M04: LINEAR ALGEBRA I

Thomas Mettler

Fall Semester 2024

Version 29 Aug 2024

FernUni.ch
UniDistance.ch

Contents

Acknowledgements

I am grateful to Micha Wasem for undertaking the meticulous task of proofreading the complete set of lecture notes and offering insights for enhancing pedagogy. Micha's contributions also include crafting exercises with solutions, formulating multiple-choice questions, creating figures, and coding animations. I am also grateful to Keegan Flood for contributing insightful multiple-choice questions.

Furthermore, I would like to express my appreciation to a group of diligent students for spotting typos, in particular to Stéphane Billeter, Daniele Bolla, Johanna Bühler and Liborio Costa.

These lectures notes are inspired by the following sources:

- *Algebra* by Michael Artin, Birkhäuser Grundstudium der Mathematik.
- *Linear Algebra Done Right* by Sheldon Axler, Springer Undergraduate Texts in Mathematics.
- *Linear Algebra* by Emmanuel Kowalski, lecture notes available from his home page at ETH Zurich.
- *Introduction to Linear Algebra* by Gilbert Strang, Wellesley-Cambridge Press

HTML Version

These lecture notes are also available in an HTML version and in app form.

https://apptest.fernuni.ch

The HTML version contains the lectures notes and additionally animations, solutions to the exercises and multiple choice questions.

Part 1

Linear Algebra I

CHAPTER 1

Fields and complex numbers

1.1 Fields

WEEK 1

A field $\mathbb{K}$ is roughly speaking a number system in which we can add and multiply numbers, so that the expected properties hold. We will only briefly state the basic facts about fields. For a more detailed account, we refer to the algebra module.

Definition 1.1 A *field* consists of a set $\mathbb{K}$ containing distinguished elements $0_{\mathbb{K}} \neq 1_{\mathbb{K}}$, as well as two binary operations, *addition* $+_{\mathbb{K}} : \mathbb{K} \times \mathbb{K} \to \mathbb{K}$ and *multiplication* $\cdot_{\mathbb{K}} : \mathbb{K} \times \mathbb{K} \to \mathbb{K}$, so that the following properties hold:

- Commutativity of addition

$$x +_{\mathbb{K}} y = y +_{\mathbb{K}} x \quad \text{for all } x, y \in \mathbb{K}.$$

- Commutativity of multiplication

$$x \cdot_{\mathbb{K}} y = y \cdot_{\mathbb{K}} x \quad \text{for all } x, y \in \mathbb{K}. \tag{1.1}$$

- Associativity of addition

$$(x +_{\mathbb{K}} y) +_{\mathbb{K}} z = x +_{\mathbb{K}} (y +_{\mathbb{K}} z) \quad \text{for all } x, y, z \in \mathbb{K}. \tag{1.2}$$

- Associativity of multiplication

$$(x \cdot_{\mathbb{K}} y) \cdot_{\mathbb{K}} z = x \cdot_{\mathbb{K}} (y \cdot_{\mathbb{K}} z) \quad \text{for all } x, y, z \in \mathbb{K}. \tag{1.3}$$

- $0_{\mathbb{K}}$ is the identity element of addition

$$x +_{\mathbb{K}} 0_{\mathbb{K}} = 0_{\mathbb{K}} +_{\mathbb{K}} x = x \quad \text{for all } x \in \mathbb{K}. \tag{1.4}$$

- $1_{\mathbb{K}}$ is the identity element of multiplication

$$x \cdot_{\mathbb{K}} 1_{\mathbb{K}} = 1_{\mathbb{K}} \cdot_{\mathbb{K}} x = x \quad \text{for all } x \in \mathbb{K}. \tag{1.5}$$

- For any $x \in \mathbb{K}$ there exists a unique element, denoted by $(-x)$ and called the *additive inverse* of x, such that

$$x +_{\mathbb{K}} (-x) = (-x) +_{\mathbb{K}} x = 0_{\mathbb{K}}. \tag{1.6}$$

- For any $x \in \mathbb{K} \setminus \{0_{\mathbb{K}}\}$ there exists a unique element, denoted by x^{-1} or $\frac{1}{x}$ and called the *multiplicative inverse* of x, such that

$$x \cdot_{\mathbb{K}} \frac{1}{x} = \frac{1}{x} \cdot_{\mathbb{K}} x = 1_{\mathbb{K}}. \tag{1.7}$$

- Distributivity of multiplication over addition

$$(x +_{\mathbb{K}} y) \cdot_{\mathbb{K}} z = x \cdot_{\mathbb{K}} z +_{\mathbb{K}} y \cdot_{\mathbb{K}} z \quad \text{for all } x, y, z \in \mathbb{K}. \tag{1.8}$$

Remark 1.2

(i) It is customary to simply speak of a field $\mathbb{K}$, without explicitly mentioning $0_{\mathbb{K}}$, $1_{\mathbb{K}}$ and $+_{\mathbb{K}}, \cdot_{\mathbb{K}}$.

(ii) When $\mathbb{K}$ is clear from the context, we often simply write 0 and 1 instead of $0_\mathbb{K}$ and $1_\mathbb{K}$. Likewise, it is customary to write $+$ instead of $+_\mathbb{K}$ and $\cdot$ instead of $\cdot_\mathbb{K}$. Often $\cdot_\mathbb{K}$ is omitted entirely so that we write xy instead of $x \cdot_\mathbb{K} y$.

(iii) We refer to the elements of a field as *scalars*.

(iv) The set $\mathbb{K} \setminus \{0_\mathbb{K}\}$ is usually denoted by $\mathbb{K}^*$.

(v) For all $x, y \in \mathbb{K}$ we write $x - y = x +_\mathbb{K} (-y)$ and for all $x \in \mathbb{K}$ and $y \in \mathbb{K}^*$ we write $\frac{x}{y} = x \cdot_\mathbb{K} \frac{1}{y} = x \cdot_\mathbb{K} y^{-1}$.

(vi) A field $\mathbb{K}$ containing only finitely many elements is called *finite*. Algorithms in cryptography are typically based on finite fields.

Example 1.3

(i) The rational numbers or quotients $\mathbb{Q}$, the real numbers $\mathbb{R}$ and the complex numbers $\mathbb{C}$ – that we will study more carefully below – equipped with the usual addition and multiplication are examples of fields.

(ii) The integers $\mathbb{Z}$ (with usual addition and multiplication) are not a field, as only 1 and -1 admit a multiplicative inverse.

(iii) Considering a set $\mathbb{F}_2$ consisting of only two elements that we may denote by 0 and 1, we define $+_{\mathbb{F}_2}$ and $\cdot_{\mathbb{F}_2}$ via the following tables

$+_{\mathbb{F}_2}$	0	1
0	0	1
1	1	0

and

$\cdot_{\mathbb{F}_2}$	0	1
0	0	0
1	0	1

For instance, we have $1 +_{\mathbb{F}_2} 1 = 0$ and $1 \cdot_{\mathbb{F}_2} 1 = 1$. Then, one can check that $\mathbb{F}_2$ equipped with these operations is indeed a field. A way to remember these tables is to think of 0 as representing the even numbers, while 1 represents the odd numbers. So for instance, a sum of two odd numbers is even and a product of two odd numbers is odd. Alternatively, we may think of 0 and 1 representing the boolean values *FALSE* and *TRUE*. In doing so, $+_{\mathbb{F}_2}$ corresponds to the logical *XOR* and $\cdot_{\mathbb{F}_2}$ corresponds to the logical *AND*.

(iv) Considering a set $\mathbb{F}_4$ consisting of four elements, say $\{0, 1, a, b\}$, we define $+_{\mathbb{F}_4}$ and $\cdot_{\mathbb{F}_4}$ via the following tables

$+_{\mathbb{F}_4}$	0	1	a	b
0	0	1	a	b
1	1	0	b	a
a	a	b	0	1
b	b	a	1	0

and

$\cdot_{\mathbb{F}_4}$	0	1	a	b
0	0	0	0	0
1	0	1	a	b
a	0	a	b	1
b	0	b	1	a

Again one can check that $\mathbb{F}_4$ equipped with these operations is indeed a field.

Lemma 1.4 (Field properties) *In a field $\mathbb{K}$ we have the following properties:*

(i) $0_\mathbb{K} \cdot_\mathbb{K} x = 0_\mathbb{K}$ *for all* $x \in \mathbb{K}$.

(ii) $-x = (-1_\mathbb{K}) \cdot_\mathbb{K} x$ *for all* $x \in \mathbb{K}$.

(iii) For all $x, y \in \mathbb{K}$*, if* $x \cdot_\mathbb{K} y = 0_\mathbb{K}$*, then* $x = 0_\mathbb{K}$ *or* $y = 0_\mathbb{K}$.

(iv) $-0_\mathbb{K} = 0_\mathbb{K}$.

(v) $(1_\mathbb{K})^{-1} = 1_\mathbb{K}$.

(vi) $(-(-x)) = x$ *for all* $x \in \mathbb{K}$.

(vii) $(-x) \cdot_\mathbb{K} y = x \cdot_\mathbb{K} (-y) = -(x \cdot_\mathbb{K} y)$.

(viii) $(x^{-1})^{-1} = x$ *for all* $x \in \mathbb{K}^*$.

Proof We will only prove some of the items, the rest are an exercise for the reader.

(i) Using (1.4), we obtain $0_{\mathbb{K}} +_{\mathbb{K}} 0_{\mathbb{K}} = 0_{\mathbb{K}}$. Hence for all $x \in \mathbb{K}$ we have

$$x \cdot_{\mathbb{K}} 0_{\mathbb{K}} = x \cdot_{\mathbb{K}} (0_{\mathbb{K}} + 0_{\mathbb{K}}) = x \cdot_{\mathbb{K}} 0_{\mathbb{K}} +_{\mathbb{K}} x \cdot_{\mathbb{K}} 0_{\mathbb{K}},$$

where the second equality uses (1.8). Adding the additive inverse of $x \cdot_{\mathbb{K}} 0_{\mathbb{K}}$, we get

$$x \cdot_{\mathbb{K}} 0_{\mathbb{K}} - x \cdot_{\mathbb{K}} 0_{\mathbb{K}} = (x \cdot_{\mathbb{K}} 0_{\mathbb{K}} +_{\mathbb{K}} x \cdot_{\mathbb{K}} 0_{\mathbb{K}}) - x \cdot_{\mathbb{K}} 0_{\mathbb{K}}$$

using the associativity of addition (1.2) and (1.6), this last equation is equivalent to

$$0_{\mathbb{K}} = x \cdot_{\mathbb{K}} 0_{\mathbb{K}}$$

as claimed.

(iii) Let $x, y \in \mathbb{K}$ such that $x \cdot_{\mathbb{K}} y = 0_{\mathbb{K}}$. If $x = 0_{\mathbb{K}}$ then we are done, so suppose $x \neq 0_{\mathbb{K}}$. Using (1.7), we have $1_{\mathbb{K}} = x^{-1} \cdot_{\mathbb{K}} x$. Multiplying this equation with y we obtain

$$y = y \cdot_{\mathbb{K}} 1_{\mathbb{K}} = y \cdot_{\mathbb{K}} (x \cdot_{\mathbb{K}} x^{-1}) = (y \cdot_{\mathbb{K}} x) \cdot_{\mathbb{K}} x^{-1} = 0_{\mathbb{K}} \cdot_{\mathbb{K}} x^{-1} = 0_{\mathbb{K}}$$

where we have used (1.5), the commutativity (1.1) and associativity (1.3) of multiplication as well as (i) from above.

(v) By (1.5), we have $1_{\mathbb{K}} \cdot_{\mathbb{K}} 1_{\mathbb{K}} = 1_{\mathbb{K}}$, hence $1_{\mathbb{K}}$ is the multiplicative inverse of $1_{\mathbb{K}}$ and since the multiplicative inverse is unique, it follows that $(1_{\mathbb{K}})^{-1} = 1_{\mathbb{K}}$. □

For a positive integer $n \in \mathbb{N}$ and an element x of a field $\mathbb{K}$, we write

$$nx = \underbrace{x +_{\mathbb{K}} x +_{\mathbb{K}} x +_{\mathbb{K}} \cdots +_{\mathbb{K}} x}_{n \text{ summands}}.$$

The field $\mathbb{F}_2$ has the property that $2x = 0$ for all $x \in \mathbb{F}_2$. In this case we say the $\mathbb{F}_2$ has characteristic 2. More generally, the smallest positive integer p such that $px = 0_{\mathbb{K}}$ for all $x \in \mathbb{K}$ is called the *characteristic of the field.* In the case where no such integer exists the field is said to have characteristic 0. So $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are fields of characteristic 0. It can be shown that the characteristic of any field is either 0 or a prime number.

A subset $\mathbb{F}$ of a field $\mathbb{K}$ that is itself a field, when equipped with the multiplication and addition of $\mathbb{K}$, is called a *subfield of* $\mathbb{K}$.

Example 1.5

(i) The rational numbers $\mathbb{Q}$ form a subfield of the real numbers $\mathbb{R}$. Furthermore, as we will see below, the real numbers $\mathbb{R}$ can be interpreted as a subfield of the complex numbers $\mathbb{C}$.

(ii) $\mathbb{F}_2$ may be thought of as the subfield of $\mathbb{F}_4$ consisting of $\{0, 1\}$.

Throughout your studies in mathematics, you will encounter various mappings having names ending in *morphism*, such as *homomorphism*, *isomorphism*, *endomorphism*, *automorphism*. This is quite confusing and to make things worse, the precise meaning of $\star$-morphism depends on the structure of the set between which the mapping is defined. But don't worry, we will introduce one $\star$-morphism at a time, starting with *homomorphism*. Broadly speaking, a *homomorphism* between sets $\mathcal{X}$ and $\mathcal{Y}$ that are equipped with some extra structure of the same type is a map $f : \mathcal{X} \to \mathcal{Y}$ that *respects* the extra structure.

In the case of a field $\mathbb{K}$, the extra structure consists of addition $+_{\mathbb{K}}$, multiplication $\cdot_{\mathbb{K}}$, the identity element of multiplication $1_{\mathbb{K}}$ and the identity element of addition $0_{\mathbb{K}}$. A field homomorphism respects this structure. More precisely:

Definition 1.6 (Field homomorphism) Let $\mathbb{F}$ and $\mathbb{K}$ be fields. A *field homomorphism* is a mapping $\chi : \mathbb{F} \to \mathbb{K}$ satisfying $\chi(1_{\mathbb{F}}) = 1_{\mathbb{K}}$ as well as

$$\chi(x +_{\mathbb{F}} y) = \chi(x) +_{\mathbb{K}} \chi(y) \qquad \text{and} \qquad \chi(x \cdot_{\mathbb{F}} y) = \chi(x) \cdot_{\mathbb{K}} \chi(y)$$

for all $x, y \in \mathbb{F}$.

Example 1.7 From the above tables we see that $\chi : \mathbb{F}_2 \to \mathbb{F}_4$ defined by $\chi(1_{\mathbb{F}_2}) = 1_{\mathbb{F}_4}$ and $\chi(0_{\mathbb{F}_2}) = 0_{\mathbb{F}_4}$ is a field homomorphism.

Remark 1.8

(i) We certainly also want that a field homorphism $\chi : \mathbb{F} \to \mathbb{K}$ satisfies $\chi(0_{\mathbb{F}}) = 0_{\mathbb{K}}$. It turns out that we don't have to ask for this in the definition of a field homomorphism, it is automatically satisfied with Definition 1.6. Indeed, we have

$$\chi(0_{\mathbb{F}}) = \chi(0_{\mathbb{F}} +_{\mathbb{F}} 0_{\mathbb{F}}) = \chi(0_{\mathbb{F}}) +_{\mathbb{K}} \chi(0_{\mathbb{F}}).$$

Adding the additive inverse of $\chi(0_{\mathbb{F}})$ in $\mathbb{K}$, we conclude that $0_{\mathbb{K}} = \chi(0_{\mathbb{F}})$.

(ii) A field homomorphism is injective. Suppose $x, y \in \mathbb{F}$ satisfy $\chi(x) = \chi(y)$ so that $\chi(x - y) = 0_{\mathbb{K}}$. Assume $w = x - y \neq 0_{\mathbb{F}}$, then $\chi(w) \cdot_{\mathbb{K}} \chi(w^{-1}) = \chi(1_{\mathbb{F}}) = 1_{\mathbb{K}}$. Since by assumption $\chi(w) = 0_{\mathbb{K}}$, we thus obtain $0_{\mathbb{K}} \cdot_{\mathbb{K}} \chi(w^{-1}) = 1_{\mathbb{K}}$, contradicting Lemma 1.4 (i). It follows that $x = y$ and hence χ is injective.

1.2 Complex numbers

Video Complex numbers

Historically the complex numbers arose from an interest to make sense of the square root of a negative number. We may picture the rational numbers $\mathbb{Q}$ as elements of an infinite number line with an origin 0. Positive numbers extending to the right of the origin and negative numbers to the left. Mathematicians have observed early on that this line of numbers contains elements, such as π or $\sqrt{2}$, that are not quotients. Phrased differently, the rational numbers do not fill out the whole number line, there are gaps consisting of *irrational numbers*. In a sense to be made precise in the Analysis module, the real numbers may be thought of as the union of the rational numbers and the gaps on the number line, resulting in a gap less line of numbers, known as the *complete field of real numbers*.

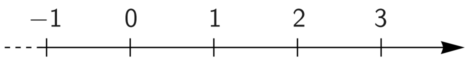

FIGURE 1.1. The real number line.

The square x^2 of a real number x is a non-negative real number, $x^2 \geqslant 0$, hence if we want to define what the square root of a negative number ought to be, we are in trouble, since there are no numbers left on the line of numbers that we might use. The solution is to consider pairs of real numbers instead. A complex number is an ordered pair (x, y) of real numbers $x, y \in \mathbb{R}$. We denote the set of complex numbers by $\mathbb{C}$. We equip $\mathbb{C}$ with the addition defined by the rule

$$(x_1, y_1) +_{\mathbb{C}} (x_2, y_2) = (x_1 + x_2, y_1 + y_2)$$

for all (x_1, y_1) and $(x_2, y_2) \in \mathbb{C}$ and where $+$ on the right denotes the usual addition $+_\mathbb{R}$ of real numbers. Furthermore, we equip $\mathbb{C}$ with the multiplication defined by the rule

$$(x_1, y_1) \cdot_\mathbb{C} (x_2, y_2) = (x_1 \cdot x_2 - y_1 \cdot y_2, x_1 \cdot y_2 + y_1 \cdot x_2). \tag{1.9}$$

for all (x_1, y_1) and $(x_2, y_2) \in \mathbb{C}$ and where $\cdot$ on the right denotes the usual multiplication $\cdot_\mathbb{R}$ of real numbers.

Definition 1.9 (Complex numbers) The set $\mathbb{C}$ together with the operations $+_\mathbb{C}, \cdot_\mathbb{C}$ and $0_\mathbb{C} = (0, 0)$ and $1_\mathbb{C} = (1, 0)$ is called the *field of complex numbers*.

The mapping $\chi : \mathbb{R} \to \mathbb{C}, x \mapsto (x, 0)$ is a field homomorphism. Indeed,

$$\chi(x_1 +_\mathbb{R} x_2) = (x_1 +_\mathbb{R} x_2, 0) = (x_1, 0) +_\mathbb{C} (x_2, 0) = \chi(x_1) +_\mathbb{C} \chi(x_2),$$
$$\chi(x_1 \cdot_\mathbb{R} x_2) = (x_1 \cdot_\mathbb{R} x_2, 0) = (x_1, 0) \cdot_\mathbb{C} (x_2, 0) = \chi(x_1) \cdot_\mathbb{C} \chi(x_2),$$

for all $x_1, x_2 \in \mathbb{R}$ and $\chi(1) = (1, 0) = 1_\mathbb{C}$.

This allows to think of the real numbers $\mathbb{R}$ as the subfield $\{(x, 0)|x \in \mathbb{R}\}$ of the complex numbers $\mathbb{C}$. Because of the injectivity of χ, it is customary to identify x with $\chi(x)$, hence abusing notation, we write $(x, 0) = x$.

Notice that $(0, 1)$ satisfies $(0, 1) \cdot_\mathbb{C} (0, 1) = (-1, 0)$ and hence is a square root of the real number $(-1, 0) = -1$. The number $(0, 1)$ is called the *imaginary unit* and usually denoted by i. Sometimes the notation $\sqrt{-1}$ is also used. Every complex number $(x, y) \in \mathbb{C}$ can now be written as

$$(x, y) = (x, 0) +_\mathbb{C} (0, y) = (x, 0) +_\mathbb{C} \mathrm{i} \cdot_\mathbb{C} (y, 0) = x + \mathrm{i}y,$$

where we follow the usual custom of omitting $\cdot_\mathbb{C}$ and writing $+$ instead of $+_\mathbb{C}$ on the right hand side. With this convention, complex numbers can be manipulated as real numbers, we just need to keep in mind that i satisfies $\mathrm{i}^2 = -1$. For instance, the multiplication of complex numbers $x_1 + \mathrm{i}y_1$ and $x_2 + \mathrm{i}y_2$ gives

$$(x_1 + \mathrm{i}y_1)(x_2 + \mathrm{i}y_2) = x_1x_2 + \mathrm{i}^2y_1y_2 + \mathrm{i}(x_1y_2 + y_1x_2) = x_1x_2 - y_1y_2 + \mathrm{i}(x_1y_2 + y_1x_2)$$

in agreement with (1.9). Here we also follow the usual custom of omitting $\cdot_\mathbb{R}$ on the right hand side.

Definition 1.10 For a complex number $z = x + \mathrm{i}y \in \mathbb{C}$ with $x, y \in \mathbb{R}$ we call

- $\mathrm{Re}(z) = x$ its *real part*;
- $\mathrm{Im}(z) = y$ its *imaginary part*;
- $\bar{z} = x - \mathrm{i}y$ the *complex conjugate of* z;
- $|z| = \sqrt{z\bar{z}} = \sqrt{x^2 + y^2}$ the *absolute value or modulus of* z.

The mapping $z \mapsto \bar{z}$ is called *complex conjugation*.

Remark 1.11

(i) For $z \in \mathbb{C}$ the following statements are equivalent

$$z \in \mathbb{R} \quad \iff \quad \mathrm{Re}(z) = z \quad \iff \quad \mathrm{Im}(z) = 0 \quad \iff \quad z = \bar{z}.$$

(ii) We have $|z| = 0$ if and only if $z = 0$.

Example 1.12 Let $z = \frac{2+5i}{6-i}$. Then

$$z = \frac{(2+5i)\overline{(6-i)}}{(6-i)\overline{(6-i)}} = \frac{(2+5i)(6+i)}{|6-i|^2} = \frac{1}{37}(7+32i),$$

so that $\mathrm{Re}(z) = \frac{7}{37}$ and $\mathrm{Im}(z) = \frac{32}{37}$. Moreover,

$$|z| = \sqrt{\left(\frac{7}{37}\right)^2 + \left(\frac{32}{37}\right)^2} = \sqrt{\frac{29}{37}}.$$

Remark 1.13

(i) We may think of a complex number $z = a + ib$ as a point or a vector in the plane $\mathbb{R}^2$ with x-coordinate a and y-coordinate b.

(ii) The real numbers form the horizontal coordinate axis (the real axis) and the *purely imaginary complex numbers* $\{iy | y \in \mathbb{R}\}$ form the vertical coordinate axis (the imaginary axis).

(iii) The point $\overline{z}$ is obtained by reflecting z along the real axis.

(iv) $|z|$ is the distance of z to the origin $0_{\mathbb{C}} = (0,0) \in \mathbb{C}$

(v) The addition of complex numbers corresponds to the usual vector addition.

(vi) For the geometric significance of the multiplication, we refer the reader to the Analysis module.

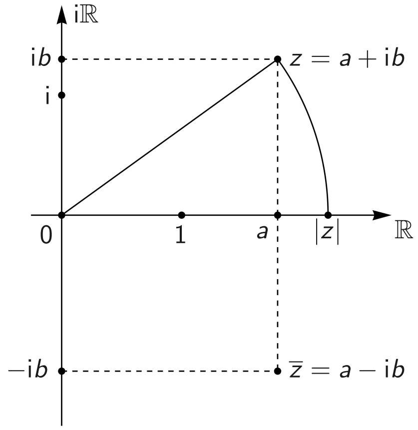

FIGURE 1.2. The complex number plane $\mathbb{C}$

We have the following elementary facts about complex numbers:

Proposition 1.14 *For all $z, w \in \mathbb{C}$ we have*

(i) $\mathrm{Re}(z) = \frac{z+\overline{z}}{2}$, $\mathrm{Im}(z) = \frac{z-\overline{z}}{2i}$;

(ii) $\mathrm{Re}(z+w) = \mathrm{Re}(z) + \mathrm{Re}(w)$, $\mathrm{Im}(z+w) = \mathrm{Im}(z) + \mathrm{Im}(w)$;

(iii) $\overline{z+w} = \overline{z} + \overline{w}$, $\overline{zw} = \overline{z}\,\overline{w}$, $\overline{\overline{z}} = z$;

(iv) $|z|^2 = |\overline{z}|^2 = z\overline{z} = \mathrm{Re}(z)^2 + \mathrm{Im}(z)^2$;

(v) $|zw| = |z||w|$.

Proof Exercise. □

Exercises

Exercise 1.15 Check that $\mathbb{C}$ is indeed a field.

CHAPTER 2

Matrices

2.1 Definitions

WEEK 2

A matrix (plural matrices) is simply a rectangular block of numbers. As we will see below, every matrix gives rise to a mapping sending a finite list of numbers to another finite list of numbers. Mappings arising from matrices are called *linear* and linear mappings are among the most fundamental objects in mathematics. In the Linear Algebra modules we develop the theory of linear maps as well as the theory of *vector spaces*, the natural habitat of linear maps. While this theory may come accross as quite abstract, it is in fact at the heart of many real world applications, including optics and quantum physics, radio astronomy, MP3 and JPEG compression, X-ray crystallography, MRI scans and machine learning, just to name a few.

Throughout the Linear Algebra modules, $\mathbb{K}$ stands for either the real numbers $\mathbb{R}$ or the complex numbers $\mathbb{C}$, but almost all statements are also valid over arbitrary fields.

We start with some definitions. In this chapter, $m, n, \tilde{m}, \tilde{n}$ denote natural numbers.

Definition 2.1 (Matrix)

- A rectangular block of scalars $A_{ij} \in \mathbb{K}$, $1 \leqslant i \leqslant m$, $1 \leqslant j \leqslant n$

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix} \tag{2.1}$$

  is called an $m \times n$ *matrix* with entries in $\mathbb{K}$.
- We also say that $\mathbf{A}$ is an m-by-n matrix, that $\mathbf{A}$ has *size* $m \times n$ and that $\mathbf{A}$ has m *rows* and n *columns*.
- The entry A_{ij} of $\mathbf{A}$ is said to have *row index* i where $1 \leqslant i \leqslant m$, *column index* j where $1 \leqslant j \leqslant n$ and will be referred to as the (i, j)-th entry of $\mathbf{A}$.
- A shorthand notation for (2.1) is $\mathbf{A} = (A_{ij})_{1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n}$.
- For matrices $\mathbf{A} = (A_{ij})_{1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n}$ and $\mathbf{B} = (B_{ij})_{1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n}$ we write $\mathbf{A} = \mathbf{B}$, provided $A_{ij} = B_{ij}$ for all $1 \leqslant i \leqslant m$ and all $1 \leqslant j \leqslant n$.

Definition 2.2 (Set of matrices)

- The set of m-by-n matrices with entries in $\mathbb{K}$ will be denoted by $M_{m,n}(\mathbb{K})$.
- The elements of the set $M_{m,1}(\mathbb{K})$ are called *column vectors of length* m and the elements of the set $M_{1,n}(\mathbb{K})$ are called *row vectors of length* n.

- We will use the Latin alphabet for column vectors and decorate them with an arrow. For a column vector
$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \in M_{m,1}(\mathbb{K})$$
we also use the shorthand notation $\vec{x} = (x_i)_{1 \leqslant i \leqslant m}$ and we write $[\vec{x}]_i$ for the i-th entry of $\vec{x}$, so that $[\vec{x}]_i = x_i$ for all $1 \leqslant i \leqslant m$.
- We will use the Greek alphabet for row vectors and decorate them with an arrow. For a row vector
$$\vec{\xi} = \begin{pmatrix} \xi_1 & \xi_2 & \cdots & \xi_n \end{pmatrix} \in M_{1,n}(\mathbb{K})$$
we also use the shorthand notation $\vec{\xi} = (\xi_i)_{1 \leqslant i \leqslant n}$ and we write $[\vec{\xi}]_i$ for the i-th entry of $\vec{\xi}$, so that $[\vec{\xi}]_i = \xi_i$ for all $1 \leqslant i \leqslant n$.

Remark 2.3 (Notation)

(i) A matrix is always denoted by a bold capital letter, such as $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$.
(ii) The entries of the matrix are denoted by $A_{ij}, B_{ij}, C_{ij}, D_{ij}$, respectively.
(iii) We may think of an $m \times n$ matrix as consisting of n column vectors of length m. The column vectors of the matrix are denoted by $\vec{a}_i, \vec{b}_i, \vec{c}_i, \vec{d}_i$, respectively.
(iv) We may think of an $m \times n$ matrix as consisting of m row vectors of length n. The row vectors of the matrix are denoted by $\vec{\alpha}_i, \vec{\beta}_i, \vec{\gamma}_i, \vec{\delta}_i$, respectively.
(v) For a matrix $\mathbf{A}$ we also write $[\mathbf{A}]_{ij}$ for the (i,j)-th entry of $\mathbf{A}$. So for $\mathbf{A} = (A_{ij})_{1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n}$, we have $[\mathbf{A}]_{ij} = A_{ij}$ for all $1 \leqslant i \leqslant m$, $1 \leqslant j \leqslant n$.

Example 2.4 For
$$\mathbf{A} = \begin{pmatrix} \pi & \sqrt{2} \\ -1 & 5/3 \\ \log 2 & 3 \end{pmatrix} \in M_{3,2}(\mathbb{R}),$$
we have for instance $[\mathbf{A}]_{32} = 3$, $[\mathbf{A}]_{12} = \sqrt{2}$, $[\mathbf{A}]_{21} = -1$ and
$$\vec{a}_1 = \begin{pmatrix} \pi \\ -1 \\ \log 2 \end{pmatrix}, \quad \vec{a}_2 = \begin{pmatrix} \sqrt{2} \\ 5/3 \\ 3 \end{pmatrix}, \quad \vec{\alpha}_2 = \begin{pmatrix} -1 & 5/3 \end{pmatrix}, \quad \vec{\alpha}_3 = \begin{pmatrix} \log 2 & 3 \end{pmatrix}.$$

Recall that for sets $\mathcal{X}$ and $\mathcal{Y}$ we write $\mathcal{X} \times \mathcal{Y}$ for the Cartesian product of $\mathcal{X}$ and $\mathcal{Y}$, defined as the set of ordered pairs (x, y) with $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. Moreover, $\mathcal{X} \times \mathcal{X}$ is usually denoted as $\mathcal{X}^2$. Likewise, for a natural number $n \in \mathbb{N}$, we write $\mathcal{X}^n$ for the set of ordered lists consisting of n elements of $\mathcal{X}$. We will also refer to ordered lists consisting of n elements as *n-tuples*. The elements of $\mathcal{X}^n$ are denoted by $(x_1, x_2, \ldots, x_n)$ with $x_i \in \mathcal{X}$ for all $1 \leqslant i \leqslant n$. In particular, for all $n \in \mathbb{N}$ we have a bijective map from $\mathbb{K}^n$ to $M_{n,1}(\mathbb{K})$ given by

$$(x_1, \ldots, x_n) \mapsto \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}. \tag{2.2}$$

For this reason, we also write $\mathbb{K}^n$ for the set of column vectors of length n with entries in $\mathbb{K}$. The set of row vectors of length n with entries in $\mathbb{K}$ will be denoted by $\mathbb{K}_n$.

Definition 2.5 (Special matrices and vectors)

- The *zero matrix* $\mathbf{0}_{m,n}$ is the $m \times n$ matrix whose entries are all zero. We will also write $\mathbf{0}_n$ for the $n \times n$-matrix whose entries are all zero.
- Matrices with equal number n of rows and columns are known as *square matrices*.
- An entry A_{ij} of a square matrix $\mathbf{A} \in M_{n,n}(\mathbb{K})$ is said to be a *diagonal entry* if $i = j$ and an *off-diagonal entry* otherwise. A matrix whose off-diagonal entries are all zero is said to be *diagonal*.
- We write $\mathbf{1}_n$ for the diagonal $n \times n$ matrix whose diagonal entries are all equal to 1. Using the so-called *Kronecker delta* defined by the rule

$$\delta_{ij} = \begin{cases} 1 & i = j, \\ 0 & i \neq j, \end{cases}$$

  we have $[\mathbf{1}_n]_{ij} = \delta_{ij}$ for all $1 \leqslant i, j \leqslant n$. The matrix $\mathbf{1}_n$ is called the *unit matrix or identity matrix* of size n.
- The *standard basis of* $\mathbb{K}^n$ is the set $\{\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n\}$ consisting of the column vectors of the identity matrix $\mathbf{1}_n$ of size n.
- The *standard basis of* $\mathbb{K}_n$ is the set $\{\vec{\varepsilon}_1, \vec{\varepsilon}_2, \dots, \vec{\varepsilon}_n\}$ consisting of the row vectors of the identity matrix $\mathbf{1}_n$ of size n.

Example 2.6

(i) Special matrices:

$$\mathbf{0}_{2,3} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \mathbf{1}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \mathbf{1}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

(ii) The standard basis of $\mathbb{K}^3$ is $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$, where

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \vec{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

(iii) The standard basis of $\mathbb{K}_3$ is $\{\vec{\varepsilon}_1, \vec{\varepsilon}_2, \vec{\varepsilon}_3\}$, where

$$\vec{\varepsilon}_1 = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}, \quad \vec{\varepsilon}_2 = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad \vec{\varepsilon}_3 = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}.$$

2.2 Matrix operations

We can multiply a matrix $\mathbf{A} \in M_{m,n}(\mathbb{K})$ with a *scalar* $s \in \mathbb{K}$. This amounts to multiplying each entry of $\mathbf{A}$ with s:

Definition 2.7 Scalar multiplication in $M_{m,n}(\mathbb{K})$ is the map

$$\cdot_{M_{m,n}(\mathbb{K})} : \mathbb{K} \times M_{m,n}(\mathbb{K}) \to M_{m,n}(\mathbb{K}), \qquad (s, \mathbf{A}) \mapsto s \cdot_{M_{m,n}(\mathbb{K})} \mathbf{A}$$

defined by the rule

$$s \cdot_{M_{m,n}(\mathbb{K})} \mathbf{A} = (s \cdot_{\mathbb{K}} A_{ij})_{1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n} \in M_{m,n}(\mathbb{K}), \tag{2.3}$$

where $s \cdot_{\mathbb{K}} A_{ij}$ denotes the field multiplication of scalars $s, A_{ij} \in \mathbb{K}$.

Remark 2.8 Here we multiply with s *from the left*. Likewise, we define $\mathbf{A} \cdot_{M_{m,n}(\mathbb{K})} s = (A_{ij} \cdot_{\mathbb{K}} s)_{1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n}$, that is, we multiply *from the right*. Of course, since multiplication of scalars is commutative, we have $s \cdot_{M_{m,n}(\mathbb{K})} \mathbf{A} = \mathbf{A} \cdot_{M_{m,n}(\mathbb{K})} s$, that is, left multiplication and right multiplication gives the same matrix. Be aware that this is not true in every number system. An example that you might encounter later on are the so-called *quaternions*, where multiplication fails to be commutative.

The sum of matrices **A** and **B** of *identical size* is defined as follows:

Definition 2.9 Addition in $M_{m,n}(\mathbb{K})$ is the map

$$+_{M_{m,n}(\mathbb{K})} : M_{m,n}(\mathbb{K}) \times M_{m,n}(\mathbb{K}) \to M_{m,n}(\mathbb{K}), \qquad (\mathbf{A}, \mathbf{B}) \mapsto \mathbf{A} +_{M_{m,n}(\mathbb{K})} \mathbf{B}$$

defined by the rule

$$\mathbf{A} +_{M_{m,n}(\mathbb{K})} \mathbf{B} = (A_{ij} +_{\mathbb{K}} B_{ij})_{1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n} \in M_{m,n}(\mathbb{K}), \tag{2.4}$$

where $A_{ij} +_{\mathbb{K}} B_{ij}$ denotes the field addition of scalars $A_{ij}, B_{ij} \in \mathbb{K}$.

Remark 2.10 (Abusing notation)

- Field addition takes two scalars and produces another scalar, thus it is a map $\mathbb{K} \times \mathbb{K} \to \mathbb{K}$, whereas addition of matrices is a map $M_{m,n}(\mathbb{K}) \times M_{m,n}(\mathbb{K}) \to M_{m,n}(\mathbb{K})$. For this reason we wrote $+_{M_{m,n}(\mathbb{K})}$ above in order to distinguish matrix addition from field addition of scalars. Of course, it is quite cumbersome to always write $+_{M_{m,n}(\mathbb{K})}$ and $+_{\mathbb{K}}$, so we follow the usual custom of writing $+$, both for field addition of scalars and for matrix addition, trusting that the reader is aware of the difference.
- Likewise, we simply write $\cdot$ instead of $\cdot_{M_{m,n}(\mathbb{K})}$ or omit the dot entirely, so that $s \cdot \mathbf{A} = s\mathbf{A} = s \cdot_{M_{m,n}(\mathbb{K})} \mathbf{A}$ for $s \in \mathbb{K}$ and $\mathbf{A} \in M_{m,n}(\mathbb{K})$.

Example 2.11

- Multiplication of a matrix by a scalar:

$$5\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} 5 = \begin{pmatrix} 5 \cdot 1 & 5 \cdot 2 \\ 5 \cdot 3 & 5 \cdot 4 \end{pmatrix} = \begin{pmatrix} 5 & 10 \\ 15 & 20 \end{pmatrix}.$$

- Addition of matrices:

$$\begin{pmatrix} 3 & -5 \\ -2 & 8 \end{pmatrix} + \begin{pmatrix} -3 & 8 \\ 7 & 10 \end{pmatrix} = \begin{pmatrix} 0 & 3 \\ 5 & 18 \end{pmatrix}.$$

If the number of columns of a matrix **A** is *equal* to the number of rows of a matrix **B**, we define the matrix product **AB** of **A** and **B** as follows:

Definition 2.12 (Matrix multiplication — Video) Let $\mathbf{A} \in M_{m,n}(\mathbb{K})$ be an m-by-n matrix and $\mathbf{B} \in M_{n,\tilde{m}}(\mathbb{K})$ be an n-by-$\tilde{m}$ matrix. The *matrix product* of **A** and **B** is the

m-by-$\tilde{m}$ matrix $\mathbf{AB} \in M_{m,\tilde{m}}(\mathbb{K})$ whose entries are defined by the rule

$$[\mathbf{AB}]_{ik} = A_{i1}B_{1k} + A_{i2}B_{2k} + \cdots + A_{in}B_{nk} = \sum_{j=1}^{n} A_{ij}B_{jk} = \sum_{j=1}^{n} [\mathbf{A}]_{ij}[\mathbf{B}]_{jk}.$$

for all $1 \leqslant i \leqslant m$ and all $1 \leqslant k \leqslant \tilde{m}$.

Remark 2.13 (Pairing of row and column vectors) We may define a pairing $\mathbb{K}_n \times \mathbb{K}^n \to \mathbb{K}$ of a row vector of length n and a column vector of length n by the rule

$$(\vec{\xi}, \vec{x}) \mapsto \vec{\xi}\vec{x} = \xi_1 x_1 + \xi_2 x_2 + \cdots + \xi_n x_n$$

for all $\vec{\xi} = (\xi_i)_{1 \leqslant i \leqslant n} \in \mathbb{K}_n$ and for all $\vec{x} = (x_i)_{1 \leqslant i \leqslant n} \in \mathbb{K}^n$. So we multiply the first entry of $\vec{\xi}$ with the first entry of $\vec{x}$, add the product of the second entry of $\vec{\xi}$ and the second entry of $\vec{x}$ and continue in this fashion until the last entry of $\vec{\xi}$ and $\vec{x}$.

The (i, j)-th entry of the matrix product of $\mathbf{A} \in M_{m,n}(\mathbb{K})$ and $\mathbf{B} \in M_{n,\tilde{m}}(\mathbb{K})$ is then given by the pairing

$$[\mathbf{AB}]_{ij} = \vec{\alpha}_i \vec{b}_j$$

of the i-th row vector $\vec{\alpha}_i$ of $\mathbf{A}$ and the j-th column vector $\vec{b}_j$ of $\mathbf{B}$.

Remark 2.14 (Matrix multiplication is not commutative — Video) If $\mathbf{A}$ is a m-by-n matrix and $\mathbf{B}$ a n-by-m matrix, then both $\mathbf{AB}$ and $\mathbf{BA}$ are defined, but in general $\mathbf{AB} \neq \mathbf{BA}$ since $\mathbf{AB}$ is an m-by-m matrix and $\mathbf{BA}$ is an n-by-n matrix. Even when $n = m$ so that both $\mathbf{A}$ and $\mathbf{B}$ are square matrices, it is false in general that $\mathbf{AB} = \mathbf{BA}$.

The matrix operations have the following properties:

Proposition 2.15 (Properties of matrix operations)

- $\mathbf{0}_{m,n} + \mathbf{A} = \mathbf{A}$ *for all* $\mathbf{A} \in M_{m,n}(\mathbb{K})$;
- $\mathbf{1}_m\mathbf{A} = \mathbf{A}$ *and* $\mathbf{A}\mathbf{1}_n = \mathbf{A}$ *for all* $\mathbf{A} \in M_{m,n}(\mathbb{K})$;
- $\mathbf{0}_{\tilde{m},m}\mathbf{A} = \mathbf{0}_{\tilde{m},n}$ *and* $\mathbf{A}\mathbf{0}_{n,\tilde{m}} = \mathbf{0}_{m,\tilde{m}}$ *for all* $\mathbf{A} \in M_{m,n}(\mathbb{K})$;
- $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$ *and* $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$ *for all* $\mathbf{A}, \mathbf{B}, \mathbf{C} \in M_{m,n}(\mathbb{K})$;
- $0 \cdot \mathbf{A} = \mathbf{0}_{m,n}$ *for all* $\mathbf{A} \in M_{m,n}(\mathbb{K})$;
- $(s_1 s_2)\mathbf{A} = s_1(s_2\mathbf{A})$ *for all* $\mathbf{A} \in M_{m,n}(\mathbb{K})$ *and all* $s_1, s_2 \in \mathbb{K}$;
- $\mathbf{A}(s\mathbf{B}) = s(\mathbf{AB}) = (s\mathbf{A})\mathbf{B}$ *for all* $\mathbf{A} \in M_{m,n}(\mathbb{K})$ *and all* $\mathbf{B} \in M_{n,\tilde{m}}(\mathbb{K})$ *and all* $s \in \mathbb{K}$;
- $s(\mathbf{A} + \mathbf{B}) = s\mathbf{A} + s\mathbf{B}$ *for all* $\mathbf{A}, \mathbf{B} \in M_{m,n}(\mathbb{K})$ *and* $s \in \mathbb{K}$;
- $(s_1 + s_2)\mathbf{A} = s_1\mathbf{A} + s_2\mathbf{A}$ *for all* $\mathbf{A} \in M_{m,n}(\mathbb{K})$ *and for all* $s_1, s_2 \in \mathbb{K}$;
- $(\mathbf{B} + \mathbf{C})\mathbf{A} = \mathbf{BA} + \mathbf{CA}$ *for all* $\mathbf{B}, \mathbf{C} \in M_{\tilde{m},m}(\mathbb{K})$ *and for all* $\mathbf{A} \in M_{m,n}(\mathbb{K})$;
- $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$ *for all* $\mathbf{A} \in M_{\tilde{m},m}(\mathbb{K})$ *and for all* $\mathbf{B}, \mathbf{C} \in M_{m,n}(\mathbb{K})$.

Proof We only show the second and the last property. The proofs of the remaining ones are similar and/or elementary consequences of the properties of addition and multiplication of scalars.

To show the second property consider $\mathbf{A} \in M_{m,n}(\mathbb{K})$. Then, by definition, we have for all $1 \leqslant k \leqslant m$ and all $1 \leqslant j \leqslant n$

$$[\mathbf{1}_m\mathbf{A}]_{kj} = \sum_{i=1}^{m} [\mathbf{1}_m]_{ki}[\mathbf{A}]_{ij} = \sum_{i=1}^{m} \delta_{ki} A_{ij} = A_{kj} = [\mathbf{A}]_{kj},$$

where the second last equality uses that δ_{ki} is 0 unless $i = k$, in which case $\delta_{kk} = 1$. We conclude that $\mathbf{1}_m\mathbf{A} = \mathbf{A}$. Likewise, we obtain for all $1 \leqslant i \leqslant m$ and all $1 \leqslant k \leqslant n$

$$[\mathbf{A}\mathbf{1}_n]_{ik} = \sum_{j=1}^{n} [\mathbf{A}]_{ij}[\mathbf{1}_n]_{jk} = \sum_{j=1}^{n} A_{ij}\delta_{jk} = A_{ik} = [\mathbf{A}]_{ik}$$

so that $\mathbf{A}\mathbf{1}_n = \mathbf{A}$. The identities

$$\boxed{\sum_{i=1}^{m} \delta_{ki} A_{ij} = A_{kj} \qquad \text{and} \qquad \sum_{j=1}^{n} A_{ij}\delta_{jk} = A_{ik}}$$

are used repeatedly in Linear Algebra, so make sure you understand them.

