M04: LINEAR ALGEBRA I

David Loeffler

Spring 2025

Last updated 5th February 2025

FernUni.ch
UniDistance.ch

Contents

Acknowledgements 6
Recommended textbooks 6
HTML Version 6

Chapter 1 Fields and complex numbers 7
1.1 Reminders on sets and mappings 7
Mappings 7
Images and preimages 8
Exercises 8
1.2 Fields 9
Definitions 9
First properties 11
Mappings between fields 12
1.3 Complex numbers 12
Motivation 12
Definitions 13
Putting $\mathbb{R}$ inside $\mathbb{C}$ 13
Exercises 15

Chapter 2 Matrices, I 17
2.1 Definitions 17
2.2 Arithmetic with matrices 19
Matrix addition 19
Scalar multiplication of a matrix 20
Matrix multiplication 21
Properties 21
2.3 Transpose and inverse 23
Transpose 23
Inverses 24
Exercises 24

Chapter 3 Matrices, II 26
3.1 Row echelon form 26
Definition 26
Echelonizing a matrix via Gauss–Jordan 27
Keeping track of the transformation matrix 28
Uniqueness 29
3.2 Solving equations 30
3.3 Inverting a matrix 33
Non-square matrices 33
Square matrices 33
Exercises 34

Chapter 4 Vector spaces 36
4.1 Abstract vector spaces 36
Examples of vector spaces 38

4.2 Linear combinations 40
4.3 Vector subspaces 41
Operations on subspaces 42
4.4 Subspaces generated by sets 43
4.5 Generating sets and finite-dimensionality 44
4.6 Linear independence 45
Exercises 46

Chapter 5 Dimensions of vector spaces 48
5.1 Growing and shrinking sets 48
5.2 The fundamental inequality 49
5.3 Bases of vector spaces 50
Definition 50
The Main Theorem on Bases 51
Consequences of the Main Theorem 52
5.4 Dimensions of vector spaces 53
5.5 Computing with subspaces 54
Exercises 55

Chapter 6 Linear maps 57
6.1 Linear maps 57
Examples 58
First properties of linear maps 58
Isomorphisms 59
6.2 Images, preimages, kernels 59
Linear maps and dimension 61
6.3 The rank-nullity theorem 62
Exercises 64

Chapter 7 Linear maps and matrices 66
7.1 Linear mappings associated to matrices 66
Composition and inverses 68
7.2 Computing kernels and images 69
Image and rank 69
Kernel and nullity 70
Exercises 72

Chapter 8 Coordinate systems and changes of basis 73
8.1 Linear coordinate systems 73
8.2 The matrix of a linear map 76
8.3 Change of basis 80
Exercises 83

Chapter 9 The determinant, I 84
9.1 Axiomatic characterisation 84
9.1.1 Multilinear maps 84
9.1.2 Existence and uniqueness 86
9.2 Uniqueness of the determinant 88
9.3 Existence of the determinant 90
Exercises 93

Chapter 10 The determinant, II 95
10.1 Properties of the determinant 95
10.2 Permutations 96
10.3 The Leibniz formula 99

10.4 Cramer's rule 102
Exercises 104

Chapter 11 Endomorphisms, I 106
11.1 Matrices of endomorphisms 106
Similarity 106
Invariants of similarity classes 107
11.2 Detour: More on Subspaces 110
Sums of subspaces 110
Direct sums 110
Complements 111
11.3 Eigenvectors and eigenvalues 113
11.4 The characteristic polynomial 117
Exercises 119

Chapter 12 Endomorphisms, II 121
12.1 Properties of eigenvalues 121
12.2 Special endomorphisms 126
Involutions 126
Projections 127
Exercises 127

Chapter 13 Affine spaces and quotient vector spaces 129
13.1 Affine mappings and affine spaces 129
13.2 Quotient vector spaces 130
Exercises 132

Acknowledgements

These lecture notes are heavily based on the course taught in the Autumn 2023 semester by my colleage Thomas Mettler, and I am very grateful to Thomas for crafting such a sound foundation on which to build. I have also integrated some very useful suggestions from Sarah Zerbes, based on her “Linear Algebra 1” course at ETH Zürich.

Recommended textbooks

These lecture 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 lecture notes and additionally animations, solutions to the exercises and multiple choice questions.

CHAPTER 1

Fields and complex numbers

Contents

1.1 Reminders on sets and mappings **7**
Mappings 7
Images and preimages 8
Exercises **8**
1.2 Fields **9**
Definitions 9
First properties 11
Mappings between fields 12
1.3 Complex numbers **12**
Motivation 12
Definitions 13
Putting $\mathbb{R}$ inside $\mathbb{C}$ 13
Exercises **15**

1.1 Reminders on sets and mappings

This short section is not new material: all the concepts below are in the modules M01 Algorithmics or M02 Statistics and Discrete Structures. It is just here as a quick reminder that you can refer back to later in these lecture notes if necessary.

Mappings

Recall that for $\mathcal{X}, \mathcal{Y}$ sets, we have the notion of a **mapping** (or **map** or **function**) $f : \mathcal{X} \to \mathcal{Y}$; formally this is a subset of $\mathcal{X} \times \mathcal{Y}$ with certain properties, but we think of it as some kind of "rule" or "recipe" which produces, for each $x \in \mathcal{X}$, an element $f(x) \in \mathcal{Y}$.

- Two functions $f_1, f_2 : \mathcal{X} \to \mathcal{Y}$ are identical if they take the same values, i.e. $f_1 = f_2$ if and only if $f_1(x) = f_2(x)\ \forall x \in \mathcal{X}$.
- For any set $\mathcal{X}$ there is an **identity map** $\mathrm{Id}_{\mathcal{X}}$, defined by $\mathrm{Id}_{\mathcal{X}}(x) = x\ \forall x \in \mathcal{X}$.
- Given $f : \mathcal{X} \to \mathcal{Y}$ and $g : \mathcal{Y} \to \mathcal{Z}$, the **composition** $g \circ f$ is the function $\mathcal{X} \to \mathcal{Z}$ defined by $(g \circ f)(x) = g(f(x))\ \forall x \in \mathcal{X}$.
- We say $f : \mathcal{X} \to \mathcal{Y}$ is **injective** (or **one-to-one**) if different elements of $\mathcal{X}$ go to different elements of $\mathcal{Y}$; so for $x_1, x_2 \in \mathcal{X}$, if $f(x_1) = f(x_2)$, then we must have $x_1 = x_2$.
- We say $f : \mathcal{X} \to \mathcal{Y}$ is **surjective** (or **onto**) if, given any $y \in \mathcal{Y}$, there is some $x \in \mathcal{X}$ with $f(x) = y$.
- We say f is **bijective** if it is both injective and surjective. In this case, for every $y \in \mathcal{Y}$ there is a *unique* $x \in \mathcal{X}$ such that $f(x) = y$; sending y to this unique x defines a map $f^{-1} : \mathcal{Y} \to \mathcal{X}$, the **inverse mapping**, with the property that $f^{-1} \circ f = \mathrm{Id}_{\mathcal{X}}$ and $f \circ f^{-1} = \mathrm{Id}_{\mathcal{Y}}$.

Images and preimages

We'll also need the notion of *images* and *preimages*. 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

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

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 $\mathrm{Im}(f)$ instead of $f(\mathcal{X})$ and simply speak of *the image of* f.

Similarly, the *preimage* of $\mathcal{Z}$ under f is the set

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

consisting of all the elements in $\mathcal{X}$ which are mapped onto elements of $\mathcal{Z}$ under f.

Remark 1.1 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 $f^{-1}(\mathcal{Z})$ for the preimage is customary (and it agrees with "the image of $\mathcal{Z}$ under f^{-1}" when a function f^{-1} does exist).

Exercises

Optional, for review purposes. See website `https://apptest.fernuni.ch/` *for worked solutions*

Exercise 1.1 Which of the following functions are injective? Which are surjective? Give a simple justification (detailed proofs are not required).

(i) The function $\mathbb{N} \to \mathbb{N}$ given by $f(x) = x + 1$.
(ii) The function $\mathbb{R} \to \mathbb{R}$ given by $f(x) = x + 1$.
(iii) The function $\mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$.
(iv) The function $\mathbb{R} \to \mathbb{R}$ given by $f(x) = x^3$.
(v) The function $\mathbb{Q} \to \mathbb{Q}$ given by $f(x) = x^3$.

Exercise 1.2 Let $f : \mathcal{X} \to \mathcal{Y}$ be any function.

(i) Show that if there exists $g : \mathcal{Y} \to \mathcal{X}$ with $g \circ f = \mathrm{Id}_{\mathcal{X}}$, then f is injective. Give an example to show that f need not be surjective.
(ii) Show that if there exists $g : \mathcal{Y} \to \mathcal{X}$ with $f \circ g = \mathrm{Id}_{\mathcal{Y}}$, then f is surjective. Give an example to show that f need not be injective.
(iii) If $f : \mathcal{X} \to \mathcal{Y}$ is injective, and $g_1, g_2 : \mathcal{Y} \to \mathcal{X}$ are functions with $g_1 \circ f = g_2 \circ f = \mathrm{Id}_{\mathcal{X}}$, does it follow that $g_1 = g_2$? Give a proof or counterexample as appropriate.

Exercise 1.3 Suppose $f : \mathcal{X} \to \mathcal{Y}$ is a function.
(i) If f is injective, does there always exist a $g : \mathcal{Y} \to \mathcal{X}$ with $g \circ f = \mathrm{Id}_{\mathcal{X}}$?
(ii) If f is surjective, does there always exist a g with $f \circ g = \mathrm{Id}_{\mathcal{Y}}$?
(Warning: there is a trap for the unwary here!)

1.2 Fields

Definitions

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

Definition 1.2 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

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

- Commutativity of multiplication

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

- Associativity of addition

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

- Associativity of multiplication

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

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

(1.5) $$x +_{\mathbb{K}} 0_{\mathbb{K}} = 0_{\mathbb{K}} +_{\mathbb{K}} x = x \quad \text{for all } x \in \mathbb{K}.$$

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

(1.6) $$x \cdot_{\mathbb{K}} 1_{\mathbb{K}} = 1_{\mathbb{K}} \cdot_{\mathbb{K}} x = x \quad \text{for all } x \in \mathbb{K}.$$

- For any $x \in \mathbb{K}$ there exists an element $y \in \mathbb{K}$ such that

(1.7) $$x +_{\mathbb{K}} y = 0_{\mathbb{K}}.$$

  It follows that there is a *unique* such element, for any given x; and we denote it by $(-x)$, the *additive inverse* of x.
- For any $x \in \mathbb{K} \setminus \{0_{\mathbb{K}}\}$ there exists an element y such that

(1.8) $$x \cdot_{\mathbb{K}} y = y \cdot_{\mathbb{K}} x = 1_{\mathbb{K}}.$$

  Again, this element is necessarily uniquely determined and we denote it by x^{-1} or $\frac{1}{x}$, the *multiplicative inverse* of x,
- Distributivity of multiplication over addition

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

Remark 1.3

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

(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 1.4

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

(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.

First properties

Lemma 1.5 (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 (1.5), 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 (1.9). 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 (1.3) and (1.7), 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 (1.8), 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 (1.6), the commutativity (1.2) and associativity (1.4) of multiplication as well as (i) from above.

(v) By (1.6), 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 $(1_{\mathbb{K}})^{-1} = 1_{\mathbb{K}}$. □

Remark 1.6 (Characteristic of a field) For $n \in \mathbb{N}$ and an element x of a field $\mathbb{K}$, we write

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

(We understand this as $nx = 0_{\mathbb{K}}$ if $n = 0$.) Note that the field $\mathbb{F}_2$ has the property that $2x = 0$ for all $x \in \mathbb{F}_2$.
For a field $\mathbb{K}$, we define the *characteristic* of $\mathbb{K}$ to be the smallest positive integer p such that $px = 0_{\mathbb{K}}$ for all $x \in \mathbb{K}$, if such an integer exists. If no such integer exists the field is said to have characteristic 0.
So $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are fields of characteristic 0, while $\mathbb{F}_2$ and $\mathbb{F}_4$ both have characteristic 2. 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 1.7

(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 $\{0, 1\}$.

Mappings between fields

Generally, whenever we define some kind of 'set with extra structure' – like a group or a field – it's interesting to look at mappings which preserve these structures. This leads to the notion of a *field embedding*:

Definition 1.8 (Field embedding) Let $\mathbb{F}$ and $\mathbb{K}$ be fields. A *field embedding* is a mapping $\iota : \mathbb{F} \to \mathbb{K}$ satisfying the conditions $\iota(1_\mathbb{F}) = 1_\mathbb{K}$, $\iota(0_\mathbb{F}) = 0_\mathbb{K}$, and

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

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

Example 1.9

- The obvious inclusion of $\mathbb{Q}$ inside $\mathbb{R}$ is a field embedding.
- From the above tables we see that $\iota : \mathbb{F}_2 \to \mathbb{F}_4$ defined by $\iota(1_{\mathbb{F}_2}) = 1_{\mathbb{F}_4}$ and $\iota(0_{\mathbb{F}_2}) = 0_{\mathbb{F}_4}$ is a field embedding.

Remark 1.10

(i) A field embedding is always injective[a]. Suppose $x, y \in \mathbb{F}$ satisfy $\iota(x) = \iota(y)$ so that $\iota(x - y) = 0_\mathbb{K}$. Assume $w = x - y \neq 0_\mathbb{F}$, then $\iota(w) \cdot_\mathbb{K} \iota(w^{-1}) = \iota(1_\mathbb{F}) = 1_\mathbb{K}$. Since by assumption $\iota(w) = 0_\mathbb{K}$, we thus obtain $0_\mathbb{K} \cdot_\mathbb{K} \iota(w^{-1}) = 1_\mathbb{K}$, contradicting Lemma 1.5 (i). It follows that $x = y$ and hence ι is injective.

(ii) It turns out that we don't actually need to require the condition $\iota(0_\mathbb{F}) = 0_\mathbb{K}$ in the definition of a field embedding; it is implied by the other three conditions. Indeed, if ι satisfies the other conditions, then we have

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

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

[a] This is why the name 'field embedding' is used: in algebra, 'X embedding' generally means '*injective* map preserving X kind of structure'

1.3 Complex numbers

Motivation

You've almost certainly encountered the complex numbers, defined as something like "numbers of the form $a + b \cdot i$, where $a, b \in \mathbb{R}$ and $i^2 = -1$". However, if you think formally about this, it's problematic as a definition: what does the "$+$" in $a + bi$ mean?

We haven't defined the operations yet! So if we want to *define* the complex numbers, we're going to have to take a slightly different approach.

Definitions

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

$$(x_1, y_1) \cdot_{\mathbb{C}} (x_2, y_2) = (x_1 \cdot x_2 - y_1 \cdot y_2, x_1 \cdot y_2 + y_1 \cdot x_2). \tag{1.10}$$

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 1.11 (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*.

Let's verify that this really is a field, by verifying the field axioms are satisfied. Most of these are easy: for example, commutativity of addition – if $x = (x_1, x_2)$ and $y = (y_1, y_2)$, then we compute

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

Here we've used the commutativity-of-addition axiom for $\mathbb{R}$, which is OK, since we already know $\mathbb{R}$ is a field. We can check almost all the other axioms by similar routine calculations (exercise!).

The one which is *not* routine is existence of inverses. The trick is to notice that if $(x, y) \in \mathbb{C} \setminus \{0_{\mathbb{C}}\}$, then x and y aren't both zero; so $x^2 + y^2 > 0$ (strictly) and hence $x^2 + y^2 \neq 0$ in $\mathbb{R}$. So $(\frac{x}{x^2+y^2}, \frac{-y}{x^2+y^2})$ is a well-defined element of $\mathbb{C}$, and we can compute

$$(x, y) \cdot_{\mathbb{C}} (\tfrac{x}{x^2+y^2}, \tfrac{-y}{x^2+y^2}) = (1, 0),$$

and $(\frac{x}{x^2+y^2}, \frac{-y}{x^2+y^2})$ is an inverse of (x, y).

Putting $\mathbb{R}$ inside $\mathbb{C}$

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

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

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

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 ι, it is customary to identify[1] x with $\iota(x)$, 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

[1] We can't completely forget, though, that $\mathbb{R}$ has its own separate definition: if we tried to define $\mathbb{R}$ as a subfield of $\mathbb{C}$, *and* to define $\mathbb{C}$ as the set of ordered pairs of elements of $\mathbb{R}$, then that would be circular logic.

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 (1.10). Here we also follow the usual custom of omitting $\cdot_{\mathbb{R}}$ on the right hand side. We can now understand where the funny formula for inverses came from:

$$\frac{1}{x + iy} = \frac{x - iy}{(x + iy)(x - iy)} = \frac{x - iy}{x^2 + y^2}.$$

Remark 1.12 This last manipulation is a good way of *remembering* the formula for inverses, but it's not a *proof* in itself: we need to prove that inverses exist in $\mathbb{C}$ before we can legally write down a fraction!

