + (-b_m)y_m = 0.$$

This gives a relation as announced in (ii), for $n = m + 1 \geq 1$, for the n vectors $v, y_1, \ldots, y_m$ in L and for the n non-zero scalars $1, -b_1, \ldots, -b_m$. Indeed, since the $y_1, \ldots, y_m$ are all distinct, the only possible repetition among $v, y_1, \ldots, y_m$ would be to have $v = y_i$ for some i but that is excluded since $y_i \in L'$ and $v \notin L'$.

(ii)⇒(i): This is the place where we use for the first time that $\mathbb{F}$ is a field! Suppose that $x_1, \ldots, x_n \in L$ are all distinct and $a_1x_1 + \cdots + a_nx_n = 0$ for $a_1, \ldots, a_n \in \mathbb{F}$ all non-zero. Since $n \geq 1$, our equation can be rewritten as

$$a_1 \cdot x_1 = (-a_2) \cdot x_2 + \cdots + (-a_n) \cdot x_n = \sum_{i=2}^{n} (-a_i) \cdot x_i.$$

Since $a_1 \neq 0$, we can divide this equation by a_1 (multiply by a_1^{-1}) and get

$$x_1 = a_1^{-1} \cdot \sum_{i=2}^{n} (-a_i) \cdot x_i = \sum_{i=2}^{n} \big(- a_1^{-1}\, a_i \big) \cdot x_i.$$

Whatever the coefficients on the right-hand side are, we just proved that x_1 is a linear combination of $x_2, \ldots, x_n$. Since x_1 is distinct from all of $x_2, \ldots, x_n$, we have shown that x_1 is a linear combination of vectors $x_2, \ldots, x_n$ in $L \smallsetminus \{x_1\}$. This reads $\{x_1\} \subseteq \operatorname{Span}(L \smallsetminus \{x_1\})$. If we baptize $L' = L \smallsetminus \{x_1\}$ and $L'' = \{x_1\}$, we have established that $L'' \subseteq \operatorname{Span}(L')$. By Proposition 1.4.10 this implies that $\operatorname{Span}(L') = \operatorname{Span}(L' \cup L'')$. In our case, this means $\operatorname{Span}(L \smallsetminus \{x_1\}) = \operatorname{Span}(L)$ since $L' \cup L'' = (L \smallsetminus \{x_1\}) \cup \{x_1\} = L$. In summary, $\operatorname{Span}(L)$ can be spanned by the strictly smaller subset $L' = L \smallsetminus \{x_1\}$, which is the meaning of L linearly dependent (Definition 1.5.1). Hence we proved (i). □

We can rewrite the above in the 'traditional' formulations of linear independence that is sometimes so hard for some students to remember.

1.5.8. Corollary. *Let $L \subseteq V$ be a subset of an $\mathbb{F}$-vector space. The following are equivalent:*

(i) *The set L is linearly independent in V.*
(ii) *For every $n \geq 1$ distinct vectors $x_1, \ldots, x_n \in L$ and n scalars $a_1, \ldots, a_n \in \mathbb{F}$ such that $a_1x_1 + \cdots + a_nx_n = 0$, we have $a_1 = \ldots = a_n = 0$.*

Proof. This is contraposition on Proposition 1.5.7. Suppose (i): L is linearly independent. Let $x_1, \dots, x_n \in L$ be $n \geq 1$ distinct vectors and $a_1, \dots, a_n \in \mathbb{F}$ be such that $a_1x_1 + \cdots + a_nx_n = 0$. If *ab absurdo* $a_1, \dots, a_n$ are not all zero, up to removing the zero terms, we can assume that *all* a_i are non-zero (see Lemma 1.4.16). By Proposition 1.5.7 (ii)⇒(i), this means L is linearly dependent, a contradiction. So all $a_i = 0$ as claimed.

Conversely, suppose (ii). If *ab absurdo* L was linearly dependent then by Proposition 1.5.7 (i)⇒(ii) we could find $n \geq 1$ distinct vectors in L and $a_1, \dots, a_n \in \mathbb{F}$ all non-zero such that $a_1x_1 + \cdots + a_nx_n = 0$. This contradicts our (ii). So L linearly dependent is absurd, meaning it is linearly independent. □

When L is finite, we can simply write the above as follows.

1.5.9. Corollary. *Let $n \geq 1$ and $x_1, \dots, x_n$ be n distinct vectors in an $\mathbb{F}$-vector space V. Then the set $\{x_1, \dots, x_n\}$ is linearly dependent if and only if there exists $a_1, \dots, a_n \in \mathbb{F}^n$* not all zero *such that $a_1x_1 + \cdots + a_nx_n = 0$.*

Proof. Apply Proposition 1.5.7 to $L = \{x_1, \dots, x_n\}$. □

1.5.10. Remark. For the sake of beginners, let us make the above argument explicit. There is one difference between Proposition 1.5.7 and Corollary 1.5.9. In (ii) of the proposition, we say the $a_1, \dots, a_n$ are 'all non-zero' but in the corollary we say that they are 'not all zero'. This is of course different. But if you apply Proposition 1.5.7 carefully to $L = \{x_1, \dots, x_n\}$, part (ii) only says that one can find $m \geq 1$ distinct vectors *among* $\{x_1, \dots, x_n\}$, perhaps not all of them, and a linear combination of those m vectors with only non-zero coefficients that yields zero. Such a linear combination can be expanded into a linear combination of all x_i by just taking $a_i = 0$ for the other i. For instance, if $n = 5$ and $m = 2$ and we have $a_2\,x_2 + a_4\,x_4 = 0$ with $a_2 \neq 0$ and $a_4 \neq 0$ then we also have $a_1\,x_1 + \cdots + a_5\,x_5 = 0$ if we set $a_1 = a_3 = a_5 = 0$. We then only know that the $a_1, a_2, \dots, a_n$ are *not all zero* (in our example, a_2 and a_4 remain non-zero) but we cannot say they are all non-zero anymore.

The contraposition of Corollary 1.5.9 is:

1.5.11. Corollary. *Let $n \geq 1$ and $x_1, \dots, x_n$ be n distinct vectors in an $\mathbb{F}$-vector space V. Then the set $\{x_1, \dots, x_n\}$ is linearly independent if and only if the only solution $(a_1, \dots, a_n) \in \mathbb{F}^n$ of the equation $a_1x_1 + \cdots + a_nx_n = 0$ is the trivial solution $(a_1, \cdots, a_n) = (0, \dots, 0)$.*

Proof. Apply Corollary 1.5.8 to $L = \{x_1, \dots, x_n\}$. □

1.5.12. Example. By Example 1.4.11, the vectors $x_1 = (1, 2, 3)$ and $x_2 = (4, 5, 6)$ and $x_3 = (7, 8, 9)$ are linearly dependent in $V = \mathbb{R}^3$, over $\mathbb{F} = \mathbb{R}$. Indeed, they satisfy the non-trivial relation $1 \cdot x_1 + (-2) \cdot x_2 + 1 \cdot x_3 = 0$.

Let us make the following easy observations:

1.5.13. Corollary. *Let $L_1 \subseteq L_2$ be two subsets of an $\mathbb{F}$-vector space V.*

(1) *If L_1 is linearly dependent then the larger L_2 is linearly dependent.*
(2) *If L_2 is linearly independent then the smaller L_1 is linearly independent.*

Proof. Part (1) is direct from Proposition 1.5.7. Indeed, in that proposition for $L = L_1$, (i)⇒(ii) gives us $n \geq 1$ distinct vectors $x_1, \dots, x_n \in L_1$ and non-zero scalars $a_1, \ldots, a_n \in \mathbb{F}$ such that $\sum_{i=1}^n a_i\, x_i = 0$. Those x_i are also in L_2. Using the same Proposition 1.5.7 (ii)⇒(i) for $L = L_2$ tells us that L_2 is linearly dependent.

Part (2) is direct by contraposition: We saw in (1) that 'L_1 linearly dependent' implies 'L_2 linearly dependent'. So 'L_2 not linearly dependent' implies 'L_1 not linearly dependent'. This is (2). □

1.5.14. Exercise. Let $L \subset V$ be linearly independent, possibly infinite. Let $G \subseteq V$ be a generating set. Suppose that $\mathrm{Span}(L) \neq V$. Show that there exists $g \in G$ that does not belong to L and such that $L \cup \{g\}$ remains linearly independent.

1.6. Bases and dimension

We continue to have $\mathbb{F}$ a fixed ground field and for most purposes, V is an $\mathbb{F}$-vector space that is fixed as well.

We are going to prove our first important result about general vector spaces. It involves the notion of basis, that combines the idea of generators (Definition 1.4.13) and of linear independence (Definition 1.5.1 and Corollary 1.5.11).

1.6.1. Definition. Let V be an $\mathbb{F}$-vector space. A *basis* B of V is a linearly independent subset $B \subset V$ that generates V. (We do not ask B to be a finite set.)

1.6.2. Remark. Unpacking the definitions in terms of Span(...) as in Definition 1.4.4, to say that B is a basis of V means two things:

(1) $V = \mathrm{Span}(B)$: Every vector in V is a linear combination of vectors in B.
(2) $V \neq \mathrm{Span}(L)$ for every $L \subsetneq B$: No proper subset of B generates V.

In other words, B generates the whole of V and does so in an optimal way.

1.6.3. Example. The $\mathbb{F}$-vector space $V = \mathbb{F}^n$, written in columns today, admits a *canonical* (or *standard*) basis $e_1, \ldots, e_n$ our old friends

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \; e_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \; \ldots, \; e_n = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

More precisely, the j-th entry of the vector e_i is δ_{ij} (Notation A.1.1). Indeed, every vectors $\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$ in $\mathbb{F}^n$ can be written as $a_1e_1+\cdots+a_ne_n$, hence $\mathbb{F}^n = \mathrm{Span}(e_1, \ldots, e_n)$. And if $a_1, \ldots, a_n \in \mathbb{F}$ satisfy $a_1e_1 + \cdots + a_ne_n = 0$ this equation reads $\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$ and thus all $a_i = 0$. In other words, $e_1, \ldots, e_n$ are linearly independent.

We can generalize the above.

1.6.4. Exercise. Let I be a set. For every $i \in I$, define e_i in $\mathbb{F}^I$ to be the function $e_i \colon I \to \mathbb{F}$ defined by $e_i(j) = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$ for every $j \in I$. In another formula (also to get used to the notation) here is the function e_i

$$\begin{aligned} e_i \colon \quad I &\longrightarrow \mathbb{F} \\ j &\longmapsto \delta_{ij}. \end{aligned}$$

Show that the vectors $\{e_i\}_{i\in I}$ form a basis of the subspace $\mathbb{F}^{(I)}$ of $\mathbb{F}^I$. That is, $\{e_i\}_{i\in I}$ is a basis of the free vector space on the set I of Definition 1.3.17.

Note that the $\{e_i\}_{i\in I}$ do not form a basis of $\mathbb{F}^I$ itself, when I is infinite. Give a simple explicit example showing why.

1.6.5. Exercise. Show that $\{1, X, X^2, \ldots\} = \big\{\, X^i \,\big|\, i \in \mathbb{N} \,\big\}$ is a basis of $V = \mathbb{F}[X]$.

1.6.6. Exercise. Show that $\{1, X-1, X^2-1, \ldots\} = \{1\} \cup \big\{\, X^i - 1 \,\big|\, i \in \mathbb{N}, i \geq 1 \,\big\}$ is also a basis of $V = \mathbb{F}[X]$.

1.6.7. Remark. You should remember the (infinitely) many bases that you constructed in $\mathbb{R}^2$ and $\mathbb{R}^3$ in introductory linear algebra. For instance, $\mathbb{R}^3$ has the canonical basis $\{e_1, e_2, e_3\}$ but also the basis $\{\left[\begin{smallmatrix}1\\2\\3\end{smallmatrix}\right], \left[\begin{smallmatrix}4\\5\\6\end{smallmatrix}\right], \left[\begin{smallmatrix}7\\8\\10\end{smallmatrix}\right]\}$ for instance.

The questions we want to address are: First, does every vector space admit a basis? And second, must two different bases have the same number of elements?

1.6.8. Proposition. *Let V be an $\mathbb{F}$-vector space that is generated by a finite set G. Then V admits a finite basis B that is also a subset of G.*

Proof. Let $n = |G|$ be the number of elements of the finite set G. We proceed by induction on n. To start, suppose that $n = 0$, so $G = \varnothing$. Then G is linearly independent (and $V = 0$) and so G is a basis. We can take $B = \varnothing = G$ in that case. The reader can also discuss the case $n = 1$ as an exercise, if the empty set makes them nervous. Suppose now that $n \geq 1$ is such that we know the result for all generating sets of at most $n-1$ elements. Let G be a generating set of V with n elements. There are two cases. If G is linearly independent, then we are done and we can take $B = G$. Otherwise, G is linearly dependent. This means by Definition 1.5.1 that there is a proper subset $G' \subsetneq G$ such that $\operatorname{Span}(G') = \operatorname{Span}(G) = V$. This generating set G' has less elements than G, *i.e.* at most $n-1$ elements. By induction hypothesis, we know that there exists a subset $B \subseteq G'$ which is a basis of V. And clearly $B \subseteq G' \subset G$, so we found a basis B of V inside G. □

1.6.9. Exercise. Recall the cooking recipe from introductory linear algebra. Let $x_1, \ldots, x_n$ be n vectors in $V = \mathbb{F}^m$, with possible repetitions. Consider the system of m linear equations $a_1x_1 + \cdots + a_nx_n = 0$ in the n variables $a_1, \ldots, a_n$. Beware! The variables are not called x here. The x_i are given. They will give you the matrix A of the system by writing them in columns $A = (x_1|x_2|\cdots|x_n)$. Apply Gauss-Jordan to this $(m \times n)$-matrix A. Some of the columns will contain a 'pivot' in the row-reduced-echelon form (a leading 1 in its row). Let $I = \big\{\, i \in \{1, \ldots, n\} \,\big|\, \text{there is a pivot in the } i\text{-th colum} \,\big\}$ be the indices of the columns where there is such a pivot. Show that $\{x_i\}_{i\in I}$ is a basis of $\operatorname{Span}(x_1, \ldots, x_n)$. [Pitfall: Do not use the i-th columns ($i \in I$) in the row-reduction of A but those in the original A! Otherwise you will always get a subset of the canonical basis.]

1.6.10. Example. Let us do a numerical calculation. Take x_1, x_2, x_3 the three vector in $\mathbb{R}^3$ of Example 1.4.11. We want to find a basis of $\operatorname{Span}(x_1, x_2, x_3)$. Form the matrix $A = (x_1|x_2|x_3)$ by writing these generators in columns. Applying row-reduction to this matrix, we find

$$A = \left(\begin{smallmatrix}1&4&7\\2&5&8\\3&6&9\end{smallmatrix}\right) \rightsquigarrow \left(\begin{smallmatrix}1&4&7\\0&-3&-6\\0&-6&-12\end{smallmatrix}\right) \rightsquigarrow \left(\begin{smallmatrix}1&4&7\\0&1&2\\0&-6&-12\end{smallmatrix}\right) \rightsquigarrow \left(\begin{smallmatrix}1&0&-1\\0&1&2\\0&0&0\end{smallmatrix}\right).$$

(Note how we divided by -3 in the second step, because $-3 \neq 0$ in $\mathbb{F} = \mathbb{R}$.) We see pivots in the first and second columns. In the above notation, $= \{1, 2\}$. So x_1 and x_2 are linearly independent. They form a basis $\{x_1, x_2\}$ of $\operatorname{Span}(x_1, x_2, x_3)$.

1.6.11. Remark. One can prove that Proposition 1.6.8 has an analogue without the finiteness assumption: If G is a generating set of V then there exists a basis $B \subseteq G$ contained in it. ([6])

We can also give an intuitive formulation of being a basis as 'maximal linearly independent set' or 'minimal set of generators'. We need a preparation.

1.6.12. Lemma. *Let L be linearly independent in V and $\ell' \in V$ be a vector that does not belong to* $\operatorname{Span}(L)$. *Then $L \cup \{\ell'\}$ remains linearly independent.*

Proof. Note that $\ell' \notin L$ since $\ell' \notin \operatorname{Span}(L)$. Suppose *ab absurdo* that $L' = L \cup \{\ell'\}$ is linearly dependent. Then Proposition 1.5.7 applied to L' gives us a linear combination $a_1\, x_1 + \cdots + a_n\, x_n = 0$ for n distinct vectors $x_1, \ldots, x_n$ in $L \cup \{\ell'\}$ such that all $a_i \neq 0$. These $x_1, \ldots, x_n$ cannot be all in L, for L is linearly independent. So one of the x_i is ℓ', say $x_1 = \ell'$ up to renumbering, and then $x_2, \ldots, x_n \in L \cup \{\ell'\}$ being distinct from $x_1 = \ell'$ must all be in L. But then we can can extract $\ell' = x_1$ from the equation $\sum_{i=1}^n a_i\, x_i = 0$ as $\ell' = x_1 = -a_1^{-1} \cdot (\sum_{i=2}^n a_i x_i)$. This shows that $\ell' \in \operatorname{Span}(x_2, \ldots, x_n) \subseteq \operatorname{Span}(L)$ a contradiction with the choice of ℓ'. □

1.6.13. Proposition. *Let V be an $\mathbb{F}$-vector space.*

(1) *Let $L \subset V$ be a maximal linearly independent subset (maximal means that whenever $L \subseteq L'$ with L' linearly independent then $L = L'$). Then L is a basis.*
(2) *Let $G \subset V$ be a minimal generating subset (minimal means that whenever $G' \subseteq G$ with G' generating V then $G = G'$). Then G is a basis.*

Proof. In Part (1), we need to show $\operatorname{Span}(L) = V$. Suppose *ab absurdo* that $\operatorname{Span}(L) \neq V$. Then there exists $\ell' \in V$ outside of $\operatorname{Span}(L)$ and in particular ℓ' does not belong to L. Then $L' = L \cup \{\ell'\}$ is strictly larger than L and remains linearly independent by Lemma 1.6.12. This contradicts the maximality of L.

In Part (2), we need to show that G is linearly independent. By minimality of G we cannot find $G' \subsetneq G$ such that $\operatorname{Span}(G') = \operatorname{Span}(G)$ for such G' would span V. Therefore G is linearly independent by Definition 1.5.1. □

Let us now turn to our second question about bases: Can we say something about the number of vectors in a basis? The fundamental trick to answer that question is the following:

1.6.14. Theorem. *Let V be an $\mathbb{F}$-vector space that is generated by a* finite *set G. Let $L \subset V$ be a finite set that is linearly independent. Then L has at most as many elements as G, which we write $|L| \leq |G|$.*

More precisely, there exists a subset $H \subseteq G$ such that

(a) *The subset H has the same number of elements as L, in our notation $|H| = |L|$.*

[6] Here is a quick outline of this argument. The key fact is that a maximal linearly independent subset $B \subseteq G$ exists by set-theory (Zorn's Lemma). This is actually equivalent to the so-called Axiom of Choice. Maximality of B means that if $B \subseteq B' \subseteq G$ and B' is also linearly independent then $B = B'$. Such a maximal linearly independent $B \subseteq G$ must span $\operatorname{Span}(G)$ otherwise we can enlarge B to a bigger linearly independent subset B' with $B \subsetneq B' \subseteq G$ by adding to B a suitably chosen element of G (Exercise 1.5.14). This gives a contradiction with the maximality of B. So our B is a linearly independent generating set, hence a basis.

(b) *The space V is generated by $L \cup (G \smallsetminus H)$, in our notation* $\operatorname{Span}(L \cup (G \smallsetminus H)) = V$.

This is sometimes called the *Replacement Theorem* for we managed to remove in the generating set G a bunch of vectors (those in H), keeping the rest $G \smallsetminus H$, and replace those missing vectors by those of L, to create the new generating set $L \cup (G \smallsetminus H)$. The number $|H|$ of generators we remove is exactly $|L|$, the number we put back.

Proof. First note that the second part of the statement is indeed 'more precise' than just saying $|L| \leq |G|$. If we find $H \subseteq G$ that satisfies (a) then of course $|L| = |H| \leq |G|$. Part (b) is crucial to perform induction. We prove more than we need because it is easier. So let us prove the result by induction on $n = |L|$.

Suppose that $n = 0$, meaning $L = \varnothing$. Then there is nothing to do: Take $H = \varnothing$ for we have no choice if we want (a). Then check (b): $L \cup (G \smallsetminus H) = \varnothing \cup (G \smallsetminus \varnothing) = G$ still generates V. Nothing has changed.

The rest of the proof is the induction step. Suppose that $n \geq 1$ and that we know the result for any linearly independent set L' with at most $n-1$ elements. Take now our L, with $|L| = n$. There is an obvious way to trigger the induction hypothesis. Just pick some $\ell \in L$ and set $L' = L \smallsetminus \{\ell\}$. This L' remains linearly independent (Corollary 1.5.13 (2)) and has $n-1$ elements.

By induction hypothesis applied to L', we can find a subset $H' \subseteq G$ with $n-1$ elements such that $L' \cup (G \smallsetminus H')$ generates V:

$$\operatorname{Span}(L' \cup (G \smallsetminus H')) = V. \tag{1.6.15}$$

This is not yet what we want: We need to find $H \subseteq G$ with n elements, whereas H' has only $n-1$ elements. We just need to find one more vector to 'replace' by the last vector in $L \smallsetminus L'$, namely ℓ. In other words, we are going to define H as $H' \cup \{g\}$ for a judicious choice of $g \in G$, not in H'.

Let us now use what we know from the induction hypothesis. We know (1.6.15). Surely the vector ℓ is in V, so it is a linear combination of vectors in this generating set $L' \cup (G \smallsetminus H')$. This means that we can find $x_1 \ldots, x_r \in L'$ and $y_1, \ldots, y_s \in G \smallsetminus H'$ and scalars $a_1, \ldots, a_r$ and $b_1, \ldots, b_s$ such that

$$\ell = \sum_{i=1}^{r} a_i \cdot x_i \ + \sum_{j=1}^{s} b_j \cdot y_j. \tag{1.6.16}$$

We simply wrote ℓ as a linear combination of vectors in the union $L' \cup (G \smallsetminus H')$, regrouping in the first sum the generators in L' (we called them x_i) and in the second sum the generators in $G \smallsetminus H'$ (we called them y_j). As usual, we can assume that *the vectors $x_1, \ldots, x_r, y_1, \ldots, y_s$ are all distinct and the coefficients $a_1, \ldots, a_r, b_1, \ldots, b_s$ are all non-zero*, otherwise we can regroup the terms and remove those with coefficient zero (as in the proof of Lemma 1.4.16) at the possible cost of making r or s smaller. But so far, we have not excluded $r = 0$ or $s = 0$ anyway.

Now comes the trick. We cannot have $s = 0$ in (1.6.16). Indeed, this absurd assumption would mean that $\ell = \sum_{i=1}^{r} a_i\, x_i$ which would contradict the linear independence of $L = L' \cup \{\ell\}$ since $\ell \notin L'$ and since $x_1, \ldots, x_r \in L'$ are all distinct.

So we know that $s \geq 1$ in (1.6.16), with all $b_j \neq 0$. In particular there is this scalar $b_1 \neq 0$ in the field $\mathbb{F}$. We know what to do. First multiply everything in sight by b_1^{-1} and then think. We rearrange (1.6.16) to extract y_1 as follows:

$$y_1 = b_1^{-1} \cdot \Big(-\sum_{i=1}^{r} a_i \cdot x_i \ + \ell - \sum_{j=2}^{s} b_j \cdot y_j \Big).$$

Let us not panic about the exact coefficients. What this equation says is that y_1 is a linear combination of $x_1, \ldots, x_r, \ell, y_2, \ldots, y_s$. Let us record this fact for later:

$$y_1 \in \operatorname{Span}(x_1, \ldots, x_r, \ell, y_2, \ldots, y_s). \tag{1.6.17}$$

We are almost done! We just need to do the bookkeeping right. First, recall the goal. We want a subset $H \subseteq G$ with $n = |L|$ elements. We define it to be

$$H := H' \cup \{y_1\}.$$

Since H' has $n-1$ elements, H will have n elements as long as $y_1 \notin H'$. But remember where y_1 was taken from: Like all y_j it belongs to $G \smallsetminus H'$, see before (1.6.16). In particular $y_1 \notin H'$ as wanted. So we have (a). Let us prove (b).

The claim is that $W := \operatorname{Span}(L \cup (G \smallsetminus H))$ is the whole of V. We shall use the generators of V given in (1.6.15). First note that

$$L' \subseteq L \subseteq \operatorname{Span}(L \cup (G \smallsetminus H)) = W \tag{1.6.18}$$

and in particular the x_i which were in L' and ℓ which was in L all belong to W:

$$x_1 \ldots, x_r, \ell \in W. \tag{1.6.19}$$

We now want to show that $(G \smallsetminus H') \subseteq W$. Take $y \in G \smallsetminus H'$. If $y \neq y_1$ then y belongs to the elements of G that are neither in H' nor in $\{y_1\}$ so $y \in G \smallsetminus H$ by definition of $H = H' \cup \{y_1\}$. For instance, all the $y_2, \ldots, y_r$ that belong to $G \smallsetminus H'$ and are distinct from y_1 belong to $G \smallsetminus H$. So

$$y_2, \ldots, y_s \in (G \smallsetminus H) \subseteq \operatorname{Span}(L \cup (G \smallsetminus H)) = W. \tag{1.6.20}$$

We now deduce from (1.6.17), (1.6.19) and (1.6.20) that

$$y_1 \in \operatorname{Span}(x_1 \ldots, x_r, \ell, y_2, \ldots, y_s) \subseteq W.$$

So we just proved two things about the elements $y \in G \smallsetminus H'$. When $y \neq y_1$ then $y \in (G \smallsetminus H) \subseteq W$. And when $y = y_1$ then $y \in W$ too! In short, they are all in W:

$$(G \smallsetminus H') \subseteq W.$$

Combining the latter with (1.6.18), we see that

$$L' \cup (G \smallsetminus H') \subseteq W.$$

But the set $L' \cup (G \smallsetminus H')$ was a generating set of V by (1.6.15). So we deduce that $W = V$, as was left to prove to get (b). The induction step is complete. □

1.6.21. Corollary. *Let V be an $\mathbb{F}$-vector space generated by a finite set G and L be linearly independent. Then L is finite and $|L| \leq |G|$.*

Proof. Let $m = |G|$. By Theorem 1.6.14, every finite subset $L' \subseteq L$ being linearly independent (Corollary 1.5.13 (2)) has at most m elements. So L cannot contain $m+1$ different elements. Hence L is finite and $|L| \leq m$. □

The Replacement Theorem 1.6.14 has a cascade of consequences. The most important result of this chapter is one of those consequences:

1.6.22. Corollary. *Let V be an $\mathbb{F}$-vector space with a finite basis B. Then any other basis B' of V is finite as well and has the same number of elements as B:*

$$|B| = |B'|.$$

Proof. Applying Corollary 1.6.21 to $G = B$ and $L = B'$ tells us that B' is finite and $|B'| \leq |B|$. Applying Corollary 1.6.21 to $G = B'$ and $L = B$ tells us that $|B| \leq |B'|$. Combining the two, we get $|B| = |B'|$. Voilà! □

1.6.23. Remark. The above proof is misleadingly simple. It relies on the non-trivial argument of Theorem 1.6.14. Note in particular that we can have two very different subsets B and B', for instance with $B \cap B' = \varnothing$. See Remark 1.6.7.

Now comes the most important definition of this chapter, beyond the concept of vector space.

1.6.24. Definition. Let V be an $\mathbb{F}$-vector space.

(1) If V admits a finite basis (or a finite generating set by Proposition 1.6.8), we say that V is *finite-dimensional over* $\mathbb{F}$. In that case, its *dimension over* $\mathbb{F}$, denoted $\dim(V) = \dim_{\mathbb{F}}(V)$, is the number of elements in *any* basis B of V.

(2) If V does not admit a finite basis we say that V is *infinite-dimensional.* In that case, all bases of V are infinite by Corollary 1.6.22. ([7])

1.6.25. Example. The zero vector space $\{0\}$ is the only vector space with dimension zero (*i.e.* empty basis).

1.6.26. Example. For any $n \in \mathbb{N}$, we have $\dim_{\mathbb{F}}(\mathbb{F}^n) = n$. Indeed, the canonical basis has n vectors. Therefore all other bases of $\mathbb{F}^n$ must contain exactly n vectors.

1.6.27. Exercise. For any $p, q \geq 1$, we have $\dim_{\mathbb{F}}(\mathrm{M}_{p\times q}(\mathbb{F})) = p \cdot q$.

1.6.28. Remark. It is important to make sure the field is known. For instance, $\dim_{\mathbb{C}}(\mathbb{C}) = 1$ but $\mathbb{C}$ is also a real vector space, namely $\mathbb{R}^2$ and $\dim_{\mathbb{R}}(\mathbb{C}) = 2$. A real basis of $\mathbb{C}$ is given by $\{1, i\}$. In particular, those two vectors are linearly independent over $\mathbb{R}$. But they are linearly dependent over $\mathbb{C}$ as $i \cdot 1 + (-1) \cdot i = 0$. A complex basis of $\mathbb{C}$ would be $\{1\}$ (canonical) or $\{i\}$ for instance, or any non-zero $\{z\}$. ([8])

1.6.29. Example. The $\mathbb{F}$-vector space $\mathbb{F}[X]$ is infinite-dimensional, as one basis $\{\, X^i \,\big|\, i \in \mathbb{N} \,\}$ is infinite, and therefore so are all others.

1.6.30. Exercise. Let $d \geq 1$. The subspace $W \subset \mathbb{F}[X]$ of polynomials of degree at most d has dimension $d + 1$.

Let us return to the general consequences of Theorem 1.6.14.

1.6.31. Corollary. *Let V be a finite-dimensional $\mathbb{F}$-vector space of dimension n.*

(1) *Any linearly independent subset $L \subset V$ has at most n elements. It has exactly n elements if and only if L is a basis of V.*

(2) *Any generating set G of V has at least n elements. It has exactly n elements if and only if G is a basis of V.*

[7] There is a way to distinguish between infinite sets. We say that two sets have the *same cardinal* if there exists a bijection between them. One can prove with a little bit more set theory that two bases of a possibly infinite-dimensional vector space have the same cardinal. That cardinal is the general definition of dimension, and it is finer than just saying 'infinite'.

[8] It gets wilder. For instance $\mathbb{R}$ is infinite-dimensional over $\mathbb{Q}$. And that 'infinite' is not 'countable', that is of cardinal larger than the cardinal of $\mathbb{N}$.

Proof. Let B be a basis of V. We have $|B| = n$.

For (1), we have $|L| \leq |B|$ by Corollary 1.6.21 applied to $G = B$. If now L has exactly n elements, let us apply the full Replacement Theorem 1.6.14 to the generating set $G = B$. It tells us that we can find $H \subseteq B$ such that $|H| = |L|$ and $L \cup (B \smallsetminus H)$ generates V. But $|H| = |L| = n = |B|$ and $H \subseteq B$ forces $H = B$. Thus the generating set is $L \cup (B \smallsetminus H) = L \cup \varnothing = L$. So L is a basis by definition.

For (2), every basis vector $b \in B$ belongs to $V = \mathrm{Span}(G)$, hence belongs to $\mathrm{Span}(G_b)$ for some finite subset $G_b \subseteq G$. Since B is finite, putting all those finite G_b together, we see that there exists a finite $G' \subseteq G$ that spans V. By Proposition 1.6.8, there exists a basis $B' \subseteq G'$ inside G'. Hence $n = |B'| \leq |G'| \leq |G|$. If moreover, $|G| = n$ then the subset $B' \subseteq G$ must be equal to G and $G = B'$ is a basis. □

1.6.32. Corollary. *Let V be a finite-dimensional $\mathbb{F}$-vector space and $W \subseteq V$ be a subspace. Then the following statements hold:*

(1) *The subspace W is finite-dimensional and* $\dim(W) \leq \dim(V)$.
(2) *We have* $\dim(W) = \dim(V)$ *if and only if $W = V$.*
(3) *We can* extend *any basis of W into a basis of V: If C is a basis of W then there exists a basis B of V that contains C.*

Proof. Let $L \subset W$ be a linearly independent subset of W. Note that it remains linearly independent in V. (Direct from Definition 1.5.1 or Corollary 1.5.11.) Hence $|L| \leq \dim(V)$ by Corollary 1.6.31 (1). So there are only finitely many numbers $\{\, m \in \mathbb{N} \mid \text{there exists } L \subset W \text{ linearly independent with } |L| = m \,\}$ all between zero and $\dim(V)$. Pick m maximal in that set and pick $L \subseteq W$ linearly independent with m elements. This L is a maximal linearly independent subset of the vector space W in the sense of Proposition 1.6.13 (1). Therefore L is a basis of W. Therefore $\dim(W) = |L| = m \leq \dim(V)$ as claimed in (1).

For (2), if $\dim(W) = \dim(V)$ then a basis C of W would be a linearly independent subset of V with $\dim(V)$ elements, hence a basis of V by Corollary 1.6.31. In particular C would generate V and thus $V = \mathrm{Span}(C) = W$.

For (3), we apply the Replacement Theorem 1.6.14 to $L = C$ and G any basis of V. There exists $H \subseteq G$ with $|H| = |C|$ and such that $B := C \cup (G \smallsetminus H)$ generates V. Clearly $C \subseteq B$ and we claim that B is a basis. By construction B has at most $|C| + |G \smallsetminus H|$ elements (there could be an intersection between C and $(G \smallsetminus H)$) and that number is $|C| + (|G| - |H|) = |G|$ since $|C| = |H|$. In summary, B is a generating set of V with at most $|G| = \dim(V)$ elements since G is a basis. By Corollary 1.6.31 (2) we know that B is a basis of V as wanted. □

1.6.33. Exercise. In the spirit of Exercise 1.6.9, let us recall another cooking recipe from introductory linear algebra. Let $x_1, \ldots, x_\ell$ be ℓ linearly independent vectors in $V = \mathbb{F}^n$, for instance a basis of subspace. How to find a basis of V that extends $C = \{x_1, \ldots, x_\ell\}$? Consider the system of n linear equations $a_1 x_1 + \cdots + a_\ell x_\ell + a_{\ell+1} e_1 + \cdots + a_{\ell+n} e_n = 0$ in the $\ell + n$ variables $a_1, \ldots, a_{\ell+n}$. (The number of variables is not really an issue: we rarely write the variables, just the matrix.) Let A be the $(n \times (\ell + n))$-matrix the system $A = (x_1 | x_2 | \cdots | x_\ell | e_1 | \cdots | e_n)$. Apply Gauss-Jordan and look at the columns where there is a pivot, as in Exercise 1.6.9. Let $I = \{\, i \in \{1, \ldots, \ell{+}n\} \mid \text{there is a pivot in the } i\text{-th colum} \,\}$. Show that the first ℓ columns have a pivot: $\{1, \ldots, \ell\} \subseteq I$ and that $\{x_1, \ldots, x_\ell\} \cup \{\, e_i \mid \ell + i \in I \,\}$ is a basis of V. In other words, we can complete C by picking some e_j in the canonical basis $e_1, \ldots, e_n$ (for those j corresponding to the pivots). There is nothing specific about the canonical basis and one can use any basis B of V instead.

1.6.34. Exercise. Let W_1 and W_2 be two finite-dimensional subspaces of an $\mathbb{F}$-vector space V. Show that $W_1 + W_2$ and $W_1 \cap W_2$ are finite-dimensional and that

$$\dim(W_1 + W_2) = \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2).$$

We now discuss the topic of coordinates, that will be important to turn our beautiful abstract linear algebra into aesthetically-challenged introductory linear algebra, column vectors and matrices. To do that carefully, we need to distinguish a set with n elements from an ordered sequence $x_1, \ldots, x_n$.

1.6.35. Definition. Let V be an $\mathbb{F}$-vector space of finite dimension n. An *ordered basis* is a numbered sequence of vectors $x_1, \ldots, x_n$, all distinct, such that the set $\{x_1, \ldots, x_n\}$ is a basis of V. We shall mildly abuse notation and write "$B = \{x_1, \ldots, x_n\}$ is an ordered basis" in this situation.

1.6.36. Proposition. *Let $x_1, \ldots, x_n$ be an ordered basis of an $\mathbb{F}$-vector space V. Then for every $v \in V$ there exists a* unique *n-tuple of scalars $(a_1, \ldots, a_n) \in \mathbb{F}^n$, depending on v, such that $a_1x_1 + \cdots + a_nx_n = v$.*

PROOF. Such a tuple $(a_1, \ldots, a_n)$ exists by definition of $\operatorname{Span}(x_1, \ldots, x_n) = V$. For uniqueness, let $(b_1, \ldots, b_n)$ be such that $b_1x_1 + \cdots + b_nx_n = v$. We compute

$$\sum_{i=1}^{n}(a_i - b_i) \cdot x_i = \sum_{i=1}^{n} a_i \cdot x_i - \sum_{i=1}^{n} b_i \cdot x_i = v - v = 0.$$

Since the vectors $x_1, \ldots, x_n$ are linearly independent, those coefficients $(a_i - b_i)$ must all be zero, for all $1 \leq i \leq n$ (Corollary 1.5.11). Thus $(a_1, \ldots, a_n) = (b_1, \ldots, b_n)$. □

1.6.37. Definition. Let $B = \{x_1, \ldots, x_n\}$ be an ordered basis of V. Let $v \in V$. The *coordinates* of v in the ordered basis B is the unique n-tuple $(a_1, \ldots, a_n)$ in $\mathbb{F}^n$ of Proposition 1.6.36 such that $a_1 \cdot x_1 + \cdots + a_n \cdot x_n = v$. We write

$$\left[\ v\ \right]_B = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

to mean "the coordinates of v in the ordered basis B are $a_1, \ldots, a_n$, in that order". We use column vectors because of the matrix calculus that ensues.

1.6.38. Exercise. Let B be an ordered basis of V. Show that we have $[v + w]_B = [v]_B + [w]_B$ and $[a \cdot v]_B = a \cdot [v]_B$ for all $v, w \in V$ and $a \in \mathbb{F}$.

1.6.39. Exercise. Let $B = \{e_1, \ldots, e_n\}$ be the canonical basis of $V = \mathbb{F}^n$. Let $v \in \mathbb{F}^n$. What are the coordinates of v in the basis B?

1.6.40. Exercise. Let $W = \operatorname{Span}(x_1, x_2)$ in $V = \mathbb{R}^3$ for $x_1 = (1, 2, 3)$ and $x_2 = (4, 5, 6)$. We know that $\{x_1, x_2\}$ is an (ordered) basis of W. What are the coordinates of $v = (7, 8, 9)$ in that basis?

1.6.41. Remark. We shall rarely use coordinates with an infinite basis. However, they are quite easy, especially if we adopt a flexible notation. The notation is mildly dangerous (don't hurt yourself!) but very convenient. Let B be a basis of V, possibly infinite. The fact that B is a set of generators means that every $v \in V$ can be written as

$$v = \sum_{b \in B} a_b \cdot b \tag{1.6.42}$$

for scalars $a_b \in \mathbb{F}$ indexed by $b \in B$ and that are *almost all zero*, meaning the set of indices $\left\{ b \in B \,\middle|\, a_b \neq 0 \right\}$ is finite. So the apparently infinite sum in (1.6.42) only involves finitely many non-zero terms, so it should be read as the finite sum

$$\sum_{\substack{b \in B \\ a_b \neq 0}} a_b \cdot b\,.$$

The fact that B is linearly independent really means that those coefficients a_b are unique (as a function $B \to \mathbb{F}$), for a given v. The argument of Proposition 1.6.36 goes through. These scalars $(a_b)_{b \in B}$ are the *coordinates* of v in the basis B. [Exercise: Check that this boils down to the earlier definition when B is finite.]

CHAPTER 2

Linear transformations

In Chapter 1, we studied one vector space at a time. We now want to move from one vector space to another. For the whole chapter $\mathbb{F}$ is a fixed field.

2.1. Linear transformations

2.1.1. Definition. Let V and V' be two vector spaces on the *same* field $\mathbb{F}$. A *linear transformation* $T\colon V \to V'$ is a function

$$\begin{array}{rccc} T\colon & V & \longrightarrow & V' \\ & x & \longmapsto & T(x) \end{array}$$

that preserves the two operations, the addition and the action of $\mathbb{F}$, *i.e.* we require that T satisfies the following two axioms:

(LT1) We have $T(x+y) = T(x) + T(y)$ in V' for every x, y in V.

(LT2) We have $T(a \cdot x) = a \cdot T(x)$ in V' for every x in V and every a in $\mathbb{F}$.

Equivalently, this means that T preserves linear combinations: We have

$$T\Big(\sum_{i=1}^{n} a_i \cdot x_i\Big) = \sum_{i=1}^{n} a_i \cdot T(x_i) \tag{2.1.2}$$

in V' for every $n \in \mathbb{N}$, vectors $x_1, \ldots, x_n$ in V and scalars $a_1, \ldots, a_n$ in $\mathbb{F}$.

Let us begin with the trivial examples, from very trivial to less so.

2.1.3. Example. For any V and V' there is at least the *zero transformation* $0\colon V \to V'$ defined by $0(x) = 0_{V'}$ for every $x \in V$.

2.1.4. Example. For any vector space V there is the *identity transformation* $\mathrm{Id}_V\colon V \to V$ defined by $\mathrm{Id}_V(x) = x$ for every $x \in V$.

2.1.5. Example. A slight variation on the previous example is the inclusion of a subspace. Let $W \subseteq V$ be a subspace. Then $\mathrm{incl}\colon W \to V$ defined by $\mathrm{incl}(x) = x$ for every $x \in W$ is a linear transformation from W to V.

Following our pattern, after trivial examples, we should import all examples from introductory linear algebra.

2.1.6. Example. Let $p, q \in \mathbb{N}$ with $p, q \geq 1$. We want to define linear transformations $\mathbb{F}^p \to \mathbb{F}^q$. So here $V = \mathbb{F}^p$ and $V' = \mathbb{F}^q$ and we shall write vectors in columns to use matrix multiplication. Let $A \in \mathrm{M}_{q\times p}(\mathbb{F})$ be a matrix of size $q \times p$ – stress here the 'reversal' of the order of p and q. Define

$$\begin{array}{rccc} T_A\colon & \mathbb{F}^p & \longrightarrow & \mathbb{F}^q \\ & x & \longmapsto & T_A(x) = A \cdot x \end{array} \tag{2.1.7}$$

where the latter $\cdot$ is matrix multiplication. In other words, if $x = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}$ in $\mathbb{F}^p$ then $T_A(x) = \begin{bmatrix} y_1 \\ \vdots \\ y_q \end{bmatrix}$ is given by the following formula, for $1 \leq i \leq q$:

$$y_i = \sum_{j=1}^{p} A_{ij}\, x_j. \tag{2.1.8}$$

It is an easy but important exercise to verify linearity of this transformation T_A.

Recall that these T_A are in fact the *only* linear transformations from $\mathbb{F}^p$ to $\mathbb{F}^q$. We shall reprove this in Corollary 2.4.9.

Let us give further examples.

2.1.9. Exercise. Let $n \in \mathbb{N}$, $n \geq 1$. Consider $V = \mathrm{M}_{n\times n}(\mathbb{F})$ and $V' = \mathbb{F}^1$. The *trace* $\mathrm{tr}\colon \mathrm{M}_{n\times n}(\mathbb{F}) \to \mathbb{F}$ is defined by $\mathrm{tr}(X) = X_{11} + X_{22} + \cdots + X_{nn}$ for every $(n \times n)$-matrix X. Verify that $\mathrm{tr}\colon V \to V'$ is linear.

2.1.10. Exercise. Is the *determinant* $\det\colon \mathrm{M}_{n\times n}(\mathbb{F}) \to \mathbb{F}$ linear?

Let now discuss examples involving infinite-dimensional vector spaces.

2.1.11. Exercise. Let $V = \mathbb{F}^{\mathbb{N}}$ be the vector space of sequences $(x_i)_{i\in\mathbb{N}}$ in $\mathbb{F}$. Consider the transformation

$$\begin{array}{cccccl} T\colon & V & \longrightarrow & V & \\ & (x_i)_{i\in\mathbb{N}} & \longmapsto & (y_i)_{i\in\mathbb{N}} & \text{where } y_i = x_{i+1} \text{ for all } i \in \mathbb{N} \end{array}$$

that 'moves a sequence one notch to the left' $T(x_0, x_1, x_2, \ldots) = (x_1, x_2, x_3 \ldots)$ and 'drops' x_0. Prove that T is linear.

2.1.12. Exercise. Let $V = \mathbb{F}^{\mathbb{N}}$ as above and consider the transformation

$$\begin{array}{cccccl} S\colon & V & \longrightarrow & V & \\ & (x_i)_{i\in\mathbb{N}} & \longmapsto & (y_i)_{i\in\mathbb{N}} & \text{where } y_i = \begin{cases} x_{i-1} & \text{if } i \geq 1 \\ 0 & \text{if } i = 0. \end{cases} \end{array}$$

that 'moves one notch to the right' $S(x_0, x_1, x_2, x_3, \ldots) = (0, x_0, x_1, x_2, \ldots)$ and 'squeezes in' 0 at the start. Prove that S is linear as well. Prove that 0 was the only option to 'squeeze in' in the first slot.

2.1.13. Exercise. Let $c \in \mathbb{F}$ be fixed. Define *evaluation at c*

$$\mathrm{ev}_c\colon \mathbb{F}[X] \to \mathbb{F}$$

at every $P = a_0 + a_1 X + \cdots + a_n X^n$ by $\mathrm{ev}_c(P) = P(c) = a_0 + a_1 c + \cdots + a_n c^n$ (in colloquial terms, 'plug $X = c$'). Show that ev_c is linear.

2.1.14. Remark. It is sometimes necessary to say $\mathbb{F}$*-linear*, to emphasize the field of scalars. For instance, consider $V = \mathbb{C}$ and $T\colon V \to V$ complex conjugation $T(x + yi) = x - yi$. Then V is a real vector space and T is $\mathbb{R}$-linear. However, V is also a complex vector space but T is not $\mathbb{C}$-linear since $T(i \cdot 1) = -i \neq i = i \cdot T(1)$.

2.1.15. Proposition. *Let $B = \{x_1, \ldots, x_n\}$ be an ordered basis of V. Then we have a linear transformation $[-]_B\colon V \to \mathbb{F}^n$ defined by mapping $v \in V$ to its coordinates $[v]_B \in \mathbb{F}^n$ in the basis.*

PROOF. Exercise 1.6.38. □

2.1.16. Remark. Note that a linear transformation $T\colon V \to V'$ is in particular a function, hence belongs to the vector space $(V')^V$. (See Exercise 1.2.20.) We can thus add linear transformations and multiply them by scalars in the obvious way and verify that they remain linear. Let us spell this out.

2.1.17. Proposition. *Let V, V' be $\mathbb{F}$-vector spaces.*

(1) *If $T_1, T_2\colon V \to V'$ are linear then $T_1 + T_2\colon V \to V'$ remains linear, where we recall that $(T_1 + T_2)(x) = T_1(x) + T_2(x)$ for every $x \in V$.*

(2) *If $T\colon V \to V'$ is linear and $a \in \mathbb{F}$ is a scalar then $a \cdot T\colon V \to V'$ remains linear, where we recall that $(a \cdot T)(x) = a \cdot (T(x))$ for every $x \in V$.*

(3) *The set $\mathrm{Lin}_{\mathbb{F}}(V, V') = \big\{T\colon V \to V' \,\big|\, T \text{ is linear}\big\}$ is a vector space with the above operations. (It is a subspace of $(V')^V$.)*

PROOF. Exercise. □

2.1.18. Notation. In advanced algebra, $\mathrm{Lin}_{\mathbb{F}}(V, V')$ is often denoted $\mathrm{Hom}_{\mathbb{F}}(V, V')$ instead. Here Hom stands to 'homomorphisms'. We shall keep $\mathrm{Lin}_{\mathbb{F}}$ in these notes, sometimes dropping the $\mathbb{F}$ when clear from context.