For the last property, applying the definition of matrix multiplication gives

$$\mathbf{AB} = \left(\sum_{i=1}^{m} A_{ki}B_{ij}\right)_{1\leqslant k\leqslant \tilde{m}, 1\leqslant j\leqslant n} \qquad \text{and} \qquad \mathbf{AC} = \left(\sum_{i=1}^{m} A_{ki}C_{ij}\right)_{1\leqslant k\leqslant \tilde{m}, 1\leqslant j\leqslant n},$$

so that

$$\begin{aligned}\mathbf{AB} + \mathbf{AC} &= \left(\sum_{i=1}^{m} A_{ki}B_{ij} + \sum_{i=1}^{m} A_{ki}C_{ij}\right)_{1\leqslant k\leqslant \tilde{m}, 1\leqslant j\leqslant n} \\ &= \left(\sum_{i=1}^{m} A_{ki}\left(B_{ij} + C_{ij}\right)\right)_{1\leqslant k\leqslant \tilde{m}, 1\leqslant j\leqslant n} = \mathbf{A}(\mathbf{B} + \mathbf{C}),\end{aligned}$$

where we use that

$$\mathbf{B} + \mathbf{C} = (B_{ij} + C_{ij})_{1\leqslant i\leqslant m, 1\leqslant j\leqslant n}.$$

□

Finally, we may flip a matrix along its "diagonal entries", that is, we interchange the role of rows and columns. More precisely:

Definition 2.16 (Transpose of a matrix)

- The *transpose* of a matrix $\mathbf{A} \in M_{m,n}(\mathbb{K})$ is the matrix $\mathbf{A}^T \in M_{n,m}(\mathbb{K})$ satisfying

$$\left[\mathbf{A}^T\right]_{ij} = [\mathbf{A}]_{ji}$$

  for all $1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant m$.
- A square matrix $\mathbf{A} \in M_{n,n}(\mathbb{K})$ that satisfies $\mathbf{A} = \mathbf{A}^T$ is called *symmetric*.
- A square matrix $\mathbf{A} \in M_{n,n}(\mathbb{K})$ that satisfies $\mathbf{A} = -\mathbf{A}^T$ is called *anti-symmetric*.

Example 2.17 If

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}, \quad \text{then} \quad \mathbf{A}^T = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}.$$

Remark 2.18 (Properties of the transpose)

(i) For $\mathbf{A} \in M_{m,n}(\mathbb{K})$ we have by definition $(\mathbf{A}^T)^T = \mathbf{A}$.

(ii) For $\mathbf{A} \in M_{m,n}(\mathbb{K})$ and $\mathbf{B} \in M_{n,\tilde{m}}(\mathbb{K})$, we have

$$(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T.$$

Indeed, by definition we have for all $1 \leqslant i \leqslant \tilde{m}$ and all $1 \leqslant j \leqslant m$

$$\left[(\mathbf{AB})^T\right]_{ij} = [\mathbf{AB}]_{ji} = \sum_{k=1}^{n}[\mathbf{A}]_{jk}[\mathbf{B}]_{ki} = \sum_{k=1}^{n}\left[\mathbf{B}^T\right]_{ik}\left[\mathbf{A}^T\right]_{kj} = \left[\mathbf{B}^T\mathbf{A}^T\right]_{ij}.$$

2.3 Mappings associated to matrices

Definition 2.19 (Mapping associated to a matrix) For an $(m \times n)$-matrix $\mathbf{A} = (A_{ij})_{1\leqslant i\leqslant m, 1\leqslant j\leqslant n} \in M_{m,n}(\mathbb{K})$ with column vectors $\vec{a}_1, \dots, \vec{a}_n \in \mathbb{K}^m$ we define a mapping

$$f_{\mathbf{A}} : \mathbb{K}^n \to \mathbb{K}^m, \qquad \vec{x} \mapsto \mathbf{A}\vec{x},$$

where the column vector $\mathbf{A}\vec{x} \in \mathbb{K}^m$ is obtained by matrix multiplication of the matrix $\mathbf{A} \in M_{m,n}(\mathbb{K})$ and the column vector $\vec{x} = (x_i)_{1\leqslant i\leqslant n} \in \mathbb{K}^n$

$$\mathbf{A}\vec{x} = \vec{a}_1x_1 + \vec{a}_2x_2 + \cdots + \vec{a}_nx_n = \begin{pmatrix} A_{11}x_1 + A_{12}x_2 + \cdots + A_{1n}x_n \\ A_{21}x_1 + A_{22}x_2 + \cdots + A_{2n}x_n \\ \vdots \\ A_{m1}x_1 + A_{m2}x_2 + \cdots + A_{mn}x_n \end{pmatrix}.$$

Recall that if $f : \mathcal{X} \to \mathcal{Y}$ and $g : \mathcal{X} \to \mathcal{Y}$ are mappings from a set $\mathcal{X}$ into a set $\mathcal{Y}$, then we write $f = g$ if $f(x) = g(x)$ for all elements $x \in \mathcal{X}$.

The matrix $\mathbf{A} \in M_{m,n}(\mathbb{K})$ uniquely determines the mapping $f_{\mathbf{A}}$:

Proposition 2.20 *Let $\mathbf{A}, \mathbf{B} \in M_{m,n}(\mathbb{K})$. Then $f_{\mathbf{A}} = f_{\mathbf{B}}$ if and only if $\mathbf{A} = \mathbf{B}$.*

Proof If $\mathbf{A} = \mathbf{B}$, then $A_{ij} = B_{ij}$ for all $1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n$, hence we conclude that $f_{\mathbf{A}} = f_{\mathbf{B}}$. In order to show the converse direction we consider the standard basis $\vec{e}_i = (\delta_{ij})_{1\leqslant j\leqslant n}$, $i = 1, \dots, n$ of $\mathbb{K}^n$. Now by assumption

$$f_{\mathbf{A}}(\vec{e}_i) = \begin{pmatrix} A_{1i} \\ A_{2i} \\ \vdots \\ A_{mi} \end{pmatrix} = f_{\mathbf{B}}(\vec{e}_i) = \begin{pmatrix} B_{1i} \\ B_{2i} \\ \vdots \\ B_{mi} \end{pmatrix}.$$

Since this holds for all $i = 1, \dots, n$, we conclude $A_{ij} = B_{ij}$ for all $j = 1, \dots, m$ and $i = 1, \dots, n$. Therefore, we have $\mathbf{A} = \mathbf{B}$, as claimed. □

Recall that if $f : \mathcal{X} \to \mathcal{Y}$ is a mapping from a set $\mathcal{X}$ into a set $\mathcal{Y}$ and $g : \mathcal{Y} \to \mathcal{Z}$ a mapping from $\mathcal{Y}$ into a set $\mathcal{Z}$, we can consider the *composition* of g and f

$$g \circ f : \mathcal{X} \to \mathcal{Z}, \qquad x \mapsto g(f(x)).$$

The motivation for the Definition 2.12 of matrix multiplication is given by the following theorem which states that the mapping $f_{\mathbf{AB}}$ associated to the matrix product $\mathbf{AB}$ is the composition of the mapping $f_{\mathbf{A}}$ associated to the matrix $\mathbf{A}$ and the mapping $f_{\mathbf{B}}$ associated to the matrix $\mathbf{B}$. More precisely:

Theorem 2.21 *Let $\mathbf{A} \in M_{m,n}(\mathbb{K})$ and $\mathbf{B} \in M_{n,\tilde{m}}(\mathbb{K})$ so that $f_{\mathbf{A}} : \mathbb{K}^n \to \mathbb{K}^m$ and $f_{\mathbf{B}} : \mathbb{K}^{\tilde{m}} \to \mathbb{K}^n$ and $f_{\mathbf{AB}} : \mathbb{K}^{\tilde{m}} \to \mathbb{K}^m$. Then $f_{\mathbf{AB}} = f_{\mathbf{A}} \circ f_{\mathbf{B}}$.*

Proof For $\vec{x} = (x_k)_{1 \leqslant k \leqslant \tilde{m}} \in \mathbb{K}^{\tilde{m}}$ we write $\vec{y} = f_{\mathbf{B}}(\vec{x})$. Then, by definition, $\vec{y} = \mathbf{B}\vec{x} = (y_j)_{1 \leqslant j \leqslant n}$ where

$$y_j = B_{j1}x_1 + B_{j2}x_2 + \cdots + B_{j\tilde{m}}x_{\tilde{m}} = \sum_{k=1}^{\tilde{m}} B_{jk}x_k. \tag{2.5}$$

Hence writing $\vec{z} = f_{\mathbf{A}}(\vec{y}) = \mathbf{A}\vec{y}$, we have $\vec{z} = (z_i)_{1 \leqslant i \leqslant m}$, where

$$\begin{aligned} z_i &= A_{i1}y_1 + A_{i2}y_2 + \cdots + A_{in}y_n = \sum_{j=1}^{n} A_{ij}y_j = \sum_{j=1}^{n} A_{ij} \sum_{k=1}^{\tilde{m}} B_{jk}x_k \\ &= \sum_{k=1}^{\tilde{m}} \left(\sum_{j=1}^{n} A_{ij}B_{jk} \right) x_k \end{aligned}$$

and where have used (2.5). Since $\mathbf{AB} = (C_{ik})_{1 \leqslant i \leqslant m, 1 \leqslant k \leqslant \tilde{m}}$ with

$$C_{ik} = \sum_{j=1}^{n} A_{ij}B_{jk},$$

we conclude that $\vec{z} = f_{\mathbf{AB}}(\vec{x})$, as claimed. □

Combining Theorem 2.21 and Proposition 2.20, we also obtain:

Corollary 2.22 *Let $\mathbf{A} \in M_{m,n}(\mathbb{K})$, $\mathbf{B} \in M_{n,\tilde{m}}(\mathbb{K})$ and $\mathbf{C} \in M_{\tilde{m},\tilde{n}}(\mathbb{K})$. Then*

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}),$$

that is, the matrix product is associative.

Proof Using Proposition 2.20 it is enough to show that

$$f_{\mathbf{AB}} \circ f_{\mathbf{C}} = f_{\mathbf{A}} \circ f_{\mathbf{BC}}.$$

Using Theorem 2.21, we get for all $\vec{x} \in \mathbb{K}^{\tilde{n}}$

$$(f_{\mathbf{AB}} \circ f_{\mathbf{C}})(\vec{x}) = f_{\mathbf{AB}}(f_{\mathbf{C}}(\vec{x})) = f_{\mathbf{A}}(f_{\mathbf{B}}(f_{\mathbf{C}}(\vec{x}))) = f_{\mathbf{A}}(f_{\mathbf{BC}}(\vec{x})) = (f_{\mathbf{A}} \circ f_{\mathbf{BC}})(\vec{x}).$$

□

Remark 2.23 For all $\mathbf{A} \in M_{m,n}(\mathbb{K})$, the mapping $f_{\mathbf{A}} : \mathbb{K}^n \to \mathbb{K}^m$ satisfies the following two *very important* properties

$$\begin{aligned} f_{\mathbf{A}}(\vec{x} + \vec{y}) &= f_{\mathbf{A}}(\vec{x}) + f_{\mathbf{A}}(\vec{y}), && \text{(additivity)}, \\ f_{\mathbf{A}}(s \cdot \vec{x}) &= s \cdot f_{\mathbf{A}}(\vec{x}), && \text{(1-homogeneity)}, \end{aligned} \tag{2.6}$$

for all $\vec{x}, \vec{y} \in \mathbb{K}^n$ and $s \in \mathbb{K}$. Indeed, using Proposition 2.15 we have

$$f_{\mathbf{A}}(\vec{x} + \vec{y}) = \mathbf{A}(\vec{x} + \vec{y}) = \mathbf{A}\vec{x} + \mathbf{A}\vec{y} = f_{\mathbf{A}}(\vec{x}) + f_{\mathbf{A}}(\vec{y})$$

and

$$f_{\mathbf{A}}(s \cdot x) = \mathbf{A}(sx) = s \cdot (\mathbf{A}x) = s \cdot f_{\mathbf{A}}(x).$$

Mappings satisfying (2.6) are called *linear*.

Example 2.24 Notice that "most" functions $\mathbb{R} \to \mathbb{R}$ are neither additive nor 1-homogeneous. As an example, consider a mapping $f : \mathbb{R} \to \mathbb{R}$ which satisfies the 1-homogeneity property. Let $a = f(1) \in \mathbb{R}$. Then the 1-homogeneity implies that for all $x \in \mathbb{R} = \mathbb{R}^1$ we have

$$f(x) = f(x \cdot 1) = x \cdot f(1) = a \cdot x,$$

showing that the only 1-homogeneous mappings from $\mathbb{R} \to \mathbb{R}$ are of the form $x \mapsto ax$, where a is a real number. In particular, $\sin$, $\cos$, $\tan$, $\log$, $\exp$, $\sqrt{\ }$ *and all polynomials of degree higher than one are not linear*.

CHAPTER 3

Vector spaces and linear maps

3.1 Vector spaces

WEEK 3

We have seen that to every matrix $\mathbf{A} \in M_{m,n}(\mathbb{K})$ we can associate a mapping $f_{\mathbf{A}} : \mathbb{K}^n \to \mathbb{K}^m$ which is additive and 1-homogeneous. Another example of a mapping which is additive and 1-homogeneous is the derivative. Consider $\mathsf{P}(\mathbb{R})$, the set of polynomial functions in one real variable, which we denote by x, with real coefficients. That is, an element $p \in \mathsf{P}(\mathbb{R})$ is a function

$$p : \mathbb{R} \to \mathbb{R}, \qquad x \mapsto a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = \sum_{k=0}^{n} a_k x^k,$$

where $n \in \mathbb{N}$ and the *coefficients* $a_k \in \mathbb{R}$ for $k = 0, 1, \ldots, n$. The largest $m \in \mathbb{N} \cup \{0\}$ such that $a_m \neq 0$ is called the *degree* of p. Notice that we consider polynomials of arbitrary, but *finite degree*. A *power series* $x \mapsto \sum_{k=0}^{\infty} a_k x^k$, that you encounter in the Analysis module, is not a polynomial, unless only finitely many of its coefficients are different from zero.

Clearly, we can multiply p with a real number $s \in \mathbb{R}$ to obtain a new polynomial $s \cdot_{\mathsf{P}(\mathbb{R})} p$

$$s \cdot_{\mathsf{P}(\mathbb{R})} p : \mathbb{R} \to \mathbb{R}, \qquad x \mapsto s \cdot p(x) \tag{3.1}$$

so that $(s \cdot_{\mathsf{P}(\mathbb{R})} p)(x) = \sum_{k=0}^{n} s a_k x^k$ for all $x \in \mathbb{R}$. Here $s \cdot p(x)$ is the usual multiplication of the real numbers s and $p(x)$. If we consider another polynomial

$$q : \mathbb{R} \to \mathbb{R}, \qquad x \mapsto \sum_{k=0}^{n} b_k x^k$$

with $b_k \in \mathbb{R}$ for $k = 0, 1, \ldots, n$, the sum of the polynomials p and q is the polynomial

$$p +_{\mathsf{P}(\mathbb{R})} q : \mathbb{R} \to \mathbb{R}, \qquad x \mapsto p(x) + q(x) \tag{3.2}$$

so that $(p +_{\mathsf{P}(\mathbb{R})} q)(x) = \sum_{k=0} (a_k + b_k) x^k$ for all $x \in \mathbb{R}$. Here $p(x) + q(x)$ is the usual addition of the real numbers $p(x)$ and $q(x)$. We will henceforth omit writing $+_{\mathsf{P}(\mathbb{R})}$ and $\cdot_{\mathsf{P}(\mathbb{R})}$ and simply write $+$ and $\cdot$.

We may think of the derivative with respect to the variable x as a mapping

$$\frac{\mathrm{d}}{\mathrm{d}x} : \mathsf{P}(\mathbb{R}) \to \mathsf{P}(\mathbb{R}).$$

Now recall that the derivative satisfies

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x}(p+q) &= \frac{\mathrm{d}}{\mathrm{d}x}(p) + \frac{\mathrm{d}}{\mathrm{d}x}(q) && \text{(additivity)}, \\ \frac{\mathrm{d}}{\mathrm{d}x}(s \cdot p) &= s \cdot \frac{\mathrm{d}}{\mathrm{d}x}(p) && \text{(1-homogeneity)}. \end{aligned} \tag{3.3}$$

Comparing (2.6) with (3.3) we notice that the polynomials p, q take the role of the vectors $\vec{x}, \vec{y}$ and the derivative takes the role of the mapping $f_{\mathbf{A}}$. This suggests that the mental image of a vector being an arrow in $\mathbb{K}^n$ is too narrow and that we ought to come up with a generalisation of the space $\mathbb{K}^n$ whose elements are *abstract vectors*.

Video Vector spaces

In order to define the notion of a space of abstract vectors, we may ask what key structure the set of (column) vectors $\mathbb{K}^n$ carries. On $\mathbb{K}^n$, we have two fundamental operations,

$$\begin{aligned} + &: \mathbb{K}^n \times \mathbb{K}^n \to \mathbb{K}^n \quad (\vec{x}, \vec{y}) \mapsto \vec{x} + \vec{y}, \quad \text{(vector addition)}, \\ \cdot &: \mathbb{K} \times \mathbb{K}^n \to \mathbb{K}^n, \quad (s, \vec{x}) \mapsto s \cdot \vec{x}, \quad \text{(scalar multiplication)}. \end{aligned}$$

A *vector space* is roughly speaking a set where these two operations are defined and obey the expected properties. More precisely:

Definition 3.1 (Vector space) A $\mathbb{K}$-*vector space*, or *vector space over* $\mathbb{K}$ is a set V with a distinguished element 0_V (called the zero vector) and two operations

$$+_V : V \times V \to V \quad (v_1, v_2) \mapsto v_1 +_V v_2 \quad \text{(vector addition)}$$

and

$$\cdot_V : \mathbb{K} \times V \to V \quad (s, v) \mapsto s \cdot_V v \quad \text{(scalar multiplication)},$$

so that the following properties hold:

- Commutativity of vector addition

$$v_1 +_V v_2 = v_2 +_V v_1 \quad \text{(for all } v_1, v_2 \in V);$$

- Associativity of vector addition

$$v_1 +_V (v_2 +_V v_3) = (v_1 +_V v_2) +_V v_3 \quad \text{(for all } v_1, v_2, v_3 \in V);$$

- Identity element of vector addition

$$0_V +_V v = v +_V 0_V = v \quad \text{(for all } v \in V); \tag{3.4}$$

- Identity element of scalar multiplication

$$1 \cdot_V v = v \quad \text{(for all } v \in V);$$

- Scalar multiplication by zero

$$0 \cdot_V v = 0_V \quad \text{(for all } v \in V); \tag{3.5}$$

- Compatibility of scalar multiplication with field multiplication

$$(s_1 s_2) \cdot_V v = s_1 \cdot_V (s_2 \cdot_V v) \quad \text{(for all } s_1, s_2 \in \mathbb{K}, v \in V);$$

- Distributivity of scalar multiplication with respect to vector addition

$$s \cdot_V (v_1 +_V v_2) = s \cdot_V v_1 +_V s \cdot_V v_2 \quad \text{(for all } s \in \mathbb{K}, v_1, v_2 \in V);$$

- Distributivity of scalar multiplication with respect to field addition

$$(s_1 + s_2) \cdot_V v = s_1 \cdot_V v +_V s_2 \cdot_V v \quad \text{(for all } s_1, s_2 \in \mathbb{K}, v \in V).$$

The elements of V are called *vectors*.

Example 3.2 (Field) A field $\mathbb{K}$ is a $\mathbb{K}$-vector space. We may take $V = \mathbb{K}$, $0_V = 0_{\mathbb{K}}$ and equip V with addition $+_V = +_{\mathbb{K}}$ and scalar multiplication $\cdot_V = \cdot_{\mathbb{K}}$. Then the properties of a field imply that $V = \mathbb{K}$ is a $\mathbb{K}$-vector space.

Example 3.3 (Vector space of matrices) Let $V = M_{m,n}(\mathbb{K})$ denote the set of $m \times n$-matrices with entries in $\mathbb{K}$ and $0_V = \mathbf{0}_{m,n}$ denote the zero vector. It follows from Proposition 2.15 that V equipped with addition $+_V : V \times V \to V$ defined by (2.4) and scalar multiplication $\cdot_V : \mathbb{K} \times V \to V$ defined by (2.3) is a $\mathbb{K}$-vector space. In particular, the set of column vectors $\mathbb{K}^n = M_{n,1}(\mathbb{K})$ is a $\mathbb{K}$-vector space as well.

Example 3.4 (Vector space of polynomials) The set $\mathsf{P}(\mathbb{R})$ of polynomials in one real variable and with real coefficients is an $\mathbb{R}$-vector space, when equipped with addition and scalar multiplication as defined in (3.1) and (3.2) and when the zero vector $0_{\mathsf{P}(\mathbb{R})}$ is defined to be the *zero polynomial* $o : \mathbb{R} \to \mathbb{R}$, that is, the polynomial satisfying $o(x) = 0$ for all $x \in \mathbb{R}$.

More generally, functions form a vector space:

Example 3.5 (Vector space of functions) We follow the convention of calling a mapping with values in $\mathbb{K}$ a *function*. Let $I \subset \mathbb{R}$ be an interval and let $o : I \to \mathbb{K}$ denote the *zero function* defined by $o(x) = 0$ for all $x \in I$. We consider $V = \mathsf{F}(I, \mathbb{K})$, the set of functions from I to $\mathbb{K}$ with zero vector $0_V = o$ given by the zero function and define addition $+_V : V \times V \to V$ as in (3.2) and scalar multiplication $\cdot_V : \mathbb{K} \times V \to V$ as in (3.1). It now is a consequence of the properties of addition and multiplication of scalars that $\mathsf{F}(I, \mathbb{K})$ is a $\mathbb{K}$-vector space. (The reader is invited to check this assertion!)

Example 3.6 (Vector space of sequences) A mapping $x : \mathbb{N} \to \mathbb{K}$ from the natural numbers into a field $\mathbb{K}$ called a sequence in $\mathbb{K}$ (or simply a sequence, when $\mathbb{K}$ is clear from the context). It is common to write x_n instead of $x(n)$ for $n \in \mathbb{N}$ and to denote a sequence by $(x_n)_{n\in\mathbb{N}} = (x_1, x_2, x_3, \ldots)$. We write $\mathbb{K}^\infty$ for the set of sequences in $\mathbb{K}$. For instance, taking $\mathbb{K} = \mathbb{R}$, we may consider the sequence

$$\left(\frac{1}{n}\right)_{n\in\mathbb{N}} = \left(1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \ldots\right)$$

or the sequence

$$\left(\sqrt{n}\right)_{n\in\mathbb{N}} = \left(1, \sqrt{2}, \sqrt{3}, 2, \sqrt{5}, \ldots\right).$$

If we equip $\mathbb{K}^\infty$ with the zero vector given by the zero sequence $(0, 0, 0, 0, 0, \ldots)$, addition given by $(x_n)_{n\in\mathbb{N}} + (y_n)_{n\in N} = (x_n + y_n)_{n\in\mathbb{N}}$ and scalar multiplication given by $s \cdot (x_n)_{n\in\mathbb{N}} = (sx_n)_{n\in\mathbb{N}}$ for $s \in \mathbb{K}$, then $\mathbb{K}^\infty$ is a $\mathbb{K}$-vector space.

Example 3.7 (Zero vector space) Consider a set $V = \{x\}$ consisting of a single element. We define $0_V = x$, addition by $x +_V x = x$ and scalar multiplication by $s \cdot_V x = x$. Then all the properties of Definition 3.1 are satisfied. We write $V = \{0_V\}$ or simply $V = \{0\}$ and call V the zero vector space (over $\mathbb{K}$).

The notion of a vector space is an example of an *abstract space*. Later in your studies you will encounter further examples, like *topological spaces, metric spaces and manifolds*.

Remark 3.8 (Notation & Definition) Let V be a $\mathbb{K}$-vector space.

- For $v \in V$ we write $-v = (-1) \cdot_V v$ and for $v_1, v_2 \in V$ we write $v_1 - v_2 = v_1 +_V (-v_2)$. In particular, using the properties from Definition 3.1 we have (check which properties we do use!)

$$v - v = v +_V (-v) = v +_V (-1) \cdot_V v = (1-1) \cdot_V v = 0 \cdot_V v = 0_V$$

For this reason we call $-v$ the *additive inverse of* v.
- Again, it is too cumbersome to always write $+_V$, for this reason we often write $v_1 + v_2$ instead of $v_1 +_V v_2$.
- Likewise, we will often write $s \cdot v$ or sv instead of $s \cdot_V v$.
- It is also customary to write 0 instead of 0_V.

Lemma 3.9 (Elementary properties of vector spaces) *Let V be a $\mathbb{K}$-vector space. Then we have:*

(i) The zero vector is unique, that is, if $0'_V$ is another vector such that $0'_V + v = v + 0'_V = v$ for all $v \in V$, then $0'_V = 0_V$.
(ii) The additive inverse of every $v \in V$ is unique, that is, if $w \in V$ satisfies $v + w = 0_V$, then $w = -v$.
(iii) For all $s \in \mathbb{K}$ we have $s0_V = 0_V$.
(iv) For $s \in \mathbb{K}$ and $v \in V$ we have $sv = 0_V$ if and only if either $s = 0$ or $v = 0_V$.

Proof (The reader is invited to check which property of Definition 3.1 is used in each of the equality signs below)

(i) We have $0'_V = 0'_V + 0_V = 0_V$.
(ii) Since $v + w = 0_V$, adding $-v$, we obtain $(-v) + v + w = 0_V + (-v) = -v = w$.
(iii) We compute $s0_V = s(0_V + 0_V) = s0_V + s0_V$ so that $s0_V - s0_V = 0_V = s0_V$.
(iv) $\Leftarrow$ If $v = 0_V$, then $sv = 0_V$ by (iii). If $s = 0$, then $sv = 0_V$ by (3.5).
$\Rightarrow$ Let $s \in \mathbb{K}$ and $v \in V$ such that $sv = 0_V$. It is sufficient to show that if $s \neq 0$, then $v = 0_V$. Since $s \neq 0$ we can multiply $sv = 0_V$ with $1/s$ so that

$$\frac{1}{s}(sv) = \left(\frac{1}{s}s\right) v = v = \frac{1}{s}0_V = 0_V.$$

□

3.2 Linear maps

Throughout this section, V, W denote $\mathbb{K}$-vector spaces.

Previously we saw that the mapping $f_{\mathbf{A}} : \mathbb{K}^n \to \mathbb{K}^m$ associated to a matrix $M_{m,n}(\mathbb{K})$ is additive and 1-homogeneous. These notions also make sense for mappings between vector spaces.

Definition 3.10 (Linear map) A mapping $f : V \to W$ is called *linear* if it is additive and 1-homogeneous, that is, if it satisfies

$$f(s_1 v_1 + s_2 v_2) = s_1 f(v_1) + s_2 f(v_2) \tag{3.6}$$

for all $s_1, s_2 \in \mathbb{K}$ and for all $v_1, v_2 \in V$.

The reader is invited to check that the condition (3.6) is indeed equivalent to f being additive and 1-homogeneous.

Example 3.11 As we have seen in Remark 2.23, the mapping $f_{\mathbf{A}} : \mathbb{K}^n \to \mathbb{K}^m$ associated to a matrix $\mathbf{A} \in M_{m,n}(\mathbb{K})$ is linear. In Lemma 3.18 below we will see that in fact any linear map $\mathbb{K}^n \to \mathbb{K}^m$ is of this form.

Example 3.12 The derivative $\frac{\mathrm{d}}{\mathrm{d}x} : \mathsf{P}(\mathbb{R}) \to \mathsf{P}(\mathbb{R})$ is linear, see (3.3).

Example 3.13 The matrix transpose is a map $M_{m,n}(\mathbb{K}) \to M_{n,m}(\mathbb{K})$ and this map is linear. Indeed, for all $s, t \in \mathbb{K}$ and $\mathbf{A}, \mathbf{B} \in M_{m,n}(\mathbb{K})$, we have

$$(s\mathbf{A} + t\mathbf{B})^T = (sA_{ji} + tB_{ji})_{1\leqslant j\leqslant n, 1\leqslant i\leqslant m} = s(A_{ji})_{1\leqslant j\leqslant n, 1\leqslant i\leqslant m} + \\ t(B_{ji})_{1\leqslant j\leqslant n, 1\leqslant i\leqslant m} = s\mathbf{A}^T + t\mathbf{B}^T.$$

Example 3.14 If $\mathcal{X}$ is set, the mapping $\mathrm{Id}_{\mathcal{X}} : \mathcal{X} \to \mathcal{X}$ which returns its input is called the *identity mapping*. Let V be a $\mathbb{K}$-vector space and $\mathrm{Id}_V : V \to V$ the identity mapping so that $\mathrm{Id}_V(v) = v$ for all $v \in V$. The identity mapping is linear since for all $s_1, s_2 \in \mathbb{K}$ and $v_1, v_2 \in V$ we have

$$\mathrm{Id}_V(s_1v_1 + s_2v_2) = s_1v_1 + s_2v_2 = s_1\mathrm{Id}_V(v_1) + s_2\mathrm{Id}_V(v_2).$$

A necessary condition for linearity of a mapping is that it maps the zero vector onto the zero vector:

Lemma 3.15 *Let $f : V \to W$ be a linear map, then $f(0_V) = 0_W$.*

Proof Since $f : V \to W$ is linear, we have

$$f(0_V) = f(0 \cdot 0_V) = 0 \cdot f(0_V) = 0_W.$$

□

Proposition 3.16 *Let V_1, V_2, V_3 be $\mathbb{K}$-vector spaces and $f : V_1 \to V_2$ and $g : V_2 \to V_3$ be linear maps. Then the composition $g \circ f : V_1 \to V_3$ is linear. Furthermore, if $f : V_1 \to V_2$ is bijective, then the inverse function $f^{-1} : V_2 \to V_1$ (satisfying $f^{-1} \circ f = f \circ f^{-1} = \mathrm{Id}_{V_1}$) is linear.*

Proof Let $s, t \in \mathbb{K}$ and $v, w \in V_1$. Then

$$\begin{aligned}(g \circ f)(sv + tw) &= g(f(sv + tw)) = g(sf(v) + tf(w)) \\ &= sg(f(v)) + tg(f(w)) = s(g \circ f)(v) + t(g \circ f)(w),\end{aligned}$$

where we first use the linearity of f and then the linearity of g. It follows that $g \circ f$ is linear.

Now suppose $f : V_1 \to V_2$ is bijective with inverse function $f^{-1} : V_2 \to V_1$. Let $s, t \in \mathbb{K}$ and $v, w \in V_2$. Since f is bijective there exist unique vectors $v', w' \in V_1$ with $f(v') = v$ and $f(w') = w$. Hence we can write

$$\begin{aligned} f^{-1}(sv + tw) &= f^{-1}(sf(v') + tf(w')) = f^{-1}(f(sv' + tw')) \\ &= (f^{-1} \circ f)(sv' + tw') = sv' + tw', \end{aligned}$$

where we use the linearity of f. Since we also have $v' = f^{-1}(v)$ and $w' = f^{-1}(w)$, we obtain

$$f^{-1}(sv + tw) = sf^{-1}(v) + tf^{-1}(w),$$

thus showing that $f^{-1} : V_2 \to V_1$ is linear. □

We also have:

Proposition 3.17 *Let $\mathbf{A} \in M_{m,n}(\mathbb{K})$ and $f_\mathbf{A} : \mathbb{K}^n \to \mathbb{K}^m$ the associated linear map. Then $f_\mathbf{A}$ is bijective if and only if there exists a matrix $\mathbf{B} \in M_{n,m}(\mathbb{K})$ satisfying $\mathbf{BA} = \mathbf{1}_n$ and $\mathbf{AB} = \mathbf{1}_m$. In this case, the matrix $\mathbf{B}$ is unique and will be denoted by $\mathbf{A}^{-1}$. We refer to $\mathbf{A}^{-1}$ as the inverse of $\mathbf{A}$ and call $\mathbf{A}$ invertible.*

In order to prove Proposition 3.17 we need the following lemma:

Lemma 3.18 *A mapping $g : \mathbb{K}^m \to \mathbb{K}^n$ is linear if and only if there exists a matrix $\mathbf{B} \in M_{n,m}(\mathbb{K})$ so that $g = f_\mathbf{B}$.*

Proof Let $\mathbf{B} \in M_{n,m}(\mathbb{K})$, then $f_\mathbf{B}$ is linear by Remark 2.23. Conversely, let $g : \mathbb{K}^m \to \mathbb{K}^n$ be linear. Let $\{\vec{e}_1, \dots, \vec{e}_m\}$ denote the standard basis of $\mathbb{K}^m$. Write

$$g(\vec{e}_i) = \begin{pmatrix} B_{1i} \\ \vdots \\ B_{ni} \end{pmatrix} \quad \text{for} \quad i = 1, \dots, m$$

and consider the matrix

$$\mathbf{B} = \begin{pmatrix} B_{11} & \cdots & B_{1m} \\ \vdots & \ddots & \vdots \\ B_{n1} & \cdots & B_{nm} \end{pmatrix} \in M_{n,m}(\mathbb{K}).$$

For $i = 1, \dots, m$ we obtain

$$f_\mathbf{B}(\vec{e}_i) = \mathbf{B}\vec{e}_i = g(\vec{e}_i). \tag{3.7}$$

Any vector $\vec{v} = (v_i)_{1 \leqslant i \leqslant m} \in \mathbb{K}^m$ can be written as

$$\vec{v} = v_1\vec{e}_1 + \cdots + v_m\vec{e}_m$$

for (unique) scalars v_i, $i = 1, \dots, m$. Hence using the linearity of g and $f_\mathbf{B}$, we compute

$$\begin{aligned} g(\vec{v}) - f_\mathbf{B}(\vec{v}) &= g(v_1\vec{e}_1 + \cdots + v_m\vec{e}_m) - f_\mathbf{B}(v_1\vec{e}_1 + \cdots + v_m\vec{e}_m) \\ &= v_1(g(\vec{e}_1) - f_\mathbf{B}(\vec{e}_1)) + \cdots + v_m(g(\vec{e}_m) - f_\mathbf{B}(\vec{e}_m)) = 0_{\mathbb{K}^n}, \end{aligned}$$

where the last equality uses (3.7). Since the vector $\vec{v}$ is arbitrary, it follows that $g = f_\mathbf{B}$, as claimed. □

Proof of Proposition 3.17 First, notice that the mapping $f_{\mathbf{1}_n} : \mathbb{K}^n \to \mathbb{K}^n$ associated to the unit matrix is the identity mapping on $\mathbb{K}^n$, that is, for all $n \in \mathbb{N}$, we have $f_{\mathbf{1}_n} = \mathrm{Id}_{\mathbb{K}^n}$.

Let $\mathbf{A} \in M_{m,n}(\mathbb{K})$ and suppose that $f_\mathbf{A} : \mathbb{K}^n \to \mathbb{K}^m$ is bijective with inverse function $(f_\mathbf{A})^{-1} : \mathbb{K}^m \to \mathbb{K}^n$. By Proposition 3.16, the mapping $(f_\mathbf{A})^{-1}$ is linear and hence of the form $(f_{\mathbf{A}})^{-1} = f_{\mathbf{B}}$ for some matrix $\mathbf{B} \in M_{n,m}(\mathbb{K})$ by the previous Lemma 3.18. Using Theorem 2.21, we obtain

$$(f_{\mathbf{A}})^{-1} \circ f_{\mathbf{A}} = \mathrm{Id}_{\mathbb{K}^n} = f_{\mathbf{B}} \circ f_{\mathbf{A}} = f_{\mathbf{BA}} = f_{\mathbf{1}_n}$$

hence Proposition 2.20 implies that $\mathbf{BA} = \mathbf{1}_n$. Likewise we have

$$f_{\mathbf{A}} \circ (f_{\mathbf{A}})^{-1} = \mathrm{Id}_{\mathbb{K}^m} = f_{\mathbf{A}} \circ f_{\mathbf{B}} = f_{\mathbf{AB}} = f_{\mathbf{1}_m}$$

so that $\mathbf{AB} = \mathbf{1}_m$.

Conversely, let $\mathbf{A} \in M_{m,n}(\mathbb{K})$ and suppose the matrix $\mathbf{B} \in M_{n,m}(\mathbb{K})$ satisfies $\mathbf{AB} = \mathbf{1}_m$ and $\mathbf{BA} = \mathbf{1}_n$. Then, as before, we have

$$f_{\mathbf{AB}} = f_{\mathbf{1}_m} = \mathrm{Id}_{\mathbb{K}^m} = f_{\mathbf{A}} \circ f_{\mathbf{B}} \quad \text{and} \quad f_{\mathbf{BA}} = f_{\mathbf{1}_n} = \mathrm{Id}_{\mathbb{K}^n} = f_{\mathbf{B}} \circ f_{\mathbf{A}}$$

showing that $f_{\mathbf{A}} : \mathbb{K}^n \to \mathbb{K}^m$ is bijective with inverse function $f_{\mathbf{B}} : \mathbb{K}^m \to \mathbb{K}^n$.

Finally, to verify the uniqueness of $\mathbf{B}$, we assume that there exists $\mathbf{B}' \in M_{n,m}(\mathbb{K})$ with $\mathbf{AB}' = \mathbf{1}_m$ and $\mathbf{B}'\mathbf{A} = \mathbf{1}_n$. Then

$$\mathbf{B}' = \mathbf{B}'\mathbf{1}_m = \mathbf{B}'\mathbf{AB} = (\mathbf{B}'\mathbf{A})\mathbf{B} = \mathbf{1}_n\mathbf{B} = \mathbf{B},$$

showing that $\mathbf{B}' = \mathbf{B}$, hence $\mathbf{B}$ is unique. □

Exercises

Exercise 3.19 Let $f : V \to W$ be a linear map, $k \geqslant 2$ a natural number and $s_1, \ldots, s_k \in \mathbb{K}$ and $v_1, \ldots, v_k \in V$. Then $f : V \to W$ satisfies

$$f(s_1 v_1 + \cdots + s_k v_k) = s_1 f(v_1) + \cdots + s_k f(v_k)$$

or written with the sum symbol

$$\boxed{f\left(\sum_{i=1}^{k} s_i v_i\right) = \sum_{i=1}^{k} s_i f(v_i).}$$

This identity is used frequently in Linear Algebra, so make sure you understand it.

Exercise 3.20 Let $a, b, c, d \in \mathbb{K}$ and

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_{2,2}(\mathbb{K}).$$

Show that $\mathbf{A}$ has an inverse $\mathbf{A}^{-1}$ if and only if $ad - bc \neq 0$. For $ad - bc \neq 0$, compute the inverse $\mathbf{A}^{-1}$.

3.3 Vector subspaces and isomorphisms

WEEK 4

3.3.1 Vector subspaces

A vector subspace of a vector space is a subset that is itself a vector space, more precisely:

Definition 3.21 (Vector subspace) Let V be a $\mathbb{K}$-vector space. A subset $U \subset V$ is called a *vector subspace* of V if U is non-empty and if

$$s_1 \cdot_V v_1 +_V s_2 \cdot_V v_2 \in U \quad \text{for all } s_1, s_2 \in \mathbb{K} \text{ and all } v_1, v_2 \in U. \tag{3.8}$$

Video Subspaces

Remark 3.22

(i) Observe that since U is non-empty, it contains an element, say u. Since $0 \cdot_V u = 0_V \in U$ it follows that the zero vector 0_V lies in U. A vector subspace U is itself a vector space when we take $0_U = 0_V$ and borrow vector addition and scalar multiplication from V. Indeed, all of the properties in Definition 3.1 of $+_V$ and $\cdot_V$ hold for all elements of V and all scalars, hence also for all elements of $U \subset V$ and all scalars. We only need to verify that we cannot fall out of U by vector addition and scalar multiplication, but this is precisely what the condition (3.8) states.

(ii) A vector subspace is also called a *linear subspace* or simply a subspace.

The prototypical example of a vector subspace are lines and planes through the origin in $\mathbb{R}^3$:

Example 3.23 (Lines through the origin) Let $\vec{w} \neq 0_{\mathbb{R}^3}$, then the line

$$U = \{s\vec{w} \,|\, s \in \mathbb{R}\} \subset \mathbb{R}^3$$

is a vector subspace. Indeed, taking $s = 0$ it follows that $0_{\mathbb{R}^3} \in U$ so that U is non-empty. Let $\vec{u}_1, \vec{u}_2$ be vectors in U so that $\vec{u}_1 = t_1\vec{w}$ and $\vec{u}_2 = t_2\vec{w}$ for scalars $t_1, t_2 \in \mathbb{R}$. Let $s_1, s_2 \in \mathbb{R}$, then

$$s_1\vec{u}_1 + s_2\vec{u}_2 = s_1t_1\vec{w} + s_2t_2\vec{w} = (s_1t_1 + s_2t_2)\,\vec{w} \in U$$

so that $U \subset \mathbb{R}^3$ is a subspace.

Example 3.24 (Zero vector space) Let V be a $\mathbb{K}$-vector space and $U = \{0_V\}$ the zero vector space arising from 0_V. Then, by Definition 3.21 and the properties of Definition 3.1, it follows that U is a vector subspace of V.

Example 3.25 (Periodic functions) Taking $I = \mathbb{R}$ and $\mathbb{K} = \mathbb{R}$ in Example 3.5, we see that the functions $f : \mathbb{R} \to \mathbb{R}$ form an $\mathbb{R}$-vector space $V = \mathrm{F}(\mathbb{R}, \mathbb{R})$. Consider the subset

$$U = \{f \in \mathrm{F}(\mathbb{R}, \mathbb{R}) \,|\, f \text{ is periodic with period } 2\pi\}$$

consisting of 2π-periodic functions, that is, an element $f \in U$ satisfies $f(x+2\pi) = f(x)$ for all $x \in \mathbb{R}$. Notice that U is not empty, as $\cos : \mathbb{R} \to \mathbb{R}$ and $\sin : \mathbb{R} \to \mathbb{R}$ are elements of U. Suppose $f_1, f_2 \in U$ and $s_1, s_2 \in \mathbb{R}$. Then, we have for all $x \in \mathbb{R}$

$$\begin{aligned}(s_1 f_1 + s_2 f_2)(x+2\pi) &= s_1 f_1(x+2\pi) + s_2 f_2(x+2\pi) = s_1 f_1(x) + s_2 f_2(x) \\ &= (s_1 f_1 + s_2 f_2)(x)\end{aligned}$$

showing that $s_1 f_1 + s_2 f_2$ is periodic with period 2π. By Definition 3.21, it follows that U is a vector subspace of $\mathsf{F}(\mathbb{R}, \mathbb{R})$.

Recall, if $\mathcal{X}, \mathcal{W}$ are sets, $\mathcal{Y} \subset \mathcal{X}$, $\mathcal{Z} \subset \mathcal{W}$ subsets and $f : \mathcal{X} \to \mathcal{W}$ a mapping, then the *image* of $\mathcal{Y}$ under f is the set

$$f(\mathcal{Y}) = \{w \in \mathcal{W} \,|\, \text{there exists an element } y \in \mathcal{Y} \text{ with } f(y) = w\}$$

consisting of all the elements in $\mathcal{W}$ which are hit by an element of $\mathcal{Y}$ under the mapping f. In the special case where $\mathcal{Y}$ is all of $\mathcal{X}$, that is, $\mathcal{Y} = \mathcal{X}$, it is also customary to write $\operatorname{Im}(f)$ instead of $f(\mathcal{X})$ and simply speak of the image of f. Similarly, the *preimage* of $\mathcal{Z}$ under f is the set

$$f^{-1}(\mathcal{Z}) = \{x \in \mathcal{X} \,|\, f(x) \in \mathcal{Z}\}$$

consisting of all the elements in $\mathcal{X}$ which are mapped onto elements of $\mathcal{Z}$ under f. Notice that f is not assumed to be bijective, hence the inverse mapping $f^{-1} : \mathcal{W} \to \mathcal{X}$ does not need to exist (and in fact the definition of the preimage does not involve the inverse mapping). Nonetheless the notation $f^{-1}(\mathcal{Z})$ is customary.

It is natural to ask how the image and preimage of subspaces look like under a linear map:

Proposition 3.26 *Let V, W be $\mathbb{K}$-vector spaces, $U \subset V$ and $Z \subset W$ be vector subspaces and $f : V \to W$ a linear map. Then the image $f(U)$ is a vector subspace of W and the preimage $f^{-1}(Z)$ is a vector subspace of V.*

Proof Since U is a vector subspace, we have $0_V \in U$. By Lemma 3.15, $f(0_V) = 0_W$, hence $0_W \in f(U)$. For all $w_1, w_2 \in f(U)$ there exist $u_1, u_2 \in U$ with $f(u_1) = w_1$ and $f(u_2) = w_2$. Hence for all $s_1, s_2 \in \mathbb{K}$ we obtain

$$s_1 w_1 + s_2 w_2 = s_1 f(u_1) + s_2 f(u_2) = f(s_1 u_1 + s_2 u_2),$$

where we use the linearity of f. Since U is a subspace, $s_1 u_1 + s_2 u_2$ is an element of U as well. It follows that $s_1 w_1 + s_2 w_2 \in f(U)$ and hence applying Definition 3.21 again, we conclude that $f(U)$ is a subspace of W. The second claim is left to the reader as an exercise. □

Vector subspaces are stable under intersection in the following sense:

Proposition 3.27 *Let V be a $\mathbb{K}$-vector space, $n \geqslant 2$ a natural number and $U_1, \ldots, U_n$ vector subspaces of V. Then the intersection*

$$U' = \bigcap_{j=1}^{n} U_j = \{v \in V \,|\, v \in U_j \text{ for all } j = 1, \ldots, n\}$$

is a vector subspace of V as well.

Proof Since U_j is a vector subspace, $0_V \in U_j$ for all $j = 1, \dots, n$. Therefore, $0_V \in U'$, hence U' is not empty. Let $u_1, u_2 \in U'$ and $s_1, s_2 \in \mathbb{K}$. By assumption, $u_1, u_2 \in U_j$ for all $j = 1, \dots, n$. Since U_j is a vector subspace for all $j = 1, \dots, n$ it follows that $s_1u_1+s_2u_2 \in U_j$ for all $j = 1, \dots, n$ and hence $s_1u_1 + s_2u_2 \in U'$. By Definition 3.21, it follows that U' is a vector subspace of V. □

Remark 3.28 Notice that the union of subspaces need not be a subspace. Let $V = \mathbb{R}^2$, $\{\vec{e}_1, \vec{e}_2\}$ its standard basis and

$$U_1 = \{s\vec{e}_1 \,|\, s \in \mathbb{R}\} \quad \text{and} \quad U_2 = \{s\vec{e}_2 \,|\, s \in \mathbb{R}\}.$$

Then $\vec{e}_1 \in U_1 \cup U_2$ and $\vec{e}_2 \in U_1 \cup U_2$, but $\vec{e}_1 + \vec{e}_2 \notin U_1 \cup U_2$.