Definition 1.13 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 1.14

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

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

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

Example 1.15 Let $z = \frac{2+5\mathrm{i}}{6-\mathrm{i}}$. Then

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

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 1.16

(i) We may think of a complex number $z = a + \mathrm{i}b$ 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* $\{\mathrm{i}y : y \in \mathbb{R}\}$ 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 Calculus module.

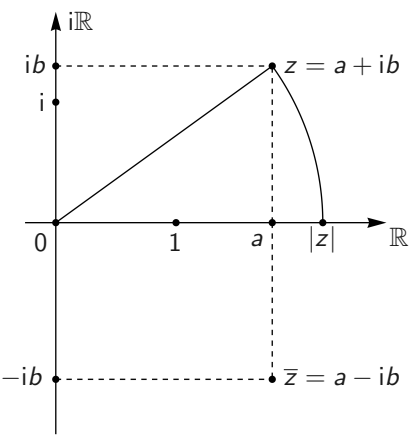

FIGURE 1.1. The complex number plane $\mathbb{C}$

We have the following elementary facts about complex numbers:

Proposition 1.17 *For all $z, w \in \mathbb{C}$ we have*

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

See website `https://apptest.fernuni.ch/` *for worked solutions*

Exercise 1.4 Check that $\mathbb{C}$ is indeed a field.

Exercise 1.5 Show that the set of pairs (x, y) with $x, y \in \mathbb{F}_2$, and addition and multiplication defined as in (1.10) above, is *not* a field.

Exercise 1.6 Let $z \in \mathbb{C}$ with $|z| = 1$. Show that there is a unique $\theta \in [0, 2\pi)$ such that $z = \cos\theta + \mathrm{i}\sin\theta$.

Exercise 1.7 Prove that there is no complex number z with $|z| = z + \mathrm{i}$.

CHAPTER 2

Matrices, I

Contents

2.1	**Definitions**	**17**
2.2	**Arithmetic with matrices**	**19**
	Matrix addition	19
	Scalar multiplication of a matrix	20
	Matrix multiplication	21
	Properties	21
2.3	**Transpose and inverse**	**23**
	Transpose	23
	Inverses	24
Exercises		**24**

In this chapter we'll recall some things you learned in "Algorithmics" about matrices and vectors, and learn some new properties.

Throughout the rest of this module, $\mathbb{K}$ stands for an arbitrary field. It won't ever matter *which* field it is; so you can assume $\mathbb{K} = \mathbb{R}$ (or $\mathbb{C}$) throughout if you prefer.

2.1 Definitions

We start with some definitions. In this chapter, m, n, p, q, r denote natural numbers.

A *matrix* (plural *matrices*) is simply a rectangular block of numbers. More precisely:

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
$$\vec{a}_1 = \begin{pmatrix} \pi \\ -1 \\ \log 2 \end{pmatrix}, \quad \vec{a}_2 = \begin{pmatrix} \sqrt{2} \\ 5/3 \\ 3 \end{pmatrix}, \quad \vec{\alpha}_2 = \begin{pmatrix} -1 & 5/3 \end{pmatrix}, \quad \vec{\alpha}_3 = \begin{pmatrix} \log 2 & 3 \end{pmatrix}.$$

We'll use the shorthand notation $\mathbb{K}^n$ for column vectors with n entries (i.e. $M_{n,1}(\mathbb{K})$), and $\mathbb{K}_n$ for row vectors $M_{1,n}(\mathbb{K})$. We can of course go back and forth between them, since there's a bijective map from $\mathbb{K}_n$ to $\mathbb{K}^n$ sending $\begin{pmatrix} x_1 & \dots & x_n \end{pmatrix}$ to $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$; but it's helpful to think of them as separate, since they will interact differently with matrix multiplication.

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*. (The notion of "diagonal" sort of makes sense for non-square matrices too, but it's most useful in the diagonal case.)
- 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 Arithmetic with matrices

Matrix addition

The sum of matrices $\mathbf{A}$ and $\mathbf{B}$ of *identical size* is defined as follows:

Definition 2.7 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

$$\mathbf{A} +_{M_{m,n}(\mathbb{K})} \mathbf{B} = (A_{ij} +_{\mathbb{K}} B_{ij})_{1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n} \in M_{m,n}(\mathbb{K}), \tag{2.2}$$

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

Remark 2.8 (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.
(Similarly, the notations $\vec{\varepsilon}_1$ etc for standard basis vectors are slightly ambiguous since we haven't specified n, but $(1 \quad 0)$ and $(1 \quad 0 \quad 0)$ are not literally the same vector. Whenever we use these notations it will always be clear from context what n is.)

Scalar multiplication of a matrix

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 2.9 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 2.10 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. However, in a moment we'll encounter a more general kind of multiplication where this isn't true.

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:

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

In particular, since row vectors and column vectors are just special types of matrix, we can multiply those by scalars; and we have the following easy but useful lemma:

Lemma 2.12 *For any row vector* $\vec{\xi} = (\xi_1 \quad \xi_2 \quad \dots \quad \xi_n) \in \mathbb{K}_n$ *we have* $\vec{\xi} = \sum_{j=1}^{n} \xi_j \vec{\varepsilon}_j$*, and similarly for column vectors.*

Proof Clear. □

Matrix multiplication

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

Definition 2.13 (Matrix multiplication) Let $\mathbf{A} \in M_{m,n}(\mathbb{K})$ be an m-by-n matrix and $\mathbf{B} \in M_{n,r}(\mathbb{K})$ be an n-by-r matrix. The *matrix product* of $\mathbf{A}$ and $\mathbf{B}$ is the m-by-r matrix $\mathbf{AB} \in M_{m,r}(\mathbb{K})$ 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 $1 \leqslant k \leqslant r$.

This definition might seem a little weird and arbitrary at first sight, but will turn out to give us a nice theory. (Perhaps the best motivation for it will come much later in the module, in Theorem 7.6.)

Remark 2.14 (Matrix multiplication is not commutative) 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 $\mathbf{AB} \neq \mathbf{BA}$. In general $\mathbf{AB}$ and $\mathbf{BA}$ aren't even the same size, since $\mathbf{AB}$ is an m-by-m matrix and $\mathbf{BA}$ is an n-by-n matrix. Even when $n = m$, so that $\mathbf{AB}$ and $\mathbf{BA}$ are the same size, it is still false in general that $\mathbf{AB} = \mathbf{BA}$.

Remark 2.15 (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

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

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 $\mathbf{B} \in M_{n,r}(\mathbb{K})$ 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 $\mathbf{B}$.

Properties

Here is a long list of (mostly quite straightforward) properties of these matrix operations. Here $\mathbb{K}$ is any field, and m, n, r, s are natural numbers.

Proposition 2.16 (Properties of matrix operations)

(i) $\mathbf{0}_{m,n} + \mathbf{A} = \mathbf{A}$ *for all* $\mathbf{A} \in M_{m,n}(\mathbb{K})$*;*
(ii) $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$ *and* $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$ *for all* $\mathbf{A}, \mathbf{B}, \mathbf{C} \in M_{m,n}(\mathbb{K})$*;*
(iii) $\mathbf{1}_m\mathbf{A} = \mathbf{A}$ *and* $\mathbf{A}\mathbf{1}_n = \mathbf{A}$ *for all* $\mathbf{A} \in M_{m,n}(\mathbb{K})$*;*
(iv) $\mathbf{0}_{r,m}\mathbf{A} = \mathbf{0}_{r,n}$ *and* $\mathbf{A}\mathbf{0}_{n,r} = \mathbf{0}_{m,r}$ *for all* $\mathbf{A} \in M_{m,n}(\mathbb{K})$*;*
(v) ("Associativity of multiplication") $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$ *for all* $\mathbf{A} \in M_{m,p}(\mathbb{K})$*,* $\mathbf{B} \in M_{p,q}(\mathbb{K})$ *and* $\mathbf{C} \in M_{q,n}(\mathbb{K})$*;*
(vi) $(\mathbf{B} + \mathbf{C})\mathbf{A} = \mathbf{BA} + \mathbf{CA}$ *for all* $\mathbf{B}, \mathbf{C} \in M_{r,m}(\mathbb{K})$ *and for all* $\mathbf{A} \in M_{m,n}(\mathbb{K})$*;*
(vii) $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$ *for all* $\mathbf{A} \in M_{r,m}(\mathbb{K})$ *and for all* $\mathbf{B}, \mathbf{C} \in M_{m,n}(\mathbb{K})$*.*
(viii) $0_{\mathbb{K}} \cdot \mathbf{A} = \mathbf{0}_{m,n}$ *for all* $\mathbf{A} \in M_{m,n}(\mathbb{K})$*;*
(ix) $(s_1 s_2)\mathbf{A} = s_1(s_2\mathbf{A})$ *for all* $\mathbf{A} \in M_{m,n}(\mathbb{K})$ *and all* $s_1, s_2 \in \mathbb{K}$*;*
(x) $\mathbf{A}(s\mathbf{B}) = s(\mathbf{AB}) = (s\mathbf{A})\mathbf{B}$ *for all* $\mathbf{A} \in M_{m,n}(\mathbb{K})$ *and all* $\mathbf{B} \in M_{n,r}(\mathbb{K})$ *and all* $s \in \mathbb{K}$*;*
(xi) $s(\mathbf{A} + \mathbf{B}) = s\mathbf{A} + s\mathbf{B}$ *for all* $\mathbf{A}, \mathbf{B} \in M_{m,n}(\mathbb{K})$ *and* $s \in \mathbb{K}$*;*
(xii) $(s_1 + s_2)\mathbf{A} = s_1\mathbf{A} + s_2\mathbf{A}$ *for all* $\mathbf{A} \in M_{m,n}(\mathbb{K})$ *and for all* $s_1, s_2 \in \mathbb{K}$*.*

Remark 2.17 Some of these only involve one of our three matrix operations, and some involve several. (*Exercise*: make a table showing which properties involve which operations!)
Because of (ii) and (v), we don't need to write brackets when we deal with sums (or products) of three or more matrices – we can just write $\mathbf{A}+\mathbf{B}+\mathbf{C}$ or $\mathbf{ABC}$, assuming the matrices are of compatible sizes so the operations make sense (and similarly for four or more matrices).

Proof As a sample, we show properties (iii) and (vii), which are quite easy, and (v), which is slightly harder. The proofs of the remaining ones are similar and/or elementary consequences of the properties of addition and multiplication of scalars.

To show property (iii), 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 property (vii), applying the definition of matrix multiplication gives

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

so that

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

where we use that

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

For property (v), if $A_{ij}=[\mathbf{A}]_{ij}$ etc, we have

$$[(\mathbf{AB})\mathbf{C}]_{ij}=\sum_{b=1}^{q}[(\mathbf{AB})]_{ib}[\mathbf{C}]_{bj}=\sum_{b=1}^{q}\left(\sum_{a=1}^{p}A_{ia}B_{ab}\right)C_{bj}$$

This is the sum over all possible products $A_{ia}B_{ab}C_{bj}$ (with i,j fixed and a,b varying); we can group together these terms in whichever order we like, so we take the sum over a to the outside:

$$\cdots=\sum_{a=1}^{p}A_{ia}\left(\sum_{b=1}^{q}B_{ab}C_{bj}\right)=\sum_{a=1}^{p}[\mathbf{A}]_{ia}[\mathbf{BC}]_{aj}=[\mathbf{A}(\mathbf{BC})]_{ij}$$

as required. □

2.3 Transpose and inverse

Transpose

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

Definition 2.18 (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 2.19 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 2.20 (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 $\mathbf{B}\in M_{n,r}(\mathbb{K})$, we have
$$(\mathbf{AB})^T=\mathbf{B}^T\mathbf{A}^T.$$

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

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

Transposes won't play a very big role in this course, but they will be much more important in Linear Algebra II when you start studying *bilinear* mappings.

Inverses

Definition 2.21 Let $\mathbf{A} \in M_{m,n}(\mathbb{K})$ be a matrix. If there exists a matrix $\mathbf{B} \in M_{n,m}(\mathbb{K})$ with $\mathbf{AB} = \mathbf{1}_m$ and $\mathbf{BA} = \mathbf{1}_n$, then we say $\mathbf{A}$ is *invertible*.

If such a matrix exists (for a given $\mathbf{A}$) then it's unique (see Exercises). So we can denote this unique matrix (if it exists!) by $\mathbf{A}^{-1}$, and we call it the *inverse* of $\mathbf{A}$.

Proposition 2.22 *If $\mathbf{A} \in M_{m,n}(\mathbb{K})$ and $\mathbf{B} \in M_{n,r}(\mathbb{K})$ are both invertible, then $\mathbf{AB}$ is invertible and $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.*

Proof We compute (using part (v) of Proposition 2.16 repeatedly) that $(\mathbf{B}^{-1}\mathbf{A}^{-1})(\mathbf{AB}) = \mathbf{B}^{-1}(\mathbf{A}^{-1}\mathbf{A})\mathbf{B} = \mathbf{B}^{-1}\mathbf{1}_n\mathbf{B} = \mathbf{B}^{-1}\mathbf{B} = \mathbf{1}_r$, and similarly $(\mathbf{AB}) \cdot (\mathbf{B}^{-1}\mathbf{A}^{-1}) = \mathbf{1}_m$. □

Remark 2.23 You saw these definitions already in Algorithmics, assuming $\mathbb{K} = \mathbb{R}$ and, more importantly, that $m = n$. The definition still makes logical sense if $m \neq n$, but we'll see later that it is *never satisfied* – it is a theorem that a non-square matrix cannot be invertible.

Exercises

See website `https://apptest.fernuni.ch/` *for worked solutions*

Exercise 2.1 Find two 2×2 matrices $\mathbf{A}$ and $\mathbf{B}$ with $\mathbf{AB} \neq \mathbf{BA}$.

Exercise 2.2 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 $\mathbf{A}^{-1}$.

Exercise 2.3 Let $\mathbf{A} = (1 \quad 0) \in M_{1,2}(\mathbb{R})$.

(i) Find a matrix $\mathbf{B} \in M_{2,1}(\mathbb{R})$ with $\mathbf{AB} = \mathbf{1}_1$.

(ii) Is the matrix $\mathbf{B}$ from (i) uniquely determined?

(iii) Show that there does *not* exist a matrix $\mathbf{B} \in M_{2,1}(\mathbb{R})$ with $\mathbf{BA} = \mathbf{1}_2$.

(We say that $\mathbf{A}$ has a *right inverse* but not a *left inverse*.)

Exercise 2.4 Let $\mathbf{A} \in M_{m,n}(\mathbb{K})$. Suppose there exist matrices $\mathbf{B}, \mathbf{B}' \in M_{n,m}(\mathbb{K})$ such that $\mathbf{AB} = \mathbf{1}_m$ and $\mathbf{B}'\mathbf{A} = \mathbf{1}_n$. Show that we must have $\mathbf{B} = \mathbf{B}'$. (Hint: Consider $\mathbf{B}'\mathbf{AB}$.) Hence show that the inverse of a matrix is unique if it exists.

CHAPTER 3

Matrices, II

Contents

3.1 Row echelon form **26**
Definition 26
Echelonizing a matrix via Gauss–Jordan 27
Keeping track of the transformation matrix 28
Uniqueness 29
3.2 Solving equations **30**
3.3 Inverting a matrix **33**
Non-square matrices 33
Square matrices 33
Exercises **34**

3.1 Row echelon form

We'll now introduce an important way of calculating with matrices – an extension of the *Gaussian elimination* which you saw in the Algorithmics module – which is going to be the key to virtually all the computations we'll do in linear algebra.

Definition

Definition 3.1 Let $\mathbf{M}$ and $\mathbf{N}$ be matrices in $M_{m,n}(\mathbb{K})$. We say $\mathbf{M}$ and $\mathbf{N}$ are *left-equivalent* if there exists an *invertible* $\mathbf{A} \in M_{m,m}(\mathbb{K})$ such that $\mathbf{AM} = \mathbf{N}$.

One can show that "left equivalence" is an *equivalence relation* in the sense of Algorithmics, Section 2; that is, we have the following properties:

- (Reflexivity) Every matrix is left-equivalent to itself.
- (Symmetry) If $\mathbf{M}$ is left-equivalent to $\mathbf{N}$, then $\mathbf{N}$ is left-equivalent to $\mathbf{M}$.
- (Transitivity) If $\mathbf{M}$ is left-equivalent to $\mathbf{N}$, and $\mathbf{N}$ is left-equivalent to $\mathbf{R}$, then $\mathbf{M}$ is left-equivalent to $\mathbf{R}$.

The idea of this section is to show that among all the matrices that are left-equivalent to a given $\mathbf{M}$, there is a unique "nicest" one.

Definition 3.2 For each row in a matrix, if the row does not consist of zeros only, then the leftmost nonzero entry is called the *leading entry* of that row.

Definition 3.3 (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 below all the non-zero rows;
- the leading entry of a nonzero row is always strictly to the right of the leading entry 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 entries are equal to 1;
- in every column containing a leading entry, all of the other entries in that column are zero.

The basic theorem about REF and RREF is the following one:

Theorem 3.4 *For any $\mathbf{M} \in M_{m,n}(\mathbb{K})$, there exists a **unique** matrix $\mathbf{N} \in M_{m,n}(\mathbb{K})$ with the following two properties:*

- $\mathbf{N}$ *is in reduced row echelon form, and*
- $\mathbf{N}$ *is left-equivalent to* $\mathbf{M}$.