We can *compose* linear transformations.

2.1.19. Definition. Let V, V' and V'' be three $\mathbb{F}$-vector spaces and let $T\colon V \to V'$ and $S\colon V' \to V''$ be two linear transformations. Their *composition* $S \circ T\colon V \to V''$ (as functions) is defined as follows

$$\begin{array}{cccccccc} S \circ T\colon & V & \xrightarrow{T} & V' & \xrightarrow{S} & V'' & & \\ & x & \longmapsto & & & (S \circ T)(x) & = & S(T(x)). \end{array}$$

To check that $S \circ T\colon V \to V''$ remains linear, we compute for all possible $n \in \mathbb{N}$, all vectors $x_1, \ldots, x_n \in V$ and all scalars $a_1, \ldots, a_n \in \mathbb{F}$

$$\begin{aligned} (S \circ T)\big(\textstyle\sum_{i=1}^n a_i \cdot x_i\big) &= S\big(T\big(\textstyle\sum_{i=1}^n a_i \cdot x_i\big)\big) && \text{by definition of } S \circ T \\ &= S\big(\textstyle\sum_{i=1}^n a_i \cdot T(x_i)\big) && \text{by linearity of } T \\ &= \textstyle\sum_{i=1}^n a_i \cdot S\big(T(x_i)\big) && \text{by linearity of } S \\ &= \textstyle\sum_{i=1}^n a_i \cdot \big((S \circ T)(x_i)\big) && \text{by definition of } S \circ T. \end{aligned}$$

Hence $S \circ T$ is indeed linear.

2.1.20. Remark. For any $T\colon V \to V'$, we have $T \circ \mathrm{Id}_V = T$ and $\mathrm{Id}_{V'} \circ T = T$. For three consecutive linear transformations, T, S, R, we also have $R \circ (S \circ T) = (R \circ S) \circ T$. [Exercise: Draw the 'diagram' with the vector spaces and the arrows.]

2.1.21. Exercise. Consider the two linear transformations $S, T\colon V \to V$ of Examples 2.1.11 and 2.1.12, on the vector space $V = \mathbb{F}^{\mathbb{N}}$ of sequences. Compute the composites $S \circ T$ and $T \circ S$. For every $m \in \mathbb{N}$ describe $T^m = T \circ T \circ \cdots \circ T$ (the composition of m copies of T) and describe S^m.

2.1.22. Exercise. Let $S\colon V' \to V''$ be linear. Show that $S \circ -\colon \mathrm{Lin}_{\mathbb{F}}(V, V') \to \mathrm{Lin}_{\mathbb{F}}(V, V'')$, $T \mapsto S \circ T$, is linear. State (and prove) an analogous result for $- \circ T$.

2.1.23. Theorem. *Let V and V' be two $\mathbb{F}$-vector spaces. Let B be a basis of V. Suppose given a vector y_b in V' for every basis vector $b \in B$.*

(1) *We can extend the assignment $b \mapsto y_b$ (given for $b \in B$) to a linear transformation on the whole of V. In other words, there exists a* linear *transformation $T\colon V \to V'$ such that $T(b) = y_b$ for every $b \in B$.*

(2) *The linear transformation T of Part (1) is unique: If $T, T'\colon V \to V'$ are two linear transformations such that $T(b) = T'(b)$ for all $b \in B$ then $T = T'$.*

Proof. Let us prove uniqueness (2) first. Suppose that $T\colon V \to V'$ is linear and satisfies $T(b) = y_b$ for all $b \in B$. Pick now a vector $v \in V$. Since B is a basis, we can write $v = \sum_{b\in B} a_b \cdot b$ for scalars $a_b \in \mathbb{F}$, with the abuse of notation of (1.6.42), that is, the a_b are almost all zero and this apparently infinite sum (when B is infinite) is always finite. Then by linearity of T we must have

$$T(v) = T(\sum_{b\in B} a_b \cdot b) = \sum_{b\in B} a_b \cdot T(b) = \sum_{b\in B} a_b \cdot y_b \tag{2.1.24}$$

where the last equality holds because we want $T(b)$ to be the given vector y_b for $b \in B$. As we see from (2.1.24), the image $T(v)$ is entirely forced once we know the coordinates of v in the basis B and the image $T(b) = y_b$. (Another T' would give the same $T'(v)$ as on the right-hand side of (2.1.24).)

We can now prove (1) by means of (2.1.24). Define $T\colon V \to V'$ by the following process. For every $v \in V$, write $v = \sum_{b\in B} a_b \cdot b$ for scalars $a_b \in \mathbb{F}$ almost all zero and set $T(v) = \sum_{b\in B} a_b \cdot y_b$. This is a well-defined process as there is only *one* such collection $(a_b)_{b\in B}$ of scalars, since B is a basis. (See Remark 1.6.41.) We need to verify two things: That T is linear and that $T(b) = y_b$ for all $b \in B$. The latter is easy since $b = 1 \cdot b$ has obvious coordinates $(\delta_{b,b'})_{b'\in B}$ hence $T(b) = 1 \cdot y_b = y_b$. For linearity, let us check (LT 1). Suppose that $v, \tilde{v}$ are two vectors in V and write them in coordinates: $v = \sum_{b\in B} a_b \cdot b$ and $\tilde{v} = \sum_{b\in B} \tilde{a}_b \cdot b$ for scalars $a_b, \tilde{a}_b \in \mathbb{F}$. The coordinates of $v + \tilde{v}$ are easy: $v + \tilde{v} = \sum_{b\in B}(a_b + \tilde{a}_b) \cdot b$. Again, all these sums are truly finite. By definition of T, we have set $T(v)$, $T(\tilde{v})$ and $T(v + \tilde{v})$ to be respectively

$$\sum_{b\in B} a_b \cdot y_b\,, \quad \sum_{b\in B} \tilde{a}_b \cdot y_b \quad \text{and} \quad \sum_{b\in B} (a_b + \tilde{a}_b) \cdot y_b\,.$$

It is then easy to see that $T(v) + T(\tilde{v}) = T(v + \tilde{v})$. This proves (LT 1). We leave the (easier) verification of (LT 2) to the reader. □

2.1.25. Exercise. Redo the above proof when V is finite-dimensional and $B = \{x_1, \ldots, x_n\}$ is an ordered basis. [In that case, we do not write a_{x_i} for $x_i \in B$ but simply a_i.] Show that given $y_1, \ldots, y_n$ in V' the unique $T\colon V \to V'$ such that $T(x_i) = y_i$ on the basis $(i = 1 \ldots, n)$ is given by

$$T(v) = \sum_{i=1}^{n} a_i \cdot y_i \qquad \text{where} \qquad [v]_B = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}. \tag{2.1.26}$$

2.1.27. Remark. The main take-home information of Theorem 2.1.23 is the following. A linear transformation $T\colon V \to V'$ which a priori seems to involve a massive amount of information (all $T(v)$ for the usually infinitely many $v \in V$) can be completely and uniquely described by deciding where T maps the vectors of a given basis of V. Even better, if V is finite-dimensional with basis $\{x_1, \ldots, x_n\}$ then the information necessary to describe T consists of *finitely* many vectors $T(x_1), \ldots, T(x_n)$ in V'. Even even better, if V' is also finite-dimensional, say of dimension m, then those n vectors $T(x_1), \dots, T(x_n)$ themselves are known once we know their coordinates in some basis of V'. So the knowledge of T boils down to n sets of m scalars. We follow-up on this idea in Section 2.4.

2.1.28. Remark. There is a simpler way to formulate Theorem 2.1.23 as follows. The transformation $T\colon V \to V'$ that we construct is simply the composition $T = S \circ [-]_B$ of two linear transformations

$$\begin{tikzcd} V \arrow[r, "{[-]_B}"] & \mathbb{F}^{(B)} \arrow[r, "S"] & V'. \end{tikzcd}$$

The first one is 'the coordinates' $[-]_B\colon V \to \mathbb{F}^{(B)}$ with respect to B. The second, $S\colon \mathbb{F}^{(B)} \to V'$, maps a function $a\colon B \to \mathbb{F}$ to $\sum_{b\in B} a(b)\cdot y_b$ where the latter sum is finite, since $a \in \mathbb{F}^{(B)}$ means that $a(b) \neq 0$ only for finitely many $b \in B$.

When the basis $B = \{x_1, \dots, x_n\}$ is finite, the transformation S simply maps every $a = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$ to $a_1 \cdot y_1 + \cdots + a_n \cdot y_n$. In both cases, the only thing we need is the knowledge of y_b in V' for each $b \in B$ (respectively of $y_i = y_{x_i}$ for each $i = 1, \dots, n$).

2.1.29. Remark. In fact, linear transformations $T\colon \mathbb{F}^{(I)} \to V$, out of the free $\mathbb{F}$-vector space $\mathbb{F}^{(I)}$ over a set I into a given vector space V, are exactly the same as functions $I \to V$, as sets. This is the meaning of the phrase 'free vector space over the set I'. In other words, we have (what we'll call in Section 2.3) an isomorphism

$$\mathrm{Lin}(\mathbb{F}^{(I)}, V) \simeq V^I.$$

To see this, note that we have a bijection $\epsilon\colon i \mapsto e_i$ between I and the canonical basis $\{\, e_i \,|\, i \in I \,\}$ of $\mathbb{F}^{(I)}$. If $T\colon \mathbb{F}^{(I)} \to V$ is linear, we can associate to it the function $f = \Phi(T)\colon I \to V$ mapping i to $T(e_i)$. (In short, $\Phi(T) = T \circ \epsilon$.) Conversely, given a function $f\colon I \to V$ as sets, we use Theorem 2.1.23 to extend linearly the assignment $e_i \mapsto f(i)$ to a linear transformation $T = \Psi(f)\colon \mathbb{F}^{(I)} \to V$. Explicitly, this T is simply given by $T(a) = \sum_{i\in I} a(i)\cdot f(i)$ in V for every $a\colon I \to \mathbb{F}$ that belongs to $\mathbb{F}^{(I)}$. As usual, this apparently infinite sum $\sum_{i\in I} \cdots$ is well-defined since our $a\colon I \to \mathbb{F}$ is zero except at finitely many points. One can verify that $\Phi\colon \mathrm{Lin}(\mathbb{F}^{(I)}, V) \to V^I$ and $\Psi\colon V^I \to \mathrm{Lin}(\mathbb{F}^{(I)}, V)$ are inverses isomorphisms: They are linear and $\Phi \circ \Psi = \mathrm{Id}$ and $\Psi \circ \Phi = \mathrm{Id}$.

2.2. Kernel and Image

In this section, we consider a linear transformation $T\colon V \to V'$ between $\mathbb{F}$-vector spaces V and V' and associate to it two subspaces, one of V and one of V'.

2.2.1. Definition. The *kernel* (or *nullspace*) of a linear transformation $T\colon V \to V'$ is defined to be the following subspace of V:

$$\mathrm{Ker}(T) = \{\, v \in V \,|\, T(v) = 0 \,\}$$

consisting of all vectors that get mapped to zero under T.

The dimension of this subspace is sometimes called the *nullity* of T and denoted $\mathrm{null}(T) = \dim(\mathrm{Ker}(T))$.

2.2.2. Exercise. Verify that the subset $\mathrm{Ker}(T)$ is indeed a sub*space* of V.

2.2.3. Definition. The *image* (or *range*) of a linear transformation $T\colon V \to V'$ is defined to be the following subspace of V':

$$\operatorname{Im}(T) = \left\{ T(x) \,\middle|\, x \in V \right\} = \left\{ y \in V' \,\middle|\, \text{there exists } x \in V \text{ with } T(x) = y \right\}$$

consisting of all vectors in V' that are the image of at least one vector in V.

The dimension of this subspace if always called the *rank* of T and denoted

$$\operatorname{rank}(T) = \dim(\operatorname{Im}(T)).$$

2.2.4. Remark. It is also easy to check that $\operatorname{Im}(T)$ is a subspace of V'. Let us show that $\operatorname{Im}(T)$ is closed under arbitrary linear combinations (using Proposition 1.4.3). Let $n \in \mathbb{N}$, $y_1, \ldots, y_n \in \operatorname{Im}(T)$ and $a_1, \ldots, a_n \in \mathbb{F}$. For each $i = 1, \ldots, n$ the definition of $y_i \in \operatorname{Im}(T)$ guarantees the existence of some $x_i \in V$ such that $y_i = T(x_i)$. Then we get by linearity of T that

$$\sum_{i=1}^{n} a_i \cdot y_i = \sum_{i=1}^{n} a_i \cdot T(x_i) = T(\sum_{i=1}^{n} a_i \cdot x_i).$$

This relation shows that our linear combination $\sum_{i=1}^{n} a_i \cdot y_i$ in V' is the image of the vector $\sum_{i=1}^{n} a_i \cdot x_i$ in V, hence $\sum_{i=1}^{n} a_i \cdot y_i$ belongs to $\operatorname{Im}(T)$ by definition.

2.2.5. Exercise. Let $T\colon V \to V'$ be a linear transformation and $G \subseteq V$ be a generating set, *i.e.* $\operatorname{Span}(G) = V$. Show that $T(G) = \left\{ T(x) \,\middle|\, x \in G \right\}$ is a generating set of the image of T. In formula: $T(\operatorname{Span}(G)) = \operatorname{Span}(T(G))$.

We can now prove the following celebrated result:

2.2.6. Theorem (Rank-Nullity)**.** *Let $T\colon V \to V'$ be a linear transformation of $\mathbb{F}$-vector spaces. Suppose that V is finite-dimensional. Then so are $\operatorname{Ker}(T)$ and $\operatorname{Im}(T)$. Furthermore, the dimension of V is the sum of the rank of T and the nullity of T:*

$$\dim(\operatorname{Im}(T)) + \dim(\operatorname{Ker}(T)) = \dim(V).$$

Proof. By Corollary 1.6.32, a subspace of a finite-dimensional space is finite-dimensional. Hence $\operatorname{Ker}(T) \subseteq V$ is finite-dimensional. (By Exercise 2.2.5 and the fact that V is generated by a finite basis, we know that $\operatorname{Im}(T)$ is generated by a finite set hence is finite-dimensional by Proposition 1.6.8. But we can prove the result without this step. Note that V' is not assumed finite-dimensional!)

Let $n = \dim(V)$ and $m = \dim(\operatorname{Ker}(T)) \leq n$. Let $C = \{x_1, \ldots, x_m\}$ be a basis of $\operatorname{Ker}(T)$. We know by Corollary 1.6.32 (3) that C can be completed into a basis $B = \{x_1, \ldots, x_m, x_{m+1}, \ldots, x_n\}$ of V. We now claim that the $n - m$ vectors $T(x_{m+1}), \ldots, T(x_n)$ are all distinct and form a basis of $\operatorname{Im}(T)$. This claim gives $\dim(V) = n = (n-m) + m = \dim(\operatorname{Im}(T)) + \dim(\operatorname{Ker}(T))$ hence the theorem.

Suppose that $a_{m+1}, \ldots, a_n \in \mathbb{F}$ are such that $a_{m+1}T(x_{m+1}) + \cdots + a_nT(x_n) = 0$. We need to show that $a_{m+1} = \cdots = a_n = 0$. Then $T(\sum_{i=m+1}^{n} a_i x_i) = 0$ by linearity of T and therefore $\sum_{i=m+1}^{n} a_i x_i \in \operatorname{Ker}(T) = \operatorname{Span}(x_1, \ldots, x_m)$. So we can write $\sum_{i=m+1}^{n} a_i x_i = \sum_{i=1}^{m} a_i x_i$ for some $a_1, \ldots, a_m \in \mathbb{F}$. By linear independence of $x_1, \ldots, x_m, x_{m+1}, \ldots, x_n$ this relation forces all $a_i = 0$, and in particular $a_{m+1} = \cdots = a_n = 0$ as wanted. We have thus shown that $T(x_{m+1}), \ldots, T(x_n)$ are all distinct and linearly independent. We claim they also span $\operatorname{Im}(T)$. This is easy. If $y \in \operatorname{Im}(T)$ then $y = T(x)$ for some $x \in V$. Writing x in coordinates in the basis B, we have $x = \sum_{i=1}^{n} b_i \cdot x_i$ for $b_1, \ldots, b_n \in \mathbb{F}$. Then $y = T(x) = \sum_{i=1}^{n} b_i \cdot T(x_i)$ by linearity. But the first m vectors $x_1, \ldots, x_m \in \operatorname{Ker}(T)$ map to zero under T, so the first m terms in our sum vanish: $y = \sum_{i=1}^{n} b_i \cdot T(x_i) = \sum_{i=m+1}^{n} b_i \cdot T(x_i)$ and therefore $y \in \operatorname{Span}(T(x_{m+1}), \ldots, T(x_n))$. In short $\operatorname{Im}(T) \subseteq \operatorname{Span}(T(x_{m+1}), \ldots, T(x_n))$ hence this inclusion is an equality since all $T(x_i)$ belong to $\operatorname{Im}(T)$. □

2.2.7. Exercise. If V is infinite-dimensional, then one of $\operatorname{Ker}(T)$ or $\operatorname{Im}(T)$ is infinite-dimensional but not necessarily both. Give three examples of $T\colon V \to V'$ illustrating each situation (1: finite nullity but infinite rank; 2: infinite nullity but finite rank; 3: infinite nullity and infinite rank).

2.2.8. Exercise. Let $x_1, \ldots, x_n$ be vectors of V and $T\colon V \to V'$ be linear. Show that if $T(x_1), \ldots, T(x_n)$ are distinct and linearly independent in V' then $x_1, \ldots, x_n$ are distinct and linearly independent in V. Give an easy example to show that the converse is not always true.

Recall from Appendix B that $T\colon V \to V'$ is called *surjective* or *onto* if for every $y \in V'$ there exists $x \in V$ such that $T(x) = y$. This is equivalent to $\operatorname{Im}(T) = V'$.

Recall that $T\colon V \to V'$ is called *injective* or *one-to-one* if $T(x) = T(x')$ forces $x = x'$. For linear T we can rephrase this as follows:

2.2.9. Proposition. *Let $T\colon V \to V'$ be a linear transformation. Then T is injective (one-to-one) if and only if* $\operatorname{Ker}(T) = \{0\}$.

Proof. Suppose that T is injective and pick $x \in \operatorname{Ker}(T)$. Then $T(x) = 0 = T(0)$. Hence $x = 0$ by injectivity. In short, $\operatorname{Ker}(T) = \{0\}$. That was trivial.

The interesting implication is the converse. Suppose that $\operatorname{Ker}(T) = \{0\}$. Let now $x, x' \in V$ such that $T(x) = T(x')$. We want to show that $x = x'$. Since T is linear we have $T(x - x') = T(x) - T(x') = 0$. So $x - x' \in \operatorname{Ker}(T) = \{0\}$ which means $x - x' = 0$ and therefore $x = x'$ as wanted. □

2.2.10. Remark. Of course linearity is essential here (at least 'additivity'). For instance the function $f\colon \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ has $\{\, x \in \mathbb{R} \mid f(x) = 0 \,\}$ reduced to $\{0\}$ but is not injective, as $f(-1) = f(1)$ for instance. So this function is not linear, as we know: $(x+y)^2 \neq x^2 + y^2$. The subset $f^{-1}(\{0\}) = \{\, x \in \mathbb{R} \mid f(x) = 0 \,\}$ is not called the kernel for non-linear functions, just the preimage of 0.

2.2.11. Exercise. What are the possible ranks and nullities of linear transformations $T\colon \mathbb{F}^2 \to \mathbb{F}^5$. Provide examples of T for all possible pairs (rank,nullity) in your list. Same question about $T\colon \mathbb{F}^6 \to \mathbb{F}^2$.

2.2.12. Exercise. Let $T\colon \mathbb{F}^n \to \mathbb{F}^n$ be a linear transformation such that $\operatorname{Ker}(T) = \operatorname{Im}(T)$. Show that n must be even. Conversely, if n is even, show that there exists a linear transformation $T\colon \mathbb{F}^n \to \mathbb{F}^n$ such that $\operatorname{Ker}(T) = \operatorname{Im}(T)$.

2.3. Isomorphisms

Recall that the field $\mathbb{F}$ is fixed.

In mathematics, we say that two structures are *isomorphic* (from 'iso' meaning 'same' and 'morphe' meaning 'form') if there are morphisms (homomorphisms) both ways, whose compositions are the respective identities. Here, we use linear transformations as (homo)morphisms, so the notion specializes to:

2.3.1. Definition. Let V and V' be two $\mathbb{F}$-vector spaces. We say that V and V' are *isomorphic* if there exists two linear transformations $T\colon V \to V'$ and $T'\colon V' \to V$ such that $T' \circ T = \mathrm{Id}_V$ and $T \circ T' = \mathrm{Id}_{V'}$, meaning that $T'(T(x)) = x$ for every $x \in V$ and $T(T'(y)) = y$ for every $y \in V'$. We write $V \simeq V'$ or $T\colon V \overset{\sim}{\to} V'$ and we say that $T\colon V \overset{\sim}{\to} V'$ is an *isomorphism* between V and V'. Alternatively, one often says that a linear transformation $T\colon V \to V'$ is *invertible* if it is an isomorphism.

2.3.2. Example. The 'identifications' we wanted to do in Chapter 1 are in fact (canonical) isomorphisms. For instance, $\mathrm{M}_{p\times q}(\mathbb{F}) \simeq \mathbb{F}^{pq}$ by re-arranging indices. Similarly $\mathbb{F}[X] \simeq \mathbb{F}^{(\mathbb{N})}$ and $\left\{ P \in \mathbb{F}[X] \,\middle|\, \deg(P) \le d \right\} \simeq \mathbb{F}^{d+1}$ via obvious isomorphisms $(a_0 + a_1X + a_2X^2 + \cdots) \mapsto (a_0, a_1, a_2, \ldots)$.

2.3.3. Example. Recall from elementary linear algebra that a square matrix $A \in \mathrm{M}_{n\times n}(\mathbb{F})$ is *invertible* if there exists a matrix $C \in \mathrm{M}_{n\times n}(\mathbb{F})$ such that $A \cdot C = C \cdot A = I_n$, where $I_n = [\delta_{ij}]_{1\le i,j\le n}$ is the identity matrix. If it exists, this matrix C is unique and denoted A^{-1}. If $A \in \mathrm{M}_{n\times n}(\mathbb{F})$ is invertible, with inverse A^{-1}, then $T_A\colon \mathbb{F}^n \overset{\sim}{\to} \mathbb{F}^n$ is an isomorphism with inverse $(T_A)^{-1} = T_{A^{-1}}$ since $T_{A^{-1}} \circ T_A = T_{A^{-1}A} = T_{I_n} = \mathrm{Id}$ and similarly $T_A \circ T_{A^{-1}} = \mathrm{Id}$.

Perhaps it is good to refresh the calculation of the inverse.

2.3.4. Exercise. Let $A \in \mathrm{M}_{n\times n}(\mathbb{F})$. Apply Gauss-Jordan to the $n \times (2n)$-matrix $[A|I_n]$ obtained by copying the identity next to A. Show that the output is of the form $[I_n|B]$ (*i.e.* the n first columns contain a pivot) if and only if A is invertible. Show that in that case, B is the inverse of A. [Hint: See how this is solving n linear systems simultaneously, namely $A \cdot x_i = e_i$ for $i = 1, \ldots, n$.]

2.3.5. Remark. Recall the set-theoretic notion of bijection (Appendix B). It follows from Definition 2.3.1 that $T\colon V \overset{\sim}{\to} V'$ being an isomorphism forces T to be a bijection between the sets V and V'. Its set-theoretic inverse $T^{-1}\colon V' \to V$ is the T' of Definition 2.3.1, which is required to be linear. In fact, this is automatic.

2.3.6. Lemma. *Suppose that a linear transformation $T\colon V \to V'$ is bijective and let $T^{-1}\colon V' \to V$ be its inverse as a function. Then T^{-1} is linear as well.*

Proof. Let $n \in \mathbb{N}$, $y_1, \ldots, y_n \in V'$ and $a_1, \ldots, a_n \in \mathbb{F}$. We want to show that

$$T^{-1}\big(\sum_{i=1}^{n} a_i \cdot y_i\big) = \sum_{i=1}^{n} a_i \cdot T^{-1}(y_i) \tag{2.3.7}$$

in V. Since $T\colon V \to V'$ is injective, it suffices to check that the image under T of those two vectors are the same in V'. And indeed, we compute in V'

$$\begin{aligned} T\Big(T^{-1}\big(\textstyle\sum_{i=1}^{n} a_i \cdot y_i\big)\Big) &= \textstyle\sum_{i=1}^{n} a_i \cdot y_i && \text{since } T \circ T^{-1} = \mathrm{Id} \\ &= \textstyle\sum_{i=1}^{n} a_i \cdot T(T^{-1}(y_i)) && \text{since } T \circ T^{-1} = \mathrm{Id} \\ &= T\big(\textstyle\sum_{i=1}^{n} a_i \cdot T^{-1}(y_i)\big) && \text{since } T \text{ is linear.} \end{aligned}$$

This proves the equality (2.3.7). Thus T^{-1} is linear. □

2.3.8. Proposition. *Let $T\colon V \to V'$ be a linear transformation of $\mathbb{F}$-vector spaces. The following are equivalent:*

(i) *T is an isomorphism (*i.e.* T is invertible).*
(ii) *T is a bijection as a map of sets (*i.e.* it is one-to-one and onto).*

(iii) $\mathrm{Ker}(T) = \{0\}$ *and* $\mathrm{Im}(T) = V'$.

PROOF. For (ii)⇔(iii), we have seen that T is injective $\iff$ $\mathrm{Ker}(T) = \{0\}$ in Proposition 2.2.9 and we have T surjective $\iff$ $\mathrm{Im}(T) = V'$ by definition.

(i)⇒(ii) is immediate by definition of isomorphism. (See Remark 2.3.5.) For (ii)⇒(i), suppose that T is a bijection. Lemma 2.3.6 tells us that T^{-1} is linear, hence T is an isomorphism by Definition 2.3.1 with $T' = T^{-1}$. □

2.3.9. Exercise. Being isomorphic $\simeq$ is an equivalence relation. ([1]) Prove:

(1) For any V, the identity Id_V is an isomorphism. Hence reflexivity: $V \simeq V$. [Beware that we are absolutely *not* saying that *any* linear $T\colon V \to V$ is an isomorphism! Remember to test silly guesses with $T = 0$ for instance.]

(2) If $T\colon V \overset{\sim}{\to} V'$ is an isomorphism then $T^{-1}\colon V' \to V$ is also an isomorphism with $(T^{-1})^{-1} = T$. Hence symmetry: If $V \simeq V'$ then $V' \simeq V$.

(3) Show that if $T\colon V \overset{\sim}{\to} V'$ and $S\colon V' \overset{\sim}{\to} V''$ are isomorphisms then their composition $S \circ T\colon V \to V''$ is an isomorphism as well and that
$$(S \circ T)^{-1} = T^{-1} \circ S^{-1}.$$
Hence transitivity: If $V \simeq V'$ and if $V' \simeq V''$ then $V \simeq V''$.

We can use the notion of isomorphism to revisit what it means to be a (finite, ordered) basis.

2.3.10. Theorem. *Let B be an ordered collection (an n-tuple) of vectors $x_1, \ldots, x_n$ in a vector space V and consider the following linear transformation* ([2])

$$\begin{array}{cccc} S_B\colon & \mathbb{F}^n & \longrightarrow & V \\ & \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} & \longmapsto & a_1 \cdot x_1 + \cdots + a_n \cdot x_n. \end{array}$$

(Here, we write vectors of $\mathbb{F}^n$ in columns.) Then we have:

(1) *S_B is surjective (onto) if and only if the vectors $x_1, \ldots, x_n$ generate V.*

(2) *S_B is injective (one-to-one) if and only if the vectors $x_1, \ldots, x_n$ are all distinct and linearly independent.*

(3) *$S_B\colon \mathbb{F}^n \overset{\sim}{\to} V$ is an isomorphism if and only if B is an ordered basis of V.*

Moreover, when B is a basis of V, then the inverse $V \overset{\sim}{\to} \mathbb{F}^n$ of the isomorphism $S_B\colon \mathbb{F}^n \overset{\sim}{\to} V$ is nothing but the coordinate function with respect to B:
$$(S_B)^{-1} = [-]_B.$$

PROOF. For (1), asking that every $y \in V$ can be written as $S_B(\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix})$ for some $a_1, \ldots, a_n \in \mathbb{F}$ is both the surjectivity of S_B and the fact that B spans V.

For (2), asking that S_B is one-to-one is the same as $\mathrm{Ker}(S_B) = \{0\}$ and this is exactly saying that the only $\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$ such that $S_B(\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}) = 0$ is $\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$. The latter precisely says that the $x_1, \ldots, x_n$ are all distinct and linearly independent.

[1] Except that all $\mathbb{F}$-vector spaces taken together do not form a set, but a proper class.

[2] In the notation of Theorem 2.1.23, the linear transformation S_B is characterized by the requirement that $S_B(e_i) = x_i$ for all $i = 1, \ldots, n$, where $e_1, \ldots, e_n$ is the canonical basis of $\mathbb{F}^n$.

For (3), we combine (1) and (2). Now for the description of $(S_B)^{-1}$, let us unpack the definition of the inverse: For each $v \in V$ we set $S_B^{-1}(v)$ to be the unique $\underline{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \in \mathbb{F}^n$ such that $S_B(\underline{a}) = v$, which means $a_1 \cdot x_1 + \cdots + a_n \cdot x_n = v$. This n-tuple $\underline{a}$ is exactly the definition of $[v]_B$, the coordinates of v in the basis B. □

2.3.11. Exercise. The finite dimension was not essential above. If B is a basis of V then we can define $S_B \colon \mathbb{F}^{(B)} \to V$ in the same way (see Remark 2.1.29). Namely, we let $S_B(a) = \sum_{b \in B} a(b) \cdot b$ for any $a \in \mathbb{F}^{(B)}$, that is a function $a \colon B \to \mathbb{F}$ which is zero almost everywhere. Again, S_B is surjective because B generates and injective because B is linearly independent. Hence S_B is an isomorphism between the free $\mathbb{F}$-vector space over the set B and the vector space V. The inverse isomorphism $[-] \colon V \to \mathbb{F}^{(B)}$ is again the coordinate function, by definition.

There is a strong connection between isomorphism and dimension. Let us gather all information in one statement.

2.3.12. Theorem. *Let V and V' be vector spaces over $\mathbb{F}$. Then we have:*

(1) *If $V \simeq V'$ are isomorphic, they have the same dimension: Either they are both infinite-dimensional or they are both finite-dimensional and* $\dim(V) = \dim(V')$*.*

(2) *If V has finite dimension n then V is isomorphic to $\mathbb{F}^n$. More precisely, if B is a basis of V then we have an isomorphism $[-]_B \colon V \overset{\sim}{\to} \mathbb{F}^n$ defined by sending a vector to its coordinates $x \mapsto [x]_B$.*

(3) *Suppose that V and V' are finite-dimensional. Then V and V' are isomorphic if and only if* $\dim(V) = \dim(V')$*.*

Proof. Part (1) follows from the meta-statement: *Isomorphic spaces share all linear properties.* Specifically, if $T \colon V \overset{\sim}{\to} V'$ is an isomorphism, and $L \subseteq V$ is a subset and $L' = T(L)$ is the corresponding subset of V' then L is linearly independent in V if and only if L' is linearly independent in V'. Similarly, L spans V if and only if L' spans V'. Etc. These are good easy exercises using (2.1.2) and bijectivity of T. In particular, L is a basis of V if and only if its image L' is a basis of V'. As $T \colon L \to L'$ is a bijection, L and L' are simultaneously finite, and when finite they have the same number of elements. This gives (1).

Part (2) is direct from Theorem 2.3.10. We have an isomorphism $S_B \colon \mathbb{F}^n \overset{\sim}{\to} V$ and $[-]_B \colon V \overset{\sim}{\to} \mathbb{F}^n$ is its inverse.

Part (3) follows easily. We have already seen one direction in (1). Conversely, if V and V' have the same finite dimension n then $V \simeq \mathbb{F}^n$ by (2) and $\mathbb{F}^n \simeq V'$ by (2) and therefore $V \simeq V'$ by transitivity of $\simeq$. See Exercise 2.3.9 if necessary. Unpacking the proof, we can take a basis $x_1, \ldots, x_n$ of V, a basis $x'_1, \ldots, x'_n$ of V' (for the same n since we assume $\dim(V) = \dim(V')$) and define an isomorphism $T \colon V \to V'$ by extending $\mathbb{F}$-linearly the assignment on the basis $T(x_i) = x'_i$ for all i, as in Theorem 2.1.23. Its inverse $T' \colon V' \to V$ is defined by extending $T'(x'_i) = x_i$ for all i. We have $T' \circ T = \mathrm{Id}_V$ and $T \circ T' = \mathrm{Id}_{V'}$ since these relations hold on the basis (by the uniqueness part of Theorem 2.1.23). □

2.3.13. Remark. If $A \in \mathrm{M}_{p \times q}(\mathbb{F})$ and $C \in \mathrm{M}_{q \times p}(\mathbb{F})$ satisfy $A \cdot C = I_p$ and $C \cdot A = I_q$ then $T_A \colon \mathbb{F}^q \to \mathbb{F}^p$ is an isomorphism with inverse $T_C \colon \mathbb{F}^p \to \mathbb{F}^q$. In that case Theorem 2.3.12 (1) forces $p = \dim(\mathbb{F}^p) = \dim(\mathbb{F}^q) = q$. So only square matrices should be considered for invertibility as we learned in introductory linear algebra (see Example 2.3.3).

In the spirit of Corollary 1.6.31, we can show that knowing that 'the dimension matches' cuts our work in half:

2.3.14. Corollary. *Let V and V' be two $\mathbb{F}$-vector spaces of the* same *finite dimension:* $\dim(V) = \dim(V') = n < \infty$. *Let $T\colon V \to V'$ be a linear transformation. The following are equivalent:*

(i) *T is injective (one-to-one), or equivalently* $\operatorname{Ker}(T) = \{0\}$.
(ii) *T is surjective (onto), or equivalently* $\operatorname{Im}(T) = V'$.
(iii) *T has rank n.*
(iv) *T is an isomorphism.*

Proof. The equivalent formulations of (i) are given in Proposition 2.2.9. The implications (iv)⇒(i) and (iv)⇒(ii)⇒(iii) are very easy, almost by definition. If (iii) holds, since $\operatorname{Im}(T) \subseteq V'$ is a subspace, knowing that $\dim(\operatorname{Im}(T)) = \operatorname{rank}(T) = n = \dim(V')$ forces $\operatorname{Im}(T) = V'$ by Corollary 1.6.32 (ii). So (iii)⇒(ii) as well. So

$$\text{(i)} \Longleftarrow \text{(iv)} \Longrightarrow \text{(ii)} \Longleftrightarrow \text{(iii)}.$$

The proofs that (i)⇒(iv) and (ii)⇒(iv) are essentially the same. Let us show that, under our hypothesis that $\dim(V) = \dim(V')$, the transformation T is bijective as soon as it is injective or surjective. By Rank-Nullity Theorem 2.2.6, we have

$$\dim(\operatorname{Ker}(T)) + \dim(\operatorname{Im}(T)) = n. \tag{2.3.15}$$

So if we suppose (i) then we get from (2.3.15) that $\operatorname{rank}(T) = \dim(\operatorname{Im}(T)) = n - 0 = n$ as in (iii), hence T is surjective and therefore bijective. On the other hand, if we suppose (ii) then we get from (2.3.15) that $\dim(\operatorname{Ker}(T)) = n - n = 0$, hence T is injective and therefore bijective, again. In both cases, we know from Proposition 2.3.8 that T bijective implies that T is an isomorphism. So both (i)⇒(iv) and (ii)⇒(iv) and we have established all implications. □

2.3.16. Exercise. The assumption that the dimension n is finite is essential in Corollary 2.3.14. Show using the linear transformations of Exercises 2.1.11 and 2.1.12 that on an infinite-dimensional vector space V we can have $T\colon V \to V$ that is injective but not surjective, or surjective but not injective, although clearly V and V have the same dimension (which is infinite here).

2.3.17. Remark. We can use the notion of isomorphism to clarify the notation $W_1 \oplus W_2$ for subspaces $W_1, W_2 \subseteq V$. We defined $W_1 \oplus W_2$ in Remark 1.2.26, for any vector spaces W_1, W_2 (subspaces are spaces!) and it meant the 'external' cartesian product $W_1 \oplus W_2 := W_1 \times W_2 = \{\,(w_1, w_2) \,|\, w_i \in W_i\,\}$. On the other hand, we can construct $W_1 + W_2 = \{\,w_1 + w_2 \,|\, w_i \in W_i\,\}$ 'internally' to V. We have a linear map $S\colon W_1 \times W_2 \to W_1 + W_2$ which sums the components $S(w_1, w_2) = w_1 + w_2$. It is surjective by definition of $W_1 + W_2$ and its kernel is isomorphic to $W_1 \cap W_2$ via $W_1 \cap W_2 \to \operatorname{Ker}(S)$ given by $w \mapsto (w, -w)$ and inverse $\operatorname{Ker}(S) \to W_1 \cap W_2$ given by $(w_1, w_2) \mapsto w_1$. (This is an easy exercise.) So when $W_1 \cap W_2 = 0$, we have an isomorphism $S\colon W_1 \oplus W_2 = W_1 \times W_2 \overset{\sim}{\to} W_1 + W_2$ and one often writes $W_1 \oplus W_2$ for the latter (inside V). In other words, when you see $W_1 \oplus W_2 \subseteq V$ for subspaces $W_1, W_2 \subseteq V$ it refers to the subspace $W_1 + W_2$ but it also tells you that $W_1 \cap W_2 = 0$. Similarly for $W_1 \oplus \cdots \oplus W_n$ inside V, for any $n \geq 2$.

2.4. Matrix of linear transformation

2.4.1. Remark. We proved in Theorem 2.1.23 that every linear transformation $T\colon V \to V'$ is entirely characterized by what it does on a basis of V. We combine this with a basis of V' and reduce the information about T to a table of scalars.

2.4.2. Definition. Let V and V' be two $\mathbb{F}$-vector spaces of finite dimension $p = \dim(V)$ and $q = \dim(V')$. Let $B = \{b_1, \ldots, b_p\}$ be an ordered basis of V and $B' = \{b'_1, \ldots, b'_q\}$ be an ordered basis of V'. Let $T\colon V \to V'$ be a linear transformation.

For every $j = 1, \ldots, p$ we can write the vector $T(b_j)$ of V' in coordinates in the basis B'; this gives a column vector $[T(b_j)]_{B'}$ in $\mathbb{F}^q$. (See Definition 1.6.37 if necessary.) Let us call this coordinate vector

$$\begin{bmatrix} a_{1j} \\ \vdots \\ a_{qj} \end{bmatrix} := [T(b_j)]_{B'} .$$

We cannot just call it $\begin{bmatrix} a_1 \\ \vdots \\ a_q \end{bmatrix}$ as it of course depends on which j we have chosen, *i.e.* which vector b_j we consider in the basis B. By definition of $[-]_{B'}$ this means that we found scalars $a_{1j}, \ldots, a_{qj} \in \mathbb{F}$ such that the following equality holds in V'

$$T(b_j) = \sum_{i=1}^{q} a_{ij} \cdot b'_i. \tag{2.4.3}$$

Recall that j was any index between 1 and p. So we obtain p such columns in $\mathbb{F}^q$; in other words we obtain a $(\mathbf{q} \times \mathbf{p})$-matrix whose j-th column is the above $\begin{bmatrix} a_{1j} \\ \vdots \\ a_{qj} \end{bmatrix}$:

$$A = \Big(a_{ij}\Big)_{\substack{1\le i\le q\\ 1\le j\le p}} = \begin{pmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1p} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{q1} & \cdots & a_{qj} & \cdots & a_{qp} \end{pmatrix}.$$

We call this the *matrix of T with respect to the bases B and B'* and write it as

$$[T]_{B',B} \quad \text{in} \quad \mathrm{M}_{q\times p}(\mathbb{F}).$$

If we use the notation $A = \big(* \mid * \mid \cdots \mid * \big)$ to indicate the columns of a matrix A then the definition of $[T]_{B',B}$ in condensed form is:

$$[T]_{B',B} = \Big(\big[T(b_1)\big]_{B'} \;\Big|\; \big[T(b_2)\big]_{B'} \;\Big|\; \cdots \;\Big|\; \big[T(b_p)\big]_{B'} \Big) \quad \text{where} \quad B = \{b_1, \ldots, b_p\}.$$

2.4.4. Remark. Some authors will denote $[T]_{B',B}$ as $[T]_B^{B'}$. It is fine (and students should be allowed to use either notation). We like the notation $[T]_{B',B}$ for it poetically evokes a matrix of size $B' \times B$. We return to this in Remark 2.4.20.

2.4.5. Example. Let $T = T_A$ be a linear transformation $V = \mathbb{F}^p \to V' = \mathbb{F}^q$ already given by a matrix $A \in \mathrm{M}_{q\times p}(\mathbb{F})$. Is the matrix of T_A equal to A? The answer is no in general! It depends on the bases. But *if* you use the canonical bases $B = \{e_1, \ldots, e_p\}$ and $B' = \{e'_1, \ldots, e'_q\}$ ([3]) of $V = \mathbb{F}^p$ and $V' = \mathbb{F}^q$ then indeed $\big[T_A\big]_{B',B} = A$. This is easy to verify since coordinates in canonical bases are so straightforward and since $A \cdot e_j$ is simply the j-th column of A.

[3] Here we pay the price for writing $e_1, e_2, \ldots$ independently of the n of $\mathbb{F}^n$.

2.4.6. Exercise. With notation as in Definition 2.4.2, show that $[T]_{B',B}$ is linear *in* T, meaning that $[T_1+T_2]_{B',B} = [T_1]_{B',B} + [T_2]_{B',B}$ and $[a \cdot T]_{B',B} = a \cdot [T]_{B',B}$ for every $T, T_1, T_2 \colon V \to V'$ linear and $a \in \mathbb{F}$. In other words, we obtain a linear transformation $[-]_{B',B} \colon \mathrm{Lin}(V,V') \to \mathrm{M}_{q\times p}(\mathbb{F})$.

Now in full generality, we can recover T from its matrix.

2.4.7. Proposition. *Let V, V' be $\mathbb{F}$-vector spaces of finite dimension p and q respectively. Let B be a basis of V and B' a basis of V'. Let $T \colon V \to V'$ be linear and let $A = [T]_{B',B} \in \mathrm{M}_{q\times p}(\mathbb{F})$ be the matrix of T with respect to those bases.*

Let $v \in V$. Let $x = [v]_B \in \mathbb{F}^p$ and $y = [T(v)]_{B'} \in \mathbb{F}^q$ be the coordinates of v and of its image $T(v)$ with respect to our bases. Then we can compute y in terms of $A = \big(a_{ij}\big)_{1\le i\le q, 1\le j\le p}$ and x as follows:

$$y_i = \sum_{j=1}^{p} a_{ij} \cdot x_j$$

for every $1 \le i \le q$. In condensed form, we have for every $v \in V$

$$\big[T(v)\big]_{B'} = \big[T\big]_{B',B} \cdot \big[v\big]_B \tag{2.4.8}$$

where the right-hand $\cdot$ is matrix multiplication.

Proof. We adopt the notation of Definition 2.4.2. In particular, the bases are $B = \{b_1, \ldots, b_p\}$ and $B' = \{b'_1, \ldots, b'_q\}$ and the entries a_{ij} of $A = [T]_{B',B}$ are characterized by (2.4.3). The meaning of $x = [v]_B$ is that we have $v = \sum_{j=1}^{p} x_j \cdot b_j$ in V. Applying T to this equality, we compute in V'

$$\begin{aligned} T(v) &= \textstyle\sum_{j=1}^{p} x_j \cdot T(b_j) && \text{by linearity of } T \\ &= \textstyle\sum_{j=1}^{p} x_j \cdot (\sum_{i=1}^{q} a_{ij} \cdot b'_i) && \text{by construction of the } a_{ij} \text{ in (2.4.3)} \\ &= \textstyle\sum_{j=1}^{p} \sum_{i=1}^{q} (x_j a_{ij}) \cdot b'_i && \text{by distributivity, etc} \\ &= \textstyle\sum_{i=1}^{q} \big(\sum_{j=1}^{p} (a_{ij} x_j)\big) \cdot b'_i && \text{by rearranging terms.} \end{aligned}$$

The last equation describes $T(v)$ as a linear combination of $b'_1, \ldots, b'_q$ hence we can read the (unique) coordinate $y_1, \ldots, y_q$ of $T(v)$ in the basis B' from it. It says precisely $y_i = \sum_{j=1}^{p} (a_{ij} x_j)$ for every $i = 1, \ldots, q$ as announced in the statement. □

2.4.9. Corollary. *Let $T \colon \mathbb{F}^p \to \mathbb{F}^q$ be a linear transformation. Then $T = T_A$ for a unique $(q \times p)$-matrix A. It is given by $A = [T]_{B',B}$ with respect to the canonical bases B and B' of $V = \mathbb{F}^p$ and $V' = \mathbb{F}^q$ respectively.*

Proof. The beauty of canonical bases is that (if we write $x \in \mathbb{F}^p$ in columns) then $[x]_B = x$. Similarly, $[y]_{B'} = y$ for any $y \in \mathbb{F}^q$. Reading (2.4.8) in those canonical bases then tells us that $T(v) = [T(v)]_{B'} = [T]_{B',B} \cdot [v]_B = [T]_{B',B} \cdot v$. The latter is simply $T_A(v)$ for $A = [T]_{B',B}$. For uniqueness, suppose that $T = T_A$ for some A and note that $[T]_{B',B} = A$ since $A \cdot e_j$ is simply the j-th column of A. So we can recover A from the linear transformation T_A, by computing its matrix in the canonical bases. □

2.4.10. Remark. The matrix $[T]_{B',B}$ is the *unique* matrix that satisfies (2.4.8). In other words, in the notation of Proposition 2.4.7, if $A \in \mathrm{M}_{q\times p}(\mathbb{F})$ is some matrix that gives the coordinates of the image of *every* vector $v \in V$ by the recipe

$$[T(v)]_{B'} = A \cdot [v]_B$$

then $A = [T]_{B',B}$ is the matrix of T constructed in Definition 2.4.2. To see that, it suffices to plug $v = b_j$ in the above equation, for each $j = 1, \ldots, p$ in turn. The left-hand side gives our friend $[T(b_j)]_{B'}$ and the right-hand side gives $A \cdot [b_j]_B = A \cdot e_j$ where e_j is the j-th canonical vector. And $A \cdot e_j$ is the j-th column of A. In short, the j-th column of A must be $[T(b_j)]_{B'}$ for every $j = 1, \ldots, p$, as in Definition 2.4.2.