The kernel of a linear map $f : V \to W$ consists of those vectors in V that are mapped onto the zero vector of W:

Definition 3.29 (Kernel) The *kernel* of a linear map $f : V \to W$ is the preimage of $\{0_W\}$ under f, that is,

$$\mathrm{Ker}(f) = \{v \in V \,|\, f(v) = 0_W\} = f^{-1}(\{0_W\}).$$

Example 3.30 The kernel of the linear map $\frac{\mathrm{d}}{\mathrm{d}x} : \mathsf{P}_n(\mathbb{R}) \to \mathsf{P}_{n-1}(\mathbb{R})$ consists of the constant polynomials satisfying $f(x) = c$ for all $x \in \mathbb{R}$ and where $c \in \mathbb{R}$ is some constant.

We can characterise the injectivity of a linear map $f : V \to W$ in terms of its kernel:

Lemma 3.31 *A linear map $f : V \to W$ is injective if and only if* $\mathrm{Ker}(f) = \{0_V\}$.

Proof Let $f : V \to W$ be injective. Suppose $f(v) = 0_W$. Since $f(0_V) = 0_W$ by Lemma 3.15, we have $f(v) = f(0_V)$, hence $v = 0_V$ by the injectivity assumption. It follows that $\mathrm{Ker}(f) = \{0_V\}$. Conversely, suppose $\mathrm{Ker}(f) = \{0_V\}$ and let $v_1, v_2 \in V$ be such that $f(v_1) = f(v_2)$. Then by the linearity we have $f(v_1) - f(v_2) = 0_W = f(v_1 - v_2)$. Hence $v_1 - v_2$ is in the kernel of f so that $v_1 - v_2 = 0_V$ or $v_1 = v_2$. □

An immediate consequence of Proposition 3.26 is:

Corollary 3.32 *Let $f : V \to W$ be a linear map, then its image* $\mathrm{Im}(f)$ *is a vector subspace of W and its kernel* $\mathrm{Ker}(f)$ *is a vector subspace of V.*

3.3.2 Isomorphisms

Definition 3.33 (Vector space isomorphism) A bijective linear map $f : V \to W$ is called a (vector space) *isomorphism*. If an isomorphism $f : V \to W$ exists, then the $\mathbb{K}$-vector spaces V and W are called *isomorphic*.

By the definition of surjectivity, a map $f : V \to W$ is surjective if and only if $\mathrm{Im}(f) = W$. Combining this with Lemma 3.31 gives:

Proposition 3.34 *A linear map $f : V \to W$ is an isomorphism if and only if* $\mathrm{Ker}(f) = \{0_V\}$ *and* $\mathrm{Im}(f) = W$.

3.4 Generating sets

Definition 3.35 (Linear combination) Let V be a $\mathbb{K}$-vector space, $k \in \mathbb{N}$ and $\{v_1, \dots, v_k\}$ a set of vectors from V. A *linear combination* of the vectors $\{v_1, \dots, v_k\}$ is a vector of the form

$$w = s_1 v_1 + \cdots + s_k v_k = \sum_{i=1}^{k} s_i v_i$$

for some $s_1, \dots, s_k \in \mathbb{K}$.

Example 3.36 For $n \in \mathbb{N}$ with $n \geqslant 2$ consider $V = \mathsf{P}_n(\mathbb{R})$ and the polynomials $p_1, p_2, p_3 \in \mathsf{P}_n(\mathbb{R})$ defined by the rules $p_1(x) = 1$, $p_2(x) = x$, $p_3(x) = x^2$ for all $x \in \mathbb{R}$. A linear combination of $\{p_1, p_2, p_3\}$ is a polynomial of the form $p(x) = ax^2 + bx + c$ where $a, b, c \in \mathbb{R}$.

Definition 3.37 (Subspace generated by a set) Let V be a $\mathbb{K}$-vector space and $\mathcal{S} \subset V$ be a non-empty subset. The *subspace generated by* $\mathcal{S}$ is the set $\mathrm{span}(\mathcal{S})$ whose elements are linear combinations of finitely many vectors in $\mathcal{S}$. The set $\mathrm{span}(\mathcal{S})$ is called the *span of* $\mathcal{S}$. Formally, we have

$$\mathrm{span}(\mathcal{S}) = \left\{ v \in V \,\middle|\, v = \sum_{i=1}^{k} s_i v_i,\, k \in \mathbb{N}, s_1, \dots, s_k \in \mathbb{K}, v_1, \dots, v_k \in \mathcal{S} \right\}.$$

Remark 3.38 The notation $\langle \mathcal{S} \rangle$ for the span of $\mathcal{S}$ is also in use.

Proposition 3.39 *Let V be a $\mathbb{K}$-vector space and $\mathcal{S} \subset V$ be a non-empty subset. Then* $\mathrm{span}(\mathcal{S})$ *is a vector subspace of V.*

Proof Since $\mathcal{S}$ is non-empty it contains some element, say u. Since u itself is a linear combination of $\{u\}$, it follows that $\mathrm{span}(\mathcal{S})$ is non-empty. Let $k \in \mathbb{N}$ and $v_1 = t_1 w_1 + \cdots + t_k w_k$ for $t_1, \dots t_k \in \mathbb{K}$ and $w_1, \dots, w_k \in \mathcal{S}$ be a linear combination of vectors in $\mathcal{S}$. Furthermore, let $j \in \mathbb{N}$ and $v_2 = \hat{t}_1 \hat{w}_1 + \cdots + \hat{t}_j \hat{w}_j$ for $\hat{t}_1, \dots, \hat{t}_j$ and $\hat{w}_1, \dots, \hat{w}_j \in \mathcal{S}$ be another linear combination of vectors in $\mathcal{S}$. By Definition 3.21, it suffices to show that for all $s_1, s_2 \in \mathbb{K}$ the vector $s_1 v_1 + s_2 v_2$ is a linear combination of vectors in $\mathcal{S}$. Since

$$\begin{aligned} s_1 v_1 + s_2 v_2 &= s_1(t_1 w_1 + \cdots + t_k w_k) + s_2(\hat{t}_1 \hat{w}_1 + \cdots + \hat{t}_j \hat{w}_j) \\ &= s_1 t_1 w_1 + \cdots + s_1 t_k w_k + s_2 \hat{t}_1 \hat{w}_1 + \cdots + s_2 \hat{t}_j \hat{w}_j \end{aligned}$$

is a linear combination of the vectors $\{w_1, \ldots, w_k, \hat{w}_1, \ldots, \hat{w}_j\}$ in $\mathcal{S}$, the claim follows. □

Remark 3.40 For a subset $\mathcal{S} \subset V$, we may alternatively define $\operatorname{span}(\mathcal{S})$ to be the smallest vector subspace of V that contains $\mathcal{S}$. This has the advantage of $\mathcal{S}$ being allowed to be empty, in which case $\operatorname{span}(\emptyset) = \{0_V\}$, that is, the empty set is a generating set for the zero vector space.

Definition 3.41 Let V be a $\mathbb{K}$-vector space. A subset $\mathcal{S} \subset V$ is called a *generating set* if $\operatorname{span}(\mathcal{S}) = V$. The vector space V is called *finite dimensional* if V admits a generating set with finitely many elements (also called a finite set). A vector space that is not finite dimensional will be call *infinite dimensional*.

Example 3.42 Thinking of a field $\mathbb{K}$ as a $\mathbb{K}$-vector space, the set $\mathcal{S} = \{1_{\mathbb{K}}\}$ consisting of the identity element of multiplication is a generating set for $V = \mathbb{K}$. Indeed, for every $x \in \mathbb{K}$ we have $x = x \cdot_V 1_{\mathbb{K}}$.

Example 3.43 The standard basis $\mathcal{S} = \{\vec{e}_1, \ldots, \vec{e}_n\}$ is a generating set for $\mathbb{K}^n$, since for all $\vec{x} = (x_i)_{1 \leqslant i \leqslant n} \in \mathbb{K}^n$, we can write $\vec{x} = x_1\vec{e}_1 + \cdots + x_n\vec{e}_n$ so that $\vec{x}$ is a linear combination of elements of $\mathcal{S}$.

Example 3.44 Let $\mathbf{E}_{k,l} \in M_{m,n}(\mathbb{K})$ for $1 \leqslant k \leqslant m$ and $1 \leqslant l \leqslant n$ denote the m-by-n matrix satisfying $\mathbf{E}_{k,l} = (\delta_{ki}\delta_{lj})_{1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n}$. For example, for $m = 2$ and $n = 3$ we have

$$\mathbf{E}_{1,1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \mathbf{E}_{1,2} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \mathbf{E}_{1,3} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

and

$$\mathbf{E}_{2,1} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \qquad \mathbf{E}_{2,2} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \qquad \mathbf{E}_{2,3} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Then $\mathcal{S} = \{\mathbf{E}_{k,l}\}_{1 \leqslant k \leqslant m, 1 \leqslant l \leqslant n}$ is a generating set for $M_{m,n}(\mathbb{K})$, since a matrix $\mathbf{A} \in M_{m,n}(\mathbb{K})$ can be written as

$$\mathbf{A} = \sum_{k=1}^{m} \sum_{l=1}^{n} A_{kl} \mathbf{E}_{k,l}$$

so that $\mathbf{A}$ is a linear combination of the elements of $\mathcal{S}$.

Example 3.45 The vector space $\mathsf{P}(\mathbb{R})$ of polynomials is infinite dimensional. In order to see this, consider a finite set of polynomials $\{p_1, \ldots, p_n\}$, $n \in \mathbb{N}$ and let d_i denote the degree of the polynomial p_i for $i = 1, \ldots, n$. We set $D = \max\{d_1, \ldots, d_n\}$. Since a linear combination of the polynomials $\{p_1, \ldots, p_n\}$ has degree at most D, any polynomial q whose degree is strictly larger than D will satisfy $q \notin \operatorname{span}\{p_1, \ldots, p_n\}$. It follows that $\mathsf{P}(\mathbb{R})$ cannot be generated by a finite set of polynomials.

Lemma 3.46 *Let $f : V \to W$ be linear and $\mathcal{S} \subset V$ a generating set. If f is surjective, then $f(\mathcal{S})$ is a generating set for W. Furthermore, if f is bijective, then V is finite dimensional if and only if W is finite dimensional.*

Proof Let $w \in W$. Since f is surjective there exists $v \in V$ such that $f(v) = w$. Since $\operatorname{span}(\mathcal{S}) = V$, there exists $k \in \mathbb{N}$, as well as elements $v_1, \dots, v_k \in \mathcal{S}$ and scalars $s_1, \dots, s_k$ such that $v = \sum_{i=1}^{k} s_i v_i$ and hence $w = \sum_{i=1}^{k} s_i f(v_i)$, where we use the linearity of f. We conclude that $w \in \operatorname{span}(f(\mathcal{S}))$ and since w is arbitrary, it follows that $W = \operatorname{span}(f(\mathcal{S}))$.

For the second claim suppose V is finite dimensional, hence we have a finite set $\mathcal{S}$ with $\operatorname{span}(\mathcal{S}) = V$. The set $f(\mathcal{S})$ is finite as well and satisfies $\operatorname{span}(f(\mathcal{S})) = W$ by the previous argument, hence W is finite dimensional as well. Conversely suppose W is finite dimensional with generating set $\mathcal{T} \subset W$. Since f is bijective there exists an inverse mapping $f^{-1} : W \to V$ which is surjective, hence $V = \operatorname{span}(f^{-1}(\mathcal{T}))$ so that V is finite dimensional as well. □

3.5 Linear independence and bases

A set of vectors where no vector can be expressed as a linear combination of the other vectors is called linearly independent. More precisely:

Definition 3.47 (Linear independence) Let $\mathcal{S} \subset V$ be a non-empty finite subset so that $\mathcal{S} = \{v_1, \dots, v_k\}$ for distinct vectors $v_i \in V$, $i = 1, \dots, k$. We say $\mathcal{S}$ is *linearly independent* if

$$s_1 v_1 + \cdots + s_k v_k = 0_V \quad \iff \quad s_1 = \cdots = s_k = 0,$$

where $s_1, \dots, s_k \in \mathbb{K}$. If $\mathcal{S}$ is not linearly independent, then $\mathcal{S}$ is called *linearly dependent*. Furthermore, we call a subset $\mathcal{S} \subset V$ linearly independent if every finite subset of $\mathcal{S}$ is linearly independent. We will call distinct vectors $v_1, \dots, v_k$ linearly independent/dependent if the set $\{v_1, \dots, v_k\}$ is linearly independent/dependent.

Remark 3.48 Instead of *distinct*, many authors write *pairwise distinct*, which means that all pairs of vectors v_i, v_j with $i \neq j$ satisfy $v_i \neq v_j$. Of course, this simply means that the list $v_1, \dots, v_k$ of vectors is not allowed to contain a vector more than once.

Notice that if the vectors $v_1, \dots, v_k \in V$ are linearly dependent, then there exist scalars $s_1, \dots, s_k$, not all zero, so that $\sum_{i=1}^{k} s_i v_i = 0_V$. After possibly changing the numbering of the vectors and scalars, we can assume that $s_1 \neq 0$. Therefore, we can write

$$v_1 = -\sum_{i=2}^{k} \left(\frac{s_i}{s_1}\right) v_i,$$

so that v_1 is a linear combination of the vectors $v_2, \dots, v_k$.

Also, observe that a subset $\mathcal{T}$ of a linearly independent set $\mathcal{S}$ is itself linearly independent. (Why?)

Example 3.49 We consider the polynomials $p_1, p_2, p_3 \in \mathsf{P}(\mathbb{R})$ defined by the rules $p_1(x) = 1$, $p_2(x) = x$, $p_3(x) = x^2$ for all $x \in \mathbb{R}$. Then $\{p_1, p_2, p_3\}$ is linearly independent. In order to see this, consider the condition

$$s_1 p_1 + s_2 p_2 + s_3 p_3 = 0_{\mathsf{P}(\mathbb{R})} = o \tag{3.9}$$

where $o : \mathbb{R} \to \mathbb{R}$ denotes the zero polynomial. Since (3.9) means that

$$s_1 p_1(x) + s_2 p_2(x) + s_3 p_3(x) = o(x),$$

for all $x \in \mathbb{R}$, we can evaluate this condition for any choice of real number x. Taking $x = 0$ gives

$$s_1 p_1(0) + s_2 p_2(0) + s_3 p_3(0) = o(0) = 0 = s_1.$$

Taking $x = 1$ and $x = -1$ gives

$$0 = s_2 p_2(1) + s_3 p_3(1) = s_2 + s_3,$$
$$0 = s_2 p_2(-1) + s_3 p_3(-1) = -s_2 + s_3,$$

so that $s_2 = s_3 = 0$ as well. It follows that $\{p_1, p_2, p_3\}$ is linearly independent.

Remark 3.50 By convention, the empty set is linearly independent.

Definition 3.51 (Basis) A subset $\mathcal{S} \subset V$ which is a generating set of V and also linearly independent is called a *basis* of V.

Video Basis

Example 3.52 Thinking of a field $\mathbb{K}$ as a $\mathbb{K}$-vector space, the set $\{1_{\mathbb{K}}\}$ is linearly independent, since $1_{\mathbb{K}} \neq 0_{\mathbb{K}}$. Example 3.42 implies that $\{1_{\mathbb{K}}\}$ is a basis of $\mathbb{K}$.

Example 3.53 Clearly, the standard basis $\{\vec{e}_1, \ldots, \vec{e}_n\}$ of $\mathbb{K}^n$ is linearly independent since

$$s_1\vec{e}_1 + \cdots + s_n\vec{e}_n = \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix} = 0_{\mathbb{K}^n} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \qquad \iff \qquad s_1 = \cdots = s_n = 0.$$

It follows together with Example 3.43 that the standard basis of $\mathbb{K}^n$ is indeed a basis in the sense of Definition 3.51.

Example 3.54 The matrices $\mathbf{E}_{k,l} \in M_{m,n}(\mathbb{K})$ for $1 \leqslant k \leqslant m$ and $1 \leqslant l \leqslant n$ are linearly independent. Suppose we have scalars $s_{kl} \in \mathbb{K}$ such that

$$\sum_{k=1}^{m}\sum_{l=1}^{n} s_{kl}\mathbf{E}_{k,l} = \mathbf{0}_{m,n} = \begin{pmatrix} s_{11} & \cdots & s_{1n} \\ \vdots & \ddots & \vdots \\ s_{m1} & \cdots & s_{mn} \end{pmatrix} = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix}$$

so $s_{kl} = 0$ for all $1 \leqslant k \leqslant m$ and all $1 \leqslant l \leqslant n$. It follows together with Example 3.44 that $\{\mathbf{E}_{k,l}\}_{1\leqslant k\leqslant m, 1\leqslant l\leqslant n}$ is a basis of $M_{m,n}(\mathbb{K})$. We refer to $\{\mathbf{E}_{k,l}\}_{1\leqslant k\leqslant m, 1\leqslant l\leqslant n}$ as the *standard basis of $M_{m,n}(\mathbb{K})$*.

Example 3.55 Combining Remark 3.40 and Remark 3.50 we conclude that the empty set is a basis for the zero vector space $\{0\}$.

Lemma 3.56 *Let $f : V \to W$ be an injective linear map. Suppose $\mathcal{S} \subset V$ is linearly independent, then $f(\mathcal{S}) \subset W$ is also linearly independent.*

Proof Let $\{w_1, \ldots, w_k\} \subset f(\mathcal{S})$ be a finite subset for some $k \in \mathbb{N}$ some and distinct vectors $w_i \in W$, where $1 \leqslant i \leqslant k$. Then there exist vectors $v_1, \ldots, v_k$ with $f(v_i) = w_i$ for $1 \leqslant i \leqslant k$. Suppose there exist scalars $s_1, \ldots, s_k$ such that $s_1 w_1 + \cdots + s_k w_k = 0_W$. Using the linearity of f, this implies

$$0_W = s_1 w_1 + \cdots + s_k w_k = s_1 f(v_1) + \cdots + s_k f(v_k) = f(s_1 v_1 + \cdots + s_k v_k).$$

Since f is injective we have $\operatorname{Ker}(f) = \{0_V\}$ by Lemma 3.31. Since $s_1 v_1 + \cdots + s_k v_k \in \operatorname{Ker} f$ it follows that $s_1 v_1 + \cdots + s_k v_k = 0_V$, hence $s_1 = \cdots = s_k = 0$ by the linear independence of $\mathcal{S}$. It follows that $f(\mathcal{S})$ is linearly independent as well. $\square$

Exercises

Exercise 3.57 Let $U \subset V$ be a vector subspace and $k \in \mathbb{N}$ with $k \geqslant 2$. Show that for $u_1, \ldots, u_k \in U$ and $s_1, \ldots, s_k \in \mathbb{K}$, we have $s_1 u_1 + \cdots + s_k u_k \in U$.

Exercise 3.58 (Planes through the origin) Let $\vec{w}_1, \vec{w}_2 \neq 0_{\mathbb{R}^3}$ and $\vec{w}_1 \neq s\vec{w}_2$ for all $s \in \mathbb{R}$. Show that the plane

$$U = \{s_1 \vec{w}_1 + s_2 \vec{w}_2 \,|\, s_1, s_2 \in \mathbb{R}\}$$

is a vector subspace of $\mathbb{R}^3$.

Exercise 3.59 (Polynomials) Let $n \in \mathbb{N} \cup \{0\}$ and $\mathsf{P}_n(\mathbb{R})$ denote the subset of $\mathsf{P}(\mathbb{R})$ consisting of polynomials of degree *at most* n. Show that $\mathsf{P}_n(\mathbb{R})$ is a subspace of $\mathsf{P}(\mathbb{R})$ for all $n \in \mathbb{N} \cup \{0\}$.

Exercise 3.60 Show that the $\mathbb{K}$-vector space $\mathbb{K}^n$ of column vectors with n entries is isomorphic to the $\mathbb{K}$-vector space $\mathbb{K}_n$ of row vectors with n entries.

Exercise 3.61 Show that the $\mathbb{R}$-vector spaces $\mathsf{P}_n(\mathbb{R})$ and $\mathbb{R}^{n+1}$ are isomorphic for all $n \in \mathbb{N} \cup \{0\}$.

Exercise 3.62 Show that for a non-empty subset $\mathcal{S}$ of a $\mathbb{K}$-vector space V, the set $\operatorname{span}(\mathcal{S})$ as defined in Definition 3.37 is the same as the set $\operatorname{span}(\mathcal{S})$ as defined in Remark 3.40. In particular, Proposition 3.39 remains true when removing the assumption that $\mathcal{S}$ is non-empty.

Exercise 3.63 Show that a subset $\{v\}$ consisting of a single vector $v \in V$ is linearly independent if and only if $v \neq 0_V$.

3.6 The dimension

WEEK 5

3.6.1 Defining the dimension

Intuitively, we might define the dimension of a finite dimensional vector space V to be the number of elements of any basis of V, so that a line is 1-dimensional, a plane is 2-dimensional and so on. Of course, this definition only makes sense if we know that there always exists a basis of V and that the number of elements in the basis is independent of the chosen basis. Perhaps surprisingly, these facts take quite a bit of work to prove.

Theorem 3.64 *Let V be a $\mathbb{K}$-vector space.*

(i) *Any subset $\mathcal{S} \subset V$ generating V admits a subset $\mathcal{T} \subset \mathcal{S}$ that is a basis of V.*
(ii) *Any subset $\mathcal{S} \subset V$ that is linearly independent in V is contained in a subset $\mathcal{T} \subset V$ that is a basis of V.*
(iii) *If $\mathcal{S}_1, \mathcal{S}_2$ are bases of V, then there exists a bijective map $f : \mathcal{S}_1 \to \mathcal{S}_2$.*
(iv) *If V is finite dimensional, then any basis of V is a finite set and the number of elements in the basis is independent of the choice of the basis.*

Corollary 3.65 *Every $\mathbb{K}$-vector space V admits at least one basis.*

Proof Since V is a generating set for V, we can apply (i) from Theorem 3.64 to $\mathcal{S} = V$ to obtain a basis of V. □

Remark 3.66 Let $\mathcal{X}$ be a set with finitely many elements. We write $\mathrm{Card}(\mathcal{X})$ – for *cardinality* – for the number of elements of $\mathcal{X}$.

Definition 3.67 The dimension of a finite dimensional $\mathbb{K}$-vector space V, denoted by $\dim(V)$ or $\dim_{\mathbb{K}}(V)$, is the number of elements of any basis of V.

Example 3.68

(i) The zero vector space $\{0\}$ has the empty set as a basis and hence is 0-dimensional.
(ii) A field $\mathbb{K}$ – thought of as a $\mathbb{K}$-vector space – has $\{1_{\mathbb{K}}\}$ as a basis and hence is 1-dimensional.
(iii) The vector space $\mathbb{K}^n$ has $\{\vec{e}_1, \ldots, \vec{e}_n\}$ as a basis and hence is n-dimensional.
(iv) The vector space $M_{m,n}(\mathbb{K})$ has $\mathbf{E}_{k,l}$ for $1 \leqslant k \leqslant m$ and $1 \leqslant l \leqslant n$ as a basis, hence it is mn-dimensional.

We will only prove Theorem 3.64 for finite dimensional vector spaces. This will be done with the help of three lemmas.

Lemma 3.69 *Let V be a $\mathbb{K}$-vector space, $\mathcal{S} \subset V$ linearly independent and $v_0 \in V$. Suppose $v_0 \notin \mathrm{span}(\mathcal{S})$, then $\mathcal{S} \cup \{v_0\}$ is linearly independent.*

Proof Let $\mathcal{T}$ be a finite subset of $\mathcal{S} \cup \{v_0\}$. If $v_0 \notin \mathcal{T}$, then $\mathcal{T}$ is linearly independent, as $\mathcal{S}$ is linearly independent. So suppose $v_0 \in \mathcal{T}$. There exist distinct elements $v_1, \ldots, v_n$ of $\mathcal{S}$ so that $\mathcal{T} = \{v_0, v_1, \ldots, v_n\}$. Suppose $s_0 v_0 + s_1 v_1 + \cdots + s_n v_n = 0_V$ for some scalars $s_0, s_1, \ldots, s_n \in \mathbb{K}$. If $s_0 \neq 0$, then we can write

$$v_0 = -\sum_{i=1}^{n} \frac{s_i}{s_0} v_i,$$

contradicting the assumption that $v_0 \notin \operatorname{span}(\mathcal{S})$. Hence we must have $s_0 = 0$. Since $s_0 = 0$ it follows that $s_1 v_1 + \cdots + s_n v_n = 0_V$ so that $s_1 = \cdots = s_n = 0$ by the linear independence of $\mathcal{S}$. We conclude that $\mathcal{S} \cup \{v_0\}$ is linearly independent. □

Lemma 3.70 *Let V be a $\mathbb{K}$-vector space and $\mathcal{S} \subset V$ a generating set. If $v_0 \in \operatorname{span}(\mathcal{S} \setminus \{v_0\})$, then $\mathcal{S} \setminus \{v_0\}$ is a generating set.*

Proof Since $v_0 \in \operatorname{span}(\mathcal{S} \setminus \{v_0\})$, there exist vectors $v_1, \ldots, v_n \in \mathcal{S}$ with $v_i \neq v_0$ and scalars $s_1, \ldots, s_n$ so that $v_0 = s_1 v_1 + \cdots + s_n v_n$. Suppose $v \in V$. Since $\mathcal{S}$ is a generating set, there exist vectors $w_1, \ldots, w_k \in \mathcal{S}$ and scalars $t_1, \ldots, t_k$ so that $v = t_1 w_1 + \cdots + t_k w_k$. If $\{w_1, \ldots, w_k\}$ does not contain v_0, then $v \in \operatorname{span}(\mathcal{S} \setminus \{v_0\})$, so assume that $v_0 \in \{w_1, \ldots, w_k\}$. After possibly relabelling the elements of $\{w_1, \ldots, w_k\}$ we can assume that $v_0 = w_1$. Hence we have

$$v = t_1 (s_1 v_1 + \cdots + s_n v_n) + t_2 w_2 + \cdots + t_k w_k$$

with $v_0 \neq v_i$ for $1 \leqslant i \leqslant n$ and $v_0 \neq w_j$ for $2 \leqslant j \leqslant k$. It follows that $v \in \operatorname{span}(\mathcal{S} \setminus \{v_0\})$, as claimed. □

Lemma 3.71 *Let V be a finite dimensional $\mathbb{K}$-vector space and $\mathcal{S} \subset V$ a finite set with n elements which generates V. If $\mathcal{T} \subset V$ has more than n elements, then $\mathcal{T}$ is linearly dependent.*

Proof We show that if $\mathcal{T}$ has exactly $n+1$ elements, then it is linearly dependent. In the other cases, $\mathcal{T}$ contains a subset with exactly $n+1$ elements and if this subset is linearly dependent, then so is $\mathcal{T}$.

We prove the claim by induction on $n \geqslant 0$. Let $\mathcal{A}(n)$ be the following statement: "For any $\mathbb{K}$-vector space V, if there exists a generating subset $\mathcal{S} \subset V$ with n elements, then all subsets of V with $n+1$ elements are linearly dependent."

We first show that $\mathcal{A}(0)$ is true. A subset with zero elements is the empty set $\emptyset$. Hence $V = \operatorname{span}(\emptyset) = \{0_V\}$ is the zero vector space. The only subset of $\{V\}$ with 1 element is $\{0_V\}$. Since $s0_V = 0_V$ for all $s \in \mathbb{K}$, the set $\{0_V\}$ is linearly dependent, thus showing that $\mathcal{A}(0)$ is correct.

Suppose $n \geqslant 1$ and that $\mathcal{A}(n-1)$ is true. We want to argue that $\mathcal{A}(n)$ is true as well. Suppose V is generated by the set $\mathcal{S} = \{v_1, \ldots, v_n\}$ with n elements. Let $\mathcal{T} = \{w_1, \ldots, w_{n+1}\}$ be a subset with $n+1$ elements. We need to show that $\mathcal{T}$ is linearly dependent. Since $\mathcal{S}$ is generating, we have scalars $s_{ij} \in \mathbb{K}$ with $1 \leqslant i \leqslant n+1$ and $1 \leqslant j \leqslant n$ so that

$$w_i = \sum_{j=1}^{n} s_{ij} v_j \tag{3.10}$$

for all $1 \leqslant i \leqslant n+1$. We now consider two cases:

Case 1. If $s_{11} = \cdots = s_{n+1,1} = 0$, then (3.10) gives for all $1 \leqslant i \leqslant n+1$

$$w_i = \sum_{j=2}^{n} s_{ij} v_j.$$

Notice that the summation now starts at $j = 2$. This implies that $\mathcal{T} \subset W$, where $W = \operatorname{span}\{v_2, \ldots, v_n\}$. We can now apply $\mathcal{A}(n-1)$ to the vector space W, the generating set $\mathcal{S}_1 = \{v_2, \ldots, v_n\}$ and the subset with n elements being $\mathcal{T}_1 = \{w_1, \ldots, w_n\}$. It follows that $\mathcal{T}_1$ is linearly dependent and hence so is $\mathcal{T}$, as it contains $\mathcal{T}_1$.

Case 2. Suppose there exists i so that $s_{i1} \neq 0$. Then, after possibly relabelling the vectors, we can assume that $s_{11} \neq 0$. For $2 \leqslant i \leqslant n+1$ we thus obtain from (3.10)

$$\begin{aligned}
w_i - \frac{s_{i1}}{s_{11}} w_1 = w_i - \frac{s_{i1}}{s_{11}} \left(\sum_{j=1}^{n} s_{1j} v_j \right) &= \sum_{j=1}^{n} s_{ij} v_j - \frac{s_{i1}}{s_{11}} \left(\sum_{j=1}^{n} s_{1j} v_j \right) \\
&= \sum_{j=1}^{n} \left(s_{ij} - \frac{s_{i1}}{s_{11}} s_{1j} \right) v_j \\
&= \underbrace{\left(s_{i1} - \frac{s_{i1}}{s_{11}} s_{11} \right)}_{=0} v_1 + \sum_{j=2}^{n} \left(s_{ij} - \frac{s_{i1}}{s_{11}} s_{1j} \right) v_j \\
&= \sum_{j=2}^{n} \left(s_{ij} - \frac{s_{i1}}{s_{11}} s_{1j} \right) v_j .
\end{aligned}$$

Hence, setting

$$\hat{w}_i = w_i - \frac{s_{i1}}{s_{11}} w_1 \tag{3.11}$$

for $2 \leqslant i \leqslant n+1$ and $\hat{s}_{ij} = s_{ij} - \frac{s_{i1}}{s_{11}} s_{1j}$ for $2 \leqslant i \leqslant n+1$ and $2 \leqslant j \leqslant n$, we obtain the relations

$$\hat{w}_i = \sum_{j=2}^{n} \hat{s}_{ij} v_j$$

for all $2 \leqslant i \leqslant n+1$. Therefore, the set $\hat{\mathcal{T}} = \{\hat{w}_2, \ldots, \hat{w}_{n+1}\}$ with n elements is contained in W which is generated by $n-1$ elements. Applying $\mathcal{A}(n-1)$, we conclude that $\hat{\mathcal{T}}$ is linearly dependent. It follows that we have scalars $t_2, \ldots, t_{n+1}$ not all zero so that

$$t_2 \hat{w}_2 + \cdots + t_{n+1} \hat{w}_{n+1} = 0_V .$$

Using (3.11), we get

$$\sum_{i=2}^{n+1} t_i \left(w_i - \frac{s_{i1}}{s_{11}} w_1 \right) = -\left(\sum_{i=2}^{n+1} t_i \frac{s_{i1}}{s_{11}} \right) w_1 + t_2 w_2 + \cdots + t_{n+1} w_{n+1} = 0_V .$$

Since not all scalars $t_2, \ldots, t_{n+1}$ are zero, it follows that $w_1, \ldots, w_{n+1}$ are linearly dependent and hence so is $\mathcal{T}$. □

Proof of Theorem 3.64 We restrict to the case where V is finite dimensional. Hence there exists an integer $n \geqslant 0$ so that V has a generating set $\mathcal{S}_0$ with n elements.

(i) Let $\mathcal{S} \subset V$ be a subset generating V. We consider the set $\mathcal{X}$ consisting of those integers $d \geqslant 0$ for which there exists a linearly independent subset $\mathcal{T} \subset \mathcal{S}$ with d elements. Since $\emptyset \subset \mathcal{S}$, we have $0 \in \mathcal{X}$, so $\mathcal{X}$ is non-empty. Furthermore, $\mathcal{X}$ is a finite set, as it cannot contain any integer greater than n by Lemma 3.71. Let $m \in \mathcal{X}$ be the largest integer and $\mathcal{T} \subset \mathcal{S}$ a set with m elements. We want to argue that $\mathcal{T}$ is a basis of V. Suppose $\mathcal{T}$ is not a basis of V. Then there exists an element $v_0 \in \mathcal{S}$ so that $v_0 \notin \operatorname{span}(\mathcal{T})$, since if no such element exists, we have $\mathcal{S} \subset \operatorname{span}(\mathcal{T})$ and hence $V = \operatorname{span}(\mathcal{S}) \subset \operatorname{span}(\mathcal{T})$ contradicting the assumption that $\mathcal{T}$ is not a basis of V. Applying Lemma 3.69, we conclude that

$\hat{\mathcal{T}} = \{v_0\} \cup \mathcal{T} \subset \mathcal{S}$ is linearly independent. Since $\hat{\mathcal{T}}$ has $m+1$ elements, we have $m+1 \in \mathcal{X}$, contradicting the fact that m is the largest integer in $\mathcal{X}$. It follows that $\mathcal{T}$ must be a basis of V.

(ii) Let $\mathcal{S} \subset V$ be a subset that is linearly independent in V. Let $\hat{\mathcal{X}}$ denote the set consisting of those integers $d \geqslant 0$ for which there exists a subset $\mathcal{T} \subset V$ with d elements, which contains $\mathcal{S}$ and which is a generating set of V. Notice that $\mathcal{S} \cup \mathcal{S}_0$ is such a set, hence $\hat{\mathcal{X}}$ is not empty. Let m denote the smallest element of $\hat{\mathcal{X}}$ and $\mathcal{T}$ be a generating subset of V containing $\mathcal{S}$ and with m elements. We want to argue that $\mathcal{T}$ is basis for V. By assumption, $\mathcal{T}$ generates V, hence we need to check that $\mathcal{T}$ is linearly independent in V. Suppose $\mathcal{T}$ is linearly dependent and write $\mathcal{T} = \{v_1, \ldots, v_m\}$ for distinct elements of V. Suppose $\mathcal{S} = \{v_1, \ldots, v_k\}$ for some $k \leqslant m$. This holds true since $\mathcal{S} \subset \mathcal{T}$. Since $\mathcal{T}$ is linearly dependent we have scalars $s_1, \ldots, s_m$ so that

$$s_1 v_1 + \cdots + s_m v_m = 0_V.$$

There must exist a scalar s_i with $i > k$ such that $s_i \neq 0$. Otherwise $\mathcal{S}$ would be linearly dependent. After possibly relabelling the vectors, we can assume that $s_{k+1} \neq 0$ so that

$$v_{k+1} = -\frac{1}{s_{k+1}}\left(s_1 v_1 + \cdots + s_k v_k + s_{k+2} v_{k+2} + \cdots + s_m v_m\right). \tag{3.12}$$

Let $\hat{\mathcal{T}} = \{v_1, \ldots, v_k, v_{k+2}, \ldots, v_m\}$. Then $\mathcal{S} \subset \hat{\mathcal{T}}$ and (3.12) shows that $v_{k+1} \in \operatorname{span}(\hat{\mathcal{T}})$. Lemma 3.70 shows that $\hat{\mathcal{T}}$ generates V, contains $\mathcal{S}$ and has $m-1$ elements, contradicting the minimality of m.

(iii) Suppose $\mathcal{S}_1$ is a basis of V with n_1 elements and $\mathcal{S}_2$ is a basis of V with n_2 elements. Since $\mathcal{S}_2$ is linearly independent and $\mathcal{S}_1$ generates V, Lemma 3.71 implies that $n_2 \leqslant n_1$. Likewise, we conclude that $n_2 \geqslant n_1$. It follows that $n_1 = n_2$ and hence there exists a bijective mapping from $\mathcal{S}_1$ to $\mathcal{S}_2$ as these are finite sets with the same number of elements.

(iv) is an immediate consequence of (iii). □

3.6.2 Properties of the dimension

Lemma 3.72 *Isomorphic finite dimensional vector spaces have the same dimension.*

Proof Let V, W be finite dimensional $\mathbb{K}$-vector spaces and $f : V \to W$ an isomorphism. Let $\mathcal{S} \subset V$ be a basis of V, then $f(\mathcal{S}) \subset W$ is a basis of W, by combining Lemma 3.46 and Lemma 3.56. Since $\mathcal{S}$ and $f(\mathcal{S})$ have the same number of elements, we have $\dim(V) = \dim(W)$. □

Lemma 3.73 *A subspace of a finite dimensional $\mathbb{K}$-vector space is finite dimensional as well.*

Proof Let V be a finite dimensional $\mathbb{K}$-vector space and $U \subset V$ a subspace. Let $\mathcal{S} = \{v_1, \ldots, v_n\}$ be a basis of V. For $1 \leqslant i \leqslant n$, we define $U_i = U \cap \operatorname{span}\{v_1, \ldots, v_i\}$. By construction, each U_i is a subspace and $U_1 \subset U_2 \subset \cdots \subset U_n = U$, since $\mathcal{S}$ is a basis of V.

We will show inductively that all U_i are finite dimensional. Notice that U_1 is a subspace of $\operatorname{span}\{v_1\}$. The only subspaces of $\operatorname{span}\{v_1\}$ are $\{0_V\}$ and $\{tv_1 \,|\, t \in \mathbb{R}\}$, both are finite dimensional, hence U_1 is finite dimensional.

Assume $i \geqslant 2$. We will show next that if U_{i-1} is finite dimensional, then so is U_i. Let $\mathcal{T}_{i-1}$ be a basis of U_{i-1}. If $U_i = U_{i-1}$, then U_i is finite dimensional as well, so assume there exists a non-zero vector $w \in U_i \setminus U_{i-1}$. Since $\mathcal{S}$ is a basis of V and since $w \in \operatorname{span}\{v_1, \ldots, v_i\}$, there exist scalars $s_1, \ldots, s_i$ so that $w = s_1v_1 + \cdots + s_iv_i$. By assumption, $w \notin U_{i-1}$, hence $s_i \neq 0$. Any vector $v \in U_i$ can be written as $v = t_1v_1 + \cdots + t_iv_i$ for scalars $t_1, \ldots, t_i$. We now compute

$$\begin{aligned} v - \frac{t_i}{s_i}w &= \sum_{k=1}^{i} t_kv_k - \frac{t_i}{s_i}\left(\sum_{k=1}^{i} s_kv_k\right) = \sum_{k=1}^{i}\left(t_k - \frac{t_i}{s_i}s_k\right)v_k \\ &= \sum_{k=1}^{i-1}\left(t_k - \frac{t_i}{s_i}s_k\right)v_k \end{aligned}$$

so that $v - (t_i/s_i)w$ can be written as a linear combination of the vectors $v_1, \ldots, v_{i-1}$, hence is an element of U_{i-1}. Recall that $\mathcal{T}_{i-1}$ is a basis of U_{i-1}, hence $v - (t_i/s_i)w$ is a linear combination of elements of $\mathcal{T}_{i-1}$. It follows that any vector $v \in U_i$ is a linear combination of elements of $\mathcal{T}_{i-1} \cup \{w\}$, that is, $\mathcal{T}_{i-1} \cup \{w\}$ generates U_i. Since $\mathcal{T}_{i-1} \cup \{w\}$ contains finitely many vectors, it follows that U_i is finite dimensional. □

Proposition 3.74 *Let V be a finite dimensional $\mathbb{K}$-vector space. Then for any subspace $U \subset V$*

$$0 \leqslant \dim(U) \leqslant \dim(V).$$

Furthermore $\dim(U) = 0$ *if and only if* $U = \{0_V\}$ *and* $\dim(U) = \dim(V)$ *if and only if* $V = U$.

Proof By Lemma 3.73, U is finite dimensional and hence by Corollary 3.65 admits a basis $\mathcal{S}$. By Theorem 3.64 (ii), there is a basis $\mathcal{T}$ of V which contains $\mathcal{S}$. Therefore

$$0 \leqslant \dim(U) = \operatorname{Card}(\mathcal{S}) \leqslant \operatorname{Card}(\mathcal{T}) = \dim(V).$$

Suppose $\dim(V) = \dim(U)$, then $\operatorname{Card}(\mathcal{S}) = \operatorname{Card}(\mathcal{T})$ and hence $\mathcal{S} = \mathcal{T}$ since every element of $\mathcal{S}$ is an element of $\mathcal{T}$ and $\mathcal{S}$ and $\mathcal{T}$ have the same number of elements. Therefore, we get $U = \operatorname{span}(\mathcal{S}) = \operatorname{span}(\mathcal{T}) = V$. Since $\dim U = 0$ if and only if the empty set is a basis for U we have $\dim U = 0$ if and only if $U = \{0_V\}$. □

Definition 3.75 (Rank of a linear map and matrix) Let V, W be $\mathbb{K}$-vector spaces with W finite dimensional. The *rank* of a linear map $f : V \to W$ is defined as

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

If $\mathbf{A} \in M_{m,n}(\mathbb{K})$ is a matrix, then we define

$$\operatorname{rank}(\mathbf{A}) = \operatorname{rank}(f_{\mathbf{A}}).$$

The *nullity* of a linear map $f : V \to W$ is the dimension of its kernel, $\operatorname{nullity}(f) = \dim \operatorname{Ker}(f)$. The following *important* theorem establishes a relation between the nullity and the rank of a linear map. It states something that is intuitively not surprising, namely that the dimension of the image of a linear map $f : V \to W$ is the dimension of the vector space V minus the dimension of the subspace of vectors that we “lose”, that is, those that are mapped onto the zero vector of W. More precisely:

Theorem 3.76 (Rank–nullity theorem) *Let V, W be finite dimensional $\mathbb{K}$-vector spaces and $f : V \to W$ a linear map. Then we have*

$$\dim(V) = \dim \operatorname{Ker}(f) + \dim \operatorname{Im}(f) = \operatorname{nullity}(f) + \operatorname{rank}(f).$$

Proof Let $d = \dim \operatorname{Ker}(f)$ and $n = \dim V$, so that $d \leqslant n$ by Proposition 3.74. Let $\{v_1, \dots, v_d\}$ be a basis of $\mathcal{S} = \operatorname{Ker}(f)$. By Theorem 3.64 (ii) we can find linearly independent vectors $\hat{\mathcal{S}} = \{v_{d+1}, \dots, v_n\}$ so that $\mathcal{T} = \mathcal{S} \cup \hat{\mathcal{S}}$ is a basis of V. Now $U = \operatorname{span}(\hat{\mathcal{S}})$ is a subspace of V of dimension $n - d$. We consider the linear map

$$g : U \to \operatorname{Im}(f), \qquad v \mapsto f(v).$$

We want to show that g is an isomorphism, since then $\dim \operatorname{Im}(f) = \dim(U) = n - d$, so that

$$\dim \operatorname{Im}(f) = n - d = \dim(V) - \dim \operatorname{Ker}(f),$$

as claimed.

We first show that g is injective. Assume $g(v) = 0_W$. Since $v \in U$, we can write $v = s_{d+1}v_{d+1} + \cdots + s_n v_n$ for scalars $s_{d+1}, \dots, s_n$. Since $g(v) = 0_W$ we have $v \in \operatorname{Ker}(f)$, hence we can also write $v = s_1 v_1 + \cdots + s_d v_d$ for scalars $s_1, \dots, s_d$, subtracting the two expressions for v, we get

$$0_V = s_1 v_1 + \cdots + s_d v_d - s_{d+1} v_{d+1} - \cdots - s_n v_n.$$

Since $\{v_1, \dots, v_n\}$ is a basis, it follows that all the coefficients s_i vanish, where $1 \leqslant i \leqslant n$. Therefore we have $v = 0_V$ and g is injective.

Second, we show that g is surjective. Suppose $w \in \operatorname{Im}(f)$ so that $w = f(v)$ for some vector $v \in V$. We write $v = \sum_{i=1}^n s_i v_i$ for scalars $s_1, \dots, s_n$. Using the linearity of f, we compute

$$w = f(v) = f\left(\sum_{i=1}^n s_i v_i\right) = f\Big(\underbrace{\sum_{i=d+1}^n s_i v_i}_{=\hat{v}}\Big) = f(\hat{v})$$

where $\hat{v} \in U$. We thus have an element $\hat{v}$ with $g(\hat{v}) = w$. Since w was arbitrary, we conclude that g is surjective. □

Corollary 3.77 *Let V, W be finite dimensional $\mathbb{K}$-vector spaces with $\dim(V) = \dim(W)$ and $f : V \to W$ a linear map. Then the following statements are equivalent:*

(i) f is injective;
(ii) f is surjective;
(iii) f is bijective.

Proof (i) ⇒ (ii) By Lemma 3.31, the map f is injective if and only if $\operatorname{Ker}(f) = \{0_V\}$ so that $\dim \operatorname{Ker}(f) = 0$ by Example 3.68 (i). Theorem 3.76 implies that $\dim \operatorname{Im}(f) = \dim(V) = \dim(W)$ and hence Proposition 3.74 implies that $\operatorname{Im}(f) = W$, that is, f is surjective.

(ii) ⇒ (iii) Since f is surjective $\operatorname{Im}(f) = W$ and hence $\dim \operatorname{Im}(f) = \dim(W) = \dim(V)$. Theorem 3.76 implies that $\dim \operatorname{Ker}(f) = 0$ so that $\operatorname{Ker}(f) = \{0_V\}$ by Proposition 3.74. Applying Lemma 3.31 again shows that f is injective and hence bijective.

(iii) ⇒ (i) Since f is bijective, it is also injective. □

Corollary 3.78 *Let V, W be finite dimensional $\mathbb{K}$-vector spaces and $f : V \to W$ a linear map. Then* $\operatorname{rank}(f) \leqslant \min\{\dim(V), \dim(W)\}$ *and*

$$\operatorname{rank}(f) = \dim(V) \iff f \text{ is injective,}$$
$$\operatorname{rank}(f) = \dim(W) \iff f \text{ is surjective.}$$

Proof For the first claim it is sufficient to show that $\operatorname{rank}(f) \leqslant \dim(V)$ and $\operatorname{rank}(f) \leqslant \dim(W)$. By definition, $\operatorname{rank}(f) = \dim \operatorname{Im}(f)$ and since $\operatorname{Im}(f) \subset W$, we have $\operatorname{rank}(f) = \dim \operatorname{Im}(f) \leqslant \dim(W)$ with equality if and only if f is surjective, by Proposition 3.74.