We say $\mathbf{N}$ is the reduced row echelon form of $\mathbf{M}$. Moreover, there is an explicit algorithm for computing $\mathbf{N}$, given $\mathbf{M}$.

Remark 3.5 By the definition of left-equivalence, there must be some invertible $\mathbf{A}$ that multiplies $\mathbf{M}$ into its RREF; but this $\mathbf{A}$ is *not* unique in general. It's somehow a miracle that $\mathbf{N}$ is unique, even though $\mathbf{A}$ isn't. The extreme case is when $\mathbf{M} = \mathbf{0}_{mn}$; then $\mathbf{M}$ is already in RREF, but $\mathbf{AM} = \mathbf{M}$ for any matrix $\mathbf{A}$, so we could take $\mathbf{A}$ to be any invertible matrix we like.

Echelonizing a matrix via Gauss–Jordan

We'll first prove the "existence" part of the theorem. You already saw how to convert a matrix into row echelon form (Gaussian elimination); to get *reduced* row echelon form, we'll use a slight refinement, Gauss–Jordan elimination. This relies on the following tools:

Proposition 3.6 (Elementary row operations) *Let $\mathbf{M} \in M_{m,n}(\mathbb{K})$. If $\mathbf{M}'$ is obtained from $\mathbf{M}$ by any one of the following operations, then $\mathbf{M}'$ is left-equivalent to $\mathbf{M}$:*

- *Interchanging the i-th and j-th rows of $\mathbf{M}$, for any $i \neq j \in \{1, \dots, m\}$;*
- *Multiplying all the entries of the i-th row by λ, for some $\lambda \neq 0 \in \mathbb{K}$;*
- *Adding λ times the j-th row of $\mathbf{M}$ to the i-th row, for any $i \neq j \in \{1, \dots, m\}$ and $\lambda \in \mathbb{K}$.*

Proof Each of these operations corresponds to left-multiplying $\mathbf{M}$ by one of the three kinds of *elementary matrices* which you learned about in the M01 course. For instance, adding λ times the j-th row of $\mathbf{M}$ to the i-th row corresponds to multiplying by the matrix with 1's down the diagonal, λ in the (i, j) position, and all other entries zero. □

Proof of existence of RREF Using the "transitivity" property above, if we apply any finite sequence of elementary row operations to $\mathbf{M}$, we still get a matrix left-equivalent to

$\mathbf{M}$. We use this to gradually transform $\mathbf{M}$, one column at a time, to get it into RREF. More precisely, we'll prove the following:

> For any $0 \leqslant r \leqslant n$, we can apply a finite sequence of row operations to $\mathbf{M}$ to obtain a matrix $\mathbf{M}_r$ with the following property: the $m \times r$ matrix given by the first r columns of $\mathbf{M}_r$ is in reduced row echelon form.

We'll prove this by induction on r. The statement is vacuous for $r = 0$: a matrix without any columns is certainly in RREF, so $\mathbf{M}_0 = \mathbf{M}$ will do. So let's assume that we have already transformed $\mathbf{M}$ into a matrix $\mathbf{M}_r$ whose first r columns are in RREF, for some $r < n$, and try to deal with the $(r+1)$-st column.

Let $h \leqslant m$ be the number of nonzero rows in the left-hand $m \times r$ submatrix, so the first r columns are all zero below the h-th entry. Then we can visualise $\mathbf{M}_r$ as follows:

$$\mathbf{M}_r = \left(\begin{array}{c|c} \overbrace{\begin{pmatrix}\text{RREF,}\\ \text{no zero rows}\end{pmatrix}}^{r} & \overbrace{\begin{matrix}(?)\end{matrix}}^{(n-r)} \\ \hline (\mathbf{0}) & (?) \end{array}\right) \begin{matrix} \\ \}h \\ \}(m-h) \end{matrix}$$

If $h = m$ (which is certainly possible), then the whole matrix $\mathbf{M}_r$ is in RREF already, so we can set $\mathbf{M}_{r+1} = \mathbf{M}_r$ and go on. If not, then we home in on the first column in the bottom right $(m-h) \times (n-r)$ submatrix of $\mathbf{M}_r$. Two things can now happen:

- Case A: all these entries are zero. In this case, the first $(r+1)$ columns of $\mathbf{M}_r$ are already in RREF; so we can set $\mathbf{M}_{r+1} = \mathbf{M}_r$, and the induction step is complete.
- Case B: at least one of these $(m-h)$ entries is non-zero (so in particular $h < m$).

  Swapping the $(h+1)$-st row with one of the rows further down if necessary, we can assume that this nonzero entry is the top left corner entry of our submatrix (i.e. in position $(h+1, r+1)$ of $\mathbf{M}_r$). By multiplying the $(h+1)$-th row by a suitable scalar, we can make this nonzero entry be 1. This row swap and scaling doesn't change anything in the leftmost r columns – we're just moving zeroes around – so the first r columns are still in RREF.

  We now kill off all the other non-zero entries in the $(r+1)$-th column, by subtracting a suitable multiple of the $(h+1)$-th row. Again, this doesn't change anything in the first r columns, because the $(h+1)$-th row has zeroes in these positions; so the resulting matrix $\mathbf{M}_{r+1}$ has its first $(r+1)$ columns in RREF.

After n steps of this process we end up with a matrix which is fully in RREF. □

Remark 3.7 Notice that this proof doesn't just abstractly show you that the RREF exists; it gives you a completely explicit recipe for finding it.
Since calculating the RREF of a matrix is such a basic tool in linear algebra calculations, any half-decent computer mathematics package will include a RREF routine; often several different ones, tailored to matrices of particular specific shapes or with entries in specific fields.

Keeping track of the transformation matrix

In principle, whenever you Gauss–Jordan eliminate a matrix, you can make a list of the elementary row operations you used, and the corresponding elementary matrices; this gives you a list of elementary matrices $(\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_k)$ such that $\mathbf{A}_k \mathbf{A}_{k-1} \dots \mathbf{A}_1 \mathbf{M} = \mathbf{N}$

is in RREF. However, this quickly gets quite tedious to do, and there's a much better method.

Definition 3.8 For $\mathbf{A} \in M_{m,n}(\mathbb{K})$ and $\mathbf{B} \in M_{m,r}(\mathbb{K})$, we write $(\mathbf{A} \mid \mathbf{B})$ for the $m \times (n+r)$ matrix given by joining $\mathbf{A}$ and $\mathbf{B}$ together (so the (i,j) entry is A_{ij} for $1 \leqslant j \leqslant n$, and $B_{i,j-n}$ for $n+1 \leqslant j \leqslant n+r$).

E.g. if we have

$$\mathbf{A} = \begin{pmatrix} 1 & 3 & 2 \\ 2 & 0 & 1 \\ 5 & 2 & 2 \end{pmatrix} \qquad \mathbf{B} = \begin{pmatrix} 4 \\ 3 \\ 1 \end{pmatrix},$$

then

$$(\mathbf{A} \mid \mathbf{B}) = \left(\begin{array}{ccc|c} 1 & 3 & 2 & 4 \\ 2 & 0 & 1 & 3 \\ 5 & 2 & 2 & 1 \end{array}\right).$$

It's common to write a line in between the entries, as above, to remind ourselves which entries came from where.

Proposition 3.9 *For any matrices $\mathbf{A} \in M_{m,n}(\mathbb{K})$, $\mathbf{B} \in M_{n,r}(\mathbb{K})$, and $\mathbf{C} \in M_{n,s}(\mathbb{K})$, we have*

$$\mathbf{A} \cdot (\mathbf{B} \mid \mathbf{C}) = (\mathbf{AB} \mid \mathbf{AC}).$$

Proof The j-th column of $\mathbf{A} \cdot (\mathbf{B} \mid \mathbf{C})$ is given by $\mathbf{A} \cdot \vec{v_j}$ where $\vec{v_j}$ is the j-th column of $(\mathbf{B} \mid \mathbf{C})$. Considering the cases $1 \leqslant j \leqslant r$ and $r+1 \leqslant j \leqslant r+s$ separately, we obtain either a column of $\mathbf{AB}$ or a column of $\mathbf{AC}$. □

Proposition 3.10 *Given any $\mathbf{M} \in M_{m,n}(\mathbb{K})$, let $\tilde{\mathbf{M}} = (\mathbf{M} \mid \mathbf{1}_m)$, and let $\tilde{\mathbf{N}}$ be the RREF of $\tilde{\mathbf{M}}$. Then we have $\tilde{\mathbf{N}} = (\mathbf{N} \mid \mathbf{A})$, where $\mathbf{N}$ is the RREF of $\mathbf{M}$, and $\mathbf{A}$ is a matrix such that $\mathbf{AM} = \mathbf{N}$.*

Proof Let $\mathbf{A}$ be any invertible matrix such that $\tilde{\mathbf{N}} = \mathbf{A}\tilde{\mathbf{M}}$ is in RREF. From the proposition, we have $\mathbf{A}\tilde{\mathbf{M}} = (\mathbf{AM} \mid \mathbf{A}\mathbf{1}_m) = (\mathbf{AM} \mid \mathbf{A})$. However, if $\tilde{\mathbf{N}}$ is an RREF matrix, then its first n columns are also an RREF matrix, so $\mathbf{AM}$ must be the unique RREF matrix left-equivalent to $\mathbf{M}$. □

Remark 3.11 We can take a slight shortcut here: since Gauss–Jordan elimination transforms a matrix into RREF one column at a time, we can stop as soon as the first n columns of $\tilde{M}$ are in RREF – we don't have to keep going all the way up to the $(m+n)$-th column.

Uniqueness

We'll now prove the uniqueness property of reduced row echelon form. This proof is **non-examinable** (marked by a dark green line down the margin in the printed notes); but you should definitely be aware of the statement of the theorem!

We use the following easy remark:

Proposition 3.12 *Assume that* $\mathbf{M}$ *and* $\mathbf{N}$ *are left-equivalent, and let* $j \in \{1, \dots, n\}$. *Then the matrices* $\mathbf{M}'$ *and* $\mathbf{N}'$ *given by deleting the* j*-th columns from* $\mathbf{M}$ *and* $\mathbf{N}$ *are also left-equivalent.*

Proof Exercise. □

Proof of uniqueness of RREF Now suppose $\mathbf{M}$ and $\mathbf{N}$ are left-equivalent matrices, both in reduced row echelon form, with $\mathbf{M} \neq \mathbf{N}$. Let the t-th column, for $1 \leqslant t \leqslant n$, be the first (leftmost) column where the two matrices differ; and consider the new matrices $\mathbf{M}'$, $\mathbf{N}'$ given by deleting all columns strictly to the right of the t-th, and all columns to the left which don't contain a leading entry. Then we have $\mathbf{M}' \neq \mathbf{N}'$; both are still in RREF; and we have

$$\mathbf{M}' = \begin{pmatrix} \mathbf{1}_h & \vec{r} \\ 0 & \vec{s} \end{pmatrix}, \qquad \mathbf{N}' = \begin{pmatrix} \mathbf{1}_h & \vec{r'} \\ 0 & \vec{s'} \end{pmatrix}$$

for some $h < t$ and some column vectors $\vec{r}, \vec{s}, \vec{r'}, \vec{s'}$.

Since $\mathbf{M}$ and $\mathbf{N}$ are left-equivalent, so are $\mathbf{M}'$ and $\mathbf{N}'$ (by the last Proposition), so there exists an invertible $\mathbf{A}$ with $\mathbf{A}\mathbf{M}' = \mathbf{N}'$. Let's think about what $\mathbf{A}\mathbf{M}'$ looks like. We can divide $\mathbf{A}$ up into blocks $\begin{pmatrix} \mathbf{R} & \mathbf{S} \\ \mathbf{T} & \mathbf{U} \end{pmatrix}$, with $\mathbf{R} \in M_{h,h}(\mathbb{K})$ etc; and the product is then

$$\begin{pmatrix} \mathbf{R} & \mathbf{S} \\ \mathbf{T} & \mathbf{U} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{1}_h & \vec{r} \\ 0 & \vec{s} \end{pmatrix} = \begin{pmatrix} \mathbf{R} & \mathbf{R}\vec{r} + \mathbf{S}\vec{s} \\ \mathbf{T} & \mathbf{T}\vec{r} + \mathbf{U}\vec{s} \end{pmatrix}.$$

Let's suppose this is equal to $\mathbf{N}'$. Then, comparing the top-left and bottom-left blocks, we must have $\mathbf{R} = \mathbf{1}_h$ and $\mathbf{T} = \mathbf{0}_{m-h,h}$; so this becomes

$$\begin{pmatrix} \mathbf{1}_h & \vec{r} + \mathbf{S}\vec{s} \\ 0 & \mathbf{U}\vec{s} \end{pmatrix} = \begin{pmatrix} \mathbf{1}_h & \vec{r'} \\ 0 & \vec{s'} \end{pmatrix}.$$

If $\vec{s} \neq 0$, then $\mathbf{U}\vec{s} \neq 0$ also, since the invertibility of $\mathbf{A}$ implies that $\mathbf{U}$ is also invertible (see Exercise below). Since $\mathbf{M}'$ and $\mathbf{N}'$ are in RREF, both $\vec{r}$ and $\vec{r'}$ have to be zero, and $\vec{s}$ and $\vec{s'}$ are both equal to the standard basis vector $\vec{e}_1$; so in fact $\mathbf{M}' = \mathbf{N}'$, contradicting our assumptions.

On the other hand, if $\vec{s} = 0$, then $\vec{s'} = \mathbf{U}\vec{s} = 0$ as well; and substituting this into the top right block, we have $\vec{r'} = \vec{r} + \mathbf{S}\vec{s} = \vec{r}$. So again $\mathbf{M}' = \mathbf{N}'$, contradiction. So we can conclude that it is impossible for two distinct RREF matrices to be left-equivalent. □

3.2 Solving equations

You already saw in Algorithmics that matrices can be used to "package together" systems of linear equations. Suppose we have a system of simultaneous equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1, \\ \vdots \qquad &= \quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m, \end{aligned}$$

with (known) values $a_{ij} \in \mathbb{K}$ and $b_i \in \mathbb{K}$, and we want to solve this for the unknowns x_j. Then we can write this simply as $\mathbf{A}\vec{x} = \vec{b}$, with $\mathbf{A} = (A_{ij}) \in M_{m,n}(\mathbb{K})$, $\vec{b} = (b_i)$ and

$\vec{x} = (x_j)$; and we write $\mathrm{Sol}(\mathbf{A}, \vec{b})$ for the set of solutions to this equation, i.e.

$$\mathrm{Sol}(\mathbf{A}, \vec{b}) = \{\vec{x} \in \mathbb{K}^n : \mathbf{A}\vec{x} = \vec{b}\}.$$

If $\mathbf{B}$ is an invertible square matrix, then we have $\mathrm{Sol}(\mathbf{BA}, \mathbf{B}\vec{b}) = \mathrm{Sol}(\mathbf{A}, \vec{b})$. So, if we want to compute $\mathrm{Sol}(\mathbf{A}, \vec{b})$, it's a good idea to choose some $\mathbf{B}$ such that $\mathbf{BA}$ and $\mathbf{B}\vec{b}$ are as simple as possible. So we form the augmented matrix $(\mathbf{A} \,|\, \vec{b})$ and put that into echelon form; this gives a new, echelonized system of equations where we can immediately read off the solutions.

Proposition 3.13 *Suppose $\mathbf{A}$ is in REF (not necessarily RREF), and let $r \leqslant m$ be the number of non-zero rows of $\mathbf{A}$. Then the equation $\mathbf{A}\vec{x} = \vec{b}$ has a solution if, and only if, we have $b_i = 0$ for all i with $r < i \leqslant m$.*

(This condition is vacuously satisfied if $r = m$, since there are no such i, so solutions always exist in this case.) It's clear that "$b_i = 0$ for all i with $r < i \leqslant m$" is a *necessary* condition for solutions to exist, since for i in this range the i-th equation in our system is just $0 = b_i$. What's less obvious is that it is a *sufficient* condition, which is shown in Chapter 9 of Algorithmics.

Remark 3.14 We say the system of equations is *inconsistent* if one of $b_{r+1}, \ldots, b_m$ is nonzero, so the solution set is empty.

no solution

my mind is a matrix
that has been reduced
into row echelon form
and proven to be
— inconsistent

(from *More Math Poems* by Eileen Tupaz)

Of course, we don't just want to know whether solutions exist: we want to find them! Having found a REF for $\mathbf{A}$, we say x_j is a *free variable* if there is no row of the REF whose leading term is in the j-th column. Then it's easy to check that for *any* values of the free variables, there is a unique way to fill in the rest of the variables to get a solution $\vec{x}$.