2.4.11. Remark. We can express Proposition 2.4.7 by saying that the diagram

$$\begin{tikzcd} V \arrow[r, "\sim", "{[-]_B}"'] \arrow[d, "T"'] & \mathbb{F}^p \arrow[d, dotted, "T_A"] \\ V' \arrow[r, "\sim", "{[-]_{B'}}"'] & \mathbb{F}^q \end{tikzcd} \tag{2.4.12}$$

'commutes'. Let us explain this. Horizontally, we have written the isomorphisms of Theorem 2.3.12 (2) that turn an n-dimensional $\mathbb{F}$-vector space into $\mathbb{F}^n$ using coordinates. Vertically, on the left, we have our linear transformation T which has no particular property. We can then follow the diagram by composing linear transformations, starting from $\mathbb{F}^p$ in the upper right corner, going left to V via the inverse of $[-]_B$ (that we called S_B in Theorem 2.3.10), then moving down to V' via T and finally moving right to $\mathbb{F}^q$ via $[-]_{B'}$. [The reader *should* follow this path with their finger on the above diagram!] This composite $\mathbb{F}^p \to \mathbb{F}^q$ must be given by multiplication by a matrix, that is by T_A, for some $(q \times p)$-matrix A, by Corollary 2.4.9. That matrix is unique. It depends on our transformation T but also on the isomorphisms $[-]_B$ and $[-]_{B'}$, that is, on our choices of bases. By Proposition 2.4.7 this A is precisely our $[T]_{B',B}$. The commutativity of the diagram means that if you start with an element $v \in V$ in the upper-left corner and compose the maps along either sides of the diagram you get the same answer in the bottom-right corner. Indeed, going down and then right gives $[T(v)]_{B'}$ and going right and then down gives $A \cdot [v]_B$ for $A = [T]_{B',B}$. The fact that the two agree is exactly (2.4.8).

Another way to summarize our discussion is the following:

2.4.13. Corollary. *Let V, V' be $\mathbb{F}$-vector spaces, with respective bases B, B' and dimensions $p = \dim(V)$, $q = \dim(V')$, as in Proposition 2.4.7. Then the function*

$$\begin{array}{rccc} [-]_{B',B'}\colon & \mathrm{Lin}(V,V') & \longrightarrow & \mathrm{M}_{q\times p}(\mathbb{F}) \\ & T & \longmapsto & [T]_{B',B} \end{array}$$

is an isomorphism.

Proof. The linearity of the construction is Exercise 2.4.6. Injectivity is direct from Equation (2.4.8): If $[T]_{B',B} = 0$ then $T(v) = 0$ for all $v \in V$, hence $T = 0$. For surjectivity, choose $A = [a_{ij}] \in \mathrm{M}_{q\times p}(\mathbb{F})$ and define $T\colon V \to V'$ as the composite $S_{B'} \circ T_A \circ [-]_B$ in the spirit of Remark 2.4.11 (but now T_A is known and we get T out of it). Explicitly, define the images $T(b_j) \in V'$ of the basis vectors $b_j \in B$ by giving them in coordinates in the basis B', that is, by setting $T(b_j) = \sum_{i=1}^{q} a_{ij} \cdot b'_i$ as in (2.4.3). This assignment admits a (unique) extension $T\colon V \to V'$ that is linear by Theorem 2.1.23 again. By construction $[T(b_j)]_{B'}$ is the j-th column of A for all $j = 1, \ldots, p$, hence $[T]_{B',B} = A$. □

2.4.14. Corollary. *Let V and V' be finite-dimensional $\mathbb{F}$-vector spaces. Then* $\operatorname{Lin}(V,V')$ *is finite-dimensional and* $\dim(\operatorname{Lin}(V,V')) = \dim(V)\cdot\dim(V')$. □

2.4.15. Remark. So we see that matrix multiplication do not really exist out there and are then used to do linear transformations. It is rather the other way around. Linear transformations are out there and, on finite-dimensional vector spaces, they can be completely encoded in a finite table. We can think of $[T]_{B',B}$ as the 'coordinates' of the linear transformation. When we want to recover the transformation from that table, we obtain the formula of Proposition 2.4.7 in coordinates. We then call this type of formulas matrix multiplication. It is the reason we do matrix multiplication, its justification. Almost! You may argue that matrix multiplication is defined more generally than for just a matrix and a column vector as in (2.4.8). Well, it is true but this also has a conceptual background, as we now explain.

2.4.16. Theorem. *Let $T\colon V \to V'$ and $S\colon V' \to V''$ be two linear transformations between $\mathbb{F}$-vector spaces V, V', V'' of finite dimensions p, q, r respectively. Let B be a basis of V, let B' be a basis of V' and let B'' be a basis of V''. Then the matrix of $S \circ T\colon V \to V''$ is given by matrix multiplication*

$$\big[S\circ T\big]_{B'',B} = \big[S\big]_{B'',B'}\cdot\big[T\big]_{B',B}. \tag{2.4.17}$$

Proof. Let us name the basis vectors as $B' = \{b_1,\ldots,b_p\}$, $B' = \{b'_1,\ldots,b'_q\}$ and $B'' = \{b''_1,\ldots,b''_r\}$. Let us write $A = \big(a_{ij}\big) \in \mathrm{M}_{q\times p}(\mathbb{F})$ for the matrix $[T]_{B',B}$ and since B is already taken, let us write $C = \big(c_{k\ell}\big) \in \mathrm{M}_{r\times q}(\mathbb{F})$ for the matrix $[S]_{B'',B'}$. We want to show that $C\cdot A$ is the matrix $[S\circ T]_{B'',B}$. We have by construction of $[T]_{B',B}$ and $[S]_{B'',B'}$, following (2.4.3), that

$$T(b_j) = \sum_{i=1}^{q} a_{ij}\cdot b'_i \qquad \text{for all } j = 1,\ldots,p \tag{2.4.18}$$

$$S(b'_i) = \sum_{k=1}^{r} c_{ki}\cdot b''_k \qquad \text{for all } i = 1,\ldots,q. \tag{2.4.19}$$

We want to compute $[S\circ T]_{B'',B}$ one column at a time, as in Definition 2.4.2. Let $1 \le j \le p$ and consider $(S\circ T)(b_j) = S(T(b_j))$ that we need to write in coordinates in the basis B''. Let's go, we compute in V'':

$$\begin{aligned}
S(T(b_j)) &= S(\textstyle\sum_{i=1}^{q} a_{ij}\cdot b'_i) && \text{by (2.4.18)}\\
&= \textstyle\sum_{i=1}^{q} a_{ij}\cdot S(b'_i) && \text{by linearity of } S\\
&= \textstyle\sum_{i=1}^{q} a_{ij}\cdot\big(\sum_{k=1}^{r} c_{ki}\cdot b''_k\big) && \text{by (2.4.19)}\\
&= \textstyle\sum_{i=1}^{q}\sum_{k=1}^{r}(a_{ij}c_{ki})\cdot b''_k && \text{by distributivity}\\
&= \textstyle\sum_{k=1}^{r}\big(\sum_{i=1}^{q}(c_{ki}a_{ij})\big)\cdot b''_k && \text{by rearranging.}
\end{aligned}$$

And we see that the k-th coordinate of $(S\circ T)(b_j)$ in the basis $B'' = \{b''_1,\ldots,b''_r\}$ is $\sum_{i=1}^{q}(c_{ki}a_{ij})$ which is indeed the (k,j)-entry of $C\cdot A$, as announced. □

2.4.20. Remark. In the important formula (2.4.17) our notation $[T]_{B',B}$ reminds us of the rules about matrix multiplication: *Only multiply matrices of size $r\times q$ and $q\times p$ for matching q and then the product will be an $(r\times p)$-matrix.* This numerology is now explained: The intermediate basis (called B' here) should be the same for the incoming transformation $\cdots\to V'$ and the outgoing one $V'\to\cdots$.

2.4.21. Remark. One can give another proof of Theorem 2.4.16, via the associativity of matrix multiplication. (The latter is relatively easy to check from the explicit formulas.) In colloquial terms, Remark 2.4.10 says that if we can prove that the product matrix $[S] \cdot [T]$ makes (2.4.8) 'work' for the linear transformation $S \circ T$ then by uniqueness, this product $[S] \cdot [T]$ must be $[S \circ T]$. Writing more carefully, Remark 2.4.10 says that in order to show that the product $[S]_{B'',B'} \cdot [T]_{B',B}$ is indeed $[S \circ T]_{B'',B}$, it suffices to check that for every $v \in V$ we have

$$[(S \circ T)(v)]_{B''} = ([S]_{B'',B'} \cdot [T]_{B',B}) \cdot [v]_B. \tag{2.4.22}$$

We have by Proposition 2.4.7, that for every $v \in V$

$$[T(v)]_{B'} = [T]_{B',B} \cdot [v]_B. \tag{2.4.23}$$

Applying the same Proposition 2.4.7 to $S\colon V' \to V''$ we get

$$[S(w)]_{B''} = [S]_{B'',B'} \cdot [w]_{B'} \tag{2.4.24}$$

for every $w \in V'$, in particular for $w = T(v)$. If we combine these matrix equations we get for every $v \in V$ that

$$\begin{aligned} [(S \circ T)(v)]_{B''} &= [S(T(v))]_{B''} && \text{by definition of } S \circ T \\ &= [S]_{B'',B'} \cdot [T(v)]_{B'} && \text{by (2.4.24) for } w = T(v) \\ &= [S]_{B'',B'} \cdot ([T]_{B',B} \cdot [v]_B) && \text{by (2.4.23)} \\ &= ([S]_{B'',B'} \cdot [T]_{B',B}) \cdot [v]_B && \text{by associativity.} \end{aligned}$$

This is the wanted relation and we conclude that $[S]_{B'',B'} \cdot [T]_{B',B} = [S \circ T]_{B'',B}$.

2.4.25. Exercise. Consider $T = \mathrm{Id}_V\colon V \to V$ and two finite bases B and B' of V. (Here $V' = V$.) Give necessary and sufficient conditions for $[T]_{B',B}$ to be the identity matrix I_n.

2.4.26. Corollary. *Let $T\colon V \to V'$ be a linear transformation between vector spaces of finite dimension n. Then the following are equivalent:*

(i) *The transformation T is invertible.*

(ii) *For all bases B of V and B' of V', the matrix $[T]_{B',B} \in \mathrm{M}_{n\times n}(\mathbb{F})$ of T with respect to the bases B and B' is invertible.*

(iii) *There exists bases B of V and B' of V' such that $[T]_{B',B}$ is invertible.*

In that case, with notation as in (ii) or (iii), the matrix of $T^{-1}\colon V' \overset{\sim}{\to} V$ is

$$[T^{-1}]_{B,B'} = \big([T]_{B',B}\big)^{-1}. \tag{2.4.27}$$

Proof. (i)⇒(ii): Apply Theorem 2.4.16 twice, to the relations $T^{-1} \circ T = \mathrm{Id}_V$ and $T \circ T^{-1} = \mathrm{Id}_{V'}$. We get matrix equalities $[T^{-1}]_{B,B'} \cdot [T]_{B',B} = [\mathrm{Id}_V]_{B,B} = I_n$ and $[T]_{B',B} \cdot [T^{-1}]_{B,B'} = [\mathrm{Id}_{V'}]_{B',B'} = I_n$. (In both cases, it is important to have the matrix of the identity with respect to twice the same basis.) Hence $A = [T]_{B',B}$ and $C = [T^{-1}]_{B,B'}$ satisfy $A \cdot C = C \cdot A = I_n$. Which proves (ii), and (2.4.27).

(ii)⇒(iii) is obvious: Pick *any* bases B and B'.

(iii)⇒(i): Suppose that $A = [T]_{B',B}$ is invertible. Recall from Corollary 2.4.13 that $[-]_{B,B'}\colon \mathrm{Lin}(V',V) \overset{\sim}{\to} \mathrm{M}_{n\times n}(\mathbb{F})$ is an isomorphism. So there exists a unique linear transformation $T'\colon V' \to V$ such that $[T']_{B,B'} = A^{-1}$. Then by Theorem 2.4.16 we know that $[T' \circ T]_{B,B} = A^{-1} \cdot A = I_n = [\mathrm{Id}_V]_{B,B}$ which means (Corollary 2.4.13 again) that $T' \circ T = \mathrm{Id}_V$. In the same way, we prove $T \circ T' = \mathrm{Id}_{V'}$. Hence T is invertible and $T' = T^{-1}$. □

2.4.28. Remark. Recall from introductory linear algebra that we have at least two methods to decide when an $n \times n$-matrix A is invertible. The first is to do Gauss-Jordan to it and check that we eventually get the identity (that is, the system $A \cdot x = 0$ has only $x = 0$ as solution). We can also compute the *determinant* $\det(A)$. And we know that A is invertible if and only if $\det(A) \neq 0$. This is reviewed in Appendix D.

2.4.29. Exercise. Let $B = \{b_1, \dots, b_p\}$ be a basis of V and $B' = \{b'_1, \dots, b'_q\}$ a basis of V'. Give a basis $C = \{E_{ij}\}_{1 \leq i \leq q, 1 \leq j \leq p}$ of $\operatorname{Lin}(V, V')$ where $E_{ij}(b_k) = \delta_{jk} \cdot b'_i$. What are the coordinates of a linear transformation $T \colon V \to V'$ in the basis C?

2.5. Base change

The results of Section 2.4 depend heavily on the choice of the bases. In this section, we want to analyze what happens when we change said bases. We could already ask this question about coordinates, at the end of Section 1.6, but we did not have the tools to answer it then. Now we do:

2.5.1. Definition. Let $B = \{b_1, \dots, b_n\}$ and $C = \{c_1, \dots, c_n\}$ be two bases of the same n-dimensional vector space V. The *base-change matrix from B to C* (or, in full, *for change of coordinates with respect to B into coordinates with respect to C*) is the matrix of the identity tranformation $\mathrm{Id}_V \colon V \to V$ with respect to those bases:

$$Q = Q_{C,B} := \big[\, \mathrm{Id}_V \,\big]_{C,B}.$$

Expanding Definition 2.4.2, the matrix $Q_{C,B} \in \mathrm{M}_{n \times n}(\mathbb{F})$ is constructed one column at a time, the j-th column being given by $[b_j]_C$ the coordinates of the j-th basis vector of B in the basis C. In detail, it means that if $Q_{C,B} = [Q_{ij}]_{1 \leq i,j \leq n}$ then

$$b_j = \sum_{i=1}^{n} Q_{ij} \cdot c_i \tag{2.5.2}$$

for every $1 \leq j \leq n$.

2.5.3. Proposition. *Let $B = \{b_1, \dots, b_n\}$ and $C = \{c_1, \dots, c_n\}$ be two bases of the vector space V and let $Q = Q_{C,B}$ the base-change matrix from B to C. Then*

(1) *For every vector $v \in V$, we have*

$$[v]_C = Q \cdot [v]_B.$$

(2) *The matrix Q is invertible and $Q^{-1} = Q_{B,C}$.*

Proof. Part (1) is simply the equation $[T(v)]_{B'} = [T]_{B',B} \cdot [v]_B$ of Proposition 2.4.7 applied to $T = \mathrm{Id}_V$ and $V' = V$ and $B' = C$. Part (2) is simply Corollary 2.4.26 applied to $T = \mathrm{Id}_V$, which here tells us that Q is invertible and $(Q_{C,B})^{-1} = ([\mathrm{Id}_V]_{C,B})^{-1} = [\mathrm{Id}_V^{-1}]_{B,C} = [\mathrm{Id}_V]_{B,C} = Q_{B,C}$. □

The reader has done such base-changes in introductory linear algebra and all those examples remain valid here. Let us do another elementary example in slightly different clothes.

2.5.4. Example. We want to write any vector $aX^2 + bX + c$ in coordinates in the ordered basis $B = \{X + 1, X^2 + 1, X^2 + X\}$ of the real vector space $V = \big\{ P \in \mathbb{R}[X] \,\big|\, \deg(P) \leq 2 \big\}$. We also have a canonical basis $C = \{1, X, X^2\}$ of V in which it is very easy to write coordinates. So $Q = Q_{C,B} = \left(\begin{smallmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{smallmatrix}\right)$ is easy to produce. By Exercise 2.3.4 we compute Q^{-1} by doing Gauss-Jordan on $[Q|I_n]$:

$$\left(\begin{smallmatrix} 1&1&0&1&0&0 \\ 1&0&1&0&1&0 \\ 0&1&1&0&0&1 \end{smallmatrix}\right) \rightsquigarrow \left(\begin{smallmatrix} 1&1&0&1&0&0 \\ 0&-1&1&-1&1&0 \\ 0&1&1&0&0&1 \end{smallmatrix}\right) \rightsquigarrow \left(\begin{smallmatrix} 1&0&1&0&1&0 \\ 0&1&-1&1&-1&0 \\ 0&0&2&-1&1&1 \end{smallmatrix}\right) \rightsquigarrow \left(\begin{smallmatrix} 1&0&0&0.5&0.5&-0.5 \\ 0&1&0&0.5&-0.5&0.5 \\ 0&0&1&-0.5&0.5&0.5 \end{smallmatrix}\right).$$

Hence $Q_{B,C} = Q_{C,B}^{-1} = \frac{1}{2}\left(\begin{smallmatrix} 1&1&-1 \\ 1&-1&1 \\ -1&1&1 \end{smallmatrix}\right)$. So if $v = aX^2 + bX + c$ then $[v]_C = \left[\begin{smallmatrix} c \\ b \\ a \end{smallmatrix}\right]$ and therefore $[v]_B = Q_{B,C} \cdot [v]_C = \frac{1}{2}\left(\begin{smallmatrix} 1&1&-1 \\ 1&-1&1 \\ -1&1&1 \end{smallmatrix}\right) \cdot \left[\begin{smallmatrix} c \\ b \\ a \end{smallmatrix}\right] = \frac{1}{2}\left[\begin{smallmatrix} -a+b+c \\ a-b+c \\ a+b-c \end{smallmatrix}\right]$. We can verify that these are indeed the coordinates of v in the basis $B = \{X+1, X^2+1, X^2+X\}$:

$$\frac{-a+b+c}{2}\cdot(X+1) + \frac{a-b+c}{2}\cdot(X^2+1) + \frac{a+b-c}{2}\cdot(X^2+X) = aX^2+bX+c = v.$$

2.5.5. Theorem. *Let $T\colon V \to V'$ be a linear transformation between finite-dimensional $\mathbb{F}$-vector spaces. Let B and C be two bases of V. Let B' and C' be two bases of V'. Then we have the following relation between the two matrices of T, obtained by using respectively the pair B, B' and the pair C, C' of bases:*

$$[T]_{C',C} = Q_{C',B'} \cdot [T]_{B',B} \cdot Q_{B,C}$$

where $Q_{B,C}$ and $Q_{C',B'}$ are the base-change matrices of Definition 2.5.1.

Proof. We apply Theorem 2.4.16 to the composition $T = \mathrm{Id}_{V'} \circ T \circ \mathrm{Id}_V$. This looks rather trivial but we let the bases vary as we go. It reads:

$$[T]_{C',C} = [\mathrm{Id}_{V'}]_{C',B'} \cdot [T]_{B',B} \cdot [\mathrm{Id}_V]_{B,C}. \tag{2.5.6}$$

This is the announced formula since $Q_{C',B'} = [\mathrm{Id}_{V'}]_{C',B'}$ and $Q_{B,C} = [\mathrm{Id}_V]_{B,C}$. □

2.5.7. Remark. It is important not to mix up $Q_{B,C}$ and $Q_{C,B}$ in those formulas! Beware of calling all of them Q without paying attention. Also, it is important to remember that we read matrix multiplication 'from right to left': The right-most matrix is the first one to 'hit' the column vector. Reading (2.5.6) in this order, we see that to describe T from coordinates in the basis C to coordinates in the basis C', we first change coordinates from the basis C to the basis B, then use $[T]_{B',B}$ (at which point we 'are' in coordinates in the basis B') and finally change coordinates from B' to C'.

2.5.8. Corollary. *Let $T\colon V \to V$ be a linear transformation from V to itself. Let B and C be two bases of V and $Q = Q_{C,B}$ the base-change matrix from B to C. Then*

$$[T]_{C,C} = Q \cdot [T]_{B,B} \cdot Q^{-1}.$$

Proof. This is immediate from the previous proposition with $V' = V$, $B' = B$ and $C' = C$. Thus $Q_{C',B'} = Q_{C,B}$ is our Q and $Q_{B,C} = Q_{C,B}^{-1}$ is our Q^{-1}. □

2.5.9. Remark. It is convenient, although slightly abusive, to write $[T]_B$ instead of $[T]_{B,B}$ when we consider an *operator* $T\colon V \to V$ (*i.e.* a linear transformation from V to itself) and the same basis B at the two ends of it (*i.e.* $B' = B$). In that case, the above formula becomes

$$[T]_C = Q \cdot [T]_B \cdot Q^{-1} \quad \text{where} \quad Q = Q_{B,C}.$$

2.5.10. Exercise. Let $V = \left\{ P \in \mathbb{R}[X] \,\middle|\, \deg(P) \le 2 \right\}$ as in Example 2.5.4. Let $T\colon V \to V$ be the derivative with respect to X. Compute the matrix $[T]_C$ of T with respect to the canonical basis C and compute $[T]_B$ for $B = \{X+1, X^2+1, X^2+X\}$.

2.5.11. Definition. Recall that two square matrices $A, \tilde{A} \in \mathrm{M}_{n,n}(\mathbb{F})$ are called *similar* if there exists an invertible $Q \in \mathrm{M}_{n\times n}(\mathbb{F})$ such that $\tilde{A} = Q \cdot A \cdot Q^{-1}$.

2.5.12. Exercise. Write $A \sim \tilde{A}$ to say that A and $\tilde{A}$ are similar. Show that $\sim$ is an equivalence relation on $\mathrm{M}_{n\times n}(\mathbb{F})$.

2.6. Dual space*

We have seen in Proposition 2.1.17 that, given $\mathbb{F}$-vector spaces V and V', the set $\mathrm{Lin}_{\mathbb{F}}(V, V')$ of $\mathbb{F}$-linear transformations from V to V' is itself an $\mathbb{F}$-vector space. The special case of this construction where $V' = \mathbb{F}$ is particularly useful.

2.6.1. Definition. Let V be an $\mathbb{F}$-vector space. The *dual* of V is the vector space $V^{\#} = \mathrm{Lin}_{\mathbb{F}}(V, \mathbb{F})$. Its vectors are linear transformations $\ell\colon V \to \mathbb{F}$. We add them and we multiply them by scalars, as we do with functions: $(\ell + \ell')(v) = \ell(v) + \ell'(v)$ and $(a \cdot \ell)(v) = a \cdot \ell(v)$ for all $v \in V$. For instance, the zero element $0_{V^{\#}}$ is the linear map $V \to \mathbb{F}$ that maps all vectors to zero: $0_{V^{\#}}(v) = 0_{\mathbb{F}}$ for all $v \in V$.

2.6.2. Remark. Some authors will write V^* or $V^\vee$ for the dual.

The vector space $V^{\#}$ is defined abstractly. It does not mean it is difficult to use. It just means it is difficult to describe as a simple vector space even when V itself is rather simple, like $\mathbb{F}^n$. We can do it up to isomorphism though.

2.6.3. Proposition. *If V has finite dimension n then $V^{\#}$ has finite dimension n.*

Proof. This is Corollary 2.4.14 with $V' = \mathbb{F}$ has dimension one. □

2.6.4. Example. Let $V = \mathbb{F}^n$ for some $n \geq 1$. For every $\underline{a} = (a_1, \ldots, a_n) \in \mathbb{F}^n$ we can define $\ell_{\underline{a}}\colon \mathbb{F}^n \to \mathbb{F}$ by $\ell_{\underline{a}}(\underline{x}) = a_1x_1 + \cdots + a_nx_n$ for all $\underline{x} = (x_1, \ldots, x_n)$ in $\mathbb{F}^n$. It is easy to see that $\ell_{\underline{a}}\colon V \to \mathbb{F}$ is linear. Hence $\ell_{\underline{a}}$ is a vector in $(\mathbb{F}^n)^{\#}$. One can also verify that $T\colon V \to V^{\#}$ defined by $\underline{a} \mapsto T(\underline{a}) = \ell_{\underline{a}}$ is linear (in $\underline{a}$) and is actually an isomorphism. Indeed, note that $\ell_{\underline{a}}(e_i) = a_i$ where $e_1, \ldots, e_n$ is the canonical basis. Hence $T(\underline{a}) = 0$ forces $a_1, \ldots, a_n = 0$, hence $\underline{a} = 0$. So $\mathrm{Ker}(T) = \{0\}$ and $T\colon V \to V^{\#}$ is injective. It follows that T is an isomorphism by Corollary 2.3.14 since we already know that $\dim((\mathbb{F}^n)^{\#}) = n = \dim(\mathbb{F}^n)$. We can also make surjectivity explicit: Given $\ell\colon V \to \mathbb{F}$ linear, it is characterized by the images $a_1 := \ell(e_1), \ldots, a_n := \ell(e_n)$ in $\mathbb{F}$. These scalars are simply the entries of the matrix $[\ell]_{C,B}$ of ℓ with resepct to the canonical bases B and $C = \{1\}$ (the canonical basis of $V' = \mathbb{F}$). We know that $\ell = T_{\underline{a}}$ for instance via Corollary 2.4.9. Hence we have an isomorphism

$$\begin{array}{cccc} T\colon & \mathbb{F}^n & \xrightarrow{\sim} & (\mathbb{F}^n)^{\#} \\ & \underline{a} & \longmapsto & \ell_{\underline{a}}. \end{array}$$

The above was the case of a free vector space on a finite set with n elements. The proof actually gives:

2.6.5. Exercise. Let I be an arbitrary set and $V = \mathbb{F}^{(I)}$. Show that the dual $V^{\#} = (\mathbb{F}^{(I)})^{\#}$ is (canonically) isomorphic to the vector space $\mathbb{F}^I$ of functions $I \to \mathbb{F}$.

2.6.6. Definition. Let B be a basis of a vector space V. For every $b \in B$ denote by $b^{\#}\colon V \to \mathbb{F}$ the unique linear transformation (Theorem 2.1.23 again) such that $b^{\#}(b') = \delta_{b,b'}$ for every $b' \in B$. This defines an element $b^{\#} \in V^{\#}$ that we call the *dual vector* of the basis vector b (with respect to that basis B).

2.6.7. Remark. In other words, $b^{\#}(\sum_{b\in B} a_b \cdot b) = a_b$. The transformation $b^{\#}\colon V \to \mathbb{F}$ sends a vector v to its b-th coordinate. This notation $b^{\#}$ is slightly misleading as $b^{\#}$ does not only depend on b but really on the whole basis, as the coordinates of v do depend on the whole of B.

2.6.8. Proposition. *Let V be a finite-dimensional $\mathbb{F}$-vector space with basis $B = \{b_1, \ldots, b_n\}$. Then $\{b_1^{\#}, \ldots, b_n^{\#}\}$ is a basis of $V^{\#}$. (It is called the* dual basis *of B.) We have an isomorphism $V \overset{\sim}{\to} V^{\#}$ mapping each b_j in B to $b_j^{\#}$.*(4)

Proof. This is similar to Example 2.6.4. By our friend Theorem 2.1.23 we can define $T\colon V \to V^{\#}$ by deciding what $T(b_j)$ is. Here we pick $T(b_j) = b_j^{\#}$ for all $j = 1, \ldots, n$. We claim that T is an isomorphism. Since $\dim(V) = \dim(V^{\#})$ by Proposition 2.6.3, it suffices to show that T is injective (Corollary 2.3.14). Let $v \in V$ such that $T(v) = 0$. Writing $v = a_1 \cdot b_1 + \cdots + a_n \cdot b_n$, our claim is that $a_1 \cdot b_1^{\#} + \cdots + a_n \cdot b_n^{\#} = 0$ in $V^{\#}$ and we want to show that all a_i are zero (*i.e.* that $v = 0$). Just evaluate this linear transformation $0_{V^{\#}}\colon V \to \mathbb{F}$ on b_i. It gives in $\mathbb{F}$

$$\begin{aligned} 0 &= 0_{V^{\#}}(b_i) && \text{by definition of 0 in } V^{\#} \\ &= (\textstyle\sum_{j=1}^{n} a_j \cdot b_j^{\#})(b_i) && \text{since we assume } \textstyle\sum_{j=1}^{n} c_j \cdot b_j^{\#} = 0 \\ &= \textstyle\sum_{j=1}^{n} a_j \cdot b_j^{\#}(b_i) && \text{unpacking the linear combination} \\ &= \textstyle\sum_{j=1}^{n} a_j \cdot \delta_{ji} && \text{by definition of } b_j^{\#} \\ &= 0 + \cdots + 0 + a_i \cdot 1 + 0 + \cdots + 0 = a_i \end{aligned}$$

and this holds for all $i = 1, \ldots, n$. Hence the result. □

2.6.9. Remark. The construction $V \mapsto V^{\#}$ is what we call *contravariant.*(5) It means the following. Given a linear transformation $T\colon V_1 \to V_2$ (yes, V_2, we are not going to deal with $V'^{\#}$), we can define a new linear transformation $T^{\#}$ between the duals of the given spaces. However, it goes *backwards*:

$$\begin{array}{cccc} T^{\#}\colon & V_2^{\#} & \longrightarrow & V_1^{\#} \\ & \ell & \longmapsto & \ell \circ T. \end{array}$$

Indeed, spelling this out slowly, a vector $\ell \in V_2^{\#}$ is by definition of $V^{\#} = \operatorname{Lin}(V, \mathbb{F})$ just a linear transformation $\ell\colon V_2 \to \mathbb{F}$. The composition $\ell \circ T$ is then a linear transformation $V_1 \to V_2 \to \mathbb{F}$, that is, an element of $V_1^{\#}$. We call it $T^{\#}(\ell) = \ell \circ T$. This $T^{\#}$ is linear (in ℓ) since $- \circ T$ is linear (composition $- \circ -$ is linear in each variable, separately, *i.e.* when the other is fixed). We have the formulas – beware the reversal of order in the second

$$(\mathrm{Id}_V)^{\#} = \mathrm{Id}_{V^{\#}} \qquad \text{and} \qquad (S \circ T)^{\#} = T^{\#} \circ S^{\#}.$$

The first is clear. For the second, we have $(S \circ T)^{\#}(\ell) = \ell \circ (S \circ T) = (\ell \circ S) \circ T = (S^{\#}(\ell)) \circ T = T^{\#}(S^{\#}(\ell)) = (T^{\#} \circ S^{\#})(\ell)$.

[4] This isomorphism is *non-canonical* meaning that it depends on the choice of the basis B in a significant way: Other bases will give different isomorphisms.

[5] To be precise it is a contravariant *functor*.

2.6.10. Remark. Because of the ‘reversal of arrows’ discussed in the previous remark, the isomorphism $V \overset{\sim}{\to} V^{\#}$ of Proposition 2.6.8 is not *natural*. This means that the following diagram does not commute in general (for some T):

$$\begin{tikzcd}
V \arrow[r, "\sim"] \arrow[d, "T"'] \arrow[dr, phantom, "\times"] & V^{\#} \\
W \arrow[r, "\sim"] & W^{\#} \arrow[u, "T^{\#}"']
\end{tikzcd}$$

where the horizontal isomorphisms are those of Proposition 2.5.3 for some choice of bases of V and W, assumed finite-dimensional.

However, we can say something about the matrix of the dual transformation $T^{\#}$ as long as we use the dual bases of Definition 2.6.6 and Proposition 2.6.8.

2.6.11. Proposition. *Let V and W be finite-dimensional $\mathbb{F}$-vector spaces. Let B be a basis of V and C a basis of W. Recall the dual bases, $B^{\#}$ of $V^{\#}$, and $C^{\#}$ of $W^{\#}$. Let $T\colon V \to W$ be linear and $T^{\#}\colon W^{\#} \to V^{\#}$ be the dual transformation ($T^{\#}(\ell) = \ell \circ T$) of Remark 2.6.9. Then the matrix of $T^{\#}$ with respect to $C^{\#}$ and $B^{\#}$ (watch the order) is the* transpose *of the matrix of T with respect to B and C:*

$$\left[T^{\#}\right]_{B^{\#},C^{\#}} = \Big(\left[T\right]_{C,B}\Big)^{t}.$$

Proof. Write $B = \{b_1, \dots, b_m\}$ and $C = \{c_1, \dots, c_n\}$ where $m = \dim(V)$ and $n = \dim(W)$. To construct $[T]_{C,B} = A = [a_{ij}]_{i,j} \in \mathrm{M}_{m\times n}(\mathbb{F})$ as in (2.4.3), we write each $T(b_j)$ in coordinates, that is, we have for every $j = 1, \dots, m$

$$T(b_j) = \sum_{i=1}^{n} a_{ij} \cdot c_i. \tag{2.6.12}$$

The transpose $A^t \in \mathrm{M}_{m\times n}$ is given by $(A^t)_{p,q} = a_{q,p}$. So the claim is that for every $q = 1, \dots, n$ we should have

$$T^{\#}(c_q^{\#}) = \sum_{p=1}^{m} (A^t)_{pq} \cdot b_p^{\#} = \sum_{p=1}^{m} a_{qp} \cdot b_p^{\#}. \tag{2.6.13}$$

This is an equality in $V^{\#}$, *i.e.* between two linear transformations $V \to \mathbb{F}$. So it suffices to test it on each basis vector b_j of V, for $j = 1, \dots, m$. We have (in $\mathbb{F}$)

$$\begin{aligned}
(T^{\#}(c_q^{\#}))(b_j) &= (c_q^{\#} \circ T)(b_j) && \text{by definition of } T^{\#} \\
&= c_q^{\#}(T(b_j)) && \text{by definition of } \circ \\
&= c_q^{\#}(\textstyle\sum_{i=1}^{n} a_{ij}\, c_i) && \text{by (2.6.12)} \\
&= \textstyle\sum_{i=1}^{n} a_{ij}\, \delta_{qi} && \text{by definition of } c_q^{\#} \\
&= a_{qj} && \text{by definition of } \delta_{qi} \\
&= \textstyle\sum_{p=1}^{n} a_{qp}\, \delta_{pj} && \text{by definition of } \delta_{pj} \\
&= \textstyle\sum_{p=1}^{n} a_{qp}\, b_p^{\#}(b_j) && \text{by definition of } b_p^{\#}\colon V \to \mathbb{F}. \\
&= (\textstyle\sum_{p=1}^{n} a_{qp} b_p^{\#})(b_j) && \text{rewriting the linear combination.}
\end{aligned}$$

This establishes the claim (2.6.13) and the result. □

Is there such a thing as too much of a good thing?

2.6.14. Proposition. *Let V be an $\mathbb{F}$-vector space. Its* double-dual *is the dual of its dual, $V^{\#\#} := (V^{\#})^{\#}$. There is a canonical $\mathbb{F}$-linear transformation*

$$\alpha = \alpha_V \colon V \longrightarrow V^{\#\#}$$

that maps a vector $v \in V$ to $\alpha(v) \colon V^{\#} \to \mathbb{F}$ defined by $(\alpha(v))(\ell) = \ell(v)$. (In other words, $\alpha(v)$ is 'evaluation at v'). Moreover:

(1) *$\alpha_V \colon V \to V^{\#\#}$ is injective.* ([6])

(2) *α_V is an isomorphism if V is finite dimensional.*

(3) *α is* natural *in V, meaning that for every linear transformation $T \colon V \to W$, with dual $T^{\#} \colon W^{\#} \to V^{\#}$ and double-dual $T^{\#\#} = (T^{\#})^{\#} \colon V^{\#\#} \to W^{\#\#}$, the following diagram of vector spaces commute:*

$$\begin{tikzcd}
V \arrow[r, "\alpha_V"] \arrow[d, "T"'] & V^{\#\#} \arrow[d, "T^{\#\#}"] \\
V \arrow[r, "\alpha_W"'] & W^{\#\#}\,.
\end{tikzcd}$$

Proof. It is easy to verify that α is well-defined (*i.e.* $\alpha(v) \colon V^{\#} \to \mathbb{F}$ is indeed linear for every v) and that α is linear (in v). We leave these verifications to the reader.

For (1), let B be a basis of V. Let $v \in V$ such that $\alpha(v) = 0$. This means that $\ell(v) = 0$ for every $\ell \colon V \to \mathbb{F}$. In particular, this holds for $\ell = b^{\#}$ with $b \in B$ (Definition 2.6.6). If we write $v = \sum_{b \in B} a_b \cdot b$ with $a_b \in \mathbb{F}$ almost all zero, then $a_b = b^{\#}(v) = 0$ for all b. Hence $v = 0$. In short, $\mathrm{Ker}(\alpha) = \{0\}$, so α is injective.

For (2), we know by Proposition 2.6.3 that when V is finite-dimensional then so is $V^{\#}$ and $\dim(V^{\#}) = \dim(V)$ and thus, applying this to $V^{\#}$, we know that $V^{\#\#}$ is finite-dimensional and $\dim(V^{\#\#}) = \dim(V^{\#}) = \dim(V)$. Hence the injective $\alpha \colon V \to V^{\#\#}$ is automatically an isomorphism by Corollary 2.3.14.

Part (3) is an exercise on the definitions. Let $v \in V$ and consider its two images under the two paths in the square: either $T^{\#\#}(\alpha_V(v))$ or $\alpha_W(T(v))$. These are two vectors in $W^{\#\#}$, in other words, these are linear transformations $W^{\#} \to \mathbb{F}$. So we compare them by evaluating them on all $\ell \in W^{\#}$:

$$\begin{aligned}
\big(T^{\#\#}(\alpha_V(v))\big)(\ell) &= (\alpha_V(v) \circ T^{\#})(\ell) && \text{by definition of } (T^{\#})^{\#} \\
&= (\alpha_V(v))(T^{\#}(\ell)) \\
&= (T^{\#}(\ell))(v) && \text{by definition of } \alpha_V(v) \\
&= (\ell \circ T)(v) && \text{by definition of } T^{\#} \\
&= \ell(T(v)) \\
&= \big(\alpha_W(T(v))\big)(\ell) && \text{by definition of } \alpha_W(T(v)).
\end{aligned}$$

As ℓ was arbitrary, we have the claim. □

[6] At least when V has a basis, which is always true with the Axiom of Choice.

2.7. Quotients*

In this section V is an $\mathbb{F}$-vector space.

2.7.1. Remark. Given a subspace $W \subseteq V$ we may wonder if there is a way of producing a new vector space V', that receives V by a linear transformation $V \to V'$ in such a way that all vectors of W are mapped to *zero* in V'. The answer is trivial: Just map everything to zero. So the better question is: Can we do this in a 'optimal' way: We want to send W to zero but nothing else. Well, mapping W to zero has consequences on other vectors too. Indeed, if two vectors x and y in V are such that $x-y$ belongs to W then necesarily x and y will be sent to the same image by T since $T(x) - T(y) = T(x-y)$ by linearity. So 'killing W' forces many apparently different vectors to 'become the same'. Let us formalize this.

2.7.2. Lemma. *Let $W \subseteq V$ be a subspace of V. Define a relation $\sim$ on the vectors of V by setting $x \sim y$ if $x - y \in W$. Then $\sim$ is an equivalence relation:*

(a) *for all $x \in V$ we have $x \sim x$;*
(b) *for all $x, y \in V$, if $x \sim y$ then $y \sim x$;*
(c) *for all $x, y, z \in V$ if $x \sim y$ and $y \sim z$ then $x \sim z$.*

Proof. (a) holds because $0 \in W$. (b) holds because $(y-x) = -(x-y)$ and $-W \subseteq W$. And (c) holds because $(x-z) = (x-y)+(y-z)$ and $W + W \subseteq W$. □

2.7.3. Example. Of course the equivalence relation depends on W. For instance:

(1) If $W = 0$ then $x \sim y$ if and only if $x = y$.
(2) If $W = V$ then $x \sim y$ for all $x, y \in V$.

2.7.4. Definition. Let $W \subseteq V$ be a subspace. Define V/W as the set of equivalence classes with respect to $\sim$:

$$V/W := V/\sim = \{\, [x]_W \,|\, x \in V \,\}.$$

Formally this is a subset of the set of subsets of V, where $[x]_W = \{\, y \in V \,|\, y \sim x \,\}$ is the equivalence class of x, sometimes denoted $[x]_\sim$. Different $x \in V$ can give the same class $[x]_W$. By definition $[x]_W = [y]_W$ if and only if $x \sim y$. Explicitly, the class of x is given by $[x]_W = x + W := \{\, x + w \,|\, w \in W \,\}$ and some authors will write everywhere $x + W$ instead of $[x]_W$.

This set V/W is called the *quotient* of V by W (or later the *quotient space* once we know that V/W is itself a vector space). When $x \sim y$, that is, when $x - y \in W$, we also say that x and y are *congruent modulo W*.

Note that we have an obvious function $V \to V/W$, that we call the *projection*, which is defined by

$$\begin{aligned} \mathrm{proj} = \mathrm{proj}_W \colon \quad V &\longrightarrow V/W \\ x &\longmapsto [x]_W. \end{aligned}$$

It simply sends every vector to its class modulo W. It is surjective by definition.

2.7.5. Example. Let us review our trivial examples. For $W = 0$, the quotient $V/0 = V$ is the same as V since in that case $\sim$ is just equality. And of course, the projection is the identity.

For $W = V$, all vectors are equivalent, for instance, all equivalent to zero. So $V/V = \{[0]_V\}$ has a unique element. (And there is no much choice for proj!)

In both cases, V/W is a vector space. This is true in general.

2.7.6. Proposition. *Let $W \subseteq V$ be a subspace and consider the quotient V/W. Then V/W admits a unique structure of $\mathbb{F}$-vector space, that is, a unique addition and scalar action, such that the projection $\mathrm{proj}_W \colon V \to V/W$ is linear. Explicitly, $[x]_W + [y]_W = [x+y]_W$ and $a \cdot [x]_W = [a \cdot x]_W$ for all $x, y \in V$ and $a \in \mathbb{F}$.*

Proof. If we want $\mathrm{proj}\colon V \to V/W$, which is defined by $x \mapsto [x]_W$, to be linear then indeed we *must* have the formulas $[x]_W +_{V/W} [y]_W = [x+y]_W$ and $a \cdot_{V/W} [x]_W = [a \cdot x]_W$ of the statement. So there is at most one such $\mathbb{F}$-vector space structure on V/W. We need to see that these $+_{V/W}$ and $\cdot_{V/W}$ are well-defined and they do satisfy Axioms (VS1)-(VS8) of Definition 1.2.1. The axioms are actually very easy and will be left to the reader. The critical point is that those operations are *well-defined*. Let us explain this slowly. We write $[-]$ instead of $[-]_W$ for readability. If we take two elements $X, Y \in V/W$ and we want to define $X+Y$ as we did above, we write $X = [x]$ as the equivalence class of *some* $x \in V$ and $Y = [y]$ as the equivalence class of *some* $y \in V$ and we decide that $X + Y$ will be $[x+y]$ the equivalence class of $x+y$, added in V. The problem is that this might depend on the choice of x and y. So we need to prove that if $X = [x']$ and $Y = [y']$ for other representatives x' and y' in their class then $x'+y'$ will give the same answer. In general, $x+y$ and $x'+y'$ will be different. This is not important, what we need is that $[x+y] = [x'+y']$. In short we need to prove that for all $x, x'y, y' \in V$ the following holds true:

$$\text{If} \quad x \sim x' \text{ and } y \sim y' \quad \text{then} \quad x+y \sim x'+y'.$$

The assumptions are that $x - x' \in W$ and $y - y' \in W$. Then $(x+y)-(x'+y') = (x-x')+(y-y')$ also belongs to W since $W + W \subseteq W$ by (SS2). So the addition $[x]+[y] = [x+y]$ is well defined on V/W. Similarly, the verification that $a\cdot[x] = [a\cdot x]$ is well-defined for $a \in \mathbb{F}$ boils down to check that if $x \sim x'$ then $a \cdot x \sim a \cdot x'$. The hypothesis is $x - x' \in W$ and then $a \cdot x - a \cdot x' = a \cdot (x - x')$ is still in W since $\mathbb{F} \cdot W \subseteq W$ by (SS3). □

2.7.7. Remark. In particular, the zero vector in V/W is simply $[0]_W$ the class of zero (which explicitly is $[0]_W = 0 + W = W$ as a subset of V).

Let us identify some examples.

2.7.8. Example. Let V_1 and V_2 be two $\mathbb{F}$-vector spaces and consider the (external) direct sum $V_1 \oplus V_2$ of Exercise 1.2.25. Let $W = V_1 \oplus 0$ (which is really just V_1 seen inside $V_1 \oplus V_2 = V_1 \times V_2$ as $V_1 \times \{0\}$). So we simply write $W = V_1 \subseteq V_1 \oplus V_2$. Then $(V_1 \oplus V_2)/V_1$ is isomorphic to V_2. Indeed, saying that $(x_1, x_2) \sim (x_1', x_2')$ means that $(x_1 - x_1', x_2 - x_2') \in V_1 \times \{0\}$, that is, $x_2 = x_2'$. So we get a well-defined linear transformation $T\colon (V_1 \oplus V_2)/V_1 \to V_2$, $[(x_1, x_2)] \mapsto x_2$. It is an isomorphism with inverse $T'\colon V_2 \to (V_1 \oplus V_2)/V_1$ defined by $T'(y) = [(0, y)]$.

This examples illustrates how V/W is in spirit the *difference* of V and W. Here is another fact supporting this intuition:

2.7.9. Proposition. *Let $W \subseteq V$ be a subspace. The kernel of the surjective linear transformation $\mathrm{proj}\colon V \to V/W$ is exactly W. Consequently, if V is finite dimensional then so is V/W and $\dim(V/W) = \dim(V) - \dim(W)$.*

Proof. We already know that $\mathrm{proj}\colon V \to V/W$ is linear and surjective. For $x \in V$ we have $x \in W$ if and only if $x \sim 0$ if and only if $[x]_W = [0]_W$ if and only if $x \in W$. Hence $\mathrm{Ker}(\mathrm{proj}) = W$ indeed. If $\dim(V) < \infty$ then Rank-Nullity Theorem 2.2.6

applied to $T = \mathrm{proj}$ tells us that $\mathrm{Im}(\mathrm{proj})$ is finite-dimensional and $\dim(\mathrm{Ker}(\mathrm{proj}))+\dim(\mathrm{Im}(\mathrm{proj})) = \dim(V)$. Since $\mathrm{Ker}(\mathrm{proj}) = W$ by the above and since $\mathrm{Im}(\mathrm{proj}) = V/W$ by surjectivity, we have the second claim of the statement. □

We now want to apply our quotient construction to a particular subspace: the kernel of a linear transformation T.

2.7.10. Proposition. *Let $T\colon V \to V'$ be a linear transformation. Then T induces a well-defined isomorphism*

$$\begin{array}{cccc} \bar{T}\colon & V/\mathrm{Ker}(T) & \xrightarrow{\sim} & \mathrm{Im}(T) \\ & [x]_{\mathrm{Ker}(T)} & \longmapsto & T(x). \end{array}$$

Proof. Let $W = \mathrm{Ker}(T)$. Let us see that $\bar{T}$ is well-defined. Suppose that $[x]_W = [x']_W$, that is, $x \sim x'$ are two representatives of the same equivalence class modulo $W = \mathrm{Ker}(T)$. We need to show that $T(x) = T(x')$. But indeed, $x \sim x'$ means $x - x' \in \mathrm{Ker}(T)$ hence $0 = T(x - x') = T(x) - T(x')$ by linearity of T. Since clearly $\bar{T}([x]) = T(x)$ belongs to $\mathrm{Im}(T)$ in V', the function $\bar{T}\colon V/\mathrm{Ker}(T) \to \mathrm{Im}(T)$ is well-defined. It is linear. Let us do this verification to familiarize ourselves with the structure on V/W. Let $n \in \mathbb{N}$ and $[x_1], \ldots, [x_n] \in V/W$ be n vectors in the source of $\bar{T}$ and $a_1, \ldots, a_n$ be n scalars. Compute in $\mathrm{Im}(T)$, that is, in V':

$$\begin{aligned} \bar{T}\big(\textstyle\sum_{i=1}^n a_i \cdot [x_i]\big) &= \bar{T}\big([\textstyle\sum_{i=1}^n a_i \cdot x_i]\big) && \text{by structure of } V/W \\ &= T(\textstyle\sum_{i=1}^n a_i \cdot x_i) && \text{by definition of } \bar{T} \\ &= \textstyle\sum_{i=1}^n a_i \cdot T(x_i) && \text{by linearity of } T \\ &= \textstyle\sum_{i=1}^n a_i \cdot \bar{T}([x_i]) && \text{by definition of } \bar{T}. \end{aligned}$$

It can be useful to some readers to see that $\bar{T}$ is the (unique) factorization of T via $V/\mathrm{Ker}(T)$, in that the following diagram of linear transformations commutes:

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "\mathrm{proj}"'] & \mathrm{Im}(T) \subseteq V'. \\ V/\mathrm{Ker}(T) \arrow[ur, "\bar{T}"'] & \end{tikzcd}$$

Clearly, $\bar{T}$ is surjective onto $\mathrm{Im}(T)$. (We do not claim that $\bar{T}$ is surjective on V'.) If $y \in \mathrm{Im}(T)$ then $y = T(x)$ for some $x \in V$ and then $y = \bar{T}([x]) \in \mathrm{Im}(\bar{T})$.