Theorem 3.76 implies that $\operatorname{rank}(f) = \dim \operatorname{Im}(f) = \dim(V) - \dim \operatorname{Ker}(f) \leqslant \dim(V)$ with equality if and only if $\dim \operatorname{Ker}(f) = 0$, that is, when f is injective (as we have just seen in the proof of the previous corollary). □

Corollary 3.79 *Let V, W be finite dimensional $\mathbb{K}$-vector spaces and $f : V \to W$ a linear map. Then we have*

(i) If $\dim(V) < \dim(W)$, then f is not surjective;

(ii) If $\dim(V) > \dim(W)$, then f is not injective. In particular, there exist non-zero vectors $v \in V$ with $f(v) = 0_W$.

Proof (i) Suppose $\dim(V) < \dim(W)$, then by Theorem 3.76

$$\operatorname{rank}(f) = \dim(V) - \dim \operatorname{Ker}(f) \leqslant \dim(V) < \dim(W)$$

and the claim follows from Corollary 3.78.

(ii) Suppose $\dim(V) > \dim(W)$, then

$$\operatorname{rank}(f) \leqslant \dim(W) < \dim(V)$$

and the claim follows from Corollary 3.78. □

Proposition 3.80 *Let V, W be finite dimensional $\mathbb{K}$-vector spaces. Then there exists an isomorphism $\Theta : V \to W$ if and only if $\dim(V) = \dim(W)$.*

Proof $\Rightarrow$ This was already proved in Lemma 3.72.

$\Leftarrow$ Let $\dim(V) = \dim(W) = n \in \mathbb{N}$. Choose a basis $\mathcal{T} = \{w_1, \ldots, w_n\}$ of W and consider the linear map

$$\Theta : \mathbb{K}^n \to W, \quad \vec{x} \mapsto x_1 w_1 + \cdots + x_n w_n,$$

where $\vec{x} = (x_i)_{1 \leqslant i \leqslant n}$ Notice that Θ is injective. Indeed, if $\Theta(\vec{x}) = x_1 w_1 + \cdots + x_n w_n = 0_W$, then $x_1 = \cdots = x_n = 0$, since $\{w_1, \ldots, w_n\}$ are linearly independent. We thus conclude $\operatorname{Ker}\Theta = \{0_V\}$ and hence Lemma 3.31 implies that Θ is injective and therefore bijective by Corollary 3.77. The map Θ is linear and bijective, thus an isomorphism. Likewise, for a choice of basis $\mathcal{S} = \{v_1, \ldots, v_n\}$ of V, we obtain an isomorphism $\Phi : \mathbb{K}^n \to V$. Since the composition of bijective maps is again bijective, the map $\Theta \circ \Phi^{-1} : V \to W$ is bijective and since by Proposition 3.16 the composition of linear maps is again linear, the map $\Theta \circ \Phi^{-1} : V \to W$ is an isomorphism. □

Corollary 3.81 *Suppose* $\mathbf{A} \in M_{m,n}(\mathbb{K})$ *is invertible with inverse* $\mathbf{A}^{-1} \in M_{n,m}(\mathbb{K})$*. Then* $n = m$*, hence* $\mathbf{A}$ *is a square matrix.*

Proof Consider $f_{\mathbf{A}} : \mathbb{K}^n \to \mathbb{K}^m$. By Proposition 3.17, $f_{\mathbf{A}}$ is bijective and hence an isomorphism. Proposition 3.80 implies that $n = m$. □

3.7 Matrix representation of linear maps

WEEK 6

Notice that Proposition 3.80 implies that every finite dimensional $\mathbb{K}$-vector space V is isomorphic to $\mathbb{K}^n$, where $n = \dim(V)$. Choosing an isomorphism from V to $\mathbb{K}^n$ allows to uniquely describe each vector of V in terms of n scalars, its *coordinates*.

Definition 3.82 (Linear coordinate system) Let V be a $\mathbb{K}$-vector space of dimension $n \in \mathbb{N}$. A *linear coordinate system* is an injective linear map $\varphi : V \to \mathbb{K}^n$. The entries of the vector $\varphi(v) \in \mathbb{K}^n$ are called the *coordinates* of the vector $v \in V$ with respect to the coordinate system φ.

We only request that φ is injective, but the mapping φ is automatically bijective by Corollary 3.77.

Example 3.83 (Standard coordinates) On the vector space $\mathbb{K}^n$ we have a linear coordinate system defined by the identity mapping, that is, we define $\varphi(\vec{v}) = \vec{v}$ for all $\vec{v} \in \mathbb{K}^n$. We call this coordinate system the *standard coordinate system of* $\mathbb{K}^n$.

Example 3.84 (Non-linear coordinates) In Linear Algebra we only consider linear coordinate systems, but in other areas of mathematics *non-linear coordinate systems* are also used. An example are the so-called *polar coordinates*

$$\rho : \mathbb{R}^2 \setminus \{0_{\mathbb{R}^2}\} \to (0, \infty) \times (-\pi, \pi] \subset \mathbb{R}^2, \qquad \vec{x} \mapsto \begin{pmatrix} r \\ \phi \end{pmatrix} = \begin{pmatrix} \sqrt{(x_1)^2 + (x_2)^2} \\ \arg(\vec{x}) \end{pmatrix},$$

where $\arg(\vec{x}) = \arccos(x_1/r)$ for $x_2 \geqslant 0$ and $\arg(\vec{x}) = -\arccos(x_1/r)$ for $x_2 < 0$. Notice that the polar coordinates are only defined on $\mathbb{R}^2 \setminus \{0_{\mathbb{R}^2}\}$. For further details we refer to the Analysis module.

A convenient way to visualise a linear coordinate system $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ is to consider the preimage $\varphi^{-1}(\mathcal{C})$ of the *standard coordinate grid*

$$\mathcal{C} = \{s\vec{e}_1 + k\vec{e}_2 | s \in \mathbb{R}, k \in \mathbb{Z}\} \cup \{k\vec{e}_1 + s\vec{e}_2 | s \in \mathbb{R}, k \in \mathbb{Z}\} \tag{3.13}$$

under φ. The first set in the union (3.13) of sets are the *horizontal coordinate lines* and the second set the *vertical coordinate lines*.

Example 3.85 (see Figure 3.1) The vector $\vec{v} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ has coordinates $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ with respect to the standard coordinate system of $\mathbb{R}^2$. The same vector has coordinates $\varphi(\vec{v}) = \begin{pmatrix} 4 \\ -1 \end{pmatrix}$ with respect to the coordinate system $\varphi\left(\begin{pmatrix} v_1 \\ v_2 \end{pmatrix}\right) = \begin{pmatrix} v_1 + 2v_2 \\ -v_1 + v_2 \end{pmatrix}$.

While $\mathbb{K}^n$ is equipped with the standard coordinate system, in an abstract vector space V there is no preferred linear coordinate system and a choice of linear coordinate system amounts to choosing a so-called ordered basis of V.

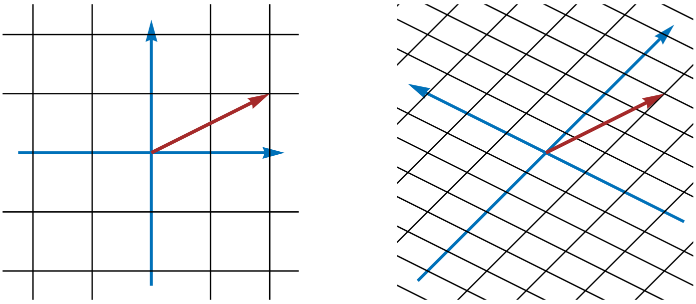

FIGURE 3.1. The coordinates of a vector with respect to different coordinate systems.

Definition 3.86 (Ordered basis) Let V be a finite dimensional $\mathbb{K}$-vector space. An (ordered) n-tuple $\mathbf{b} = (v_1, \ldots, v_n)$ of vectors from V is called an *ordered basis of V* if the set $\{v_1, \ldots, v_n\}$ is a basis of V.

That there is a bijective correspondence between ordered bases of V and linear coordinate systems on V is a consequence of the following very important lemma which states in particular that two linear maps $f, g : V \to W$ are the same if and only if they agree on a basis of V.

Lemma 3.87 *Let V, W be finite dimensional $\mathbb{K}$-vector spaces.*

(i) Suppose $f, g : V \to W$ are linear maps and $\mathbf{b} = (v_1, \ldots, v_n)$ is an ordered basis of V. Then $f = g$ if and only if $f(v_i) = g(v_i)$ for all $1 \leqslant i \leqslant n$.

(ii) If $\dim V = \dim W$ and $\mathbf{b} = (v_1, \ldots, v_n)$ is an ordered basis of V and $\mathbf{c} = (w_1, \ldots, w_n)$ an ordered basis of W, then there exists a unique isomorphism $f : V \to W$ such that $f(v_i) = w_i$ for all $1 \leqslant i \leqslant n$.

Proof (i) $\Rightarrow$ If $f = g$ then $f(v_i) = g(v_i)$ for all $1 \leqslant i \leqslant n$. $\Leftarrow$ Let $v \in V$. Since $\mathbf{b}$ is an ordered basis of V there exist unique scalars $s_1, \ldots, s_n \in \mathbb{K}$ such that $v = \sum_{i=1}^n s_i v_i$. Using the linearity of f and g, we compute

$$f(v) = f\left(\sum_{i=1}^n s_i v_i\right) = \sum_{i=1}^n s_i f(v_i) = \sum_{i=1}^n s_i g(v_i) = g\left(\sum_{i=1}^n s_i v_i\right) = g(v)$$

so that $f = g$.

(ii) Let $v \in V$. Since $\{v_1, \ldots, v_n\}$ is a basis of V there exist unique scalars $s_1, \ldots, s_n$ such that $v = \sum_{i=1}^n s_i v_i$. We define $f(v) = \sum_{i=1}^n s_i w_i$, so that in particular $f(v_i) = w_i$ for $1 \leqslant i \leqslant n$. Since $\{w_1, \ldots, w_n\}$ are linearly independent we have $f(v) = 0_W$ if and only if $s_1 = \cdots = s_n = 0$, that is $v = 0_V$. Lemma 3.31 implies that f is injective and hence an isomorphism by Corollary 3.77. The uniqueness of f follows from (i). □

Remark 3.88 Notice that Lemma 3.87 is wrong for maps that are not linear. Consider

$$f : \mathbb{R}^2 \to \mathbb{R}, \qquad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto x_1 x_2$$

and

$$g : \mathbb{R}^2 \to \mathbb{R} \quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto (x_1 - 1)(x_2 - 1).$$

Then $f(\vec{e}_1) = g(\vec{e}_1)$ and $f(\vec{e}_2) = g(\vec{e}_2)$, but $f \neq g$.

Given an ordered basis $\mathbf{b} = (v_1, \ldots, v_n)$ of V, the previous lemma implies that there is a unique linear coordinate system $\boldsymbol{\beta} : V \to \mathbb{K}^n$ such that

$$\boldsymbol{\beta}(v_i) = \vec{e}_i \tag{3.14}$$

for $1 \leqslant i \leqslant n$, where $\{\vec{e}_1, \ldots, \vec{e}_n\}$ denotes the standard basis of $\mathbb{K}^n$. Conversely, if $\boldsymbol{\beta} : V \to \mathbb{K}^n$ is a linear coordinate system, we obtain an ordered basis of V

$$\mathbf{b} = (\boldsymbol{\beta}^{-1}(\vec{e}_1), \ldots, \boldsymbol{\beta}^{-1}(\vec{e}_n))$$

and these assignments are inverse to each other. Notice that for all $v \in V$ we have

$$\boldsymbol{\beta}(v) = \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix} \qquad \iff \qquad v = s_1 v_1 + \cdots + s_n v_n.$$

Remark 3.89 (Notation) We will denote an ordered basis by an upright bold Roman letter, such as $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$ or $\mathbf{e}$. We will denote the corresponding linear coordinate system by the corresponding bold Greek letter $\boldsymbol{\beta}$,$\boldsymbol{\gamma}$,$\boldsymbol{\delta}$ or $\boldsymbol{\varepsilon}$, respectively.

Example 3.90 Let $V = \mathbb{K}^3$ and $\mathbf{e} = (\vec{e}_1, \vec{e}_2, \vec{e}_3)$ denote the ordered standard basis. Then for all $\vec{x} = (x_i)_{1 \leqslant i \leqslant 3} \in \mathbb{R}^3$ we have

$$\boldsymbol{\varepsilon}(\vec{x}) = \vec{x}.$$

where $\boldsymbol{\varepsilon}$ denotes the linear coordinate system corresponding to $\mathbf{e}$. Notice that $\boldsymbol{\varepsilon}$ is the standard coordinate system on $\mathbb{K}^n$. Considering instead the ordered basis $\mathbf{b} = (\vec{v}_1, \vec{v}_2, \vec{v}_3) = (\vec{e}_1 + \vec{e}_3, \vec{e}_3, \vec{e}_2 - \vec{e}_1)$, we obtain

$$\boldsymbol{\beta}(\vec{x}) = \begin{pmatrix} x_1 + x_2 \\ x_3 - x_1 - x_2 \\ x_2 \end{pmatrix}$$

since

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = (x_1 + x_2) \underbrace{\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}}_{=\vec{v}_1} + (x_3 - x_1 - x_2) \underbrace{\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}}_{=\vec{v}_2} + x_2 \underbrace{\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}}_{=\vec{v}_3}.$$

Fixing linear coordinate systems – or equivalently ordered bases – on finite dimensional vector spaces V, W allows to describe each linear map $g : V \to W$ in terms of a matrix:

Definition 3.91 (Matrix representation of a linear map — Video) Let V, W be finite dimensional $\mathbb{K}$-vector spaces, $\mathbf{b}$ an ordered basis of V and $\mathbf{c}$ an ordered basis of W. The *matrix representation of a linear map $g : V \to W$ with respect to the ordered bases* $\mathbf{b}$ *and* $\mathbf{c}$ is the unique matrix $\mathbf{M}(g, \mathbf{b}, \mathbf{c}) \in M_{m,n}(\mathbb{K})$ such that

$$f_{\mathbf{M}(g,\mathbf{b},\mathbf{c})} = \boldsymbol{\gamma} \circ g \circ \boldsymbol{\beta}^{-1},$$

where β and γ denote the linear coordinate systems corresponding to $\mathbf{b}$ and $\mathbf{c}$, respectively.

The role of the different mappings can be summarised in terms of the following diagram:

$$\begin{array}{ccc} V & \xrightarrow{\;g\;} & W \\ {\scriptstyle \beta^{-1}}\uparrow & & \downarrow{\scriptstyle \gamma} \\ \mathbb{K}^n & \xrightarrow{\;f_{\mathbf{M}(g,\mathbf{b},\mathbf{c})}\;} & \mathbb{K}^m \end{array}$$

In practise, we can compute the matrix representation of a linear map as follows:

Proposition 3.92 *Let V, W be finite dimensional $\mathbb{K}$-vector spaces, $\mathbf{b} = (v_1, \ldots, v_n)$ an ordered basis of V, $\mathbf{c} = (w_1, \ldots, w_m)$ an ordered basis of W and $g : V \to W$ a linear map. Then there exist unique scalars $A_{ij} \in \mathbb{K}$, where $1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n$ such that*

$$g(v_j) = \sum_{i=1}^{m} A_{ij} w_i, \qquad 1 \leqslant j \leqslant n. \tag{3.15}$$

Furthermore, the matrix $\mathbf{A} = (A_{ij})_{1\leqslant i\leqslant m, 1\leqslant j\leqslant n}$ satisfies

$$f_{\mathbf{A}} = \gamma \circ g \circ \beta^{-1}$$

and hence is the matrix representation of g with respect to the ordered bases $\mathbf{b}$ and $\mathbf{c}$.

Remark 3.93 Notice that we sum over the first index of A_{ij} in (3.15).

Proof of Proposition 3.92 For all $1 \leqslant j \leqslant n$ the vector $g(v_j)$ is an element of W and hence a linear combination of the vectors $\mathbf{c} = (w_1, \ldots, w_m)$, as $\mathbf{c}$ is an ordered basis of W. We thus have scalars $A_{ij} \in \mathbb{K}$ with $1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n$ such that $g(v_j) = \sum_{i=1}^m A_{ij} w_i$. If $\hat{A}_{ij} \in \mathbb{K}$ with $1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n$ also satisfy $g(v_j) = \sum_{i=1}^m \hat{A}_{ij} w_i$, then subtracting the two equations gives

$$g(v_j) - g(v_j) = 0_W = \sum_{i=1}^{m} (A_{ij} - \hat{A}_{ij}) w_i$$

so that $0 = A_{ij} - \hat{A}_{ij}$ for $1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n$, since the vectors $(w_1, \ldots, w_m)$ are linearly independent. It follows that the scalars A_{ij} are unique.

We want to show that $f_{\mathbf{A}} \circ \beta = \gamma \circ g$. Using Lemma 3.87 it is sufficient to show that $(f_{\mathbf{A}} \circ \beta)(v_j) = (\gamma \circ g)(v_j)$ for $1 \leqslant j \leqslant n$. Let $\{\vec{e}_1, \ldots, \vec{e}_n\}$ denote the standard basis of $\mathbb{K}^n$ so that $\beta(v_j) = \vec{e}_j$ and $\{\vec{d}_1, \ldots, \vec{d}_m\}$ the standard basis of $\mathbb{K}^m$ so that $\gamma(w_i) = \vec{d}_i$. We compute

$$\begin{aligned}(f_{\mathbf{A}} \circ \beta)(v_j) &= f_{\mathbf{A}}(\vec{e}_j) = \mathbf{A}\vec{e}_j = \sum_{i=1}^{m} A_{ij} \vec{d}_i = \sum_{i=1}^{m} A_{ij} \gamma(w_i) = \gamma\left(\sum_{i=1}^{m} A_{ij} w_i\right) \\ &= \gamma(g(v_j)) = (\gamma \circ g)(v_j)\end{aligned}$$

where we have used the linearity of γ and (3.15). □

This all translates to a simple recipe for calculating the matrix representation of a linear map, which we now illustrate in some examples.

Example 3.94 Let $V = \mathrm{P}_2(\mathbb{R})$ and $W = \mathrm{P}_1(\mathbb{R})$ and $g = \frac{\mathrm{d}}{\mathrm{d}x}$. We consider the ordered basis $\mathbf{b} = (v_1, v_2, v_3) = ((1/2)(3x^2 - 1), x, 1)$ of V and $\mathbf{c} = (w_1, w_2) = (x, 1)$ of W.

(i) Compute the image under g of the elements v_i of the ordered basis $\mathbf{b}$.

$$g\left(\frac{1}{2}(3x^2 - 1)\right) = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{2}(3x^2 - 1)\right) = 3x$$
$$g\,(x) = \frac{\mathrm{d}}{\mathrm{d}x}(x) = 1$$
$$g\,(1) = \frac{\mathrm{d}}{\mathrm{d}x}(1) = 0.$$

(ii) Write the image vectors as linear combinations of the elements of the ordered basis $\mathbf{c}$.

$$\begin{aligned} 3x &= 3 \cdot w_1 + 0 \cdot w_2 \\ 1 &= 0 \cdot w_1 + 1 \cdot w_2 \\ 0 &= 0 \cdot w_1 + 0 \cdot w_2 \end{aligned} \tag{3.16}$$

(iii) Taking the transpose of the matrix of coefficients appearing in (3.16) gives the matrix representation

$$\mathbf{M}\left(\frac{\mathrm{d}}{\mathrm{d}x}, \mathbf{b}, \mathbf{c}\right) = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

of the linear map $g = \frac{\mathrm{d}}{\mathrm{d}x}$ with respect to the bases $\mathbf{b}, \mathbf{c}$.

Example 3.95 Let $\mathbf{e} = (\vec{e}_1, \ldots, \vec{e}_n)$ and $\mathbf{d} = (\vec{d}_1, \ldots, \vec{d}_m)$ denote the ordered standard basis of $\mathbb{K}^n$ and $\mathbb{K}^m$, respectively. Then for $\mathbf{A} \in M_{m,n}(\mathbb{K})$, we have

$$\mathbf{A} = \mathbf{M}(f_\mathbf{A}, \mathbf{e}, \mathbf{d}),$$

that is, the matrix representation of the mapping $f_\mathbf{A} : \mathbb{K}^n \to \mathbb{K}^m$ with respect to the standard bases is simply the matrix $\mathbf{A}$. Indeed, we have

$$f_\mathbf{A}(\vec{e}_j) = \mathbf{A}\vec{e}_j = \begin{pmatrix} A_{1j} \\ \vdots \\ A_{mj} \end{pmatrix} = \sum_{i=1}^{m} A_{ij}\vec{d}_i.$$

Example 3.96 Let $\mathbf{e} = (\vec{e}_1, \vec{e}_2)$ denote the ordered standard basis of $\mathbb{R}^2$. Consider the matrix

$$\mathbf{A} = \begin{pmatrix} 5 & 1 \\ 1 & 5 \end{pmatrix} = \mathbf{M}(f_\mathbf{A}, \mathbf{e}, \mathbf{e}).$$

We want to compute $\mathrm{Mat}(f_\mathbf{A}, \mathbf{b}, \mathbf{b})$, where $\mathbf{b} = (\vec{v}_1, \vec{v}_2) = (\vec{e}_1 + \vec{e}_2, \vec{e}_2 - \vec{e}_1)$ is *not* the standard basis of $\mathbb{R}^2$. We obtain

$$f_\mathbf{A}(\vec{v}_1) = A\vec{v}_1 = \begin{pmatrix} 5 & 1 \\ 1 & 5 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 6 \\ 6 \end{pmatrix} = 6 \cdot \vec{v}_1 + 0 \cdot \vec{v}_2$$
$$f_\mathbf{A}(\vec{v}_2) = A\vec{v}_2 = \begin{pmatrix} 5 & 1 \\ 1 & 5 \end{pmatrix}\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -4 \\ 4 \end{pmatrix} = 0 \cdot \vec{v}_1 + 4 \cdot \vec{v}_2$$

Therefore, we have

$$\mathbf{M}(f_\mathbf{A}, \mathbf{b}, \mathbf{b}) = \begin{pmatrix} 6 & 0 \\ 0 & 4 \end{pmatrix}.$$

Proposition 3.97 *Let V, W be finite dimensional $\mathbb{K}$-vector spaces, $\mathbf{b}$ an ordered basis of V with corresponding linear coordinate system β, $\mathbf{c}$ an ordered basis of W with corresponding linear coordinate system γ and $g : V \to W$ a linear map. Then for all $v \in V$ we have*

$$\gamma(g(v)) = \mathbf{M}(g, \mathbf{b}, \mathbf{c})\beta(v).$$

Proof By definition we have for all $\vec{x} \in \mathbb{K}^n$ and $\mathbf{A} \in M_{m,n}(\mathbb{K})$

$$\mathbf{A}\vec{x} = f_{\mathbf{A}}(\vec{x}).$$

Combining this with Definition 3.91, we obtain for all $v \in V$

$$\mathbf{M}(g, \mathbf{b}, \mathbf{c})\beta(v) = f_{\mathbf{M}(g,\mathbf{b},\mathbf{c})}(\beta(v)) = (\gamma \circ g \circ \beta^{-1})(\beta(v)) = \gamma(g(v)),$$

as claimed. □

Remark 3.98 Explicitly, Proposition 3.97 states the following. Let $\mathbf{A} = \mathbf{M}(g, \mathbf{b}, \mathbf{c})$ and let $v \in V$. Since $\mathbf{b}$ is an ordered basis of V, there exist unique scalars $s_i \in \mathbb{K}$, $1 \leqslant i \leqslant n$ such that

$$v = s_1 v_1 + \cdots + s_n v_n.$$

Then we have

$$g(v) = t_1 w_1 + \cdots + t_m w_m,$$

where

$$\begin{pmatrix} t_1 \\ \vdots \\ t_m \end{pmatrix} = \mathbf{A} \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix}.$$

Example 3.99 (Example 3.94 continued) With respect to the ordered basis $\mathbf{b} = \left(\frac{1}{2}(3x^2 - 1), x, 1\right)$, the polynomial $a_2x^2 + a_1x + a_0 \in V = \mathsf{P}_2(\mathbb{R})$ is represented by the vector

$$\beta(a_2x^2 + a_1x + a_0) = \begin{pmatrix} \frac{2}{3}a_2 \\ a_1 \\ \frac{a_2}{3} + a_0 \end{pmatrix}$$

Indeed

$$a_2x^2 + a_1x + a_0 = \frac{2}{3}a_2\left(\frac{1}{2}(3x^2 - 1)\right) + a_1x + \left(\frac{a_2}{3} + a_0\right)1.$$

Computing $\mathbf{M}(\frac{\mathrm{d}}{\mathrm{d}x}, \mathbf{b}, \mathbf{c})\beta(a_2x^2 + a_1x + a_0)$ gives

$$\begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{2}{3}a_2 \\ a_1 \\ \frac{a_2}{3} + a_0 \end{pmatrix} = \begin{pmatrix} 2a_2 \\ a_1 \end{pmatrix}$$

and this vector represents the polynomial $2a_2 \cdot x + a_1 \cdot 1 = \frac{\mathrm{d}}{\mathrm{d}x}(a_2x^2 + a_1x + a_0)$ with respect to the basis $\mathbf{c} = (x, 1)$ of $\mathsf{P}_1(\mathbb{R})$.

As a corollary to Proposition 3.92 we obtain:

Corollary 3.100 *Let V_1, V_2, V_3 be finite dimensional $\mathbb{K}$-vector spaces and $\mathbf{b}_i$ an ordered basis of V_i for $i = 1, 2, 3$. Let $g_1 : V_1 \to V_2$ and $g_2 : V_2 \to V_3$ be linear maps.*

Then

$$\mathbf{M}(g_2 \circ g_1, \mathbf{b}_1, \mathbf{b}_3) = \mathbf{M}(g_2, \mathbf{b}_2, \mathbf{b}_3)\mathbf{M}(g_1, \mathbf{b}_1, \mathbf{b}_2).$$

Proof Let us write $\mathbf{C} = \mathbf{M}(g_2 \circ g_1, \mathbf{b}_1, \mathbf{b}_3)$ and $\mathbf{A}_1 = \mathbf{M}(g_1, \mathbf{b}_1, \mathbf{b}_2)$ as well as $\mathbf{A}_2 = \mathbf{M}(g_2, \mathbf{b}_2, \mathbf{b}_3)$. Using Proposition 2.20 and Theorem 2.21 it suffices to show that $f_{\mathbf{C}} = f_{\mathbf{A}_2\mathbf{A}_1} = f_{\mathbf{A}_2} \circ f_{\mathbf{A}_1}$. Now Proposition 3.92 gives

$$f_{\mathbf{A}_2} \circ f_{\mathbf{A}_1} = \boldsymbol{\beta}_3 \circ g_2 \circ \boldsymbol{\beta}_2^{-1} \circ \boldsymbol{\beta}_2 \circ g_1 \circ \boldsymbol{\beta}_1^{-1} = \boldsymbol{\beta}_3 \circ g_2 \circ g_1 \circ \boldsymbol{\beta}_1^{-1} = f_{\mathbf{C}}.$$

□

Proposition 3.101 *Let V, W be finite dimensional $\mathbb{K}$-vector spaces, $\mathbf{b}$ an ordered basis of V and $\mathbf{c}$ an ordered basis of W. A linear map $g : V \to W$ is bijective if and only if $\mathbf{M}(g, \mathbf{b}, \mathbf{c})$ is invertible. Moreover, in the case where g is bijective we have*

$$\mathbf{M}(g^{-1}, \mathbf{c}, \mathbf{b}) = (\mathbf{M}(g, \mathbf{b}, \mathbf{c}))^{-1}.$$

Proof Let $n = \dim(V)$ and $m = \dim(W)$.

$\Rightarrow$ Let $g : V \to W$ be bijective so that g is an isomorphism and hence $n = \dim(V) = \dim(W) = m$ by Proposition 3.80. Then Corollary 3.100 gives

$$\mathbf{M}(g^{-1}, \mathbf{c}, \mathbf{b})\mathbf{M}(g, \mathbf{b}, \mathbf{c}) = \mathbf{M}(g^{-1} \circ g, \mathbf{b}, \mathbf{b}) = \mathbf{M}(\mathrm{Id}_V, \mathbf{b}, \mathbf{b}) = \mathbf{1}_n$$

and

$$\mathbf{M}(g, \mathbf{b}, \mathbf{c})\mathbf{M}(g^{-1}, \mathbf{c}, \mathbf{b}) = \mathbf{M}(g \circ g^{-1}, \mathbf{c}, \mathbf{c}) = \mathbf{M}(\mathrm{Id}_W, \mathbf{c}, \mathbf{c}) = \mathbf{1}_n$$

so that $\mathbf{M}(g, \mathbf{b}, \mathbf{c})$ is invertible with inverse $\mathbf{M}(g^{-1}, \mathbf{c}, \mathbf{b})$.

$\Leftarrow$ Conversely suppose $\mathbf{A} = \mathbf{M}(g, \mathbf{b}, \mathbf{c})$ is invertible with inverse $\mathbf{A}^{-1}$. It follows that $n = m$ by Corollary 3.81. We consider $h = \boldsymbol{\beta}^{-1} \circ f_{\mathbf{A}^{-1}} \circ \boldsymbol{\gamma} : W \to V$ and since $f_{\mathbf{A}} = \boldsymbol{\gamma} \circ g \circ \boldsymbol{\beta}^{-1}$ by Proposition 3.92, we have

$$g \circ h = \boldsymbol{\gamma}^{-1} \circ f_{\mathbf{A}} \circ \boldsymbol{\beta} \circ \boldsymbol{\beta}^{-1} \circ f_{\mathbf{A}^{-1}} \circ \boldsymbol{\gamma} = \boldsymbol{\gamma}^{-1} \circ f_{\mathbf{A}\mathbf{A}^{-1}} \circ \boldsymbol{\gamma} = \mathrm{Id}_W.$$

Likewise, we have

$$h \circ g = \boldsymbol{\beta}^{-1} \circ f_{\mathbf{A}^{-1}} \circ \boldsymbol{\gamma} \circ \boldsymbol{\gamma}^{-1} \circ f_{\mathbf{A}} \circ \boldsymbol{\beta} = \boldsymbol{\beta}^{-1} \circ f_{\mathbf{A}^{-1}\mathbf{A}} \circ \boldsymbol{\beta} = \mathrm{Id}_V,$$

showing that g admits an inverse mapping $h : W \to V$ and hence g is bijective. □

Recall that a mapping $f : \mathcal{X} \to \mathcal{Y}$ between sets $\mathcal{X}, \mathcal{Y}$ is said to admit a *left inverse* if there exists a mapping $g : \mathcal{Y} \to \mathcal{X}$ such that $g \circ f = \mathrm{Id}_{\mathcal{X}}$. Likewise, a *right inverse* is a mapping $h : \mathcal{Y} \to \mathcal{X}$ such that $f \circ h = \mathrm{Id}_{\mathcal{Y}}$.

We now have:

Proposition 3.102 *Let $n \in \mathbb{N}$ and $\mathbf{A} \in M_{n,n}(\mathbb{K})$ a square matrix. Then the following statements are equivalent:*

(i) The matrix $\mathbf{A}$ admits a left inverse, that is, a matrix $\mathbf{B} \in M_{n,n}(\mathbb{K})$ such that $\mathbf{BA} = \mathbf{1}_n$;
(ii) The matrix $\mathbf{A}$ admits a right inverse, that is, a matrix $\mathbf{B} \in M_{n,n}(\mathbb{K})$ such that $\mathbf{AB} = \mathbf{1}_n$;
(iii) The matrix $\mathbf{A}$ is invertible.

Proof By the definition of the invertability of a matrix, (iii) implies both (i) and (ii).

(i) $\Rightarrow$ (iii) Since $\mathbf{BA} = \mathbf{1}_n$ we have $f_{\mathbf{B}} \circ f_{\mathbf{A}} = f_{\mathbf{1}_n} = \mathrm{Id}_{\mathbb{K}^n}$ by Theorem 2.21 and hence $f_{\mathbf{B}}$ is a left inverse for $f_{\mathbf{A}}$. Therefore, by the above exercise, $f_{\mathbf{A}}$ is injective. Corollary 3.77 implies that $f_{\mathbf{A}}$ is also bijective. Denoting the ordered standard basis of $\mathbb{K}^n$ by $\mathbf{e}$, we have $\mathbf{M}(f_{\mathbf{A}}, \mathbf{e}, \mathbf{e}) = \mathbf{A}$ and hence Proposition 3.101 implies that $\mathbf{A}$ is invertible.

(ii) $\Rightarrow$ (iii) is completely analogous to (i) $\Rightarrow$ (iii). □

3.7.1 Change of basis

It is natural to ask how the choice of bases affects the matrix representation of a linear map.

Definition 3.103 (Change of basis matrix) Let V be a finite dimensional $\mathbb{K}$-vector space and $\mathbf{b}, \mathbf{b}'$ be ordered bases of V with corresponding linear coordinate systems $\boldsymbol{\beta}, \boldsymbol{\beta}'$. The *change of basis matrix from* $\mathbf{b}$ *to* $\mathbf{b}'$ is the matrix $\mathbf{C} \in M_{n,n}(\mathbb{K})$ satisfying

$$f_{\mathbf{C}} = \boldsymbol{\beta}' \circ \boldsymbol{\beta}^{-1}$$

We will write $\mathbf{C}(\mathbf{b}, \mathbf{b}')$ for the change of basis matrix from $\mathbf{b}$ to $\mathbf{b}'$.

Remark 3.104 Notice that by definition

$$\mathbf{C}(\mathbf{b}, \mathbf{b}') = \mathbf{M}(\mathrm{Id}_V, \mathbf{b}, \mathbf{b}').$$

Since the identity map $\mathrm{Id}_V : V \to V$ is bijective with inverse $(\mathrm{Id}_V)^{-1} = \mathrm{Id}_V$, Proposition 3.101 implies that the change of basis matrix $\mathbf{C}(\mathbf{b}, \mathbf{b}')$ is invertible with inverse

$$\mathbf{C}(\mathbf{b}, \mathbf{b}')^{-1} = \mathbf{C}(\mathbf{b}', \mathbf{b}).$$

Example 3.105 Let $V = \mathbb{R}^2$ and $\mathbf{e} = (\vec{e}_1, \vec{e}_2)$ be the ordered standard basis and $\mathbf{b} = (\vec{v}_1, \vec{v}_2) = (\vec{e}_1 + \vec{e}_2, \vec{e}_2 - \vec{e}_1)$ another ordered basis. According to the recipe mentioned in Example 3.94, if we want to compute $\mathbf{C}(\mathbf{e}, \mathbf{b})$ we simply need to write each vector of $\mathbf{e}$ as a linear combination of the elements of $\mathbf{b}$. The transpose of the resulting coefficient matrix is then $\mathbf{C}(\mathbf{e}, \mathbf{b})$. We obtain

$$\vec{e}_1 = \frac{1}{2}\vec{v}_1 - \frac{1}{2}\vec{v}_2,$$
$$\vec{e}_2 = \frac{1}{2}\vec{v}_1 + \frac{1}{2}\vec{v}_2,$$

so that

$$\mathbf{C}(\mathbf{e}, \mathbf{b}) = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Reversing the role of $\mathbf{e}$ and $\mathbf{b}$ gives $\mathbf{C}(\mathbf{b}, \mathbf{e})$

$$\vec{v}_1 = 1\vec{e}_1 + 1\vec{e}_2,$$
$$\vec{v}_2 = -1\vec{e}_1 + 1\vec{e}_2,$$

so that

$$\mathbf{C}(\mathbf{b}, \mathbf{e}) = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

Notice that indeed we have

$$\mathbf{C}(\mathbf{e},\mathbf{b})\mathbf{C}(\mathbf{b},\mathbf{e}) = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

so that $\mathbf{C}(\mathbf{e},\mathbf{b})^{-1} = \mathbf{C}(\mathbf{b},\mathbf{e})$.

Theorem 3.106 *Let V, W be finite dimensional $\mathbb{K}$-vector spaces and $\mathbf{b}, \mathbf{b}'$ ordered bases of V and $\mathbf{c}, \mathbf{c}'$ ordered bases of W. Let $g : V \to W$ be a linear map. Then we have*

$$\mathbf{M}(g,\mathbf{b}',\mathbf{c}') = \mathbf{C}(\mathbf{c},\mathbf{c}')\mathbf{M}(g,\mathbf{b},\mathbf{c})\mathbf{C}(\mathbf{b}',\mathbf{b})$$

In particular, for a linear map $g : V \to V$ we have

$$\mathbf{M}(g,\mathbf{b}',\mathbf{b}') = \mathbf{C}\,\mathbf{M}(g,\mathbf{b},\mathbf{b})\,\mathbf{C}^{-1},$$

where we write $\mathbf{C} = \mathbf{C}(\mathbf{b},\mathbf{b}')$.

Proof We write $\mathbf{A} = \mathbf{M}(g,\mathbf{b},\mathbf{c})$ and $\mathbf{B} = \mathbf{M}(g,\mathbf{b}',\mathbf{c}')$ and $\mathbf{C} = \mathbf{C}(\mathbf{b},\mathbf{b}')$ and $\mathbf{D} = \mathbf{C}(\mathbf{c},\mathbf{c}')$. By Remark 3.104 we have $\mathbf{C}^{-1} = \mathbf{C}(\mathbf{b}',\mathbf{b})$, hence applying Proposition 2.20 and Theorem 2.21 and Corollary 2.22, we need to show that

$$f_{\mathbf{B}} = f_{\mathbf{D}} \circ f_{\mathbf{A}} \circ f_{\mathbf{C}^{-1}}.$$

By Definition 3.91 we have

$$f_{\mathbf{A}} = \gamma \circ g \circ \beta^{-1},$$
$$f_{\mathbf{B}} = \gamma' \circ g \circ (\beta')^{-1}$$

and by Definition 3.103 we have

$$f_{\mathbf{C}^{-1}} = \beta \circ (\beta')^{-1},$$
$$f_{\mathbf{D}} = \gamma' \circ \gamma^{-1}.$$

Hence we obtain

$$f_{\mathbf{D}} \circ f_{\mathbf{A}} \circ f_{\mathbf{C}^{-1}} = \gamma' \circ \gamma^{-1} \circ \gamma \circ g \circ \beta^{-1} \circ \beta \circ (\beta')^{-1} = \gamma' \circ g \circ (\beta')^{-1} = f_{\mathbf{B}},$$

as claimed. The second statement follows again by applying Remark 3.104. □

Example 3.107 (Example 3.96 and Example 3.105 continued) Let $\mathbf{e} = (\vec{e}_1, \vec{e}_2)$ denote the ordered standard basis of $\mathbb{R}^2$ and

$$\mathbf{A} = \begin{pmatrix} 5 & 1 \\ 1 & 5 \end{pmatrix} = \mathbf{M}(f_{\mathbf{A}},\mathbf{e},\mathbf{e}).$$

Let $\mathbf{b} = (\vec{e}_1 + \vec{e}_2, \vec{e}_2 - \vec{e}_1)$. We computed that

$$\mathbf{M}(f_{\mathbf{A}},\mathbf{b},\mathbf{b}) = \begin{pmatrix} 6 & 0 \\ 0 & 4 \end{pmatrix}$$

as well as

$$\mathbf{C}(\mathbf{e},\mathbf{b}) = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \quad \text{and} \quad \mathbf{C}(\mathbf{b},\mathbf{e}) = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

According to Theorem 3.106 we must have

$$\mathbf{M}(f_{\mathbf{A}},\mathbf{b},\mathbf{b}) = \mathbf{C}(\mathbf{e},\mathbf{b})\mathbf{M}(f_{\mathbf{A}},\mathbf{e},\mathbf{e})\mathbf{C}(\mathbf{b},\mathbf{e})$$

and indeed

$$\begin{pmatrix} 6 & 0 \\ 0 & 4 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}\begin{pmatrix} 5 & 1 \\ 1 & 5 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

Finally, we observe that every invertible matrix can be realised as a change of basis matrix:

Lemma 3.108 *Let V be a finite dimensional $\mathbb{K}$-vector space, $\mathbf{b} = (v_1, \dots, v_n)$ an ordered basis of V and $\mathbf{C} \in M_{n,n}(\mathbb{K})$ an invertible $n \times n$-matrix. Define $v'_j = \sum_{i=1}^{n} C_{ij} v_i$ for $1 \leqslant i \leqslant n$. Then $\mathbf{b}' = (v'_1, \dots, v'_n)$ is an ordered basis of V and* $\mathbf{C}(\mathbf{b}', \mathbf{b}) = \mathbf{C}$.

Proof It is sufficient to prove that the vectors $\{v'_1, \dots, v'_n\}$ are linearly independent. Indeed, if they are linearly independent, then they span a subspace U of dimension n and Proposition 3.74 implies that $U = V$, so that $\mathbf{b}'$ is an ordered basis of V. Suppose we have scalars $s_1, \dots, s_n$ such that

$$0_V = \sum_{j=1}^{n} s_j v'_j = \sum_{j=1}^{n} \sum_{i=1}^{n} s_j C_{ij} v_i = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} C_{ij} s_j \right) v_i.$$

Since $\{v_1, \dots, v_n\}$ is a basis of V we must have $\sum_{j=1}^{n} C_{ij} s_j = 0$ for all $i = 1, \dots, n$. In matrix notation this is equivalent to the conditon $\mathbf{C}\vec{s} = 0_{\mathbb{K}^n}$, where $\vec{s} = (s_i)_{1 \leqslant i \leqslant n}$. Since $\mathbf{C}$ is invertible, we can multiply this last equation from the left with $\mathbf{C}^{-1}$ to obtain $\mathbf{C}^{-1}\mathbf{C}\vec{s} = \mathbf{C}^{-1} 0_{\mathbb{K}^n}$ which is equivalent to $\vec{s} = 0_{\mathbb{K}^n}$. It follows that $\mathbf{b}'$ is an ordered basis of V. By definition we have $\mathbf{C}(\mathbf{b}', \mathbf{b}) = \mathbf{C}$. □

Exercises

Exercise 3.109 Let $\mathrm{Id}_V : V \to V$ denote the identity mapping of the finite dimensional $\mathbb{K}$-vector space V and let $\mathbf{b} = (v_1, \dots, v_n)$ be any ordered basis of V. Show that $\mathbf{M}(\mathrm{Id}_V, \mathbf{b}, \mathbf{b}) = \mathbf{1}_n$.

Exercise 3.110 Show that $f : \mathcal{X} \to \mathcal{Y}$ admits a left inverse if and only if f is injective and that $f : \mathcal{X} \to \mathcal{Y}$ admits a right inverse if and only if f is surjective.

Exercise 3.111 Let V be a finite dimensional $\mathbb{K}$-vector space and $\mathbf{b}, \mathbf{b}'$ be ordered bases of V. Show that for all $v \in V$ we have

$$\beta'(v) = \mathbf{C}(\mathbf{b}, \mathbf{b}')\beta(v).$$

CHAPTER 4

Applications of Gaussian elimination

4.1 Gaussian elimination

WEEK 7

In the Algorithmics module M01 you learned how to use Gaussian elimination to solve a system of equations of the form

$$\mathbf{A}\vec{x} = \vec{b} \tag{4.1}$$

for some given matrix $\mathbf{A} \in M_{m,n}(\mathbb{K})$, vector $\vec{b} \in \mathbb{K}^m$ and unknown $\vec{x} \in \mathbb{K}^n$. Many concrete problems in Linear Algebra lead to systems of the form (4.1). A few sample problems that can be solved with Gaussian elimination are discussed below.

Solving equation of the type (4.1) hinges on the elementary observation that a vector $\vec{x} \in \mathbb{K}^n$ solves $\mathbf{A}\vec{x} = \vec{b}$ if and only if it solves $\mathbf{B}\mathbf{A}\vec{x} = \mathbf{B}\vec{b}$, where $\mathbf{B} \in M_{m,m}(\mathbb{K})$ is any invertible m-by-m matrix.

In the Gaussian elimination algorithm, the matrix $\mathbf{B}$ is chosen among three types of matrices:

Definition 4.1 (Elementary matrices — Video) Let $m \in \mathbb{N}$. The *elementary matrices of size m* are the square matrices

$$\begin{aligned}
\mathbf{L}_{k,l}(s) &= \mathbf{1}_m + s\mathbf{E}_{k,l},\\
\mathbf{D}_k(s) &= \mathbf{1}_m + (s-1)\mathbf{E}_{k,k},\\
\mathbf{P}_{k,l} &= \mathbf{1}_m - \mathbf{E}_{k,k} - \mathbf{E}_{l,l} + \mathbf{E}_{k,l} + \mathbf{E}_{l,k},
\end{aligned}$$

where $1 \leqslant k, l \leqslant m$ with $k \neq l$, $\mathbf{E}_{k,l} \in M_{m,m}(\mathbb{K})$ and $s \in \mathbb{K}$ with $s \neq 0$.

Example 4.2 For $m = 4$ we have for instance

$$\mathbf{L}_{2,3}(s) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & s & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad \mathbf{D}_4(s) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & s \end{pmatrix}$$

and

$$\mathbf{P}_{2,4} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

As an exercise in matrix multiplication, we compute the effect of left multiplication with elementary matrices.

For $\mathbf{A} = (A_{ij})_{1\leqslant i,j\leqslant m} \in M_{m,n}(\mathbb{K})$, we obtain

$$[\mathbf{L}_{k,l}(s)\mathbf{A}]_{ij} = \sum_{r=1}^{m} (\delta_{ir} + s\delta_{ik}\delta_{lr})\, A_{rj} = A_{ij} + s\delta_{ik}A_{lj} = \begin{cases} A_{ij} + sA_{lj} & i = k \\ A_{ij} & i \neq k \end{cases},$$

where we use that $[\mathbf{1}_m]_{ir} = \delta_{ir}$ and $[\mathbf{E}_{k,l}]_{ir} = \delta_{ik}\delta_{lr}$. Therefore, multiplying the matrix $\mathbf{A}$ with $\mathbf{L}_{k,l}(s)$ from the left, adds s times the l-th row of $\mathbf{A}$ to the k-th row of $\mathbf{A}$ and leaves $\mathbf{A}$ unchanged otherwise.

Likewise, we obtain

$$[\mathbf{D}_k(s)\mathbf{A}]_{ij} = \sum_{r=1}^{m} (\delta_{ir} + (s-1)\delta_{ik}\delta_{kr})\, A_{rj} = \begin{cases} sA_{ij} & i = k \\ A_{ij} & i \neq k \end{cases}.$$

Therefore, multiplying the matrix $\mathbf{A}$ with $\mathbf{D}_k(s)$ from the left, multiplies the k-th row of $\mathbf{A}$ with s and leaves $\mathbf{A}$ unchanged otherwise.