Example 3.15 *"Find all real numbers $x_1, \ldots, x_4$ satisfying the system of linear equations*

$$\begin{aligned} -9x_2 + 3x_3 + 4x_4 &= 9 \\ x_1 + 4x_2 - x_4 &= 5 \\ 2x_1 + 6x_2 - x_3 + 5x_4 &= -5\,." \end{aligned}$$

The augmented matrix is

$$\left(\begin{array}{cccc|c} 0 & -9 & 3 & 4 & 9 \\ 1 & 4 & 0 & -1 & 5 \\ 2 & 6 & -1 & 5 & -5 \end{array}\right)$$

Let's walk through the steps of echelonizing this:

- Swap rows 1 and 2 to get $\left(\begin{array}{cccc|c} 1 & 4 & 0 & -1 & 5 \\ 0 & -9 & 3 & 4 & 9 \\ 2 & 6 & -1 & 5 & -5 \end{array}\right)$
- Add -2 times row 1 to row 3 to get $\left(\begin{array}{cccc|c} 1 & 4 & 0 & -1 & 5 \\ 0 & -9 & 3 & 4 & 9 \\ 0 & -2 & -1 & 7 & -15 \end{array}\right)$
- Multiply row 2 by $-\frac{1}{9}$ to get $\left(\begin{array}{cccc|c} 1 & 4 & 0 & -1 & 5 \\ 0 & 1 & -\frac{1}{3} & -\frac{4}{9} & -1 \\ 0 & -2 & -1 & 7 & -15 \end{array}\right)$
- Add -4 times row 2 to row 1, and 2 times row 2 to row 3, to get $\left(\begin{array}{cccc|c} 1 & 0 & \frac{4}{3} & \frac{7}{9} & 9 \\ 0 & 1 & -\frac{1}{3} & -\frac{4}{9} & -1 \\ 0 & 0 & -\frac{5}{3} & \frac{55}{9} & -17 \end{array}\right)$
- Multiply row 3 by $-\frac{3}{5}$ to get $\left(\begin{array}{cccc|c} 1 & 0 & \frac{4}{3} & \frac{7}{9} & 9 \\ 0 & 1 & -\frac{1}{3} & -\frac{4}{9} & -1 \\ 0 & 0 & 1 & -\frac{11}{3} & \frac{51}{5} \end{array}\right)$
- Add $-\frac{4}{3}$ of row 3 to row 1, and $\frac{1}{3}$ of row 3 to row 2, to get $\left(\begin{array}{cccc|c} 1 & 0 & 0 & \frac{17}{3} & -\frac{23}{5} \\ 0 & 1 & 0 & -\frac{5}{3} & \frac{12}{5} \\ 0 & 0 & 1 & -\frac{11}{3} & \frac{51}{5} \end{array}\right)$

So the solutions to the original system are the same as those of the echelonized system

$$\begin{aligned} x_1 + \tfrac{17}{3}x_4 &= -\tfrac{23}{5} \\ x_2 - \tfrac{5}{3}x_4 &= \tfrac{12}{5} \\ x_3 - \tfrac{11}{3}x_4 &= \tfrac{51}{5} \end{aligned}$$

Clearly, we can choose any value of x_4 we like (let's call it λ), and then read off the values of x_1, x_2, x_3 from that, so the solutions are given by

$$\vec{x} = \begin{pmatrix} -\frac{23}{5} - \frac{17}{3}\lambda \\ \frac{12}{5} + \frac{5}{3}\lambda \\ \frac{51}{5} + \frac{11}{3}\lambda \\ \lambda \end{pmatrix}$$

for any $\lambda \in \mathbb{R}$.

Here is a variation on RREF, where we have unknown parameters, not numbers, in the last column. (I am grateful to Sarah Zerbes for this example.)

Example 3.16 *"Consider the system of equations*

$$\begin{pmatrix} 1 & -2 & 1 \\ 2 & 1 & 1 \\ 0 & 5 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

where $b_1, b_2, b_3 \in \mathbb{R}$. For which values of the b_i is this solvable?"

We take the augmented matrix

$$\left(\begin{array}{ccc|c} 1 & -2 & 1 & b_1 \\ 2 & 1 & 1 & b_2 \\ 0 & 5 & -1 & b_3 \end{array}\right)$$

and apply Gauss–Jordan elimination to echelonize the first three columns. After clearing the first column we get

$$\left(\begin{array}{ccc|c} 1 & -2 & 1 & b_1 \\ 0 & 5 & -1 & -2b_1 + b_2 \\ 0 & 5 & -1 & b_3 \end{array}\right)$$

and continuing to clear the second column we get

$$\left(\begin{array}{ccc|c} 1 & 0 & \frac{3}{5} & \frac{1}{5}b_1 + \frac{2}{5}b_2 \\ 0 & 1 & -\frac{1}{5} & -\frac{2}{5}b_1 + \frac{1}{5}b_2 \\ 0 & 0 & 0 & 2b_1 - b_2 + b_3 \end{array}\right).$$

So the original system has solutions if and only if $2b_1 - b_2 + b_3 = 0$.

3.3 Inverting a matrix

Non-square matrices

We can now make good on a promise from the last chapter, by showing that a non-square matrix cannot be invertible. This follows from the following more specific theorem:

Proposition 3.17 *Suppose* $\mathbf{A} \in M_{m,n}(\mathbb{K})$.

(i) If $n > m$ *(so* $\mathbf{A}$ *has strictly more columns than it has rows), then there does not exist a matrix* $\mathbf{B} \in M_{n,m}(\mathbb{K})$ *with* $\mathbf{BA} = \mathbf{1}_n$.

(ii) If $n < m$, *then there does not exist a matrix* $\mathbf{B} \in M_{n,m}(\mathbb{K})$ *with* $\mathbf{AB} = \mathbf{1}_m$.

Proof First suppose $n > m$. Consider the system of equations $\mathbf{A} \cdot \vec{x} = \mathbf{0}$. This obviously has the solution $\vec{x} = \mathbf{0}$. But it must have other solutions too, since $\mathbf{A}$ has only m rows, so at most m of the columns of the RREF can contain a leading entry (possibly less, if there are some zero rows). Thus the general solution has some free variables in it (at least $n - m$ of them). So there exists a solution $\vec{x} \neq \mathbf{0}$.

But then we have a contradiction, since

$$\begin{aligned} \vec{x} &= \mathbf{1}_n \vec{x} \\ &= (\mathbf{B} \cdot \mathbf{A}) \cdot \vec{x} \\ &= \mathbf{B} \cdot (\mathbf{A} \cdot \vec{x}) \\ &= \mathbf{B} \cdot \mathbf{0} = \mathbf{0}. \end{aligned}$$

This proves (i).

Now let's consider $m > n$. If $\mathbf{AB} = \mathbf{1}_m$, then $\mathbf{B}^T\mathbf{A}^T = (\mathbf{1}_m)^T = \mathbf{1}_m$, and we obtain a contradiction by applying (i) to $\mathbf{A}^T \in M_{n,m}(\mathbb{K})$. □

Square matrices

For square matrices, we can use Gauss–Jordan elimination to determine if a matrix is invertible, and to compute its inverse if so. This relies on the following fact:

Proposition 3.18 *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 RREF of $\mathbf{A}$ *is the identity.*

Proof Suppose $\mathbf{A}$ is invertible. Then $\mathbf{A}^{-1}$ is itself invertible (with inverse $\mathbf{A}$), so the equation $\mathbf{A}^{-1}\mathbf{A} = \mathbf{1}_n$ shows that $\mathbf{A}$ is left-equivalent to $\mathbf{1}_n$. As $\mathbf{1}_n$ is obviously in RREF, this must be the (unique) RREF of $\mathbf{A}$.

Conversely, if the RREF of $\mathbf{A}$ is $\mathbf{1}_n$, then we know there is an invertible $\mathbf{B}$ such that $\mathbf{BA} = \mathbf{1}_n$. However, since $\mathbf{B}$ is invertible, this implies $\mathbf{A} = \mathbf{B}^{-1}$ (using Exercise 2.4), so we have $\mathbf{AB} = \mathbf{B}^{-1}\mathbf{B} = \mathbf{1}_n$ as well. Thus $\mathbf{A}$ is invertible. □

Note that this also shows that when $\mathbf{A}$ is invertible, the inverse of $\mathbf{A}$ is the same as the matrix which puts it into RREF. So if we apply Gauss–Jordan elimination to the augmented matrix $(\mathbf{A} \mid \mathbf{1}_n)$, then the rightmost n columns of the echelonized matrix will be the inverse of $\mathbf{A}$.

Example 3.19 (Inverse of a matrix) 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

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

The first two columns are now the identity, so $\mathbf{A}$ is invertible; and the last two columns give us the inverse, so

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

Exercises

Exercise 3.1 Check that left equivalence of matrices is an equivalence relation.

Exercise 3.2 Consider the system of equations from Example 3.16, and suppose $b_3 = b_2 - 2b_1$, so the system is consistent.
Find formulae (in terms of b_1 and b_2) for vectors $\vec{c}$ and $\vec{d}$ such that the general solution of the above system of equations is given by

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \{\vec{c} + \lambda\vec{d} \mid \lambda \in \mathbb{R}\}.$$

Exercise 3.3 Let $\mathbf{A} \in M_{n,n}(\mathbb{K})$ be an upper-triangular matrix (so its entries satisfy $A_{ij} = 0$ if $i > j$).

(i) Show that $\mathbf{A}$ is invertible if and only if all the diagonal entries A_{ii} are non-zero.

(ii) Show that if the condition of part (a) is satisfied, then $\mathbf{A}^{-1}$ is also upper-triangular.

Exercise 3.4 Justify the claim made in the proof of uniqueness of RREF that if $\mathbf{A} = \begin{pmatrix} \mathbf{1}_h & \mathbf{S} \\ \mathbf{0}_{m-h,h} & \mathbf{U} \end{pmatrix}$ is invertible, then $\mathbf{U}$ is itself invertible.

Exercise 3.5 Show that if $\mathbf{A}$ is an invertible square matrix, then there is a finite sequence of elementary matrices $\mathbf{B}_1\mathbf{B}_2 \ldots \mathbf{B}_k$ whose product is $\mathbf{A}$. (*Hint*: what is the RREF of $\mathbf{A}^{-1}$?)

CHAPTER 4

Vector spaces

Contents

4.1	Abstract vector spaces	36
	Examples of vector spaces	38
4.2	Linear combinations	40
4.3	Vector subspaces	41
	Operations on subspaces	42
4.4	Subspaces generated by sets	43
4.5	Generating sets and finite-dimensionality	44
4.6	Linear independence	45
Exercises		46

4.1 Abstract vector spaces

Let $\mathbb{K}$ be any field, and $n \geqslant 1$. We've seen that the space $\mathbb{K}^n$ of column vectors has two fundamental operations,

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

It turns out that there are lots of other mathematical structures where these two operations also make sense.

Example 4.1 Consider $\mathrm{P}(\mathbb{R})$, the set of polynomial functions in one real variable, which we denote by x, with real coefficients. That is, an element $p \in \mathrm{P}(\mathbb{R})$ 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 $k = 0, 1, \dots, n$. The largest $m \in \mathbb{N}$ 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 Calculus 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 $s \cdot_{\mathrm{P}(\mathbb{R})} p$

$$s \cdot_{\mathrm{P}(\mathbb{R})} p : \mathbb{R} \to \mathbb{R}, \qquad x \mapsto s \cdot p(x) \tag{4.1}$$

so that $(s \cdot_{\mathrm{P}(\mathbb{R})} p)(x) = \sum_{k=0}^{n} s a_k x^k$ 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 $k = 0, 1, \dots, n$, the sum of the polynomials p and q is the polynomial

$$p +_{\mathrm{P}(\mathbb{R})} q : \mathbb{R} \to \mathbb{R}, \qquad x \mapsto p(x) + q(x) \tag{4.2}$$

so that $(p +_{\mathrm{P}(\mathbb{R})} q)(x) = \sum_{k=0}(a_k + b_k)x^k$ 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 $+_{\mathrm{P}(\mathbb{R})}$ and $\cdot_{\mathrm{P}(\mathbb{R})}$ and simply write $+$ and $\cdot$.

This suggests that just working with the explicit spaces of column vectors $\mathbb{K}^n$ is too limiting. Instead, we're going to make a list of rules which capture how addition and scalar multiplication behave on $\mathbb{K}^n$; and we'll allow ourselves to work with *any* structure which has "addition" and "scalar multiplication" operations that play by these rules (much as we did in Chapter 1 with the axiomatic definition of a field). We'll think of the elements of these structures as "abstract vectors".

Definition 4.2 (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

$$0_V +_V v = v +_V 0_V = v \quad (\text{for all } v \in V); \tag{4.3}$$

- Identity element of scalar multiplication

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

- Scalar multiplication by zero

$$0 \cdot_V v = 0_V \quad (\text{for all } v \in V); \tag{4.4}$$

- 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*.

Examples of vector spaces

Example 4.3 (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 4.4 (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 2.16 that V equipped with addition $+_V : V \times V \to V$ defined by (2.2) 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 4.5 (Vector space of polynomials) The set $\mathrm{P}(\mathbb{R})$ 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 (4.1) and (4.2) and when the zero vector $0_{\mathrm{P}(\mathbb{R})}$ 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 4.6 (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 $V = \mathrm{F}(I, \mathbb{K})$, 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 (4.2) and scalar multiplication $\cdot_V : \mathbb{K} \times V \to V$ as in (4.1). It now is a consequence of the properties of addition and multiplication of scalars that $\mathrm{F}(I, \mathbb{K})$ is a $\mathbb{K}$-vector space. (The reader is invited to check this assertion!)

Example 4.7 (Vector space of sequences) A mapping $x : \mathbb{N} \to \mathbb{K}$, from the natural numbers into a field $\mathbb{K}$, is 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 $(x_n)_{n\in\mathbb{N}} = (x_0, x_1, x_2, \ldots)$. 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

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

or the sequence

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

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 4.8 (Zero vector space) Consider a set V whose only element is a formal symbol $\star$. We define $0_V = \star$, addition by $\star +_V \star = \star$ and scalar multiplication by $s \cdot_V \star = \star$. Then all the properties of Definition 4.2 are satisfied. We write $V = \{\star\}$, or simply $V = \{0\}$, and call V the zero vector space (over $\mathbb{K}$).

Example 4.9 (Field embeddings) If $\mathbb{F}$ and $\mathbb{K}$ are fields, and $\iota : \mathbb{F} \to \mathbb{K}$ is a field embedding, then $\mathbb{K}$ is an $\mathbb{F}$-vector space in a natural way: the addition is the native field addition of $\mathbb{K}$, and the scalar multiplication being given by $s \cdot x = \iota(s) \cdot_{\mathbb{K}} x$. (Exercise: check that the axioms are satisfied!) In effect, we are throwing away some of the structure from $\mathbb{K}$ – we are "forgetting" how to multiply elements of $\mathbb{K}$, except when one of them is in the image of ι.
In particular, $\mathbb{R}$ **is a** $\mathbb{Q}$**-vector space.** This is a really strange and puzzling concept, and shows that sometimes we have to live with our definitions having unexpected consequences!

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 4.10 (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 4.2 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 4.11 (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 4.2 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 (4.4).

$\Rightarrow$ 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

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

4.2 Linear combinations

Definition 4.12 (Linear combination) Let V be a $\mathbb{K}$-vector space and $\mathcal{S}$ a set of vectors from V. A *linear combination* of the vectors in $\mathcal{S}$ is a vector $w \in V$ which can be written in the form

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

for some $k \in \mathbb{N}$, scalars $s_1, \ldots, s_k \in \mathbb{K}$, and vectors $v_1, \ldots, v_k \in \mathcal{S}$.

Note that we don't have to use *all* of the elements in $\mathcal{S}$; indeed $\mathcal{S}$ doesn't have to be finite, and a linear combination is only allowed to mention finitely many elements (the defininition of a vector space doesn't give us any way of making sense of infinite sums). On the other hand, if $\mathcal{S}$ is finite, then it suffices to consider linear combinations which *do* involve all the elements of $\mathcal{S}$, simply by introducing extra terms with $s_i = 0$.

Remark 4.13 When $k = 0$, we understand the empty sum—the sum of no elements of V—to mean 0_V. In particular, 0_V is a linear combination of vectors in $\mathcal{S}$ for any $\mathcal{S}$ (even if $\mathcal{S}$ is the empty set).

Example 4.14 For $n \in \mathbb{N}$ with $n \geqslant 2$ consider $V = \mathrm{P}_n(\mathbb{R})$ and the polynomials $p_1, p_2, p_3 \in \mathrm{P}_n(\mathbb{R})$ 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 $a, b, c \in \mathbb{R}$.

For vectors in the column-vector space $\mathbb{K}^n$, "being a linear combination" is expressible by a linear system of equations:

Example 4.15 *"Is* $\begin{pmatrix}1\\2\\1\end{pmatrix} \in \mathbb{R}^3$ *a linear combination of* $\left\{\begin{pmatrix}3\\-1\\0\end{pmatrix}, \begin{pmatrix}0\\1\\-1\end{pmatrix}\right\}$*?"*

We're asking if there are s_1, s_2 such that $s_1\begin{pmatrix}3\\-1\\0\end{pmatrix} + s_2\begin{pmatrix}0\\1\\-1\end{pmatrix} = \begin{pmatrix}1\\2\\1\end{pmatrix}$, which is the same as the equations

$$\begin{aligned} 3s_1 + 0s_2 &= 1 \\ -s_1 + s_2 &= 2 \\ 0s_1 - s_2 &= 1 \end{aligned}, \qquad \text{i.e.} \qquad \left(\begin{array}{cc|c} 3 & 0 & 1 \\ -1 & 1 & 2 \\ 0 & -1 & 1 \end{array}\right).$$

Since the RREF of this matrix is $\left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right)$, there are no solutions. So the answer to the original question is **no**.