Finally, $\bar{T}$ is injective by construction. If $[x] \in \mathrm{Ker}(\bar{T})$ then $T(x) = \bar{T}([x]) = 0$ by assumption, hence $x \in \mathrm{Ker}(T) = W$ which says that $[x] = 0$. □

2.7.11. Corollary. *Let $T\colon V \to V'$ be a* surjective *linear transformation. Then T induces an isomorphism $\bar{T}\colon V/\mathrm{Ker}(T) \overset{\sim}{\to} V'$.* □

In summary, V/W is characterized by the fact that it comes with a surjective linear transformation $V \to V/W$ whose kernel is W. Indeed, this holds with $\mathrm{proj}\colon V \to V/W$ by Proposition 2.7.9. And conversely, if $T\colon V \to V'$ is another surjective linear transformation with $\mathrm{Ker}(T) = W$ then the above corollary says that $V' \simeq V/W$. (In fact, since $\bar{T} \circ \mathrm{proj} = T$, even the T is uniquely characterized.)

2.7.12. Remark. For every subspace $W \subseteq V$ we have a (non-canonical) isomorphism $V \simeq W \oplus (V/W)$. More precisely, suppose given a section $S\colon V/W \to V$ of

$\mathrm{proj}\colon V \twoheadrightarrow V/W$, that is, a linear transformation $S\colon V/W \to V$ such that $\mathrm{proj} \circ S = \mathrm{Id}_{V/W}$. (Such an S always exists if we admit that V/W has a basis B, *e.g.* under the axiom of choice or if V, hence V/W, is finite dimensional. It suffices to choose a representative $S(b)$ in V of each basis element $b \in B$, that is, write $b = [S(b)]$ for $S(b) \in V$, for all $b \in B$, and use Theorem 2.1.23 to extend S to the whole of V/W as usual.) Then we claim that $W' = \mathrm{Im}(S)$ satisfies $W \oplus W' = V$, or in other words (Remark 2.3.17) $W + W' = V$ and $W \cap W' = 0$. Indeed, for every $x \in V$ we have $x = (x - S(\mathrm{proj}(x))) + S(\mathrm{proj}(x))$ and $x - S(\mathrm{proj}(x)) \in W$ since $\mathrm{proj}(x - S(\mathrm{proj}(x))) = \mathrm{proj}(x) - \mathrm{proj}(S(\mathrm{proj}(x))) = \mathrm{proj}(x) - \mathrm{proj}(x) = 0$ in V/W. Also, if $x \in W \cap W'$ then $x = S(y)$ for some $y \in V/W$ which is $y = \mathrm{proj}(S(y)) = \mathrm{proj}(x) = 0$ since $x \in W$ and so $x = S(y) = S(0) = 0$.

2.8. Tensor product*

A useful construction with vector spaces is the *tensor product* $V \otimes W$. It formalizes some 'rules' we want to have. Let us make a 'wish list' and then see how this is satisfied.

2.8.1. Remark. Let V and W be two $\mathbb{F}$-vector spaces. We want to construct a new vector space $V \otimes W$ that is 'natural' enough and has dimension the *product* of the dimensions of V and W. In other words, we would like that if $B = \{b_i\}_{i \in I}$ is a basis of V and $C = \{c_j\}_{j \in J}$ is a basis of W then $V \otimes W$ would have a basis consisting of $B \times C$: all pairs of (b_i, c_j) for $i \in I$ and $j \in J$. We shall denote this basis element $b_i \otimes c_j$ in $V \otimes W$. More generally, we wish to have elements $v \otimes w$ in $V \otimes W$ for all $v \in V$ and $w \in W$. If we think of $\otimes$ as a product then these (as of yet undefined) elements $v \otimes w$ should satisfy some rules, like for instance:

$$(v_1 + v_2) \otimes w = v_1 \otimes w + v_2 \otimes w \quad \text{and} \quad v \otimes (w_1 + w_2) = v \otimes w_1 + v \otimes w_2.$$

Also, we would like for every scalar $a \in \mathbb{F}$ that

$$(a \cdot v) \otimes w = v \otimes (a \cdot w)$$

for the latter is probably going to be $a \cdot (v \otimes w)$. Can we make that happen?

2.8.2. Definition. Let V and W be $\mathbb{F}$-vector spaces. Consider the free $\mathbb{F}$-vector space $\mathbb{F}^{(V \times W)}$ on the *set* $V \times W$, ignoring the vector space structures of V and W for now. Recall that $\mathbb{F}^{(V \times W)}$ consists of functions $f\colon V \times W \to \mathbb{F}$ that are zero almost everywhere: $\big\{ (v, w) \in V \times W \,\big|\, f(v, w) \neq 0 \big\}$ is finite.

For each pair $(v, w) \in V \times W$ consider the basis vector $e_{v,w} \in \mathbb{F}^{(V \times W)}$

$$\begin{array}{rccc} e_{v,w}\colon & V \times W & \longrightarrow & \mathbb{F} \\ & (v', w') & \longmapsto & \delta_{v,v'} \cdot \delta_{w,w'} = \begin{cases} 1 & \text{if } (v', w') = (v, w) \\ 0 & \text{otherwise.} \end{cases} \end{array}$$

Consider now the subspace $R \subseteq \mathbb{F}^{(V \times W)}$ spanned by all the following elements:

(a) $e_{v+v',w} - e_{v,w} - e_{v',w}$ for all $v, v' \in V$ and $w \in W$;
(b) $e_{v,w+w'} - e_{v,w} - e_{v,w'}$ for all $v \in V$ and $w, w' \in W$;
(c) $e_{a \cdot v,w} - a \cdot e_{v,w}$ for all $v \in V$, $w \in W$ and $a \in \mathbb{F}$;
(d) $e_{v,a \cdot w} - a \cdot e_{v,w}$ for all $v \in V$, $w \in W$ and $a \in \mathbb{F}$.

Then define the *tensor product* $V \otimes W$ as the quotient

$$V \otimes W = (\mathbb{F}^{(V \otimes W)})/R.$$

(The letter R stands for 'relations', not for real numbers.) One sometimes writes $V \otimes_{\mathbb{F}} W$ to keep track of which field we work with. By construction, $V \otimes W$ is an $\mathbb{F}$-vector space.

For every $v \in V$ and $w \in W$ we denote by $v \otimes w$ the class $[e_{v,w}]_R$ of the vector $e_{v,w} \in \mathbb{F}^{(V \times W)}$ modulo the subspace R.

2.8.3. Remark. At this stage, $V \otimes W$ could be anything: It could be a monster – after all we took the free vector space on a typically huge set $V \times W$ – or it could be zero – how do we know that R is not the whole of $\mathbb{F}^{(V \times W)}$ for instance? Let us address both concerns, to get comfortable with this mysterious space $V \otimes W$.

2.8.4. Proposition. *In $V \otimes W$ we have the following formulas:*

(a) $(v + v') \otimes w = v \otimes w + v' \otimes w$ *for all* $v, v' \in V$ *and* $w \in W$*;*
(b) $v \otimes (w + w') = v \otimes w + v \otimes w'$ *for all* $v \in V$ *and* $w, w' \in W$*;*
(c) $a \cdot (v \otimes w) = (a \cdot v) \otimes w = v \otimes (a \cdot w)$ *for all* $v \in V$*,* $w \in W$ *and* $a \in \mathbb{F}$*.*

Proof. In each equation, the difference between the representative of the left-hand side and the representative of the right-hand side belongs to R. Let us check (a) for instance. The meaning of $(v + v') \otimes w$ is the class $[e_{v+v',w}]_R$ modulo R. Similarly, the meaning of $v \otimes w + v' \otimes w$ is $[e_{v,w}]_R + [e_{v',w}]_R$ and the latter is $[e_{v,w} + e_{v',w}]_R$ by definition of the sum in the quotient. Now the difference between those representatives of the two classes $e_{v+v',w} - (e_{v,w} + e_{v',w})$ is precisely one of the generators of R by Definition 2.8.2 (a). In particular, this difference belongs to R and therefore the two classes are equal: $[e_{v+v',w}]_R = [e_{v,w}]_R + [e_{v',w}]_R$ by definition of the quotient. □

2.8.5. Proposition. *Any vector in $V \otimes W$ is a finite sum $(v_1 \otimes w_1) + \cdots + (v_r \otimes w_r)$ for $r \in \mathbb{N}$ and $v_1, \ldots, v_r \in V$ and $w_1, \ldots, w_r \in W$.*

Proof. The vectors $e_{v,w}$ form a basis of $\mathbb{F}^{(V \times W)}$ by Exercise 1.6.4. So every vector of $\mathbb{F}^{(V \times W)}$ is a linear combination of the form $a_1 \cdot e_{v_1,w_1} + \cdots + a_r \cdot e_{v_r,w_r}$. This remains true in the quotient (after all $\mathbb{F}^{(V \times W)} \to \mathbb{F}^{(V \times W)}/R$ is linear). However, in the quotient one more thing happens: $[a \cdot e_{v,w}]_R = a \cdot (v \otimes w) = (a \cdot v) \otimes w$ by Proposition 2.8.4 (c). (We could also absorb a in the second factor.) In any case, it follows that every vector in $V \otimes W$ is of the form $(a_1 \cdot v_1) \otimes w_1 + \cdots + (a_r \cdot v_r) \otimes w_r$ as announced. □

2.8.6. Remark. It is a common mistake to think of vectors in $V \otimes W$ as just so-called *simple tensors* $v \otimes w$. Proposition 2.8.5 tells us that any vector is a *finite sum* of simple tensors.

Here is the so-called *universal property* of $V \otimes W$. The function $V \times W \to V \otimes W$, $(v, w) \mapsto v \otimes w)$ is bilinear (*i.e.* satisfies properties (a)-(d) of Proposition 2.8.4, or is linear in each variable separately). In fact, $V \otimes W$ is the 'best' recipient of a bilinear function out of $V \times W$. Here is what this means.

2.8.7. Proposition. *Let U be an $\mathbb{F}$-vector space and let $t \colon V \times W \to U$ be a function that is bilinear,* i.e.

(i) *For each fixed $w \in W$, the function $V \to U$ given by $v \mapsto t(v, w)$ is linear.*

(ii) *For each fixed $v \in V$, the function $W \to U$ given by $w \mapsto t(v, w)$ is linear.*

Then there exists a unique *linear $T\colon V \otimes W \to U$ such that $T(v \otimes w) = t(v, w)$ for all $v \in V$ and $w \in W$.*

Proof. The transformation T is unique by Proposition 2.8.5: Once we have $T(x + y) = T(x) + T(y)$ and $T(v \otimes w) = t(v, w)$ then we know T on any sum of simple tensors.

Let us construct such a T. Since $\{\, e_{v,w} \,|\, (v, w) \in V \times W \,\}$ is a basis of $\mathbb{F}^{(V\times W)}$ we can define (Theorem 2.1.23) a linear transformation $T_0\colon \mathbb{F}^{(V\times W)} \to U$ by deciding that $T_0(e_{v,w}) = t(v, w)$. To show that T_0 descends to the quotient by R

$$\begin{tikzcd}
\mathbb{F}^{(V\times W)} \arrow[r, "T_0"] \arrow[d, two heads] & U \\
\mathbb{F}^{(V\times W)}/R = V \otimes W \arrow[ur, dashed, "T"'] &
\end{tikzcd}$$

we need to show that $T_0(R) = 0$. Since T_0 is linear it suffices to verify that T_0 sends all generators of R (see Definition 2.8.2) to zero. This is immediate by bilinearity of t. For instance, for a generator of type (c) we have $T_0(e_{a\cdot v,w} - a \cdot e_{v,w}) = T_0(e_{a\cdot v,w}) - a \cdot T_0(e_{v,w}) = t(a \cdot v, w) - a \cdot t(v, w) = 0$ by (i). Hence we can define $T([x]_R) = T_0(x)$ since the latter does not depend on the representative x by $T_0(R) = 0$. (See the 'summary' at the end of Section 2.7.) Direct computation gives $T(v \otimes w) = T([e_{v,w}]_R) = T_0(e_{v,w}) = t(v, w)$ as wanted. □

2.8.8. Proposition. *Let B be a basis of V and C be a basis of W. Then $B \otimes C = \{\, b \otimes c \,|\, b \in B, c \in C \,\}$ is a basis of $V \otimes W$ that is in bijection with $B \times C$ (i.e. all $b \otimes c$ are distinct in $V \otimes W$).* ([7]) *Hence if V and W are finite-dimensional then*

$$\dim(V \otimes W) = \dim(V) \cdot \dim(W).$$

Proof. Let us verify that all $b\otimes c$ are distinct. Let $b \neq b'$ be in B and $c, c' \in C$ and let us see that $b \otimes c \neq b' \otimes c'$. (A similar argument proves that $b \otimes c \neq b' \otimes c'$ if $c \neq c'$.) Consider the dual vectors $b^{\#}\colon V \to \mathbb{F}$ and $c^{\#}\colon W \to \mathbb{F}$, which pick the '$b$-th' coordinate and '$c$-th' coordinate respectively. Let $t\colon V \times W \to \mathbb{F}$ be defined by $t(v, w) = b^{\#}(v) \cdot c^{\#}(w)$. It is easy to see that t is bilinear. So by Proposition 2.8.7 there exists a linear transformation $T\colon V \otimes W \to \mathbb{F}$ such that $T(v \otimes w) = t(v, w) = b^{\#}(v) \cdot c^{\#}(w)$. But then $T(b \otimes c) = 1 \cdot 1 = 1$ whereas $T(b' \otimes c') = 0 \cdot c^{\#}(c') = 0$. In particular $b \otimes c \neq b' \otimes c'$.

We can use the same construction to show that $B \otimes C$ is linearly independent. Suppose that $(b_1, c_1), \ldots, (b_r, c_r) \in B\times C$ are r distinct pairs and that $a_1, \ldots, a_r \in \mathbb{F}$ are such that

$$a_1 \cdot (b_1 \otimes c_1) + \cdots + a_r \cdot (b_r \otimes c_r) = 0 \tag{2.8.9}$$

in $V \otimes W$. (Note that we are not going to spell out what it means for a class to be zero in the quotient modulo R. That would be a nightmare.) We construct $T\colon V \otimes W \to \mathbb{F}$ as in the first part of the proof for $b = b_1$ and $c = c_1$. And we apply T to the linear combination (2.8.9). Since $T(b_i \otimes c_i) = b_1^{\#}(b_i) \cdot c_1^{\#}(c_i)$ is zero unless both $b_i = b_1$ and $c_i = c_1$, we see that $a_1 = 0$. Repeating for every $b_j \otimes c_j$ we get $a_j = 0$ for all $j = 1, \ldots, r$.

[7] Beware that $B \otimes C$ is not a tensor product of vector spaces of course, just a notation for the set of all $b \otimes c$ in $V \otimes W$, for all $b \in B$ and $c \in C$.

Finally, $B\otimes C$ spans $V\otimes W$ is easy. We already know that every vector in $V\otimes W$ is a sum of simple tensors. Now for $v \in V$ and $w \in W$ we have $v = \sum_{i=1}^{m} x_i \cdot b_i$ for some $x_1, \ldots, x_m \in \mathbb{F}$ and $b_1, \ldots, b_m \in B$ and similarly $w = \sum_{j=1}^{n} y_j \cdot c_j$ for some $y_1, \ldots, y_n \in \mathbb{F}$ and $c_1, \ldots, c_n \in C$. Then by bilinearity of $\otimes$, we have

$$v \otimes w = \sum_{i=1}^{m} \sum_{j=1}^{n} (x_i y_j) \cdot b_i \otimes c_j.$$

This finishes the proof that $B \otimes C$ is a basis of $V \otimes W$ and that $B \times C \to B \otimes C$, $(b, c) \mapsto b \otimes c$ is a bijection. □

2.8.10. Example. We have $\mathbb{F}^m \otimes \mathbb{F}^n \cong \mathbb{F}^{mn}$. In fact, if we think of $\mathbb{F}^{mn}$ as $m \times n$ matrices then this isomorphism can be made very explicit: It maps $e_i \otimes e'_j$ to E_{ij}, where $e_1, \ldots, e_m$ and $e'_1, \ldots, e'_n$ are the canonical bases of $\mathbb{F}^n$ and $\mathbb{F}^n$ respectively and E_{ij} the canonical basis of $\mathrm{M}_{m\times n}(\mathbb{F})$.

2.8.11. Exercise. Show that the tensor product is additive in each variable, *i.e.*

$$(V_1 \oplus V_2) \otimes W \cong (V_1 \otimes W) \oplus (V_2 \otimes W)$$

and similarly on the other side, and prove that the tensor is symmetric, *i.e.*

$$V \otimes W \cong W \otimes V$$

for all vector spaces V, V_1, V_2, W.

2.8.12. Exercise. Let V and W be $\mathbb{F}$-vector spaces. Show that we have a canonical linear transformation

$$S = S_{V,W} \colon V^{\#} \otimes W \to \mathrm{Lin}(V, W)$$

mapping simple tensors $\ell\otimes w$ (for $\ell \in V^{\#} = \mathrm{Lin}(V, \mathbb{F})$ and $w \in W$) to $T_{\ell,w} \colon V \to W$ defined by $T_{\ell,w}(x) = \ell(x) \cdot w$, for all $x \in V$. (For instance, in the notation of Exercise 2.4.29, show that E_{ij} is $S(b_j^{\#} \otimes b'_i)$.) Show that $S_{V,W}$ is an isomorphism for all W if and only if V is finite-dimensional. [Hint: Finite-rank transformations $\left\{ T \colon V \to W \,\middle|\, \dim(\mathrm{Im}(T)) < \infty \right\}$ form a sub*space* of $\mathrm{Lin}(V, W)$ containing $\mathrm{Im}(S)$.]

CHAPTER 3

Diagonalization

For the whole chapter, $\mathbb{F}$ is a fixed field. Without explicit mention, $T\colon V \to V$ is a linear transformation from a finite-dimensional $\mathbb{F}$-vector space V to itself.

3.1. Diagonalizable operators

3.1.1. Definition. Let V be an $\mathbb{F}$-vector space. A linear transformation $T\colon V \to V$ from V to itself is often called a *(linear) operator* on V.

3.1.2. Example. Any square matrix $A \in \mathrm{M}_{n\times n}(\mathbb{F})$ defines a linear operator T_A on $\mathbb{F}^n = \mathrm{M}_{n\times 1}(\mathbb{F})$, where as usual $T_A(x) = A \cdot x$ is given by matrix multiplication.

Conversely, the *choice of a basis* B of a finite-dimensional $\mathbb{F}$-vector space V yields a square matrix

$$[T]_B := [T]_{B,B}$$

for every operator $T\colon V \to V$. Note that we use the *same* basis B in the source and target space(s) of T, hence the abbreviated notation $[T]_B$. The matrix $[T]_B$ depends on B and we discussed in Section 2.5 how to relate $[T]_B$ and $[T]_C$ for different bases B and C of the same V. We have $[T]_B = Q \cdot [T]_C \cdot Q^{-1}$ where $Q = Q_{B,C} = [\mathrm{Id}_V]_{B,C}$ and $Q^{-1} = Q_{C,B} = [\mathrm{Id}_V]_{C,B}$ are invertible.

3.1.3. Remark. Operators $T\colon V \to V,\ x \mapsto T(x)$ have the property that they can be iterated, since the output $T(x)$ belongs to the source space of T, meaning that $T(T(x))$ makes sense, and so on. This is easy to motivate in applications: If your mathematical model of a concrete problem is that T describes what happens to the $x \in V$ after 1 year (or 1 second) then we may want to know what happens after 2 years, 3 years, etc (or 2 seconds, 3 seconds, etc). This amounts to re-apply T to $T(x)$, and again, and again, etc.

For every $i \in \mathbb{N}$ we let $T^i = T \circ T \circ \cdots \circ T$ be the operator $V \to V$ defined by composing T with itself i times. In other words, for every $x \in V$, we have $T^i(x) = T(T(\cdots T(T(x))\cdots))$ where T appears i times. Note that for $i = 0$, this means $T^0 = \mathrm{Id}_V$. (If we do not apply T at all, every $x \in V$ stays where it is.) We have for all $i, j \in \mathbb{N}$ the following equality of operators $V \to V$

$$T^i \circ T^j = T^{i+j}.$$

More generally, if $P = a_0 + a_1X + \cdots + a_dX^d = \sum_{i=0}^{d} a_iX^i \in \mathbb{F}[X]$ is a polynomial with coefficients in $\mathbb{F}$ then $P(T)\colon V \to V$ is the operator obtained by ‘evaluating’ the polynomial P at T (*i.e.* replacing X by T, using $T^0 = \mathrm{Id}_V$)

$$P(T) = \sum_{i=0}^{d} a_i \cdot T^i = a_0 \cdot \mathrm{Id}_V + a_1T + \cdots + a_dT^d.$$

In explicit terms, for every $x \in V$ we have the following equality in V

$$(P(T))(x) = \sum_{i=0}^{d} a_i \cdot T^i(x) = a_0 \cdot x + a_1 \cdot T(x) + \cdots + a_d \cdot T^d(x).$$

3.1.4. Example. Suppose that $V = \mathbb{F}^n$ and $T = T_A$ for $A \in \mathrm{M}_{n\times n}(\mathbb{F})$. Then $P(T) = T_{P(A)}$ where $P(A)$ is the polynomial $P = a_0 + a_1X + \cdots + a_dX^d$ evaluated at the matrix A, that is, $P(A) = a_0 \cdot I_n + a_1 \cdot A + \cdots + a_d \cdot A^d$. Note that iterated powers of matrices $A^i = A \cdot A \cdots A$ can be rather complicated, as each step is a matrix multiplication (a sum of n products of n elements, for each of the $n \times n$ entries, the whole thing repeated $i - 1$ times).

We can commute 'take the matrix' and 'evaluate a polynomial':

3.1.5. Proposition. *Let $T\colon V \to V$ be an operator on V of finite dimension and let B be a basis of V. Let $P \in \mathbb{F}[X]$ be a polynomial and consider $P(T)\colon V \to V$. Then the matrix $[P(T)]_B = [P(T)]_{B,B}$ of the operator $P(T)$ with respect to B is*

$$[P(T)]_B = P(A)$$

where $A = [T]_B = [T]_{B,B}$ is the matrix of T with respect to B.

Proof. We have by Theorem 2.4.16 and induction on $i \in \mathbb{N}$ that $[T^i]_B = [T \circ \cdots \circ T]_{B,B} = [T]_{B,B} \cdots [T]_{B,B} = ([T]_B)^i$. Hence if $P = \sum_{i=0}^d a_iX^i$ we have

$$\begin{aligned} [P(T)]_B &= [\textstyle\sum_{i=0}^d a_iT^i]_B && \text{by definition of } P(T) \\ &= \textstyle\sum_{i=0}^d a_i[T^i]_B && \text{by linearity of } [-]_{B,B}\colon \mathrm{Lin}(V,V) \to \mathrm{M}_{n\times n}(\mathbb{F}) \\ &= \textstyle\sum_{i=0}^d a_iA^i && \text{since we saw above that } [T^i]_B = ([T]_B)^i = A^i \\ &= P(A) && \text{by definition of } P(A). \end{aligned}$$

This is the claim. □

Let us now turn attention to diagonal matrices. We recall the notation.

3.1.6. Definition. A square matrix $A \in \mathrm{M}_{n\times n}(\mathbb{F})$ is called *diagonal* if $A_{ij} = 0$ for all $1 \leq i, j \leq n$ with $i \neq j$. We write $\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ for the diagonal matrix A such that $A_{ii} = \lambda_i$ for all $1 \leq i \leq n$:

$$\mathrm{diag}(\lambda_1, \ldots, \lambda_n) = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix}.$$

Polynomials are much easier to evaluate on diagonal matrices. In particular, they remain diagonal.

3.1.7. Proposition. *Let $A = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ be a diagonal matrix. Then for every $i \in \mathbb{N}$ we have $A^i = \mathrm{diag}(\lambda_1^i, \lambda_2^i, \ldots, \lambda_n^i)$. More generally, if $P \in \mathbb{F}[X]$ then we have $P(A) = \mathrm{diag}(P(\lambda_1), P(\lambda_2), \ldots, P(\lambda_n))$.*

Proof. Both claims are easy to verify from the formulas for matrix multiplication. We have $A^0 = I_n = \mathrm{diag}(1, \ldots, 1) = \mathrm{diag}(\lambda_1^0, \ldots, \lambda_n^0)$ and by induction

$$A^{i+1} = A \cdot A^i = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix} \cdot \begin{pmatrix} \lambda_1^i & 0 & \cdots & 0 \\ 0 & \lambda_2^i & & 0 \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n^i \end{pmatrix} = \begin{pmatrix} \lambda_1^{i+1} & 0 & \cdots & 0 \\ 0 & \lambda_2^{i+1} & & 0 \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n^{i+1} \end{pmatrix}$$

which is indeed $\mathrm{diag}(\lambda_1^{i+1}, \ldots, \lambda_n^{i+1})$. The result for $P(A)$ follows linearly. □

3.1.8. Remark. Let us emphasize how simpler it is to compute $P(A)$ when $A = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ is diagonal, compared to the general case. ([1])

3.1.9. Corollary. *Suppose that B is a basis of V such that $[T]_B = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ is diagonal. Then $[P(T)]_B = \operatorname{diag}(P(\lambda_1), \ldots, P(\lambda_n))$ is diagonal as well.*

Proof. Simply combine Propositions 3.1.5 and 3.1.7. □

3.1.10. Remark. If $A, D \in \mathrm{M}_{n\times n}(\mathbb{F})$ are *similar*, $A \sim D$, meaning that $A = Q{\cdot}D{\cdot}Q^{-1}$ for an invertible matrix Q, then $P(A) \sim P(D)$ and more precisely $P(A) = Q \cdot P(D) \cdot Q^{-1}$. Indeed, $A^i = (QDQ^{-1})^i = QDQ^{-1}QDQ^{-1} \cdot \ldots \cdot QDQ^{-1}QDQ^{-1} = QD^iQ^{-1}$ by induction on i. Thus if $P = \sum_{i=0}^{d} a_i X^i$ then $P(A) = \sum_{i=0}^{d} a_i A^i = \sum_{i=0}^{d} a_i\, Q \cdot D^i \cdot Q^{-1} = Q \cdot (\sum_{i=0}^{d} a_i D^i \cdot Q^{-1}) = Q \cdot (\sum_{i=0}^{d} a_i D^i) \cdot Q^{-1} = Q \cdot P(D) \cdot Q^{-1}$. In particular, if $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ is diagonal then $P(A) = Q \cdot \operatorname{diag}(P(\lambda_1), \ldots, P(\lambda_n)) \cdot Q^{-1}$ is very easy to compute.

In view of Corollary 3.1.9, it is interesting to know *when* the matrix of an operator $T\colon V \to V$ can be diagonal, *i.e.* when such a basis B exists. Now note that this heavily depends on B, as diagonal matrices do *not* commute with other matrices in general. (Diagonal matrices of the form $a \cdot I_n = \operatorname{diag}(a, a, \ldots, a)$ *do* commute with any Q. And indeed $[a \cdot \mathrm{Id}_V]_B = a \cdot I_n$ for all B.) So it can happen that $[T]_B$ is diagonal for a nice basis B but that $[T]_C$ is not diagonal for another. We have $[T]_C = Q \cdot [T]_B \cdot Q^{-1}$ where $Q = [\mathrm{Id}]_{B,C}$ but this product cannot be turned around to cancel Q and Q^{-1} in general. Hence the notion of diagonal*izable* operator, not of diagonal operator. This is the fundamental definition of the chapter.

3.1.11. Definition. Let $T\colon V \to V$ be a linear operator on a finite-dimensional $\mathbb{F}$-vector space V. We say that T is *diagonalizable* if there exists a basis B of V such that the matrix $[T]_B = [T]_{B,B}$ is diagonal.

A matrix $A \in \mathrm{M}_{n\times n}(\mathbb{F})$ is called *diagonalizable* if $T_A\colon \mathbb{F}^n \to \mathbb{F}^n$ is diagonalizable. By Corollary 2.5.8 this means that there exists an invertible matrix $Q \in \mathrm{GL}_n(\mathbb{F})$ and a diagonal matrix $D = \operatorname{diag}(d_1, \ldots, d_n)$ such that $A = Q \cdot D \cdot Q^{-1}$. In other words, A is similar to a diagonal matrix.

3.1.12. Example. Let $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ be the (orthogonal) symmetry of the plane with respect to some fixed line $L = \operatorname{Span}(\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right])$ for $\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right] \neq 0$. Then T is diagonalizable. Indeed, let $B = \{v_1, v_2\}$ be the basis of $\mathbb{R}^2$ consisting of $v_1 = \left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$ and $v_2 = \left[\begin{smallmatrix} -b \\ a \end{smallmatrix}\right]$ for instance. The vectors v_1, v_2 are chosen so that $v_1 \in L$ and v_2 is orthogonal to v_1 (and non-zero). Indeed, the scalar product $\langle v_1, v_2 \rangle = \langle \left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right], \left[\begin{smallmatrix} -b \\ a \end{smallmatrix}\right] \rangle = a{\cdot}(-b) + b{\cdot}a = 0$. Now we have $T(v_1) = v_1$ since $v_1 \in L$ and $T(v_2) = -v_2$ since v_2 is orthogonal to L. Hence $[T]_B = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ is diagonal.

We can also compute the matrix $[T]_C$ in the canonical basis $C = \{e_1, e_2\}$ although computing $T(e_1)$ and $T(e_2)$ form the geometric definition of T is a little tricky. For this, note that $Q := Q_{C,B} = \left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right)$ and therefore $Q^{-1} = \frac{1}{a^2+b^2} \left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right)$. See Example D.2.19. Then

$$[T]_C = Q \cdot [T]_B \cdot Q^{-1} = \frac{1}{a^2+b^2} \left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right) \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right) \left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right) = \frac{1}{a^2+b^2} \left(\begin{smallmatrix} a^2-b^2 & 2ab \\ 2ab & -a^2+b^2 \end{smallmatrix}\right).$$

[1] Morally, for every reasonable function f for which $f(A)$ makes sense for matrices A, one can claim that $f(\operatorname{diag}(\lambda_1, \ldots, \lambda_n)) = \operatorname{diag}(f(\lambda_1), \ldots, f(\lambda_n))$. Indeed, if we have a metric, say for $\mathbb{F} = \mathbb{R}$ for instance, and if $f(X) = a_0 + a_1X + a_2X^2 + \cdots$ admits a Taylor expansion as an infinite series, then this claim is 'just' the limit over $\deg(P) \to \infty$ of Proposition 3.1.7. For instance, we can define $\exp(A)$ using $f(X) = \sum_{i=0}^{\infty} \frac{1}{i!} X^i$. In that case, $\exp(\operatorname{diag}(\lambda_1, \ldots, \lambda_n)) = \operatorname{diag}(e^{\lambda_1}, \ldots, e^{\lambda_n})$.

3.1.13. Exercise. Let $a, b \in \mathbb{R}$ not both zero and let $A = \frac{1}{a^2+b^2}\begin{pmatrix} a^2-b^2 & 2ab \\ 2ab & -a^2+b^2 \end{pmatrix}$. Compute A^2 directly. Then compare to the above example.

3.2. Eigenvalues and characteristic polynomial

For this section, V is a finite-dimensional $\mathbb{F}$-vector space and $T\colon V \to V$ is a linear operator. In the previous section, we motivated the search for a basis B of V such that $A = [T]_B$ would be diagonal, say

$$A = \operatorname{diag}(\lambda_1, \dots, \lambda_n) = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix}.$$

Let us unpack the meaning of this equation, with the basis vectors explicitly named

$$[T]_B = \operatorname{diag}(\lambda_1, \dots, \lambda_n) \qquad \text{where} \qquad B = \{b_1, \dots, b_n\}. \tag{3.2.1}$$

By the definition of $[T]_B = [T]_{B,B}$ this exactly means that for each $j = 1, \dots, n$ the coordinates of $T(b_j)$ in the basis B is the j-th column of $\operatorname{diag}(\lambda_1, \dots, \lambda_n)$, that is $\lambda_j \cdot e_j$. By definition of the coordinates of a vector in the basis $B = \{b_1, \dots, b_n\}$, this relation $[T(b_j)]_B = \lambda_j \cdot e_j$ means that $T(b_j) = 0 \cdot b_1 + \cdot + 0 \cdot b_{j-1} + \lambda_j \cdot b_j + 0 \cdot b_{j+1} + \cdots + 0 \cdot b_n = \lambda_j \cdot b_j$. So we have proved:

3.2.2. Proposition. *An operator $T\colon V \to V$ and a basis $B = \{b_1, \dots, b_n\}$ satisfy $[T]_B = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ if and only if $T(b_j) = \lambda_j \cdot b_j$ for all $j = 1, \dots, n$.* □

3.2.3. Remark. In other words, the vectors b_j of the basis B are remarkable (with respect to T) in that their image under the operator is a multiple of themselves: $T(b_j) \in \operatorname{Span}(b_j)$. Note that there is an unremarkable vector v such that $T(v) \in \operatorname{Span}(v)$ for all T and that is $v = 0$. But our basis vector b_j cannot be zero.

Let us rephrase the above. What is remarkable about T being diagonalizable is that we can build a basis of V with (non-zero) vectors b_j that each satisfy $T(b_j) = \lambda_j \cdot b_j$ for some $\lambda_j \in \mathbb{F}$.

It is actually easier to isolate which λ_j can appear in this way and then to look for the b_j after the fact. Let us formalize this.

3.2.4. Definition. Let $T\colon V \to V$ be an operator on a finite-dimensional $\mathbb{F}$-vector space V. Let $\lambda \in \mathbb{F}$ be a scalar. We say that λ is an *eigenvalue of* T if there exists a *non-zero* vector $v \in V$ such that $T(v) = \lambda \cdot v$.

For a matrix $A \in \mathrm{M}_{n\times n}(\mathbb{F})$, we say that λ is an *eigenvalue of* A if it is an eigenvalue of T_A, that is, if there exists a non-zero $x \in \mathbb{F}^n$ such that $A \cdot x = \lambda x$.

3.2.5. Remark. Again, appreciate how these definitions would be silly if we did not insist on v (or x) being *non-zero.* Indeed $T(0) = \lambda \cdot 0$ for all $\lambda \in \mathbb{F}$.

3.2.6. Proposition. *Let $T\colon V \to V$ be a linear operator on a finite-dimensional $\mathbb{F}$-vector space V and let $\lambda \in \mathbb{F}$ be a scalar. The following are equivalent:*

(i) *The scalar λ is an eigenvalue of T*
(ii) *The operator $(T - \lambda \cdot \mathrm{Id}_V)\colon V \to V$ has a non-zero kernel.*
(iii) *The operator $(T - \lambda \cdot \mathrm{Id}_V)\colon V \to V$ is not invertible.*
(iv) *For every basis C of V, the scalar λ is an eigenvalue of the matrix $A = [T]_C$.*

(v) *There exists a basis C of V such that λ is an eigenvalue of $[T]_C$.*
(vi) *For every basis C of V, we have* $\det(A - \lambda \cdot I_n) = 0$*, where* $A = [T]_C$.
(vii) *There exists a basis C of V such that* $\det(A - \lambda \cdot I_n) = 0$*, where* $A = [T]_C$.

Proof. Note that $(T-\lambda\cdot\mathrm{Id}_V)(v) = T(v)-\lambda\cdot v$. So the equivalence of (i) and (ii) is direct from Definition 3.2.4. And (ii) is equivalent to (iii) since $\dim(V) = \dim(V)$ (duh!) and we can use Corollary 2.3.14 with $V' = V$. To show that (iii)⇔(vi)⇔(vii) note that if $A = [T]_C$ then $[T - \lambda \cdot \mathrm{Id}_V]_C = [T]_C - \lambda \cdot [\mathrm{Id}_V]_C = A - \lambda \cdot I_n$. So we can use Corollary 2.4.26 which says that $T - \lambda\,\mathrm{Id}_V$ is an isomorphism if and only if its matrix is invertible in some (or in every) basis. Of course, we express invertibility by saying that the determinant is non-zero (Theorem D.2.15). Finally, to pass to (iv) and (v), we note as above that the square matrix $A - \lambda \cdot I_n$ is non-invertible is the same thing as saying that it is not injective, *i.e.* is the same thing as saying that $\mathrm{Ker}(A - \lambda \cdot I_n)$ is non-zero. This is equivalent to the very definition of λ being an eigenvalue of A since $x \in \mathrm{Ker}(A - \lambda \cdot I_n)$ is just saying $A \cdot x = \lambda \cdot x$. So we have (iv)⇔(vi) and (v)⇔(vii). □

3.2.7. Remark. Here is a mostly cosmetic remark but honesty requires us to make it. The definition of the determinant $\det(A)$ of a square matrix A (*e.g.* as in Definition D.2.3) does not use that the matrix has entries in a field. We just need what is called a commutative ring: We need to be able to add and multiply and have associativity and distributivity, as for instance in axioms (F 1)-(F 8) – without asking for (F 9) and (F 10). In particular, we can compute $\det(A)$ for a matrix whose entries are in the *commutative ring* of polynomials $\mathbb{F}[X]$ in one variable X. We know how to add polynomials (the addition of the $\mathbb{F}$-vector space $\mathbb{F}[X]$) but we also know how to multiply them by using distributivity and commutativity and the rule $X^i \cdot X^j = X^{i+j}$. Note that $\mathbb{F}[X]$ is not a field though since for instance the polynomial X has no inverse. In any case, the following definition makes sense.

3.2.8. Definition. Let $A \in \mathrm{M}_{n\times n}(\mathbb{F})$ be a square matrix. Its *characteristic polynomial* $P_A \in \mathbb{F}[X]$ is defined as

$$P_A(X) = \det(A - X \cdot I_n) = \det \begin{pmatrix} A_{11}-\mathbf{X} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22}-\mathbf{X} & \ddots & \vdots \\ \vdots & \ddots & \ddots & A_{n-1,n} \\ A_{n1} & \cdots & A_{n,n-1} & A_{n,n}-\mathbf{X} \end{pmatrix}.$$

See Appendix D.2 for a reminder on the determinant.

3.2.9. Lemma. *The characteristic polynomial $P_A(X)$ is a polynomial of degree n in the variable X, with leading term $(-1)^n$ and constant term* $\det(A)$*, that is,*

$$P_A(X) = (-1)^n X^n + a_{n-1}X^{n-1} + \cdots + a_1 X + \det(A).$$

We can also check that $a_{n-1} = (-1)^{n-1}\,\mathrm{tr}(A)$.

Proof. The constant term is obvious: $P_A(0) = \det(A - 0 \cdot I_n) = \det(A)$. Let us determine the coefficients of X^d in $P_A(X)$ for $d \geq n-1$. With the basic definition (Definition D.2.3) it is an unpleasant exercise by induction on n. With the 'better' definition of the determinant using permutations (Theorem D.2.34), we get

$$\begin{aligned} \det(A - X \cdot I_n) &= \textstyle\sum_{\sigma\in S_n} \mathrm{sgn}(\sigma) \prod_{i=1}^n (A - X \cdot I_n)_{i,\sigma(i)} \\ &= \textstyle\prod_{i=1}^n (A_{i,i} - X) + \sum_{\substack{\sigma\in S_n \\ \sigma\neq \mathrm{id}}} \mathrm{sgn}(\sigma) \prod_{i=1}^n (A - X \cdot I_n)_{i,\sigma(i)} \end{aligned}$$

by separating the term $\sigma = \mathrm{id}$ from the rest. Expanding the first term it reads

$$(3.2.10)\quad (A_{11} - X) \cdots (A_{nn} - X) = (-X)^n + (A_{11} + \cdots + A_{nn})(-X)^{n-1} + Q(X)$$

where $Q(X)$ is a polynomial of degree at most $n-2$ in X. Similarly, all products $\prod_{i=1}^{n}(A - X \cdot I_n)_{i,\sigma(i)}$ for $\sigma \neq \mathrm{id}$ in the rest of the above sum

$$(3.2.11)\qquad \sum_{\substack{\sigma \in S_n \\ \sigma \neq \mathrm{id}}} \mathrm{sgn}(\sigma) \prod_{i=1}^{n} (A - X \cdot I_n)_{i,\sigma(i)}$$

are polynomials of degree at most $n-2$ in X. Indeed each factor $(A - X \cdot I_n)_{ij}$ is either the constant A_{ij} for $i \neq j$ (for $(I_n)_{i,j} = 0$ in that case) or is the degree one polynomial $(A_{ii} - X)$ for $i = j$. Now if $\sigma \neq \mathrm{id}$ then $\sigma \colon \{1, \ldots, n\} \xrightarrow{\sim} \{1, \ldots, n\}$ moves at least *two* indices, say j, k. Then $(A - X \cdot I_n)_{i,\sigma(i)} = A_{i,\sigma(i)}$ is a constant polynomial when $i \in \{j, k\}$. So the product $\prod_{i=1}^{n}(A - X \cdot I_n)_{i,\sigma(i)}$ is a product of n polynomials of degree at most one in X, at least two of which are actually constant. Hence this product is a polynomial of degree at most $n-2$ in X. So the coefficients of degree n and $n-1$ (and higher) will not be changed when we add (3.2.11) to (3.2.10). This gives the announced result. □

3.2.12. Proposition. *Let $T \colon V \to V$ be an operator on a finite-dimensional $\mathbb{F}$-vector space. Let $A = [T]_C$ be its matrix with respect to some basis C of V. Then the characteristic polynomial P_A is independent of the choice of said basis C. We call it the* characteristic polynomial of *T:*

$$P_T = \det([T]_C - X \cdot I_n) \qquad \textit{for } \text{any} \textit{ basis } C \textit{ of } V.$$

Proof. Let C' be another basis of V and $A' = [T]_{C'}$. We need to show that A and A' have the same characteristic polynomial. We know that $A' = Q^{-1}AQ$ where $Q = [\mathrm{Id}]_{C,C'}$. But then $A' - X \cdot I_n = Q^{-1}A\,Q - X \cdot I_n \cdot (Q^{-1}Q) = Q^{-1}A\,Q - Q^{-1}X \cdot I_n\,Q = Q^{-1}(A\,Q - X \cdot I_n\,Q) = Q^{-1}(A - X \cdot I_n)Q$, where the second equality used that the matrix $X \cdot I_n = \mathrm{diag}(X, X, \ldots, X)$ commutes with Q. And then those similar matrices have the same determinant (Corollary D.2.18):

$$P_{A'} = \det(A' - X \cdot I_n) = \det(A - X \cdot I_n) = P_A$$

as was to be shown. □

3.2.13. Corollary. *Let $\lambda \in \mathbb{F}$ and $T \colon V \to V$. Then λ is an eigenvalue of T if and only if $P_T(\lambda) = 0$, that is, λ is a root of the characteristic polynomial of T.*

Proof. Choose a basis C of V and let $A = [T]_C$. We have seen in Proposition 3.2.6 that λ is an eigenvalue of T if and only if $P_A(\lambda) = \det(A - \lambda I_n)$ vanishes. This means that λ is a root of $P_A = P_T$. □

At this stage, the reader uncomfortable with polynomials, roots, etc, should revise these prerequisites. (See Appendix D.1.)

3.2.14. Corollary. *An operator $T \colon V \to V$ has at most* $\dim(V)$ *distinct eigenvalues.*

Proof. We saw that the characteristic polynomial (that is, the characteristic polynomial of the matrix of T with respect to any basis of V) has degree $\dim(V)$. A polynomial of degree d has at most d distinct roots. (See Proposition D.1.3.) □

3.2.15. Example. Let $T = T_A\colon \mathbb{R}^2 \to \mathbb{R}^2$ be given by $A = \frac{1}{a^2+b^2}\begin{pmatrix} a^2-b^2 & 2ab \\ 2ab & -a^2+b^2 \end{pmatrix}$. as in Example 3.1.12. Then $P_A(X) = \det(A - X \cdot I_2)$ is

$$(\frac{1}{a^2+b^2})^2 \det(\begin{pmatrix} a^2-b^2-X(a^2+b^2) & 2ab \\ 2ab & -a^2+b^2-X(a^2+b^2) \end{pmatrix})$$

but can also be computed as $P_A(X) = X^2 - \operatorname{tr}(A) \cdot X + \det(A) = X^2 - 1$ since $\det(A) = (\frac{1}{a^2+b^2})^2(-a^4 + 2a^2b^2 - b^4 - 4a^2b^2) = -1$ Hence $P_A(X) = (X-1)(X+1)$ and A (and T_A) has two eigenvalues: 1 and -1 (in $\mathbb{R}$). This is fitting for a symmetry. After all, T was constructed in such a way that $T(v_1) = v_1$ and $T(v_2) = -v_2$ in the notation of Example 3.1.12. So we already knew that 1 and -1 are eigenvalues and since $\dim(V) = 2$ they are the only two.

3.2.16. Example. Let $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ the matrix of the rotation of the plane by angle $\frac{\pi}{2}$ around the origin, in the canonical basis. We see by a direct geometric reasoning that A cannot have eigenvalues as there is no non-zero vector $v \in \mathbb{R}^2$ such that $A \cdot v \in \operatorname{Span}(v)$ since $A \cdot v$ makes a right angle with v. We can also check this algebraically. We have $P_A(X) = \det\begin{pmatrix} -X & 1 \\ -1 & -X \end{pmatrix} = X^2 + 1$ and this polynomial has no real root.

3.2.17. Corollary. *Let $T\colon V \to V$ be a diagonalizable operator. Then P_T splits completely: $P_T(X) = (-1)^n(X - \lambda_1)\cdots(X - \lambda_n)$ for $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$.*

Proof. The characteristic polynomial $P_T(X)$ can be computed as $P_A(X)$ for the matrix $A = [T]_B$ of T in *any* basis. Since T is diagonalizable, let us choose B so that $[T]_B = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. The result follows. □

3.2.18. Remark. Beware that the complete decomposition of the polynomial P_T over $\mathbb{F}$ is necessary but not sufficient for T to be diagonalizable. Consider $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $T = T_A\colon \mathbb{F}^2 \to \mathbb{F}^2$. Then $P_T(X) = P_A(X) = X^2$ and 0 is the only eigenvalue. However, $A \neq 0$ so it cannot be similar to $\operatorname{diag}(0,0) = 0$. Therefore T is not diagonalizable even though $P_T(X)$ is completely split.