Finally,

$$\begin{aligned}
[\mathbf{P}_{k,l}\mathbf{A}]_{ij} &= \sum_{r=1}^{m} (\delta_{ir} - \delta_{ik}\delta_{kr} - \delta_{il}\delta_{lr} + \delta_{ik}\delta_{lr} + \delta_{il}\delta_{rk})A_{rj} \\
&= A_{ij} - \delta_{ik}A_{kj} - \delta_{il}A_{lj} + \delta_{ik}A_{lj} + \delta_{il}A_{kj} \\
&= A_{ij} + \delta_{ik}\,(A_{lj} - A_{kj}) + \delta_{il}\,(A_{kj} - A_{lj}) = \begin{cases} A_{lj} & i = k \\ A_{kj} & i = l \\ A_{ij} & i \neq k, i \neq l \end{cases}.
\end{aligned}$$

Therefore, multiplying the matrix $\mathbf{A}$ with $\mathbf{P}_{k,l}$ from the left, swaps the k-th row of $\mathbf{A}$ with the l-th row of $\mathbf{A}$ and leaves $\mathbf{A}$ unchanged otherwise.

These calculations immediately imply:

Proposition 4.3 *The elementary matrices are invertible with*

$$\mathbf{L}_{k,l}(s)^{-1} = \mathbf{L}_{k,l}(-s) \quad \textit{and} \quad \mathbf{D}_k(s)^{-1} = \mathbf{D}_k(1/s) \quad \textit{and} \quad (\mathbf{P}_{k,l})^{-1} = \mathbf{P}_{k,l}.$$

The sceptical reader may also verify this fact by direct computation with the help of the following lemma:

Lemma 4.4 *Let $m \in \mathbb{N}$. For $1 \leqslant k, l, p, q \leqslant m$, we have*

$$\mathbf{E}_{k,l}\mathbf{E}_{p,q} = \begin{cases} \mathbf{E}_{k,q} & p = l \\ \mathbf{0}_{m,m} & p \neq l \end{cases}$$

Proof By definition, we have

$$\mathbf{E}_{k,l}\mathbf{E}_{p,q} = \left(\sum_{r=1}^{m} \delta_{ik}\delta_{lr}\delta_{rp}\delta_{qj}\right)_{1\leqslant i,j\leqslant m} = \delta_{lp}\,(\delta_{ik}\delta_{qj})_{1\leqslant i,j\leqslant m} = \begin{cases} \mathbf{E}_{k,q} & p = l \\ \mathbf{0}_{m,m} & p \neq l \end{cases}.$$

□

For each row in a matrix, if the row does not consist of zeros only, then the leftmost nonzero entry is called the leading coefficient of that row.

Definition 4.5 (Row echelon form) A matrix $\mathbf{A} \in M_{m,n}(\mathbb{K})$ is said to be in *row echelon form* (REF) if

- all rows consisting of only zeros are at the bottom;
- the leading coefficient of a nonzero row is always strictly to the right of the leading coefficient of the row above it.

The matrix $\mathbf{A}$ is said to be in *reduced row echelon form* (rREF) if furthermore

- all of the leading coefficients are equal to 1;
- in every column containing a leading coefficient, all of the other entries in that column are zero.

Gaussian elimination from the Algorithmics module M01 implies the following statement:

Theorem 4.6 (Gauss–Jordan elimination) *Let $\mathbf{A} \in M_{m,n}(\mathbb{K})$ then there exists $N \in \mathbb{N}$ and an N-tuple of elementary matrices $(\mathbf{B}_1, \ldots, \mathbf{B}_N)$ such that the matrix $\mathbf{B}_N\mathbf{B}_{N-1}\cdots\mathbf{B}_2\mathbf{B}_1\mathbf{A}$ is in reduced row echelon form.*

Proof Applying Gaussian elimination implies the existence of $\hat{N} \in \mathbb{N}$ and elementary matrices $\mathbf{B}_1, \ldots, \mathbf{B}_{\hat{N}}$ so that $\mathbf{B}_{\hat{N}}\mathbf{B}_{\hat{N}-1}\cdots\mathbf{B}_2\mathbf{B}_1\mathbf{A}$ is REF. After possibly further multiplying this matrix from the left with elementary matrices of the type $\mathbf{D}_k(s)$, we can assume that all leading coefficients are 1. By choosing suitable left multiplications with matrices of the type $\mathbf{L}_{k,l}(s)$, we find a natural number $N \geqslant \hat{N}$ and elementary matrices $(\mathbf{B}_1, \ldots, \mathbf{B}_N)$ so that $\mathbf{B}_N\mathbf{B}_{N-1}\cdots\mathbf{B}_2\mathbf{B}_1\mathbf{A}$ is in reduced row echelon form. □

4.2 Applications

4.2.1 Compute the inverse of a matrix

An algorithm using Gaussian elimination for computing the inverse of an invertible matrix relies on the following fact:

Proposition 4.7 *Let $\mathbf{A} \in M_{n,n}(\mathbb{K})$ be a square matrix. Then the following statements are equivalent:*

(i) $\mathbf{A}$ is invertible;
(ii) the row vectors of $\mathbf{A}$ are linearly independent;
(iii) the column vectors of $\mathbf{A}$ are linearly independent.

Proof Part of an exercise sheet. □

Suppose the matrix $\mathbf{A} \in M_{n,n}(\mathbb{K})$ is invertible. Applying Gauss–Jordan elimination to $\mathbf{A}$, we cannot encounter a zero row, since the occurrence of a zero row corresponds to a non-trivial linear combination of row vectors which gives the zero vector. This is excluded by the above proposition. Having no zero row vectors, the Gauss–Jordan elimination applied to $\mathbf{A}$ must give the identity matrix $\mathbf{1}_n$. Thus we can find a sequence of elementary matrices $\mathbf{B}_1, \ldots, \mathbf{B}_N$, $N \in \mathbb{N}$, so that

$$\mathbf{1}_n = \mathbf{B}_N\mathbf{B}_{N-1}\cdots\mathbf{B}_2\mathbf{B}_1\mathbf{A}.$$

In other words, $\mathbf{B}_N\mathbf{B}_{N-1}\cdots\mathbf{B}_2\mathbf{B}_1$ is the inverse of $\mathbf{A}$. This gives the following recipe for computing the inverse of $\mathbf{A}$:

We write the matrix $\mathbf{A}$ and $\mathbf{1}_n$ next to each other, say $\mathbf{A}$ on the left and $\mathbf{1}_n$ on the right. We then perform Gauss–Jordan elimination on $\mathbf{A}$. At each step, we also perform the Gauss–Jordan elimination step to the matrix on the right. Once Gauss–Jordan elimination terminates, we thus obtain $\mathbf{B}_N\mathbf{B}_{N-1}\cdots\mathbf{B}_2\mathbf{B}_1\mathbf{A}$ on the left and $\mathbf{B}_N\mathbf{B}_{N-1}\cdots\mathbf{B}_2\mathbf{B}_1\mathbf{1}_n$ on the right. But since $\mathbf{B}_N\mathbf{B}_{N-1}\cdots\mathbf{B}_2\mathbf{B}_1\mathbf{1}_n = \mathbf{B}_N\mathbf{B}_{N-1}\cdots\mathbf{B}_2\mathbf{B}_1$ (notice the absence of $\mathbf{1}_n$ after the equality sign), the right hand side is the inverse of $\mathbf{A}$.

Example 4.8 (Inverse of a matrix — Video) We want to compute the inverse of

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

Write

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

Adding 3-times the first row to the second row gives

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

Dividing the second row by -2 gives

$$\left(\begin{array}{cc|cc} 1 & -2 & 1 & 0 \\ 0 & 1 & -\frac{3}{2} & -\frac{1}{2} \end{array}\right).$$

Finally, adding the second row twice to the first row gives

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

so that

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

4.2.2 Compute a basis of a subspace

Gaussian elimination can also be used to compute a basis for a vector subspace U of a finite dimensional $\mathbb{K}$-vector space V. We assume that $U = \operatorname{span}\{v_1, \ldots, v_k\}$ for some vectors $v_i \in V, 1 \leqslant i \leqslant k$. We assume that $\dim U \geqslant 1$ so that not all vectors are the zero vector.

We first consider the special case where V is the space $\mathbb{K}_n$ of row vectors of length n and with entries in $\mathbb{K}$. Recall that we denote the row vectors by small Greek letters. We write $\mathbb{K}_n^m$ for the m-fold Cartesian product $(\mathbb{K}_n)^m$ of $\mathbb{K}_n$. Clearly, we have a bijective mapping

$$\Omega : \mathbb{K}_n^m \to M_{m,n}(\mathbb{K}), \quad (\vec{\nu}_1, \ldots, \vec{\nu}_m) \mapsto \begin{pmatrix} \vec{\nu}_1 \\ \vdots \\ \vec{\nu}_m \end{pmatrix}$$

which simply writes the row vectors $(\vec{\nu}_1, \ldots, \vec{\nu}_m)$ into a matrix with the k-th row vector from the m-tuple of row vectors becoming the k-th row of the matrix.

Example 4.9

$$\Omega\left(\begin{pmatrix} 1 & 2 & 3 \end{pmatrix}, \begin{pmatrix} 4 & 5 & 6 \end{pmatrix}\right) = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}.$$

We have

$$\mathbf{L}_{k,l}(s)\Omega(\vec{\nu}_1,\dots,\vec{\nu}_m) = \Omega\left(\vec{\nu}_1,\dots,\vec{\nu}_{k-1},\vec{\nu}_k + s\vec{\nu}_l,\vec{\nu}_{k+1},\dots,\vec{\nu}_m\right),$$
$$\mathbf{D}_k(s)\Omega(\vec{\nu}_1,\dots,\vec{\nu}_m) = \Omega\left(\vec{\nu}_1,\dots,\vec{\nu}_{k-1},s\vec{\nu}_k,\vec{\nu}_{k+1},\dots,\vec{\nu}_m\right),$$
$$\mathbf{P}_{k,l}\Omega(\vec{\nu}_1,\dots,\vec{\nu}_m) = \Omega\left(\vec{\nu}_1,\dots,\vec{\nu}_{k-1},\vec{\nu}_l,\vec{\nu}_{k+1},\dots,\vec{\nu}_{l-1},\vec{\nu}_k,\vec{\nu}_{l+1},\dots,\vec{\nu}_m\right).$$

Notice that all these operations do not change the span of the vectors $\vec{\nu}_1,\dots,\vec{\nu}_m$. More precisely, if $(\vec{\nu}_1,\dots,\vec{\nu}_m)$ is an n-tuple of row vectors and if $\Omega\left(\vec{\omega}_1,\dots,\vec{\omega}_m\right) = \mathbf{B}\Omega(\vec{\nu}_1,\dots,\vec{\nu}_m)$ for some elementary matrix $\mathbf{B}$, then

$$\operatorname{span}\{\vec{\nu}_1,\dots,\vec{\nu}_m\} = \operatorname{span}\{\vec{\omega}_1,\dots,\vec{\omega}_m\}.$$

Applying Gaussian elimination to the matrix $\Omega(\vec{\nu}_1,\dots,\vec{\nu}_m)$ gives a list of elementary matrices $\mathbf{B}_1,\dots,\mathbf{B}_N$ such that

$$\mathbf{B}_N\mathbf{B}_{N-1}\cdots\mathbf{B}_2\mathbf{B}_1\Omega(\vec{\nu}_1,\dots,\vec{\nu}_m) = \Omega(\vec{\omega}_1,\dots,\vec{\omega}_r,0_{\mathbb{K}_n},\dots,0_{\mathbb{K}_n})$$

where $1 \leqslant r \leqslant m$ and $0_{\mathbb{K}_n}$ denotes the zero vector in $\mathbb{K}_n$. By construction, the matrix $\mathbf{A} = \Omega(\vec{\omega}_1,\dots,\vec{\omega}_r,0_{\mathbb{K}_n},\dots,0_{\mathbb{K}_n})$ is REF. Since the leading coefficient of $\vec{\omega}_i$ is always strictly to the right of the leading coefficient of $\vec{\omega}_{i-1}$, it follows that the vectors $\vec{\omega}_1,\dots,\vec{\omega}_r$ are linearly independent. Therefore, a basis of $\operatorname{span}\{\vec{\nu}_1,\dots,\vec{\nu}_m\}$ is given by $\{\vec{\omega}_1,\dots,\vec{\omega}_r\}$.

The general case can be treated with the help of the following facts:

Proposition 4.10 *Let V, W be finite dimensional $\mathbb{K}$-vector spaces and $\Phi : V \to W$ an isomorphism. Then $\mathcal{S} \subset V$ is a basis of V if and only if $\Phi(\mathcal{S})$ is a basis of W.*

Proof $\Rightarrow$ Since $\mathcal{S}$ is a basis, the set $\mathcal{S}$ is linearly independent and since Φ is injective, so is $\Phi(\mathcal{S})$ by Lemma 3.56. Since $\mathcal{S}$ is a basis, $\mathcal{S}$ is a generating set and since Φ is surjective, the subset $\Phi(\mathcal{S}) \subset W$ is a generating set for W by Lemma 3.46.

$\Leftarrow$ We apply the above implication to $\Phi^{-1} : W \to V$ and the basis $\Phi(\mathcal{S}) \subset W$. □

Corollary 4.11 *Let $\hat{V}, \hat{W}$ be finite dimensional $\mathbb{K}$-vector spaces, $\Theta : \hat{V} \to \hat{W}$ an isomorphism and $U \subset \hat{V}$ a vector subspace. Then $\mathcal{S} \subset U$ is a basis of U if and only if $\Theta(\mathcal{S})$ is a basis of $\Theta(U)$.*

Proof Apply Proposition 4.10 to the vector space $V = U$, the vector space $W = \Theta(U)$ and the isomorphism $\Phi = \Theta|_U : V \to W$. □

We now describe a recipe to treat the general case of a subset $U = \operatorname{span}\{v_1,\dots,v_m\}$ of a finite dimensional $\mathbb{K}$-vector space V:

(i) Fix an isomorphism $\Phi : V \to \mathbb{K}_n$ and write $\vec{\nu}_i = \Phi(v_i)$ for $1 \leqslant i \leqslant m$.
(ii) Apply Gaussian elimination to the matrix $\Omega(\vec{\nu}_1,\dots,\vec{\nu}_m)$ to obtain a set of new vectors $(\vec{\omega}_1,\dots,\vec{\omega}_r,0_{\mathbb{K}_n},\dots,0_{\mathbb{K}_n})$ for some $r \in \mathbb{N}$.
(iii) Apply the inverse isomorphism Φ^{-1} to the obtained list of vectors. This gives the desired basis $\{\Phi^{-1}(\vec{\omega}_1),\dots,\Phi^{-1}(\vec{\omega}_r)\}$ of U.

Example 4.12 (Basis of a subspace — Video) Let $V = \mathrm{P}_3(\mathbb{R})$ so that $\dim(V) = 4$ and

$$U = \operatorname{span}\{x^3+2x^2-x, 4x^3+8x^2-4x-3, x^2+3x+4, 2x^3+5x+x+4\}.$$

We want to compute a basis of U. We choose the isomorphism $\Phi : V \to \mathbb{R}_4$ defined by

$$\Phi(a_3x^3 + a_2x^2 + a_1x + a_0) = \begin{pmatrix} a_3 & a_2 & a_1 & a_0 \end{pmatrix}.$$

We thus have $\vec{\nu}_1 = \begin{pmatrix} 1 & 2 & -1 & 0 \end{pmatrix}$, $\vec{\nu}_2 = \begin{pmatrix} 4 & 8 & -4 & -3 \end{pmatrix}$, $\vec{\nu}_3 = \begin{pmatrix} 0 & 1 & 3 & 4 \end{pmatrix}$ and $\vec{\nu}_4 = \begin{pmatrix} 2 & 5 & 1 & 4 \end{pmatrix}$.
Applying Gaussian elimination to the matrix

$$\Omega(\vec{\nu}_1, \vec{\nu}_2, \vec{\nu}_3, \vec{\nu}_4) = \begin{pmatrix} 1 & 2 & -1 & 0 \\ 4 & 8 & -4 & -3 \\ 0 & 1 & 3 & 4 \\ 2 & 5 & 1 & 4 \end{pmatrix}$$

yields

$$\begin{pmatrix} 1 & 0 & -7 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Here we applied Gauss-Jordan elimination, but Gaussian elimination is good enough. This gives the vectors $\vec{\omega}_1 = \begin{pmatrix} 1 & 0 & -7 & 0 \end{pmatrix}$, $\vec{\omega}_2 = \begin{pmatrix} 0 & 1 & 3 & 0 \end{pmatrix}$, $\vec{\omega}_3 = \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix}$.
Our basis of U is thus

$$\{\Phi^{-1}(\vec{\omega}_1), \Phi^{-1}(\vec{\omega}_2), \Phi^{-1}(\vec{\omega}_3)\} = \{x^3 - 7x, x^2 + 3x, 1\},$$

where we use that

$$\Phi^{-1}\left(\begin{pmatrix} a_3 & a_2 & a_1 & a_0 \end{pmatrix}\right) = a_3x^3 + a_2x^2 + a_1x + a_0.$$

4.2.3 Compute the image and rank of a linear map

Let V, W be finite dimensional $\mathbb{K}$-vector spaces and $f : V \to W$ a linear map. By computing the image of a linear map f, we mean computing a basis of $\operatorname{Im}(f)$.

In order to compute a basis for $\operatorname{Im}(f)$ we use the following lemma:

Lemma 4.13 *Let V, W be finite dimensional $\mathbb{K}$-vector spaces and $f : V \to W$ a linear map. If $\{v_1, \ldots, v_n\}$ is a basis of V, then*

$$\operatorname{Im}(f) = \operatorname{span}\{f(v_1), \ldots, f(v_n)\}.$$

Proof Let $w \in \operatorname{Im}(f)$ so that $w = f(v)$ for some $v \in V$. We have scalars s_i for $1 \leqslant i \leqslant n$ so that $v = \sum_{i=1}^n s_i v_i$. We obtain

$$w = f(v) = f\left(\sum_{i=1}^n s_i v_i\right) = \sum_{i=1}^n s_i f(v_i)$$

so that w is a linear combination of the vectors $\{f(v_1), \ldots, f(v_n)\}$. On the other hand, a linear combination of the vectors $f(v_i) \in \operatorname{Im}(f)$ lies in the image of f as well, since $\operatorname{Im}(f)$ is a vector subspace. Hence we have $\operatorname{Im}(f) = \operatorname{span}\{f(v_1), \ldots, f(v_n)\}$, as claimed. □

Knowing that $\operatorname{Im}(f) = \operatorname{span}\{f(v_1), \ldots, f(v_n)\}$ we can apply the recipe from Section 4.2.2 to $U = \operatorname{span}\{f(v_1), \ldots, f(v_n)\}$. By definition, the number of basis vectors for $\operatorname{Im}(f)$ is the rank of f.

Example 4.14 Let

$$\mathbf{A} = \begin{pmatrix} 1 & -2 & 0 & 4 \\ 3 & 1 & 1 & 0 \\ -1 & -5 & -1 & 8 \\ 3 & 8 & 2 & -12 \end{pmatrix}$$

Compute a basis for the image of $f_{\mathbf{A}} : \mathbb{R}^4 \to \mathbb{R}^4$ and the rank of $f_{\mathbf{A}}$. By Lemma 4.13 we have

$$U = \operatorname{Im}(f_{\mathbf{A}}) = \operatorname{span}\{\mathbf{A}\vec{e}_1, \mathbf{A}\vec{e}_2, \mathbf{A}\vec{e}_3, \mathbf{A}\vec{e}_4\} = \operatorname{span}\{\vec{a}_1, \vec{a}_2, \vec{a}_3, \vec{a}_4\},$$

where $\{\vec{e}_i\}_{1\leqslant i\leqslant 4}$ denotes the standard basis of $\mathbb{R}^4$ and $\{\vec{a}_i\}_{1\leqslant i\leqslant 4}$ the column vectors of $\mathbf{A}$. Comparing with the general setup described above, we are in the case where $V = \mathbb{R}^4$ and $v_i = \mathbf{A}\vec{e}_i$ for $i = 1, 2, 3, 4$.

(i) For the isomorphism $\Phi : V = \mathbb{R}^4 \to \mathbb{R}_4$ we usually choose the transpose (but any other isomorphism would work too). We thus have $\vec{\nu}_1 = \begin{pmatrix} 1 & 3 & -1 & 3 \end{pmatrix}$, $\vec{\nu}_2 = \begin{pmatrix} -2 & 1 & -5 & 8 \end{pmatrix}$, $\vec{\nu}_3 = \begin{pmatrix} 0 & 1 & -1 & 2 \end{pmatrix}$ and $\vec{\nu}_4 = \begin{pmatrix} 4 & 0 & 8 & -12 \end{pmatrix}$.

(ii) Applying Gaussian elimination to the matrix

$$\Omega(\vec{\nu}_1, \vec{\nu}_2, \vec{\nu}_3, \vec{\nu}_4) = \mathbf{A}^T = \begin{pmatrix} 1 & 3 & -1 & 3 \\ -2 & 1 & -5 & 8 \\ 0 & 1 & -1 & 2 \\ 4 & 0 & 8 & -12 \end{pmatrix}$$

yields

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

Here again, we applied Gauss-Jordan elimination, but Gaussian elimination is good enough. This gives the vectors $\vec{\omega}_1 = \begin{pmatrix} 1 & 0 & 2 & -3 \end{pmatrix}$, $\vec{\omega}_2 = \begin{pmatrix} 0 & 1 & -1 & 2 \end{pmatrix}$.

(iii) Our basis of $\operatorname{Im}(f)$ is thus

$$\{\Phi^{-1}(\vec{\omega}_1), \Phi^{-1}(\vec{\omega}_2)\} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 2 \\ -3 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \\ 2 \end{pmatrix} \right\},$$

where we use that the transpose is its own inverse. We also conclude that $\operatorname{rank}(f_{\mathbf{A}}) = 2$.

Remark 4.15 In the special case where we want to compute a basis for the image of $f_{\mathbf{A}}$ for some matrix $\mathbf{A}$, the recipe thus reduces to the following steps. Take the transpose of $\mathbf{A}$, perform Gauss elimination, take the transpose again, write down the nonzero column vectors. This gives the desired basis.

4.2.4 Compute the kernel and nullity of a linear map

In order to find a recipe for computing the kernel and nullity of a linear map, we first start with a related problem. Let $\mathbf{A} \in M_{n,m}(\mathbb{K})$ be an $n \times m$-matrix and

$$U = \left\{ \vec{\xi} \in \mathbb{K}_n \,|\, \vec{\xi}\mathbf{A} = 0_{\mathbb{K}_m} \right\},$$

where $\vec{\xi}\mathbf{A}$ is defined via matrix multiplication of the row vector $\vec{\xi} \in \mathbb{K}_n = M_{1,n}(\mathbb{K})$ and the matrix $\mathbf{A} \in M_{n,m}(\mathbb{K})$. Notice that $0_{\mathbb{K}_n} \in U$ and if $\vec{\xi}_1, \vec{\xi}_2 \in U$, then $s_1\vec{\xi}_1 + s_2\vec{\xi}_2 \in U$ for all $s_1, s_2 \in \mathbb{K}$. By Definition 3.21, it follows that U is a vector subspace of $\mathbb{K}_n$. We want to compute a basis for U. Applying Gauss elimination to the matrix $\mathbf{A}$, we obtain $r \in \mathbb{N}$ and elementary matrices $\mathbf{B}_1, \dots, \mathbf{B}_N$ so that

$$\mathbf{B}_N \cdots \mathbf{B}_1\mathbf{A} = \Omega\left(\vec{\omega}_1, \dots, \vec{\omega}_r, 0_{\mathbb{K}_m}, \dots, 0_{\mathbb{K}_m}\right)$$

for some linearly independent row vectors $(\vec{\omega}_1, \dots, \vec{\omega}_r) \in \mathbb{K}_m$. Since the matrix $\mathbf{B}_N \cdots \mathbf{B}_1$ is invertible, we also obtain a basis $\{\vec{\xi}_1, \dots, \vec{\xi}_n\}$ of $\mathbb{K}_n$ so that

$$\mathbf{B}_N \cdots \mathbf{B}_1 = \Omega(\vec{\xi}_1, \dots, \vec{\xi}_n).$$

We now claim that $\mathcal{S} = \{\vec{\xi}_{r+1}, \dots, \vec{\xi}_n\}$ is a basis of U. The set $\mathcal{S}$ is linearly independent, hence we only need to show that $\operatorname{span}(\mathcal{S}) = U$. Since we have

$$\Omega(\vec{\xi}_1, \dots, \vec{\xi}_n)\mathbf{A} = \Omega\left(\vec{\omega}_1, \dots, \vec{\omega}_r, 0_{\mathbb{K}_m}, \dots, 0_{\mathbb{K}_m}\right),$$

the definition of matrix multiplication implies that $\vec{\xi}_i\mathbf{A} = \vec{\omega}_i$ for $1 \leqslant i \leqslant r$ and $\vec{\xi}_i\mathbf{A} = 0_{\mathbb{K}_m}$ for $r+1 \leqslant i \leqslant n$. Any vector in U can be written as $\vec{\nu} = \sum_{i=1}^n s_i\vec{\xi}_i$. The condition $\vec{\nu}\mathbf{A} = 0_{\mathbb{K}_m}$ then implies that $s_1 = \cdots = s_r = 0$, hence $\mathcal{S}$ is generating.

We can use this observation to compute the kernel and nullity of a linear map $\mathbb{K}^n \to \mathbb{K}^m$ because of the following lemma whose proof is left as an exercise.

Lemma 4.16 *Let* $\mathbf{C} \in M_{m,n}(\mathbb{K})$ *and* $f_\mathbf{C} : \mathbb{K}^n \to \mathbb{K}^m$ *be the associated linear map. Then* $\vec{x} \in \operatorname{Ker}(f_\mathbf{C})$ *if and only if* $\vec{x}^T\mathbf{C}^T = 0_{\mathbb{K}_m}$.

We simply apply the above procedure to the matrix $\mathbf{A} = \mathbf{C}^T$ and compute the vectors $\{\vec{\xi}_{r+1}, \dots, \vec{\xi}_n\}$. The basis of $\operatorname{Ker}(f_\mathbf{C})$ is then given by $\{\vec{\xi}^T_{r+1}, \dots, \vec{\xi}^T_n\}$.

The nullity of $f_\mathbf{C}$ is given by the number of basis vectors of $\operatorname{Ker}(f_\mathbf{C})$.

Example 4.17 (Kernel of a linear map — Video) Let

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 1 & 7 \\ -2 & -3 & 1 & 2 \\ 7 & 9 & -2 & 1 \end{pmatrix}$$

In order to compute $\operatorname{Ker}(f_\mathbf{C})$ we apply Gaussian elimination to $\mathbf{C}^T$ whilst keeping track of the relevant elementary matrices as in the algorithm for computing the inverse of a matrix. That is, we consider

$$\left(\begin{array}{ccc|cccc} 1 & -2 & 7 & 1 & 0 & 0 & 0 \\ 0 & -3 & 9 & 0 & 1 & 0 & 0 \\ 1 & 1 & -2 & 0 & 0 & 1 & 0 \\ 7 & 2 & 1 & 0 & 0 & 0 & 1 \end{array}\right).$$

Gauss–Jordan elimination (again, Gaussian elimination is enough) gives

$$\left(\begin{array}{ccc|cccc} 1 & 0 & 1 & 0 & 0 & -\frac{2}{5} & \frac{1}{5} \\ 0 & 1 & -3 & 0 & 0 & \frac{7}{5} & -\frac{1}{5} \\ 0 & 0 & 0 & 1 & 0 & \frac{16}{5} & -\frac{3}{5} \\ 0 & 0 & 0 & 0 & 1 & \frac{21}{5} & -\frac{3}{5} \end{array}\right).$$

The vectors $\vec{\xi}_3 = \begin{pmatrix} 1 & 0 & \frac{16}{5} & -\frac{3}{5} \end{pmatrix}$ and $\vec{\xi}_4 = \begin{pmatrix} 0 & 1 & \frac{21}{5} & -\frac{3}{5} \end{pmatrix}$ thus span the subspace of vectors ξ satisfying $\xi\mathbf{C}^T = 0_{\mathbb{K}_3}$. A basis $\mathcal{S}$ for the kernel of $f_\mathbf{C}$ is thus given by

$$\mathcal{S} = \left\{ \begin{pmatrix} 1 \\ 0 \\ \frac{16}{5} \\ -\frac{3}{5} \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \frac{21}{5} \\ -\frac{3}{5} \end{pmatrix} \right\}$$

and $f_{\mathbf{C}}$ satisfies $\operatorname{nullity}(f_{\mathbf{C}}) = 2$.

Remark 4.18 Section 4.2.3 and Section 4.2.4 can be combined to compute $\operatorname{Ker}(f_{\mathbf{A}})$ and $\operatorname{Im}(f_{\mathbf{A}})$ for $\mathbf{A} \in M_{m,n}(\mathbb{K})$ by a single application of Gaussian elimination.

Remark 4.19 In order to compute the kernel of a linear map $g : V \to W$ between finite dimensional vector spaces, we can fix an ordered basis $\mathbf{b}$ of V and an ordered basis $\mathbf{c}$ of W, compute $\mathbf{C} = \mathbf{M}(g, \mathbf{b}, \mathbf{c})$, apply the above procedure to the matrix $\mathbf{C}$ in order to obtain a basis $\mathcal{S}$ of $\operatorname{Ker}(f_{\mathbf{C}})$. The desired basis of $\operatorname{Ker}(g)$ is then given by $\boldsymbol{\beta}^{-1}(\mathcal{S})$. While this algorithm can always be carried out, it is computationally quite involved. In many cases it is therefore advisable to compute $\operatorname{Ker}(g)$ by some other technique.

CHAPTER 5

The determinant

5.1 Axiomatic characterisation

WEEK 8

Surprisingly, whether or not a square matrix $\mathbf{A} \in M_{n,n}(\mathbb{K})$ admits an inverse is captured by a single scalar, called the *determinant of* $\mathbf{A}$ and denoted by $\det \mathbf{A}$ or $\det(\mathbf{A})$. That is, the matrix $\mathbf{A}$ admits an inverse if and only if $\det \mathbf{A}$ is nonzero. In practice, however, it is often quicker to use Gauss–Jordan elimination to decide whether the matrix admits an inverse. The determinant is nevertheless a useful tool in linear algebra.

The determinant is an object of *multilinear algebra*, which – for $\ell \in \mathbb{N}$ – considers mappings from the ℓ-fold Cartesian product of a $\mathbb{K}$-vector space into another $\mathbb{K}$-vector space. Such a mapping f is required to be linear in each variable. This simply means that if we freeze all variables of f, except for the k-th variable, $1 \leqslant k \leqslant \ell$, then the resulting mapping g_k of one variable is required to be linear. More precisely:

Definition 5.1 (Multilinear map — Video) Let V, W be $\mathbb{K}$-vector spaces and $\ell \in \mathbb{N}$. A mapping $f : V^\ell \to W$ is called *ℓ-multilinear* (or simply multilinear) if the mapping $g_k : V \to W, v \mapsto f(v_1, \ldots, v_{k-1}, v, v_{k+1}, \ldots, v_\ell)$ is linear for all $1 \leqslant k \leqslant \ell$ and for all ℓ-tuples $(v_1, \ldots, v_\ell) \in V^\ell$.

We only need an $(\ell - 1)$-tuple of vectors to define the map g_k, but the above definition is more convenient to write down.

Two types of multilinear maps are of particular interest:

Definition 5.2 (Symmetric and alternating multilinear maps) Let V, W be $\mathbb{K}$-vector spaces and $f : V^\ell \to W$ an ℓ-multilinear map.

- The map f is called *symmetric* if exchanging two arguments does not change the value of f. That is, we have
$$f(v_1, \ldots, v_\ell) = f(v_1, \ldots, v_{i-1}, v_j, v_{i+1}, \ldots, v_{j-1}, v_i, v_{j+1}, \ldots, v_\ell)$$
for all $(v_1, \ldots, v_\ell) \in V^\ell$.
- The map f is called *alternating* if $f(v_1, \ldots, v_\ell) = 0_W$ whenever at least two arguments agree, that is, there exist $i \neq j$ with $v_i = v_j$. Alternating ℓ-multilinear maps are also called *W-valued ℓ-forms* or simply *ℓ-forms* when $W = \mathbb{K}$.

1-multilinear maps are simply linear maps. 2-multilinear maps are called *bilinear* and 3-multilinear maps are called *trilinear*. Most likely, you are already familiar with two examples of bilinear maps:

Example 5.3 (Bilinear maps)

(i) The first one is the *scalar product* of two vectors in $\mathbb{R}^3$ (or more generally $\mathbb{R}^n$). So $V = \mathbb{R}^3$ and $W = \mathbb{R}$. Recall that the scalar product is the mapping

$$V^2 = \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}, \quad (\vec{x}, \vec{y}) \mapsto \vec{x} \cdot \vec{y} = x_1 y_1 + x_2 y_2 + x_3 y_3,$$

where we write $\vec{x} = (x_i)_{1 \leqslant i \leqslant 3}$ and $\vec{y} = (y_i)_{1 \leqslant i \leqslant 3}$. Notice that for all $s_1, s_2 \in \mathbb{R}$ and all $\vec{x}_1, \vec{x}_2, \vec{y} \in \mathbb{R}^3$ we have

$$(s_1 \vec{x}_1 + s_2 \vec{x}_2) \cdot \vec{y} = s_1(\vec{x}_1 \cdot \vec{y}) + s_2(\vec{x}_2 \cdot \vec{y}),$$

so that the scalar product is linear in the first variable. Furthermore, the scalar product is symmetric, $\vec{x} \cdot \vec{y} = \vec{y} \cdot \vec{x}$. It follows that the scalar product is also linear in the second variable, hence it is bilinear or 2-multilinear.

(ii) The second one is the *cross product* of two vectors in $\mathbb{R}^3$. Here $V = \mathbb{R}^3$ and $W = \mathbb{R}^3$. Recall that the cross product is the mapping

$$V^2 = \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3, \quad (\vec{x}, \vec{y}) \mapsto \vec{x} \times \vec{y} = \begin{pmatrix} x_2 y_3 - x_3 y_2 \\ x_3 y_1 - x_1 y_3 \\ x_1 y_2 - x_2 y_1 \end{pmatrix}.$$

Notice that for all $s_1, s_2 \in \mathbb{R}$ and all $\vec{x}_1, \vec{x}_2, \vec{y} \in \mathbb{R}^3$ we have

$$(s_1 \vec{x}_1 + s_2 \vec{x}_2) \times \vec{y} = s_1(\vec{x}_1 \times \vec{y}) + s_2(\vec{x}_2 \times \vec{y}),$$

so that the cross product is linear in the first variable. Likewise, we can check that the cross product is also linear in the second variable, hence it is bilinear or 2-multilinear. Observe that the cross product is alternating.

Example 5.4 (Multilinear map) Let $V = \mathbb{K}$ and consider $f : V^\ell \to \mathbb{K}$, $(x_1, \ldots, x_\ell) \mapsto x_1 x_2 \cdots x_\ell$. Then f is ℓ-multilinear and symmetric.

Example 5.5 Let $\mathbf{A} \in M_{n,n}(\mathbb{R})$ be a symmetric matrix, $\mathbf{A}^T = \mathbf{A}$. Notice that we obtain a symmetric bilinear map

$$f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, \quad (x, y) \mapsto \vec{x}^T \mathbf{A} \vec{y},$$

where on the right hand side all products are defined by matrix multiplication.

The Example 5.5 gives us a wealth of symmetric bilinear maps on $\mathbb{R}^n$. As we will see shortly, the situation is quite different if we consider alternating n-multilinear maps on $\mathbb{K}_n$ (notice that we have the same number n of arguments as the dimension of $\mathbb{K}_n$).

Let $\{\vec{\varepsilon}_1, \ldots, \vec{\varepsilon}_n\}$ denote the standard basis of $\mathbb{K}_n$ so that $\Omega(\vec{\varepsilon}_1, \ldots, \vec{\varepsilon}_n) = \mathbf{1}_n$.

Theorem 5.6 *Let $n \in \mathbb{N}$. Then there exists a unique alternating n-multilinear map $f_n : (\mathbb{K}_n)^n \to \mathbb{K}$ satisfying $f_n(\vec{\varepsilon}_1, \ldots, \vec{\varepsilon}_n) = 1$.*

Recall that we have bijective mapping $\Omega : (\mathbb{K}_n)^n \to M_{n,n}(\mathbb{K})$ which forms an $n \times n$-matrix from n row vectors of length n. For the choice $V = \mathbb{K}_n$, the notion of n-multilinearity thus also makes sense for a mapping $f : M_{n,n}(\mathbb{K}) \to \mathbb{K}$ which takes an $n \times n$ matrix as an input. Here the multilinearity means the the mapping is linear in each row of the matrix. Since $\Omega(\vec{\varepsilon}_1, \ldots, \vec{\varepsilon}_n) = \mathbf{1}_n$, we may phrase the above theorem equivalently as:

Theorem 5.7 (Existence and uniqueness of the determinant) *Let $n \in \mathbb{N}$. Then there exists a unique alternating n-multilinear map $f_n : M_{n,n}(\mathbb{K}) \to \mathbb{K}$ satisfying $f_n(\mathbf{1}_n) = 1$.*

Definition 5.8 (Determinant — Video) The mapping $f_n : M_{n,n}(\mathbb{K}) \to \mathbb{K}$ provided by Theorem 5.7 is called the *determinant* and denoted by $\det$. For $\mathbf{A} \in M_{n,n}(\mathbb{K})$ we say $\det(\mathbf{A})$ is the determinant of the matrix $\mathbf{A}$.

Remark 5.9 (Abuse of notation) It would be more precise to write $\det_n$ since the determinant is a family of mappings, one mapping $\det_n : M_{n,n}(\mathbb{K}) \to \mathbb{K}$ for each $n \in \mathbb{N}$. It is however common to simply write $\det$.

Example 5.10 For $n = 1$ the condition that a 1-multilinear (i.e. linear) map $f_1 : M_{1,1}(\mathbb{K}) \to \mathbb{K}$ is alternating is vacuous. So the Theorem 5.7 states that there is a unique linear map $f_1 : M_{1,1}(\mathbb{K}) \to \mathbb{K}$ that satisfies $f_1((1)) = 1$. Of course, this is just the map defined by the rule $f_1((a)) = a$, where $(a) \in M_{1,1}(\mathbb{K})$ is any 1-by-1 matrix.

Example 5.11 For $n = 2$ and $a, b, c, d \in \mathbb{K}$ we consider the mapping $f_2 : M_{2,2}(\mathbb{K}) \to \mathbb{K}$ defined by the rule

$$f_2\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = ad - cb. \tag{5.1}$$

We claim that f_2 is bilinear in the rows and alternating. The condition that f_2 is alternating simplifies to $f(\mathbf{A}) = 0$ whenever the two rows of $\mathbf{A} \in M_{2,2}(\mathbb{K})$ agree. Clearly, f_2 is alternating, since

$$f_2\left(\begin{pmatrix} a & b \\ a & b \end{pmatrix}\right) = ab - ab = 0.$$

Furthermore, f_2 needs to be linear in each row. The additivity condition applied to the first row gives that we must have

$$f_2\left(\begin{pmatrix} a_1 + a_2 & b_1 + b_2 \\ c & d \end{pmatrix}\right) = f_2\left(\begin{pmatrix} a_1 & b_1 \\ c & d \end{pmatrix}\right) + f_2\left(\begin{pmatrix} a_2 & b_2 \\ c & d \end{pmatrix}\right)$$

for all $a_1, a_2, b_1, b_2, c, d \in \mathbb{K}$. Using the definition (5.1), we obtain

$$\begin{aligned} f_2\left(\begin{pmatrix} a_1 + a_2 & b_1 + b_2 \\ c & d \end{pmatrix}\right) &= (a_1 + a_2)d - c(b_1 + b_2) \\ &= a_1 d - cb_1 + a_2 d - cb_2 \\ &= f_2\left(\begin{pmatrix} a_1 & b_1 \\ c & d \end{pmatrix}\right) + f_2\left(\begin{pmatrix} a_2 & b_2 \\ c & d \end{pmatrix}\right), \end{aligned}$$

so that f_2 is indeed additive in the first row. The 1-homogeneity condition applied to the first row gives that we must have

$$f_2\left(\begin{pmatrix} sa & sb \\ c & d \end{pmatrix}\right) = sf_2\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right)$$

for all $a, b, c, d \in \mathbb{K}$ and $s \in \mathbb{K}$. Indeed, using the definition (5.1), we obtain

$$f_2\left(\begin{pmatrix} sa & sb \\ c & d \end{pmatrix}\right) = sad - csb = s(ad - cb) = sf_2\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right),$$

so that f_2 is also 1-homogeneous in the first row. We conclude that f_2 is linear in the first row. Likewise, the reader is invited to check that f_2 is also linear in the second row. Furthermore, we can easily compute that $f_2(\mathbf{1}_2) = 1$. The mapping f_2 thus satisfies all the properties of Theorem 5.7, hence by the uniqueness statement we must have $f_2 = \det$ and we obtain the formula

$$\det\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = ad - cb \tag{5.2}$$

for all $a, b, c, d \in \mathbb{K}$.

5.2 Uniqueness of the determinant

So far we have only shown that the determinant exists for $n = 1$ and $n = 2$. However, we need to show the existence and uniqueness part of Theorem 5.7 in general. We first show the uniqueness part. We start by deducing some consequences from the alternating property:

Lemma 5.12 *Let V, W be $\mathbb{K}$-vector spaces and $\ell \in \mathbb{N}$. An alternating ℓ-multilinear map $f : V^\ell \to W$ satisfies:*

(i) interchanging two arguments of f leads to a minus sign. That is, for $1 \leqslant i, j \leqslant \ell$ and $i \neq j$ we obtain

$$f(v_1, \ldots, v_\ell) = -f(v_1, \ldots, v_{i-1}, v_j, v_{i+1}, \ldots, v_{j-1}, v_i, v_{j+1}, \ldots, v_\ell)$$

for all $(v_1, \ldots, v_\ell) \in V^\ell$;

(ii) if the vectors $(v_1, \ldots, v_\ell) \in V^\ell$ are linearly dependent, then $f(v_1, \ldots, v_\ell) = 0_W$;

(iii) for all $1 \leqslant i \leqslant \ell$, for all ℓ-tuples of vectors $(v_1, \ldots, v_\ell) \in V^\ell$ and scalars $s_1, \ldots, s_\ell \in \mathbb{K}$, we have

$$f(v_1, \ldots, v_{i-1}, v_i + w, v_{i+1}, \ldots, v_\ell) = f(v_1, \ldots, v_\ell)$$

where $w = \sum_{j=1, j\neq i}^{\ell} s_j v_j$. That is, adding a linear combination of vectors to some argument of f does not change the output, provided the linear combination consists of the remaining arguments.

Proof (i) Since f is alternating, we have for all $(v_1, \ldots, v_\ell) \in V^\ell$

$$f(v_1, \ldots, v_{i-1}, v_i + v_j, v_{i+1}, \ldots, v_{j-1}, v_i + v_j, v_{j+1}, \ldots, v_\ell) = 0_W.$$

Using the linearity in the i-th argument, this gives

$$\begin{aligned} 0_W = {} & f(v_1, \ldots, v_{i-1}, v_i, v_{i+1}, \ldots, v_{j-1}, v_i + v_j, v_{j+1}, \ldots, v_\ell) \\ & + f(v_1, \ldots, v_{i-1}, v_j, v_{i+1}, \ldots, v_{j-1}, v_i + v_j, v_{j+1}, \ldots, v_\ell). \end{aligned}$$

Using the linearity in the j-th argument, we obtain

$$\begin{aligned} 0_W = {} & f(v_1, \ldots, v_{i-1}, v_i, v_{i+1}, \ldots, v_{j-1}, v_i, v_{j+1}, \ldots, v_\ell) \\ & + f(v_1, \ldots, v_{i-1}, v_i, v_{i+1}, \ldots, v_{j-1}, v_j, v_{j+1}, \ldots, v_\ell) \\ & + f(v_1, \ldots, v_{i-1}, v_j, v_{i+1}, \ldots, v_{j-1}, v_i, v_{j+1}, \ldots, v_\ell) \\ & + f(v_1, \ldots, v_{i-1}, v_j, v_{i+1}, \ldots, v_{j-1}, v_j, v_{j+1}, \ldots, v_\ell). \end{aligned}$$

The first summand has a double occurrence of v_i and hence vanishes by the alternating property. Likewise, the fourth summand has a double occurrence of v_j and hence vanishes as well. Since the second summand equals $f(v_1, \dots, v_\ell)$, we thus obtain

$$f(v_1, \dots, v_\ell) = -f(v_1, \dots, v_{i-1}, v_j, v_{i+1}, \dots, v_{j-1}, v_i, v_{j+1}, \dots, v_\ell)$$

as claimed.

(ii) Suppose $\{v_1, \dots, v_\ell\}$ are linearly dependent so that we have scalars $s_j \in \mathbb{K}$ not all zero, $1 \leqslant j \leqslant \ell$, so that $s_1 v_1 + \cdots + s_\ell v_\ell = 0_V$. Suppose $s_i \neq 0$ for some index $1 \leqslant i \leqslant \ell$. Then

$$v_i = -\sum_{j=1, j\neq i}^{\ell} \left(\frac{s_j}{s_i}\right) v_j$$

and hence by the linearity in the i-th argument, we obtain

$$\begin{aligned} f&\left(v_1, \dots, v_{i-1}, -\sum_{j=1, j\neq i}^{\ell} \left(\frac{s_j}{s_i}\right) v_j, v_{i+1}, \dots, v_\ell\right) \\ &= -\sum_{j=1, j\neq i}^{\ell} \left(\frac{s_j}{s_i}\right) f\left(v_1, \dots, v_{i-1}, v_j, v_{i+1}, \dots, v_\ell\right) = 0_W, \end{aligned}$$

where we use that for each $1 \leqslant j \leqslant \ell$ with $j \neq i$, the expression

$$f(v_1, \dots, v_{i-1}, v_j, v_{i+1}, \dots, v_\ell)$$

has a double occurrence of v_j and thus vanishes by the alternating property.