4.3 Vector subspaces

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

Definition 4.16 (Vector subspace) Let V be a $\mathbb{K}$-vector space. A subset $U \subset V$ is called a *vector subspace* of V if the restriction to U of the addition and scalar-multiplication operations of V make U into a vector space; that is,

- $0_V \in U$,
- $v_1 +_V v_2 \in U$ for all $v_1, v_2 \in U$,
- $s \cdot_V v \in U$ for all $s \in \mathbb{K}$ and $v \in U$.

One can check that these conditions are equivalent to the following easier-to-check condition:

- U is non-empty, and

$$s_1 \cdot_V v_1 +_V s_2 \cdot_V v_2 \in U \quad \text{for all } s_1, s_2 \in \mathbb{K} \text{ and all } v_1, v_2 \in U. \tag{4.5}$$

Remark 4.17

(i) Let's check that this simpler condition implies the ones in Definition 4.16. Since U is non-empty, it contains an element, say u. Taking $s_1 = s_2 = 0$ and $v_1 = v_2 = u$ in (4.5), we see that $0 \cdot_V u + 0 \cdot_V u = 0_V \in U$. Thus the zero vector 0_V lies in U. Taking $s_1 = s_2 = 1$ and v_1, v_2 arbitrary, we get that U is closed under sums; and taking $s_2 = 0_{\mathbb{K}}$ we see that U is closed under scalar multiplication.

(On the other hand, we can't drop the condition that U be non-empty, since the empty set vacuously satisfies (4.5) but is not a subspace.)

(ii) We'll see in the exercises that a subspace is automatically *closed under linear combinations*; that is, if U is a vector subspace, then any linear combination of elements of U is in U.

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

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

Example 4.18 (Lines through the origin) Let $\vec{w} \neq 0_{\mathbb{R}^3}$, then the line

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

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 4.19 (Zero subspace) Let V be a $\mathbb{K}$-vector space and $U = \{0_V\}$ the set consisting of the zero vector of V. Then, by Definition 4.16 and the properties of Definition 4.2, it follows that U is a vector subspace of V: the zero subspace $\{0_V\}$. If V and W are different vector spaces over $\mathbb{K}$, then $\{0_V\}$ and $\{0_W\}$ may or may not be exactly the same – it depends on what V and W are – but they have the same vector-space structure (they are *isomorphic*, a concept we'll see later).

Example 4.20 (Periodic functions) Taking $I = \mathbb{R}$ and $\mathbb{K} = \mathbb{R}$ in Example 4.6, 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

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

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 $x \in \mathbb{R}$

$$\begin{aligned}(s_1f_1 + s_2f_2)(x + 2\pi) &= s_1f_1(x + 2\pi) + s_2f_2(x + 2\pi) = s_1f_1(x) + s_2f_2(x)\\ &= (s_1f_1 + s_2f_2)(x)\end{aligned}$$

showing that $s_1f_1 + s_2f_2$ is periodic with period 2π. By Definition 4.16, it follows that U is a vector subspace of $\mathrm{F}(\mathbb{R}, \mathbb{R})$.

Operations on subspaces

Vector subspaces are stable under intersection in the following sense:

Proposition 4.21 *Let V be a $\mathbb{K}$-vector space, $n \geqslant 1$ a natural number and $U_1, \dots, U_n$ vector subspaces of V. Then the intersection*

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

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 4.16, it follows that U' is a vector subspace of V. □

Remark 4.22 The last proposition is also true for $n = 0$, if we understand "the intersection of no subspaces of V" to mean the whole of V. Infinite intersections work too: if I is any set (can be finite, can be infinite, can be empty) and we have

a mapping $I \to \{\text{vector subspaces of } V\}$, $i \mapsto V_i$, then $\bigcap_{i\in I} V_i = \{v \in V : v \in V_i \, \forall i \in I\}$ is a well-defined subset of V and it is a subspace.

Remark 4.23 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 $\vec{e}_1 + \vec{e}_2 \notin U_1 \cup U_2$.

4.4 Subspaces generated by sets

Definition 4.24 (Subspace generated by a set) Let V be a $\mathbb{K}$-vector space and $\mathcal{S} \subset V$ be a subset. The *subspace generated by* $\mathcal{S}$, or the *span* of $\mathcal{S}$, is the set $\operatorname{span}(\mathcal{S})$ whose elements are linear combinations of vectors in $\mathcal{S}$. Formally, we have

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

Remark 4.25 The notation $\langle \mathcal{S} \rangle$ for the span of $\mathcal{S}$ is also in use.

Proposition 4.26 *Let V be a $\mathbb{K}$-vector space and $\mathcal{S} \subset V$ be a non-empty subset. Then $\operatorname{span}(\mathcal{S})$ is a vector subspace of V.*

Proof Clearly $\operatorname{span}(\mathcal{S})$ cannot be empty, since it always contains 0_V. So it suffices to show that if $v_1, v_2 \in \operatorname{span}(\mathcal{S})$ and $s_1, s_2 \in \mathbb{K}$, then $s_1 v_1 + s_2 v_2 \in \operatorname{span}(\mathcal{S})$. By assumption, we can write $v_1 = t_1 w_1 + \cdots + t_k w_k$ for some $k \in \mathbb{N}$, $t_1, \dots t_k \in \mathbb{K}$ and $w_1, \dots, w_k \in \mathcal{S}$; and similarly $v_2 = \hat{t}_1 \hat{w}_1 + \cdots + \hat{t}_j \hat{w}_j$ for some j, scalars $\hat{t}_1, \dots, \hat{t}_j$ and $\hat{w}_1, \dots, \hat{w}_j \in \mathcal{S}$.

But then

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

also a linear combination of vectors in $\mathcal{S}$ and the claim follows. □

Remark 4.27 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 $\mathcal{S}$" (but then we have to justify the claim that such a subspace exists). Another alternative is "the intersection of all subspaces of V that contain S" (but then it's not so clear what its elements are).

4.5 Generating sets and finite-dimensionality

Definition 4.28 Let V be a $\mathbb{K}$-vector space. A subset $\mathcal{S} \subset V$ is called a *generating set* (or *spanning set*) of V if $\text{span}(\mathcal{S}) = V$.

Example 4.29 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 $x = x \cdot_V 1_{\mathbb{K}}$.

Definition 4.30 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*.

Remark 4.31 Notice that we're playing a devious notational trick: the definition of "finite-dimensional" is not " the dimension is finite" – we haven't defined the notion of "dimension" yet! It would be logically better to call these "finitely generated vector spaces", but this is not the convention, so we won't do it.

Example 4.32 The standard basis $\mathcal{S} = \{\vec{e}_1, \dots, \vec{e}_n\}$ 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 $\mathcal{S}$. Thus $\mathbb{K}^n$ is finite-dimensional.

Example 4.33 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 4.34 The vector space $\mathrm{P}(\mathbb{R})$ of polynomials is infinite dimensional. In order to see this, consider a finite set of polynomials $\{p_1, \dots, p_n\}$, $n \in \mathbb{N}$ and let d_i denote the degree of the polynomial p_i for $i = 1, \dots, n$. We set $D =$

$\max\{d_1, \dots, d_n\}$. Since a linear combination of the polynomials $\{p_1, \dots, p_n\}$ has degree at most D, any polynomial q whose degree is strictly larger than D will satisfy $q \notin \operatorname{span}\{p_1, \dots, p_n\}$. It follows that $\mathrm{P}(\mathbb{R})$ cannot be generated by a finite set of polynomials.

4.6 Linear independence

Recall that *spanning* was about *existence* of linear combinations: which elements of V we can "hit" with linear combinations of $\mathcal{S}$. We have a complementary notion which is about *uniqueness* of linear combinations – "how many" ways we can hit a given element:

Definition 4.35 (Linear independence) Let $\mathcal{S} \subset V$ be a subset. We say $\mathcal{S}$ is *linearly independent* if there is no non-trivial way of writing 0_V as a linear combination of vectors in $\mathcal{S}$. That is, for all natural numbers $k \geqslant 1$, all $s_1, \dots, s_k \in \mathbb{K}$ and all *distinct* $v_1, \dots, v_k \in \mathcal{S}$, we have the implication

$$s_1 v_1 + \cdots + s_k v_k = 0_V \qquad \Longleftrightarrow \qquad s_1 = \cdots = s_k = 0.$$

If $\mathcal{S}$ is not linearly independent (so there exists a non-trivial linear combination of elements of $\mathcal{S}$ equal to 0) then $\mathcal{S}$ is called *linearly dependent*.

Remark 4.36 Observe that a subset $\mathcal{T}$ of a linearly independent set $\mathcal{S}$ is itself linearly independent (exercise); and the empty set is linearly independent.

Example 4.37 We consider the polynomials $p_1, p_2, p_3 \in \mathrm{P}(\mathbb{R})$ 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

$$s_1 p_1 + s_2 p_2 + s_3 p_3 = 0_{\mathrm{P}(\mathbb{R})} = o \tag{4.6}$$

where $o : \mathbb{R} \to \mathbb{R}$ denotes the zero polynomial. Since (4.6) 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.

For vectors in $\mathbb{K}^n$, linear independence can be checked using row-echelon form.

Example 4.38 Consider the vectors in $\mathbb{R}^3$ given by

$$v_1 = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, v_2 = \begin{pmatrix} -9 \\ 4 \\ 6 \end{pmatrix}, v_3 = \begin{pmatrix} 3 \\ 0 \\ -1 \end{pmatrix}, v_4 = \begin{pmatrix} 4 \\ -1 \\ 5 \end{pmatrix}.$$

These are linearly independent if and only if the system of equations

$$\begin{pmatrix} 0 & -9 & 3 & 4 \\ 1 & 4 & 0 & -1 \\ 2 & 6 & -1 & 5 \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

has only the zero solution. We already calculated the RREF of this matrix in a previous example; so we know that this system is equivalent to

$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & \frac{17}{3} & 0 \\ 0 & 1 & 0 & -\frac{5}{3} & 0 \\ 0 & 0 & 1 & -\frac{11}{3} & 0 \end{array}\right).$$

Obviously this system is consistent (because the rightmost column is entirely zeroes), but we care about *uniqueness* of solutions; and since none of the rows has its leading entry in the fourth column, s_4 is a free variable, and hence the solution is not unique.

Remark 4.39 Of course, this was bound to happen, since there are four columns to the left of the dividing line, but only three rows. So there can't possibly be a leading entry in every column. What this proves is: **any set of vectors in $\mathbb{K}^n$ containing more than n elements must be linearly dependent**. This is an instance of the *Fundamental Inequality*, one of the main theorems of this module, which we'll prove in the next section.

Exercises

Exercise 4.1 Use the method of Example 3.16 to show that a vector $\vec{b} \in \mathbb{R}^3$ is a linear combination of $\left\{ \begin{pmatrix} 3 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \right\}$ if and only if $\begin{pmatrix} 1 & 3 & 3 \end{pmatrix} \cdot \vec{b} = 0$.

(*) Can you relate this to the RREF of the matrix

$$\left(\begin{array}{cc|ccc} 3 & 0 & 1 & 0 & 0 \\ -1 & 1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 1 \end{array}\right) \quad ?$$

Exercise 4.2 Does there exist a field $\mathbb{F}$ and a vector space V over $\mathbb{F}$ which *can* be written as the union of two proper subspaces? What about three proper subspaces?

Exercise 4.3 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 4.4 (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

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

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

Exercise 4.5 (Polynomials) Let $n \in \mathbb{N}$ and $\mathrm{P}_n(\mathbb{R})$ denote the subset of $\mathrm{P}(\mathbb{R})$ consisting of polynomials of degree *at most* n. Show that $\mathrm{P}_n(\mathbb{R})$ is a subspace of $\mathrm{P}(\mathbb{R})$ for all $n \in \mathbb{N}$.

Exercise 4.6 Show that if $\mathcal{S} \subset \mathcal{T} \subset V$, and $\mathcal{S}$ is a generating set of V, then $\mathcal{T}$ is also a generating set.

Exercise 4.7 Show that for a non-empty subset $\mathcal{S}$ of a $\mathbb{K}$-vector space V, the set $\mathrm{span}(\mathcal{S})$ as defined in Definition 4.24 is the same as either of the two definitions given in Remark 4.27.

Exercise 4.8 Show that a subset $\{v\}$ consisting of a single vector $v \in V$ is linearly independent if and only if $v \neq 0_V$.

CHAPTER 5

Dimensions of vector spaces

Contents

5.1	**Growing and shrinking sets**	**48**
5.2	**The fundamental inequality**	**49**
5.3	**Bases of vector spaces**	**50**
	Definition	50
	The Main Theorem on Bases	51
	Consequences of the Main Theorem	52
5.4	**Dimensions of vector spaces**	**53**
5.5	**Computing with subspaces**	**54**
Exercises		**55**

In this chapter, we'll prove two of the key theorems of this module – the *Fundamental Inequality* and the *Main Theorem on Bases*. This chapter is quite abstract (and quite difficult), but we'll get back to more concrete computations later!

5.1 Growing and shrinking sets

We've seen that if $\mathcal{S}$ is a generating set in V, and we modify $\mathcal{S}$ by putting some more elements into it, then it's still a generating set. On the other hand, if we take some elements out, it might not be generating any more. Linear independence, on the other hand, has the opposite behaviour: if $\mathcal{S}$ is LI, and we take some elements out, then the modified set is still LI; but if we put some more elements in, it might not be LI any more.

The next two lemmas give criteria for when we *can* shrink a generating set without breaking the generating property, or grow an LI set without breaking the LI property. These will be crucial in the next section.

Lemma 5.1 (Shrinking generating sets) *Let V be a $\mathbb{K}$-vector space and $\mathcal{S} \subset V$ a generating set. If $v_0 \in \mathcal{S}$ satisfies $v_0 \in \mathrm{span}(\mathcal{S} \setminus \{v_0\})$, then $\mathcal{S} \setminus \{v_0\}$ is a generating set.*

Proof Since $v_0 \in \mathrm{span}(\mathcal{S} \setminus \{v_0\})$, there exist vectors $v_1, \dots, v_n \in \mathcal{S}$ with $v_i \neq v_0$ and scalars $s_1, \dots, s_n$ so that $v_0 = s_1 v_1 + \cdots + s_n v_n$.

Suppose we are given $v \in V$. Since $\mathcal{S}$ is a generating set, there exist vectors $w_1, \dots, w_k \in \mathcal{S}$ and scalars $t_1, \dots, t_k$ so that $v = t_1 w_1 + \cdots + t_k w_k$. If $\{w_1, \dots, w_k\}$ does not contain v_0, then $v \in \mathrm{span}(\mathcal{S} \setminus \{v_0\})$, so assume that $v_0 \in \{w_1, \dots, w_k\}$. After possibly relabelling the elements of $\{w_1, \dots, w_k\}$ 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 $v \in \mathrm{span}(\mathcal{S} \setminus \{v_0\})$, as claimed. □

Lemma 5.2 (Growing LI sets) *Let V be a $\mathbb{K}$-vector space, $\mathcal{S} \subset V$ linearly independent and $v_0 \in V$. Suppose $v_0 \notin \text{span}(\mathcal{S})$, 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 $v_1, \dots, v_n$ of $\mathcal{S}$ so that $\mathcal{T} = \{v_0, v_1, \dots, v_n\}$. Suppose $s_0 v_0 + s_1 v_1 + \cdots + s_n v_n = 0_V$ for some scalars $s_0, s_1, \dots, s_n \in \mathbb{K}$. 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 $v_0 \notin \text{span}(\mathcal{S})$. 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. □

5.2 The fundamental inequality

We'll now prove the following important lemma, which is going to be the key to understanding how finite-dimensional vector spaces "work":

Lemma 5.3 (Fundamental inequality) *Let V be a vector space, and suppose V has a finite generating set W. Then any linearly independent set U in V is finite and satisfies $|U| \leqslant |W|$.*

The proof of this lemma (everything from here to the end of Section 5.2) is **non-examinable**; but you should *definitely* make sure you are aware of the statement!

We'll deduce the Fundamental Inequality from the following stronger statement:

Lemma 5.4 (The Steinitz Exchange Lemma) *Suppose V is a vector space and U, W subsets of V such that*

- *W is finite,*
- *U is linearly independent,*
- *W spans V.*

Then $|U| \leqslant |W|$ (so U is also finite); and there is a subset $W' \subseteq W$, with $|W'| = |W| - |U|$ and $W' \cap U = \varnothing$, such that $U \cup W'$ also spans V.

In other words, we can exchange some of the elements of W for the elements of U, without breaking the generating property.