3.3. Eigenvectors and eigenspaces

As before, V is a finite-dimensional $\mathbb{F}$-vector space and $T\colon V \to V$ is a linear operator. We saw in the previous section what eigenvalues were and how to find them. But we should remember Proposition 3.2.2: To show that T is diagonalizable we need a whole basis B of V consisting of vectors b such that $T(b) = \lambda \cdot b$ for an eigenvalue λ. Let us give those a name.

3.3.1. Definition. Let $\lambda \in \mathbb{F}$ be an eigenvalue of $T\colon V \to V$. Recall that it means that there exists a *non-zero* $v \in V$ such that $T(v) = \lambda v$. The entire subspace on which T acts via λ is called the *eigenspace* of T for the eigenvalue λ:

$$E_\lambda(T) = \left\{ v \in V \,\middle|\, T(v) = \lambda \cdot v \right\} = \operatorname{Ker}(T - \lambda \cdot \operatorname{Id}_V).$$

Vectors in $E_\lambda(T)$ (or sometimes only non-zero vectors in there) are called *eigenvectors* of T for the eigenvalue λ. These are the (non-zero) $v \in V$ such that $T(v) = \lambda v$. For λ to be an eigenvalue means that T admits a non-zero eigenvector for the eigenvalue λ. (2)

[2] For readers preoccupied by exams, grades, etc, you will not be penalized for saying that 0 is an eigenvector. However you will lose 100% of the points of any given problem if you even

3.3.2. Remark. Note that $E_\lambda(T)$ is a sub*space*! So it has a dimension, it has bases, etc.

3.3.3. Example. Let us continue Example 3.2.15 with $T = T_A\colon \mathbb{R}^2 \to \mathbb{R}^2$ the symmetry with $A = \frac{1}{a^2+b^2}\left(\begin{smallmatrix} a^2-b^2 & 2ab \\ 2ab & -a^2+b^2 \end{smallmatrix}\right)$. We found two eigenvalues 1 and -1. We can compute $E_1(A) = \operatorname{Ker}(A - I_2)$. Since $a^2 + b^2 \neq 0$, this amounts to

$$\operatorname{Ker}(\left(\begin{smallmatrix} a^2-b^2-(a^2+b^2) & 2ab \\ 2ab & -a^2+b^2-(a^2+b^2) \end{smallmatrix}\right)) = \operatorname{Ker}(\left(\begin{smallmatrix} -2b^2 & 2ab \\ 2ab & -2a^2 \end{smallmatrix}\right)) = \operatorname{Span}([\begin{smallmatrix} a \\ b \end{smallmatrix}]).$$

Of course, we guess $E_1(T)$ from the geometry but we can really solve the above by Gauss-Jordan. Indeed, if $b \neq 0$, the matrix $A - I_2$ can be row-reduced to $\left(\begin{smallmatrix} 1 & -\frac{a}{b} \\ 0 & 0 \end{smallmatrix}\right)$, whose kernel is $\operatorname{Span}([\begin{smallmatrix} \frac{a}{b} \\ 1 \end{smallmatrix}]) = \operatorname{Span}([\begin{smallmatrix} a \\ b \end{smallmatrix}])$ since $b \neq 0$. And if $b = 0$, the matrix $A - I_2$ can be row-reduced to $\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$, whose kernel is $\operatorname{Span}([\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}]) = \operatorname{Span}([\begin{smallmatrix} a \\ b \end{smallmatrix}])$ since $a \neq 0 = b$ in this case. In both cases we recover $\operatorname{Span}([\begin{smallmatrix} a \\ b \end{smallmatrix}])$ which was the line L of Example 3.1.12. A direct verification shows that $E_{-1}(T) = \operatorname{Span}([\begin{smallmatrix} -b \\ a \end{smallmatrix}]) = L^\perp$.

3.3.4. Remark. If we resume our quest of a basis B of V in which $[T]_B$ is diagonal, we can construct bases of each eigenspace $E_\lambda(T)$ and then try to create the basis B of V. The good news is that there is no risk of 'collision' between those separate bases.

3.3.5. Lemma. *Let $\lambda_1, \dots, \lambda_r$ be $r \geq 1$* distinct *eigenvalues of an operator $T\colon V \to V$. Suppose that $x_1 \in E_{\lambda_1}(T), \dots, x_r \in E_{\lambda_r}(T)$ are respective eigenvectors such that*

$$x_1 + \cdots + x_r = 0. \tag{3.3.6}$$

Then they are all trivial $x_1 = \ldots = x_r = 0$.

Proof. Let us prove this by induction on $r \geq 1$. The case $r = 1$ is a reading exercise. Suppose $r \geq 2$ and the lemma known for $r - 1$. Applying T to (3.3.6), using linearity of T and the fact that $T(x_i) = \lambda_i \cdot x_i$ we have

$$\lambda_1 \cdot x_1 + \lambda_2 \cdot x_2 + \cdots + \lambda_r \cdot x_r = 0.$$

But we can subtract λ_1 times (3.3.6) to the latter and get

$$(\lambda_2 - \lambda_1) \cdot x_2 + \cdots + (\lambda_r - \lambda_1) \cdot x_r = 0$$

since the first term disappeared: $\lambda_1 x_1 - \lambda_1 x_1 = 0$. Now this is a zero sum of $r - 1$ eigenvectors $(\lambda_i - \lambda_1) \cdot x_i \in E_{\lambda_i}(T)$ – recall that eigenspaces are subspaces, hence closed by multiplication by scalars. By induction hypothesis, if $r - 1$ eigenvectors for distinct eigenvalues (here $\lambda_2, \dots, \lambda_r$) add up to zero, they are all zero. So

$$(\lambda_i - \lambda_1) \cdot x_i = 0$$

for all $i = 2, \dots, r$. Now $\lambda_1 \neq \lambda_i$ for $i \geq 2$ (since the λ_i are all distinct). Therefore the last relation forces $x_i = 0$ for all $i = 2, \dots, r$. Finally $x_1 = 0$ by (3.3.6) and the already proved $x_2 = \ldots = x_r = 0$. □

3.3.7. Proposition. *Let $\lambda_1, \dots, \lambda_r$ be $r \geq 1$* distinct *eigenvalues of an operator $T\colon V \to V$. Let B_i be a basis of $E_{\lambda_i}(T)$ for each $i = 1, \dots, r$. Then these are all disjoint ($B_i \cap B_j = \varnothing$ for all $i \neq j$) and $B_1 \cup \cdots \cup B_r$ is linearly independent.*

remotely imply that an eigenvalue is a scalar $\lambda \in \mathbb{F}$ such that there exists $v \in V$ with $T(v) = \lambda v$. It is crucial that such a v exists that is also non-zero! Once we know that this happens, the eigenspace $E_\lambda(T)$ should be a subspace; so it contains zero; nothing to write home about.

Proof. The fact that $B_i \cap B_j = \varnothing$ for all $i \neq j$ is immediate from Lemma 3.3.5 with $r = 2$. An x in the intersection would give $x + (-x) = 0$. Let us show linear independence. Suppose that we have a trivial linear combination of vectors in the union $B_1 \cup \cdots \cup B_r$. Regrouping those vectors by eigenspace, we have distinct vectors $x_{1,1}, \ldots, x_{s_1,1} \in B_1 \subset E_{\lambda_1}(T), \ldots, x_{1,r}, \ldots, x_{s_r,r} \in B_r \subset E_{\lambda_r}(T)$ and scalars $a_{1,1}, \ldots, a_{s_1,1}, \ldots, a_{1,r}, \ldots, a_{s_r,r} \in \mathbb{F}$ such that

$$\sum_{j=1}^{r} \sum_{i=1}^{s_j} a_{i,j} \cdot x_{i,j} = 0.$$

This reads $y_1 + \cdots + y_r = 0$ where $y_j = \sum_{i=1}^{s_j} a_{i,j} \cdot x_{i,j}$ for each $j = 1, \ldots, r$. Note that y_j belongs to $E_{\lambda_j}(T)$ since all $x_{i,j}$ do. By Lemma 3.3.5 this forces $y_j = 0$ for all j. But this reads $\sum_{i=1}^{s_j} a_{i,j} \cdot x_{i,j} = 0$ and the $x_{1,j}, \ldots, x_{s_j,j}$ are linearly independent (they belong to the basis B_j of $E_{\lambda_j}(T)$). We conclude that $a_{i,j} = 0$ for all $i = 1, \ldots, s_j$, and this is true for all $j = 1, \ldots, r$. Hence the linear independence of $B_1 \cup \cdots \cup B_r$. □

This gives us the criterion for diagonalization: T is diagonalizable if and only if the sum of the dimensions of its eigenspaces equals the dimension of V.

3.3.8. Theorem. *Let $T\colon V \to V$ be a linear operator on a finite-dimensional $\mathbb{F}$-vector space. Then the following are equivalent:*

(i) *T is diagonalizable.*
(ii) *Let $\lambda_1, \ldots, \lambda_r \in \mathbb{F}$ be a complete list of* all *the r distinct eigenvalues of T in $\mathbb{F}$. Then $\sum_{i=1}^{r} \dim(E_{\lambda_i}(T)) = \dim(V)$.*
(iii) *There exists r distinct eigenvalues of T in $\mathbb{F}$ such that $\sum_{i=1}^{r} \dim(E_{\lambda_i}(T)) = \dim(V)$.*

Proof. Let $n = \dim(V)$. Suppose (i), *i.e.* T is diagonalizable. Then there exists a basis B such that $[T]_B$ is diagonal. Up to re-ordering B, we can assume that all eigenvectors for the same eigenvalue come ‘together’, that is,

$$[T]_B = \operatorname{diag}(\lambda_1, \ldots, \lambda_1, \lambda_2, \ldots, \lambda_2, \ldots, \lambda_r, \ldots, \lambda_r)$$

where $\lambda_1, \ldots, \lambda_r$ are the distinct eigenvalues ($\lambda_i \neq \lambda_j$ for all $i \neq j$). Say we have $m_1 \geq 1$ times λ_1, then m_2 times λ_2, ..., up to m_r times λ_r. The characteristic polynomial of T can be computed with the diagonal matrix $[T]_B$ which gives $P_T(X) = (-1)^n (X - \lambda_1)^{m_1} \cdots (X - \lambda_r)^{m_r}$. Then the matrix of $T - \lambda_i \cdot \mathrm{Id}_V$ in the basis B has exactly m_i times a zero entry on the diagonal (the $\lambda_i - \lambda_i$ of the ‘i-th group’) and all other entries $(\lambda_j - \lambda_i)$ non-zero. The rank of this diagonal matrix is $n - m_i$, hence its kernel has dimension m_i. In other words, for each $i = 1, \ldots, r$ the dimension of $E_{\lambda_i}(T) = \operatorname{Ker}(T - \lambda_i \mathrm{Id}_V)$ is m_i. But then $\sum_{i=1}^{r} \dim(E_{\lambda_i}(T)) = \sum_{i=1}^{r} m_i = \deg(P_T(X)) = n$. This gives (ii).

The implication (ii)⇒(iii) is trivial.

The implication (iii)⇒(i) is the interesting direction but we have done all the preparation to make it short. Suppose that $\lambda_1, \ldots, \lambda_r$ are r distinct eigenvalues and $\sum_{i=1}^{r} \dim(E_{\lambda_i}(T)) = \dim(V)$. Pick a finite basis B_i of $E_{\lambda_i}(T)$ for each $i = 1, \ldots, r$ (after all, each $E_{\lambda_i}(T)$ is a subspace of a finite-dimensional $\mathbb{F}$-vector space, hence has finite dimension). Form the union $B = B_1 \cup \cdots \cup B_r$. By Proposition 3.3.7, the B_i are all disjoints, so B has $|B_1| + \cdots + |B_r|$ elements. By hypothesis this adds up to $\dim(V)$. By the same Proposition 3.3.7 we know that B is linearly independent. As it has the right number of vectors, Corollary 1.6.31 tells us that B is a basis.

We have therefore constructed a basis of V consisting of eigenvectors (each B_i is in $E_{\lambda_i}(T)$) and we conclude by Proposition 3.2.2. $\square$

We can rephrase Theorem 3.3.8 in terms of so-called multiplicities.

3.3.9. Definition. Let $T\colon V \to V$ be an operator and $\lambda \in \mathbb{F}$ be an eigenvalue.

(1) The *algebraic multiplicity* $\mathrm{mult}_T(\lambda)$ is the multiplicity m of λ as a root of the characteristic polynomial of T, that is, $P_T(X) = (X-\lambda)^m \cdot Q(X)$ and $Q(\lambda) \neq 0$.
(2) The *geometric multiplicity* of λ is simply $\dim(E_\lambda(T))$.

The geometric multiplicity is always bounded by the algebraic multiplicity:

3.3.10. Proposition. *Let $\lambda \in \mathbb{F}$ be an eigenvalue of $T\colon V \to V$. Then we have*

$$1 \leq \dim(E_\lambda(T)) \leq \mathrm{mult}_T(\lambda).$$

Proof. Choose $C = \{c_1, \dots, c_d\}$ a basis of $E_\lambda(T)$ where $d = \dim(E_\lambda(T))$ is the geometric multiplicity of λ. Complete C into a basis B of V. (Corollary 1.6.32.) In the basis B, the matrix of T has the following form, directly from the definition

$$[T]_B = \begin{pmatrix} [T]_C & A' \\ 0 & A'' \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} \lambda & & 0 \\ & \ddots & \\ 0 & & \lambda \end{pmatrix} & A' \\ 0 & A'' \end{pmatrix}$$

where the second equality uses $[T]_C = \mathrm{diag}(\lambda, \dots, \lambda) \in \mathrm{M}_{d\times d}(\mathbb{F})$ since every $c \in C$ belongs to $E_\lambda(T)$, *i.e.* satisfies $T(c) = \lambda \cdot c$. We can compute $P_T(X)$ with the latter matrix and expanding $\det([T]_B - X \cdot I_n)$ along the first columns, we see that $P_T(X) = (\lambda - X)^d \cdot P_{A''}(X)$. In particular, the multiplicity of λ in $P_T(X)$ is at least d, as was to be shown. $\square$

3.3.11. Example. In the already encountered example $A = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$ we have $P_A(X) = X^2$ and $\lambda = 0$ is the only eigenvalue. Its algebraic multiplicity is 2 but its geometric multiplicity is the dimension of $E_0(A) = \mathrm{Ker}(A) = \mathrm{Span}(\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right])$ which is 1.

In fact, Proposition 3.3.10 is sharp in that we can produce examples of operators with arbitrary geometric and algebraic multiplicities, as long as they satisfy the conclusion of Proposition 3.3.10.

3.3.12. Exercise. Let $1 \leq d \leq m$ be integers. Let $\lambda \in \mathbb{F}$. Define a matrix $J = J(\lambda; m, d)$ in $\mathrm{M}_{m\times m}(\mathbb{F})$ that has only (possibly) non-zero entries on the diagonal and just above the diagonal, as follows, for all $1 \leq i, j \leq m$:

$$J_{ij} = \begin{cases} \lambda & \text{if } j = i \\ 1 & \text{if } j = i+1 \text{ and } 1 \leq i \leq m-d \\ 0 & \text{otherwise.} \end{cases}$$

For instance $J(\lambda; 3, 1) = \left(\begin{smallmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{smallmatrix}\right)$, $J(\lambda; 3, 2) = \left(\begin{smallmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{smallmatrix}\right)$ and $J(\lambda; 3, 3) = \left(\begin{smallmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{smallmatrix}\right)$. Write down the matrices $J(\lambda; 4, d)$ for $d = 1, 2, 3, 4$. Compute the geometric and algebraic multiplicities of λ for $T = T_J$ in a general $J = J(\lambda; m, d)$.

3.3.13. Exercise. Find a block-diagonal matrix $A = \begin{pmatrix} J_1 & 0 & & 0 \\ 0 & J_2 & & 0 \\ & & \ddots & \\ 0 & & & J_r \end{pmatrix}$ for $J_i \in \mathrm{M}_{m_i\times m_i}(\mathbb{F})$ as in Exercise 3.3.12 to produce a matrix A with any choice of r distinct eigenvalues $\lambda_1, \dots, \lambda_r$ of prescribed multiplicities $d_i = \dim(E_{\lambda_i}(A))$ and

$\mathrm{mult}_A(\lambda_i) = m_i$ for all $i = 1, \ldots, r$, where the r distinct scalars $\lambda_1, \ldots, \lambda_r$ and the numbers $1 \leq d_i \leq m_i$ are chosen by the opponent.

3.3.14. Exercise. Same problem as in Exercise 3.3.13 but where $A \in \mathrm{M}_{n\times n}(\mathbb{F})$ could have $n > m_1 + \cdots + m_r$, that is, where the characteristic polynomial $P_A(X)$ is not split. (Here $\mathbb{F}$ is a non-algebraically closed field, say $\mathbb{F} = \mathbb{R}$ for instance.)

We can now rephrase Theorem 3.3.8 in terms of multiplicities.

3.3.15. Corollary. *Let $T\colon V \to V$ be a linear operator on an $\mathbb{F}$-vector space V of dimension n. The following are equivalent:*

(i) *The operator T is diagonalizable.*

(ii) *The characteristic polynomial of T is completely split in $\mathbb{F}[X]$, i.e. we have $P_T(X) = (-1)^n(X - \lambda_1)^{m_1} \cdots (X - \lambda_r)^{m_r}$ for r distinct $\lambda_1, \ldots, \lambda_r \in \mathbb{F}$ and for $m_1, \ldots, m_r \geq 1$, and every eigenvalue of T has maximal geometric multiplicity, i.e. we have $\dim(E_{\lambda_i}(T)) = m_i$ for all $i = 1, \ldots, r$.*

Proof. The proof of (i)⇒(ii) is an easy exercise about diagonal matrices. Let us see why (ii) suffices to know that T is diagonalizable. This is easy from Theorem 3.3.8. Indeed, we need to check that $\sum_{i=1}^{r} \dim(E_{\lambda_i}(T)) = \dim(V) = n$. Recall that $n = \deg(P_T(X)) = m_1 + \ldots + m_r$ since P_T is split. So the assumption $\dim(E_{\lambda_i}(T)) = m_i$ for all $i = 1 \ldots, r$ gives us the result. □

3.3.16. Remark. In the previous statement, it is important to assume $P_T(X)$ completely split, for the equality between multiplicities $\dim(E_\lambda(T)) = \mathrm{mult}_T(\lambda)$ for all eigenvalues is not enough. For instance $A = \left(\begin{smallmatrix} \lambda & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{smallmatrix}\right)$ has characteristic polynomial $P_A(X) = -(X - \lambda)(X^2 + 1)$. So $T_A\colon \mathbb{R}^3 \to \mathbb{R}^3$ has only λ as eigenvalue in $\mathbb{F} = \mathbb{R}$ and $\dim(E_\lambda(T_A)) = m_T(\lambda) = 1$ but $1 \neq 3$.

3.3.17. Corollary. *Let $T\colon V \to V$ be a linear operator on a finite-dimensional $\mathbb{F}$-vector space V. Suppose that the field $\mathbb{F}$ is* algebraically closed*. Then T is diagonalizable if and only if $\dim(E_\lambda(T)) = m_T(\lambda)$ for all eigenvalues λ of T.*

Proof. Indeed $P_T(X)$ is completely split, as all polynomials over $\mathbb{F}$ are. □

3.3.18. Corollary. *Let $T\colon V \to V$ be a linear operator on an $\mathbb{F}$-vector space V of dimension n. Suppose that T has n distinct eigenvalues. Then T is diagonalizable.*

Proof. The n distinct eigenvalues are n distinct roots of the characteristic polynomial P_T, which has degree n. This forces P_T to be completely split and each eigenvalue to have algebraic multiplicity 1, hence geometric multiplicity 1 as well by Proposition 3.3.10. The sum of the geometric multiplicities is then $\sum_{i=1}^{n} 1 = n$ and we conclude by Corollary 3.3.15. □

3.4. Cayley-Hamilton Theorem*

We want to prove a very simple result: If $T\colon V \to V$ is a linear operator on a finite-dimensional $\mathbb{F}$-vector space V then $P_T(T) = 0$. Recall that $P_T(X)$ is the characteristic polynomial of T (Section 3.2). The expression $P_T(T)$ is an operator on V (as explained in Remark 3.1.3) and we want to show it is the zero operator, meaning that $(P_T(T))(v) = 0$ for every $v \in V$. To this end, it is convenient to do a little preparation.

3.4.1. Definition. Let $T\colon V \to V$ be an operator and $W \subseteq V$ be a subspace. We say that W is *T-invariant* if $T(W) \subseteq W$, that is, if $T(w) \in W$ for every $w \in W$.

3.4.2. Remark. Beware not to confuse this property with W being T-fixed, which would be $T(w) = w$ for every $w \in W$. The operator T still 'moves' a T-invariant subspace W, but lands inside of W. In particular, we can restrict T to an operator

$$\begin{array}{rccc} T_{|_W}\colon & W & \longrightarrow & W \\ & w & \longmapsto & T(w). \end{array}$$

This operator $T_{|_W}\colon W \to W$ is called the *restriction of* T to the subspace W.

3.4.3. Example. Of course, $\{0\}$ and V itself are always invariant subspaces. There is also a way to 'generate' an invariant subspace. Pick $v \in V$ and consider $\{v, T(v), T^2(v), \ldots\} = \left\{ T^i(v) \,\middle|\, i \in \mathbb{N} \right\} \subseteq V$. This subset is T-invariant but it is not a sub*space*. Then $W = \operatorname{Span}(\left\{ T^i(v) \,\middle|\, i \in \mathbb{N} \right\})$ is a T-invariant subspace. It is the smallest T-invariant subspace of V that contains the vector v.

3.4.4. Exercise. Let $G \subseteq V$ be a subset and $T\colon V \to V$ be an operator. Describe the smallest T-invariant subspace of V that contains G.

3.4.5. Example. Let $V = \mathbb{F}[X]$ be the vector space of polynomials (which is not finite dimensional, of course). Let $T\colon V \to V$ be the derivative of polynomials. Let $d \in \mathbb{N}$ and $W = \left\{ P \in V \,\middle|\, \deg(P) \le d \right\}$. Then W is T-invariant. This W is however not S-invariant if $S\colon V \to V$ is defined by $S(P) = X \cdot P$.

We can relate the characteristic polynomial of T and the characteristic polynomial of the restriction $T_{|_W}$ to a T-invariant subspace W as follows.

3.4.6. Proposition. *Let V be a finite-dimensional $\mathbb{F}$-vector space and $T\colon V \to V$ be an operator. Let $W \subseteq V$ be a T-invariant subspace and $T_{|_W}\colon W \to W$ the restriction. Then the characteristic polynomial of $T_{|_W}$ divides the characteristic polynomial of T.*

Proof. Let $n = \dim(V)$ and $d = \dim(W) \le n$. Set $e := n - d$. Choose a basis C of W and complete it into a basis B of V. (We order B by putting the vectors of C first.) In the ordered basis B, the matrix of T has the following form

$$[T]_B = \left(\begin{smallmatrix} A' & A'' \\ 0 & A''' \end{smallmatrix}\right)$$

that is, it is upper-triangular by blocks, where $A' \in \mathrm{M}_{d\times d}(\mathbb{F})$, and $A'' \in \mathrm{M}_{e\times d}(\mathbb{F})$ and $A''' \in \mathrm{M}_{e\times e}(\mathbb{F})$. Indeed, when we compute the j-th column of $[T]_B = [T]_{B,B}$ for $j \le d$ then we consider $b_j \in B$ with $j \le d$ which is in C hence in W. Thus by T-invariance of W, we have $T(b_j) \in W = \operatorname{Span}(b_1, \ldots, b_d)$. So the coordinates of $T(b_j)$ in the basis $B = \{b_1, \ldots, b_d, b_{d+1}, \ldots, b_n\}$ are all zero beyond the d-th one. Note in particular that $A' = [T_{|_W}]_C$ is the matrix of $T_{|_W}\colon W \to W$ with respect to the basis C of W. Subtracting $X \cdot I_n = \left(\begin{smallmatrix} X\cdot I_d & 0 \\ 0 & X\cdot I_e \end{smallmatrix}\right)$ to the above matrix and computing the determinant, we get

$$\begin{aligned} P_T(X) = \det([T]_B - X\cdot I_n) &= \det\left(\left(\begin{smallmatrix} A'-X\cdot I_d & A'' \\ 0 & A'''-X\cdot I_e \end{smallmatrix}\right)\right) \\ &= \det(A' - X\cdot I_d)\cdot\det(A''' - X\cdot I_e) = P_{A'}(X)\cdot P_{A'''}(X). \end{aligned}$$

See Exercise D.2.8. Since $A' = [T_{|_W}]_C$, the characteristic polynomial $P_{A'}$ of $T_{|_W}$ divides that of T as was to be shown. □

Let us describe the characteristic polynomial of $T_{|W}$ that appears above, in the case where W is spanned by a single vector, as in Example 3.4.3.

3.4.7. Lemma. *Let $T\colon V \to V$ be an operator on a finite-dimensional V. Let $v \in V$ be non-zero and let $W = \operatorname{Span}(v, T(v), \ldots, T^i(v), \ldots)$ be the T-invariant subspace generated by v. Let $d = \dim(W) \leq \dim(V)$. Then we have:*

(1) *The vectors $v, T(v), \ldots, T^{d-1}(v)$ form a basis of W.*
(2) *If $a_0, \ldots, a_{d-1} \in \mathbb{F}$ are the coordinates of $T^d(v)$ in the basis $v, T(v), \ldots, T^{d-1}(v)$ of (1), then the characteristic polynomial of the operator $T_{|W}\colon W \to W$ is $(-1)^d \cdot (X^d - a_{d-1}X^{d-1} - \cdots - a_1X - a_0)$.*

Proof. Since V is finite dimensional, we cannot have infinitely many linearly independent vectors in the list $v, T(v), \ldots, T^i(v), \ldots$. So there is a largest integer $k \geq 0$ such that the first k vectors $v, T(v), \ldots, T^k(v)$ are linearly independent. Then $T^{k+1}(v) \in \operatorname{Span}(v, T(v), \ldots, T^k(v))$ otherwise $v, T(v), \ldots, T^{k+1}(v)$ would be linearly independent (Lemma 1.6.12) contradicting the maximality of k. Then by induction on $i \geq 1$ it is easy to see that $T^{k+i}(v) \in \operatorname{Span}(v, T(v), \ldots, T^k(v))$. It follows that $\operatorname{Span}(v, T(v), \ldots, T^k(v)) = W$ and since the $k+1$ vectors $v, T(v), \ldots, T^k(v)$ are linearly independent, they form a basis of W. Hence $k = d-1$ as claimed in (1).

For (2), the assumption means $T^d(v) = a_0 \cdot v + a_1T^1(v) + \cdots + a_{d-1}T^{d-1}(v)$. Write the matrix of T in the basis $B = \{v, \ldots, T^{d-1}(v)\}$. It is

$$[T]_B = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 & a_0 \\ 1 & 0 & \cdots & 0 & 0 & a_1 \\ 0 & 1 & \ddots & 0 & 0 & a_2 \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & 0 & & 1 & 0 & a_{d-2} \\ 0 & 0 & \cdots & 0 & 1 & a_{d-1} \end{pmatrix} \mathrm{M}_{d\times d}(\mathbb{F}).$$

The characteristic polynomial of this matrix is the determinant of

$$[T]_B - X \cdot I_n = \begin{pmatrix} -X & 0 & \cdots & 0 & 0 & a_0 \\ 1 & -X & \cdots & 0 & 0 & a_1 \\ 0 & 1 & \ddots & 0 & 0 & a_2 \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & 0 & & 1 & -X & a_{d-2} \\ 0 & 0 & \cdots & 0 & 1 & a_{d-1} - X \end{pmatrix}.$$

Expanding it along the last column (the d-th one) and using that the determinant of a triangular matrix is the product of the diagonal elements, we get

$$\begin{aligned} P_T(X) &= \sum_{i=1}^{d-1} (-1)^{i+d} a_{i-1} (-X)^{i-1} + (a_{d-1} - X) \cdot (-X)^{d-1} \\ &= (-1)^d (X^d - a_{d-1}X^{d-1} - \cdots - a_1X - a_0) \end{aligned}$$

as claimed. □

We are ready to prove our goal in this section:

3.4.8. Theorem (Cayley-Hamilton)**.** *Let $T\colon V \to V$ be a linear operator on a finite-dimensional $\mathbb{F}$-vector space V. Let $P_T(X) \in \mathbb{F}[X]$ be the characteristic polynomial of T. Then $P_T(T) = 0$ in $\mathrm{Lin}(V, V)$.*

Proof. Let $v \in V$. We need to prove that $(P_T(T))(v) = 0$. If $v = 0$, that is clear. So we can assume that $v \neq 0$. Consider $W = \mathrm{Span}(v, T(v), \ldots)$ the T-invariant subspace of V generated by v as in Lemma 3.4.7. Adopt the notation of that lemma, so $d = \dim(W)$ and $T^d(v) = a_0 \cdot v + a_1 \cdot T(v) + \cdots + a_{d-1}T^{d-1}(v)$. To lighten the notation, let $S = T_{|_W} : W \to W$ the operator T restricted to W. We have seen in Lemma 3.4.7 that $P_S(X) = (-1)^d(X^d - a_{d-1}X^{d-1} - \cdots - a_1X - a_0)$. Note right away that $P_S(T)\colon W \to W$ vanishes on v by construction:

$$(P_S(T))(v) = (-1)^d\big(T^d(v) - a_{d-1}T^{d-1}(v) - \cdots - a_1T(v) - a_0v\big) = 0$$

since $T^d(v) = a_0v + a_1T(V) + \cdots + a_{d-1}T^{d-1}(v)$. On the other hand, we saw in Proposition 3.4.6 that $P_T(X) = Q(X) \cdot P_S(X)$ for some polynomial $Q \in \mathbb{F}[X]$. Then we use that $(Q \cdot P)(T) = Q(T) \circ P(T)$ in $\mathrm{Lin}(V, V)$ to compute in V

$$(P_T(T))(v) = (Q(T) \circ P_S(T))(v) = (Q(T))\big((P_S(T))(v)\big) = (Q(T))(0) = 0$$

using the already established $(P_S(T))(v) = 0$. □

3.4.9. Corollary. *Let $A \in \mathrm{M}_{n\times n}(\mathbb{F})$ and $P_A(X)$ its characteristic polynomial. Then the matrix $P_A(A)$ is zero in $\mathrm{M}_{n\times n}(\mathbb{F})$.*

Proof. Apply Theorem 3.4.8 to $T = T_A$. □

3.5. Nilpotent operators*

This section is of a slightly technical nature and constitutes a preparation for the Jordan canonical form in the next section. Throughout, all $\mathbb{F}$-vector spaces V are finite-dimensional.

3.5.1. Definition. Let $T\colon V \to V$ be a linear operator. We say that T is *nilpotent* if there exists $n \in \mathbb{N}$ large enough such that $T^n = 0$. For brevity we shall call a *nil-space* a pair (V, T) where V is a finite-dimensional $\mathbb{F}$-vector space and $T\colon V \to V$ is a nilpotent operator. (This is a local terminology only!)

3.5.2. Example. Let $V = \mathbb{F}^n$ and $T = T_A$ for $A = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & & 1 & 0 \end{pmatrix}$ the square matrix of size n with 1's just below the diagonal and zero everywhere else. In other words, $T(e_1) = e_2$, …, $T(e_j) = e_{j+1}$,…, $T(e_{n-1}) = e_n$ and $T(e_n) = 0$. So clearly $T^n(e_j) = 0$ for all j and therefore $T^n = 0$. One can also check that A^i has 1's on a 'small diagonal' i places below *the* diagonal, and eventually $A^n = 0$. (As an exercise, take $n = 4$ for instance and compute $A^2, A^3, \ldots$)

3.5.3. Definition. A *cyclic nil-space* will be one as in the easy example above: So $(V, T\colon V \to V)$ is a cyclic nilspace if $T^n = 0$ and there exists $v \in V$ such that $v, T(v), \ldots, T^{n-1}(v)$ form a basis of V. Indeed, in that basis the matrix of T is exactly the matrix A of Example 3.5.2.

Our goal is to show that every nil-space is just a direct sum of cyclic ones. Let us clarify what the direct sum of nil-spaces means.

3.5.4. Remark. Let $(V_1, T_1), \ldots, (V_r, T_r)$ be nil-spaces, *i.e.* each V_i is an $\mathbb{F}$-vector space and each $T_i \colon V_i \to V_i$ is a nilpotent operator, say $T_i^{n_i} = 0$. Define $V = V_1 \oplus \cdots \oplus V_r$ as in Remark 1.2.26, that is simply the cartesian product and define $T \colon V \to V$ by mapping $(v_1, \ldots, v_r)$ to $(T_1(v_1), \ldots, T_r(v_r))$. In other words, T is defined as T_i in the i-th component. This defines a nilpotent operator since $T^n = 0$ for any $n \geq \max\{n_1, \ldots, n_r\}$. In matrix form, if B_i is a basis of V_i and $A_i = [T_i]_{B_i}$ is the matrix of T_i in that basis then the matrix of T is diagonal by blocks

$$[T]_B = \begin{pmatrix} A_1 & 0 & & 0 \\ 0 & A_2 & & \\ & & \ddots & \\ 0 & & & A_r \end{pmatrix} \tag{3.5.5}$$

in the basis B of V obtained from putting the bases $B_1, \ldots, B_r$ 'together' (where $b_1 \in B_1$ is viewed as $(b_1, 0, \ldots, 0)$ in V, where $b_2 \in B_2$ is viewed as $(0, b_2, 0, \ldots 0)$, etc, as in the proof of $\dim(V_1 \oplus \cdots \oplus V_r) = \dim(V_1) + \cdots + \dim(V_r)$).

3.5.6. Remark. Nil-spaces (V, T) and (V', T') can be called 'isomorphic' if the following happens: Suppose that $S \colon V \xrightarrow{\sim} V'$ is an isomorphism and that $T' = S \circ T \circ S^{-1}$, that is, T' is T 'transported' to V' along the isomorphism $V \xrightarrow{\sim} V'$

$$\begin{tikzcd} V \arrow[r, "\cong", "S"'] \arrow[d, "T"'] & V' \arrow[d, "T'"] \\ V \arrow[r, "\cong", "S"'] & V'. \end{tikzcd}$$

In this situation, properties of (V, T) pass to an isomorphic (V', T') in the above sense. For instance, if (V, T) is cyclic, generated by v (Definition 3.5.3) then (V', T') is also cyclic, generated by $v' = S(v)$. Similarly, if $(V, T) = (V_1, T_1) \oplus \cdots \oplus (V_r, T_r)$ is a direct sum as in Remark 3.5.4 then we can transport this decomposition up to isomorphism: $(V', T') = (V'_1, T'_1) \oplus \cdots \oplus (V'_r, T'_r)$ where $V'_i = S(V_i)$ and $T'_i = S \circ T_i \circ S^{-1}$. Combining those two remarks, we see that if (V, T) is a direct sum of cyclic nil-spaces, then so is the isomorphic (V', T').

3.5.7. Proposition. *Let $T \colon V \to V$ be a nilpotent operator, say $T^N = 0$ for $N \geq 1$ large enough. Then the characteristic polynomial of T is $(-1)^n X^n$ where $n = \dim(V)$. In particular, T has only zero as eigenvalue and $T^n = 0$.*

Proof. We can do this in matrix form (with respect to any basis of V). Let $A \in \mathrm{M}_{n\times n}(\mathbb{F})$ such that $A^N = 0$ for $N \geq 1$ large enough and let us show that $P_A(X) = (-1)^n X^n$. The consequences follow since 0 is the only root of X^n and $A^n = 0$ follows from Cayley-Hamilton Theorem 3.4.8. We can assume that $\mathbb{F}$ is algebraically closed, as $\mathrm{M}_{n\times n}(\mathbb{F}) \subseteq \mathrm{M}_{n\times n}(\bar{\mathbb{F}})$ preserves the characteristic polynomial, for any algebraic closure $\mathbb{F} \subseteq \bar{\mathbb{F}}$ (like for instance under $\mathbb{R} \subset \mathbb{C}$). In the case where $\mathbb{F}$ is algebraically closed, $P_A(X)$ is completely split $P_A(X) = (-1)^n (X - \lambda_1) \cdots (X - \lambda_n)$ and it suffices to show that all the eigenvalues λ_i are zero. Let $\lambda \in \mathbb{F}$ such that $A \cdot x = \lambda x$ for some $x \in \mathbb{F}^n$ non-zero. Then by induction on $i \geq 0$ we have $A^i \cdot x = \lambda^i \cdot x$ for all $i \in \mathbb{N}$. Since $A^N = 0$ we have $\lambda^N \cdot x = 0$ and $x \neq 0$ in V. Hence $\lambda^N = 0$ in $\mathbb{F}$ and therefore $\lambda = 0$ since $\mathbb{F}$ is a field. So $P_A(X) = (-1)^n (X - \lambda_1) \cdots (X - \lambda_n) = (-1)^n X^n$ as claimed. □

3.5.8. Corollary. *Let $T \colon V \to V$ be nilpotent and $v \in V$ be a non-zero vector that generates V as T-invariant subspace, that is, $V = \mathrm{Span}(v, T(v), T^2(v), \ldots)$. Let $n = \dim(V)$. Then $T^n(v) = 0$ and $v, T(v), \ldots, T^{n-1}(v)$ is a basis of V.*

Proof. By Lemma 3.4.7 (1) the vectors $v, \dots, T^{n-1}(v)$ form a basis of V and by Proposition 3.5.7 we have $T^n = 0$ and in particular $T^n(v) = 0$. □

3.5.9. Theorem. *Let V be a finite-dimensional $\mathbb{F}$-vector space and $T\colon V \to V$ a nilpotent operator (i.e. (V, T) is a 'nil-space'). Then V is a direct sum of T-invariant subspaces $V_1, \dots, V_r$ each of which is cyclic as a nil-space $(V_i, T_{|V_i})$ in the sense of Definition 3.5.3. In particular, V admits a basis B in which the matrix of T is diagonal by blocks as in (3.5.5), with each A_i having only 1's below the diagonal and zero elsewhere, as in Example 3.5.2.*

Proof. We proceed by induction on $\dim(V)$. If $\dim(V) = 1$ there is hardly anything to prove: A nilpotent T must be zero in that case. So we suppose that we know the result for all nil-spaces (V', T') of dimension strictly less than $\dim(V)$.

For each non-zero $v \in V$, there is a smallest $k \geq 1$ such that $T^k(v) = 0$. Let us call this k the *order* of v. (It should be the order of nilpotence of T at v but well, life is short.) If we have $T^n = 0$ then the order of every v is at most n. In particular, we can pick a non-zero $w \in V$ of *maximal order* d in V. By Corollary 3.5.8, this d is also $\dim(W)$ where W is the T-invariant subspace generated by w

$$W := \operatorname{Span}(w, T(w), \dots, T^{d-1}(w)). \tag{3.5.10}$$

So far we have created a T-invariant subspace $W \subseteq V$ of our nil-space (V, T) and $(W, T_{|W})$ is a cyclic nil-space by construction, generated by $w \in V$ of order $d = \dim(W)$ maximal among all the orders of non-zero vectors in V. If $W = V$ we are done: V is itself cyclic. Otherwise $1 \leq \dim(W) < \dim(V)$ and $\dim(V/W) = \dim(V) - \dim(W)$ is strictly less than $\dim(V)$. We claim that the quotient V/W is a nil-space in a natural way. The following picture could help some readers:

$$\begin{tikzcd}
W \arrow[r, hook, "\text{incl}"] \arrow[d, "T_{|W}"'] & V \arrow[r, two heads, "\text{proj}"] \arrow[d, "T"] & V/W \arrow[d, dotted, "\bar{T}"] \\
W \arrow[r, hook, "\text{incl}"'] & V \arrow[r, two heads, "\text{proj}"'] & V/W
\end{tikzcd}$$

The operator $\bar{T}\colon V/W \to V/W$ is simply defined by $\bar{T}([x]_W) = [T(x)]_W$. In other words, $\bar{T} \circ \operatorname{proj} = \operatorname{proj} \circ T$. This is well-defined since $T(W) \subseteq W$ so when $[x] = [y]$, that is, when x and y are two representatives of the same class, then $x - y \in W$ and therefore $T(x) - T(y) = T(x - y) \in T(W) \subseteq W$ and thus $[T(x)] = [T(y)]$. We clearly have $\bar{T}^n = 0$ since $\bar{T}^n([x]) = [T^n(x)] = [0] = 0$ for all $[x] \in V/W$.

We can therefore use the statement for the nil-space $(V/W, \bar{T})$ by induction hypothesis. Hence we have $(V/W, \bar{T}) = (\bar{V}_1, \bar{T}_1) \oplus \cdots \oplus (\bar{V}_r, \bar{T}_r)$ for $\bar{T}$-invariant subspaces $\bar{V}_i \subseteq V/W$ which are moreover cyclic when viewed as nil-spaces for the restricted operator $\bar{T}_i := \bar{T}_{|\bar{V}_i}$. We now claim the following:

Claim: There exists a linear transformation $S\colon V/W \to V$ such that $\operatorname{proj} \circ S = \operatorname{Id}_{V/W}$ *and such that* $T \circ S = S \circ \bar{T}$.

Since $V/W = \bar{V}_1 \oplus \cdots \oplus \bar{V}_r$ it suffices to construct $S_i\colon \bar{V}_i \to V$ one subspace at a time in such a way that $\operatorname{proj} \circ S_i = \operatorname{Id}_{\bar{V}_i}$ and $T \circ S_i = S_i \circ \bar{T}_i$. Indeed, we can then define $S\colon V/W = \bar{V}_1 \oplus \cdots \oplus \bar{V}_r \to V$ by setting $S(\bar{v}_1, \dots, \bar{v}_r) = S_1(\bar{v}_1) + \cdots + S_r(\bar{v}_r)$ and check the claim for this S.

Fix one $i \in \{1, \ldots, r\}$ for a moment. To construct our $S_i \colon \bar{V}_i \to V$ we are going to use that $(\bar{V}_i, \bar{T}_i)$ is cyclic. Let $\bar{v} \in \bar{V}_i$ be a generator, so that if we set $\ell := \dim(\bar{V}_i) \geq 1$ then $\bar{v}$ has order ℓ and $\bar{V}_i$ has basis

$$\bar{v}, \bar{T}(\bar{v}), \ldots, \bar{T}^{\ell-1}(\bar{v}) \tag{3.5.11}$$

by Corollary 3.5.8. To construct $S_i \colon \bar{V}_i \to V$ it suffices to construct S_i on the above basis elements. The condition $\operatorname{proj} \circ S_i = \mathrm{Id}$ means that $S_i(\bar{v})$ must be some representative of the class $\bar{v} \in V/W$, and similarly for all other basis vectors $\bar{T}^j(\bar{v})$. But we also want $S_i \circ \bar{T} = T \circ S_i$ and therefore $S_i \circ \bar{T}^j = T^j \circ S_i$ for all j. In other words, once we decide what the image $v := S_i(\bar{v})$ of $\bar{v}$ should be, then all the images of all $\bar{T}^i(\bar{v})$ are forced: Each $\bar{T}^j(\bar{v})$ must be mapped to $T^j(v)$:

$$\begin{tikzcd}
\bar{v} \arrow[r, mapsto, bend left, "\bar{T}"] \arrow[d, mapsto, "S_i"'] & \bar{T}(\bar{v}) \arrow[r, mapsto, bend left, "\bar{T}"] \arrow[d, mapsto, "S_i"'] & \bar{T}^2(\bar{v}) \arrow[r, mapsto, bend left, "\bar{T}"] \arrow[d, mapsto, "S_i"'] & \cdots \arrow[r, mapsto, bend left, "\bar{T}"] & \bar{T}^{\ell-1}(\bar{v}) \arrow[r, mapsto, bend left, "\bar{T}"] \arrow[d, mapsto, "S_i"'] & \bar{T}^{\ell}(\bar{v}) = 0 \arrow[d, mapsto, "S_i"'] \\
v \arrow[r, mapsto, bend right, "T"'] & T(v) \arrow[r, mapsto, bend right, "T"'] & T^2(v) \arrow[r, mapsto, bend right, "T"'] & \cdots \arrow[r, mapsto, bend right, "T"'] & T^{\ell-1}(v) \arrow[r, mapsto, bend right, "T"'] & T^{\ell}(v) \underset{?}{=} 0 \\
\text{chosen} \arrow[u, dotted] & \text{forced} \arrow[u, dotted] & \text{forced} \arrow[u, dotted] & \cdots & \text{forced} \arrow[u, dotted] & \text{oopsy-daisy} \arrow[u, dotted]
\end{tikzcd}$$

All this is well and would yield $S_i \colon \bar{V}_i \to V$ such that $\operatorname{proj} \circ S_i = \mathrm{Id}$ and almost such that $T \circ S_i = S_i \circ \bar{T}$ *except* that for the very last basis vector $b = \bar{T}^{\ell-1}(\bar{v})$ in (3.5.11), we have $S_i(b) = T^{\ell-1}(v)$ by definition and we have $\bar{T}(b) = \bar{T}^{\ell}(\bar{v}) = 0$ and therefore $S_i \circ \bar{T}(b) = 0$. However, it is not clear why $T \circ S_i(b)$ should be zero (as indicated by the question mark '?' above). Indeed $T(S_i(b)) = T(T^{\ell-1}(v)) = T^{\ell}(v)$ and we have no guarantee that the chosen representative v of $\bar{v}$ satisfies $T^{\ell}(v) = 0$.

Let us start with an imperfect $v \in V$ such that $\bar{v} = [v]$ and let us 'correct' it. We have $0 = \bar{T}^{\ell}(\bar{v}) = [T^{\ell}(v)]$, which means that $T^{\ell}(v) \in W = \operatorname{Span}(w, \ldots, T^{d-1}(w))$. See (3.5.10). We can therefore find $a_0, \ldots, a_{d-1} \in \mathbb{F}$ such that

$$T^{\ell}(v) = a_0 w + a_1 T(w) + \cdots + a_{d-1} T^{d-1}(w). \tag{3.5.12}$$

Let k be the order of v in V. Since $\bar{T}^k(\bar{v}) = [T^k(v)] = 0$ and since $\bar{v}$ has order ℓ, we see that $k \geq \ell$. Since d was the maximal order in V, we also have $k \leq d$. In short we have $\ell \leq k \leq d$. Applying $T^{d-\ell}$ to (3.5.12) we get

$$T^d(v) = \sum_{i=0}^{d-1} a_i T^{d-\ell+i}(w).$$

Using that $T^d(v) = 0$ since $d \geq k$ and $T^j(w) = 0$ for all $j \geq d$, many terms become zero above (namely $T^{d-\ell+i}(w) = 0$ if $d - \ell + i \geq d$, that is, $i \geq \ell$) we are left with

$$0 = \sum_{i=0}^{\ell-1} a_i T^{d-\ell+i}(w).$$

But in that range, $j \in \{d-\ell, \ldots, d-1\}$ the $T^j(w)$ are linearly independent. So we have established

$$a_0 = \ldots = a_{\ell-1} = 0.$$

Since $\ell \geq 1$ we have at least some information here! We can rewrite (3.5.12) as

$$T^\ell(v) = a_\ell T^\ell(w) + a_{\ell+1}T^{\ell+1}(w) + \cdots + a_{d-1}T^{d-1}(w).$$

By linearity, this means that $T^\ell(v') = 0$ where

$$v' = v - \big(a_\ell w + a_{\ell+1}T(w) + \cdots + a_{d-1}T^{d-\ell-1}(w)\big).$$

Note that this v' is just v modified by a vector of W. In particular $[v'] = [v] = \bar{v}$ is our generator of $\bar{V}_i$. In other words, v' is as good as v but satisfies in addition $T^\ell(v') = 0$. By the above discussion, we can now construct $S_i\colon \bar{V}_i \to V$ by sending the basis vectors (2.4.22) to $T(v'), \ldots, T^{\ell-1}(v')$ respectively.