(iii) Let $(v_1, \dots, v_\ell) \in V^\ell$ and $(s_1, \dots, s_\ell) \in \mathbb{K}^\ell$. Then, using the linearity in the i-th argument, we compute

$$\begin{aligned} f&(v_1, \dots, v_{i-1}, v_i + \sum_{j=1, j\neq i}^{\ell} s_j v_j, v_{i+1}, \dots, v_\ell) \\ &= f(v_1, \dots, v_\ell) + \sum_{j=1, j\neq i}^{\ell} s_j f(v_1, \dots, v_{i-1} v_j, v_{i+1}, \dots, v_\ell) = f(v_1, \dots, v_\ell), \end{aligned}$$

where the last equality follows exactly as in the proof of (ii). □

The alternating property of an n-multilinear map $f_n : M_{n,n}(\mathbb{K}) \to \mathbb{K}$ together with the condition $f_n(\mathbf{1}_n) = 1$ uniquely determines the value of f_n on the elementary matrices:

Lemma 5.13 *Let $n \in \mathbb{N}$ and $f_n : M_{n,n}(\mathbb{K}) \to \mathbb{K}$ an alternating n-multilinear map satisfying $f_n(\mathbf{1}_n) = 1$. Then for all $1 \leqslant k, l \leqslant n$ with $k \neq l$ and all $s \in \mathbb{K}$, we have*

$$f_n(\mathbf{D}_k(s)) = s, \qquad f_n(\mathbf{L}_{k,l}(s)) = 1, \qquad f_n(\mathbf{P}_{k,l}) = -1. \tag{5.3}$$

Moreover, for $\mathbf{A} \in M_{n,n}(\mathbb{K})$ and an elementary matrix $\mathbf{B}$ of size n, we have

$$f_n(\mathbf{BA}) = f_n(\mathbf{B}) f_n(\mathbf{A}). \tag{5.4}$$

Proof Recall that $\mathbf{D}_k(s)$ applied to a square matrix $\mathbf{A}$ multiplies the k-th row of $\mathbf{A}$ with s and leaves $\mathbf{A}$ unchanged otherwise. We write $\mathbf{A} \in M_{n,n}(\mathbb{K})$ as $\mathbf{A} = \Omega(\vec{\alpha}_1, \dots, \vec{\alpha}_n)$ for $\vec{\alpha}_i \in \mathbb{K}_n$, $1 \leqslant i \leqslant n$. Hence we obtain

$$\mathbf{D}_k(s)\mathbf{A} = \Omega(\vec{\alpha}_1, \dots, \vec{\alpha}_{k-1}, s\vec{\alpha}_k, \vec{\alpha}_{k+1}, \dots, \vec{\alpha}_n).$$

The linearity of f in the k-th row thus gives $f_n(\mathbf{D}_k(s)\mathbf{A}) = s f_n(\mathbf{A})$. In particular, the choice $\mathbf{A} = \mathbf{1}_n$ together with $f_n(\mathbf{1}_n) = 1$ implies that $f_n(\mathbf{D}_k(s)) = f_n(\mathbf{D}_k(s)\mathbf{1}_n) = s f_n(\mathbf{1}_n) = s$.

Therefore, we have

$$f_n(\mathbf{D}_k(s)\mathbf{A}) = f_n(\mathbf{D}_k(s))f_n(\mathbf{A}).$$

Likewise we obtain

$$\mathbf{L}_{k,l}(s)\mathbf{A} = \Omega(\vec{\alpha}_1, \dots, \vec{\alpha}_{k-1}, \vec{\alpha}_k + s\vec{\alpha}_l, \vec{\alpha}_{k+1}, \dots, \vec{\alpha}_n)$$

and we can apply property (iii) of Lemma 5.12 for the choice $w = s\vec{\alpha}_l$ to conclude that $f_n(\mathbf{L}_{k,l}(s)\mathbf{A}) = f_n(\mathbf{A})$. In particular, the choice $\mathbf{A} = \mathbf{1}_n$ together with $f_n(\mathbf{1}_n) = 1$ implies $f_n(\mathbf{L}_{k,l}(s)) = f_n(\mathbf{L}_{k,l}(s)\mathbf{1}_n) = f_n(\mathbf{1}_n) = 1$.

Therefore, we have

$$f_n(\mathbf{L}_{k,l}(s)\mathbf{A}) = f_n(\mathbf{L}_{k,l}(s))f_n(\mathbf{A}).$$

Finally, we have

$$\mathbf{P}_{k,l}\mathbf{A} = \Omega(\vec{\alpha}_1, \dots, \vec{\alpha}_{k-1}, \vec{\alpha}_l, \vec{\alpha}_{k+1}, \dots, \vec{\alpha}_{l-1}, \vec{\alpha}_k, \vec{\alpha}_{l+1}, \dots, \vec{\alpha}_n)$$

so that property (ii) of Lemma 5.12 immediately gives that

$$f_n(\mathbf{P}_{k,l}\mathbf{A}) = -f_n(\mathbf{A}).$$

In particular, the choice $\mathbf{A} = \mathbf{1}_n$ together with $f_n(\mathbf{1}_n) = 1$ implies $f_n(\mathbf{P}_{k,l}) = f_n(\mathbf{P}_{k,l}\mathbf{1}_n) = -f_n(\mathbf{1}_n) = -1$.

Therefore, we have $f_n(\mathbf{P}_{k,l}\mathbf{A}) = f_n(\mathbf{P}_{k,l})f_n(\mathbf{A})$, as claimed. □

We now obtain the uniqueness part of Theorem 5.7.

Proposition 5.14 *Let $n \in \mathbb{N}$ and $f_n, \hat{f}_n : M_{n,n}(\mathbb{K}) \to \mathbb{K}$ be alternating n-multilinear maps satisfying $f_n(\mathbf{1}_n) = \hat{f}_n(\mathbf{1}_n) = 1$. Then $f_n = \hat{f}_n$.*

Proof We need to show that for all $\mathbf{A} \in M_{n,n}(\mathbb{K})$, we have $f_n(\mathbf{A}) = \hat{f}_n(\mathbf{A})$. Suppose first that $\mathbf{A}$ is not invertible. Then, by Proposition 4.7, the row vectors of $\mathbf{A}$ are linearly dependent and hence property (ii) of Lemma 5.12 implies that $f_n(\mathbf{A}) = \hat{f}_n(\mathbf{A}) = 0$.

Now suppose that $\mathbf{A}$ is invertible. Using Gauss–Jordan elimination, we obtain $N \in \mathbb{N}$ and a sequence of elementary matrices $\mathbf{B}_1, \dots, \mathbf{B}_N$ so that $\mathbf{B}_N \cdots \mathbf{B}_1 = \mathbf{A}$. We obtain

$$f_n(\mathbf{A}) = f_n(\mathbf{B}_N \cdots \mathbf{B}_1) = f_n(\mathbf{B}_N)f_n(\mathbf{B}_{N-1} \cdots \mathbf{B}_1) = \hat{f}_n(\mathbf{B}_N)f_n(\mathbf{B}_{N-1} \cdots \mathbf{B}_1),$$

where the second equality uses (5.4) and the third equality uses that (5.3) implies that $\hat{f}_n(\mathbf{B}) = f_n(\mathbf{B})$ for all elementary matrices $\mathbf{B}$. Proceeding in this fashion we get

$$\begin{aligned} f_n(\mathbf{A}) &= \hat{f}_n(\mathbf{B}_N)\hat{f}_n(\mathbf{B}_{N-1}) \cdots \hat{f}_n(\mathbf{B}_1) = \hat{f}_n(\mathbf{B}_N)\hat{f}_n(\mathbf{B}_{N-1}) \cdots \hat{f}_n(\mathbf{B}_2\mathbf{B}_1) = \cdots \\ &= \hat{f}_n(\mathbf{B}_N\mathbf{B}_{N-1} \cdots \mathbf{B}_1) = \hat{f}_n(\mathbf{A}). \end{aligned}$$

□

5.3 Existence of the determinant

It turns out that we can define the determinant recursively in terms of the determinants of certain submatrices. Determinants of submatrices are called *minors*. To this end we first define:

Definition 5.15 Let $n \in \mathbb{N}$. For a square matrix $\mathbf{A} \in M_{n,n}(\mathbb{K})$ and $1 \leqslant k, l \leqslant n$ we denote by $\mathbf{A}^{(k,l)}$ the $(n-1) \times (n-1)$ *submatrix* obtained by removing the k-th row and l-th column from $\mathbf{A}$.

Example 5.16

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad \mathbf{A}^{(1,1)} = (d), \qquad \mathbf{A}^{(2,1)} = (b).$$

$$\mathbf{A} = \begin{pmatrix} 1 & -2 & 0 & 4 \\ 3 & 1 & 1 & 0 \\ -1 & -5 & -1 & 8 \\ 3 & 8 & 2 & -12 \end{pmatrix}, \qquad \mathbf{A}^{(3,2)} = \begin{pmatrix} 1 & 0 & 4 \\ 3 & 1 & 0 \\ 3 & 2 & -12 \end{pmatrix}.$$

We use induction to prove the existence of the determinant:

Lemma 5.17 *Let $n \in \mathbb{N}$ with $n \geqslant 2$ and $f_{n-1} : M_{n-1,n-1}(\mathbb{K}) \to \mathbb{K}$ an alternating $(n-1)$-multilinear mapping satisfying $f_{n-1}(\mathbf{1}_{n-1}) = 1$. Then, for any fixed integer l with $1 \leqslant l \leqslant n$, the mapping*

$$f_n : M_{n,n}(\mathbb{K}) \to \mathbb{K}, \quad \mathbf{A} \mapsto \sum_{k=1}^{n} (-1)^{l+k} [\mathbf{A}]_{kl} f_{n-1}\left(\mathbf{A}^{(k,l)}\right)$$

is alternating, n-multilinear and satisfies $f_n(\mathbf{1}_n) = 1$.

Proof of Theorem 5.6 For $n = 1$ we have seen that $f_1 : M_{1,1}(\mathbb{K}) \to \mathbb{K}$, $(a) \mapsto a$ is 1-multilinear, alternating and satisfies $f_1(\mathbf{1}_1) = 1$. Hence Lemma 5.17 implies that for all $n \in \mathbb{N}$ there exists an n-multilinear and alternating map $f_n : M_{n,n}(\mathbb{K}) \to \mathbb{K}$ satisfying $f_n(\mathbf{1}_n) = 1$. By Proposition 5.14 there is only one such mapping for each $n \in \mathbb{N}$. □

Proof of Lemma 5.17 We take some arbitrary, but then fixed integer l with $1 \leqslant l \leqslant n$.

Step 1. We first show that $f_n(\mathbf{1}_n) = 1$. Since $[\mathbf{1}_n]_{kl} = \delta_{kl}$, we obtain

$$f_n(\mathbf{1}_n) = \sum_{k=1}^{n} (-1)^{l+k} [\mathbf{1}_n]_{kl} f_{n-1}\left(\mathbf{1}_n^{(k,l)}\right) = (-1)^{2l} f_{n-1}\left(\mathbf{1}_n^{(l,l)}\right) = f_{n-1}\left(\mathbf{1}_{n-1}\right) = 1,$$

where we use that $\mathbf{1}_n^{(l,l)} = \mathbf{1}_{n-1}$ and $f_{n-1}(\mathbf{1}_{n-1}) = 1$.

Step 2. We show that f_n is multilinear. Let $\mathbf{A} \in M_{n,n}(\mathbb{K})$ and write $\mathbf{A} = (A_{kj})_{1 \leqslant k,j \leqslant n}$. We first show that f_n is 1-homogeneous in each row. Say we multiply the i-th row of $\mathbf{A}$ with s so that we obtain a new matrix $\hat{\mathbf{A}} = (\hat{A}_{kj})_{1 \leqslant k,j \leqslant n}$ with

$$\hat{A}_{kj} = \begin{cases} A_{kj}, & k \neq i, \\ sA_{kj}, & k = i. \end{cases}$$

We need to show that $f_n(\hat{\mathbf{A}}) = sf_n(\mathbf{A})$. We compute

$$\begin{aligned} f_n(\hat{\mathbf{A}}) &= \sum_{k=1}^{n} (-1)^{l+k} \hat{A}_{kl} f_{n-1}(\hat{\mathbf{A}}^{(k,l)}) \\ &= (-1)^{l+i} sA_{il} f_{n-1}(\hat{\mathbf{A}}^{(i,l)}) + \sum_{k=1, k\neq i}^{n} (-1)^{l+k} A_{kl} f_{n-1}(\hat{\mathbf{A}}^{(k,l)}). \end{aligned}$$

Now notice that $\hat{\mathbf{A}}^{(i,l)} = \mathbf{A}^{(i,l)}$, since $\mathbf{A}$ and $\hat{\mathbf{A}}$ only differ in the i-th row, but this is the row that is removed. Since f_{n-1} is 1-homogeneous in each row, we obtain that $f_{n-1}(\hat{\mathbf{A}}^{(k,l)}) = sf_{n-1}(\mathbf{A}^{(k,l)})$ whenever $k \neq i$. Thus we have

$$\begin{aligned} f_n(\hat{\mathbf{A}}) &= s(-1)^{l+i}A_{il}f_{n-1}(\mathbf{A}^{(i,l)}) + s\sum_{k=1,k\neq i}^{n}(-1)^{l+k}A_{kl}f_{n-1}(\mathbf{A}^{(k,l)}) \\ &= s\sum_{k=1}^{n}(-1)^{l+k}A_{kl}f_{n-1}\left(\mathbf{A}^{(k,l)}\right) = sf_n(\mathbf{A}). \end{aligned}$$

We now show that f_n is additive in each row. Say the matrix $\mathbf{B} = (B_{kj})_{1\leqslant k,j\leqslant n}$ is identical to the matrix $\mathbf{A}$, except for the i-th row, so that

$$B_{kj} = \begin{cases} A_{kj} & k \neq i \\ B_j & k = i \end{cases}$$

for some scalars B_j with $1 \leqslant j \leqslant n$. We need to show that $f_n(\mathbf{C}) = f_n(\mathbf{A}) + f_n(\mathbf{B})$, where $\mathbf{C} = (C_{kj})_{1\leqslant k,j\leqslant n}$ with

$$C_{kj} = \begin{cases} A_{kj} & k \neq i \\ A_{ij} + B_j & k = i \end{cases}$$

We compute

$$f_n(\mathbf{C}) = (-1)^{l+i}(A_{il} + B_l)f_{n-1}(\mathbf{C}^{(i,l)}) + \sum_{k=1,k\neq i}^{n}(-1)^{l+k}A_{kl}f_{n-1}(\mathbf{C}^{(k,l)}).$$

As before, since $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ only differ in the i-th row, we have $\mathbf{A}^{(i,l)} = \mathbf{B}^{(i,l)} = \mathbf{C}^{(i,l)}$. Using that f_{n-1} is linear in each row, we thus obtain

$$\begin{aligned} f_n(\mathbf{C}) &= (-1)^{l+i}B_l f_{n-1}(\mathbf{B}^{(i,l)}) + \sum_{k=1,k\neq i}^{n}(-1)^{l+k}A_{kl}f_{n-1}(\mathbf{B}^{(k,l)}) \\ &\quad + (-1)^{l+i}A_{il}f_{n-1}(\mathbf{A}^{(i,l)}) + \sum_{k=1,k\neq i}^{n}(-1)^{l+k}A_{kl}f_{n-1}(\mathbf{A}^{(k,l)}) = f_n(\mathbf{A}) + f_n(\mathbf{B}). \end{aligned}$$

Step 3. We show that f_n is alternating. Suppose we have $1 \leqslant i, j \leqslant n$ with $j > i$ and so that the i-th and j-th row of the matrix $\mathbf{A} = (A_{ij})_{1\leqslant i,j\leqslant n}$ are the same. Therefore, unless $k = i$ or $k = j$, the submatrix $\mathbf{A}^{(k,l)}$ also contains two identical rows and since f_{n-1} is alternating, all summands vanish except the one for $k = i$ and $k = j$, this gives

$$\begin{aligned} f_n(\mathbf{A}) &= (-1)^{i+l}A_{il}f_{n-1}(\mathbf{A}^{(i,l)}) + (-1)^{j+l}A_{jl}f_{n-1}(\mathbf{A}^{(j,l)}) \\ &= A_{il}(-1)^l\left((-1)^i f_{n-1}(\mathbf{A}^{(i,l)}) + (-1)^j f_{n-1}(\mathbf{A}^{(j,l)})\right) \end{aligned}$$

where the second equality sign follows because we have $A_{il} = A_{jl}$ for all $1 \leqslant l \leqslant n$ (the i-th and j-th row agree). The mapping f_{n-1} is alternating, hence by the first property of the Lemma 5.12, swapping rows in the matrix $\mathbf{A}^{(j,l)}$ leads to a minus sign in $f_{n-1}(\mathbf{A}^{(j,l)})$. Moving the i-th row of $\mathbf{A}^{(j,l)}$ down by $j - i - 1$ rows (which corresponds to swapping $j - i - 1$ times), we obtain $\mathbf{A}^{(i,l)}$, hence

$$f_{n-1}(\mathbf{A}^{(j,l)}) = (-1)^{j-i-1}f_{n-1}(\mathbf{A}^{(i,l)}).$$

This gives

$$f_n(\mathbf{A}) = A_{il}(-1)^l\left((-1)^i f_{n-1}(\mathbf{A}^{(i,l)}) + (-1)^{2j-i-1}f_{n-1}(\mathbf{A}^{(i,l)})\right) = 0.$$

□

Remark 5.18 (Laplace expansion — Video) As a by-product of the proof of Lemma 5.17 we obtain the formula

$$\det(\mathbf{A}) = \sum_{k=1}^{n} (-1)^{l+k} [\mathbf{A}]_{kl} \det\left(\mathbf{A}^{(k,l)}\right), \tag{5.5}$$

known as the *Laplace expansion* of the determinant. The uniqueness statement of Theorem 5.7 thus guarantees that for every $n \times n$ matrix $\mathbf{A}$, the scalar $\sum_{k=1}^{n}(-1)^{l+k}[\mathbf{A}]_{kl}\det\left(\mathbf{A}^{(k,l)}\right)$ is independent of the choice of $l \in \mathbb{N}, 1 \leqslant l \leqslant n$. In practice, when computing the determinant, it is thus advisable to choose l such that the corresponding column contains the maximal amount of zeros.

Example 5.19 For $n = 2$ and choosing $l = 1$, we obtain

$$\det\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = a \det\left(\mathbf{A}^{(1,1)}\right) - c\det\left(\mathbf{A}^{(2,1)}\right) = ad - cb,$$

in agreement with (5.1). For $\mathbf{A} = (A_{ij})_{1 \leqslant i,j \leqslant 3} \in M_{3,3}(\mathbb{K})$ and choosing $l = 3$ we obtain

$$\begin{aligned} \det\left(\begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix}\right) &= A_{13}\det\left(\begin{pmatrix} A_{21} & A_{22} \\ A_{31} & A_{32} \end{pmatrix}\right) \\ &\quad - A_{23}\det\left(\begin{pmatrix} A_{11} & A_{12} \\ A_{31} & A_{32} \end{pmatrix}\right) + A_{33}\det\left(\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\right) \end{aligned}$$

so that

$$\begin{aligned} \det \mathbf{A} &= A_{13}(A_{21}A_{32} - A_{31}A_{22}) - A_{23}(A_{11}A_{32} - A_{31}A_{12}) \\ &\quad + A_{33}(A_{11}A_{22} - A_{21}A_{12}) \\ &= A_{11}A_{22}A_{33} - A_{11}A_{23}A_{32} - A_{12}A_{21}A_{33} \\ &\quad + A_{12}A_{23}A_{31} + A_{13}A_{21}A_{32} - A_{13}A_{22}A_{31}. \end{aligned}$$

Exercises

Exercise 5.20 (Trilinear map) Let $V = \mathbb{R}^3$ and $W = \mathbb{R}$. Show that the map

$$f : V^3 \to W, \quad (\vec{x}, \vec{y}, \vec{z}) \mapsto (\vec{x} \times \vec{y}) \cdot \vec{z}$$

is alternating and trilinear.

5.4 Properties of the determinant

WEEK 9

Proposition 5.21 (Product rule) *For matrices* $\mathbf{A}, \mathbf{B} \in M_{n,n}(\mathbb{K})$ *we have*

$$\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B}).$$

Proof We first consider the case where $\mathbf{A}$ is not invertible, then $\det(\mathbf{A}) = 0$ (see the proof of Proposition 5.14). If $\mathbf{A}$ is not invertible, then neither is $\mathbf{AB}$. Indeed, if $\mathbf{AB}$ were invertible, then there exists a matrix $\mathbf{C}$ such that $(\mathbf{AB})\mathbf{C} = \mathbf{1}_n$. But since, by Corollary 2.22, the matrix product is associative, this also gives $\mathbf{A}(\mathbf{BC}) = \mathbf{1}_n$, so that $\mathbf{BC}$ is the inverse of $\mathbf{A}$, a contradiction. Hence if $\mathbf{A}$ is not invertible, we must also have $\det(\mathbf{AB}) = 0$, which verifies that $\det(\mathbf{AB}) = 0 = \det(\mathbf{A})\det(\mathbf{B})$ for $\mathbf{A}$ not invertible.

If $\mathbf{A}$ is invertible, we can write it as a product of elementary matrices and applying the second part of Lemma 5.13, we conclude that $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$. □

Corollary 5.22 *A matrix* $\mathbf{A} \in M_{n,n}(\mathbb{K})$ *is invertible if and only if* $\det(\mathbf{A}) \neq 0$. *Moreover, in the case where* $\mathbf{A}$ *is invertible, we have*

$$\det\left(\mathbf{A}^{-1}\right) = \frac{1}{\det \mathbf{A}}.$$

Proof We have already seen that if $\mathbf{A}$ is not invertible, then $\det(\mathbf{A}) = 0$. This is equivalent to saying that if $\det(\mathbf{A}) \neq 0$, then $\mathbf{A}$ is invertible. It thus remains to show that if $\mathbf{A}$ is invertible, then $\det(\mathbf{A}) \neq 0$. Suppose $\mathbf{A}$ is invertible, then applying Proposition 5.21 gives

$$\det(\mathbf{1}_n) = \det\left(\mathbf{A}\mathbf{A}^{-1}\right) = \det(\mathbf{A})\det\left(\mathbf{A}^{-1}\right) = 1$$

so that $\det(\mathbf{A}) \neq 0$ and $\det\left(\mathbf{A}^{-1}\right) = 1/\det(\mathbf{A})$. □

Remark 5.23 (Product symbol) Recall that for scalars $x_1, \ldots, x_n \in \mathbb{K}$, we write

$$\prod_{i=1}^{n} x_i = x_1 x_2 \cdots x_n.$$

Proposition 5.24 *Let* $n \in \mathbb{N}$ *and* $\mathbf{A} = (A_{ij})_{1\leqslant i,j\leqslant n} \in M_{n,n}(\mathbb{K})$ *be an* upper triangular matrix *so that* $A_{ij} = 0$ *for* $i > j$. *Then*

$$\det(\mathbf{A}) = \prod_{i=1}^{n} A_{ii} = A_{11}A_{22}\cdots A_{nn}. \tag{5.6}$$

Proof We use induction. For $n = 1$ the condition $A_{ij} = 0$ for $i > j$ is vacuous and (5.6) is trivially satisfied, thus the statement is anchored.

Inductive step: Assume $n \in \mathbb{N}$ and $n \geqslant 2$. We want to show that if (5.6) holds for upper triangular $(n-1) \times (n-1)$-matrices, then also for upper triangular $n \times n$-matrices. Let $\mathbf{A} = (A_{ij})_{1\leqslant i,j\leqslant n} \in M_{n,n}(\mathbb{K})$ be an upper triangular matrix. Choosing $l = 1$ in the formula for $\det(\mathbf{A})$, we obtain

$$\det(\mathbf{A}) = \sum_{k=1}^{n}(-1)^{k+1}A_{k1}\det\left(\mathbf{A}^{(k,1)}\right) = A_{11}\det\left(\mathbf{A}^{(1,1)}\right) + \sum_{k=2}^{n}A_{k1}\det\left(\mathbf{A}^{(k,1)}\right)$$
$$= A_{11}\det\left(\mathbf{A}^{(1,1)}\right),$$

where the last equality uses that $A_{k1} = 0$ for $k > 1$. We have $\mathbf{A}^{(1,1)} = (A_{ij})_{2\leqslant i,j\leqslant n}$ and since $\mathbf{A}$ is an upper triangular matrix, it follows that $\mathbf{A}^{(1,1)}$ is an $(n-1)\times(n-1)$ upper triangular matrix as well. Hence by the induction hypothesis, we obtain

$$\det(\mathbf{A}^{(1,1)}) = \prod_{i=2}^{n} A_{ii}.$$

We conclude that $\det(\mathbf{A}) = \prod_{i=1}^{n} A_{ii}$, as claimed. □

5.5 Permutations

A rearrangement of the natural numbers from 1 up to n is called a permutation:

Definition 5.25 (Permutation — Video) Let $n \in \mathbb{N}$ and $\mathcal{X}_n = \{1, 2, 3, \dots, n\}$. A *permutation* is a bijective mapping $\sigma : \mathcal{X}_n \to \mathcal{X}_n$. The set of all permutations of $\mathcal{X}_n$ is denoted by S_n.

Remark 5.26 If $\tau, \sigma : \mathcal{X}_n \to \mathcal{X}_n$ are permutations, it is customary to write $\tau\sigma$ or $\tau \cdot \sigma$ instead of $\tau \circ \sigma$. Furthermore, the identity mapping $\mathrm{Id}_{\mathcal{X}_n}$ is often simply denoted by 1. A convenient way to describe a permutation $\sigma \in S_n$ is in terms of a $2 \times n$ matrix

$$\begin{pmatrix} i \\ \sigma(i)\end{pmatrix}_{1\leqslant i\leqslant n}.$$

which we denote by $\boldsymbol{\sigma}$. For instance, for $n = 4$, the matrix

$$\boldsymbol{\sigma} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4\end{pmatrix}$$

corresponds to the permutation σ satisfying $\sigma(1) = 2, \sigma(2) = 3, \sigma(3) = 1, \sigma(4) = 4$.

Permutations which only swap two natural numbers and leave all remaining numbers unchanged are known as *transpositions*:

Definition 5.27 (Transposition) Let $n \in \mathbb{N}$ and $1 \leqslant k, l \leqslant n$ with $k \neq l$. The *transposition* $\tau_{k,l} \in S_n$ is the permutation satisfying

$$\tau_{k,l}(k) = l, \qquad \tau_{k,l}(l) = k, \qquad \tau_{k,l}(i) = i \text{ if } i \notin \{k, l\}.$$

Every permutation $\sigma \in S_n$ defines a linear map $g : \mathbb{K}^n \to \mathbb{K}^n$ satisfying $g(\vec{e}_i) = \vec{e}_{\sigma(i)}$, where $\{\vec{e}_1, \dots, \vec{e}_n\}$ denotes the standard basis of $\mathbb{K}^n$. Since g is linear, there exists a unique matrix $\mathbf{P}_\sigma \in M_{n,n}(\mathbb{K})$ so that $g = f_{\mathbf{P}_\sigma}$. Observe that the column vectors of the matrix $\mathbf{P}_\sigma$ are given by $\vec{e}_{\sigma(1)}, \vec{e}_{\sigma(2)}, \dots, \vec{e}_{\sigma(n)}$.

Definition 5.28 (Permutation matrix) We call $\mathbf{P}_\sigma \in M_{n,n}(\mathbb{K})$ the *permutation matrix* associated to $\sigma \in S_n$.

Example 5.29 Let $n = 4$. For instance, we have

$$\boldsymbol{\sigma} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4 \end{pmatrix} \qquad \mathbf{P}_\sigma = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and

$$\boldsymbol{\tau}_{2,4} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 3 & 2 \end{pmatrix} \qquad \mathbf{P}_{\tau_{2,4}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

Remark 5.30 Notice that $\mathbf{P}_{\tau_{k,l}} = \mathbf{P}_{k,l}$, where $\mathbf{P}_{k,l}$ belongs to the elementary matrices of size n, c.f. Definition 4.1.

Assigning to a permutation its permutation matrix turns composition of permutations into matrix multiplication:

Proposition 5.31 *Let $n \in \mathbb{N}$. Then $\mathbf{P}_1 = \mathbf{1}_n$ and for all $\sigma, \pi \in S_n$ we have*

$$\mathbf{P}_{\pi\cdot\sigma} = \mathbf{P}_\pi \mathbf{P}_\sigma.$$

In particular, for all $\sigma \in S_n$, the permutation matrix $\mathbf{P}_\sigma$ is invertible with $(\mathbf{P}_\sigma)^{-1} = \mathbf{P}_{\sigma^{-1}}$.

Example 5.32 Considering $n = 3$. For

$$\boldsymbol{\sigma} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\pi} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \quad \text{we have} \quad \boldsymbol{\pi}\cdot\boldsymbol{\sigma} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix},$$

as well as

$$\mathbf{P}_\sigma = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \qquad \mathbf{P}_\pi = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{P}_{\pi\cdot\sigma} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Thus we obtain

$$\mathbf{P}_{\pi\cdot\sigma} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} = \mathbf{P}_\pi \mathbf{P}_\sigma,$$

as claimed by Proposition 5.31.

Proof of Proposition 5.31 The matrix $\mathbf{P}_1$ has column vectors given by $\vec{e}_1, \ldots, \vec{e}_n$, hence $\mathbf{P}_1 = \mathbf{1}_n$.

Using Proposition 2.20 and Theorem 2.21 it is sufficient to show that for all $\pi, \sigma \in S_n$ we have $f_{\mathbf{P}_{\pi\cdot\sigma}} = f_{\mathbf{P}_\pi} \circ f_{\mathbf{P}_\sigma}$. For all $1 \leqslant i \leqslant n$, we obtain

$$f_{\mathbf{P}_\pi}\left(f_{\mathbf{P}_\sigma}(\vec{e}_i)\right) = f_{\mathbf{P}_\pi}\left(\vec{e}_{\sigma(i)}\right) = \vec{e}_{\pi(\sigma(i))} = \vec{e}_{(\pi\cdot\sigma)(i)} = f_{\mathbf{P}_{\pi\cdot\sigma}}(\vec{e}_i).$$

The two maps thus agree on the ordered basis $\mathbf{e} = (\vec{e}_1, \ldots, \vec{e}_n)$ of $\mathbb{K}^n$, so that the second claim follows by applying Lemma 3.87.

We have

$$\mathbf{P}_{\sigma\cdot\sigma^{-1}} = \mathbf{P}_1 = \mathbf{1}_n = \mathbf{P}_\sigma \mathbf{P}_{\sigma^{-1}}$$

showing that $\mathbf{P}_\sigma$ is invertible with inverse $(\mathbf{P}_\sigma)^{-1} = \mathbf{P}_{\sigma^{-1}}$. □

Definition 5.33 (Signature of a permutation) For $\sigma \in S_n$ we call $\operatorname{sgn}(\sigma) = \det(\mathbf{P}_\sigma)$ its *signature*.

Remark 5.34

(i) Combining Proposition 5.21 and Proposition 5.31, we conclude that

$$\operatorname{sgn}(\pi\cdot\sigma) = \operatorname{sgn}(\pi)\operatorname{sgn}(\sigma)$$

for all $\pi, \sigma \in S_n$.

(ii) Since $\mathbf{P}_{\tau_{k,l}} = \mathbf{P}_{k,l}$ and $\det \mathbf{P}_{k,l} = -1$ by Lemma 5.13, we conclude that

$$\operatorname{sgn}(\tau_{k,l}) = -1$$

for all transpositions $\tau_{k,l} \in S_n$.

Similarly to elementary matrices being the building blocks of invertible matrices, transpositions are the building blocks of permutations:

Proposition 5.35 *Let $n \in \mathbb{N}$ and $\sigma \in S_n$. Then there exists $m \geqslant 0$ and m transpositions $\tau_{k_1,l_1}, \ldots, \tau_{k_m,l_m} \in S_n$ such that $\sigma = \tau_{k_m,l_m} \cdots \tau_{k_1,l_1}$, where we use the convention that 0 transpositions corresponds to the identity permutation.*

Example 5.36 Let $n = 6$ and σ the permutation defined by the matrix

$$\boldsymbol{\sigma} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 5 & 2 & 4 & 6 & 1 \end{pmatrix}.$$

To express it as a product of transposition, we write

$$\begin{array}{cccccc|l} 3 & 5 & 2 & 4 & 6 & 1 & \\ 3 & 2 & 5 & 4 & 6 & 1 & \tau_{2,3} \\ 1 & 2 & 5 & 4 & 6 & 3 & \tau_{1,6} \\ 1 & 2 & 5 & 4 & 3 & 6 & \tau_{5,6} \\ 1 & 2 & 3 & 4 & 5 & 6 & \tau_{3,5} \end{array}$$

so that $\sigma = \tau_{3,5}\tau_{5,6}\tau_{1,6}\tau_{2,3}$.

Proof of Proposition 5.35 We use induction. For $n = 1$ we have $\mathcal{X}_n = \{1\}$ and the only permutation is the identity permutation 1, so the statement is trivially true and hence anchored.

Inductive step: Assume $n \in \mathbb{N}$ and $n \geqslant 2$. We want to show that if the claim holds for S_{n-1}, then also for S_n. Let $\sigma \in S_n$ and define $k = \sigma(n)$. Then the permutation $\sigma_1 = \tau_{n,k}\sigma$ satisfies $\sigma_1(n) = \tau_{n,k}\sigma(n) = \tau_{n,k}(k) = n$ and hence does not permute n. Restricting σ_1 to the first $n-1$ elements, we obtain a permutation of $\{1, \dots, n-1\}$. By the induction hypothesis, we thus have $\tilde{m} \in \mathbb{N}$ and $\tau_{k_1,l_1}, \dots \tau_{k_{\tilde{m}}}, \tau_{l_{\tilde{m}}} \in S_n$ such that

$$\sigma_1 = \tau_{k_{\tilde{m}},l_{\tilde{m}}} \cdots \tau_{k_1,l_1} = \tau_{n,k}\sigma.$$

Since $\tau_{n,k}^2 = 1$, multiplying from the left with $\tau_{n,k}$ gives $\sigma = \tau_{n,k}\tau_{k_{\tilde{m}},l_{\tilde{m}}} \cdots \tau_{k_1,l_1}$, the claim follows with $m = \tilde{m} + 1$. ☐

Combining Definition 5.33, Remark 5.34 and Proposition 5.35, we conclude:

Proposition 5.37 *Let $n \in \mathbb{N}$ and $\sigma \in S_n$. Then $\operatorname{sgn}(\sigma) = \pm 1$. If σ is a product of m transpositions, then $\operatorname{sgn}(\sigma) = (-1)^m$.*

Remark 5.38 Permutations with $\operatorname{sgn}(\sigma) = 1$ are called *even* and permutations with $\operatorname{sgn}(\sigma) = -1$ are called *odd*, since they arise from the composition of an even or odd number of transpositions, respectively.

5.6 The Leibniz formula

Besides the Laplace expansion, there is also a formula for the determinant which relies on permutations. As a warm-up, we first consider the case $n = 2$. Using the linearity of the determinant in the first row, we obtain

$$\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \det\begin{pmatrix} a & 0 \\ c & d \end{pmatrix} + \det\begin{pmatrix} 0 & b \\ c & d \end{pmatrix},$$

where $a, b, c, d \in \mathbb{K}$. Using the linearity of the determinant in the second row, we can further decompose the two above summands

$$\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \underbrace{\det\begin{pmatrix} a & 0 \\ c & 0 \end{pmatrix} + \det\begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}}_{=\det\begin{pmatrix} a & 0 \\ c & d \end{pmatrix}} + \underbrace{\det\begin{pmatrix} 0 & b \\ c & 0 \end{pmatrix} + \det\begin{pmatrix} 0 & b \\ 0 & d \end{pmatrix}}_{=\det\begin{pmatrix} 0 & b \\ c & d \end{pmatrix}}$$

The first and fourth summand are *always zero* due to the occurrence of a zero column. The second and third summand are *possibly nonzero* (it might still happen that they are zero in the case where some of a, b, c, d are zero). In any case, we get

$$\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \det\begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix} + \det\begin{pmatrix} 0 & b \\ c & 0 \end{pmatrix}.$$

We can proceed analogously in general. Let $\mathbf{A} = (A_{ij})_{1\leqslant i,j\leqslant n} \in M_{n,n}(\mathbb{K})$. We denote the rows of $\mathbf{A}$ by $\{\vec{\alpha}_1, \ldots, \vec{\alpha}_n\}$. Using the linearity of det in the first row, we can write

$$\det \mathbf{A} = \det \begin{pmatrix} A_{11} & 0 & 0 & \cdots & 0 \\ & & \vec{\alpha}_2 & & \\ & & \vdots & & \\ & & \vec{\alpha}_n & & \end{pmatrix} + \det \begin{pmatrix} 0 & A_{12} & 0 & \cdots & 0 \\ & & \vec{\alpha}_2 & & \\ & & \vdots & & \\ & & \vec{\alpha}_n & & \end{pmatrix} + \cdots$$

$$\cdots + \det \begin{pmatrix} 0 & 0 & 0 & \cdots & A_{1n} \\ & & \vec{\alpha}_2 & & \\ & & \vdots & & \\ & & \vec{\alpha}_n & & \end{pmatrix}.$$

We can now use the linearity in the second row and proceed in the same fashion with each of the above summands. We continue this procedure until the n-th row. As a result, we can write

$$\det \mathbf{A} = \sum_{k=1}^{n^n} \det \mathbf{M}_k \tag{5.7}$$

where each of the matrices $\mathbf{M}_k$ has exactly one possibly nonzero entry in each row. As above, some of the matrices $\mathbf{M}_k$ will have a zero column so that their determinant vanishes. The matrices $\mathbf{M}_k$ without a zero column must have exactly one possibly nonzero entry in each row and each column. We can thus write the matrices $\mathbf{M}_k$ with possibly nonzero determinant as

$$\mathbf{M}_k = \sum_{i=1}^{n} A_{\sigma(i)i}\mathbf{E}_{\sigma(i),i}$$

for some permutation $\sigma \in S_n$. Every permutation of $\{1, \ldots, n\}$ occurs precisely once in the expansion (5.7), hence we can write

$$\det \mathbf{A} = \sum_{\sigma\in S_n} \det \left(\sum_{i=1}^{n} A_{\sigma(i)i}\mathbf{E}_{\sigma(i),i} \right),$$

where the notation $\sum_{\sigma\in S_n}$ means that we sum over all possible permutations of $\{1, \ldots, n\}$. We will next write the matrix $\sum_{i=1}^{n} A_{\sigma(i)i}\mathbf{E}_{\sigma(i),i}$ differently. To this end notice that for all $\sigma \in S_n$, the permutation matrix $\mathbf{P}_\sigma$ can be written as $\mathbf{P}_\sigma = \sum_{i=1}^{n} \mathbf{E}_{\sigma(i),i}$. Furthermore, the diagonal matrix

$$\mathbf{D}_\sigma = \begin{pmatrix} A_{\sigma(1)1} & & & \\ & A_{\sigma(2)2} & & \\ & & \ddots & \\ & & & A_{\sigma(n)n} \end{pmatrix}$$

can be written as $\mathbf{D}_\sigma = \sum_{j=1}^{n} A_{\sigma(j)j}\mathbf{E}_{j,j}$. Therefore, using Lemma 4.4, we obtain

$$\mathbf{P}_\sigma\mathbf{D}_\sigma = \sum_{i=1}^{n} \mathbf{E}_{\sigma(i),i} \sum_{j=1}^{n} A_{\sigma(j)j}\mathbf{E}_{j,j} = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{\sigma(j)j}\mathbf{E}_{\sigma(i),i}\mathbf{E}_{j,j} = \sum_{i=1}^{n} A_{\sigma(i)i}\mathbf{E}_{\sigma(i),i},$$

We thus have the formula

$$\det \mathbf{A} = \sum_{\sigma\in S_n} \det (\mathbf{P}_\sigma\mathbf{D}_\sigma) = \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \det(\mathbf{D}_\sigma),$$

where we use the product rule Proposition 5.21 and the definition of the signature of a permutation. By Proposition 5.24, the determinant of a diagonal matrix is the product of its diagonal entries, hence we obtain

$$\det \mathbf{A} = \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} A_{\sigma(i)i}.$$

Finally, writing $\pi = \sigma^{-1}$, we have

$$\prod_{i=1}^{n} A_{\sigma(i)i} = \prod_{j=1}^{n} A_{j\pi(j)}.$$

We have thus shown:

Proposition 5.39 (Leibniz formula for the determinant) *Let $n \in \mathbb{N}$ and $\mathbf{A} = (A_{ij})_{1\leqslant i,j\leqslant n} \in M_{n,n}(\mathbb{K})$. Then we have*

$$\det(\mathbf{A}) = \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma)\prod_{i=1}^{n} A_{\sigma(i)i} = \sum_{\pi\in S_n} \operatorname{sgn}(\pi)\prod_{j=1}^{n} A_{j\pi(j)}. \tag{5.8}$$

Example 5.40 For $n = 3$ we have six permutations

$$\boldsymbol{\sigma}_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \quad \boldsymbol{\sigma}_2 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \quad \boldsymbol{\sigma}_3 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$$
$$\boldsymbol{\sigma}_4 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \quad \boldsymbol{\sigma}_5 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}, \quad \boldsymbol{\sigma}_6 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}.$$

For $\mathbf{A} = (A_{ij})_{1\leqslant i,j\leqslant 3} \in M_{3,3}(\mathbb{K})$, the Leibniz formula gives

$$\begin{aligned}\det(\mathbf{A}) = {} & \operatorname{sgn}(\sigma_1)A_{11}A_{22}A_{33} + \operatorname{sgn}(\sigma_2)A_{11}A_{23}A_{32} + \operatorname{sgn}(\sigma_3)A_{12}A_{21}A_{33} \\ & + \operatorname{sgn}(\sigma_4)A_{12}A_{23}A_{31} + \operatorname{sgn}(\sigma_5)A_{13}A_{21}A_{32} + \operatorname{sgn}(\sigma_6)A_{13}A_{22}A_{31},\end{aligned}$$

so that in agreement with Example 5.19, we obtain

$$\begin{aligned}\det \mathbf{A} = {} & A_{11}A_{22}A_{33} - A_{11}A_{23}A_{32} - A_{12}A_{21}A_{33} \\ & + A_{12}A_{23}A_{31} + A_{13}A_{21}A_{32} - A_{13}A_{22}A_{31}.\end{aligned}$$

Remark 5.41 Exercise 5.49 has two important consequences. Since the transpose turns the rows of a matrix into columns and vice versa, we conclude:

(i) the determinant is also multilinear and alternating, when thought of as a map $(\mathbb{K}^n)^n \to \mathbb{K}$, that is, when taking n columns vectors as an input. In particular, the determinant is also linear in each column;

(ii) the Laplace expansion is also valid if we expand with respect to a row, that is, for $\mathbf{A} \in M_{n,n}(\mathbb{K})$ and $1 \leqslant l \leqslant n$, we have

$$\det(\mathbf{A}) = \sum_{k=1}^{n} (-1)^{k+l}[\mathbf{A}]_{lk} \det\left(\mathbf{A}^{(l,k)}\right).$$

Example 5.42 (♡ – not examinable) For $n \in \mathbb{N}$ and a vector $\vec{x} = (x_i)_{1\leqslant i\leqslant n} \in \mathbb{K}^n$ we can form a matrix $\mathbf{V}_{\vec{x}} = (V_{ij})_{1\leqslant i,j\leqslant n} \in M_{n,n}(\mathbb{K})$ with $V_{ij} = x_i^{j-1}$, that is,

$$\mathbf{V}_{\vec{x}} = \begin{pmatrix} 1 & x_1 & (x_1)^2 & \cdots & (x_1)^{n-1} \\ 1 & x_2 & (x_2)^2 & \cdots & (x_2)^{n-1} \\ 1 & x_3 & (x_3)^2 & \cdots & (x_3)^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & (x_n)^2 & \cdots & (x_n)^{n-1} \end{pmatrix}.$$

Such matrices are known as *Vandermonde matrices* and the determinant of a Vandermonde matrix is known as a *Vandermonde determinant*, they satisfy

$$\det(\mathbf{V}_{\vec{x}}) = \prod_{1\leqslant i<j\leqslant n} (x_j - x_i).$$

Sketch of a proof We can define a function $f : \mathbb{K}^n \to \mathbb{K}, \vec{x} \mapsto \det(\mathbf{V}_{\vec{x}})$. By the Leibniz formula, the function f is a polynomial in the variables x_i with integer coefficients. If we freeze all variables of f except the ℓ-th variable, then we obtain a function $g_\ell : \mathbb{K} \to \mathbb{K}$ of one variable x_ℓ. For $1 \leqslant i \leqslant n$ with $i \neq \ell$ we have $g_\ell(x_i) = 0$, since we compute the determinant of a matrix with two identical rows, the ℓ-th row and the i-th row. Factoring the zeros, we can thus write $g_\ell(x_\ell) = q_\ell(x_\ell)\prod_{1\leqslant i\leqslant n, i\neq\ell}(x_\ell - x_i)$ for some polynomial q_ℓ. We can repeat this argument for all ℓ and hence can write $\det(\mathbf{V}_{\vec{x}}) = q(\vec{x})\prod_{1\leqslant i<j\leqslant n}(x_j - x_i)$ for some polynomial $q(\vec{x})$. The Leibniz formula implies that the sum of the exponents of all the factors x_i in $\det(\mathbf{V}_{\vec{x}})$ must be $\frac{1}{2}n(n-1)$. The same holds true for $\prod_{1\leqslant i<j\leqslant n}$. It follows that q must be a constant. Using the Leibniz formula again, we see that the summand of $\det(\mathbf{V}_{\vec{x}})$ corresponding to the identity permutation is the product of the diagonal entries of $\mathbf{V}_{\vec{x}}$, that is, $x_2(x_3)^2\cdots(x_n)^{n-1}$. Taking the first term in all factors of $\prod_{1\leqslant i<j\leqslant n}(x_j - x_i)$, we also obtain $x_2(x_3)^2\cdots(x_n)^{n-1}$, hence $\det(\mathbf{V}_{\vec{x}}) = \prod_{1\leqslant i<j\leqslant n}(x_j - x_i)$, as claimed. □

5.7 Cramer's rule

The determinant can be used to give a formula for the solution of a linear system of equations of the form $\mathbf{A}\vec{x} = \vec{b}$ for an invertible matrix $\mathbf{A} \in M_{n,n}(\mathbb{K})$, $\vec{b} \in \mathbb{K}^n$ and unknowns $\vec{x} \in \mathbb{K}^n$. This formula is often referred to as *Cramer's rule*. In order to derive it we start with definitions:

Definition 5.43 (Adjugate matrix — Video) Let $n \in \mathbb{N}$ and $\mathbf{A} \in M_{n,n}(\mathbb{K})$ be a square matrix. The *adjugate matrix of* $\mathbf{A}$ is the $n \times n$-matrix $\mathrm{Adj}(\mathbf{A})$ whose entries are given by (notice the reverse order of i and j on the right hand side)

$$[\mathrm{Adj}(\mathbf{A})]_{ij} = (-1)^{i+j}\det\left(\mathbf{A}^{(j,i)}\right), \qquad 1 \leqslant i, j \leqslant n.$$

Example 5.44

$$\mathrm{Adj}\left(\begin{pmatrix} a & b \\ c & d\end{pmatrix}\right) = \begin{pmatrix} d & -b \\ -c & a\end{pmatrix}, \quad \mathrm{Adj}\left(\begin{pmatrix} 1 & 1 & 2 \\ 0 & 2 & 1 \\ 1 & 0 & 2\end{pmatrix}\right) = \begin{pmatrix} 4 & -2 & -3 \\ 1 & 0 & -1 \\ -2 & 1 & 2\end{pmatrix}$$

The determinant and the adjugate matrix provide a formula for the inverse of a matrix:

Theorem 5.45 *Let $n \in \mathbb{N}$ and $\mathbf{A} \in M_{n,n}(\mathbb{K})$. Then we have*

$$\mathrm{Adj}(\mathbf{A})\mathbf{A} = \mathbf{A}\,\mathrm{Adj}(\mathbf{A}) = \det(\mathbf{A})\mathbf{1}_n.$$

In particular, if $\mathbf{A}$ is invertible then

$$\mathbf{A}^{-1} = \frac{1}{\det\mathbf{A}}\mathrm{Adj}(\mathbf{A}).$$

Proof Let $\mathbf{A} = (A_{ij})_{1 \leqslant i,j \leqslant n}$. For $1 \leqslant i \leqslant n$ we obtain for the i-th diagonal entry

$$[\mathrm{Adj}(\mathbf{A})\mathbf{A}]_{ii} = \sum_{k=1}^{n} (-1)^{i+k} \det\left(\mathbf{A}^{(k,i)}\right) A_{ki} = \det(\mathbf{A}),$$

where we use the Laplace expansion (5.5) of the determinant. The diagonal entries of $\mathrm{Adj}(\mathbf{A})\mathbf{A}$ are thus all equal to $\det \mathbf{A}$. For $1 \leqslant i,j \leqslant n$ with $i \neq j$ we have

$$[\mathrm{Adj}(\mathbf{A})\mathbf{A}]_{ij} = \sum_{k=1}^{n} (-1)^{i+k} \left(\det \mathbf{A}^{(k,i)}\right) A_{kj}.$$

We would like to interpret this last expression as a Laplace expansion. We consider a new matrix $\hat{\mathbf{A}} = (\hat{A}_{ij})_{1 \leqslant i,j \leqslant n}$ which is identical to $\mathbf{A}$, except that the i-th column of $\mathbf{A}$ is replaced with the j-th column of $\mathbf{A}$, that is, for $1 \leqslant k \leqslant n$, we have

$$\hat{A}_{kl} = \begin{cases} A_{kj}, & l = i, \\ A_{kl}, & l \neq i. \end{cases} \tag{5.9}$$

Then, for all $1 \leqslant k \leqslant n$ we have $\hat{\mathbf{A}}^{(k,i)} = \mathbf{A}^{(k,i)}$, since the only column in which $\mathbf{A}$ and $\hat{\mathbf{A}}$ are different is removed in $\mathbf{A}^{(k,i)}$. Using (5.9), the Laplace expansion of $\hat{\mathbf{A}}$ with respect to the i-th column gives

$$\begin{aligned} \det \hat{\mathbf{A}} &= \sum_{k=1}^{n} (-1)^{(i+k)} \hat{A}_{ki} \det\left(\hat{\mathbf{A}}^{(k,i)}\right) = \sum_{k=1}^{n} (-1)^{i+k} \left(\det \mathbf{A}^{(k,i)}\right) A_{kj} \\ &= [\mathrm{Adj}(\mathbf{A})\mathbf{A}]_{ij} \end{aligned}$$

The matrix $\hat{\mathbf{A}}$ has a double occurrence of the i-th column, hence its column vectors are linearly dependent. Therefore $\hat{\mathbf{A}}$ is not invertible by Proposition 4.7 and so $\det \hat{\mathbf{A}} = [\mathrm{Adj}(\mathbf{A})\mathbf{A}]_{ij} = 0$ by Corollary 5.22. The off-diagonal entries of $\mathrm{Adj}(\mathbf{A})\mathbf{A}$ are thus all zero and we conclude $\mathrm{Adj}(\mathbf{A})\mathbf{A} = \det(\mathbf{A})\mathbf{1}_n$. Using the row version of the Laplace expansion we can conclude analogously that $\mathbf{A}\,\mathrm{Adj}(\mathbf{A}) = \det(\mathbf{A})\mathbf{1}_n$.