Proof Let us first prove the result assuming U is finite. Let $m = |U|$. We will argue by induction on m. If $m = 0$ then the statement is trivial; so we can assume the statement holds for $m - 1$. Thus we can suppose that $U = \{u_1, \dots, u_m\}$, with $m - 1 \leqslant n$, and $W = \{u_1, \dots, u_{m-1}, w_m, \dots, w_n\}$ for some vectors $w_m, \dots, w_n$. (We haven't yet excluded the possibility that $n = m - 1$, in which case this just means that $W = \{u_1, \dots, u_{m-1}\}$.)

We first deal with a silly special case: if $u_m \in \{w_m, \dots, w_n\}$, then we can assume $u_m = w_m$ by relabeling; then $W' = \{w_{m+1}, \dots, w_n\}$ works, since $U \cup W'$ is equal to W and we already know that W spans V. So we can suppose that u_m is not one of the w_i.

Since W spans V, and $u_m \in V$, we know that u_m can be written as a linear combination of W:

$$u_m = \sum_{i=1}^{m-1} s_i u_i + \sum_{i=m}^{n} t_i w_i, \qquad s_i, t_i \in \mathbb{K}.$$

If all the t_i's are zero, then this would contradict the linear independence of U; so there must be some i with $m \leqslant i \leqslant n$ such that $t_i \neq 0$. (This shows in particular that n must be at least m.) Reordering the w_i if necessary, we can suppose $t_m \neq 0$. Our goal will be to show that $\{u_1, \dots, u_m, w_{m+1}, \dots, w_n\}$ spans V.

We rearrange the equality above into

$$w_m = \frac{1}{t_m}\left(u_m - \sum_{i=1}^{m-1} s_i u_i - \sum_{j=m+1}^{n} t_j w_j\right).$$

Since we're not in the 'silly special case', the sum on the right doesn't involve w_m. So we've shown that w_m is in the span of $T \setminus \{w_m\}$, where $T = \{u_1, \dots, u_m, w_m, \dots, w_n\}$. But T spans V, since it contains W. So we can apply the "shrinking generating sets" lemma to conclude that $T \setminus \{w_m\}$ spans V.

This completes the proof for finite U. If U is infinite, then we can find a subset $U' \subset U$ of size $n+1$, where $n = |W|$; but U' is linearly independent (because it's contained in U) and hence applying the lemma to U' and W gives a contradiction. So the lemma holds for all U. □

Remark 5.5 Notice that the Fundamental Inequality is just one part of the Steinitz Exchange Lemma, but our proof of the Fundamental Inequality for $|U| = m$ depends on knowing the whole of the Exchange Lemma for $|U| = m-1$. It is possible to give a self-contained proof of the Fundamental Inequality (without proving the rest of Steinitz at the same time), but it needs a different method, and it's quite fiddly. This is an example of a curious paradox: sometimes a stronger (but more specific) theorem can be easier to prove than a weaker one!

5.3 Bases of vector spaces

Definition

Definition 5.6 (Basis) Let V be a vector space. A subset $\mathcal{S} \subset V$ which is a generating set of V and also linearly independent is called a *basis* of V.

Equivalently, a set $\mathcal{S}$ is a basis if every vector in V can be written *uniquely* as a linear combination of elements of $\mathcal{S}$: the "generating" condition gives existence of the linear combination, and the "linearly independent" condition gives uniqueness.

Example 5.7 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 4.29 implies that $\{1_{\mathbb{K}}\}$ is a basis of $\mathbb{K}$.

Example 5.8 Clearly, the standard basis $\{\vec{e}_1, \dots, \vec{e}_n\}$ of $\mathbb{K}^n$ is linearly independent since

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

It follows together with Example 4.32 that the standard basis of $\mathbb{K}^n$ is indeed a basis in the sense of Definition 5.6.

Example 5.9 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 4.33 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* $M_{m,n}(\mathbb{K})$.

The Main Theorem on Bases

We'll now come to one of the central theorems of this course

Theorem 5.10 (Main Theorem on Bases) *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.

We will only prove Theorem 5.10 for finite dimensional vector spaces. The proof will use the Fundamental Inequality.

Proof of Theorem 5.10 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) (Slogan: *A maximal LI set is a basis*) Let $\mathcal{S} \subset V$ be a subset generating V (we don't assume that $\mathcal{S}$ is finite). We consider the set $\mathcal{X} \subset \mathbb{N}$ 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 the Fundamental Inequality.

Let $m \in \mathcal{X}$ be the largest integer and $\mathcal{T} \subset \mathcal{S}$ an LI 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 $v_0 \notin \text{span}(\mathcal{T})$, since if no such element exists, we have $\mathcal{S} \subset \text{span}(\mathcal{T})$ and hence $V = \text{span}(\mathcal{S}) \subset \text{span}(\mathcal{T})$ contradicting the assumption that $\mathcal{T}$ is not a basis of V. Applying Lemma 5.2 (the growing-LI-sets lemma), 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) (Slogan: *A minimal generating set is a basis*) Let $\mathcal{S} \subset V$ be a subset that is linearly independent in V. Note that $\mathcal{S}$ is finite, by the Fundamental Inequality. 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 $\mathcal{T} = \{v_1, \dots, v_m\}$ for distinct elements of V. Suppose $\mathcal{S} = \{v_1, \dots, v_k\}$ for some $k \leqslant m$. This holds true since $\mathcal{S} \subset \mathcal{T}$. Since $\mathcal{T}$ is linearly dependent we have scalars $s_1, \dots, s_m$ 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

$$v_{k+1} = -\frac{1}{s_{k+1}} \left(s_1 v_1 + \cdots + s_k v_k + s_{k+2} v_{k+2} + \cdots + s_m v_m \right). \tag{5.1}$$

Let $\hat{\mathcal{T}} = \{v_1, \dots, v_k, v_{k+2}, \dots, v_m\}$. Then $\mathcal{S} \subset \hat{\mathcal{T}}$ and (5.1) shows that $v_{k+1} \in \text{span}(\hat{\mathcal{T}})$. Lemma 5.1 (the shrinking-generating-sets lemma) 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, the Fundamental Inequality 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). □

Consequences of the Main Theorem

Corollary 5.11 *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 5.10 to $\mathcal{S} = V$ to obtain a basis of V. □

Remark 5.12 We've seen that $\mathbb{R}$ is a $\mathbb{Q}$-vector space. It is *impossible* to explicitly write down a basis of this vector space, even though the corollary says that they exist!

5.4 Dimensions of vector spaces

We can now, at last, make sense of the idea of *dimension* of a vector space: we know that all vector spaces have bases, and any two bases are the same size; so we can make the following definition:

Definition 5.13 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 5.14

(i) The zero vector space $\{0\}$ has the empty set as a basis and hence is 0-dimensional. Conversely, if V is a zero-dimensional space, then the empty set is a basis of V, so we must have $V = \{0\}$.
(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 $\{\vec{e}_1, \dots, \vec{e}_n\}$ 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.

Lemma 5.15 *Let V be a finite-dimensional $\mathbb{K}$-vector space, and U a subspace of V. Then U is also finite-dimensional and we have*

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

Furthermore, $\dim(U) = 0$ iff $U = \{0_V\}$, and $\dim(U) = \dim(V)$ iff $U = V$.

Proof Let V be a finite dimensional $\mathbb{K}$-vector space and $U \subset V$ a subspace. Let's consider the set $\mathcal{X} = \{m \in \mathbb{N} : \text{there exists a LI set in } U \text{ with } m \text{ elements}\}$. Obviously $0 \in \mathcal{X}$, since $\varnothing$ is LI. On the other hand, $\mathcal{X}$ can't contain any integer larger than the dimension of V, since an LI set in U is *a fortiori* an LI subset of V.

So $\mathcal{X}$ must have a largest element, say d. Let us choose an LI subset $\mathcal{S} \subset U$ whose size is d. We will show that $\mathcal{S}$ generates U; this shows that U is finite-dimensional, since $\mathcal{S}$ is a finite generating set.

Let $u \in U$ be arbitrary. If $u \notin \operatorname{span}(\mathcal{S})$, then we can apply the growing-LI-sets lemma to see that $\mathcal{S} \cup \{u\}$ is an LI set. Since this set has size $d + 1$, and d is maximal, this is a contradiction. Thus $u \in \operatorname{span}(\mathcal{S})$ as required.

Since this set $\mathcal{S}$ is by definition LI, it is a basis of U, so $d = \dim U$. Using the Fundamental Inequality on $\mathcal{S}$ and a basis of V, we find that $\dim(U) \leqslant \dim(V)$; and if $\dim(U) = \dim(V)$, then the Exchange Lemma tells us that $\mathcal{S}$ generates V, so $U = \operatorname{span}(\mathcal{S}) = V$. On the other hand, if $d = 0$, then U is the empty set, so its span is $\{0_V\}$. □

Remark 5.16 This proof can be done much more quickly if we allow ourselves to apply the Main Theorem on Bases to conclude that U has a basis: if $\mathcal{B}$ is any basis of U, then $\mathcal{B}$ is in particular an LI set in V and hence it has size $\leqslant \dim(V)$ by the

Fundamental Inequality. The problem is that we have only proved the Main Theorem on Bases for finite-dimensional spaces, and we don't know *a priori* that U is finite-dimensional. So we have to work a bit harder.

5.5 Computing with subspaces

The above theory applies to subspaces of any abstract vector space; we'll now specialise to the concrete case of row and column vectors, and see how to compute with bases of subspaces in practice. As usual, this will reduce to computing a RREF (the "Swiss army knife" of linear algebra calculations).

We'll first investigate the case of subspaces of $\mathbb{K}_n$ (recall that the index n "downstairs" means row vectors, not column vectors).

Proposition 5.17 *Let $\vec{\alpha}_1, \dots, \vec{\alpha}_r$ be vectors in $\mathbb{K}_n$, and let $U = \operatorname{span}(\vec{\alpha}_1, \dots, \vec{\alpha}_r)$. Let $\mathbf{A}$ be the matrix with the $\vec{\alpha}_i$ as rows, and let $\mathbf{B}$ be the RREF of $\mathbf{A}$, with rows $\vec{\beta}_1, \dots, \vec{\beta}_r$. If h is the number of nonzero rows of $\mathbf{B}$, then $h = \dim U$, the vectors $\vec{\beta}_1, \dots, \vec{\beta}_h$ are a basis of U, and $\vec{\beta}_{h+1} = \cdots = \vec{\beta}_r = 0$.*

Proof Since $\mathbf{B}$ is left-equivalent to $\mathbf{A}$, every row of $\mathbf{B}$ is a linear combination of rows of $\mathbf{A}$, and vice versa. Hence the span of the β's is equal to the span of the α's. By the definition of RREF, all the non-zero rows of $\mathbf{B}$ come before all the zero rows. So it suffices to show that the non-zero rows of $\mathbf{B}$ are linearly independent.

Let $s_1, \dots, s_h$ be scalars such that $\sum s_i \vec{\beta}_i = 0$. For $1 \leqslant j \leqslant h$, let p_j be the index of the leading entry of $\vec{\beta}_j$. Then $[\vec{\beta}_j]_{p_i} = 1$, and $[\vec{\beta}_i]_{p_j} = 0$ for $i \neq j$ by the definition of row echelon form. Hence we have

$$0 = [\mathbf{0}]_{p_j} = \sum_i s_i [\vec{\beta}_i]_{p_j} = \sum_i s_i \delta_{ij} = s_j,$$

so $s_j = 0$. Since j was arbitrary, this shows that $s_1 = \cdots = s_h = 0$ and hence $\{\vec{\beta}_1, \dots, \vec{\beta}_h\}$ is an LI set. □

Of course, a vector space can have many different bases, so sometimes it can be hard to recognise when two subspaces are actually the same. The uniqueness part of RREF allows us to solve this too:

Proposition 5.18 *Let U be a subspace of $\mathbb{K}_n$. Then there is a* unique *basis $\{\vec{\beta}_1, \dots, \vec{\beta}_h\}$ of U (and a unique ordering of those basis vectors) such that the matrix with those vectors as rows is in RREF; and this basis, the* ***RREF basis of*** *U, can be computed explicity starting from any generating set of U.*

Proof We've seen that an echelon-form basis exists (and is computable), so we need to show uniqueness. But any two bases of U give matrices which are left-equivalent; hence there cannot be more than one such matrix which is in RREF. □

Thus, if we are given generating sets for two subspaces U, U' and we want to know if $U = U'$, we just compute the RREF bases of each, and check whether they're the same.

Remark 5.19 As a special case, we can check if a given vector $\vec{\xi}$ lies in U or not, since $u \in U$ iff the subspace $U' = \text{span}(\{\text{generators of } U\} \cap \{\vec{\xi}\})$ is equal to U. Concretely, we just form the matrix $\left(\begin{array}{c}\mathbf{B}\\ \hline \vec{\xi}\end{array}\right)$, where $\mathbf{B}$ is the echelon basis matrix of U, and echelonize that; if the echelon form is $\left(\begin{array}{c}\mathbf{B}\\ \hline 0\end{array}\right)$, then $\vec{u} \in U$, and otherwise not.

Of course, if we want to compute with subspaces of $\mathbb{K}^n$ instead, we can just transpose all the matrices[1] and formulate our problem in terms of $\mathbb{K}_n$.

Example 5.20 Let's revisit example Example 4.15: *"Is* $\begin{pmatrix}1\\2\\1\end{pmatrix} \in \mathbb{R}^3$ *a linear combination of* $\left\{\begin{pmatrix}3\\-1\\0\end{pmatrix}, \begin{pmatrix}0\\1\\-1\end{pmatrix}\right\}$*?"*

This is equivalent to asking: is $\vec{\gamma} = (1 \quad 2 \quad 1) \in \mathbb{R}_3$ in the subspace generated by $\vec{\alpha} = (3 \quad -1 \quad 0)$ and $\vec{\beta} = (0 \quad 1 \quad -1)$?

We compute that the echelon form of the matrix $\left(\begin{array}{ccc}3 & -1 & 0\\ 0 & 1 & -1\\ \hline 1 & 2 & 1\end{array}\right)$ is the 3×3 identity matrix. So $\text{span}(\vec{\alpha}, \vec{\beta}, \vec{\gamma})$ is 3-dimensional – it's the whole of $\mathbb{R}_3$ – whereas $\text{span}(\vec{\alpha}, \vec{\beta})$ must have dimension $\leqslant 2$. Thus $\vec{\gamma}$ is not a linear combination of the other two.

Remark 5.21 Note that this is a genuinely different method from the previous computation: previously we applied elementary row operations to $\left(\begin{smallmatrix}3 & 0 & 1\\ -1 & 1 & 2\\ 0 & -1 & 1\end{smallmatrix}\right)$, and now we're applying elementary row operations to its transpose (or elementary column operations to the original matrix). Both methods are equally valid.

Exercises

Exercise 5.1 (hard!) Show that if V is not finite-dimensional, then there exists an infinite linearly independent set in V.
[In this exercise you may only use theorems *proved* in the course, so you may not use the fact that the Main Theorem on Bases holds for infinite-dimensional spaces.]

[1] This is valid because transposing matrices sends linear combinations to linear combinations, subspaces to subspaces, etc – an example of a *vector space isomorphism*, a concept we'll meet in the next chapter.

Exercise 5.2 Consider the vector space $\mathbb{R}^\infty$ of real sequences, as in Example 4.7. Let e_i be the sequence with $(e_i)_j = 1$ if $i = j$ and 0 otherwise. Is $\{e_0, e_1, e_2, \dots\}$ a generating set of $\mathbb{R}^\infty$? Is it linearly independent?

Exercise 5.3 Find a basis of the field $\mathbb{F}_4$ from Chapter 1 as a vector space over $\mathbb{F}_2$, and show that its dimension is 2.

Exercise 5.4 Let U be the subspace of $\mathbb{R}^4$ generated by the vectors $\{\begin{pmatrix} 0 \\ -1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ -2 \\ 1 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \\ 2 \\ 1 \end{pmatrix}\}$. How many of the standard basis vectors $\{\vec{e}_1, \dots, \vec{e}_4\}$ are in U?

CHAPTER 6

Linear maps

Contents

6.1	**Linear maps**	**57**
	Examples	58
	First properties of linear maps	58
	Isomorphisms	59
6.2	**Images, preimages, kernels**	**59**
	Linear maps and dimension	61
6.3	**The rank-nullity theorem**	**62**
Exercises		**64**

6.1 Linear maps

Throughout this section, V, W denote $\mathbb{K}$-vector spaces. So we have a notion of addition and scalar multiplication; and if we have a mapping $f : V \to W$, we can ask if it respects these structures.