At this stage, we have proved the Claim. So we have $S\colon V/W \to V$ a section of $\mathrm{proj}\colon V \to V/W$ (meaning $\mathrm{proj} \circ S = \mathrm{Id}_{V/W}$), which is 'invariant' with respect to the nilpotent operators, meaning $T \circ S = S \circ \bar{T}$. It follows that S is injective hence yields an isomorphism $S\colon V/W \overset{\sim}{\to} \mathrm{Im}(S)$ onto its image in V; its inverse is given by $\mathrm{proj}\colon \mathrm{Im}(S) \to V/W$. Since $T \circ S = S \circ \bar{T}$, it follows that $\mathrm{Im}(S)$ is a T-invariant subspace of V and that S gives an isomorphism of nil-spaces as in Remark 3.5.6 between $(V/W, \bar{T})$ and $(\mathrm{Im}(S), T_{|_{\mathrm{Im}(S)}})$. In particular, that image is a sum of cyclic nil-spaces, since V/W was. As vector spaces, we have $V = W \oplus \mathrm{Im}(S)$ (Remark 2.7.12) and this is a sum of nil-spaces by the T-invariance of each subspace. The nil-space $(W, T_{|_W})$ is cyclic by construction and $(\mathrm{Im}(S), T_{|_{\mathrm{Im}(S)}})$ is a sum of cyclic nil-spaces. Hence their sum V is again a sum of cyclic nil-spaces. □

3.6. Jordan canonical form*

As before, throughout this section V is a finite-dimensional $\mathbb{F}$-vector space. Let us start with an independently useful result.

3.6.1. Lemma (Fitting's Lemma)**.** *Let $S\colon V \to V$ be a linear operator, where V is finite-dimensional. Then there exists (canonical) S-invariant subspaces W_1 and W_2 such that $V = W_1 \oplus W_2$ and $S_{|_{W_1}}\colon W_1 \to W_1$ is nilpotent, namely $S^d_{|_{W_1}} = 0$ for $d = \dim(W_1)$, and where $S_{|_{W_2}}\colon W_2 \overset{\sim}{\to} W_2$ is an isomorphism. Hence there is a basis B of V in which $[S]_B = \left(\begin{smallmatrix} A_1 & 0 \\ 0 & A_2 \end{smallmatrix}\right)$ is diagonal by blocks, with A_1 a nilpotent matrix and A_2 an invertible matrix.*

Proof. Consider the two towers of subspaces of V:

$$\mathrm{Ker}(S) \subseteq \mathrm{Ker}(S^2) \subseteq \cdots \subseteq \mathrm{Ker}(S^i) \subseteq \mathrm{Ker}(S^{i+1}) \subseteq \cdots \subseteq V$$

(indeed if $S^i(x) = 0$ then $S^{i+1}(x) = S(S^i(x)) = S(0) = 0$) and

$$V \supseteq \mathrm{Im}(S) \supseteq \mathrm{Im}(S^2) \supseteq \cdots \supseteq \mathrm{Im}(S^i) \supseteq \mathrm{Im}(S^{i+1}) \supseteq \cdots$$

(indeed if $x = S^{i+1}(y)$ for some y then $x = S^i(S(y))$). At each step in either tower, a strict inclusion means the dimension of the smaller subspace is at least one less than the dimension of the bigger one. Since $\dim(V)$ is finite there can only be *finitely many* such strict inclusions (namely, at most $\dim(V)$ strict inclusions). In other words, beyond finitely many steps in either tower, these $\subseteq$ and $\supseteq$ are all equalities. Hence there exists $\ell \geq 1$ large enough so that

$$\mathrm{Ker}(S^\ell) = \mathrm{Ker}(S^m) \qquad \text{and} \qquad \mathrm{Im}(S^\ell) = \mathrm{Im}(S^m) \tag{3.6.2}$$

for all $m \geq \ell$. We are going to use this for $m = 2\ell$ mostly. We claim that $W_1 = \mathrm{Ker}(S^\ell)$ and $W_2 = \mathrm{Im}(S^\ell)$ satisfy the statement. It is easy to check that they are S-invariant. Let us see why $W_1 \cap W_2 = 0$. Suppose that $x \in \operatorname{Ker}(S^\ell) \cap \operatorname{Im}(S^\ell)$. Then $x = S^\ell(y)$ for some $y \in V$ and $0 = S^\ell(x) = S^{2\ell}(y)$; so we just proved $y \in \operatorname{Ker}(S^{2\ell}) = \operatorname{Ker}(S^\ell)$ by (3.6.2) and therefore $0 = S^\ell(y) = x$. To see that $W_1 + W_2 = V$, pick $x \in V$ and consider $S^\ell(x) \in \operatorname{Im}(S^\ell) = \operatorname{Im}(S^{2\ell})$; this means that $S^\ell(x) = S^{2\ell}(z)$ for some $z \in V$. But then $x = w_1 + w_2$ where $w_1 = (x - S^\ell(z))$ and $w_2 = S^\ell(z)$ (duh!) and we claim that $w_i \in W_i$ for $i = 1, 2$. The second is obvious $w_2 = S^\ell(z) \in \operatorname{Im}(S^\ell) =: W_2$. The first one is easy too: $S^\ell(w_1) = S^\ell(x) - S^{2\ell}(z) = 0$ by construction of z. □

3.6.3. Corollary. *Let $T\colon V \to V$ be a linear operator on a finite-dimensional $\mathbb{F}$-vector space. Suppose that the characteristic polynomial of T is completely split, say $P_T(X) = (-1)^n(X-\lambda_1)^{m_1}\cdots(X-\lambda_r)^{m_r}$ for r distinct eigenvalues $\lambda_1, \ldots, \lambda_r \in \mathbb{F}$ with multiplicities $m_1, \ldots, m_r \geq 1$. For every $j = 1, \ldots, r$ let*

$$W_j = \left\{ x \in V \,\middle|\, \textit{there exists } \ell \geq 1 \textit{ with } (T - \lambda_j \operatorname{Id}_V)^\ell(x) = 0 \right\}.$$

the subspace of vectors on which $T - \lambda_j \operatorname{Id}_V$ is nilpotent. Then every W_j is T-invariant and $\dim(W_j) = m_j$ and $V = W_1 \oplus \cdots \oplus W_r$. In particular, there exists a basis B of V such that $[T]_B = \begin{pmatrix} A_1 & & & \\ & A_2 & 0 & \\ & 0 & \ddots & \\ & & & A_r \end{pmatrix}$ is diagonal by block with each block $A_j \in \mathrm{M}_{m_j \times m_j}(\mathbb{F})$ is such that $A_j - \lambda_j I_{m_j}$ is nilpotent.

Proof. By induction on r. If $r = 1$ then $P_T(X) = (-1)^n(X - \lambda_1)^n$ and we can take $W_1 = V$ by Cayley Hamilton Theorem 3.4.8. Suppose $r \geq 2$ and that we know the result for $r - 1$.

By Fitting's Lemma 3.6.1 applied to $S = T - \lambda_1 \operatorname{Id}_V$, we know that $V = W_1 \oplus W'$ where W_1 and W' are S-invariant and therefore T-invariant (as $T = S + \lambda_1 \operatorname{Id}_V$ and every subspace is $(\lambda_1 \operatorname{Id}_V)$-invariant). If we build a basis C by assembling a basis of W_1 and one of W', the matrix of T in that basis will be diagonal by block $[T]_C = \left(\begin{smallmatrix} A_1 & 0 \\ 0 & A' \end{smallmatrix}\right)$ with $A_1 - \lambda_1 I_d$ nilpotent and $A' - \lambda_1 I_e$ invertible, where $d = \dim(W_1)$ and $e = n - d = \dim(W')$. We want to see that $d = m_1$ as predicted by the characteristic polynomial. Indeed, we have $P_T(X) = P_1(X) \cdot P'(X)$ where $P_1 = P_{A_1} = P_{T_{|W}}$ and $P' = P_{A'} = P_{(T_{|W'})}$ (*not* the derivative). Since $A_1 - \lambda_1 I_d$ is nilpotent, we know that $P_1(X) = P_{A_1}(X) = (-1)^d(X - \lambda_1)^d$ (by Proposition 3.5.7). Therefore $P' = (-1)^e(X - \lambda_1)^{m_1 - d}(X - \lambda_2)^{m_2}\cdots(X - \lambda_r)^{m_r}$ and to show that $d = m_1$ it suffices to check that λ_1 cannot be a root of $P' = P_{A'}$. Indeed, if there was a non-zero $x \in W'$ such that $T_{|W'}(x) = \lambda_1 x$ then $S(x) = 0$; but S is invertible on W' so $x = 0$ a contradiction. So $d = m_1$ as announced. Therefore $T' = T_{|W'}\colon W' \to W'$ has a completely split polynomial $P'(X) = (-1)^e(X - \lambda_2)^{m_2}\cdots(X - \lambda_r)^{m_r}$ with $r - 1$ distinct roots. We conclude by induction hypothesis applied to W' and T'. □

The main result is now an easy corollary.

3.6.4. Theorem (Jordan Canonical Form)**.** *Let $T\colon V \to V$ be a linear operator on a finite-dimensional $\mathbb{F}$-vector space. Suppose that the characteristic polynomial of T is completely split (for instance, if $\mathbb{F}$ is algebraically closed, like $\mathbb{F} = \mathbb{C}$). Say $P_T(X) = (-1)^n(X-\lambda_1)^{m_1}\cdots(X-\lambda_r)^{m_r}$ for r distinct eigenvalues $\lambda_1, \ldots, \lambda_r \in \mathbb{F}$ with algebraic multiplicities $m_1, \ldots, m_r \geq 1$. Then V admits a basis B in which* *the matrix of T is diagonal-by-blocks*

$$[T]_B = \begin{pmatrix} A_1 & 0 & & 0 \\ 0 & A_2 & & \\ & & \ddots & \\ 0 & & & A_r \end{pmatrix}$$

and each matrix A_j is itself diagonal by blocks

$$A_j = \begin{pmatrix} A_{j,1} & 0 & & 0 \\ 0 & A_{j,2} & & \\ & & \ddots & \\ 0 & & & A_{j,s_j} \end{pmatrix}$$

with diagonal blocks of the form $A_{j,k} = \begin{pmatrix} \lambda_j & 0 & \cdots & & 0 \\ 1 & \lambda_j & 0 & \cdots & 0 \\ & & \ddots & \ddots & \\ 0 & & & 1 & \lambda_j \end{pmatrix}$. *Moreover, all this data is unique up to permutation of the $\lambda_1, \ldots, \lambda_r$ and permutation of the blocks.*

Proof. We have seen in Corollary 3.6.3 that $V = W_1 \oplus \cdots \oplus W_r$ for T-invariant subspaces W_j on which $S_j := T_{|W_j} - \lambda_j \operatorname{Id}_{W_j}$ is nilpotent. We can then apply Theorem 3.5.9 to find a basis of W_j in which S_j is diagonal by block, with blocks as in Example 3.5.2. In that basis, the matrix of $T_{|W_j} = S_j + \lambda \operatorname{Id}_{W_j}$ is the one for S_j plus λ_j times the (small) identity matrix (changing only the diagonal). The result is a matrix like A_j of the theorem. Hence the result.

Uniqueness is not too hard but we leave it as an exercise. The eigenvalues $\lambda_1, \ldots, \lambda_r$ are unique up to renumbering them: They are the root of the characteristic polynomial. Now, if we take any basis B as in the statement, one considers the dimensions $d(j, \ell) = \operatorname{Ker}((T - \lambda_j \cdot \operatorname{Id}_V)^\ell)$ for various ℓ. These increasing numbers, as ℓ increases, force the number and the size of the matrices $A_{j,k}$ for all k to be unique. Indeed, for ℓ very large, we have $d(j, \ell) = m_j$ and as ℓ goes down, the $d(j, \ell)$ go down (or stay the same). For instance, the biggest $d(j, \ell)$ smaller than m_j differs from m_j because some of the nilpotent matrices on the diagonal of S_j, as in Example 3.5.2, are not zero. This happens for the largest such matrices only. So that $d(j, \ell)$ tells us how many $A_{j,k}$ there are with maximal size. Letting ℓ go down, the next step will reveal the $A_{j,k}$ with the next smaller size, etc, until $\ell = 1$. At that point, we know the sizes of all $A_{j,k}$ uniquely in terms of T and λ_j. Since the size of $A_{j,k}$ characterizes $A_{j,k}$ entirely, we are done. Details are left to the interested reader. □

CHAPTER 4

Inner product spaces

In the whole chapter the field $\mathbb{F}$ is either the field $\mathbb{R}$ of real numbers or the field $\mathbb{C}$ of complex numbers.

We shall write complex conjugation of a scalar $c \in \mathbb{F}$ by $\bar{c}$. So, when $\mathbb{F} = \mathbb{R}$, this simply means the identity: $\bar{c} = c$ for all $c \in \mathbb{R}$. When $\mathbb{F} = \mathbb{C} = \{ a + bi \,|\, a, b \in \mathbb{R} \}$ then $\overline{a + bi} = a - bi$. Recall that $\overline{c + d} = \bar{c} + \bar{d}$ and $\overline{c \cdot d} = \bar{c} \cdot \bar{d}$. Also, if $c = a + bi$ then $c \cdot \bar{c} = a^2 + b^2 = |c|^2$. This real number $|c|$ is often called the norm, or the modulus, or the absolute value of c. If $c \neq 0$ then $|c| \neq 0$ and $c^{-1} = \frac{\bar{c}}{|c|^2} = \frac{a-bi}{a^2+b^2}$.

4.1. Inner product

We follow the pattern of previous chapters. The main definition isolates axiomatically a situation that occurs in lots and lots of examples.

4.1.1. Definition. Let V be an $\mathbb{F}$-vector space over $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. An *inner product* (or *scalar product*) on V is a function

$$\begin{array}{rccc} \langle -, - \rangle \colon & V \times V & \longrightarrow & \mathbb{F} \\ & (x, y) & \longmapsto & \langle x, y \rangle \end{array}$$

such that the following equalities hold in $\mathbb{F}$:

(IPS 1) $\langle x + x', y \rangle = \langle x, y \rangle + \langle x', y \rangle$ for all $x, x', y \in V$.
(IPS 2) $\langle c \cdot x, y \rangle = c \cdot \langle x, y \rangle$ for all $x, y \in V$ and $c \in \mathbb{F}$.
(IPS 3) $\langle y, x \rangle = \overline{\langle x, y \rangle}$ for all $x, y \in V$.

Note that plugging $x = y$ in (IPS 3) forces $\langle x, x \rangle$ to belong to $\mathbb{R}$, even when $\mathbb{F} = \mathbb{C}$.

(IPS 4) For every $x \neq 0$ in V we have $\langle x, x \rangle > 0$.

The pair $(V, \langle -, - \rangle)$ consisting of a vector space and an inner product is called an *inner product space*. It is common to simply say that V is an inner product space, the inner product $\langle -, - \rangle$ being understood.

4.1.2. Remark. Let us discuss some immediate consequences. First note that Axioms (IPS 1) and (IPS 2) tell us that for a fixed $y \in V$, the function

$$\begin{array}{rccc} \langle -, y \rangle \colon & V & \longrightarrow & \mathbb{F} \\ & x & \longmapsto & \langle x, y \rangle \end{array}$$

is linear. We say that $\langle -, - \rangle$ is linear in the first variable. It is also linear in the second variable when $\mathbb{F} = \mathbb{R}$ but it is *not* linear in the second variable when $\mathbb{F} = \mathbb{C}$. Indeed, combining with (IPS 3) we can prove:

(IPS 1') $\langle x, y + y' \rangle = \langle x, y \rangle + \langle x, y' \rangle$ for all $x, y, y' \in V$.

Indeed $\langle x, y+y'\rangle = \overline{\langle y+y', x\rangle} = \overline{\langle y, x\rangle + \langle y', x\rangle} = \overline{\langle y, x\rangle} + \overline{\langle y', x\rangle} = \langle x, y\rangle + \langle x, y'\rangle$.

(IPS 2') $\langle x, c\cdot y\rangle = \bar{c}\cdot\langle x, y\rangle$ for all $x, y \in V$ and $c \in \mathbb{F}$.

Indeed $\langle x, c\,y\rangle = \overline{\langle c\,y, x\rangle} = \overline{c\cdot\langle y, x\rangle} = \bar{c}\cdot\overline{\langle y, x\rangle} = \bar{c}\cdot\langle x, y\rangle$.

4.1.3. Remark. By (IPS 2) we have $\langle 0, 0\rangle = 0$. So we can rewrite (IPS 4) as saying that $\langle x, x\rangle \geq 0$ for all $x \in V$ and that $\langle x, x\rangle = 0$ forces $x = 0$.

4.1.4. Definition. Let $(V, \langle -, -\rangle)$ be an inner product space. The *norm* of a vector $x \in V$ is the non-negative real number $||x|| = \sqrt{\langle x, x\rangle}$. This is well-defined by (IPS 4) and Remark 4.1.3 tells us that $||x|| = 0$ if and *only if* $x = 0$.

4.1.5. Remark. The norm is compatible with the action: for all $c \in \mathbb{F}$ and $x \in V$

$$||c\cdot x|| = |c|\cdot||x||. \tag{4.1.6}$$

Indeed, $||c\cdot x||^2 = \langle c\cdot x, c\cdot x\rangle = c\cdot\bar{c}\cdot\langle x, x\rangle = |c|^2\cdot\langle x, x\rangle = |c|^2\cdot||x||^2$ and since $|c| \geq 0$ and $\langle x, x\rangle \geq 0$ we can take the square root on both sides and get (4.1.6).

However, the compatibility with the sum is less straightforward. We shall discuss this after the examples.

4.1.7. Example. Let $n \in \mathbb{N}$. Then $\mathbb{F}^n$ admits the *standard inner product*

$$\langle \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \rangle := x_1\cdot\bar{y}_1 + \cdots + x_n\cdot\bar{y}_n = \sum_{i=1}^{n} x_i\bar{y}_i.$$

In particular, if $\mathbb{F} = \mathbb{R}$, this is the usual scalar product (a. k. a. dot product) $\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i$ for every $x, y \in \mathbb{R}^n$. The associated norm is the 'usual' euclidean norm $||x|| = \sqrt{x_1^2 + \cdots + x_n^2}$ on $\mathbb{R}^n$. On $\mathbb{C}^n$ the (complex) norm is $||z|| = \sqrt{z_1\bar{z}_1 + \cdots + z_n\bar{z}_n} = \sqrt{|z_1|^2 + \cdots + |z_n|^2}$, for every $z \in \mathbb{C}^n$.

If we identify $\mathbb{F}^n = \mathrm{M}_{n\times 1}(\mathbb{F})$ with column vectors, we can write the inner product as a matrix multiplication

$$\langle x, y\rangle = x^t\cdot\bar{y} \tag{4.1.8}$$

where $x^t = [x_1 \ \cdots \ x_n]$ is the transpose and $\bar{y} = \begin{bmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_n \end{bmatrix}$ is complex-conjugation componentwise. Alternatively, using the trace $\mathrm{tr}\colon \mathrm{M}_{n\times n}(\mathbb{F}) \to \mathbb{F}$ we can write $\langle x, y\rangle$ as

$$\langle x, y\rangle = \mathrm{tr}(x\cdot\bar{y}^t) \tag{4.1.9}$$

since $x\cdot\bar{y}^t$ is the $(n\times n)$-matrix $(x_i\cdot\bar{y}_j)_{i,j}$. The latter formula can be generalized.

4.1.10. Exercise. Let $V = \mathrm{M}_{p\times q}(\mathbb{F})$. Define $\langle -, -\rangle\colon V\times V \to \mathrm{M}_{1\times 1}(\mathbb{F}) \cong \mathbb{F}$ by

$$\langle A, B\rangle = \mathrm{tr}(A\cdot\bar{B}^t)$$

for all $A, B \in V$. Verify that this defines an inner product space.

4.1.11. Exercise. Generalize Example 4.1.7 to the free space $\mathbb{F}^{(I)}$ for every set I. (And appreciate how this would not make sense for $\mathbb{F}^I$.)

4.1.12. Example. Let $V = \mathrm{Cont}([a, b], \mathbb{F}) = \big\{\, f\colon [a, b] \to \mathbb{F} \,\big|\, f \text{ is continuous} \,\big\}$ be the $\mathbb{F}$-vector space of continuous functions on the interval $[a, b] \subset \mathbb{R}$, for $a < b$.

These are either real-valued or complex-valued continuous functions. In particular, they are integrable and we can define

$$\langle f,g\rangle = \int_a^b f(t)\cdot\overline{g(t)}\ \mathrm{d}t$$

Note how the sum is replaced by an integral. The verification of (IPS 1)–(IPS 4) uses standard properties of the integral. For instance, for (IPS 4), if $f\colon [a,b]\to\mathbb{F}$ is a non-zero function then $f\cdot\bar{f}\colon [a,b]\to\mathbb{R}$ is everywhere non-negative and strictly positive at those $t\in[a,b]$ where $f(t)\neq 0$. It follows that $\int_a^b f(t)\cdot\bar{f}(t)\ \mathrm{d}t > 0$.

So far, we have reviewed the standard vector spaces of Chapter 1, free ones, vector spaces of matrices, vector spaces of functions, and seen how they typically admit an inner product space. Another source of vector spaces in Chapter 1 was the notion of subspace. The interplay with inner product is very easy:

4.1.13. Proposition. *Let $(V,\langle -,-\rangle)$ be an inner product space. Then every subspace $W\subseteq V$ is an inner product space with the restricted inner product defined by $\langle x,y\rangle_W=\langle x,y\rangle$ for all $x,y\in W$.*

Proof. Axioms (IPS 1)–(IPS 4) are trivially true in W, since true in V. □

At this point, we have an avalanche of examples of inner product spaces: All the ‘standard’ ones, and all their subspaces. We draw geometric intuition from $\mathbb{R}^2$ and $\mathbb{R}^3$ with their usual scalar product to extend some definitions, like:

4.1.14. Definition. Let V be an inner product space. We say that two vectors $x,y\in V$ are *orthogonal* and write $x\perp y$ if $\langle x,y\rangle=0$.

For a subset (often a subspace) $W\subseteq V$, the *orthogonal of* W is

$$W^{\perp}=\big\{\,x\in V\,\big|\,x\perp w \text{ for all } w\in W\,\big\}.$$

4.1.15. Exercise. Prove that $W^{\perp}$ is a subspace of V and that $W\subseteq (W^{\perp})^{\perp}$. Prove that when V is finite dimensional then $\dim(W^{\perp})=\dim(V)-\dim(W)$. Conclude that in this case $(W^{\perp})^{\perp}=W$.

The following is close to trivial but very useful.

4.1.16. Proposition. *We have $V^{\perp}=0$. In other words, if $x\in V$ is such that $\langle x,y\rangle=0$ for all $y\in V$ then $x=0$.*

Proof. Indeed, plugging $y=x$ itself, we are assuming that $\langle x,x\rangle=0$. This forces $x=0$ by (IPS 4); see Remark 4.1.3. □

Here is an important general result about inner product spaces.

4.1.17. Theorem (Cauchy-Schwarz Inequality)**.** *Let V be an inner product space and $x,y\in V$. Then the absolute value of the (real or complex) number $\langle x,y\rangle$ is bounded by the product of the norms of x and y, that is, we have (in $\mathbb{R}$)*

$$|\langle x,y\rangle|\le ||x||\cdot||y||.$$

Moreover, we have equality if and only if x,y are colinear (i.e. linearly dependent).

Proof. Suppose that x,y are linearly dependent. In that case, up to swapping x and y if one of them is zero, we can assume that $x=c\cdot y$ for some $c\in\mathbb{F}$. Then $\langle x,y\rangle=\langle cy,y\rangle=c\cdot\langle y,y\rangle$, hence $|\langle x,y\rangle|=|c|\cdot||y||^2=(c\cdot||y||)\cdot||y||=||x||\cdot||y||$.

Suppose that x, y are linearly independent and let us show the *strict* inequality $|\langle x, y\rangle| < ||x|| \cdot ||y||$. The only source of strict inequalities is Axiom (IPS 4). We are going to apply it to a vector of the form $y - c \cdot x$ for $c = \frac{\langle y,x\rangle}{\langle x,x\rangle} = \frac{\overline{\langle x,y\rangle}}{\langle x,x\rangle}$. Note that since x, y are linearly independent we have $x \neq 0$, hence $\langle x, x\rangle \neq 0$; so c is well-defined. Note also that we cannot have $y - c \cdot x = 0$ for then $y \in \operatorname{Span}(x)$ would break linear independence. Axiom (IPS 4) then tells us that in $\mathbb{R}$ we have

$$\begin{aligned}
0 \;&<\; \langle y - c\cdot x \,,\, y - c \cdot x\rangle && \text{since } y - c\cdot x \neq 0\\
&= \langle y, y\rangle - c \cdot \langle x, y\rangle - \bar{c} \cdot \langle y, x\rangle + c\bar{c}\langle x, x\rangle && \text{by (IPS 1),(IPS 1'),(IPS 2),(IPS 2')}\\
&= \langle y, y\rangle - c \cdot \langle x, y\rangle - \bar{c} \cdot \overline{\langle x, y\rangle} + c\bar{c}\langle x, x\rangle && \text{by (IPS 3)}\\
&= \langle y, y\rangle - \tfrac{|\langle x,y\rangle|^2}{\langle x,x\rangle} - \tfrac{|\langle x,y\rangle|^2}{\langle x,x\rangle} + \tfrac{|\langle x,y\rangle|^2}{\langle x,x\rangle} && \text{since } c = \tfrac{\langle y,x\rangle}{\langle x,x\rangle} \text{ and } \langle y, x\rangle = \overline{\langle x, y\rangle}.\\
&= \langle y, y\rangle - \tfrac{|\langle x,y\rangle|^2}{\langle x,x\rangle}.
\end{aligned}$$

Re-arranging the terms we get $|\langle x, y\rangle|^2 < \langle x, x\rangle \cdot \langle y, y\rangle = ||x||^2 \cdot ||y||^2$. □

We can use Cauchy-Schwarz to answer the question left open in Remark 4.1.5.

4.1.18. Corollary (Triangle Inequality)**.** *Let V be an inner product space and let $x, y \in V$. Then*

$$||x + y|| \leq ||x|| + ||y||.$$

Moreover, we have equality if and only if $y = 0$ or $x = \lambda \cdot y$ for some $\lambda \in \mathbb{R}$, $\lambda \geq 0$.

Proof. We compute in $\mathbb{R}$:

$$\begin{aligned}
||x + y||^2 &= \langle x + y, x + y\rangle && \text{by definition of the norm}\\
&= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle && \text{by (IPS 1),(IPS 1')}\\
&= \langle x, x\rangle + \langle x, y\rangle + \overline{\langle x, y\rangle} + \langle y, y\rangle && \text{by (IPS 3).}
\end{aligned}$$

Here it is important to note that the last expression uses complex numbers but the sum is real. We see the complex number $d = \langle x, y\rangle$ and its complex-conjugate $\bar{d} = \overline{\langle x, y\rangle}$. Their sum $d + \bar{d}$ is a real number: For every complex number $d = a + bi$ we have $d + \bar{d} = 2a$ and in fact $d + \bar{d} = 2a \leq 2 \cdot |a| = 2 \cdot \sqrt{a^2} \leq 2 \cdot \sqrt{a^2 + b^2} = 2 \cdot |d|$. Note *en passant* that the only way that this inequality

$$d + \bar{d} \leq 2 \cdot |d| \tag{4.1.19}$$

is an equality is to have $b = 0$ and $a \geq 0$, that is, to have $d \in \mathbb{R}_{\geq 0}$ a non-negative real number. Applying (4.1.19) to our $d = \langle x, y\rangle$ we can continue our computation

$$\begin{aligned}
||x + y||^2 &= \langle x, x\rangle + \langle x, y\rangle + \overline{\langle x, y\rangle} + \langle y, y\rangle && \text{(where we left things)}\\
&\leq \langle x, x\rangle + 2\,|\langle x, y\rangle| + \langle y, y\rangle && \text{by (4.1.19) for } d = \langle x, y\rangle\\
&\leq \langle x, x\rangle + 2\,||x|| \cdot ||y|| + \langle y, y\rangle && \text{by Cauchy-Schwarz 4.1.17}\\
&= ||x||^2 + 2\,||x|| \cdot ||y|| + ||y||^2 && \text{by definition of the norm}\\
&= (||x|| + ||y||)^2.
\end{aligned}$$

We get the result by taking square roots. Now the only way to have an equality is to have *both* inequalities to be equalities. In Cauchy-Schwarz, this happens only if x and y are linearly dependent. In (4.1.19) we have equality only if d is a nonnegative real number. If $y = 0$ there is nothing to prove. If $y \neq 0$ and $x = \lambda y$ (to get equality in Cauchy-Schwarz) for some $\lambda \in \mathbb{F}$ then we also need $d = \langle x, y\rangle \in \mathbb{R}_{\geq 0}$ but $\langle x, y\rangle = \langle \lambda y, y\rangle = \lambda \cdot ||y||^2$ and $||y||^2 > 0$. Hence $\lambda = \frac{d}{||y||^2}$ must be real and non-negative as well. So we have shown that $||x+y|| = ||x|| + ||y||$ forces $y = 0$ or $x = \lambda\, y$ for $\lambda \in \mathbb{R}_{\geq 0}$. The converse is an easy exercise. □

4.1.20. Exercise. Show that one can recover the inner product from the norm. Specifically, for $\mathbb{F} = \mathbb{R}$ we have $\langle x, y\rangle = \frac{1}{2}(||x+y||^2 - ||x||^2 - ||y||^2)$. Show that there is such a formula for $\mathbb{F} = \mathbb{C}$ as well.

4.1.21. Remark. Let $T\colon V \to V'$ be a linear transformation between two inner product spaces, with respective inner products $\langle -, -\rangle$ and $\langle -, -\rangle'$. Suppose that T preserves the inner products, *i.e.*

$$\langle T(x), T(y)\rangle' = \langle x, y\rangle$$

for all $x, y \in V$. In particular, this forces $||T(x)||' = ||x||$ for all $x \in V$ and this is in fact equivalent by Exercise 4.1.20. If this holds, we have $\mathrm{Ker}(T) = 0$ since $T(x) = 0$ would force $||x|| = ||T(x)||' = 0$ and thus $x = 0$. In particular, if $T\colon V \to V$ is an operator that preserves the inner product and if V is finite-dimensional then T is automatically an isomorphism. This explains why we usually reserve the name 'iso-metry' (same-metrics) for bijections:

4.1.22. Definition. An *isometry* between two inner product spaces V and V' is an isomorphism $T\colon V \to V'$ such that $\langle T(x), T(y)\rangle' = \langle x, y\rangle$ for all $x, y \in V$, or equivalently, $||T(x)||' = ||x||$ for all $x \in V$.

4.2. Orthonormal bases

If we continue our tour of Chapter 1 *with an inner product*, the next stop is the notion of basis. Of course, an inner product space $(V, \langle -, -\rangle)$ still has bases in the ordinary sense (infinitely many unless $V = 0$). The point is that some bases are better behaved.

4.2.1. Definition. A collection of vectors $B \subset V$ in an inner product space is called *orthonormal* if

(a) For all $b, b' \in B$ with $b \neq b'$, we have $b \perp b'$.
(b) For all $b \in B$ we have $||b|| = 1$.

We can write both conditions with the Kronecker symbol: For all $b, b' \in B$ we want

$$\langle b, b'\rangle = \delta_{b,b'} = \begin{cases} 1 & \text{if } b = b' \\ 0 & \text{if } b \neq b'. \end{cases}$$

An *orthonormal basis* B of V is a basis that is orthonormal.

4.2.2. Exercise. Verify that the canonical basis is orthonormal in $\mathbb{F}^n$ with the standard inner product.

4.2.3. Exercise. Show that the set $\left\{ f_n = (t \mapsto e^{2\pi n t i}) \,\middle|\, n \in \mathbb{Z} \right\}$ is orthonormal in the complex inner product space $\mathrm{Cont}([0,1], \mathbb{C})$ of Example 4.1.12.

4.2.4. Exercise. Show that the set $\left\{ g_n = (t \mapsto \cos(2\pi n t)) \,\middle|\, n \in \mathbb{N} \right\} \cup \left\{ h_n = (t \mapsto \sin(2\pi n t)) \,\middle|\, n \in \mathbb{N}, n \neq 0 \right\}$ is orthonormal in the real inner product space $\mathrm{Cont}([0,1], \mathbb{R})$ of Example 4.1.12.

4.2.5. Lemma. *Let $b_1, \dots, b_n$ be n orthonormal vectors in an inner product space V and let $x \in \mathrm{Span}(b_1, \dots, b_n)$. Suppose $x = \sum_{i=1}^n a_i \cdot b_i$ for $a_1, \dots, a_n \in \mathbb{F}$. Then*

$$a_i = \langle x, b_i \rangle$$

for all $i = 1, \dots, n$.

Proof. Let j be an index $1 \leq j \leq n$. We compute

$$\langle x, b_j \rangle = \langle \sum_{i=1}^n a_i b_i, b_j \rangle = \sum_{i=1}^n a_i \langle b_i, b_j \rangle = \sum_{i=1}^n a_i \delta_{i,j} = a_j.$$

Hence the result, by rewriting. □

4.2.6. Proposition. *If $B \subset V$ is orthonormal, then B is linearly independent. Hence it is an orthonormal basis of* Span(B).

Proof. Suppose that $b_1, \dots, b_n \in B$ are n distinct vectors in B and that $\sum_{i=1}^n a_i \cdot b_i = 0$ for $a_1, \dots, a_n \in \mathbb{F}$. Apply Lemma 4.2.5 for $x = 0$. We get $a_i = \langle 0, b_i \rangle = 0$ for all $i = 1, \dots, n$. □

4.2.7. Remark. We can restate the property that a basis $B = \{b_1, \dots, b_n\}$ of an inner product space V is orthonormal in terms of the isomorphism $[-]_B \colon V \overset{\sim}{\to} \mathbb{F}^n$ associated to B (see Theorem 2.3.10). Indeed, being orthonormal exactly means that $[-]_B \colon V \overset{\sim}{\to} \mathbb{F}^n$ is an *isometry* in the sense of Definition 4.1.22.

4.2.8. Proposition. *Let $B = \{b_1, \dots, b_n\}$ be a basis of an inner product space V of dimension n. Then B is orthonormal if and only if*

$$\langle x, y \rangle = \langle [x]_B \, , \, [y]_B \rangle_{\mathbb{F}^n}$$

where the right-hand side is the standard inner product of Example 4.1.7.

Proof. Easy exercise on the definitions. □

Orthonormal bases are particulary convenient to compute coordinates.

4.2.9. Proposition. *Let $B = \{b_1, \dots, b_n\}$ be an orthonormal basis of an inner product space V of finite dimension n. Then for every $x \in V$ we have*

$$x = \sum_{i=1}^n \langle x, b_i \rangle \cdot b_i. \tag{4.2.10}$$

In other words, the coordinates of x are given by $[x]_B = \begin{bmatrix} \langle x,b_1 \rangle \\ \vdots \\ \langle x,b_n \rangle \end{bmatrix}$.

Proof. This is simply rewriting Lemma 4.2.5, with the word 'coordinates' (for the a_i) and the notation $[x]_B$. □

Remarkably, the right-hand side of (4.2.10) is useful even if x does not belong to Span(B).

4.2.11. Theorem. *Let V be an inner product space and $W \subseteq V$ be a subspace of dimension $n \geq 1$. Let $B = \{b_1, \dots, b_n\} \subset W$ be a* orthonormal basis *of the subspace W. Let $x \in V$ and define $w \in W$ by*

$$w = \sum_{i=1}^n \langle x, b_i \rangle \cdot b_i.$$

Then $(x - w) \in W^\perp$. That is, we wrote $x = w + (x - w)$ in $W + W^\perp$.

In particular, we have $V = W \oplus W^\perp$ and the linear transformation

$$\begin{array}{rccc} \operatorname{proj}_W^\perp\colon & V & \longrightarrow & W \\ & x & \longmapsto & \sum_{i=1}^n \langle x, b_i\rangle \cdot b_i \end{array}$$

is the orthogonal projection of V onto W (along $W^\perp$). Its image is W. Its kernel is $W^\perp$. And we have $\operatorname{proj}_W^\perp \circ \operatorname{proj}_W^\perp = \operatorname{proj}_W^\perp$.

Proof. As often with long statements, everything is already there and the proof is easy. For the first part, apply Lemma 4.2.5 again, this time to $w = \sum_{i=1}^n \langle x, b_i\rangle \cdot b_i$ as in the statement. Lemma 4.2.5 tells us that the coefficient of b_i in this linear combination is $\langle w, b_i\rangle$. This proves that $\langle x-w, b_i\rangle = \langle x, b_i\rangle - \langle w, b_i\rangle = 0$. In other words, $(x - w) \perp b_i$ for all $i = 1, \dots, n$. It follows that $(x - w) \perp z$ for all $z \in \operatorname{Span}(b_1, \dots, b_n) = W$. Hence $(x-w) \in W^\perp$ as claimed. So $x = w+(x-w) \in W + W^\perp$.

So we proved that $W + W^\perp = V$ (as x could be taken arbitrary in the first part). We have $W \cap W^\perp = 0$ by (IPS 4) or Remark 4.1.3: If $y \in W \cap W^\perp$ then $\langle y, y\rangle = 0$ since the left-hand y belongs to W and the right-hand one belongs to $W^\perp$. In short, $V = W \oplus W^\perp$. It is easy to see that $\operatorname{proj}_W^\perp$ is well-defined (lands in $\operatorname{Span}(b_1, \dots, b_n) = W$) and linear (each $\langle -, b_i\rangle$ is).

Every $y \in W$ satisfies $\operatorname{proj}_W^\perp(y) = y$ by Proposition 4.2.9. Hence $\operatorname{proj}_W^\perp$ is onto and applying this to $y = \operatorname{proj}_W^\perp(x)$ for $x \in V$ we see that $\operatorname{proj}_W^\perp(\operatorname{proj}_W^\perp(x)) = \operatorname{proj}_W^\perp(x)$. Finally, if $\operatorname{proj}_W^\perp(x) = 0$ then $x = w+(x-w)$ with $w = 0$ and $x-0 \in W^\perp$ means $x \in W^\perp$. The converse is clear: If $x \in W^\perp$ then $\langle x, b_i\rangle = 0$ for all i since $b_i \in W$ and therefore $\operatorname{proj}_W^\perp(x) = 0$. □

In view of the convenience of orthonormal bases to compute coordinates, it is important to decide whether (finite-dimensional) inner product spaces always admit orthonormal bases, and how to construct them. Enter Gram-Schmidt.

4.2.12. Lemma. *Let $b_1, \dots, b_m$ be $m \geq 1$ orthonormal vectors in an inner product space V and let $W := \operatorname{Span}(b_1, \dots, b_m)$. Let $v \in V$.*

(1) *Define $v' = v - \operatorname{proj}_W^\perp(v) = v - \sum_{i=1}^m \langle v, b_i\rangle \cdot b_i$. Then $v' \neq 0$ if and only if $v \notin W$ if and only if $b_1, \dots, b_m, v$ are linearly independent. In that case, we have $\operatorname{Span}(b_1, \dots, b_m, v) = \operatorname{Span}(b_1, \dots, b_m, v')$.*

(2) *Suppose that $v' \neq 0$ in (1) and let $b_{m+1} = ||v'||^{-1} \cdot v'$. Then $b_1, \dots, b_{m+1}$ are orthonormal and $\operatorname{Span}(b_1, \dots, b_m, v) = \operatorname{Span}(b_1, \dots, b_m, b_{m+1})$.*

Proof. For (1), we have $v' = 0$ if and only if $v = \operatorname{proj}_W^\perp(v)$ if and only if $v \in W$ by Theorem 4.2.11. The last equivalence is Lemma 1.6.12. Finally, we can prove both inclusions $\operatorname{Span}(b_1, \dots, b_m, v) \subseteq \operatorname{Span}(b_1, \dots, b_m, v')$ and $\operatorname{Span}(b_1, \dots, b_m, v') \subseteq \operatorname{Span}(b_1, \dots, b_m, v)$ from the constructions, since $v = v' + \operatorname{proj}_W^\perp(v)$.

For (2), we already know that $b_1, \dots, b_m$ are orthonormal, so it suffices to show that $\langle b_{m+1}, b_j\rangle = 0$ for $1 \leq j \leq m$ and that $||b_{m+1}|| = 1$. The latter is obvious since we 'normed' v'. For the former, since $b_{m+1} \in \operatorname{Span}(v')$ if suffices to know that $v' \perp W = \operatorname{Span}(b_1, \dots, b_m)$. But this is Theorem 4.2.11 since $v' = v-\operatorname{proj}_W^\perp(v)$. The last equality $\operatorname{Span}(b_1, \dots, b_m, v) = \operatorname{Span}(b_1, \dots, b_m, b_{m+1})$ is easy since we already saw $\operatorname{Span}(b_1, \dots, b_m, v) = \operatorname{Span}(b_1, \dots, b_m, v')$ and b_{m+1} is a non-zero multiple of v'. □

4.2.13. Theorem (Gram-Schmidt Orthonormalization Algorithm)**.** *Let $v_1, \ldots, v_n$ be a basis of an inner product space V. We define $b_1, \ldots, b_n$ inductively. Set*

$$b_1 = \frac{1}{||v_1||} \cdot v_1.$$

Suppose defined $b_1, \ldots, b_m$ for $1 \leq m \leq n-1$ that are orthonormal and such that $\mathrm{Span}(b_1, \ldots, b_m) = \mathrm{Span}(v_1, \ldots, v_m)$. Define

$$v'_{m+1} = v_{m+1} - \mathrm{proj}^{\perp}_{\mathrm{Span}(b_1,\ldots,b_m)}(v_{m+1}) = v_{m+1} - \sum_{i=1}^{m} \langle v_{m+1}, b_i \rangle \cdot b_i.$$

Then $v'_{m+1} \neq 0$ and we can define

$$b_{m+1} = \frac{1}{||v'_{m+1}||} \cdot v'_{m+1}.$$

Eventually, we obtain an orthonormal basis $b_1, \ldots, b_n$ of V.

Proof. This is a simple induction on Lemma 4.2.12. We prove by induction on m with $m \leq n$ that $b_1, \ldots, b_m$ are well-defined and form an orthonormal basis of $\mathrm{Span}(v_1, \ldots, v_m)$. This claim for $m = n$ is the statement. This clearly holds for $m = 1$. Once this holds for some $m \leq n-1$, then we can apply Lemma 4.2.12 to $v = v_{m+1}$. Since $v_1, \ldots, v_{m+1}$ are linearly independent, we have $v_{m+1} \notin \mathrm{Span}(v_1, \ldots, v_m) = \mathrm{Span}(b_1, \ldots, b_m)$ and we can apply Part (1) of the lemma, and then Part (2). This yields b_{m+1} such that $b_1, \ldots, b_{m+1}$ form an orthonormal basis of $\mathrm{Span}(b_1, \ldots, b_m, v_{m+1}) = \mathrm{Span}(v_1, \ldots, v_m, v_{m+1})$ as claimed. □

4.2.14. Corollary. *Any finite-dimensional inner product space admits an orthonormal basis.*

Proof. Feed *any* (finite) basis into Theorem 4.2.13. □

4.2.15. Remark. As usual, subspaces are vector spaces in their own right. And subspaces of inner product spaces are inner product spaces by Proposition 4.1.13. So we can always find an orthonormal basis of any finite-dimensional subspace, as in Theorem 4.2.11. Hence we can always find an explicit formula for the orthogonal projection, at least in finite dimension.

4.2.16. Remark. If you feed n vectors $v_1, \ldots, v_n$ to Gram-Schmidt without prior verification that they were linearly independent, the process could collapse with some $v'_{m+1} = 0$ (making it impossible to define b_{m+1} since you cannot divide by $0 = ||v'_{m+1}||$ in that case). This simply indicates that $v_{m+1} \in \mathrm{Span}(b_1, \ldots, b_m) = \mathrm{Span}(v_1, \ldots, v_m)$ by Lemma 4.2.12 (1). One can then remove v_{m+1} without changing the span of $v_1, \ldots, v_n$ and restart Gram-Schmidt.

4.2.17. Exercise. Apply the Gram-Schmidt algorithm to the vectors $v_1 = \left[\begin{smallmatrix}1\\2\\3\end{smallmatrix}\right]$, $v_2 = \left[\begin{smallmatrix}4\\5\\6\end{smallmatrix}\right]$, $v_3 = \left[\begin{smallmatrix}7\\8\\9\end{smallmatrix}\right]$ of the standard inner product space $\mathbb{R}^3$. Find an orthonormal basis of $\mathrm{Span}(v_1, v_2, v_3)$.

4.2.18. Definition. A square matrix $Q \in \mathrm{M}_{n\times n}(\mathbb{F})$ is called *orthogonal* (often assuming $\mathbb{F} = \mathbb{R}$, while most authors would say *unitary* in the case $\mathbb{F} = \mathbb{C}$) if $\bar{Q}^t \cdot Q = I_n$. In other words, $Q \in \mathrm{GL}_n(\mathbb{F})$ is invertible and $Q^{-1} = \bar{Q}^t$. Here of course $\bar{A}$ is the entrywise complex conjugation $(\bar{A})_{ij} = \overline{A_{ij}}$ and A^t is the transpose.

4.2.19. Exercise. Show that Q is orthogonal (unitary) if and only if the columns $x_1, \ldots, x_n$ of $Q = [x_1|\cdots|x_n]$ are orthonormal in $\mathbb{F}^n$ with the standard scalar product.

4.2.20. Exercise. Let V be an inner product space of finite dimension $n \geq 1$. Let B be an orthonormal basis and C be another basis. Prove that the change-of-coordinates matrix $Q_{B,C}$ is orthogonal (unitary) if and only if C is also an orthonormal basis of V.

As an application of Gram-Schmidt, we have a solution to Exercise 4.1.15.

4.2.21. Corollary. *Let $W \subseteq V$ be a subspace of a finite-dimensional inner product space V. Then $\dim(W^\perp) = \dim(V) - \dim(W)$. And in particular $(W^\perp)^\perp = W$.*

Proof. We are really using Theorem 4.2.11 that tells us that $V = W \oplus W^\perp$. Except that in order to apply this theorem, we need an orthonormal basis of W, whose existence (Corollary 4.2.14) follows from Gram-Schmidt. Then $\dim(V) = \dim(W) + \dim(W^\perp)$ and we get the first statement. Consequently, $\dim((W^\perp)^\perp) = \dim(V) - \dim(W^\perp) = \dim(V) - (\dim(V) - \dim(W)) = \dim(W)$ and since $W \subseteq (W^\perp)^\perp$ is trivial those two subspaces having the same dimension must be equal. □

4.3. Adjoint of linear transformation

In this section, $T\colon V \to V'$ is a linear transformation between inner product spaces. We want to understand the interplay between T and the inner products on V and V'. When we want to distinguish them we shall write $\langle -, - \rangle$ for the inner product in V and $\langle -, - \rangle'$ for the inner product in V'. There are several questions we could ask about this topic, for instance, we could ask about $\langle T(x_1), T(x_2) \rangle'$ in terms of $\langle x_1, x_2 \rangle$. This relates with the topic of isometries (Definition 4.1.22). But let us start with something simpler: We pick a vector $x \in V$ and we consider the inner product in V' between $T(x)$ and a general vector y in V':

$$\langle T(x), y \rangle' = \langle x, ? \rangle.$$

The question is whether we can express this scalar $\langle T(x), y \rangle'$ as the inner product in V of x and some vector (denoted '?' above), probably depending on y. The construction of this '?' in terms of y is a linear construction that is called the *adjoint* of T. Let us do it in the matrix case first.