Finally, if $\mathbf{A}$ is invertible, then $\det \mathbf{A} \neq 0$ by Corollary 5.22, so that $\mathbf{A}^{-1} = \mathrm{Adj}(\mathbf{A})/\det(\mathbf{A})$, as claimed. □

As a corollary we obtain:

Corollary 5.46 *Let $n \in \mathbb{N}$ and $\mathbf{A} \in M_{n,n}(\mathbb{K})$ be an invertible upper triangular matrix. Then $\mathbf{A}^{-1}$ is also an upper triangular matrix.*

Remark 5.47 Taking the transpose also implies: Let $\mathbf{A} \in M_{n,n}(\mathbb{K})$ be an invertible lower triangular matrix. Then $\mathbf{A}^{-1}$ is also a lower triangular matrix.

Proof of Corollary 5.46 Write $\mathbf{A} = (A_{ij})_{1 \leqslant i,j \leqslant n}$. Using Theorem 5.45 it suffices to show that $\mathrm{Adj}(\mathbf{A})$ is an upper triangular matrix. If $\mathbf{A}$ is an upper triangular matrix, then $A_{ij} = 0$ for all $i > j$. By definition we have

$$[\mathrm{Adj}(\mathbf{A})]_{ij} = (-1)^{i+j} \det\left(\mathbf{A}^{(j,i)}\right), \qquad 1 \leqslant i,j \leqslant n.$$

Notice that for $i > j$ every element below the diagonal of $\mathbf{A}^{(j,i)}$ is also below the diagonal of $\mathbf{A}$ and hence must be zero. It follows that $\mathbf{A}^{(j,i)}$ is an upper triangular matrix as well. Proposition 5.24 implies that the determinant of $\mathbf{A}^{(j,i)}$ is the product of its diagonal entries. Since $\mathbf{A}^{(j,i)}$ arises from the upper triangular matrix $\mathbf{A}$ by removing a row and a column, at least one of the diagonal entries of $\mathbf{A}^{(j,i)}$ must be zero and thus $\det \mathbf{A}^{(j,i)} = 0$ for $i > j$. We conclude that $\mathbf{A}^{-1}$ is an upper triangular matrix as well. □

We now use Theorem 5.45 to obtain a formula for the solution of the linear system $\mathbf{A}\vec{x} = \vec{b}$ for an invertible matrix $\mathbf{A}$. Multiplying from the left with $\mathbf{A}^{-1}$, we get

$$\vec{x} = \mathbf{A}^{-1}\vec{b} = \frac{1}{\det \mathbf{A}} \operatorname{Adj}(\mathbf{A})\vec{b}.$$

Writing $\vec{x} = (x_i)_{1\leqslant i\leqslant n}$, multiplication with $\det \mathbf{A}$ gives for $1 \leqslant i \leqslant n$

$$x_i \det \mathbf{A} = \sum_{k=1}^{n} [\operatorname{Adj}(\mathbf{A})]_{ik} b_k = \sum_{k=1}^{n} (-1)^{i+k} \det\left(\mathbf{A}^{(k,i)}\right) b_k.$$

We can again interpret the right hand side as a Laplace expansion of the matrix $\hat{\mathbf{A}}_i$ obtained by replacing the i-th column of $\mathbf{A}$ with $\vec{b}$ and leaving $\mathbf{A}$ unchanged otherwise. Hence, we have for all $1 \leqslant i \leqslant n$

$$x_i = \frac{\det \hat{\mathbf{A}}_i}{\det \mathbf{A}}.$$

This formula is known as *Cramer's rule*. While this is a neat formula, it is rarely used in computing solutions to linear systems of equations due to the complexity of computing determinants.

Example 5.48 (Cramer's rule) We consider the system $\mathbf{A}\vec{x} = \vec{b}$ for

$$\mathbf{A} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} \quad \text{and} \quad \vec{b} = \begin{pmatrix} -2 \\ 2 \\ 4 \end{pmatrix}.$$

Here we obtain

$$\hat{\mathbf{A}}_1 = \begin{pmatrix} -2 & 1 & 1 \\ 2 & 2 & 1 \\ 4 & 1 & 2 \end{pmatrix}, \quad \hat{\mathbf{A}}_2 = \begin{pmatrix} 2 & -2 & 1 \\ 1 & 2 & 1 \\ 1 & 4 & 2 \end{pmatrix}, \quad \hat{\mathbf{A}}_3 = \begin{pmatrix} 2 & 1 & -2 \\ 1 & 2 & 2 \\ 1 & 1 & 4 \end{pmatrix}.$$

We compute $\det \mathbf{A} = 4$, $\det \hat{\mathbf{A}}_1 = -12$, $\det \hat{\mathbf{A}}_2 = 4$ and $\det \hat{\mathbf{A}}_3 = 12$ so that Cramer's rule gives indeed the correct solution

$$\vec{x} = \frac{1}{4}\begin{pmatrix} -12 \\ 4 \\ 12 \end{pmatrix} = \begin{pmatrix} -3 \\ 1 \\ 3 \end{pmatrix}.$$

Exercises

Exercise 5.49 Use the Leibniz formula to show that

$$\det(\mathbf{A}) = \det(\mathbf{A}^T)$$

for all $\mathbf{A} \in M_{n,n}(\mathbb{K})$.

CHAPTER 6

Endomorphisms

6.1 Sums, direct sums and complements

WEEK 10

In this chapter we study linear mappings from a vector space to itself.

Definition 6.1 (Endomorphism — Video) A linear map $g : V \to V$ from a $\mathbb{K}$-vector space V to itself is called an *endomorphism*. An endomorphism that is also an isomorphism is called an *automorphism*.

Before we develop the theory of endomorphisms, we introduce some notions for subspaces.

Definition 6.2 (Sum of subspaces — Video) Let V be a $\mathbb{K}$-vector space, $n \in \mathbb{N}$ and $U_1, \ldots, U_n$ be vector subspaces of V. The set

$$\sum_{i=1}^{n} U_i = U_1 + U_2 + \cdots + U_n = \{v \in V | v = u_1 + u_2 + \cdots + u_n \text{ for } u_i \in U_i\}$$

is called the *sum of the subspaces* U_i.

Recall that by Proposition 3.27, the intersection of two subspaces is again a subspace, whereas the union of two subspaces fails to be a subspace in general. However, subspaces do behave nicely with regards to sums:

Proposition 6.3 *The sum of the subspaces $U_i \subset V$, $i = 1 \ldots, n$ is again a vector subspace.*

Proof The sum $\sum_{i=1}^{n} U_i$ is non-empty, since it contains the zero vector 0_V. Let v and $v' \in \sum_{i=1}^{n} U_i$ so that

$$v = v_1 + v_2 + \cdots + v_n \qquad \text{and} \qquad v' = v_1' + v_2' + \cdots + v_n'$$

for vectors $v_i, v_i' \in U_i$, $i = 1, \ldots, n$. Each U_i is a vector subspace of V. Therefore, for all scalars $s, t \in \mathbb{K}$, the vector $sv_i + tv_i'$ is an element of U_i, $i = 1, \ldots, n$. Thus

$$sv + tv' = sv_1 + tv_1' + \cdots + sv_n + tv_n'$$

is an element of $U_1 + \cdots + U_n$. By Definition 3.21, it follows that $U_1 + \cdots + U_n$ is a vector subspace of V. □

Remark 6.4 Notice that $U_1 + \cdots + U_n$ is the smallest vector subspace of V containing all vector subspaces U_i, $i = 1, \ldots, n$.

If each vector in the sum is in a unique way the sum of vectors from the subspaces we say the subspaces are in direct sum:

Definition 6.5 (Direct sum of subspaces) Let V be a $\mathbb{K}$-vector space, $n \in \mathbb{N}$ and $U_1, \dots, U_n$ be vector subspaces of V. The subspaces $U_1, \dots, U_n$ are said to be in *direct sum* if each vector $w \in W = U_1 + \cdots + U_n$ is in a unique way the sum of vectors $v_i \in U_i$ for $1 \leqslant i \leqslant n$. That is, if $w = v_1 + v_2 + \cdots + v_n = v_1' + v_2' + \cdots + v_n'$ for vectors $v_i, v_i' \in U_i$, then $v_i = v_i'$ for all $1 \leqslant i \leqslant n$. We write

$$\bigoplus_{i=1}^{n} U_i$$

in case the subspaces $U_1, \dots, U_n$ are in direct sum.

Example 6.6 Let $n \in \mathbb{N}$ and $V = \mathbb{K}^n$ as well as $U_i = \operatorname{span}\{\vec{e}_i\}$, where $\{\vec{e}_1, \dots, \vec{e}_n\}$ denotes the standard basis of $\mathbb{K}^n$. Then $\mathbb{K}^n = \bigoplus_{i=1}^n U_i$.

Remark 6.7

(i) Two subspaces U_1, U_2 of V are in direct sum if and only if $U_1 \cap U_2 = \{0_V\}$. Indeed, suppose $U_1 \cap U_2 = \{0_V\}$ and consider $w = v_1 + v_2 = v_1' + v_2'$ with $v_i, v_i' \in U_i$ for $i = 1, 2$. We then have $v_1 - v_1' = v_2' - v_2 \in U_2$, since U_2 is a subspace. Since U_1 is a subspace as well, we also have $v_1 - v_1' \in U_1$. Since $v_1 - v_1'$ lies both in U_1 and U_2, we must have $v_1 - v_1' = 0_V = v_2' - v_2$. Conversely, suppose U_1, U_2 are in direct sum and let $w \in (U_1 \cap U_2)$. We can write $w = w + 0_V = 0_V + w$, since $w \in U_1$ and $w \in U_2$. Since U_1, U_2 are in direct sum, we must have $w = 0_V$, hence $U_1 \cap U_2 = \{0_V\}$.

(ii) Observe that if the subspaces $U_1, \dots, U_n$ are in direct sum and $v_i \in U_i$ with $v_i \neq 0_V$ for $1 \leqslant i \leqslant n$, then the vectors $\{v_1, \dots, v_n\}$ are linearly independent. Indeed, if $s_1, \dots, s_n$ are scalars such that

$$s_1 v_1 + s_2 v_2 + \cdots + s_n v_n = 0_V = 0_V + 0_V + \cdots + 0_V,$$

then $s_i v_i = 0_V$ for all $1 \leqslant i \leqslant n$. By assumption $v_i \neq 0_V$ and hence $s_i = 0$ for all $1 \leqslant i \leqslant n$.

Proposition 6.8 *Let $n \in \mathbb{N}$, V be a finite dimensional $\mathbb{K}$-vector space and $U_1, \dots, U_n$ be subspaces of V. Let $\mathbf{b}_i$ be an ordered basis of U_i for $1 \leqslant i \leqslant n$. Then we have:*

(i) The tuple of vectors obtained by listing all the vectors of the bases $\mathbf{b}_i$ is a basis of V if and only if $V = \bigoplus_{i=1}^n U_i$.

(ii) $\dim(U_1 + \cdots + U_n) \leqslant \dim(U_1) + \cdots + \dim(U_n)$ with equality if and only if the subspaces $U_1, \dots, U_n$ are in direct sum.

Proof Part of an exercise. □

Definition 6.9 (Complement to a subspace) Let V be a $\mathbb{K}$-vector space and $U \subset V$ a subspace. A subspace U' of V such that $V = U \oplus U'$ is called a *complement to* U.

Example 6.10 Notice that a complement need not be unique. Consider $V = \mathbb{R}^2$ and $U = \operatorname{span}\{\vec{e}_1\}$. Let $v \in V$. Then the subspace $U' = \operatorname{span}\{v\}$ is a complement to U, provided $\vec{e}_1, \vec{v}$ are linearly independent.

Corollary 6.11 (Existence of a complement) *Let U be a subspace of a finite dimensional $\mathbb{K}$-vector space V. Then there exists a subspace U' so that $V = U \oplus U'$.*

Proof Suppose $(v_1, \dots, v_m)$ is an ordered basis of U. By Theorem 3.64, there exists a basis $\{v_1, \dots, v_m, v_{m+1}, \dots, v_n\}$ of V. Defining $U' = \operatorname{span}\{v_{m+1}, \dots, v_n\}$, Proposition 6.8 implies the claim. □

The dimension of a sum of two subspaces equals the sum of the dimensions of the subspaces minus the dimension of the intersection:

Proposition 6.12 *Let V be a finite dimensional $\mathbb{K}$-vector space and U_1, U_2 subspaces of V. Then we have*

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

Proof Let $r = \dim(U_1 \cap U_2)$ and let $\{u_1, \dots, u_r\}$ be a basis of $U_1 \cap U_2$. These vectors are linearly independent and elements of U_1, hence by Theorem 3.64, there exist vectors $v_1, \dots, v_{m-r}$ so that $\mathcal{S}_1 = \{u_1, \dots, u_r, v_1, \dots, v_{m-r}\}$ is a basis of U_1. Likewise there exist vectors $w_1, \dots, w_{n-r}$ such that $\mathcal{S}_2 = \{u_1, \dots, u_r, w_1, \dots, w_{n-r}\}$ is a basis of U_2. Here $m = \dim U_1$ and $n = \dim U_2$.

Now consider the set $\mathcal{S} = \{u_1, \dots, u_r, v_1, \dots, v_{m-r}, w_1, \dots, w_{n-r}\}$ consisting of $r + m - r + n - r = n + m - r$ vectors. If this set is a basis of $U_1 + U_2$, then the claim follows, since then $\dim(U_1 + U_2) = n + m - r = \dim(U_1) + \dim(U_2) - \dim(U_1 \cap U_2)$.

We first show that $\mathcal{S}$ generates $U_1 + U_2$. Let $y \in U_1 + U_2$ so that $y = x_1 + x_2$ for vectors $x_1 \in U_1$ and $x_2 \in U_2$. Since $\mathcal{S}_1$ is a basis of U_1, we can write x_1 as a linear combination of elements of $\mathcal{S}_1$. Likewise we can write x_2 as a linear combination of elements of $\mathcal{S}_2$. It follows that $\mathcal{S}$ generates $U_1 + U_2$.

We need to show that $\mathcal{S}$ is linearly independent. So suppose we have scalars $s_1, \dots, s_r$, $t_1, \dots, t_{m-r}$, and $r_1, \dots, r_{n-r}$, so that

$$\underbrace{s_1 u_1 + \cdots + s_r u_r}_{=u} + \underbrace{t_1 v_1 + \cdots + t_{m-r} v_{m-r}}_{=v} + \underbrace{r_1 w_1 + \cdots + r_{n-r} w_{n-r}}_{=w} = 0_V.$$

Equivalently, $w = -u - v$ so that $w \in U_1$. Since w is a linear combination of elements of $\mathcal{S}_2$, we also have $w \in U_2$. Therefore, $w \in U_1 \cap U_2$ and there exist scalars $\hat{s}_1, \dots, \hat{s}_r$ such that

$$w = \underbrace{\hat{s}_1 u_1 + \cdots + \hat{s}_r u_r}_{=\hat{u}}$$

This is equivalent to $w - \hat{u} = 0_V$, or written out

$$r_1 w_1 + \cdots + r_{n-r} w_{n-r} - \hat{s}_1 u_1 - \cdots + \hat{s}_r u_r = 0_V.$$

Since the vectors $\{u_1, \dots, u_r, w_1, \dots, w_{n-r}\}$ are linearly independent, we conclude that $r_1 = \cdots = r_{n-r} = \hat{s}_1 = \cdots = \hat{s}_r = 0$. It follows that $w = 0_V$ and hence $u + v = 0_V$. Again, since $\{u_1, \dots, u_r, v_1, \dots, v_{n-r}\}$ are linearly independent, we conclude that $s_1 = \cdots = s_r = t_1 = \cdots = t_{m-r} = 0$ and we are done. □

6.2 Invariants of endomorphisms

Let V be a finite dimensional vector space equipped with an ordered basis $\mathbf{b}$ and $g : V \to V$ an endomorphism. Recall from Theorem 3.106 that if we consider another ordered basis $\mathbf{b}'$ of V, then

$$\mathbf{M}(g, \mathbf{b}', \mathbf{b}') = \mathbf{C}\,\mathbf{M}(g, \mathbf{b}, \mathbf{b})\,\mathbf{C}^{-1},$$

where we write $\mathbf{C} = \mathbf{C}(\mathbf{b}, \mathbf{b}')$ for the change of basis matrix. This motivates the following definition:

Definition 6.13 (Similar / conjugate matrices) Let $n \in \mathbb{N}$ and $\mathbf{A}, \mathbf{A}' \in M_{n,n}(\mathbb{K})$. The matrices $\mathbf{A}$ and $\mathbf{A}'$ are called *similar or conjugate over* $\mathbb{K}$ if there exists an invertible matrix $\mathbf{C} \in M_{n,n}(\mathbb{K})$ such that

$$\mathbf{A}' = \mathbf{C}\mathbf{A}\mathbf{C}^{-1}.$$

Similarity of matrices over $\mathbb{K}$ is an *equivalence relation*:

Proposition 6.14 *Let $n \in \mathbb{N}$ and $\mathbf{A}, \mathbf{B}, \mathbf{X} \in M_{n,n}(\mathbb{K})$. Then we have*

(i) $\mathbf{A}$ is similar to itself;
(ii) $\mathbf{A}$ is similar to $\mathbf{B}$ then $\mathbf{B}$ is similar to $\mathbf{A}$;
(iii) If $\mathbf{A}$ is similar to $\mathbf{B}$ and $\mathbf{B}$ is similar to $\mathbf{X}$, then $\mathbf{A}$ is also similar to $\mathbf{X}$.

Proof (i) We take $\mathbf{C} = \mathbf{1}_n$.

(ii) Suppose $\mathbf{A}$ is similar to $\mathbf{B}$ so that $\mathbf{B} = \mathbf{C}\mathbf{A}\mathbf{C}^{-1}$ for some invertible matrix $\mathbf{C} \in M_{n,n}(\mathbb{K})$. Multiplying with $\mathbf{C}^{-1}$ from the left and $\mathbf{C}$ from the right, we get

$$\mathbf{C}^{-1}\mathbf{B}\mathbf{C} = \mathbf{C}^{-1}\mathbf{C}\mathbf{A}\mathbf{C}^{-1}\mathbf{C} = \mathbf{A},$$

so that the similarity follows for the choice $\hat{\mathbf{C}} = \mathbf{C}^{-1}$.

(iii) We have $\mathbf{B} = \mathbf{C}\mathbf{A}\mathbf{C}^{-1}$ and $\mathbf{X} = \mathbf{D}\mathbf{B}\mathbf{D}^{-1}$ for invertible matrices $\mathbf{C}, \mathbf{D}$. Then we get

$$\mathbf{X} = \mathbf{D}\mathbf{C}\mathbf{A}\mathbf{C}^{-1}\mathbf{D}^{-1},$$

so that the similarity follows for the choice $\hat{\mathbf{C}} = \mathbf{D}\mathbf{C}$. □

Remark 6.15

- Because of (ii) in particular, one can say that two matrices $\mathbf{A}$ and $\mathbf{B}$ are similar without ambiguity.
- Theorem 3.106 shows that $\mathbf{A}$ and $\mathbf{B}$ are similar if and only if there exists an endomorphism g of $\mathbb{K}^n$ such that $\mathbf{A}$ and $\mathbf{B}$ represent g with respect to two ordered bases of $\mathbb{K}^n$.

One might wonder whether there exist functions $f : M_{n,n}(\mathbb{K}) \to \mathbb{K}$ which are *invariant* under conjugation, that is, f satisfies $f(\mathbf{C}\mathbf{A}\mathbf{C}^{-1}) = f(\mathbf{A})$ for all $\mathbf{A} \in M_{n,n}(\mathbb{K})$ and all invertible matrices $\mathbf{C} \in M_{n,n}(\mathbb{K})$. We have already seen an example of such a function, namely the determinant. Indeed using the product rule Proposition 5.21 and Corollary 5.22, we compute

$$\begin{aligned}\det\left(\mathbf{C}\mathbf{A}\mathbf{C}^{-1}\right) &= \det(\mathbf{C}\mathbf{A})\det\left(\mathbf{C}^{-1}\right) = \det(\mathbf{C})\det(\mathbf{A})\det\left(\mathbf{C}^{-1}\right)\\ &= \det(\mathbf{A}).\end{aligned} \tag{6.1}$$

Because of this fact, the following definition makes sense:

Definition 6.16 (Determinant of an endomorphism) Let V be a finite dimensional $\mathbb{K}$-vector space and $g : V \to V$ an endomorphism. We define

$$\det(g) = \det(\mathbf{M}(g, \mathbf{b}, \mathbf{b}))$$

where $\mathbf{b}$ is any ordered basis of V. By Theorem 3.106 and (6.1), the scalar $\det(g)$ is independent of the chosen ordered basis.

Another example of a scalar that we can associate to an endomorphism is the so-called *trace*. Like for the determinant, we first define the trace for matrices. Luckily, the trace is a lot simpler to define:

Definition 6.17 (Trace of a matrix) Let $n \in \mathbb{N}$ and $\mathbf{A} \in M_{n,n}(\mathbb{K})$. The sum $\sum_{i=1}^{n} [\mathbf{A}]_{ii}$ of its diagonal entries is called the *trace of* $\mathbf{A}$ and denoted by $\operatorname{Tr}(\mathbf{A})$ or $\operatorname{Tr}\mathbf{A}$.

Example 6.18 For all $n \in \mathbb{N}$ we have $\operatorname{Tr}(\mathbf{1}_n) = n$. For

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

we have $\operatorname{Tr}(\mathbf{A}) = 2 + 2 + 3 = 7$.

The trace of a product of square matrices is independent of the order of multiplication:

Proposition 6.19 *Let $n \in \mathbb{N}$ and $\mathbf{A}, \mathbf{B} \in M_{n,n}(\mathbb{K})$. Then we have*

$$\operatorname{Tr}(\mathbf{AB}) = \operatorname{Tr}(\mathbf{BA}).$$

Proof Let $\mathbf{A} = (A_{ij})_{1 \leqslant i,j \leqslant n}$ and $\mathbf{B} = (B_{ij})_{1 \leqslant i,j \leqslant n}$. Then

$$[\mathbf{AB}]_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj} \qquad \text{and} \qquad [\mathbf{BA}]_{kj} = \sum_{i=1}^{n} B_{ki} A_{ij},$$

so that

$$\operatorname{Tr}(\mathbf{AB}) = \sum_{i=1}^{n} \sum_{k=1}^{n} A_{ik} B_{ki} = \sum_{k=1}^{n} \sum_{i=1}^{n} B_{ki} A_{ik} = \operatorname{Tr}(\mathbf{BA}).$$

□

Using the previous proposition, we obtain

$$\operatorname{Tr}\left(\mathbf{CAC}^{-1}\right) = \operatorname{Tr}\left(\mathbf{AC}^{-1}\mathbf{C}\right) = \operatorname{Tr}(\mathbf{A}). \tag{6.2}$$

As for the determinant, the following definition thus makes sense:

Definition 6.20 (Trace of an endomorphism) Let V be a finite dimensional $\mathbb{K}$-vector space and $g : V \to V$ an endomorphism. We define

$$\operatorname{Tr}(g) = \operatorname{Tr}(\mathbf{M}(g, \mathbf{b}, \mathbf{b}))$$

where **b** is any ordered basis of V. By Theorem 3.106 and (6.2), the scalar $\mathrm{Tr}(g)$ is independent of the chosen ordered basis.

The trace and determinant of endomorphisms behave nicely with respect to composition of maps:

Proposition 6.21 *Let V be a finite dimensional $\mathbb{K}$-vector space. Then, for all endomorphisms $f, g : V \to V$ we have*

(i) $\mathrm{Tr}(f \circ g) = \mathrm{Tr}(g \circ f)$;

(ii) $\det(f \circ g) = \det(f)\det(g)$.

Proof (i) Fix an ordered basis **b** of V. Then, using Corollary 3.100 and Proposition 6.19, we obtain

$$\begin{aligned}\mathrm{Tr}(f \circ g) &= \mathrm{Tr}\left(\mathbf{M}(f \circ g, \mathbf{b}, \mathbf{b})\right) = \mathrm{Tr}\left(\mathbf{M}(f, \mathbf{b}, \mathbf{b})\mathbf{M}(g, \mathbf{b}, \mathbf{b})\right)\\ &= \mathrm{Tr}\left(\mathbf{M}(g, \mathbf{b}, \mathbf{b})\mathbf{M}(f, \mathbf{b}, \mathbf{b})\right) = \mathrm{Tr}\left(\mathbf{M}(g \circ f, \mathbf{b}, \mathbf{b})\right) = \mathrm{Tr}(g \circ f).\end{aligned}$$

The proof of (ii) is analogous, but we use Proposition 5.21 instead of Proposition 6.19. □

We also have:

Proposition 6.22 *Let V be a finite dimensional $\mathbb{K}$-vector space and $g : V \to V$ an endomorphism. Then the following statements are equivalent:*

(i) *g is injective;*

(ii) *g is surjective;*

(iii) *g is bijective;*

(iv) $\det(g) \neq 0$.

Proof The equivalence of the first three statements follows from Corollary 3.77. We fix an ordered basis **b** of V. Suppose g is bijective with inverse $g^{-1} : V \to V$. Then we have

$$\det(g \circ g^{-1}) = \det(g)\det\left(g^{-1}\right) = \det\left(\mathrm{Id}_V\right) = \det\left(\mathbf{M}(\mathrm{Id}_V, \mathbf{b}, \mathbf{b})\right) = \det\left(\mathbf{1}_{\dim V}\right) = 1.$$

It follows that $\det(g) \neq 0$ and moreover that

$$\det\left(g^{-1}\right) = \frac{1}{\det g}.$$

Conversely, suppose that $\det g \neq 0$. Then $\det \mathbf{M}(g, \mathbf{b}, \mathbf{b}) \neq 0$ so that $\mathbf{M}(g, \mathbf{b}, \mathbf{b})$ is invertible by Corollary 5.22 and Proposition 3.101 implies that g is bijective. □

Remark 6.23 Notice that Proposition 6.22 is wrong for infinite dimensional vector spaces. Consider $V = \mathbb{K}^\infty$, the $\mathbb{K}$-vector space of sequences from Example 3.6. The endomorphism $g : V \to V$ defined by $(x_1, x_2, x_3, \ldots) \mapsto (0, x_1, x_2, x_3, \ldots)$ is injective but not surjective.

6.3 Eigenvectors and eigenvalues

WEEK 11

Mappings g that have the same domain and codomain allow for the notion of a fixed point. Recall that an element x of a set $\mathcal{X}$ is called a *fixed point* of a mapping $g : \mathcal{X} \to \mathcal{X}$ if $g(x) = x$, that is, x agrees with its image under g. In Linear Algebra, a generalisation of the notion of a fixed point is that of an eigenvector. A vector $v \in V$ is called an *eigenvector* of the linear map $g : V \to V$ if v is merely scaled when applying g to v, that is, there exists a scalar $\lambda \in \mathbb{K}$ – called *eigenvalue* – such that $g(v) = \lambda v$. Clearly, the zero vector 0_V will satisfy this condition for every choice of scalar λ. For this reason, eigenvectors are usually required to be different from the zero vector. In this terminology, fixed points v of g are simply eigenvectors with eigenvalue 1, since they satisfy $g(v) = v = 1v$.

It is natural to ask whether a linear map $g : V \to V$ always admits an eigenvector. In the remaining part of this chapter we will answer this question and further develop our theory of linear maps, specifically endomorphisms. We start with some precise definitions.

Definition 6.24 (Eigenvector, eigenspace, eigenvalue — Video) Let $g : V \to V$ be an endomorphism of a $\mathbb{K}$-vector space V.

- An *eigenvector* with *eigenvalue* $\lambda \in \mathbb{K}$ is a *non-zero* vector $v \in V$ such that $g(v) = \lambda v$.
- If $\lambda \in \mathbb{K}$ is an eigenvalue of g, the λ-*eigenspace* $\operatorname{Eig}_g(\lambda)$ is the subspace of vectors $v \in V$ satisfying $g(v) = \lambda v$.
- The dimension of $\operatorname{Eig}_g(\lambda)$ is called the *geometric multiplicity* of the eigenvalue λ.
- The set of all eigenvalues of g is called the *spectrum of g*.
- For $\mathbf{A} \in M_{n,n}(\mathbb{K})$ we speak of eigenvalues, eigenvectors, eigenspaces and spectrum to mean those of the endomorphism $f_\mathbf{A} : \mathbb{K}^n \to \mathbb{K}^n$.

Remark 6.25 By definition, the zero vector 0_V is not an eigenvector, it is however an element of the eigenspace $\operatorname{Eig}_g(\lambda)$ for every eigenvalue λ.

Example 6.26

(i) The scalar 0 is an eigenvalue of an endomorphism $g : V \to V$ if and only if the kernel of g is different from $\{0_V\}$. In the case where the kernel of f does not only consist of the zero vector, we have $\operatorname{Ker} g = \operatorname{Eig}_g(0)$ and the geometric multiplicity of 0 is the nullity of g.

(ii) The endomorphism $f_\mathbf{D} : \mathbb{K}^n \to \mathbb{K}^n$ associated to a diagonal matrix with distinct diagonal entries

$$\mathbf{D} = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$$

has spectrum $\{\lambda_1, \dots, \lambda_n\}$ and corresponding eigenspaces $\operatorname{Eig}_{f_\mathbf{D}}(\lambda_i) = \operatorname{span}\{\vec{e}_i\}$.

(iii) Consider the $\mathbb{R}$-vector space $\mathsf{P}(\mathbb{R})$ of polynomials and $f = \frac{\mathrm{d}}{\mathrm{d}x} : \mathsf{P}(\mathbb{R}) \to \mathsf{P}(\mathbb{R})$ the derivative by the variable x. The kernel of f consists of the constant polynomials and hence 0 is an eigenvalue for f. For any non-zero scalar λ we cannot have polynomials p satisfying $\frac{\mathrm{d}}{\mathrm{d}x}p = \lambda p$, as the left hand of this last expression has a smaller degree than the right hand side.

Previously we defined the trace and determinant for an endomorphism $g : V \to V$ by observing that the trace and determinant of the matrix representation of g are independent of the chosen basis of V. Similarly, we can consider eigenvalues of g and eigenvalues of the matrix representation of g with respect to some ordered basis of V. Perhaps unsurprisingly, the eigenvalues are the same:

Proposition 6.27 *Let $g : V \to V$ be an endomorphism of a finite dimensional $\mathbb{K}$-vector space V. Let $\mathbf{b}$ be an ordered basis of V with corresponding linear coordinate system $\boldsymbol{\beta}$. Then $v \in V$ is an eigenvector of g with eigenvalue $\lambda \in \mathbb{K}$ if and only if $\boldsymbol{\beta}(v) \in \mathbb{K}^n$ is an eigenvector with eigenvalue λ of $\mathbf{M}(g, \mathbf{b}, \mathbf{b})$. In particular, conjugate matrices have the same eigenvalues.*

Proof Write $\mathbf{A} = \mathbf{M}(g, \mathbf{b}, \mathbf{b})$. Recall that by an eigenvector of $\mathbf{A} \in M_{n,n}(\mathbb{K})$, we mean an eigenvector of $f_\mathbf{A} : \mathbb{K}^n \to \mathbb{K}^n$. By Definition 3.91, we have $f_\mathbf{A} = \boldsymbol{\beta} \circ g \circ \boldsymbol{\beta}^{-1}$. Suppose $\lambda \in \mathbb{K}$ is an eigenvalue of g so that $g(v) = \lambda v$ for some non-zero vector $v \in V$. Consider the vector $\vec{x} = \boldsymbol{\beta}(v) \in \mathbb{K}^n$ which is non-zero, since $\boldsymbol{\beta} : V \to \mathbb{K}^n$ is an isomorphism. Then

$$f_\mathbf{A}(\vec{x}) = \boldsymbol{\beta}(g(\boldsymbol{\beta}^{-1}(\vec{x}))) = \boldsymbol{\beta}(g(v)) = \boldsymbol{\beta}(\lambda v) = \lambda\boldsymbol{\beta}(v) = \lambda\vec{x},$$

so that $\vec{x}$ is an eigenvector of $f_\mathbf{A}$ with eigenvalue λ.

Conversely, if λ is an eigenvalue of $f_\mathbf{A}$ with non-zero eigenvector $\vec{x}$, then it follows as above that $v = \boldsymbol{\beta}^{-1}(\vec{x}) \in V$ is an eigenvector of g with eigenvalue λ.

By Remark 6.15, if the matrices $\mathbf{A}, \mathbf{B}$ are similar, then they represent the same endomorphism $g : \mathbb{K}^n \to \mathbb{K}^n$ and hence have the same eigenvalues. □

The "nicest" endomorphisms are those for which there exists an ordered basis consisting of eigenvectors:

Definition 6.28 (Diagonalisable endomorphism)

- An endomorphism $g : V \to V$ is called *diagonalisable* if there exists an ordered basis $\mathbf{b}$ of V such that each element of $\mathbf{b}$ is an eigenvector of g.
- For $n \in \mathbb{N}$, a matrix $\mathbf{A} \in M_{n,n}(\mathbb{K})$ is called diagonalisable over $\mathbb{K}$ if the endomorphism $f_\mathbf{A} : \mathbb{K}^n \to \mathbb{K}^n$ is diagonalisable.

Example 6.29

(i) We consider $V = \mathsf{P}(\mathbb{R})$ and the endomorphism $g : V \to V$ which replaces the variable x with $2x$. For instance, we have

$$g(x^2 - 2x + 3) = (2x)^2 - 2(2x) + 3 = 4x^2 - 4x + 3.$$

Then g is diagonalisable. The vector space $\mathsf{P}(\mathbb{R})$ has an ordered basis $\mathbf{b} = (1, x, x^2, x^3, \ldots)$. Clearly, for all $k \in \mathbb{N} \cup \{0\}$ we have $g(x^k) = 2^k x^k$, so that x^k is an eigenvector of g with eigenvalue 2^k.

(ii) For $\alpha \in (0, \pi)$ consider

$$\mathbf{R}_\alpha = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}.$$

Recall that the endomorphism $f_{\mathbf{R}_\alpha} : \mathbb{R}^2 \to \mathbb{R}^2$ rotates vectors counter-clockwise around the origin $0_{\mathbb{R}^2}$ by the angle α. Since $\alpha \in (0, \pi)$, the endomorphism $f_{\mathbf{R}_\alpha}$ has no eigenvectors and hence is not diagonalisable.

Remark 6.30 Applying Proposition 6.27, we conclude that in the case of a finite dimensional $\mathbb{K}$-vector space V, an endomorphism $g : V \to V$ is diagonalisable if and only if there exists an ordered basis $\mathbf{b}$ of V such that $\mathbf{M}(g, \mathbf{b}, \mathbf{b})$ is a diagonal matrix. Moreover, $\mathbf{A} \in M_{n,n}(\mathbb{K})$ is diagonalisable if and only if $\mathbf{A}$ is similar over $\mathbb{K}$ to a diagonal matrix.

Recall, if $\mathcal{X}, \mathcal{Y}$ are sets, $f : \mathcal{X} \to \mathcal{Y}$ a mapping and $\mathcal{Z} \subset \mathcal{X}$ a subset of $\mathcal{X}$, we can consider the *restriction of f to $\mathcal{Z}$*, usually denoted by $f|_{\mathcal{Z}}$, which is the mapping

$$f|_{\mathcal{Z}} : \mathcal{Z} \to \mathcal{Y}, \quad z \mapsto f(z).$$

So we simply take the same mapping f, but apply it to the elements of the subset only.

Closely related to the notion of an eigenvector is that of a stable subspace. Let $v \in V$ be an eigenvector with eigenvalue λ of the endomorphism $g : V \to V$. The 1-dimensional subspace $U = \operatorname{span}\{v\}$ is stable under g, that is, $g(U) \subset U$. Indeed, since $g(v) = \lambda v$ and since every vector $u \in U$ can be written as $u = tv$ for some scalar $t \in \mathbb{K}$, we have $g(u) = g(tv) = tg(v) = t\lambda v \in U$. This motivates the following definition:

Definition 6.31 (Stable subspace) A subspace $U \subset V$ is called *stable or invariant under the endomorphism $g : V \to V$* if $g(U) \subset U$, that is $g(u) \in U$ for all vectors $u \in U$. In this case, the restriction $g|_U$ of g to U is an endomorphism of U.

Remark 6.32 Notice that a finite dimensional subspace $U \subset V$ is stable under g if and only if $g(v_i) \in U$ for $1 \leqslant i \leqslant m$, where $\{v_1, \ldots, v_m\}$ is a basis of U.

Example 6.33

(i) Every eigenspace of an endomorphism $g : V \to V$ is a stable subspace. By definition $g|_{\operatorname{Eig}_g(\lambda)} : \operatorname{Eig}_g(\lambda) \to \operatorname{Eig}_g(\lambda)$ is multiplication by the scalar $\lambda \in \mathbb{K}$.

(ii) We consider $V = \mathbb{R}^3$ and

$$\mathbf{R}_\alpha = \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

for $\alpha \in (0, \pi)$. The endomorphism $f_{\mathbf{R}_\alpha} : \mathbb{R}^3 \to \mathbb{R}^3$ is the rotation by the angle $\alpha \in \mathbb{R}$ around the axis spanned by $\vec{e}_3$. Then the plane $U = \{\vec{x} = (x_i)_{1 \leqslant i \leqslant 3} \in \mathbb{R}^3 | x_3 = 0\}$ is stable under $f = f_{\mathbf{R}_\alpha}$. Here $f|_\Pi : \Pi \to \Pi$ is the rotation in the plane U around the origin with angle α.

Moreover, the vector $\vec{e}_3$ is an eigenvector with eigenvalue 1 so that

$$\mathrm{Eig}_f(1) = \mathrm{span}\{\vec{e}_3\}.$$

(iii) We consider again the $\mathbb{R}$-vector space $\mathsf{P}(\mathbb{R})$ of polynomials and $f = \frac{\mathrm{d}}{\mathrm{d}x} : \mathsf{P}(\mathbb{R}) \to \mathsf{P}(\mathbb{R})$ the derivative by the variable x. For $n \in \mathbb{N}$ let U_n denote the subspace of polynomials of degree at most n. Since $U_{n-1} \subset U_n$, the subspace U_n is stable under f.

Stable subspaces correspond to zero blocks in the matrix representation of linear maps. More precisely:

Proposition 6.34 *Let V be a $\mathbb{K}$-vector space of dimension $n \in \mathbb{N}$ and $g : V \to V$ an endomorphism. Furthermore, let $U \subset V$ be a subspace of dimension $1 \leqslant m \leqslant n$ and $\mathbf{b}$ an ordered basis of U and $\mathbf{c} = (\mathbf{b}, \mathbf{b}')$ an ordered basis of V. Then U is stable under g if and only if the matrix $\mathbf{A} = \mathbf{M}(g, \mathbf{c}, \mathbf{c})$ has the form*

$$\mathbf{A} = \begin{pmatrix} \hat{\mathbf{A}} & * \\ \mathbf{0}_{n-m,m} & * \end{pmatrix}$$

for some matrix $\hat{\mathbf{A}} \in M_{m,m}(\mathbb{K})$. In the case where U is stable under g, we have $\hat{\mathbf{A}} = \mathbf{M}(g|_U, \mathbf{b}, \mathbf{b}) \in M_{m,m}(\mathbb{K})$.

Proof Write $\mathbf{b} = (v_1, \ldots, v_m)$ for vectors $v_i \in U$ and $\mathbf{b}' = (w_1, \ldots, w_{n-m})$ for vectors $w_i \in V$.

$\Rightarrow$ Since U is stable under g, we have $g(u) \in U$ for all vectors $u \in U$. Since $\mathbf{b}$ is a basis of U, there exist scalars $\hat{A}_{ij} \in \mathbb{K}$ with $1 \leqslant i, j \leqslant m$ such that

$$g(v_j) = \sum_{i=1}^{m} \hat{A}_{ij} v_i$$

for all $1 \leqslant j \leqslant m$. By Proposition 3.92, the matrix representation of g with respect to the ordered basis $\mathbf{c} = (\mathbf{b}, \mathbf{b}')$ of V thus takes the form

$$\mathbf{A} = \begin{pmatrix} \hat{\mathbf{A}} & * \\ \mathbf{0}_{n-m,m} & * \end{pmatrix}$$

where we write $\hat{\mathbf{A}} = (\hat{A}_{ij})_{1 \leqslant i,j \leqslant m} = \mathbf{M}(g|_U, \mathbf{b}, \mathbf{b})$.

$\Leftarrow$ Suppose

$$\mathbf{A} = \begin{pmatrix} \hat{\mathbf{A}} & * \\ \mathbf{0}_{n-m,m} & * \end{pmatrix} = \mathbf{M}(g, \mathbf{c}, \mathbf{c})$$

is the matrix representation of g with respect to the ordered basis $\mathbf{c}$ of V. Write $\hat{\mathbf{A}} = (\hat{A}_{ij})_{1 \leqslant i,j \leqslant m}$ Then, by Proposition 3.92, $g(v_j) = \sum_{i=1}^{m} \hat{A}_{ij} v_i \in U$ for all $1 \leqslant j \leqslant m$, hence U is stable under g, by Remark 6.32. □

From Proposition 6.34 we can conclude:

Remark 6.35 Suppose V is the direct sum of subspaces $U_1, U_2, \ldots, U_m$, all of which are stable under the endomorphism $g : V \to V$. If $\mathbf{b}_i$ is an ordered basis of U_i for $i = 1, \ldots, m$. Then the matrix representation of g with respect to the ordered basis

$\mathbf{c} = (\mathbf{b}_1, \dots, \mathbf{b}_m)$ takes the block form

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & & & \\ & \mathbf{A}_2 & & \\ & & \ddots & \\ & & & \mathbf{A}_m \end{pmatrix}$$

where $\mathbf{A}_i = \mathbf{M}(g|_{U_i}, \mathbf{b}_i, \mathbf{b}_i)$ for $i = 1, \dots, m$.