Definition 6.1 (Linear map) A mapping $f : V \to W$ is called *linear* if it satisfies the following two conditions:

- it is *additive*, i.e. for all $v_1, v_2 \in V$ we have
$$f(v_1 + v_2) = f(v_1) + f(v_2).$$
- It is 1-*homogeneous*, that is, for all $s \in \mathbb{K}$ and $v \in V$ we have
$$f(sv) = sf(v).$$

One can check (see Exercises) that a mapping $f : V \to W$ is linear iff it satisfies

$$f(s_1 v_1 + s_2 v_2) = s_1 f(v_1) + s_2 f(v_2) \tag{6.1}$$

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

Example 6.2 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.*

Examples

For instance, you already saw in Algorithmics that *matrices give linear maps*: if $\mathbf{M} \in M_{m,n}(\mathbb{K})$, then for any $\vec{x}, \vec{y} \in \mathbb{K}^n$ and $a, b \in \mathbb{K}$, we have

$$\mathbf{M} \cdot (a\vec{x} + b\vec{y}) = a\mathbf{M}\vec{x} + b\mathbf{M}\vec{y},$$

so the map $f_{\mathbf{M}} : \mathbb{K}^n \to \mathbb{K}^m$ defined by $f_{\mathbf{M}}(\vec{x}) = \mathbf{M}\vec{x}$ is linear. However, working with abstract vector spaces means we have all kinds of other exciting linear maps to study.

Example 6.3 If $\mathrm{P}(\mathbb{R})$ is the vector space of polynomials from Example 4.1, then we may think of the derivative with respect to the variable x as a mapping

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

Now recall that the derivative satisfies

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}x}(p+q) = \frac{\mathrm{d}}{\mathrm{d}x}(p) + \frac{\mathrm{d}}{\mathrm{d}x}(q) && \text{(additivity)}, \\ &\frac{\mathrm{d}}{\mathrm{d}x}(s \cdot p) = s \cdot \frac{\mathrm{d}}{\mathrm{d}x}(p) && \text{(1-homogeneity)}. \end{aligned} \tag{6.2}$$

so it is indeed linear.

Example 6.4 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 6.5 Let V be a $\mathbb{K}$-vector space. Then the identity mapping $\mathrm{Id}_V : V \to V$ is linear, since for all $s_1, s_2 \in \mathbb{K}$ and $v_1, v_2 \in V$ we have

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

First properties of linear maps

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

Lemma 6.6 *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

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

We'll now investigate what happens if you compose two such maps. 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

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

Recall also that a mapping is bijective if and only if it has an *inverse mapping*, which is a map $f^{-1} : \mathcal{Y} \to \mathcal{X}$ with $f^{-1} \circ f = \mathrm{Id}_{\mathcal{X}}$ and $f \circ f^{-1} = \mathrm{Id}_{\mathcal{Y}}$ (and this inverse mapping is unique if it exists).

Proposition 6.7 *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 map $f^{-1} : V_2 \to V_1$ is linear.*

Proof For the first statement, 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.

The second statement is more delicate. Recall that for any $v \in V_2$ we have $f(f^{-1}(v)) = v$. So for $v, w \in V_2$ we can write

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

Using linearity of f, the right-hand side can be rewritten as

$$sf(f^{-1}(v)) + tf(f^{-1}(w)) = f(sf^{-1}(v) + tf^{-1}(w)).$$

Thus

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

and since f is injective, we can cancel the f's to conclude

$$f^{-1}(sv + tw) = sf^{-1}(v) + tf^{-1}(w). \qquad \square$$

Isomorphisms

Bijective linear maps are particularly important and they have a special name:

Definition 6.8 (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*.

Remark 6.9 If two vector spaces are isomorphic, then they are "the same as vector spaces". For instance, the obvious map $\mathbb{K}_n \to \mathbb{K}^n$ given by $(x_1 \ \dots \ x_n) \mapsto \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ is an isomorphism.

6.2 Images, preimages, kernels

In this section we'll investigate how vector subspaces interact with linear maps.

Proposition 6.10 *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 6.6, $f(0_V) = 0_W$, 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 4.16 again, we conclude that $f(U)$ is a subspace of W. The second claim is left to the reader as an exercise. □

We now define some special subspaces associated to a linear map.

Definition 6.11 (Kernel) The *kernel* of a linear map $f : V \to W$ is the preimage of $\{0_W\}$ under f, that is,

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

Example 6.12 The kernel of the linear map $\frac{\mathrm{d}}{\mathrm{d}x} : \mathrm{P}_n(\mathbb{R}) \to \mathrm{P}_{n-1}(\mathbb{R})$ consists of the constant polynomials satisfying $f(x) = c$ for all $x \in \mathbb{R}$ and where $c \in \mathbb{R}$ is some constant.

An immediate consequence of Proposition 6.10 is:

Corollary 6.13 *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.*

Roughly, you can think of the kernel as being the stuff which gets "lost in the machine", and the image as the stuff which "comes out the other side". So these should, in some sense, add up to all of V; and we'll make this precise a little later (in Theorem 6.20).

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

Lemma 6.14 *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 6.6, 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 $v_1 = v_2$. □

We can characterise isomorphisms using kernels and images. 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 6.14 gives:

Proposition 6.15 *A linear map $f : V \to W$ is an isomorphism if and only if* $\mathrm{Ker}(f) = \{0_V\}$ *and* $\mathrm{Im}(f) = W$.

Linear maps and dimension

Lemma 6.16 *Let $f : V \to W$ be linear.*

(i) If $\mathcal{S} \subset V$ is a generating set, and f is surjective, then $f(\mathcal{S})$ is a generating set of W.

(ii) If f is surjective, and V is finite-dimensional, then W is finite-dimensional.

(iii) If $\mathcal{S} \subset V$ is an LI set, and f is injective, then $f(\mathcal{S})$ is an LI set in W.

(iv) If f is injective, and W is finite-dimensional, then V is finite-dimensional.

Proof (i) Let $w \in W$. Since f is surjective there exists $v \in V$ such that $f(v) = w$. Since $\mathrm{span}(\mathcal{S}) = V$, 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 $w \in \mathrm{span}(f(\mathcal{S}))$ and since w is arbitrary, it follows that $W = \mathrm{span}(f(\mathcal{S}))$.

(ii) If V is finite-dimensional then it has a finite generating set $\mathcal{S}$. Then $f(\mathcal{S})$ is a finite set in W, and by (i) it is a generating set.

(iii) Suppose, for contradiction, that $f(\mathcal{S})$ is linearly dependent. Then we can find scalars $s_1, \dots, s_k$ (not all zero) and elements $w_1, \dots, w_s$ in $f(\mathcal{S})$ with $\sum s_i w_i = 0_W$. But each w_i must be $f(v_i)$ for some $v_i \in \mathcal{S}$, and $0_W = f(0_V)$, so we have $f(\sum s_i v_i) = \sum s_i f(v_i) = 0_W = f(0_V)$. Since f is injective, we can conclude that $\sum s_i v_i = 0_V$ and thus $\mathcal{S}$ is itself linearly dependent, contradicting our assumption.

(iv) Suppose V is infinite-dimensional. Then V contains an infinite linearly independent set $\mathcal{S}$ by Exercise 5.1. By (iii), $f(\mathcal{S})$ is a linearly independent set in W; and it is still infinite, since f is injective. So W is infinite-dimensional. □

Lemma 6.17 *Let $f : V \to W$ be a vector space isomorphism. Then V is finite-dimensional if and only if W is; and if this holds, then* $\dim(V) = \dim(W)$.

Proof Parts (ii) and (iv) of Lemma 6.16 show that V is finite-dimensional iff W is. Parts (i) and (iii) show that if $\mathcal{S}$ is a basis of V, then $f(\mathcal{S})$ is a basis of W; but, since f is injective, $f(\mathcal{S})$ has the same number of elements as $\mathcal{S}$. □

Corollary 6.18 *If $m \neq n$, then $\mathbb{K}^m$ and $\mathbb{K}^n$ are not isomorphic as vector spaces.*

Proof We've seen that $\mathbb{K}^n$ has dimension n, so if an isomorphism existed for $m \neq n$ it would contradict the previous lemma. □

6.3 The rank-nullity theorem

Now that we know the notion of "dimension" is well-behaved, we're going to study the dimensions of subspaces coming from a linear map.

Definition 6.19 (Rank and nullity for linear maps and matrices) Let V, W be $\mathbb{K}$-vector spaces with W finite dimensional. The *rank* and *nullity* of a linear map $f : V \to W$ are defined as

$$\operatorname{rank}(f) = \dim \operatorname{Im}(f), \qquad \operatorname{nullity}(f) = \dim \operatorname{Ker}(f).$$

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 6.20 (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 Lemma 5.15. Let $\{v_1, \dots, v_d\}$ be a basis of $\mathcal{S} = \operatorname{Ker}(f)$. By Theorem 5.10 (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

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

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 6.21 *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) $\Rightarrow$ (ii) By Lemma 6.14, the map f is injective if and only if $\mathrm{Ker}(f) = \{0_V\}$ so that $\dim \mathrm{Ker}(f) = 0$ by Example 5.14 (i). Theorem 6.20 implies that $\dim \mathrm{Im}(f) = \dim(V) = \dim(W)$ and hence Lemma 5.15 implies that $\mathrm{Im}(f) = W$, that is, f is surjective.

(ii) $\Rightarrow$ (iii) Since f is surjective $\mathrm{Im}(f) = W$ and hence $\dim \mathrm{Im}(f) = \dim(W) = \dim(V)$. Theorem 6.20 implies that $\dim \mathrm{Ker}(f) = 0$ so that $\mathrm{Ker}(f) = \{0_V\}$ by Lemma 5.15. Applying Lemma 6.14 again shows that f is injective and hence bijective.

(iii) $\Rightarrow$ (i) Since f is bijective, it is also injective. □

Corollary 6.22 *Let V, W be finite dimensional $\mathbb{K}$-vector spaces and $f : V \to W$ a linear map. Then* $\mathrm{rank}(f) \leqslant \min\{\dim(V), \dim(W)\}$ *and*

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

Proof For the first claim it is sufficient to show that $\mathrm{rank}(f) \leqslant \dim(V)$ and $\mathrm{rank}(f) \leqslant \dim(W)$. By definition, $\mathrm{rank}(f) = \dim \mathrm{Im}(f)$ and since $\mathrm{Im}(f) \subset W$, we have $\mathrm{rank}(f) = \dim \mathrm{Im}(f) \leqslant \dim(W)$ with equality if and only if f is surjective, by Lemma 5.15.

Theorem 6.20 implies that $\mathrm{rank}(f) = \dim \mathrm{Im}(f) = \dim(V) - \dim \mathrm{Ker}(f) \leqslant \dim(V)$ with equality if and only if $\dim \mathrm{Ker}(f) = 0$, that is, when f is injective (as we have just seen in the proof of the previous corollary). □

Corollary 6.23 *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 6.20

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

and the claim follows from Corollary 6.22.

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

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

and the claim follows from Corollary 6.22. □

Proposition 6.24 *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 6.17.

$\Leftarrow$ Let $\dim(V) = \dim(W) = n \in \mathbb{N}$. Choose a basis $\mathcal{T} = \{w_1, \dots, w_n\}$ 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 $\{w_1, \dots, w_n\}$ are linearly independent. We thus conclude $\operatorname{Ker} \Theta = \{0_V\}$ and hence Lemma 6.14 implies that Θ is injective and therefore bijective by Corollary 6.21. The map Θ is linear and bijective, thus an isomorphism. Likewise, for a choice of basis $\mathcal{S} = \{v_1, \dots, v_n\}$ 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 6.7 the composition of linear maps is again linear, the map $\Theta \circ \Phi^{-1} : V \to W$ is an isomorphism. □

Exercises

Exercise 6.1 Let $f : V \to W$ be a linear map, $k \geqslant 2$ a natural number and $s_1, \dots, s_k \in \mathbb{K}$ and $v_1, \dots, v_k \in V$. Show that $f : V \to W$ 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 6.2 Show, conversely, that if a mapping $f : V \to W$ satisfies

$$f(s_1 v_1 + s_2 v_2) = s_1 f(v_1) + s_2 f(v_2)$$

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

Exercise 6.3 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 6.4 Let $f : U \to V$ and $g : V \to W$ be linear maps. Show that

$$\operatorname{rank}(g \circ f) \leqslant \min(\operatorname{rank}(f), \operatorname{rank}(g)).$$

Exercise 6.5 Show that the $\mathbb{R}$-vector spaces $\mathrm{P}_n(\mathbb{R})$ and $\mathbb{R}^{n+1}$ are isomorphic for all $n \in \mathbb{N}$.

Exercise 6.6 *(requires concepts from M03 Analysis I)* Let $C^\infty(\mathbb{R})$ be the $\mathbb{R}$-vector space of functions $\mathbb{R} \to \mathbb{R}$ which are infinitely often differentiable. Show that the map D defined by

$$D(f)(x) = f''(x) - 2f'(x) + f(x)$$

is a linear map $C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$. Is D surjective? What is the dimension of $\operatorname{Ker} D$?

CHAPTER 7

Linear maps and matrices

Contents

7.1	**Linear mappings associated to matrices**	**66**
	Composition and inverses	68
7.2	**Computing kernels and images**	**69**
	Image and rank	69
	Kernel and nullity	70
Exercises		**72**

We'll now "come down to earth" a bit, and ask how to actually *calculate* with linear maps on our standard vector space $\mathbb{K}^n$. We'll translate this into statements about matrices; so we need to introduce some new ways to calculate with those.

7.1 Linear mappings associated to matrices

Definition 7.1 (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 $\vec{x} = (x_i)_{1\leqslant i\leqslant n} \in \mathbb{K}^n$:

$$\mathbf{A}\vec{x} = \vec{a}_1 x_1 + \vec{a}_2 x_2 + \cdots + \vec{a}_n x_n = \begin{pmatrix} A_{11}x_1 + A_{12}x_2 + \cdots + A_{1n}x_n \\ A_{21}x_1 + A_{22}x_2 + \cdots + A_{2n}x_n \\ \vdots \\ A_{m1}x_1 + A_{m2}x_2 + \cdots + A_{mn}x_n \end{pmatrix}.$$

It's clear that $f_{\mathbf{A}}$ is a linear map. We'll now show that any linear map $\mathbb{K}^n \to \mathbb{K}^m$ arises this way, from a uniquely determined $\mathbf{A}$.

Proposition 7.2 *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 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 $x \in \mathcal{X}$.

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. □

Remark 7.3 Note that $[f_{\mathbf{A}}(\vec{e}_i)]_j$, the j-th entry of $f_{\mathbf{A}}(\vec{e}_i)$, is equal to $[\mathbf{A}]_{ji}$; equivalently, we have

$$f_{\mathbf{A}}(\vec{e}_i) = \sum_{j=1}^{m} A_{ji}\vec{e}_j.$$

(This looks wrong, but it is really correct as stated.)

Lemma 7.4 *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 We've just seen that for any $\mathbf{B} \in M_{n,m}(\mathbb{K})$, the map $f_{\mathbf{B}}$ is linear.

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

$$f_{\mathbf{B}}(\vec{e}_i) = \mathbf{B}\vec{e}_i = g(\vec{e}_i). \tag{7.1}$$

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

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

where the last equality uses (7.1). Since the vector $\vec{v}$ is arbitrary, it follows that $g = f_{\mathbf{B}}$, as claimed. □

Remark 7.5 Let $\mathrm{Lin}(\mathbb{K}^n, \mathbb{K}^m)$ denote the linear maps from $\mathbb{K}^n$ to $\mathbb{K}^m$. Then $\mathbf{A} \mapsto f_{\mathbf{A}}$ defines a mapping $M_{m,n}(\mathbb{K}) \to \mathrm{Lin}(\mathbb{K}^n, \mathbb{K}^m)$. Proposition 7.2 shows that this mapping is injective, and Lemma 7.4 shows that it is surjective; so it is a bijection.

(If you're not used to it, this kind of construction – *mappings between spaces of mappings* – can be a bit confusing; but one gets used to it with practice.)

Composition and inverses

The motivation for the Definition 2.13 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 7.6 *Let $\mathbf{A} \in M_{m,n}(\mathbb{K})$ and $\mathbf{B} \in M_{n,r}(\mathbb{K})$, so we have maps $f_{\mathbf{A}} : \mathbb{K}^n \to \mathbb{K}^m$, $f_{\mathbf{B}} : \mathbb{K}^r \to \mathbb{K}^n$, and $f_{\mathbf{AB}} : \mathbb{K}^r \to \mathbb{K}^m$. Then $f_{\mathbf{AB}} = f_{\mathbf{A}} \circ f_{\mathbf{B}}$.*

Proof We'll interpret this as a special case of the associativity of matrix multiplication (part (v) of Proposition 2.16). Two mappings are equal if they take the same values on any input, so we need to show that $f_{\mathbf{AB}}(\vec{x}) = f_{\mathbf{A}}(f_{\mathbf{B}}(x))$ for all $\vec{x} \in \mathbb{K}^r$. Then we have

$$\begin{aligned} f_{\mathbf{AB}}(x) &= (\mathbf{A} \cdot \mathbf{B}) \cdot \vec{x} && \text{(regarding } \vec{x} \text{ as an } r \times 1 \text{ matrix)} \\ &= \mathbf{A} \cdot (\mathbf{B} \cdot \vec{x}) && \text{(by associativity)} \\ &= \mathbf{A} \cdot (f_{\mathbf{B}}(\vec{x})) \\ &= f_{\mathbf{A}}(f_{\mathbf{B}}(\vec{x})). \end{aligned}$$

□

Proposition 7.7 *Let $\mathbf{A} \in M_{m,n}(\mathbb{K})$. Then $\mathbf{A}$ is invertible if and only if the linear map $f_{\mathbf{A}}$ is bijective. If this is the case, the inverse mapping $(f_{\mathbf{A}})^{-1}$ is the mapping associated to the inverse matrix $\mathbf{A}^{-1}$, i.e. we have the relation*

$$(f_{\mathbf{A}})^{-1} = f_{\mathbf{A}^{-1}}.$$

Proof 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 $(f_{\mathbf{A}})^{-1} : \mathbb{K}^m \to \mathbb{K}^n$. We saw in the previous chapter that $(f_{\mathbf{A}})^{-1}$ is also a linear map. Hence, by Lemma 7.4, $(f_{\mathbf{A}})^{-1} = f_{\mathbf{B}}$ for some matrix $\mathbf{B} \in M_{n,m}(\mathbb{K})$. Using Theorem 7.6, we obtain

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

hence Proposition 7.2 implies that $\mathbf{BA} = \mathbf{1}_n$. Likewise we have

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

so that $\mathbf{AB} = \mathbf{1}_m$. Thus $\mathbf{A}$ is invertible, and $\mathbf{B}$ is its inverse.