4.3.1. Example. Let $T = T_A\colon V = \mathbb{F}^p \to V' = \mathbb{F}^q$ be a linear transformation given by $A \in \mathrm{M}_{q\times p}(\mathbb{F})$. Consider the usual inner products on $\mathbb{F}^p$ and $\mathbb{F}^q$, as in Example 4.1.7. Let $x \in \mathbb{F}^p$ and $y \in \mathbb{F}^q$. We compute

$$\begin{aligned} \langle A\cdot x, y\rangle &= (A\cdot x)^t \cdot \overline{y} && \text{by (4.1.8) in } \mathbb{F}^q \\ &= x^t \cdot A^t \cdot \overline{y} && \text{since } (A_1A_2)^t = A_2^tA_1^t \\ &= x^t \cdot \overline{\bar{A}^t \cdot y} && \text{since } \bar{\bar{A}} = A \\ &= \langle x, \bar{A}^t \cdot y\rangle && \text{by (4.1.8) in } \mathbb{F}^p. \end{aligned}$$

The *adjoint* $A^* \in \mathrm{M}_{n\times m}(\mathbb{F})$ of A is simply the above conjugate-transpose

$$A^* = \bar{A}^t.$$

We just proved that for all $x \in \mathbb{F}^p$ and $y \in \mathbb{F}^q$

$$\langle A\cdot x\,,\, y\rangle = \langle x\,,\, A^*\cdot y\rangle. \tag{4.3.2}$$

We can define T^* for any linear transformation.

4.3.3. Theorem. *Let $T\colon V \to V'$ be a linear transformation between inner product spaces of finite dimension.*

(1) *There exists a unique linear transformation $T^*\colon V' \to V$ such that*

$$\langle T(x), y\rangle' = \langle x, T^*(y)\rangle \tag{4.3.4}$$

for all $x \in V$ and all $y \in V'$.

(2) *The matrix of T^* can be described as follows. If B is an orthonormal basis of V and B' is an orthonormal basis of V' and $A = [T]_{B',B}$ then*

$$[T^*]_{B,B'} = A^* = \overline{A}^t. \tag{4.3.5}$$

4.3.6. Definition. The unique linear transformation $T^*\colon V' \to V$ such that (4.3.4) holds for all $x \in V$ and all $y \in V'$ is called the *adjoint* of T (with respect to the given inner products on V and V').

Proof of Theorem 4.3.3. Let us show uniqueness in (1). Suppose that $T^*(y)$ and $T^\star(y)$ are two vectors in V such that $\langle T(x), y\rangle' = \langle x, T^*(y)\rangle = \langle x, T^\star(y)\rangle$ for all $x \in V$. Then $\langle x, T^*(y) - T^\star(y)\rangle = 0$ for all $x \in V$ and therefore $T^*(y) - T^\star(y) \in V^\perp = 0$. So $T^*(y) = T^\star(y)$ as claimed.

Let us now show that this $T^*\colon V' \to V$ exists. Part (2) predicts what the matrix of T^* should be. So we prove existence and Part (2) simultaneously. Choose some orthonormal bases B and B' of V and V', using Corollary 4.2.14. Define a linear transformation $T^*\colon V' \to V$ by the requirement that $[T^*]_{B,B'} = A^*$ where $A = [T]_{B',B}$. This means that we have

$$[T(x)]_{B'} = A \cdot [x]_B \qquad \text{and} \qquad [T^*(y)]_B = A^* \cdot [y]_{B'} \tag{4.3.7}$$

for every $x \in V$ and for every $y \in V'$. (See Proposition 2.4.7 if necessary.) Using that B and B' are orthonormal bases, we can translate the inner product computations into the usual inner product of $\mathbb{F}^n$ via Proposition 4.2.8. Now compute for every $x \in V$ and $y \in V'$ the following scalar

$$\begin{aligned}
\langle T(x), y\rangle' &= \langle [T(x)]_{B'}, [y]_{B'}\rangle_{\text{usual}} && \text{since } B' \text{ is orthonormal}\\
&= \langle A \cdot [x]_B, [y]_{B'}\rangle_{\text{usual}} && \text{by (4.3.7)}\\
&= \langle [x]_B, A^* \cdot [y]_{B'}\rangle_{\text{usual}} && \text{by Example 4.3.1}\\
&= \langle [x]_B, [T^*(y)]_B\rangle_{\text{usual}} && \text{by (4.3.7)}\\
&= \langle x\,,\, T^*(y)\rangle && \text{since } B \text{ is orthonormal.}
\end{aligned}$$

This proves the existence of $T^*\colon V' \to V$ such that (4.3.4) holds for all $x \in V$ and $y \in V'$. This actually proves Part (2). Let us be careful: We only did it for some choice of B and B' but Part (2) claims that (4.3.5) should hold for *all* orthonormal bases B and B'. However, once we know that T^* exists (which we did with the help of *some* chosen orthonormal bases) then for any orthonormal bases B and B' of V and V' we can define an auxiliary $T^\star\colon V' \to V$ satisfying $[T^\star]_{B,B'} = \overline{[T]}^t_{B',B}$ as we did above and then prove exactly as above that $\langle T(x), y\rangle' = \langle x, T^\star(y)\rangle$ for all $x \in V$ and $y \in V'$. By uniqueness, we must have $T^\star = T^*$ and therefore $[T^*]_{B,B'} = [T^\star]_{B,B'} = \overline{[T]}^t_{B',B}$ as wanted. □

4.3.8. Corollary. *Let $T = T_A\colon \mathbb{F}^p \to \mathbb{F}^q$ be the linear transformation given by the matrix $A \in \mathrm{M}_{q\times p}(\mathbb{F})$. Then its adjoint $T^*\colon \mathbb{F}^q \to \mathbb{F}^p$ with respect to the standard inner products is given by the adjoint matrix: $(T_A)^* = T_{A^*}$.*

Proof. The canonical bases are orthonormal bases for the standard inner product and the matrix of T_A in the canonical basis is A. □

4.3.9. Remark. Theorem 4.3.3 says that *with respect to orthonormal bases* the matrix of the adjoint is the adjoint of the matrix

$$[T^*]_{B,B'} = \big([T]_{B',B}\big)^*$$

when B is an orthonormal basis of V and B' is an orthonormal basis of V'. It is totally false for general (non orthonormal) bases! We can complete Diagram (2.4.12)

$$\begin{tikzcd}
V \arrow[r, "\sim", "{[-]_B}"'] \arrow[d, bend right, "T"'] & \mathbb{F}^p \arrow[d, bend left, "T_A"'] \\
V' \arrow[r, "\sim", "{[-]_{B'}}"'] & \mathbb{F}^q
\end{tikzcd}$$

and

$$\begin{tikzcd}
V \arrow[r, "\sim", "{[-]_B}"'] & \mathbb{F}^p \\
V' \arrow[r, "\sim", "{[-]_{B'}}"'] \arrow[u, bend right, "T^*"'] & \mathbb{F}^q \arrow[u, bend right, "T_{A^*}"']
\end{tikzcd}$$

The horizontal isometries come from the orthonormal bases. The linear transformation T ‘became’ multiplication by its matrix $A = [T]_{B',B}$. Similarly, the linear transformation T^* ‘becomes’ multiplication by a matrix $[T^*]_{B,B'}$ that we proved to be the conjugate-transpose A^* of A.

4.3.10. Exercise. Show that $(S+T)^* = S^* + T^*$ and that $(c \cdot T)^* = \bar{c} \cdot T^*$, for all $S, T \colon V \to V'$ and $c \in \mathbb{F}$.

4.3.11. Exercise. Show that $\mathbb{R}[X]$ is a real inner product space with scalar product $\langle P, Q\rangle = \int_{-1}^{1} P(t) \cdot Q(t)\ \mathrm{d}t$. Let V and W be the subspaces of polynomials of degree at most 2 and at most 1, respectively. Let $T \colon V \to W$ given by $T(P) = P'$. Compute the adjoint $T^* \colon W \to V$.

4.3.12. Exercise. Give the matrix of T^* in Exercise 4.3.11 with respect to the canonical bases.

4.3.13. Exercise. Exact same question as in Exercise 4.3.11 but changing the inner product to $\langle P, Q\rangle = \int_{0}^{1} P(t) \cdot Q(t)\ \mathrm{d}t$.

4.3.14. Remark. Since $(A^*)^* = A$ for matrices (and since every finite-dimensional inner product space admits an orthonormal basis) we have $(T^*)^* = T$. The double-adjoint of T is T itself. As an exercise, try to prove this from the uniqueness property of $(T^*)^*$.

4.3.15. Exercise. Let $T \colon V \to V'$ and $S \colon V' \to V''$ be linear transformations between finite-dimensional inner product spaces. Give two proofs that $(S \circ T)^* = T^* \circ S^*$, one via the matrix relation $(A_1 \cdot A_2)^* = A_2^* \cdot A_1^*$ and one via the uniqueness property of the adjoint.

Given $T \colon V \to V'$ and its adjoint $T^* \colon V' \to V$, we have four subspaces at our disposal: $\mathrm{Ker}(T)$, $\mathrm{Im}(T^*)$ in V and $\mathrm{Im}(T)$ and $\mathrm{Ker}(T^*)$ in V'. They are related via orthogonals (Definition 4.1.14).

4.3.16. Proposition. *Let V and V' be finite-dimensional inner product spaces. Let $T \colon V \to V'$ be linear and $T^* \colon V' \to V$ its adjoint. Then*

$$\mathrm{Ker}(T^*) = \mathrm{Im}(T)^{\perp} \qquad \textit{and} \qquad \mathrm{Im}(T^*) = \mathrm{Ker}(T)^{\perp}$$

and therefore

$$\mathrm{Ker}(T) = \mathrm{Im}(T^*)^{\perp} \qquad \textit{and} \qquad \mathrm{Im}(T) = \mathrm{Ker}(T^*)^{\perp}.$$

Proof. Let $y \in V'$. The condition $y \in \mathrm{Im}(T)^{\perp}$ is equivalent to saying that $\langle z, y\rangle' = 0$ for all $z \in \mathrm{Im}(T) = \{T(x) \,|\, x \in V\}$. In other words, it means $\langle T(x), y\rangle' = 0$ for all $x \in V$. By the defining property (4.3.4) of T^* the latter means $\langle x, T^*(y)\rangle = 0$ for all $x \in V$. But this equivalent to $T^*(y) \in V^{\perp} = 0$, that is $y \in \mathrm{Ker}(T^*)$. In short, we have the left-hand equality below:

$$\mathrm{Ker}(T^*) = (\mathrm{Im}(T))^{\perp} \qquad \text{and} \qquad (\mathrm{Ker}(T^*))^{\perp} = \mathrm{Im}(T).$$

The right-hand equality is simply obtained by taking the orthogonal on both sides, and using that $(\mathrm{Im}(T))^{\perp\,\perp} = \mathrm{Im}(T)$ by Corollary 4.2.21. Now applying the above two relations to the linear transformation $T^*\colon V' \to V$ we get

$$\mathrm{Ker}(T^{**}) = (\mathrm{Im}(T^*))^{\perp} \qquad \text{and} \qquad (\mathrm{Ker}(T^{**}))^{\perp} = \mathrm{Im}(T^*)$$

and we now use that $T^{**} = T$. We proved the four equalities of the statement. □

4.3.17. Corollary. *With notation as in Proposition 4.3.16, we have*

$$\mathrm{rank}(T^*) = \mathrm{rank}(T).$$

Proof. One can prove this in matrix form, since $\mathrm{rank}(A^*) = \mathrm{rank}(A)$ but we can use the above proposition too: We have $\mathrm{rank}(T^*) = \dim(\mathrm{Im}(T^*)) = \dim(\mathrm{Ker}(T)^{\perp}) = \dim(V) - \dim(\mathrm{Ker}(T))$ by Corollary 4.2.21. By Rank-Nullity Theorem 2.2.6, the latter is also $\mathrm{rank}(T)$. □

4.3.18. Proposition. *With notation as in Proposition 4.3.16, we have*

$$\mathrm{Ker}(T^* \circ T) = \mathrm{Ker}(T) \qquad \textit{and} \qquad \mathrm{Im}(T^* \circ T) = \mathrm{Im}(T^*).$$

for the operator $T^* \circ T$ on V and similarly for the operator $T \circ T^*\colon V' \to V'$

$$\mathrm{Ker}(T \circ T^*) = \mathrm{Ker}(T^*) \qquad \textit{and} \qquad \mathrm{Im}(T \circ T^*) = \mathrm{Im}(T).$$

Proof. The inclusion $\mathrm{Ker}(T) \subseteq \mathrm{Ker}(T^* \circ T)$ is trivial. Let $x \in \mathrm{Ker}(T^* \circ T)$. This means that $T^*(T(x)) = 0$. Hence $0 = \langle x, 0\rangle = \langle x, T^*(T(x))\rangle = \langle T(x), T(x)\rangle'$ by the defining property (4.3.4) of T^*. The latter means $||T(x)||' = 0$, that is, $T(x) = 0$. So $x \in \mathrm{Ker}(T)$. This proves the non-trivial inclusion $\mathrm{Ker}(T^* \circ T) \subseteq \mathrm{Ker}(T)$ and thus equality. Applying this result to $T^*\colon V' \to V$ we get the formula for $\mathrm{Ker}(T \circ T^*)$ since $(T^*)^* = T$. We then get the two formulas for the images by taking orthogonal and using Proposition 4.3.16. In so doing, we also use $(T^* \circ T)^* = T^* \circ T^{**} = T^* \circ T$; see Exercise 4.3.15. □

4.3.19. Corollary. *With notation as in Proposition 4.3.16, we have* $\mathrm{rank}(T \circ T^*) = \mathrm{rank}(T) = \mathrm{rank}(T^*) = \mathrm{rank}(T^* \circ T)$. □

4.3.20. Corollary. *Let $A \in \mathrm{M}_{p\times q}(\mathbb{R})$.*

(1) *If* $\mathrm{rank}(A) = p$ *then* $A \cdot A^t \in \mathrm{M}_{p\times p}(\mathbb{F})$ *is invertible.*
(2) *If* $\mathrm{rank}(A) = q$ *then* $A^t \cdot A \in \mathrm{M}_{q\times q}(\mathbb{F})$ *is invertible.*

Proof. We have $\mathrm{rank}(AA^t) = \mathrm{rank}(A) = p$ and AA^t is a $(p \times p)$-matrix. □

4.3.21. Remark. The last corollary is useful in two numerical applications. One is the 'least square method' which tries to approximate a solution to a linear system $A \cdot x = b$ that has no solution (typically when $p \gg q$). The other one is the 'minimal solution method' which tries to find a solution with minimal norm among the solutions of a linear system $A \cdot x = b$ that has infinitely many solutions (typically when $q \gg p$).

4.4. Normal and self-adjoint operators

Continuing our review of Chapters 1–3 in the presence of an inner product, we now turn to operators and their diagonalization.

Let $(V, \langle -, - \rangle)$ be a finite-dimensional inner product space. In this section, $T\colon V \to V$ is an operator on V. The remarkable property is the following one.

4.4.1. Definition. Let $T\colon V \to V$ be a linear operator on a finite-dimensional inner product space. Recall its adjoint $T^*\colon V \to V$ from Definition 4.3.6. We say that T is a *normal* operator if $T \circ T^* = T^* \circ T$.

4.4.2. Example. Of course, $T = 0$ and $T = \mathrm{Id}_V$ are normal. Also, if T is normal then so is $c \cdot T$ for any scalar $c \in \mathbb{F}$. Indeed, $(c\,T)^* = \bar{c}T^*$ and $(c\,T)^*(c\,T) = \bar{c}T^*\,c\,T = \bar{c}\,c\,T^*\,T = c\,\bar{c}\,T\,T^* = c\,T\,\bar{c}\,T^* = (cT)(cT)^*$.

4.4.3. Remark. We say that a matrix $A \in \mathrm{M}_{n\times n}(\mathbb{F})$ is *normal* if $A \cdot A^* = A^* \cdot A$, where we recall that the adjoint $A^* = \bar{A}^t$ is simply the conjugate-transpose. This is equivalent to the operator $T = T_A\colon \mathbb{F}^n \to \mathbb{F}^n$ being normal *for the standard inner product.* Indeed, $T_{A^*} = (T_A)^*$ by Corollary 4.3.8.

4.4.4. Proposition. *Let $T\colon V \to V$ and B be an orthonormal basis of V. Then T is normal if and only if the matrix $A = [T]_B$ is normal.*

Proof. This is immediate once we know that $[T^*]_B = A^*$ because B is orthonormal; see Theorem 4.3.3 (2). □

4.4.5. Exercise. The sum of normal operators is not necessarily normal. Find an example. However, if T is normal then $T + \lambda\,\mathrm{Id}_V$ remains normal. Indeed, $(\lambda\,\mathrm{Id}_V)^* = \bar{\lambda}\,\mathrm{Id}_V$ commutes with T and T^*.

Here is an important class of examples of normal operators:

4.4.6. Definition. An operator $T\colon V \to V$ is *self-adjoint* if $T^* = T$.

4.4.7. Example. A real matrix $A \in \mathrm{M}_{n\times n}(\mathbb{R})$ is self-adjoint for the standard inner product if and only if $A = A^t$, *i.e.* A is symmetric.

4.4.8. Example. Note that a skew-symmetric matrix $A \in \mathrm{M}_{n\times n}(\mathbb{R})$, that is, one such that $A^t = -A$, is normal too: $A^t \cdot A = -A^2 = A \cdot A^t$.

We recall that the key result of Section 3.3 was Proposition 3.3.7 that told us, in colloquial terms, that eigenspaces for different eigenvalues were 'linearly independent' – do not use that terminology! Use the precise statement of Proposition 3.3.7. For normal operators, we have actually a very clean formulation: Eigenspaces are orthogonal! Let us do some preparation.

4.4.9. Corollary. *Let $T\colon V \to V$ be a normal operator. Then* $\mathrm{Ker}(T) = \mathrm{Ker}(T^*)$.

Proof. Proposition 4.3.18 tells us that $\operatorname{Ker}(T) = \operatorname{Ker}(T^* \circ T)$ and $\operatorname{Ker}(T^*) = \operatorname{Ker}(T \circ T^*)$. If $T \circ T^* = T^* \circ T$ we are done. □

4.4.10. Corollary. *Let $\lambda \in \mathbb{F}$ be an eigenvalue of a* normal *operator $T\colon V \to V$. Then $\bar{\lambda}$ is an eigenvalue of T^* and*

$$E_\lambda(T) = E_{\bar{\lambda}}(T^*).$$

In other words, $x \in V$ is an eigenvector for T with eigenvalue λ if and only if x is an eigenvector for T^* with eigenvalue $\bar{\lambda}$.

Proof. Apply the previous corollary to the normal operator $T - \lambda \operatorname{Id}_V$. (We saw in Exercise 4.4.5 that $T - \lambda \operatorname{Id}_V$ is normal if T is.) Corollary 4.4.9 then reads

$$\operatorname{Ker}(T - \lambda \operatorname{Id}_V) = \operatorname{Ker}(T^* - \bar{\lambda} \operatorname{Id}_V).$$

Hence the left-hand side is non-zero (*i.e.* λ is an eigenvalue of T) if and only if the right-hand side is non-zero (*i.e.* $\bar{\lambda}$ is an eigenvalue of T^*). And then the above kernels are the eigenspaces of the statement. □

We can then prove the critical statement (analogous to Proposition 3.3.7).

4.4.11. Theorem. *Let $T\colon V \to V$ be a* normal *operator on a finite-dimensional inner product space. Let $\lambda \neq \mu$ be two distinct eigenvalues of T. Then $E_\lambda(T) \perp E_\mu(T)$. In other words, every eigenvector for λ is orthogonal to any eigenvector for μ.*

Proof. Let $x \in E_\lambda(T)$ and $y \in E_\mu(T)$. We want to show that $\langle x, y\rangle = 0$. By Corollary 4.4.10, we have $y \in E_{\bar{\mu}}(T^*)$. We compute in $\mathbb{F}$:

$$\begin{aligned}
\lambda \cdot \langle x, y\rangle &= \langle \lambda x, y\rangle && \text{by (IPS 2)}\\
&= \langle T(x), y\rangle && \text{since } x \in E_\lambda(T)\\
&= \langle x, T^*(y)\rangle && \text{by the key property (4.3.4) of } T^*\\
&= \langle x, \bar{\mu} y\rangle && \text{since } y \in E_{\bar{\mu}}(T^*)\\
&= \mu \cdot \langle x, y\rangle && \text{by (IPS 2').}
\end{aligned}$$

But $\lambda \neq \mu$ forces the other factor $\langle x, y\rangle$ to be zero. □

4.4.12. Example. Eigenspaces of a symmetric or skew-symmetric real matrix $A \in \mathrm{M}_{n\times n}(\mathbb{R})$ for different eigenvalues are orthogonal subspaces of $\mathbb{R}^n$, with its usual inner product.

4.5. Spectral Theorem

We can now prove a very useful result. As everywhere in this chapter, the field $\mathbb{F}$ is $\mathbb{R}$ or $\mathbb{C}$. Actually, the story will be slightly different for each case.

Let us isolate the property we are after, which is an improvement of being diagonalizable (Definition 3.1.11).

4.5.1. Definition. We say that an operator $T\colon V \to V$ on a finite-dimensional inner product space is *orthogonally diagonalizable* if V admits an orthonormal basis B such that $[T]_B$ is diagonal, *i.e.* V admits an orthonormal basis consisting of eigenvectors of T.

4.5.2. Remark. If T is orthogonally diagonalizable then choose an orthonormal basis B such that $[T]_B = D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ is diagonal. Since B in an orthonormal basis, the matrix of the adjoint $[T^*] = D^*$ is the adjoint of D, that is, $D^* = \operatorname{diag}(\bar{\lambda}_1, \ldots, \bar{\lambda}_n)$. It is easy to verify that $D^*D = DD^*$ and therefore $T^*T = TT^*$. So T is normal.

In short, orthogonally diagonalizable operators have to be normal.

Over the complex numbers the converse is true. To prove this, we can invoke a very general result, that we could already prove with the techniques of Chapter 3.

4.5.3. Lemma. *Suppose that $T\colon V \to V$ has a completely split characteristic polynomial (which is no condition for $\mathbb{F} = \mathbb{C}$). Then V admits an orthonormal basis B in which the matrix of T is upper-triangular: $[T]_B = \begin{pmatrix} \star & \star & \cdots & \star \\ 0 & \star & & \star \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \star \end{pmatrix}$.*

Proof. First note that V admits a basis C in which $[T]_C$ is upper-triangular. This follows immediately from the Jordan canonical form Theorem 3.6.4 (applied to T^*, so that we get *upper*-triangular for T). But we can also verify this directly by induction on $n = \dim(V)$. Since the characteristic polynomial $P_T(X) \in \mathbb{F}[X]$ is completely split, T admits at least one eigenvalue $\lambda \in \mathbb{F}$. We take $v_1 \in E_\lambda(T)$ a non-zero eigenvector. We can complete $v_1 \neq 0$ into a (temporary) basis $\{v_1\} \cup C'$ of V for some complement $C' = \{v_2', \ldots, v_n'\}$. So far, the matrix of T in the basis $\{v_1\} \cup C'$ has the following form:

$$A := \begin{pmatrix} \lambda & \star \\ 0 & A' \end{pmatrix}$$

where $A' \in \mathrm{M}_{(n-1)\times(n-1)}(\mathbb{F})$. We already used the argument that in this situation $P_T(X) = P_A(X) = (\lambda - X) \cdot P_{A'}(X)$ via Exercise D.2.8. Therefore $P_{A'}$ is also completely split. By induction hypothesis, $V' = \operatorname{Span}(v_2', \ldots, v_n')$ has a(nother) basis $v_2, \ldots, v_n$ in which $T_{A'}$ is upper-triangular. In the new basis $\{v_1, v_2, \ldots, v_n\}$ of V, the matrix of T is now upper-triangular.

The fact that the matrix $[T]_C$ in the basis $C = \{v_1, \ldots, v_n\}$ is upper-triangular exactly means that all the subspaces $\operatorname{Span}(v_1)$, $\operatorname{Span}(v_1, v_2)$, $\operatorname{Span}(v_1, v_2, v_3), \ldots$ are all T-invariant, *i.e.*

$$T\big(\operatorname{Span}(v_1, \ldots, v_m)\big) \subseteq \operatorname{Span}(v_1, \ldots, v_m)$$

for all $m = 1, \ldots, n$.

To get the orthonormal basis $B = \{b_1, \ldots, b_n\}$ it suffices to apply Gram-Schmidt to any basis $C = \{v_1, \ldots, v_n\}$ in which $[T]_C$ is upper-triangular. Indeed, the Gram-Schmidt process is such that $\operatorname{Span}(b_1, \ldots, b_m) = \operatorname{Span}(v_1, \ldots, v_m)$ for all $m = 1, \ldots, n$. Hence $[T]_B$ is upper-triangular as well. □

We can now prove our main result.

4.5.4. Theorem (Complex Spectral Theorem)**.** *Let $T\colon V \to V$ be a linear operator on a finite-dimensional inner product space* over the field of complex numbers $\mathbb{F} = \mathbb{C}$. *Then T is normal if and only if it is orthogonally diagonalizable.*

Proof. By Lemma 4.5.3, we have an orthonormal basis B such that $A := [T]_B$ is upper-triangular. On the other hand, since T is normal, we have $A^* \cdot A = A \cdot A^*$ by Proposition 4.4.4. So it suffices to prove the following Claim:

If $A \in \mathrm{M}_{n\times n}(\mathbb{C})$ is upper-triangular and normal then A is diagonal.

(The reader should appreciate this point: We do not need to change the orthonormal basis B obtained from Lemma 4.5.3.)

We prove this Claim by induction on n. Since A is upper-triangular, the first canonical vector $e_1 \in \mathbb{C}^n$ is an eigenvector for the eigenvalue A_{11}. Now, since A is normal, it follows from Corollary 4.4.10 that e_1 must also be an eigenvector for A^*, with eigenvalue $\bar{A}_{11}$. This means that the first column of A^* is $\begin{pmatrix} \bar{A}_{11} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, or in other words, transposing back, the first row of A is $\left(A_{11}\ 0 \cdots\ 0 \right)$. Thus A looks as follows:

$$A = \begin{pmatrix} A_{11} & 0 \\ 0 & A' \end{pmatrix}$$

for $A' \in \mathrm{M}_{(n-1)\times(n-1)}(\mathbb{C})$. Compute $A^* = \begin{pmatrix} \overline{A_{11}} & 0 \\ 0 & (A')^* \end{pmatrix}$. A direct computation of the relation $A \cdot A^* = A^* \cdot A$ in blocks gives us that A' is also normal: $(A')^* \cdot A' = A' \cdot (A')^*$. By induction hypothesis, A' is diagonal, hence so is A. □

4.5.5. Corollary. *Let $T\colon V \to V$ be a linear operator on a finite-dimensional inner product space over the complex numbers. Suppose that T is self-adjoint $T = T^*$. Then all eigenvalues of T are real and there exists an orthonormal basis of V such that $[T]_B$ is diagonal and real.*

Proof. Since T is normal, Theorem 4.5.4 gives us an orthonormal basis B such that $[T]_B = D = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$ is diagonal. But then $D = [T]_B = [T^*]_B = D^* = \bar{D}$ since B is orthonormal and $T^* = T$. Thus $\bar{\lambda}_i = \lambda_i$ for all $i = 1 \dots, n$. □

4.5.6. Corollary. *Let $A \in \mathrm{M}_{n\times n}(\mathbb{R})$ be a real matrix that is symmetric $A = A^t$. Then its characteristic polynomial is completely split over $\mathbb{R}$.*

Proof. Consider $T = T_A\colon \mathbb{C}^n \to \mathbb{C}^n$ the *complex* linear operator associated to the same matrix (for the usual inner product on $\mathbb{C}^n$). By Corollary 4.5.5 all the eigenvalues of T are real, *i.e.* the real polynomial $P_A(X)$, that can be written $P_A(X) = (-1)^n(X - \lambda_1)\cdots(X - \lambda_n)$ for $\lambda_1, \dots, \lambda_n \in \mathbb{C}$, has only real roots: all $\lambda_i \in \mathbb{R}$. Hence $P_A(X) = (-1)^n(X - \lambda_1)\cdots(X - \lambda_n)$ is split in $\mathbb{R}[X]$ already. □

We can use the complex spectral theorem to prove the following version.

4.5.7. Theorem (Real Spectral Theorem)**.** *Let $T\colon V \to V$ be a linear operator on a finite-dimensional inner product space* over the field of real numbers $\mathbb{F} = \mathbb{R}$. *Then T is orthogonally diagonalizable if and only if T is self-adjoint $T = T^*$. In particular, the characteristic polynomial of a self-adjoint real operator is completely split over $\mathbb{R}$.*

Proof. If T admits an orthonormal basis B in which $[T]_B = D$ is diagonal, say $D = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$ with *real* scalars $\lambda_1, \dots, \lambda_n$ then by Theorem 4.3.3 (2) we have $[T^*]_B = D^* = D$ and therefore $T^* = T$. Hence T is self-adjoint.

Conversely, suppose that $T = T^*$ is self-adjoint. Let C be any orthonormal basis of V and $A = [T]_C$. By Theorem 4.3.3 (2) we have $A^t = A^* = [T^*]_C = [T]_C = A$, *i.e.* A is symmetric. By Corollary 4.5.6, the characteristic polynomial $P_T(X) = P_A(X)$ is completely split over $\mathbb{R}$. We can now forget C and use Lemma 4.5.3. There exists an orthonormal basis B of V such that $D = [T]_B$ is upper-triangular. But on the other hand $D^t = D^* = [T^*]_B = [T]_B = D$ and therefore D is symmetric. But symmetric and upper-triangular means diagonal. □

4.5.8. Example. Recall our old friend $A = \left(\begin{smallmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{smallmatrix}\right) \in \mathrm{M}_{2\times 2}(\mathbb{R})$. We know that if α is not an integral multiple of π, the matrix A has no real eigenvalue. In particular it is not diagonalizable. A fortiori, it is <u>*not*</u> orthogonally diagonalizable over the real numbers! And indeed, $A^t \neq A$ unless $\sin(\alpha) = 0$ which is precisely the $\alpha \in \left\{ k \cdot \pi \,\middle|\, k \in \mathbb{Z} \right\}$ that we excluded.

However, we have $A^t = A^{-1}$. Indeed, A^t is the rotation of angle $-\alpha$, or one can directly verify $A^t \cdot A = I_2 = A \cdot A^t$. And in particular, A is *normal* as a complex matrix: $A^* \cdot A = A^t \cdot A = A \cdot A^t = A \cdot A^*$. This means that A is orthogonally diagonalizable *over the complex numbers.*

4.5.9. Exercise. Find an orthonormal basis of $\mathbb{C}^2$ in which the matrix of T_A is diagonal, where $A = \left(\begin{smallmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{smallmatrix}\right)$ and $\alpha \in \mathbb{R}$. Write $A = Q \cdot D \cdot Q^*$ for Q invertible (and unitary) and D diagonal.

Matrix forms of the spectral theorems.

We can apply the above results to the usual inner product spaces of Example 4.1.7 and the linear transformation T_A given by multiplication by a fixed square matrix. Recall that $(T_A)^* = T_{A^*}$ where $A^* = \overline{A}^t$.

In matrix form, Theorem 4.5.4 becomes:

4.5.10. Corollary. *Let $A \in \mathrm{M}_{n\times n}(\mathbb{C})$ be a complex matrix. The following are equivalent:*

(i) *A is normal, i.e. $A^* \cdot A = A \cdot A^*$.*
(ii) *A is orthogonally diagonalizable, i.e. there exists $Q \in \mathrm{M}_{n\times n}(\mathbb{C})$ unitary (i.e. invertible and $Q^{-1} = Q^*$) and $D \in \mathrm{M}_{n\times n}(\mathbb{C})$ diagonal such that $A = Q \cdot D \cdot Q^*$.*

In matrix form, Corollary 4.5.5 becomes:

4.5.11. Corollary. *Let $A \in \mathrm{M}_{n\times n}(\mathbb{C})$ be a complex matrix. The following are equivalent:*

(i) *A is self-adjoint, i.e. $A^* = A$.*
(ii) *A is orthogonally diagonalizable with only real eigenvalues, i.e. there exists $Q \in \mathrm{M}_{n\times n}(\mathbb{C})$ unitary and $D \in \mathrm{M}_{n\times n}(\mathbb{R})$ diagonal and real such that $A = Q \cdot D \cdot Q^*$.*

In matrix form, Theorem 4.5.7 becomes:

4.5.12. Corollary. *Let $A \in \mathrm{M}_{n\times n}(\mathbb{R})$ be a real matrix. The following are equivalent:*

(i) *A is self-adjoint, i.e. $A^t = A$.*
(ii) *A is orthogonally diagonalizable, i.e. there exists $Q \in \mathrm{M}_{n\times n}(\mathbb{R})$ orthogonal (i.e. invertible and $Q^{-1} = Q^t$) and $D \in \mathrm{M}_{n\times n}(\mathbb{R})$ diagonal such that $A = Q \cdot D \cdot Q^t$.*

4.5.13. Remark. In all cases, Condition (i) is very easy to verify. Also note that a matrix can remain diagonalizable in the sense of Chapter 3 without being *orthogonally* diagonalizable. So the simple Condition (i) is not necessary for ordinary diagonalization.

4.5.14. Remark. Once we verify Condition (i), we can find Q and D explicitly. We find the eigenvalues $\lambda_1, \ldots, \lambda_r$, say all distinct, by splitting the characteristic polynomial. We determine the eigenspaces $E_{\lambda_i}(A) = \mathrm{Ker}(A - \lambda_i \cdot I_n)$ by solving the linear system $(A - \lambda_i \cdot I_n) \cdot x = 0$. Once we have a basis of $E_{\lambda_i}(A)$ we apply Gram-Schmidt (Theorem 4.2.13) to find an orthonormal basis B_i of $E_{\lambda_i}(A)$. Then, automatically, $B = B_1 \cup \cdots \cup B_r$ is an orthonormal basis of $\mathbb{F}^n$ consisting of eigenvectors. (This is of course only true under the assumption that we do know that A is orthogonally diagonalizable.) Write the vectors $B = \{b_1, \ldots, b_n\}$ in a matrix $Q = [b_1 | \cdots | b_n]$. This Q is orthogonal and $A \cdot Q = Q \cdot D$ where D is the diagonal matrix whose (i, i)-entry is the eigenvalue of b_i as usual. This equation $A \cdot Q = Q \cdot D$ means $A = Q \cdot D \cdot Q^{-1} = Q \cdot D \cdot Q^t$. In other words, $Q = Q_{C,B}$ is the change-of-coordinates matrix from the orthonormal basis of eigenvectors B to the canonical basis C and $[T_A]_B = Q_{C,B} \cdot [T_A]_C \cdot Q_{C,B} = Q^{-1} \cdot A \cdot Q = Q^t \cdot A \cdot Q = D$ is diagonal.

APPENDIX A

Notations

The symbol $\in$ means 'element of'. For instance $x \in A$ means that A is a set and that x is an element of that set. Of course, if $\mathbb{F}$ is a field then $a \in \mathbb{F}$ means that a is a scalar (an element of the set $\mathbb{F}$); and if V is a vector space then $x \in V$ means that x is a vector. We do not reinvent new notation for $\in$ in each case.

The symbol $\subseteq$ means 'is contained in', for sets (or fields, vector spaces, etc). So $A \subseteq B$ simply means that A is a subset of the set B. We reserve $A \subset B$ or $A \subsetneq B$ for proper subsets, when we insist that moreover $A \neq B$.

The symbol $\forall$ means "for all". The symbol $\exists$ means "there exists". (These symbols might be used on the blackboard, for speed, but not in these notes.)

Let $A \subseteq B$ be a subset. We write $B \smallsetminus A = \left\{\, b \in B \,\middle|\, b \notin A \,\right\}$ for its complement.

Functions are denoted by expressions of the following type

$$\begin{aligned} f\colon A &\longrightarrow B \\ a &\longmapsto f(a). \end{aligned}$$

This formula means that A and B are sets and that f is the data of a chosen element $f(a)$ in B for every $a \in A$. (Try to distinguish the notation $\to$ at the level of sets and $\mapsto$ at the level of elements. This prevents confusion in tricky situations.)

The set notation $\left\{\, \text{blah} \,\middle|\, \text{blih} \,\right\}$ should be read as: *the set of all elements of the form 'blah' such that 'blih'*. The two most common uses are:

The set $\left\{\, x(a) \,\middle|\, a \in A \,\right\}$ represents the collection of all elements of the (given) form $x(a)$, where a runs through the entire set A. For instance, $\left\{\, x + y\,i \,\middle|\, x, y \in \mathbb{R} \,\right\}$ consists of all expressions $x + y\,i$ for every possible choice of x and y in $\mathbb{R}$.

The set $\left\{\, x \in B \,\middle|\, P(x) \,\right\}$ where P is a (given) property of x represents the collection of all elements x in the (given) set B such that the property $P(x)$ is true. For instance, $\left\{\, x \in \mathbb{C} \,\middle|\, x^2 = -1 \,\right\}$ is the set with two elements $\{i, -i\}$.

Formally, the latter is the correct one. Instead of the former $\left\{\, x(a) \,\middle|\, a \in A \,\right\}$ one should give a function $x\colon A \to B$ and then our $\left\{\, x(a) \,\middle|\, a \in A \,\right\}$ means the image of that function, that is, $\left\{\, b \in B \,\middle|\, \text{there exists } a \in A \text{ such that } x(a) = b \,\right\}$.

A.1.1. Notation. It is sometimes convenient to use the *Kronecker symbol*

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

APPENDIX B

Sets, functions, bijections

We use sets in the intuitive way, thinking of a set as a 'collection' of elements. ([1]) As mentioned in Appendix A, we write $a \in A$ to say that a is an element of A. We write $A \subseteq B$ to say that the set A is contained in the set B, *i.e.* that every element of the first one is also an element of the second: $\forall\, a \in A$, we have $a \in B$.

A *function* $f\colon A \to B$ assigns to every element $a \in A$ an element $f(a)$ in B. The sets A and B are often called the *source* and *target* of f. We can *compose* functions $f\colon A \to B$ and $g\colon B \to C$, when the target of the first is the source of the second (here B), to define a new function $g \circ f\colon A \to C$ via $(g \circ f)(a) = g(f(a))$ for every $a \in A$. The *identity* is the function $\mathrm{Id}_A\colon A \to A$ defined by $\mathrm{Id}_A(a) = a$ for all a. We have $f \circ \mathrm{Id}_A = f$ and $\mathrm{Id}_B \circ f = f$ for any $f\colon A \to B$.

Giving a function $f\colon A \to B$ is *not* saying that every $b \in B$ is assigned to some a. When the latter happens we say that f is *surjective* (or *onto*), *i.e.* when for every $b \in B$ there exists $a \in A$ with $f(a) = b$.

More generally, the *image* of f is the subset $\mathrm{Im}(f) = f(A) = \big\{\, f(a) \,\big|\, a \in A \,\big\} = \big\{\, b \in B \,\big|\, \exists a \in A \text{ s.t. } f(a) = b \,\big\}$ of B. So f is surjective if and only if $\mathrm{Im}(f) = B$.

Another misconception is to think that different elements a in A necessarily go to different $f(a)$ in B. This is a property of f, called being *injective* (or *one-to-one*), *i.e.* for every $a \neq a'$ in A we have $f(a) \neq f(a')$ in B. Note that this is tautologically equivalent to saying that: whenever $f(a) = f(a')$ we must have $a = a'$.

Injective and surjective are *not* each other's contrary. The function $f\colon \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is neither injective nor surjective.

A function $f\colon A \to B$ that is both injective and surjective is called *bijective*. We also say that f is a *bijection* between A and B. ([2])

If a function $f\colon A \to B$ is a bijection then for every $b \in B$ there exists at least one $a \in A$ with $f(a) = b$ by surjectivity, and there is at most one such a by injectivity. So f bijective means that for each $b \in B$ *there exists a unique* $a \in A$ such that $f(a) = b$. We call it *the preimage* of b and denote it $f^{-1}(b)$. In summary, $f^{-1}(b)$ is the unique element of A such that $f(f^{-1}(b)) = b$. This defines a new function $f^{-1}\colon B \to A$ in the opposite direction of f such that $f \circ f^{-1} = \mathrm{Id}_B$. For

[1] Formally, one can invoke the Zermelo-Fraenkel Axioms, typically including the Axiom of Choice. See for instance `https://en.wikipedia.org/wiki/Zermelo-Fraenkel`

[2] It is relatively easy to remember those names, surjective, injective, bijective. The prefix 'sur-' means 'over, above'. Being *sur*jective expresses the fact that A has enough elements to cover the whole of B via f. The word 'onto' expresses the same idea. On the other hand, the prefix 'in-' refers here the idea of 'inside, within'. Being *in*jective expresses the fact that A might be thought of as a subset of B via f. In particular, a subset $A \subseteq B$ yields an injection $A \hookrightarrow B$ simply mapping every $a \in A$ to itself $a \in B$. (The expression 'one-to-one' here might be slightly misleading, as not every 'one' in B needs to be reached by anyone and there never was any question that one $a \in A$ could have more than one image in B. So...) Finally *bi*jective. Well, nowadays we are all used to 'bi' meaning 'both'.

each $a \in A$ we also have $f^{-1}(f(a)) = a$ by applying the construction of f^{-1} to $b = f(a)$. In other words, we also have $f^{-1} \circ f = \mathrm{Id}_A$.

One can easily verify that the existence of such a map backwards $f' \colon B \to A$ with $f' \circ f = \mathrm{Id}_A$ and $f \circ f' = \mathrm{Id}_B$ is equivalent to f being a bijection. Indeed, if $f(a) = f(a')$ then $a = f'(f(a)) = f'(f(a')) = a'$, shows that f is injective and $b = f(f'(b))$ for any $b \in B$ shows that f is surjective. In that case, $f' = f^{-1}$ necessarily.

B.1.1. Exercise. (1) Show that if $f \colon A \to B$ is a bijection then so is $f^{-1} \colon B \to A$ and that $(f^{-1})^{-1} = f$.

(2) Let $f \colon A \to B$ and $g \colon B \to C$ be composable functions. Show that if two among $f, g, g \circ f$ are bijections then so is the third. Give in each case a formula for the inverse of this third function in terms of the inverses of the first two.

B.1.2. Remark. An *equivalence relation* on a set X consists of a subset of $X \times X$, that is, a collection of chosen pairs (x, y) that we understand as being the pairs of 'equivalent' elements x and y of the set X. Instead of saying that (x, y) is such a chosen pair, we write $x \sim y$. These pairs must satisfy three axioms:

- *Reflexivity*: every element $x \in X$ is equivalent to itself: $x \sim x$.
- *Symmetry*: Whenever $x \sim y$ holds then $y \sim x$ holds as well.
- *Transitivity*: Whenever $x \sim y$ and $y \sim z$ then $x \sim z$.

The simplest example of $\sim$ is simply equality: Declare $x \sim y$ when $x = y$. Another somewhat silly example is to declare every elements equivalent: Declare $x \sim y$ for all $x, y \in X$. We discuss many more examples in the text.

An important construction associated to an equivalence relation is to construct a new set $X/{\sim}$ together with a map $\pi \colon X \to X/{\sim}$ such that $x \sim y$ if and only if $\pi(x) = \pi(y)$. In other words, we turn the equivalence relation $\sim$ into the simplest one, equality. The construction of $X/{\sim}$ is as a subset of the set of subsets of X. Namely, we write $[x]_\sim = \left\{ y \in X \,\middle|\, y \sim x \right\}$ for the *equivalence class of* x and we define $X/{\sim} = \left\{ [x]_\sim \,\middle|\, x \in X \right\}$ as the set of these equivalence classes. The map $\pi \colon X \to X/{\sim}$ simply maps x to $[x]_\sim$. It is clearly surjective and by definition $x \sim y$ if and only if $[x]_\sim = [y]_\sim$, that is, if and only if $\pi(x) = \pi(y)$.

Let us see what $X/{\sim}$ is, in our two extreme examples. First if $\sim$ is equality then $X/{\sim}$ is just X itself. Indeed, for every $x \in X$ we have $[x]_= = \{x\}$: Only x itself is equivalent (equal) to x in this case. Second, if $\sim$ is the silly equality $x \sim y$ for all x, y. Then for each x, we have $[x]_\sim = X$. So there is only one equivalence class, the whole of X, and therefore $X/{\sim}$ is a set with a single element.

APPENDIX C

Techniques of proof

C.1. No!

Logic is a delicate matter, that deserves its own course. We use here only rudimentary methods, that are sufficiently intuitive to warrant a light treatment.

One topic has however proved difficult for some students: The meaning of the *negation* of a statement S. The pitfall comes from folksy logic and partisan argumentation. Namely, do not use the 'extreme opposite' of S instead of its exact negation. For instance, consider a subset $A \subseteq \mathbb{N}$ of integers and suppose that your statement $S(A)$ is "all numbers in A are even". It depends on A. Say, if $A = \{2, 4, 8\}$ the statement is true. For $A = \{2, 5, 8\}$ the statement is false. The negation of statement $S(A)$ is "at least one number in A is odd". It is *not* "all numbers in A are odd". The negation of "Bob is a Democrat" is not "Bob is a Republican"; it is "Bob is not a Democrat" (Bob could be an Independent, be apolitical, or else). Jokes about mathematicians are written around this point.

So, how to construct the exact negation like a robot? The sticky point is what to do with *quantifiers*, "for all" and "there exists":

The negation of "for all $x \in A$, property $P(x)$ is true" is "*there exists* $x \in A$ such that $P(x)$ is not true". The "for all" has become "there exists".

And vice versa. The negation of "there exists $x \in A$, such that property $P(x)$ is true" is "*for all* $x \in A$ property $P(x)$ is not true".

For instance, we saw above that (for some given subset $A \subseteq \mathbb{N}$ of integers) the negation of "for all $x \in A$ the number x is even" is "there exists some $x \in A$ such that the number x is not even (odd)". Beware also that the $\in A$ next to the quantifier is not turned in $\notin A$ just because of the negative vibe.

C.1.1. Exercise. Let A, B be two subsets of larger set C. What is the exact negation of "A is contained in B"? Verify your answer by applying negation-like-a-robot to the statement "for all $x \in A$, we have $x \in B$" (expanded form of "$A \subseteq B$").

C.2. Reasoning *ab absurdo*

In a situation where you wonder if some statement called S is true, you can prove it by assuming that its *exact negation* is true and deriving a contradiction. This means that our assumption was wrong, as it forced a contradiction. In other words, the exact negation of S was wrong, so S was true. These are proofs by contradiction, or *ab absurdo*, from the absurd. We will see many examples of this, starting with the justification of the induction method below. ([1])

[1] The postmodernist may wonder why a contradiction is such a big deal. Technically, it is not. And there is no proof that mathematics contains no contradiction! So far, none has appeared though. *If* there was a contradiction in mathematics, absolutely *all* statements would be wrong and, amusingly, all statements would be true as well. The world would be very boring.