6.4 The characteristic polynomial

The eigenvalues of an endomorphism are the solutions of a polynomial equation:

Lemma 6.36 *Let V be a finite dimensional $\mathbb{K}$-vector space and $g : V \to V$ an endomorphism. Then $\lambda \in \mathbb{K}$ is an eigenvalue of g if and only if*

$$\det(\lambda \mathrm{Id}_V - g) = 0.$$

Moreover if λ is an eigenvalue of g, then $\mathrm{Eig}_g(\lambda) = \mathrm{Ker}(\lambda \mathrm{Id}_V - g)$.

Proof Let $v \in V$. We may write $v = \mathrm{Id}_V(v)$. Hence

$$g(v) = \lambda v \quad \iff \quad 0_V = (\lambda \mathrm{Id}_V - g)(v) \quad \iff v \in \mathrm{Ker}(\lambda \mathrm{Id}_V - g)$$

It follows that $\mathrm{Eig}_g(\lambda) = \mathrm{Ker}(\lambda \mathrm{Id}_V - g)$. Moreover $\lambda \in \mathbb{K}$ is an eigenvalue of g if and only if the kernel of $\lambda \mathrm{Id}_V - g$ is different from $\{0_V\}$ or if and only if $\lambda \mathrm{Id}_V - g$ is not injective. Proposition 6.22 implies that $\lambda \in \mathbb{K}$ is an eigenvalue of g if and only if $\det(\lambda \mathrm{Id}_V - g) = 0$. □

Definition 6.37 (Characteristic polynomial — Video) Let $g : V \to V$ be an endomorphism of a finite dimensional $\mathbb{K}$-vector space V. The function

$$\mathrm{char}_g : \mathbb{K} \to \mathbb{K}, \quad x \mapsto \det(x \mathrm{Id}_V - g)$$

is called the *characteristic polynomial of the endomorphism g*.

In practice, in order to compute the characteristic polynomial of an endomorphism $g : V \to V$, we choose an ordered basis $\mathbf{b}$ of V and compute the matrix representation $\mathbf{A} = \mathbf{M}(g, \mathbf{b}, \mathbf{b})$ of g with respect to $\mathbf{b}$. We then have

$$\mathrm{char}_g(x) = \det(x \mathbf{1}_n - \mathbf{A}).$$

By the characteristic polynomial of a matrix $\mathbf{A} \in M_{n,n}(\mathbb{K})$, we mean the characteristic polynomial of the endomorphism $f_{\mathbf{A}} : \mathbb{K}^n \to \mathbb{K}^n$, that is, the function $x \mapsto \det(x \mathbf{1}_n - \mathbf{A})$.

A *zero* of a polynomial $f : \mathbb{K} \to \mathbb{K}$ is a scalar $\lambda \in \mathbb{K}$ such that $f(\lambda) = 0$. The *multiplicity of a zero* λ is the largest integer $n \geqslant 1$ such that there exists a polynomial $\hat{f} : \mathbb{K} \to \mathbb{K}$ so that $f(x) = (x - \lambda)^n \hat{f}(x)$ for all $x \in \mathbb{K}$. Zeros are also known as *roots*.

Example 6.38 The polynomial $f(x) = x^3 - x^2 - 8x + 12$ can be factorised as $f(x) = (x - 2)^2(x + 3)$ and hence has zero 2 with multiplicity 2 and -3 with multiplicity 1.

Definition 6.39 (Algebraic multiplicity) Let λ be an eigenvalue of the endomorphism $g : V \to V$. The multiplicity of the zero λ of char_g is called the *algebraic multiplicity of* λ.

Example 6.40

(i) We consider

$$\mathbf{A} = \begin{pmatrix} 1 & 5 \\ 5 & 1 \end{pmatrix}.$$

Then

$$\begin{aligned}\mathrm{char}_{\mathbf{A}}(x) = \mathrm{char}_{f_{\mathbf{A}}}(x) = \det(x\mathbf{1}_2 - A) &= \det \begin{pmatrix} x-1 & -5 \\ -5 & x-1 \end{pmatrix} \\ &= (x-1)^2 - 25 = x^2 - 2x - 24 = (x+4)(x-6).\end{aligned}$$

Hence we have eigenvalues $\lambda_1 = 6$ and $\lambda_2 = -4$, both with algebraic multiplicity 1. By definition we have

$$\mathrm{Eig}_{\mathbf{A}}(6) = \mathrm{Eig}_{f_{\mathbf{A}}}(6) = \left\{\vec{v} \in \mathbb{K}^2 | \mathbf{A}\vec{v} = 6\vec{v}\right\}$$

and we compute that

$$\mathrm{Eig}_{\mathbf{A}}(6) = \mathrm{span}\left\{\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right\}$$

Since $\dim \mathrm{Eig}_{\mathbf{A}}(6) = 1$, the eigenvalue 6 has geometric multiplicity 1. Likewise we compute

$$\mathrm{Eig}_{\mathbf{A}}(-4) = \mathrm{span}\left\{\begin{pmatrix} -1 \\ 1 \end{pmatrix}\right\}$$

so that the eigenvalue -4 has geometric multiplicity 1 as well. Notice that we have an ordered basis of eigenvectors of **A** and hence **A** is diagonalisable, c.f. Example 3.96.

(ii) We consider

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

Then $\mathrm{char}_{\mathbf{A}}(x) = (x-2)^2$ so that we have a single eigenvalue 2 with algebraic multiplicity 2. We compute

$$\mathrm{Eig}_{\mathbf{A}}(2) = \mathrm{span}\left\{\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right\}$$

so that the eigenvalue 2 has geometric multiplicity 1. Notice that we cannot find an ordered basis consisting of eigenvectors, hence **A** is not diagonalisable.

The determinant and trace of an endomorphism do appear as coefficients in its characteristic polynomial:

Lemma 6.41 *Let $g : V \to V$ be an endomorphism of a $\mathbb{K}$-vector space V of dimension n. Then char_g is a polynomial of degree n and*

$$\mathrm{char}_g(x) = x^n - \mathrm{Tr}(g)x^{n-1} + \cdots + (-1)^n \det(g).$$

Proof We fix an ordered basis $\mathbf{b}$ of V. Writing $\mathbf{M}(g,\mathbf{b},\mathbf{b}) = \mathbf{A} = (A_{ij})_{1\leqslant i,j\leqslant n}$ and using the Leibniz formula (5.8), we have

$$\operatorname{char}_g(x) = \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^n B_{i\sigma(i)},$$

where

$$B_{ij} = \begin{cases} x - A_{ii}, & i = j, \\ -A_{ij}, & i \neq j. \end{cases}$$

Therefore, char_g is a finite sum of products containing x at most n times, hence char_g is a polynomial in x of degree at most n. The identity permutation contributes the term $\prod_{i=1}^n B_{ii}$ in the Leibniz formula, hence we obtain

$$\operatorname{char}_g(x) = \prod_{i=1}^n (x - A_{ii}) + \sum_{\sigma\in S_n,\sigma\neq 1} \operatorname{sgn}(\sigma) \prod_{i=1}^n B_{i\sigma(i)}$$

We now use induction to show that

$$\prod_{i=1}^n (x - A_{ii}) = x^n - \operatorname{Tr}(\mathbf{A})x^{n-1} + C_{n-2}x^{n-2} + \cdots + c_1 x + c_0$$

for scalars $C_{n-2}, \ldots, c_0 \in \mathbb{K}$. For $n = 1$ we obtain $x - A_{11}$, so that the statement is anchored.

Inductive step: Suppose

$$\prod_{i=1}^{n-1} (x - A_{ii}) = x^{n-1} - \left(\sum_{i=1}^{n-1} A_{ii}\right) x^{n-2} + C_{n-2}x^{n-3} + \cdots + c_1 x + c_0,$$

for coefficients $C_{n-2}, \ldots, c_0$, then

$$\begin{aligned}\prod_{i=1}^n (x - A_{ii}) &= (x - A_{nn}) \left[x^{n-1} - \left(\sum_{i=1}^{n-1} A_{ii}\right) x^{n-2} + C_{n-2}x^{n-3} + \cdots + c_1 x + c_0\right] \\ &= x^n - \left(\sum_{i=1}^{n} A_{ii}\right) x^{n-1} + \text{ lower order terms in } x,\end{aligned}$$

so the induction is complete.

We next argue that $\sum_{\sigma\in S_n,\sigma\neq 1} \operatorname{sgn}(\sigma) \prod_{i=1}^n B_{i\sigma(i)}$ has at most degree $n-2$. Notice that each factor $B_{i\sigma(i)}$ of $\prod_{i=1}^n B_{i\sigma(i)}$ for which $i \neq \sigma(i)$ does not contain x. So suppose that $\sum_{\sigma\in S_n,\sigma\neq 1} \operatorname{sgn}(\sigma) \prod_{i=1}^n B_{i\sigma(i)}$ has degree bigger or equal than $n-1$. Then we have $n-1$ integers i with $1 \leqslant i \leqslant n$ such that $i = \sigma(i)$. Let j denote the remaining integer. Since σ is injective, it follows that for any $i \neq j$ we must have $i = \sigma(i) \neq \sigma(j)$. Therefore, $\sigma(j) = j$ and hence $\sigma = 1$, a contradiction.

In summary, we have shown that

$$\operatorname{char}_g(x) = x^n - \operatorname{Tr}(g)x^{n-1} + C_{n-2}x^{n-1} + \cdots + c_1 x + c_0$$

for coefficients $C_{n-2}, \ldots, c_0 \in \mathbb{K}$. It remains to show that $c_0 = (-1)^n \det(g)$. We have $c_0 = \operatorname{char}_g(0) = \det(-g) = \det(-\mathbf{A})$. Since the determinant is linear in each row of $\mathbf{A}$, this gives $\det(-\mathbf{A}) = (-1)^n \det(\mathbf{A})$, as claimed. □

Remark 6.42 In particular, for $n = 2$ we have $\operatorname{char}_g(x) = x^2 - \operatorname{Tr}(g)x + \det(g)$. Compare with Example 6.40.

6.5 Properties of eigenvalues

WEEK 12

We will argue next that an endomorphism $g : V \to V$ of a finite dimensional $\mathbb{K}$-vector space V has at most $\dim(V)$ eigenvalues. We first need:

Theorem 6.43 (Little Bézout's theorem) *For a polynomial $f \in \mathrm{P}(\mathbb{K})$ of degree $n \geqslant 1$ and $x_0 \in \mathbb{K}$, there exists a polynomial $g \in \mathrm{P}(\mathbb{K})$ of degree $n-1$ such that for all $x \in \mathbb{K}$ we have $f(x) = f(x_0) + g(x)(x - x_0)$.*

Proof We will give an explicit expression for the polynomial g. If one is not interested in such an expression, a proof using induction can also be given. Write $f(x) = \sum_{k=0}^{n} a_k x^k$ for coefficients $(a_0, \ldots, a_n) \in \mathbb{K}^{n+1}$. For $0 \leqslant j \leqslant n-1$ consider

$$b_j = \sum_{k=0}^{n-j-1} a_{k+j+1} x_0^k \tag{6.3}$$

and the polynomial

$$g(x) = \sum_{j=0}^{n-1} b_j x^j$$

of degree $n-1$. We have

$$\begin{aligned}
g(x)(x - x_0) &= \sum_{j=0}^{n-1} \sum_{k=0}^{n-j-1} \left(a_{k+j+1} x_0^k x^{j+1}\right) - \sum_{j=0}^{n-1} \sum_{k=0}^{n-j-1} \left(a_{k+j+1} x_0^{k+1} x^j\right) \\
&= \sum_{j=1}^{n} \sum_{k=0}^{n-j} \left(a_{k+j} x_0^k x^j\right) - \sum_{j=0}^{n-1} \sum_{k=1}^{n-j} \left(a_{k+j} x_0^k x^j\right) \\
&= a_n x^n + \sum_{j=1}^{n-1} a_j x^j + a_0 - a_0 - \sum_{k=1}^{n} a_k x_0^k = f(x) - f(x_0).
\end{aligned}$$

□

From this we conclude:

Proposition 6.44 *Let $f \in \mathrm{P}(\mathbb{K})$ be a polynomial of degree n. Then f has at most n (distinct) zeros or f is the zero polynomial.*

Proof We use induction. The case $n = 0$ is clear, hence the statement is anchored.

Inductive step: Suppose $f \in \mathrm{P}(\mathbb{K})$ is a polynomial of degree n with $n+1$ distinct zeros $\lambda_1, \ldots, \lambda_{n+1}$. Since $f(\lambda_{n+1}) = 0$, Theorem 6.43 implies that

$$f(x) = (x - \lambda_{n+1}) g(x)$$

for some polynomial g of degree $n-1$. For $1 \leqslant i \leqslant n$, we thus have

$$0 = f(\lambda_i) = (\lambda_i - \lambda_{n+1}) g(\lambda_i).$$

Since $\lambda_i \neq \lambda_{n+1}$ it follows that $g(\lambda_i) = 0$. Therefore, g has n distinct zeros and must be the zero polynomial by the induction hypothesis. It follows that f is the zero polynomial as well. □

This gives:

Corollary 6.45 *Let $g : V \to V$ be an endomorphism of a $\mathbb{K}$-vector space of dimension $n \in \mathbb{N}$. Then g has at most n (distinct) eigenvalues.*

Proof By Lemma 6.36 and Lemma 6.41, the eigenvalues of g are the zeros of the characteristic polynomial. The characteristic polynomial of g has degree n. The claim follows by applying Proposition 6.44. □

Proposition 6.46 (Linear independence of eigenvectors) *Let V be a finite dimensional $\mathbb{K}$-vector space and $g : V \to V$ an endomorphism. Then the eigenspaces $\operatorname{Eig}_g(\lambda)$ of g are in direct sum. In particular, if $v_1, \dots, v_m$ are eigenvectors corresponding to distinct eigenvalues of g, then $\{v_1, \dots, v_m\}$ are linearly independent.*

Proof We use induction on the number m of distinct eigenvalues of g. Let $\{\lambda_1, \dots, \lambda_m\}$ be distinct eigenvalues of g. For $m = 1$ the statement is trivially true, so the statement is anchored.

Inductive step: Assume $m - 1$ eigenspaces are in direct sum. We want to show that then m eigenspaces are also in direct sum. Let $v_i, v_i' \in \operatorname{Eig}_g(\lambda_i)$ be eigenvectors such that

$$v_1 + v_2 + \cdots + v_m = v_1' + v_2' + \cdots + v_{\tilde{m}}. \tag{6.4}$$

Applying g to this last equation gives

$$\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_m v_m = \lambda_1 v_1' + \lambda_2 v_2' + \cdots + \lambda_m v_{\tilde{m}}. \tag{6.5}$$

Subtracting λ_m times (6.4) from (6.5) gives

$$(\lambda_1 - \lambda_m)v_1 + \cdots + (\lambda_{m-1} - \lambda_m)v_{m-1} = (\lambda_1 - \lambda_m)v_1' + \cdots + (\lambda_{m-1} - \lambda_m)v_{m-1}'.$$

Since $m - 1$ eigenspaces are in direct sum, this implies that $(\lambda_i - \lambda_m)v_i = (\lambda_i - \lambda_m)v_i'$ for $1 \leqslant i \leqslant m - 1$. Since the eigenvalues are distinct, we have $\lambda_i - \lambda_m \neq 0$ for all $1 \leqslant i \leqslant m - 1$ and hence $v_i = v_i'$ for all $1 \leqslant i \leqslant m - 1$. Now (6.5) implies that $v_m = v_{\tilde{m}}$ as well and the inductive step is complete.

Since the eigenspaces are in direct sum, the linear independence of eigenvectors with respect to distinct eigenvalues follows from Remark 6.7 (ii). □

In the case where all the eigenvalues are distinct, we conclude that g is diagonalisable.

Proposition 6.47 *Let $g : V \to V$ be an endomorphism of a finite dimensional $\mathbb{K}$-vector space V. Suppose the characteristic polynomial of g has $\dim(V)$ distinct zeros (that is, the algebraic multiplicity of each eigenvalue is 1), then g is diagonalisable.*

Proof Let $n = \dim(V)$. Let $\lambda_1, \dots, \lambda_n$ denote the distinct eigenvalues of g. Let $0_V \neq v_i \in \operatorname{Eig}_g(\lambda_i)$ for $i = 1, \dots, n$. Then, by Proposition 6.46, the eigenvectors are linearly independent, it follows that $(v_1, \dots, v_n)$ is an ordered basis of V consisting of eigenvectors, hence g is diagonalisable. □

Remark 6.48 Proposition 6.47 gives a sufficient condition for an endomorphism $g : V \to V$ to be diagonalisable, it is however not necessary. The identity endomorphism is diagonalisable, but its spectrum consists of the single eigenvalue 1 with algebraic multiplicity $\dim(V)$.

Every polynomial in $\mathsf{P}(\mathbb{C})$ of degree at least 1 has at least one zero. This fact is known as the *fundamental theorem of algebra*. The name is well-established, but quite misleading, as there is no purely algebraic proof. You will encounter a proof of this statement in the module M07. As a consequence we obtain the following important existence theorem:

Theorem 6.49 (Existence of eigenvalues) *Let $g : V \to V$ be an endomorphism of a complex vector space V of dimension $n \geqslant 1$. Then g admits at least one eigenvalue. Moreover, the sum of the algebraic multiplicities of the eigenvalues of g is equal to n. In particular, if $\mathbf{A} \in M_{n,n}(\mathbb{C})$ is a matrix, then there is at least one eigenvalue of $\mathbf{A}$.*

Proof By Lemma 6.36 and Lemma 6.41, the eigenvalues of g are the zeros of the characteristic polynomial and this is an element of $\mathsf{P}(\mathbb{C})$. The first statement thus follows by applying the fundamental theorem of algebra to the characteristic polynomial of g.

Applying Theorem 6.43 and the fundamental theorem of algebra repeatedly, we find $k \in \mathbb{N}$ and multiplicities $m_1, \ldots, m_k \in \mathbb{N}$ such that

$$\mathrm{char}_g(x) = (x - \lambda_1)^{m_1}(x - \lambda_2)^{m_2} \cdots (x - \lambda_k)^{m_k}$$

where $\lambda_1, \ldots, \lambda_k$ are zeros of char_g. Since char_g has degree n, it follows that $\sum_{i=1}^{k} m_i = n$. □

Example 6.50

(i) Recall that the *discriminant* of a quadratic polynomial $x \mapsto ax^2 + bx + c \in \mathsf{P}(\mathbb{K})$ is $b^2 - 4ac$, provided $a \neq 0$. If $\mathbb{K} = \mathbb{C}$ and $b^2 - 4ac$ is non-zero, then the polynomial $ax^2 + bx + c$ has two distinct zeros. The characteristic polynomial of a 2-by-2 matrix $\mathbf{A}$ satisfies $\mathrm{char}_{\mathbf{A}}(x) = x^2 - \mathrm{Tr}(\mathbf{A})x + \det(\mathbf{A})$. Therefore, if $\mathbf{A}$ has complex entries and satisfies $(\mathrm{Tr}\,\mathbf{A})^2 - 4\det\mathbf{A} \neq 0$, then it is diagonalisable. If $\mathbf{A}$ has real entries and satisfies $(\mathrm{Tr}\,\mathbf{A})^2 - 4\det\mathbf{A} \geqslant 0$, then it has a least one eigenvalue. If $(\mathrm{Tr}\,\mathbf{A})^2 - 4\det\mathbf{A} > 0$ then it is diagonalisable.

(ii) Recall that, by Proposition 5.24, an upper triangular matrix $\mathbf{A} = (A_{ij})_{1 \leqslant i,j \leqslant n}$ satisfies $\det\mathbf{A} = \prod_{i=1}^{n} A_{ii}$. It follows that

$$\mathrm{char}_{\mathbf{A}}(x) = \prod_{i=1}^{n}(x - A_{ii}) = (x - A_{11})(x - A_{22}) \cdots (x - A_{nn}).$$

Consequently, an upper triangular matrix has spectrum $\{A_{11}, A_{22}, \ldots, A_{nn}\}$ and is diagonalisable if all its diagonal entries are distinct. Notice that by Example 6.40 (ii) not every upper triangular matrix is diagonalisable.

Example 6.51 (Fibonacci sequences) We revisit the Fibonacci sequences, now equipped with the theory of endomorphisms. A Fibonacci sequence is a sequence $\xi : \mathbb{N} \cup \{0\} \to \mathbb{K}$ satisfying the recursive relation $\xi_{n+2} = \xi_n + \xi_{n+1}$. Consider the matrix

$$\mathbf{A} = \begin{pmatrix} \xi_0 & \xi_1 \\ \xi_1 & \xi_2 \end{pmatrix}.$$

Then, using induction, we can show that

$$\mathbf{A}^n = \begin{pmatrix} \xi_{n-1} & \xi_n \\ \xi_n & \xi_{n+1} \end{pmatrix}$$

for all $n \in \mathbb{N}$. We would like to compute $\mathbf{A}^n$ for the initial conditions $\xi_0 = 0$ and $\xi_1 = 1$. Suppose we can find an invertible matrix $\mathbf{C}$ so that $\mathbf{A} = \mathbf{C}\mathbf{D}\mathbf{C}^{-1}$ for some

diagonal matrix $\mathbf{D}$. Then
$$\mathbf{A}^n = \mathbf{C}\mathbf{D}\mathbf{C}^{-1}\mathbf{C}\mathbf{D}\mathbf{C}^{-1}\cdots\mathbf{C}\mathbf{D}\mathbf{C}^{-1} = \mathbf{C}\mathbf{D}^n\mathbf{C}^{-1}$$
and we can easily compute $\mathbf{A}^n$, as the n-th power of a diagonal matrix $\mathbf{D}$ is the diagonal matrix whose diagonal entries are given by the n-th powers of diagonal entries of $\mathbf{D}$. We thus want to diagonalise the matrix
$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}.$$
We obtain $\operatorname{char}_{\mathbf{A}}(x) = x^2 - x - 1$ and hence eigenvalues $\lambda_1 = (1+\sqrt{5})/2$ and $\lambda_2 = (1-\sqrt{5})/2$. From this we compute
$$\operatorname{Eig}_{\mathbf{A}}(\lambda_1) = \operatorname{span}\left\{\begin{pmatrix} 1 \\ \lambda_1 \end{pmatrix}\right\} \qquad \text{and} \qquad \operatorname{Eig}_{\mathbf{A}}(\lambda_2) = \operatorname{span}\left\{\begin{pmatrix} 1 \\ \lambda_2 \end{pmatrix}\right\}$$
Let $\mathbf{e} = (\vec{e}_1, \vec{e}_2)$ denote the standard basis of $\mathbb{R}^2$ and consider the ordered basis
$$\mathbf{b} = \left(\begin{pmatrix} 1 \\ \lambda_1 \end{pmatrix}, \begin{pmatrix} 1 \\ \lambda_2 \end{pmatrix}\right)$$
of eigenvectors of $f_{\mathbf{A}}$. We have
$$\mathbf{M}(f_{\mathbf{A}}, \mathbf{b}, \mathbf{b}) = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = \mathbf{D}$$
and the change of base matrix is
$$\mathbf{C} = \mathbf{C}(\mathbf{b}, \mathbf{e}) = \begin{pmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{pmatrix}$$
and
$$\mathbf{C}^{-1} = \mathbf{C}(\mathbf{e}, \mathbf{b}) = \frac{1}{\lambda_2 - \lambda_1}\begin{pmatrix} \lambda_2 & -1 \\ -\lambda_1 & 1 \end{pmatrix}.$$
Therefore $\mathbf{A} = \mathbf{C}\mathbf{D}\mathbf{C}^{-1}$ and hence $\mathbf{A}^n = \mathbf{C}\mathbf{D}^n\mathbf{C}^{-1}$ so that
$$\mathbf{A}^n = \frac{1}{\lambda_2 - \lambda_1}\begin{pmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{pmatrix}\begin{pmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{pmatrix}\begin{pmatrix} \lambda_2 & -1 \\ -\lambda_1 & 1 \end{pmatrix} = \begin{pmatrix} \xi_{n-1} & \xi_n \\ \xi_n & \xi_{n+1} \end{pmatrix}.$$
This yields the formula
$$\xi_n = \frac{\lambda_1^n - \lambda_2^n}{\lambda_1 - \lambda_2}.$$

Proposition 6.52 *Let $g : V \to V$ be an endomorphism of a finite dimensional $\mathbb{K}$-vector space V of dimension $n \geqslant 1$.*

(i) Let λ be an eigenvalue of g. Then its algebraic multiplicity is at least as big as its geometric multiplicity.

(ii) If $\mathbb{K} = \mathbb{C}$, then g is diagonalisable if and only if for all eigenvalues of g, the algebraic and geometric multiplicity are the same.

Proof (i) Let $\dim \operatorname{Eig}_g(\lambda) = m$ and $\mathbf{b}$ be an ordered basis of $\operatorname{Eig}_g(\lambda)$. Furthermore, let $\mathbf{b}'$ be an ordered tuple of vectors such that $\mathbf{c} = (\mathbf{b}, \mathbf{b}')$ is an ordered basis of V. The eigenspace $\operatorname{Eig}_g(\lambda)$ is stable under g and
$$\mathbf{M}(g|_{\operatorname{Eig}_g(\lambda)}, \mathbf{b}, \mathbf{b}) = \lambda \mathbf{1}_m.$$
By Proposition 6.34, the matrix representation of g with respect to the basis $\mathbf{c}$ takes the form
$$\mathbf{M}(g, \mathbf{c}, \mathbf{c}) = \begin{pmatrix} \lambda\mathbf{1}_m & * \\ \mathbf{0}_{m-n,m} & \mathbf{B} \end{pmatrix}$$

for some matrix $\mathbf{B} \in M_{n-m,n-m}(\mathbb{K})$. We thus obtain

$$\operatorname{char}_g(x) = \det \begin{pmatrix} (x-\lambda)\mathbf{1}_m & * \\ \mathbf{0}_{m-n,m} & x\mathbf{1}_{n-m} - \mathbf{B} \end{pmatrix}$$

Applying the Laplace expansion (5.5) with respect to the first column, we have

$$\operatorname{char}_g(x) = (x-\lambda) \det \begin{pmatrix} (x-\lambda)\mathbf{1}_{m-1} & * \\ \mathbf{0}_{m-n,m-1} & x\mathbf{1}_{n-m} - \mathbf{B} \end{pmatrix}$$

Applying the Laplace expansion again with respect to the first column, m-times in total, we get

$$\operatorname{char}_g(x) = (x-\lambda)^m \det(x\mathbf{1}_{n-m} - \mathbf{B}) = (x-\lambda)^m \operatorname{char}_{\mathbf{B}}(x).$$

The algebraic multiplicity of λ is thus at least m.

(ii) Suppose $\mathbb{K} = \mathbb{C}$ and that $g : V \to V$ is diagonalisable. Hence we have an ordered basis $(v_1, \ldots, v_n)$ of V consisting of eigenvectors of g. Therefore,

$$\operatorname{char}_g(x) = \prod_{i=1}^{n} (x - \lambda_i)$$

where λ_i is the eigenvalue of the eigenvector v_i, $1 \leqslant i \leqslant n$. For any eigenvalue λ_j, its algebraic multiplicity is the number of indices i with $\lambda_i = \lambda_j$. For each such index i, the eigenvector v_i satisfies $g(v_i) = \lambda_i v_i = \lambda_j v_i$ and hence is an element of the eigenspace $\operatorname{Eig}_g(\lambda_j)$. The geometric multiplicity of each eigenvalue is thus at least as big as the algebraic multiplicity, but by the previous statement, the latter cannot be bigger than the former, hence they are equal.

Conversely, suppose that for all eigenvalues of g, the algebraic and geometric multiplicity are the same. Since $\mathbb{K} = \mathbb{C}$, by Theorem 6.49, the sum of the algebraic multiplicities is n. The sum of the geometric multiplicities is by assumption also n. Since, by Proposition 6.46, the eigenspaces with respect to different eigenvalues are in direct sum, we obtain a basis of V consisting of eigenvectors of g. □

6.6 Special endomorphisms

6.6.1 Involutions

A mapping $\iota : \mathcal{X} \to \mathcal{X}$ from a set $\mathcal{X}$ into itself is called an *involution*, if $\iota \circ \iota = \operatorname{Id}_{\mathcal{X}}$. In the case where $\mathcal{X}$ is a vector space and ι is linear, then ι is called a *linear involution*.

Example 6.53 (Involutions)

(i) Let V be a $\mathbb{K}$-vector space. Then the identity mapping $\operatorname{Id}_V : V \to V$ is a linear involution.

(ii) For all $n \in \mathbb{N}$, the transpose $M_{n,n}(\mathbb{K}) \to M_{n,n}(\mathbb{K})$ is a linear involution.

(iii) For $n \in \mathbb{N}$, let $\mathcal{X}$ denote the set of invertible $n \times n$ matrices. Then the matrix inverse ${}^{-1} : \mathcal{X} \to \mathcal{X}$ is an involution. Notice that $\mathcal{X}$ is not a vector space.

(iv) For any $\mathbb{K}$-vector space V, the mapping $\iota : V \to V, v \mapsto -v$ is a linear involution. Considering $\mathsf{F}(I, \mathbb{K})$, the $\mathbb{K}$-vector space of functions on the interval $I \subset \mathbb{R}$, we obtain a linear involution of $\mathsf{F}(V, \mathbb{K})$ by sending a function f to $f \circ \iota$.

(v) If $\mathbf{A} \in M_{n,n}(\mathbb{K})$ satisfies $\mathbf{A}^2 = \mathbf{1}_n$, then $f_{\mathbf{A}} : \mathbb{K}^n \to \mathbb{K}^n$ is a linear involution.

The spectrum of an involution is a subset of $\{-1, 1\}$.

Proposition 6.54 *Let V be a $\mathbb{K}$-vector space and $\iota : V \to V$ a linear involution. Then the spectrum of ι is contained in $\{-1, 1\}$. Moreover $V = \operatorname{Eig}_\iota(1) \oplus \operatorname{Eig}_\iota(-1)$ and ι is diagonalisable.*

Proof Suppose $\lambda \in \mathbb{K}$ is an eigenvalue of ι so that $\iota(v) = \lambda v$ for some non-zero vector $v \in V$. Then $\iota(\iota(v)) = v = \lambda\iota(v) = \lambda^2 v$. Hence $(1-\lambda^2)v = 0_V$ and since v is non-zero, we conclude that $\lambda = \pm 1$. By Proposition 6.46, the eigenspaces $\operatorname{Eig}_\iota(1)$ and $\operatorname{Eig}_\iota(-1)$ are in direct sum.

For $v \in V$ we write

$$v = \underbrace{\frac{1}{2}(v + f(v))}_{\in \operatorname{Eig}_\iota(1)} + \underbrace{\frac{1}{2}(v - f(v))}_{\in \operatorname{Eig}_\iota(-1)}$$

hence $V = \operatorname{Eig}_\iota(1) \oplus \operatorname{Eig}_\iota(-1)$. Take an ordered basis $\mathbf{b}_+$ of $\operatorname{Eig}_\iota(1)$ and an ordered basis $\mathbf{b}_-$ of $\operatorname{Eig}_\iota(1)$. Then $(\mathbf{b}_+, \mathbf{b}_-)$ is an ordered basis of V consisting of eigenvectors of ι. □

6.6.2 Projections

A linear mapping $\Pi : V \to V$ satisfying $\Pi \circ \Pi = \Pi$ is called a *projection*.

Example 6.55 Consider $V = \mathbb{R}^3$ and

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Clearly, $\mathbf{A}^2 = \mathbf{A}$ and $f_\mathbf{A} : \mathbb{R}^3 \to \mathbb{R}^3$ projects a vector $\vec{x} = (x_i)_{1 \leqslant i \leqslant 3}$ onto the plane $\{\vec{x} \in \mathbb{R}^3 | x_3 = 0\}$.

In a sense there is only one type of projection. Recall from the exercises that for a projection $\Pi : V \to V$, we have $V = \operatorname{Ker}\Pi \oplus \operatorname{Im}\Pi$. Given two subspaces U_1, U_2 of V such that $V = U_1 \oplus U_2$, there is a projection $\Pi : V \to V$ whose kernel is U_1 and whose image is U_2. Indeed, every vector $v \in V$ can be written as $v = u_1 + u_2$ for unique vectors $u_i \in U_i$ for $i = 1, 2$. Hence we obtain a projection by defining $\Pi(v) = u_2$ for all $v \in V$.

Denote by $\mathcal{X}$ the set of projections from V to V and by $\mathcal{Y}$ the set of pairs (U_1, U_2) of subspaces of V that are in direct sum and satisfy $V = U_1 \oplus U_2$. Then we obtain a mapping $\Lambda : \mathcal{X} \to \mathcal{Y}$ defined by $f \mapsto (\operatorname{Ker} f, \operatorname{Im} f)$.

Similar to Proposition 6.54, we obtain:

Proposition 6.56 *Let V be a $\mathbb{K}$-vector space and $\Pi : V \to V$ a projection. Then the spectrum of Π is contained in $\{0, 1\}$. Moreover $V = \operatorname{Eig}_\Pi(0) \oplus \operatorname{Eig}_\Pi(1)$, Π is diagonalisable and* $\operatorname{Im}\Pi = \operatorname{Eig}_\Pi(1)$.

Proof Let $v \in V$ be an eigenvector of the projection Π with eigenvalue λ. Hence we obtain $\Pi(\Pi(v)) = \lambda^2 v = \Pi(v) = \lambda v$, equivalently, $\lambda(\lambda - 1)v = 0_V$. Since v is non zero, it follows that $\lambda = 0$ or $\lambda = 1$. Since Π is a projection, we have $V = \operatorname{Ker}\Pi \oplus \operatorname{Im}\Pi$. Since $\operatorname{Ker}\Pi = \operatorname{Eig}_\Pi(0)$, we thus only need to show that $\operatorname{Im}\Pi = \operatorname{Eig}_\Pi(1)$. Let $v \in \operatorname{Im}\Pi$ so that $v = \Pi(\hat{v})$ for some vector $\hat{v} \in V$. Hence $\Pi(v) = \Pi(\Pi(\hat{v})) = \Pi(\hat{v}) = v$ and v is an eigenvector with eigenvalue 1. Conversely, suppose $v \in V$ is an eigenvector of Π with eigenvalue 1. Then $\Pi(v) = v = \Pi(\Pi(v))$ and hence $v \in \operatorname{Im}\Pi$. We thus conclude that $\operatorname{Im}\Pi = \operatorname{Eig}_\Pi(1)$. Choosing an ordered basis of $\operatorname{Ker}\Pi$ and an ordered basis of $\operatorname{Im}\Pi$ gives a basis of V consisting of eigenvectors, hence Π is diagonalisable. □

Exercises

Exercise 6.57 Derive the formula (6.3) for the coefficients b_j.

Exercise 6.58 Show that Λ is a bijection.

Exercise 6.59 Show that if $\Pi : V \to V$ is a projection then $\mathrm{Id}_V - \Pi : V \to V$ is a projection with kernel equal to the image of Π and image equal to the kernel of Π.

CHAPTER 7

Quotient vector spaces

7.1 Affine mappings and affine spaces

WEEK 13

Previously we saw that we can take the sum of subspaces of a vector space. In this final chapter of the Linear Algebra I module we introduce the concept of a quotient of a vector space by a subspace.

Translations are among the simplest non-linear mappings.

Definition 7.1 (Translation) Let V be a $\mathbb{K}$-vector space and $v_0 \in V$. The mapping

$$T_{v_0} : V \to V, \qquad v \mapsto v + v_0$$

is called the *translation* by the vector v_0.

Remark 7.2 Notice that for $v_0 \neq 0_V$, a translation is not linear, since $T_{v_0}(0_V) = 0_V + v_0 = v_0 \neq 0_V$.

Taking $s_1 = 1$ and $s_2 = -1$ in (3.6), we see that a linear map $f : V \to W$ between $\mathbb{K}$-vector spaces V, W satisfies $f(v_1 - v_2) = f(v_1) - f(v_2)$ for all $v_1, v_2 \in V$. In particular, linear maps are affine maps in the following sense:

Definition 7.3 (Affine mapping) A mapping $f : V \to W$ is called *affine* if there exists a linear map $g : V \to W$ so that $f(v_1) - f(v_2) = g(v_1 - v_2)$ for all $v_1, v_2 \in V$. We call g the *linear map associated to* f.

Affine mappings are compositions of linear mappings and translations:

Proposition 7.4 *A mapping $f : V \to W$ is affine if and only if there exists a linear map $g : V \to W$ and a translation $T_{w_0} : W \to W$ so that $f = T_{w_0} \circ g$.*

Proof $\Leftarrow$ Let $g : V \to W$ be linear and $T_{w_0} : W \to W$ be a translation for some vector $w_0 \in W$ so that $T_{w_0}(w) = w + w_0$ for all $w \in W$. Let $f = T_{w_0} \circ g$ so that $f(v) = g(v) + w_0$ for all $v \in V$. Then

$$f(v_1) - f(v_2) = g(v_1) + w_0 - g(v_2) - w_0 = g(v_1) - g(v_2) = g(v_1 - v_2),$$

hence f is affine.

$\Rightarrow$ Let $f : V \to W$ be affine and $g : V \to W$ its associated linear map. Since f is affine we have for all $v \in V$

$$f(v) - f(0_V) = g(v - 0_V) = g(v) - g(0_V) = g(v)$$

where we use the linearity of g and Lemma 3.15. Writing $w_0 = f(0_V)$ we thus have

$$f(v) = g(v) + w_0$$

so that f is the composition of the linear map g and the translation $T_{w_0} : W \to W$, $w \mapsto w + w_0$. □

Example 7.5 Let $\mathbf{A} \in M_{m,n}(\mathbb{K})$, $\vec{b} \in \mathbb{K}^m$ and

$$f_{\mathbf{A},\vec{b}} : \mathbb{K}^n \to \mathbb{K}^m, \quad \vec{x} \mapsto \mathbf{A}\vec{x} + \vec{b}.$$

Then $f_{\mathbf{A},\vec{b}}$ is an affine map whose associated linear map is $f_{\mathbf{A}}$. Conversely, combining Lemma 3.18 and Proposition 7.4, we see that every affine map $\mathbb{K}^n \to \mathbb{K}^m$ is of the form $f_{\mathbf{A},\vec{b}}$ for some matrix $\mathbf{A} \in M_{m,n}(\mathbb{K})$ and vector $\vec{b} \in \mathbb{K}^m$.

An affine subspace of a $\mathbb{K}$-vector space V is a translation of a subspace by some fixed vector v_0.

Definition 7.6 (Affine subspace) Let V be a $\mathbb{K}$-vector space. An *affine subspace of* V is a subset of the form

$$U + v_0 = \{u + v_0 | u \in U\},$$

where $U \subset V$ is a subspace and $v_0 \in V$. We call U the *associated vector space to the affine subspace* $U + v_0$ and we say that $U + v_0$ is *parallel* to U.

Example 7.7 Let $V = \mathbb{R}^2$ and $U = \operatorname{span}\{\vec{e}_1 + \vec{e}_2\} = \{s(\vec{e}_1 + \vec{e}_2) | s \in \mathbb{R}\}$ where here, as usual, $\{\vec{e}_1, \vec{e}_2\}$ denotes the standard basis of $\mathbb{R}^2$. So U is the line through the origin $0_{\mathbb{R}^2}$ defined by the equation $y = x$. By definition, for all $\vec{v} \in \mathbb{R}^2$ we have

$$U + \vec{v} = \{\vec{v} + s\vec{w} | s \in \mathbb{R}\},$$

where we write $\vec{w} = \vec{e}_1 + \vec{e}_2$. So for each $\vec{v} \in \mathbb{R}^2$, the affine subspace $U + \vec{v}$ is a line in $\mathbb{R}^2$, the translation by the vector $\vec{v}$ of the line defined by $y = x$.

7.2 Quotient vector spaces

Let U be a subspace of a $\mathbb{K}$-vector space V. We want to make sense of the notion of *dividing* V by U. It turns out that there is a natural way to do this and moreover, the quotient V/U again carries the structure of a $\mathbb{K}$-vector space. The idea is to define V/U to be the set of all translations of the subspace U, that is, we consider the *set of subsets*

$$V/U = \{U + v | v \in V\}.$$

We have to define what it means to add affine subspaces $U + v_1$ and $U + v_2$ and what it means to scale $U + v$ by a scalar $s \in \mathbb{K}$. Formally, it is tempting to define $0_{V/U} = U + 0_V$ and

(7.1) $$(U + v_1) +_{V/U} (U + v_2) = U + (v_1 + v_2)$$

for all $v_1, v_2 \in V$ as well as

$$s \cdot_{V/U} (U + v) = U + (sv) \tag{7.2}$$

for all $v \in V$ and $s \in \mathbb{K}$. However, we have to make sure that these operations are well defined. We do this with the help of the following lemma.

Lemma 7.8 *Let $U \subset V$ be a subspace. Then any vector $v \in V$ belongs to a unique affine subspace parallel to U, namely $U + v$. In particular, two affine subspaces $U + v_1$ and $U + v_2$ are either equal or have empty intersection.*

Proof Since $0_V \in U$, we have $v \in (U+v)$, hence we only need to show that if $v \in (U+\hat{v})$ for some vector $\hat{v}$, then $U + v = U + \hat{v}$. Assume $v \in (U + \hat{v})$ so that $v = u + \hat{v}$ for some vector $u \in U$. Suppose $w \in (U + \hat{v})$. We need to show that then also $w \in (U + v)$. Since $w \in (U + \hat{v})$ we have $w = \hat{u} + \hat{v}$ for some vector $\hat{u} \in U$. Using that $\hat{v} = v - u$, we obtain

$$w = \hat{u} + v - u = \hat{u} - u + v$$

Since U is a subspace we have $\hat{u} - u \in U$ and hence $w \in (U + v)$.

Conversely, suppose $w \in (U + v)$, it follows exactly as before that then $w \in (U + \hat{v})$ as well. □

We are now going to show that (7.1) and (7.2) are well defined. We start with (7.1). Let $v_1, v_2 \in V$ and $w_1, w_2 \in V$ such that

$$U + v_1 = U + w_1 \qquad \text{and} \qquad U + v_2 = U + w_2.$$

We need to show that $U + (v_1 + v_2) = U + (w_1 + w_2)$. By Lemma 7.8 it suffices to show that $w_1 + w_2$ is an element of $U + (v_1 + v_2)$. Since $U + w_1 = U + v_1$ it follows that $w_1 \in (U + v_1)$ so that $w_1 = u_1 + v_1$ for some element $u_1 \in U$. Likewise it follows that $w_2 = u_2 + v_2$ for some element $u_2 \in U$. Hence

$$w_1 + w_2 = u_1 + u_2 + v_1 + v_2.$$

Since U is a subspace, we have $u_1 + u_2 \in U$ and thus it follows that $w_1 + w_2$ is an element of $U + (v_1 + v_2)$.

For (7.2) we need to show that if $v \in V$ and $w \in V$ are such that $U + v = U + w$, then $U + (sv) = U + (sw)$ for all $s \in \mathbb{K}$. Again, applying Lemma 7.8 we only need to show that $sw \in U + (sv)$. Since $U + v = U + w$ it follows that there exists $u \in U$ with $w = u + v$. Hence $sw = su + sv$ and U being a subspace, we have $su \in U$ and thus sw lies in $U + (sv)$, as claimed.

Having equipped V/U with addition $+_{V/U}$ defined by (7.1) and scalar multiplication $\cdot_{V/U}$ defined by (7.2), we need to show that V/U with zero vector $U + 0_V$ is indeed a $\mathbb{K}$-vector space. All the properties of Definition 3.1 for V/U are however simply a consequence of the corresponding property for V. For instance commutativity of vector addition in V/U follows from the commutativity of vector in addition in V, that is, for all $v_1, v_2 \in V$ we have

$$(U + v_1) +_{V/U} (U + v_2) = U + (v_1 + v_2) = U + (v_2 + v_1) = (U + v_2) +_{V/U} (U + v_1).$$

The remaining properties follow similarly.

Notice that we have a surjective mapping

$$p : V \to V/U, \quad v \mapsto U + v.$$

which satisfies

$$p(v_1 + v_2) = U + (v_1 + v_2) = (U + v_1) +_{V/U} (U + v_2) = p(v_1) +_{V/U} p(v_2)$$

for all $v_1, v_2 \in V$ and

$$p(sv) = U + (sv) = s \cdot_{V/U} (U + v) = s \cdot_{V/U} p(v).$$

for all $v \in V$ and $s \in \mathbb{K}$. Therefore, the mapping p is linear.

Definition 7.9 (Quotient vector space) The vector space V/U is called the *quotient (vector) space of V by U*. The linear map $p : V \to V/U$ is called the *canonical surjection* from V to V/U.

The mapping $p : V \to V/U$ satisfies

$$p(v) = 0_{V/W} = U + 0_V \quad \iff \quad v \in U$$

and hence $\mathrm{Ker}(p) = U$. This gives:

Proposition 7.10 *Suppose the $\mathbb{K}$-vector space V is finite dimensional. Then V/U is finite dimensional as well and*

$$\dim(V/U) = \dim(V) - \dim(U).$$

Proof Since p is surjective it follows that V/U is finite dimensional as well. Hence we can apply Theorem 3.76 and obtain

$$\dim V = \dim \mathrm{Ker}(p) + \dim \mathrm{Im}(p) = \dim U + \dim(V/U),$$

where we use that $\mathrm{Im}(p) = V/U$ and $\mathrm{Ker}(p) = U$. □

Example 7.11 (Special cases)

(i) In the case where $U = V$ we obtain $V/U = \{0_{V/U}\}$.

(ii) In the case where $U = \{0_V\}$ we obtain that V/U is isomorphic to V.

Exercises

Exercise 7.12 Show that the image of an affine subspace under an affine map is again an affine subspace and that the preimage of an affine subspace under an affine map is again an affine subspace or empty (cf. Proposition 3.26).