C.2.1. Example. If the square n^2 of a number $n \in \mathbb{Z}$ is even then n is even. Indeed, suppose *ab absurdo* that n is not even. Then it is odd, hence $n = 2m + 1$ for some $m \in \mathbb{Z}$. Then $n^2 = 4m^2 + 4m + 1 = 2(2m^2 + 2m) + 1$ is odd, a contradiction with the hypothesis that n^2 was even. Something was wrong: The absurd assumption that n would not be even. So indeed n is even, as claimed.

C.2.2. Example. Let us prove by contradiction that the real number $\sqrt{2}$ cannot be a rational number. Suppose *ab absurdo* that we can write $\sqrt{2} = \frac{a}{b}$ for some $a, b \in \mathbb{Z}$, with $b \neq 0$ (and $a \neq 0$ of course). If a and b are both even, say $a = 2a'$ and $b = 2b'$ then we can cancel out a copy of 2 in this fraction $\frac{a}{b} = \frac{a'}{b'}$ and repeat. After canceling out all possible copies of 2, we can assume that we have $\sqrt{2} = \frac{a}{b}$ where at least one of a and b is odd (perhaps both). Then $\sqrt{2} = \frac{a}{b}$ gives $b\sqrt{2} = a$. Raising to the square we see that $b^2 \cdot 2 = a^2$. Hence a^2 is an even number. This forces a to be even (Example C.2.1). So we can write $a = 2a'$ and our equation becomes $b^2 \cdot 2 = a^2 = 4 \cdot (a')^2$. In particular $b^2 = 2 \cdot (a')^2$ is an even number. Hence b is even (Example C.2.1 again). This is a contradiction since a and b cannot be both even. Something was wrong: The absurd assumption that $\sqrt{2}$ was rational. Therefore $\sqrt{2}$ is *not* rational. It is a number in $\mathbb{R}$, not in $\mathbb{Q}$.

An important logical trick is *contraposition*: If you know that property P implies property Q then you also know that non-Q implies non-P. Indeed, you can show that *ab absurdo*. Suppose that non-Q is true and let us show non-P is true. If *ab absurdo*, this was false, that is, if non-(non-P)=P was true then Q would be true too, by our assumption that P implies Q. In summary, non-Q would be true and Q would be true too, a contradiction. What was absurd? The assumption that non-P was false. So non-P is true.

Here is an example, in the style of Example C.2.2. We can show that every even number n has an even square n^2. (That is very easy: If $n = 2m$ is even then $n^2 = 4m^2 = 2(2m^2)$ is even.) In other words, P=(n is even) implies Q=(n^2 is even). It follows that non-Q=(n^2 is odd) implies non-P=(n is odd). In words, if the square of a number n^2 is odd then the number n was odd.

C.3. Induction

A mathematical statement, let's call it $S(n)$, can depend on an integer n. For instance, $S(n) =$"*For any* $n \geq 1$, *the sum* $1+2+3+\cdots+n$ *is equal to* $\frac{n(n+1)}{2}$." This statement depends on n. We might know it for some n, not for others: You can do it in your head for $n = 4$ but you would be hard-pressed to do it for $n = 10^{80}$. [2]

One can prove such statements $S(n)$ *by induction on* n in two steps:

(1) *Starting point*: We prove that the statement $S(n_0)$ is true for a particular number n_0.
(2) *Induction step*: Assuming that $S(n-1)$ is true for some number $n \geq n_0 + 1$, we prove that $S(n)$ is true as well.

[2] Legend has it that Gauss found a proof, as a child. Call $X = 1 + 2 + \cdots + (n-1) + n$ and note that $X = n + (n-1) + \cdots + 2 + 1$. Write one sum underneath the other and add 'vertically'. You get $X + X = (1+n) + (2+(n-1)) + \cdots + ((n-1)+2) + (n+1) = (n+1) + \cdots + (n+1)$ and the latter sum has n terms. So $2 \cdot X = n \cdot (n+1)$ hence the result. Darf ich jetzt spielen gehen?

If we can do both of these steps, we can conclude that *all* the statements $S(n)$ are true for all $n \geq n_0$. The proof of (2) typically uses $S(n-1)$ somewhere. In that setting, we call $S(n-1)$ the *induction hypothesis.*

Let us justify this. If *ab absurdo* there were some bad $n \geq n_0$ such that $S(n)$ was false, choose the *smallest* one. Call it n_{bad}. So $S(n_{\text{bad}})$ is false and n_{bad} is the smallest $n \geq n_0$ such that $S(n)$ is false. By (1), n_{bad} cannot be n_0. So $n_{\text{bad}} \geq n_0+1$. And since n_{bad} is the smallest bad n, we know that $n_{\text{bad}} - 1$ is not bad, that is, $S(n_{\text{bad}} - 1)$ is true. But then (2) tells us that $S(n_{\text{bad}})$ is true, a contradiction.

Consider our example where $S(n)$ was the statement that $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$. We claim that $S(n)$ is true for all $n \geq 1$. Here we want $n_0 = 1$. We can check $S(1)$ that says $\sum_{k=1}^{1} k = 1 = \frac{1 \cdot 2}{2}$. This gives us the starting point (1) for $n_0 = 1$. Now, let us prove the induction step (2). We suppose that for some $n \geq 2$ we already know that $S(n-1)$ is true, and we want to prove $S(n)$. To do this, we compute the sum by pausing at $n-1$ and then adding the last term at the end (we skip some easy steps)

$$\sum_{k=1}^{n} k = \big(\sum_{k=1}^{n-1} k\big) + n = \frac{(n-1)((n-1)+1)}{2} + n = \frac{n^2-n}{2} + \frac{2n}{2} = \frac{n(n+1)}{2}.$$

Note that the second equality uses the induction hypothesis $S(n-1)$ to replace $1+2+\cdots+(n-1)$ by the formula we are proving, but in the case where we are allowed to! We conclude by induction on n, that $S(n)$ is true for all $n \geq 1$.

C.3.1. Exercise. Prove by induction on n that the sum of the first n squares $1+4+9+\cdots+n^2 = \sum_{k=1}^{n} k^2$ is $\frac{n(n+1)(2n+1)}{6}$.

C.3.2. Exercise. Let $W \subseteq V$ be a subset of an $\mathbb{F}$-vector space (take $V = \mathbb{R}^n$ if you have not read Section 1.2 yet). Suppose that W is closed under addition, *i.e.* for every $x, y \in W$ we have $x+y \in W$. Show that W is closed under finite sums: For every $n \geq 2$ and every $x_1, \ldots, x_n \in W$ we have $x_1 + \cdots + x_n \in W$.

C.3.3. Remark. After a while, one gets very used to induction and one tends to omit the verification of the usually easy (1). This can lead to mistakes. Also, in some problems, the induction proof only works for n_0 quite large and one then needs to check many small cases, between 1 and n_0 for instance, if we want to have a statement valid for all $n \geq 1$. Here is a silly cautionary example.

C.3.4. Exercise. Consider the statement $S(n) =$"In a classroom with n students, all students have the same name." OK, let us test. For $n = 0$ it is correct but borderline. So let us test with $n = 1$: Yes, a classroom with one student has this property. Let us pass now to the induction argument. Take n large and suppose we knew that in every classroom with $n-1$ students all students have the same name and let us consider a classroom with n students. Ask one student among the n to leave the room. You then have a classroom with $n-1$ students, all with same name by induction hypothesis. Now the question is why is the name of the person outside the room the same as the name of the other $n-1$? Simple! Ask that person to come back in and ask *another* one to leave. Again, you are down to a classroom with $n-1$ students, all with the same name by induction hypothesis, so the first person that went out has the same name as everybody else and we are done.

On the other hand, you might know that $S(n)$ is false from experience. Do we have a contradiction in mathematics (yippee!) or did we do something wrong?

C.3.5. Remark. There are variants of the induction method, where you need $S(n-1)$ and $S(n-2)$ to prove $S(n)$. In that case, the starting point should contain *two* consecutive verifications, $S(n_0)$ and $S(n_0+1)$, so that the induction step can start with proving $S(n_0+2)$ from those two, etc. And so on, with three or more.

At the extreme, it is quite common to need $S(m)$ for *all smaller* m, say, in the range $n_0 \leq m < n$, as a kind of stronger induction hypothesis (more information) in order to prove $S(n)$. In that case, modified induction proceeds as follows:

(1') *Starting Point*: Prove that $S(n_0)$ is true. (It's the same as before.)

(2') *Modified Induction Step*: Prove for every $n \geq n_0+1$ that if $S(m)$ is true for all $n_0 \leq m < n$ then $S(n)$ is true. (Before, we only used $m = n-1$.)

The conclusion is that $S(n)$ is true for all $n \geq n_0$ exactly as before. One can give the same proof as above (hit the smallest bad n), or deduce this modified induction from the first induction method for a new statement S' defined as follows. Let $S'(n)$ be "$S(m)$ is true for all $n_0 \leq m \leq n$". So $S'(n_0) = S(n_0)$ and $S'(n-1)$ is the modified induction hypothesis where we assume all previous $S(m)$. As $S'(n)$ has only $S(n)$ as new information compared to $S'(n-1)$ we know the induction step for S' thanks to (2'). By induction on n, we know that $S'(n)$ is true for all $n \geq n_0$ which is the same information as knowing $S(n)$ is true for all $n \geq n_0$. It is not needed to make this argument with S' for this situation. But it is a useful mathematical trick to make a statement 'richer' so that it passes each induction step more easily. Proving more is sometimes simpler.

APPENDIX D

Reminder from Math 33A

[Prerequisites. Will TeX more if time permits.]

D.1. Polynomials

D.1.1. Definition. Recall that a *root* (or a *zero*) of a polynomial $P(X) \in \mathbb{F}[X]$ is a scalar $\lambda \in \mathbb{F}$ such that $P(\lambda) = 0$.

We claim that

$$P(\lambda) = 0 \quad \Longrightarrow \quad P(X) = (X - \lambda) \cdot Q(X) \tag{D.1.2}$$

for a unique polynomial $Q \in \mathbb{F}[X]$. Note that if $P \neq 0$ this forces $\deg(Q) = \deg(P) - 1$. To see why (D.1.2) holds, one can use the usual division algorithm but we can also use linear algebra.

Let $\lambda \in \mathbb{F}$ and $d \geq 1$. Consider the $\mathbb{F}$-vector spaces V and W and the linear transformations $T\colon V \to \mathbb{F}$ and $S\colon W \to V$ defined as follows:

$$\begin{array}{rccc} T\colon & V := \left\{ P \in \mathbb{F}[X] \,\middle|\, \deg(P) \leq d \right\} & \longrightarrow & \mathbb{F} \\ & P & \longmapsto & P(\lambda) \end{array}$$

evaluation at λ, and multiplication by $(X - \lambda)$

$$\begin{array}{rccc} S\colon & W := \left\{ Q \in \mathbb{F}[X] \,\middle|\, \deg(Q) \leq d - 1 \right\} & \longrightarrow & V \\ & Q & \longmapsto & (X - \lambda) \cdot Q. \end{array}$$

It is easy to verify that S and T are linear. The claim is that S restricts to an isomorphism $S\colon W \xrightarrow{\sim} \mathrm{Ker}(T)$. Note that indeed $T(S(Q)) = (\lambda - \lambda)Q(\lambda) = 0$. So at least $\mathrm{Im}(S) \subseteq \mathrm{Ker}(T)$. Since $T(a_0) = a_0$ for every $a_0 \in \mathbb{F}$ we see that T is surjective. By Rank-Nullity Theorem 2.2.6 it follows that $\dim(\mathrm{Ker}(T)) = \dim(V) - 1 = (d + 1) - 1 = d$. On the other hand, S is clearly injective: If $(X - \lambda) \cdot Q = 0$ then the leading term of Q cannot be non-zero: If $Q = b_0 + b_1 X + \cdots + b_e X^e$ with $b_e \neq 0$ then $(X - \lambda)Q = b_e X^{e+1}$+terms of lower degree, hence is non-zero. By Rank-Nullity again, it follows that $\dim(\mathrm{Im}(S)) = \dim W = d$. Therefore both subspaces $\mathrm{Im}(S) \subseteq \mathrm{Ker}(T)$ have the same dimension d hence they are equal. In summary, S is a bijection $W \xrightarrow{\sim} \mathrm{Im}(S) = \mathrm{Ker}(T)$. This proves (D.1.2).

One can repeat (D.1.2): If $Q(\lambda) = 0$ as well then $Q = (X - \lambda)R$ and $P = (X - \lambda)^2 R$. Etc. Since $\deg(P) > \deg(Q) > \deg(R) > \cdots \geq 0$ this process must stop after at most $\deg(P)$ steps and we get (up to renaming Q the polynomial where this stops): If $P(\lambda) = 0$ then there exists a unique polynomial Q such that

$$P(X) = (X - \lambda)^m \cdot Q(X) \qquad \text{with} \qquad Q(\lambda) \neq 0.$$

The number $m \geq 1$ is called the *multiplicity* of the root λ.

Note also that if μ is another root of P, distinct from λ, then $P = (X-\lambda)^m Q$ forces $0 = P(\mu) = (\lambda-\mu)^m Q(\mu)$ and therefore since $(\lambda-\mu) \neq 0$ it forces $Q(\mu) = 0$. In other words, any other root of P would be one of Q. Repeating the above decomposition for Q and μ, and proceeding by induction, we obtain the following result. (Again, all this must stop as each step 'chops off one degree'.)

D.1.3. Proposition. *Let $P \in \mathbb{F}[X]$ be a polynomial of degree $d \geq 1$. If $\lambda_1, \ldots, \lambda_r$ are distinct roots of P then*

$$P(X) = (X-\lambda_1)^{m_1} \cdot \ldots \cdot (X-\lambda_r)^{m_r} \cdot Q(X)$$

for unique integers $m_1, \ldots, m_r \geq 1$ and a unique polynomial $Q \in \mathbb{F}[X]$ such that $Q(\lambda_i) \neq 0$ for all $i = 1, \ldots, r$. In particular, P has at most $d = \deg(P)$ roots, even counted with multiplicities: $\sum_{i=1}^{r} m_i \leq d$. □

D.2. Review of determinants

The 'correct' definition of the determinant is not entirely easy. Let us give a computational one instead. We start with a notation:

D.2.1. Notation. Let $A \in \mathrm{M}_{n\times n}(\mathbb{F})$ be a square matrix of size $n \geq 2$. Let $1 \leq i, j \leq n$. We denote by $A[i,j] \in \mathrm{M}_{(n-1)\times(n-1)}$ the square matrix of size one less obtained from A by removing the i-th row and the j-th column. (This is a very 'local' notation. Other authors will surely use other notation!)

D.2.2. Example. If $A = \left(\begin{smallmatrix} 1&2&3\\4&5&6\\7&8&9 \end{smallmatrix}\right)$ then $A[1,3] = \left(\begin{smallmatrix} 4&5\\7&8 \end{smallmatrix}\right)$ and $A[2,1] = \left(\begin{smallmatrix} 2&3\\8&9 \end{smallmatrix}\right)$.

D.2.3. Definition. Let $A \in \mathrm{M}_{n\times n}(\mathbb{F})$ be a square matrix. The *determinant of A (by expansion along the first column)* is defined by induction on n as follows:

$n = 1$: We define $\det([a_{11}]) = a_{11}$.

$n \geq 2$: Assume the determinant defined for all square matrices of size $n-1$. Let

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+1} \cdot A_{i,1} \cdot \det(A[i,1]). \tag{D.2.4}$$

The determinant is sometimes denoted $|A| = \det(A)$ but not in these notes.

D.2.5. Example. We have $\det\left(\begin{smallmatrix} a&c\\b&d \end{smallmatrix}\right) = a \cdot \det(d) - b \cdot \det(c) = ad - bc$.

D.2.6. Example. We have $\det\left(\begin{smallmatrix} 1&2&3\\4&5&6\\7&8&9 \end{smallmatrix}\right) = 1 \cdot \det\left(\begin{smallmatrix} 5&6\\8&9 \end{smallmatrix}\right) - 4 \cdot \det\left(\begin{smallmatrix} 2&3\\8&9 \end{smallmatrix}\right) + 7 \cdot \det\left(\begin{smallmatrix} 2&3\\5&6 \end{smallmatrix}\right) = 1 \cdot (5\cdot 9 - 8 \cdot 6) - 4 \cdot (2 \cdot 9 - 8 \cdot 3) + 7 \cdot (2 \cdot 6 - 5 \cdot 3) = -3 + 24 - 21 = 0$.

D.2.7. Example. An easy induction on n shows that if A is upper-triangular ($A_{ij} = 0$ for all $i > j$) then $\det(A) = A_{11} \det(A[1,1]) = \cdots = A_{11} \cdot A_{22} \cdots A_{nn}$ is just the product of the diagonal elements. In particular $\det(A) = A_{11} \cdot A_{22} \cdots A_{nn}$ if A is diagonal. For instance, $\det(I_n) = 1$.

D.2.8. Exercise. Let $A = \left(\begin{smallmatrix} A' & A''\\ 0 & A''' \end{smallmatrix}\right)$ be a $(n \times n)$-matrix that is upper-triangular by block, *i.e.* $A' \in \mathrm{M}_{d\times d}(\mathbb{F})$ and $A'' \in \mathrm{M}_{e\times d}(\mathbb{F})$ and $A''' \in \mathrm{M}_{e\times e}(\mathbb{F})$ where $n = d+e$. Show by induction that $\det(A) = \det(A') \cdot \det(A''')$.

Let us gather some properties of this function $\det\colon \mathrm{M}_{n\times n}(\mathbb{F}) \to \mathbb{F}$. To do this we identify $(n \times n)$-matrices with n column vectors in $\mathbb{F}^n$.

D.2.9. Theorem. *There exists a function*

$$\begin{array}{rccc} \delta\colon & (\mathbb{F}^n)^n = \mathbb{F}^n \times \cdots \times \mathbb{F}^n & \longrightarrow & \mathbb{F} \\ & (x_1, \cdots, x_n) & \longmapsto & \delta(x_1, \ldots, x_n) \end{array}$$

with the following properties:

(a) $\delta(e_1, \ldots, e_n) = 1$.

(b) δ *is* multilinear*, meaning that for all* $j = 1, \ldots, n$

$$\delta(\ldots, x_j + x_j', \ldots.) = \delta(\ldots, x_j, \ldots.) + \delta(\ldots, x_j', \ldots.)$$

and

$$\delta(\ldots, c \cdot x_j, \ldots.) = c \cdot \delta(\ldots, x_j, \ldots.)$$

for all $x_1, \ldots, x_n, x_j' \in \mathbb{F}^n$ *and* $c \in \mathbb{F}$*, where all the vectors marked* $\ldots$ *and* $\ldots.$ *are* $x_1, \ldots, x_{j-1}$ *and* $x_{j+1}, \ldots, x_n$ *respectively in all these expressions.* ([1])

(c) δ *is* alternating*, meaning that* $\delta(x_1, \ldots, x_n) = 0$ *if there exists* $i \neq j$ *such that* $x_i = x_j$.

Moreover, this function δ *is unique and is equal to* $\delta(x_1, \ldots, x_n) = \det([x_1|\cdots|x_n])$ *the determinant of the matrix whose columns are* $x_1, \ldots, x_n$.

D.2.10. Remark. Let us quickly observe that (b) and (c) imply that $\delta(x_1, \ldots, x_n)$ changes sign if we swap x_i and x_j for $i \neq j$. Let us prove this for $i = 1$ and $j = 2$ but the proof is the same everywhere. Here is an argument we could have used before as well. We use that $\delta(x, x, \ldots) = 0$ for $x = x_1$ for $x = x_2$ and for $x = x_1 + x_2$. Using (b), it gives $0 = \delta(x_1 + x_2, x_1 + x_2, \ldots) = \delta(x_1, x_1 + x_2, \ldots) + \delta(x_2, x_1 + x_2, \ldots) = \delta(x_1, x_1, \ldots) + \delta(x_1, x_2, \ldots) + \delta(x_2, x_1, \ldots) + \delta(x_2, x_2, \ldots) = \delta(x_1, x_2, \ldots) + \delta(x_2, x_1, \ldots)$. Hence $\delta(x_2, x_1, \ldots) = -\delta(x_1, x_2 \ldots)$.

Proof of Theorem D.2.9. It is an easy exercise to prove by induction on n that $\det([x_1|\cdots|x_n])$ satisfies (a) and (b). The proof of (c) is a little more delicate. One method of proof is as follows. First prove the Claim: $\det([x_1|\ldots|x_n])$ changes sign if you swap two columns. The proof of the claim can be made by induction: When swapping the first two columns, it holds by expanding twice along first columns. By induction, we get the Claim when we swap any two *consecutive* columns x_j and x_{j+1}. Then by induction on $j - i$ we get the Claim for swapping x_i and x_j for $j > i$. Indeed, swapping (i, j) is like swapping $(j-1, j)$ then $(i, j-1)$ then $(j-1, j)$ again. Each of these 3 operations changes the determinant's sign (by induction hypothesis), hence so does the swap of (i, j). Once we have the Claim, we get (c) by reducing to the case where $i = 1$ and $j = 2$. In that case, a double column-expansion (*i.e.* apply the inductive Definition D.2.3 twice) gives (c).

Let us see the converse: If $\delta\colon (\mathbb{F}^n)^n \to \mathbb{F}$ satisfies (a), (b) and (c) then δ 'is' the determinant. We proceed by induction on n and assume that every $\delta'\colon (\mathbb{F}^{n-1})^{n-1} \to \mathbb{F}$ that is 1 on the canonical basis, multilinear and alternating must be the (smaller) determinant. Consider now $\delta\colon (\mathbb{F}^n)^n \to \mathbb{F}$ satisfying (a), (b) and (c). By linearity in the first variable, we can compute $\delta(x_1, x_2, \ldots, x_n)$ as a linear combination

[1] In clear we want $\delta(x_1, \ldots, x_{j-1}, \mathbf{x_j} + \mathbf{x_j'}, x_{j+1}, \ldots, x_n) = \delta(x_1, \ldots, x_{j-1}, \mathbf{x_j}, x_{j+1}, \ldots, x_n) + \delta(x_1, \ldots, x_{j-1}, \mathbf{x_j'}, x_{j+1}, \ldots, x_n)$ and $\delta(x_1, \ldots, x_{j-1}, \mathbf{c} \cdot x_j, x_{j+1}, \ldots, x_n) = \mathbf{c} \cdot \delta(x_1, \ldots, x_n)$. In other words, the expression $\delta(x_1, \ldots, x_n)$ is linear in each x_j separately, when all the other $x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n$ are kept constant, and this for each $j = 1, \ldots, n$, one at a time.

of $\delta(e_i, x_2, \ldots, x_n)$ for $i = 1, \ldots, n$, namely, since $x_1 = x_{11}e_1 + \cdots + x_{1n}e_n$, we have

$$\delta(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_{1i} \cdot \delta(e_i, x_2, \ldots, x_n).$$

The point is that each $\delta(e_i, x_2, \ldots, x_n)$ 'ignores' the i-th entry of the column vectors $x_2, \ldots, x_n$. Let us see that for $i = 1$ and leave the other cases as an exercise. We want to show that $\delta(e_1, x_2, \ldots, x_n)$ does not depend on the first entry of the vectors $x_2, \ldots, x_n$. Indeed, write x_2 as $x_{21} \cdot e_1 + x_2'$ where $x_2' \in \operatorname{Span}(e_2, \ldots, e_n)$ has first entry zero. Then by (b) and (c), we see that

$$\begin{aligned}
\delta(e_1, \mathbf{x_2}, x_3, \ldots, x_n) &= \delta(e_1, \, \mathbf{x_{21}\, e_1 + x_2'}\,, x_3, \ldots, x_n) \\
&= \mathbf{x_{21}} \cdot \delta(e_1, \mathbf{e_1}, x_3, \ldots, x_n) + \delta(e_1, \mathbf{x_2'}, x_3, \ldots, x_n) && \text{by (b)} \\
&= \delta(e_1, x_2', x_3, \ldots, x_n) && \text{by (c).}
\end{aligned}$$

Repeating this for each column $(3, 4, \ldots, n)$ we have proved that $\delta(e_1, x_2, \ldots, x_n) = \delta(e_1, x_2', \ldots, x_n')$ where each $x_j' = x_j - x_{j1}e_1$ has first entry zero. Now, the expression $\delta(e_1, x_2', \ldots, x_n')$ is multilinear and alternating in $x_2', \ldots, x_n'$ in $\operatorname{Span}(e_2, \ldots, e_n) = \left\{ \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{F}^n \,\middle|\, y_1 = 0 \right\} = \left\{ \begin{bmatrix} 0 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{F}^n \,\middle|\, \begin{bmatrix} y_2 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{F}^{n-1} \right\} \cong \mathbb{F}^{n-1}$ and it is 1 when $x_i' = e_i$ for all $i = 2, \ldots, n$. So by induction on n, we know that $\delta(e_1, x_2', \ldots, x_n')$ is the determinant of the $((n-1)\times(n-1))$-matrix $[x_1|\ldots|x_n][1,1]$. For $\delta(e_i, x_2, \ldots, x_n)$ a similar treatment and the swap of $(i-1)$ consecutive rows shows that we have $\delta(e_i, x_2, \ldots, x_n) = (-1)^{i-1}\det([x_1|\cdots|x_n][i,1])$. In conclusion, $\delta(x_1, \ldots, x_n) = \det([x_1|\cdots|x_n])$ as in the formula of Definition D.2.3. □

A consequence (of the proof) is:

D.2.11. Corollary. *Let* $A \in \mathrm{M}_{n\times n}(\mathbb{F})$. *Then* $\det(A) = \det(A^t)$.

Proof. As in the above proof, one can verify from the formulas of Definition D.2.3 that $\delta\colon (x_1, \ldots, x_n) \mapsto \det\big((x_1|\cdots|x_n)^t\big)$ satisfies (a), (b) and (c). (Here again (c) is a little convoluted.) By uniqueness, this δ must be the determinant. □

We then have the following well-known rules for the determinant.

D.2.12. Proposition. *Let* $A \in \mathrm{M}_{n\times n}(\mathbb{F})$.

(1) *Let B be obtained from A by swapping two columns. Then* $\det(B) = -\det(A)$.
(2) *Same as (1) but for swapping two rows.*
(3) *Let C be obtained from A by adding to the i-th column of A a scalar multiple of the j-th column of A, for $i \neq j$. Then* $\det(C) = \det(A)$.
(4) *Same as (3) but with rows: Adding a multiple of a row to* another *row does not change the determinant.*
(5) *Let D be the matrix obtained from A by multiplying one column of A by $a \in \mathbb{F}$. Then* $\det(D) = a \cdot \det(A)$.
(6) *Same as (5) but with rows.*
(7) *For every $a \in \mathbb{F}$, we have* $\det(a \cdot A) = a^n \cdot \det(A)$.

Proof. We only prove the odd-numbered claimed. The even ones follow by transposition (Corollary D.2.11) or can be proved similarly.

Part (1) is Remark D.2.10.

Part (3) follows since $\delta(x, y+b\cdot x, \ldots) = \delta(x, y, \ldots) + b\cdot\delta(x, x, \ldots) = \delta(x, y, \ldots)$ since δ is alternating. And similarly if you use any two columns.

Part (5) is contained in multilinearity.

Finally, Part (7) is simply obtained by applying (5) n times. Indeed, $a \cdot A$ is obtained by multiplying every of the n columns of A by the same constant a. □

It can be useful to expand the determinant along any row or column:

D.2.13. Proposition. *Let $A \in \mathrm{M}_{n\times n}(\mathbb{F})$.*

(1) *Let $1 \leq j \leq n$. Then $\det(A)$ can be computed by 'expansion along the j-th column':*

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

(2) *Let $1 \leq i \leq n$. Then $\det(A)$ can be computed by 'expansion along the i-th row':*

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

Proof. Let us name the columns of A as $A = (x_1|\cdots|x_n)$. Define $B = (x_j|x_1|x_2|\cdots|x_{j-1}|x_{j+1}|\cdots|x_n)$, obtained by moving the j-th column of A to the front and pushing all columns from 1 to $j-1$ by one notch to the right. Note that our A can be recovered from this B by $(j-1)$ swaps of columns, putting back the j-th column to its place: Swap first column and second, then second and third, etc, until $(j-1)$-th and j-th. By Proposition D.2.12 (1), we see that $\det(A) = (-1)^{j-1}\det(B)$. But if we now expand $\det(B)$ along the first column as in Definition D.2.3, we get Part (1). (Alternatively, one can show that the formulas of the statement satisfy Conditions (a), (b) and (c) of Theorem D.2.9.) □

D.2.14. Corollary. *Let $A \in \mathrm{M}_{n\times n}(\mathbb{F})$. Suppose that we apply any number of steps of the Gauss-Jordan algorithm $A \rightsquigarrow B$ to obtain another matrix B. Then $\det(A) = 0$ if and only if $\det(B) = 0$.*

Proof. Indeed, Gauss-Jordan is a combination of (2), (4) and (6) in Proposition D.2.12 but only using non-zero $a \in \mathbb{F}$ in (6). Each step either changes the determinant by a sign, keeps it unchanged, or multiplies it by a. In any case if it was zero it stays zero and if it was non-zero it stays non-zero. □

Here is what $\det(A)$ *determines*:

D.2.15. Theorem. *Let $A \in \mathrm{M}_{n\times n}(\mathbb{F})$. Then A is invertible if and only if $\det(A)$ is non-zero.*

Proof. Recall that A is invertible if and only if it can be row-reduced $A \rightsquigarrow I_n$ to the identity matrix by Gauss-Jordan. (Indeed, this is equivalent to $\mathrm{Ker}(A) = \{0\}$ since it amounts to solving the system $A \cdot x = 0$ and finding pivots in *all* rows.) If this happens then $\det(A) \neq 0$ by Corollary D.2.14, since $\det(I_n) = 1 \neq 0$. Conversely, if $\det(A) \neq 0$ then apply Gauss-Jordan $A \rightsquigarrow B$ to get a row-reduced echelon matrix B, in particular an upper-triangular matrix B with $\det(B) \neq 0$ by Corollary D.2.14; we see that B has to be diagonal (all pivots on the diagonal) otherwise $\det(B) = 0$ by Example D.2.7. Hence $B = I_n$ and A is invertible. □

In fact, one can be more precise:

D.2.16. Theorem. *Let* $A, B \in \mathrm{M}_{n\times n}(\mathbb{F})$. *Then* $\det(A \cdot B) = \det(A) \cdot \det(B)$.

Proof. Let us do four examples.

First if A is obtained from the identity matrix by swapping the i-th row and the j-th row, for $i \neq j$. For instance, for $i = 2$ and $j = 4$ and $n = 5$, it would be $A = \begin{pmatrix} 1&0&0&0&0\\ 0&0&0&1&0\\ 0&0&1&0&0\\ 0&1&0&0&0\\ 0&0&0&0&1 \end{pmatrix}$. Note that $A \cdot B$ is simply the matrix B in which we swap the i-th row and j-th column. By Proposition D.2.12 (2), $\det(AB) = -\det(B)$ and $\det(A) = -\det(I_n) = -1$ and therefore $\det(AB) = \det(A)\cdot\det(B)$ in that case.

Second, suppose that $A = E_{ij}(b)$ for some $i \neq j$ and $b \in \mathbb{F}$ is the identity matrix, except for its (i, j)-entry where we have b instead of 0. In other words, it is the matrix obtained from I_n by adding b times the j-th row to the i-th. By Proposition D.2.12 (4) we have $\det(A) = \det(I_n) = 1$. Note also that $A \cdot B = E_{ij}(b) \cdot B$ simply is the matrix B in which we add b times the j-th row to the i-th. So the same proposition implies $\det(AB) = \det(B)$ and again $\det(AB) = \det(A)\det(B)$ in this case.

Thirdly, suppose that $A = \mathrm{diag}(1, \ldots, 1, a, 1, \ldots, 1)$ is almost the identity matrix, except that we multiplied its i-th row (where the a appears) by $a \in \mathbb{F}$. Note that $A \cdot B$ is simply B but with the i-th row multiplied by a. By Proposition D.2.12, we have $\det(A) = a \cdot \det(I_n) = a$ and $\det(AB) = a\det(B)$. Again, $\det(AB) = \det(A)\det(B)$ in that case.

In summary, at this stage, we have seen that every operation in Gauss-Jordan can be obtain by left-multiplying by a suitable matrix A and that at least for there matrices A we indeed have $\det(AB) = \det(A)\det(B)$, for all B.

Finally, the fourth example is when A is in row-reduced echelon form. Here there are two possibilities. Either $A = I_n$ and then surely $\det(AB) = \det(B) = \det(A)\det(B)$. Or A is *not* the identity, in which case its last row consists entirely of zeros. But then so does the last row of AB and by expansion along the last row we have $\det(AB) = 0 = 0 \cdot \det(B) = \det(A)\det(B)$ again.

Let us now take a general A. By Gauss-Jordan, we can row-reduce $A \rightsquigarrow A'$ to A' that is row-reduced echelon, *i.e.* A' is as in the fourth example above (the latter one). We can 'undo' Gauss-Jordan, *i.e.* we can reconstruct A from A' by multiplying on the left by matrices of the first three kinds: $A = A_1 \cdots A_\ell \cdot A'$ where each A_i is a matrix that swap rows, adds a multiple of a row to another or multiplies (or divides) a row by a non-zero scalar. By the four examples of the first part of the proof, we know that $\det(AB) = \det(A_1 \cdots A_\ell \cdot A' \cdot B) = \det(A_1)\cdots\det(A_\ell)\cdot\det(A')\cdot\det(B) = \det(A_1 \cdots A_\ell \cdot A') \cdot \det(B) = \det(A)\det(B)$. □

D.2.17. Corollary. *If* $Q \in \mathrm{GL}_n(\mathbb{F})$ *is invertible then* $\det(Q^{-1}) = \det(Q)^{-1}$.

Proof. Indeed, $\det(Q)\det(Q^{-1}) = \det(Q \cdot Q^{-1}) = \det(I_n) = 1$. □

D.2.18. Corollary. *Similar matrices have the same determinant: If* $A = Q \cdot B \cdot Q^{-1}$ *in* $\mathrm{M}_{n\times n}(\mathbb{F})$ *with* $Q \in \mathrm{GL}_n(\mathbb{F})$ *then* $\det(A) = \det(B)$.

Proof. We have by Theorem D.2.16 and the above corollary that $\det(A) = \det(Q \cdot B \cdot Q^{-1}) = \det(Q) \cdot \det(B) \cdot \det(Q)^{-1} = \det(B)$ by commutativity in $\mathbb{F}$. □

D.2.19. Example. A 2×2-matrix $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ is invertible if and only if $\det(A) = ad - bc$ is non-zero. In that case, we have an explicit formula for the inverse:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Indeed a direct computation gives $\left(\begin{smallmatrix} d & -b \\ -c & a \end{smallmatrix}\right) \cdot \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = \left(\begin{smallmatrix} ad-bc & 0 \\ 0 & ad-bc \end{smallmatrix}\right)$.

There are explicit formulas for A^{-1} in the spirit of the above, due to Cramer.

D.2.20. Exercise. Let $A \in \mathrm{M}_{n\times n}(\mathbb{F})$. Define the (adjugate) matrix $\tilde{A} \in \mathrm{M}_{n\times n}(\mathbb{F})$ by the formula

$$\tilde{A}_{ij} = (-1)^{i+j} \det(A[j,i])$$

where $A[j,i] \in \mathrm{M}_{(n-1)\times(n-1)}$ is as in Notation D.2.1. Beware that we remove the j-th row and i-th column of A when we define the (i,j)-entry of $\tilde{A}$. (No misprint!)

(1) Prove that $\tilde{A} \cdot A = A \cdot \tilde{A} = \det(A) \cdot I_n$. [Hint: Prove that $(A \cdot \tilde{A})_{ij}$ is the determinant of a certain matrix built out of A and depending on i and j.]

(2) Deduce that A is invertible if and only if $\det(A) \neq 0$ (without using Theorem D.2.15) and give an explicit formula for A^{-1}.

Permutations. We can push the analysis of the determinant a little further and produce a cleaner formula, if we speak of *permutations*.

D.2.21. Definition. Let $n \in \mathbb{N}$, $n \geq 1$. A *permutation* of n letters is simply a bijection $\sigma\colon \{1,\ldots,n\} \overset{\sim}{\to} \{1,\ldots,n\}$ on the set with n elements $\{1,\ldots,n\}$. We write $S_n = \big\{\, \sigma\colon \{1,\ldots,n\} \to \{1,\ldots,n\} \,\big|\, \sigma \text{ is bijective} \,\big\}$ for the set of all such permutations. ([2]) Note that we can compose permutations, $\sigma_2 \circ \sigma_1$, as we compose functions: $(\sigma_2 \circ \sigma_1)(i) = \sigma_2(\sigma_1(i))$ for all $1 \leq i \leq n$.

D.2.22. Example. For all n we always have an identity $\mathrm{id}\colon \{1,\ldots,n\} \overset{\sim}{\to} \{1,\ldots,n\}$, which is as usual $\mathrm{id}(i) = i$ for all $i \in \{1,\ldots,n\}$. It satisfies $\sigma \circ \mathrm{id} = \sigma = \mathrm{id} \circ \sigma$ for all $\sigma \in S_n$. For $n = 1$ this is of course the only permutation $S_1 = \{\mathrm{id}\}$.

D.2.23. Example. For $n = 2$, we have $S_2 = \{\mathrm{id}, \sigma\}$ where $\sigma\colon \{1,2\} \overset{\sim}{\to} \{1,2\}$ is given by $\sigma(1) = 2$ and $\sigma(2) = 1$. We have $\sigma \circ \sigma = \mathrm{id}$.

D.2.24. Example. For $n = 3$, we have $|S_3| = 6$. Beyond the identity, we can swap two elements and leave the last untouched (in 3 different ways) and we can permute cyclicly in two possible ways: Either $\sigma\colon 1 \mapsto 2 \mapsto 3 \mapsto 1$ which is denoted (123), and $\tau\colon 1 \mapsto 3 \mapsto 2 \mapsto 1$ which is denoted (132).

D.2.25. Example. For any $n \geq 2$ and any two elements $i \neq j$ of $\{1,\ldots,n\}$, we define the *transposition* (ij) in S_n to be the permutation of i and j that leaves the other unchanged. Explicitly $(ij) = \sigma$ where $\sigma(i) = j$ and $\sigma(j) = i$ and $\sigma(k) = k$ for all other $k \in \{1,\ldots,n\} \smallsetminus \{i,j\}$. Note that $(ij) \circ (ij) = \mathrm{id}$.

D.2.26. Exercise. Show that every permutation is the composition of a finite number of transpositions. [Hint: induction on $\big|\{i \in \{1,\ldots,n\} \mid \sigma(i) \neq i\}\big|$.]

D.2.27. Remark. One can show in general that $|S_n| = n! = n \cdot (n-1) \cdots 2 \cdot 1$.

[2] It is actually what we call a *group*, under composition.

D.2.28. Notation. Let $\sigma \in S_n$ be a permutation of n letters. We can associate to σ a matrix $Q_\sigma \in \mathrm{M}_{n\times n}(\mathbb{F})$ called a *permutation matrix* and defined by

$$(Q_\sigma)_{i,j} = \delta_{i,\sigma(j)} = \begin{cases} 1 & \text{if } i = \sigma(j) \\ 0 & \text{otherwise.} \end{cases}$$

We see that Q_σ has only one non-zero entry in each row and each column: In the j-th column, we put a 1 in the $\sigma(j)$-th row. In other words, Q_σ is the matrix obtained from the identity by permuting its rows as prescribed by σ.

This construction allows us to keep track of σ in matrix form. in other words, $\sigma \mapsto Q_\sigma$ is an injection $S_n \to \mathrm{M}_{n\times n}(\mathbb{F})$.

D.2.29. Example. We always have $Q_{\mathrm{id}} = I_n$.

For $n = 2$, we have $Q_{(12)} = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$.

For $n = 3$ (Example D.2.24) the 6 permutation matrices that we obtain are

$$\left(\begin{smallmatrix} 1&0&0\\0&1&0\\0&0&1 \end{smallmatrix}\right),\ \left(\begin{smallmatrix} 0&1&0\\1&0&0\\0&0&1 \end{smallmatrix}\right),\ \left(\begin{smallmatrix} 1&0&0\\0&0&1\\0&1&0 \end{smallmatrix}\right),\ \left(\begin{smallmatrix} 0&0&1\\0&1&0\\1&0&0 \end{smallmatrix}\right),\ \left(\begin{smallmatrix} 0&0&1\\1&0&0\\0&1&0 \end{smallmatrix}\right),\ \left(\begin{smallmatrix} 0&1&0\\0&0&1\\1&0&0 \end{smallmatrix}\right).$$

They are respectively Q_{id}, $Q_{(12)}$, $Q_{(23)}$, $Q_{(13)}$, $Q_{(123)}$ and $Q_{(132)}$.

D.2.30. Lemma. *We have $Q_{\sigma\circ\tau} = Q_\sigma \cdot Q_\tau$ for every $\sigma, \tau \in S_n$. In particular every $Q_\sigma \in \mathrm{GL}_n(\mathbb{F})$ is invertible and $(Q_\sigma)^{-1} = Q_{\sigma^{-1}}$.*

Proof. Let us show that both matrices $Q_{\sigma\circ\tau}$ and $Q_\sigma \cdot Q_\tau$ have the same (i,k)-entry, for all $i, k \in \{1, \ldots, n\}$. We have $(Q_\sigma \cdot Q_\tau)_{ik} = \sum_{j=1}^n (Q_\sigma)_{ij} \cdot (Q_\tau)_{jk} = \sum_{j=1}^n \delta_{i,\sigma(j)} \cdot \delta_{j,\tau(k)}$. Most terms in this sum as zero. If all of them are then the sum is zero. Let us see when the sum is not zero. For that we need to find j between 1 and n such that $j = \tau(k)$ (otherwise $\delta_{j,\tau(k)} = 0$) *and* also such that $i = \sigma(j)$ (otherwise $\delta_{i,\sigma(j)} = 0$). In other words, this $(Q_\sigma \cdot Q_\tau)_{ik} = 0$ is zero unless $i = \sigma(j)$ where $j = \tau(k)$, meaning unless $i = \sigma(\tau(k)) = (\sigma \circ \tau)(k)$, in which case the sum is 1 since there is only one j that fits the bill ($j = \tau(k)$). This is exactly the (i,k)-entry of $(Q_{\sigma\circ\tau})$. The rest follows since then $Q_\sigma \cdot Q_{\sigma^{-1}} = Q_{\mathrm{id}} = I_n$. □

D.2.31. Definition. The *signature* of a permutation $\sigma \in S_n$ is defined as $\mathrm{sgn}(\sigma) = \det(Q_\sigma)$. This is a somewhat artificial definition.

One should define the signature without invoking matrices, or determinants. It would be as follows (but the proof would be longer):

D.2.32. Proposition. *Let $n \geq 1$.*

(1) *We have* $\mathrm{sgn}(\sigma) = \pm 1$ *for all* $\sigma \in S_n$.
(2) *We have* $\mathrm{sgn}(\sigma \circ \tau) = \mathrm{sgn}(\sigma) \cdot \mathrm{sgn}(\tau)$ *for all* $\sigma, \tau \in S_n$.
(3) *Transposition have signature -1: For every $i \neq j$, we have* $\mathrm{sgn}((ij)) = -1$.
(4) *If σ is the composition of m transpositions as in (3) – and all permutations are like that for some m – then* $\mathrm{sgn}(\sigma) = (-1)^m$.

Proof. Part (2) follows from $Q_{\sigma\circ\tau} = Q_\sigma \cdot Q_\tau$ (Lemma D.2.30) and Theorem D.2.16. For Part (3) note that $Q_{(ij)}$ is obtained from the identity matrix by swapping two rows: the i-th and the j-th. Hence $\det(Q_{(ij)}) = -1$ by Proposition D.2.12 (2). Part (4) follows. And Part (1) is less precise than (4). □

D.2.33. Example. Returning to $S_3 = \{\mathrm{id}, (12), (23), (13), (123), (132)\}$. Then $(12) \circ (23) = (123)$ and $(12) \circ (13) = (132)$ and we see that $(12), (13), (23)$ have signature -1 and the other three have signature 1.

We can use permutations and signatures to give a non-recursive formula for the determinant. Repeating the expansion-along-a-column definition, we see that $\det(A)$ is a sum of products of entries of A. More precisely, since we remove the i-th row and j-th column 'once we have used A_{ij}' we see that $\det(A)$ is the sum of al possible products of n entries of A, where we always take exactly one entry in each row and each column. And there are signs...

D.2.34. Theorem. *Let $A \in \mathrm{M}_{n\times n}(\mathbb{F})$. Then we have*

$$\det(A) = \sum_{\sigma\in S_n} \big(\operatorname{sgn}(\sigma)\cdot \prod_{i=1}^{n} A_{i,\sigma(i)}\big).$$

Proof. One can easily verify that the formula given in the statement (as a function of the columns of A) is multilinear and takes the value 1 on the canonical basis, as in (a) and (b) of Theorem D.2.9. It only remains to check that (c) holds, *i.e.* that it is an alternating formula in the columns of A. Suppose that the j-th and the k-th columns of A coincide, say for $j < k$. Hence $A_{i,\sigma(i)} = A_{i,\tau(i)}$ whenever $\tau = (jk)\sigma$: Permuting the entries of the j-th column and of the k-th column does not change A. In other words, the expression

$$\prod_{i=1}^{n} A_{i,\sigma(i)} \qquad \text{and} \qquad \prod_{i=1}^{n} A_{i,\tau(i)}$$

are equal if $\tau = (jk)\sigma$. But now, partition S_n into two disjoint subsets $S_n = T_n \sqcup U_n$ where $T_n = \big\{\, \sigma \in S_n \,\big|\, \sigma^{-1}(j) < \sigma^{-1}(k) \,\big\}$ and $U_n = \big\{\, \tau \in S_n \,\big|\, \tau^{-1}(j) > \tau^{-1}(k) \,\big\}$. (The preimages of $j \neq k$ cannot be equal. Hence one is greater than the other.) Note also that we have a bijection $T_n \overset{\sim}{\to} U_n$ given by $\sigma \mapsto (jk)\circ\sigma$. The inverse of this bijection is given by the same formula (!) since $(jk)^2 = \mathrm{id}$. Under this bijection the signature changes sign since $\operatorname{sgn}((jk)) = -1$. Let us also abbreviate

$$b(\sigma) := \operatorname{sgn}(\sigma)\cdot \prod_{i=1}^{n} A_{i,\sigma(i)}.$$

The theorem claims that $\det(A) = \sum_{\sigma\in S_n} b(\sigma)$. So far we have seen that $S_n = T_n \sqcup U_n$, that $\sigma \mapsto (jk)\circ\sigma$ is a bijection between T_n and U_n and that $b((jk)\sigma) = -b(\sigma)$. We have all ingredients to conclude

$$\begin{aligned}
\textstyle\sum_{\sigma\in S_n} b(\sigma) &= \textstyle\sum_{\sigma\in T_n} b(\sigma) + \sum_{\tau\in U_n} b(\tau) && \text{since } S_n = T_n \sqcup U_n\\
&= \textstyle\sum_{\sigma\in T_n} b(\sigma) + \sum_{\sigma\in T_n} b((jk)\sigma) && \text{using our bijection } T_n \overset{\sim}{\to} U_n\\
&= \textstyle\sum_{\sigma\in T_n} b(\sigma) + \sum_{\sigma\in T_n} -b(\sigma) && \text{by our assumption on } A\\
&= \textstyle\sum_{\sigma\in T_n} \big(b(\sigma) - b(\sigma)\big) = 0.
\end{aligned}$$

Thus the formula of the statement has all properties that uniquely characterizes the determinant and we conclude by Theorem D.2.9. □