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 $f_{\mathbf{B}} : \mathbb{K}^m \to \mathbb{K}^n$. □

Corollary 7.8 *A non-square matrix cannot be invertible.*

Proof If $\mathbf{A} \in M_{m,n}(\mathbb{K})$ is invertible, then $f_{\mathbf{A}}$ is an isomorphism between $\mathbb{K}^n$ and $\mathbb{K}^m$, and we have seen that no such isomorphism exists unless $m = n$. □

We also get a little extra information when $m = n$:

Proposition 7.9 *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 invertibility 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 7.6 and hence $f_\mathbf{B}$ is a left inverse for $f_\mathbf{A}$. Therefore $f_\mathbf{A}$ is injective (see Review Exercises on mappings). The implication (i) $\Rightarrow$ (ii) of Corollary 6.21 implies that $f_\mathbf{A}$ is actually bijective, so by Proposition 7.7, $\mathbf{A}$ is invertible.

(ii) $\Rightarrow$ (iii) is completely analogous – since $f_\mathbf{A}$ has a right inverse, it is surjective, hence bijective by Corollary 6.21 and we conclude as before. □

7.2 Computing kernels and images

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

$$\mathrm{rank}(\mathbf{A}) = \mathrm{rank}(f_\mathbf{A}), \qquad \mathrm{nullity}(\mathbf{A}) = \mathrm{nullity}(f_\mathbf{A}).$$

We want to compute these explicitly, and compute bases for the subspaces $\ker(f_\mathbf{A}) \subset \mathbb{K}^n$ and $\mathrm{Im}(f_\mathbf{A}) \subset \mathbb{K}^m$.

Image and rank

In order to compute a basis for $\mathrm{Im}(f_\mathbf{A})$ we use the following lemma:

Lemma 7.10 *The image of $f_\mathbf{A}$ is equal to the* column space *of $\mathbf{A}$, i.e. the subspace of $\mathbb{K}^m$ spanned by the column vectors $\vec{a}_1, \dots, \vec{a}_n$.*

Proof The columns $\vec{a}_1, \dots, \vec{a}_n$ of $\mathbf{A}$ are the images of the standard basis vectors, so they are clearly contained in $\mathrm{Im}(f_\mathbf{A})$; hence $\mathrm{span}(\vec{a}_1, \dots, \vec{a}_n) \subseteq \mathrm{Im}(f_\mathbf{A})$. Conversely, for any $\vec{x} = (x_j)_{1 \leqslant j \leqslant n} \in \mathbb{K}^n$ we have $f_\mathbf{A}(\vec{x}) = \sum_j x_j f(\vec{e}_j) = \sum_j x_j \vec{a}_j \in \mathrm{span}(\vec{a}_1, \dots, \vec{a}_n)$. □

We saw in Chapter 5 how to compute the a basis of the subspace generated by a finite list of vectors, so we can just apply that here to compute the image. However, since the columns of $\mathbf{A}$ are column vectors (not row vectors), we have to rewrite them as row vectors before applying RREF.

That is, to compute the image of $\mathbf{A}$, we need to perform the following steps:

- form the transpose matrix $\mathbf{A}^T$;
- compute its RREF;
- take the non-zero rows in the RREF matrix;

- transpose these again to get column vectors.

(Equivalently, we can directly apply *column* operations to $\mathbf{A}$ to get a "reduced column echelon form" of $\mathbf{A}$ whose columns are the desired basis; but we stick with RREF for the sake of familiarity.)

Example 7.11 *"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 $\mathbf{A}$."

The transpose matrix is

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

Computing its RREF yields

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

The non-zero row vectors are $\vec{\omega}_1 = \begin{pmatrix} 1 & 0 & 2 & -3 \end{pmatrix}$, $\vec{\omega}_2 = \begin{pmatrix} 0 & 1 & -1 & 2 \end{pmatrix}$. Our basis of $\mathrm{Im}(f)$ is thus

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

Since this basis has two elements, we also conclude that $\mathrm{rank}(\mathbf{A}) = 2$.

Kernel and nullity

We already know one way of computing the kernel: since the kernel of $f_{\mathbf{A}}$ is just the set $\mathrm{Sol}(\mathbf{A}, \mathbf{0})$, we can apply the general machinery of "free variables" etc. However, there is a slightly slicker way, which allows us to compute the kernel and the image at the same time:

Proposition 7.12 *Consider the augmented matrix $\mathbf{B} = (\mathbf{A}^T \mid I_n)$. Divide up the RREF of this matrix into the shape*

$$\left(\begin{array}{c|c} \mathbf{C} & \textit{(junk)} \\ \hline \mathbf{0} & \mathbf{D} \end{array} \right)$$

where $\mathbf{C}$ has no zero rows. Then the rows of $\mathbf{C}$ (transposed into column vectors) are a basis of the image of $f_{\mathbf{A}}$, and the rows of $\mathbf{D}$ (transposed) are a basis of the kernel.

Proof The first m columns of the RREF of $\mathbf{B}$ are the RREF of $\mathbf{A}^T$, and we already know this gives a basis of the image; so we need to show that the rest of the RREF gives a basis of the kernel.

Let's write $\tilde{\mathbf{D}}$ for the square matrix given by the last n columns of the RREF, so $\tilde{\mathbf{D}} = \begin{pmatrix} (junk) \\ \mathbf{D} \end{pmatrix}$. Note that the "junk" submatrix has r rows, where $r = \text{rank}(\mathbf{A})$. Moreover, $\mathbf{D}$ is itself in RREF, so its nonzero rows are linearly independent; and it can't have any zero rows, since $\tilde{\mathbf{D}}$ is invertible.

We know that $\tilde{\mathbf{D}}$ gives the transformation matrix to put $\mathbf{A}^T$ into RREF, so the RREF of $\mathbf{A}^T$ is equal to $\tilde{\mathbf{D}}\mathbf{A}^T$. But the i-th row of the RREF is zero for $i > r = \text{rank}(\mathbf{A})$; so if δ_i are the rows of $\tilde{\mathbf{D}}$, we have $\vec{\delta_i} \cdot \mathbf{A}^T = \vec{0}$ for $r+1 \leqslant i \leqslant n$. Transposing this, we have $\mathbf{A} \cdot \vec{\delta_i}^T = 0$ for all such i, so we obtain $n - r$ linearly independent vectors in the kernel. However, since the kernel has dimension $n - r$ from the rank-nullity theorem, these vectors must in fact be a basis. □

Remark 7.13 Actually the "junk" submatrix isn't really junk: one can check that for $1 \leqslant i \leqslant r$, the vector $\vec{\delta_i}^T$ is a choice of vector in $\mathbb{K}^n$ mapping under $f_{\mathbf{A}}$ to the i-th vector in our basis of the image.

Example 7.14 (Kernel of a linear map) Let

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

In order to compute $\text{Ker}(f_{\mathbf{C}})$ 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 so $\text{nullity}(\mathbf{C}) = 2$.

Exercises

Exercise 7.1 Compute a basis for the kernel of the linear map $f_{\mathbf{A}}$ from Example 7.11.

Exercise 7.2 Prove Remark 7.13, and use it to find, for each basis vector $\vec{b}$ we found for image$(f_{\mathbf{A}})$ in Example 7.11, a vector $\vec{a}$ with $f_{\mathbf{A}}(\vec{a}) = \vec{b}$.

Exercise 7.3 Let $\mathbf{M} \in M_{m,n}(\mathbb{K})$, with $m < n$.

(i) Show that $\mathbf{M}$ cannot have a left inverse.

(ii) Show that $\mathbf{M}$ has a right inverse if and only if its rank is m.

Compute a right inverse of the matrix

$$\mathbf{M} = \begin{pmatrix} \frac{1}{2} & -1 & 0 & \frac{1}{2} \\ 0 & 0 & -1 & 1 \\ 0 & -1 & 1 & -1 \end{pmatrix} \in M_{3,4}(\mathbb{R}).$$

Exercise 7.4

(i) Show that if $\mathbf{M}$ and $\mathbf{N}$ are two $m \times n$ matrices which are right-equivalent (i.e. there is an invertible square $\mathbf{A}$ such that $\mathbf{M} = \mathbf{NA}$), then $\mathbf{M}$ and $\mathbf{N}$ have the same rank.

(ii) Show that the same holds if $\mathbf{M}$ and $\mathbf{N}$ are left-equivalent. *(Careful: this is harder than (i), since the definitions are not symmetric! You may find Exercise 6.4 useful here.)*

(iii) Hence, or otherwise, show that for any matrix $\mathbf{M}$ we have $\text{rank}(\mathbf{M}^T) = \text{rank}(\mathbf{M})$ ("row rank equals column rank").

CHAPTER 8

Coordinate systems and changes of basis

Contents

8.1 Linear coordinate systems 73
8.2 The matrix of a linear map 76
8.3 Change of basis 80
Exercises 83

8.1 Linear coordinate systems

Notice that Proposition 6.24 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 8.1 (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 6.21.

Example 8.2 (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 8.3 (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 Calculus 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*

$$\mathcal{C} = \{s\vec{e}_1 + k\vec{e}_2 | s \in \mathbb{R}, k \in \mathbb{Z}\} \cup \{k\vec{e}_1 + s\vec{e}_2 | s \in \mathbb{R}, k \in \mathbb{Z}\} \tag{8.1}$$

under φ. The first set in the union (8.1) of sets are the *horizontal coordinate lines* and the second set the *vertical coordinate lines*.

Example 8.4 (see Figure 8.1) 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 $\varphi\left(\begin{pmatrix} v_1 \\ v_2 \end{pmatrix}\right) = \begin{pmatrix} v_1 + 2v_2 \\ -v_1 + v_2 \end{pmatrix}$.

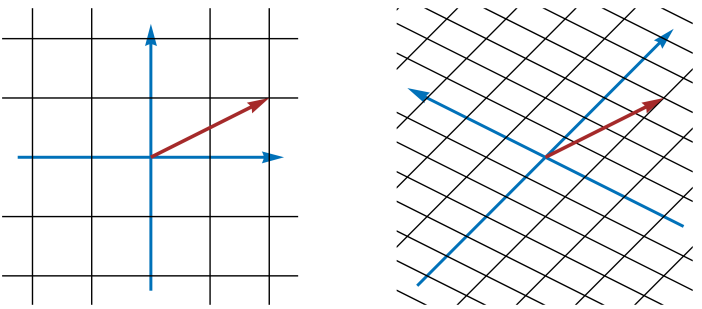

FIGURE 8.1. The coordinates of a vector with respect to different coordinate systems.

While $\mathbb{K}^n$ has an obvious "best" coordinate system – the identity map – 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.

Definition 8.5 (Ordered basis) Let V be a finite dimensional $\mathbb{K}$-vector space. An (ordered) n-tuple $\mathbf{b} = (v_1, \dots, v_n)$ of vectors from V is called an *ordered basis of V* if the set $\{v_1, \dots, v_n\}$ 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 8.6 *Let V, W be finite dimensional $\mathbb{K}$-vector spaces.*

(i) Suppose $f, g : V \to W$ are linear maps and $\mathbf{b} = (v_1, \dots, v_n)$ 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 $\mathbf{b} = (v_1, \dots, v_n)$ is an ordered basis of V and $\mathbf{c} = (w_1, \dots, w_n)$ 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 $s_1, \dots, s_n \in \mathbb{K}$ 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 $\{v_1, \dots, v_n\}$ is a basis of V there exist unique scalars $s_1, \dots, s_n$ 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 $\{w_1, \dots, w_n\}$ 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 6.14 implies that f is injective and hence an isomorphism by Corollary 6.21. The uniqueness of f follows from (i). □

Remark 8.7 Notice that Lemma 8.6 is wrong for maps that are not linear. Consider

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

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

$$\beta(v_i) = \vec{e}_i \tag{8.2}$$

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

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

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

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

Remark 8.8 (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 8.9 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

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

where ε denotes the linear coordinate system corresponding to $\mathbf{e}$. Notice that ε 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

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

since

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

8.2 The matrix of a linear map

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 8.10 (Matrix representation of a linear map) 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

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

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 8.11 *Let V, W be finite dimensional $\mathbb{K}$-vector spaces, $\mathbf{b} = (v_1, \dots, v_n)$ an ordered basis of V, $\mathbf{c} = (w_1, \dots, w_m)$ 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*

$$g(v_j) = \sum_{i=1}^{m} A_{ij} w_i, \qquad 1 \leqslant j \leqslant n. \tag{8.3}$$

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 8.12 Notice that we sum over the first index of A_{ij} in (8.3).

Proof of Proposition 8.11 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 $\mathbf{c} = (w_1, \dots, w_m)$, 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 $(w_1, \dots, w_m)$ 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 8.6 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 $\{\vec{e}_1, \dots, \vec{e}_n\}$ denote the standard basis of $\mathbb{K}^n$ so that $\beta(v_j) = \vec{e}_j$ and $\{\vec{d}_1, \dots, \vec{d}_m\}$ 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 (8.3). □

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

Example 8.13 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 $\mathbf{c}$.

$$\begin{aligned} 3x &= 3 \cdot w_1 + 0 \cdot w_2 \\ 1 &= 0 \cdot w_1 + 1 \cdot w_2 \\ 0 &= 0 \cdot w_1 + 0 \cdot w_2 \end{aligned} \tag{8.4}$$

(iii) Taking the transpose of the matrix of coefficients appearing in (8.4) 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 8.14 Let $\mathbf{e} = (\vec{e}_1, \dots, \vec{e}_n)$ and $\mathbf{d} = (\vec{d}_1, \dots, \vec{d}_m)$ 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 8.15 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 8.16 *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 8.10, 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 8.17 Explicitly, Proposition 8.16 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 8.18 (Example 8.13 continued) With respect to the ordered basis $\mathbf{b} = \left(\frac{1}{2}(3x^2 - 1), x, 1\right)$, the polynomial $a_2x^2 + a_1x + a_0 \in V = \mathrm{P}_2(\mathbb{R})$ 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 $\mathrm{P}_1(\mathbb{R})$.

As a corollary to Proposition 8.11 we obtain:

Corollary 8.19 *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 7.2 and Theorem 7.6 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 8.11 gives

$$f_{\mathbf{A}_2} \circ f_{\mathbf{A}_1} = \beta_3 \circ g_2 \circ \beta_2^{-1} \circ \beta_2 \circ g_1 \circ \beta_1^{-1} = \beta_3 \circ g_2 \circ g_1 \circ \beta_1^{-1} = f_{\mathbf{C}}. \qquad \square$$

Proposition 8.20 *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 6.24. Then Corollary 8.19 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 7.8. We consider $h = \beta^{-1} \circ f_{\mathbf{A}^{-1}} \circ \gamma : W \to V$ and since $f_{\mathbf{A}} = \gamma \circ g \circ \beta^{-1}$ by Proposition 8.11, we have

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

Likewise, we have

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

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

Remark 8.21 (Computing kernels and images) 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 $\mathbf{C} = \mathbf{M}(g, \mathbf{b}, \mathbf{c})$, and apply the methods of Section 7.2 to the matrix $\mathbf{C}$ in order to obtain a basis $\mathcal{S}$ of $\mathrm{Ker}(f_{\mathbf{C}})$. The desired basis of $\mathrm{Ker}(g)$ is then given by $\beta^{-1}(\mathcal{S})$ (and we can compute the image similarly).
(While this algorithm can always be carried out in principle, it is computationally quite involved and error-prone to do by hand.)

Example 8.22 (Basis of a subspace) Let $V = \mathrm{P}_3(\mathbb{R})$ so that $\dim(V) = 4$ and

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

We want to compute a basis of U.

The obvious ordered basis of V is $\mathbf{f} = \{x^3, x^2, x, 1\}$, and the corresponding coordinate system ϕ is the isomorphism $V \to \mathbb{R}_4$ defined by

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

The images of the given generators of U are the row vectors $\vec{\nu}_1, \ldots, \vec{\nu}_4$ given by $\vec{\nu}_1 = \begin{pmatrix} 1 & 2 & -1 & 0 \end{pmatrix}$, $\vec{\nu}_2 = \begin{pmatrix} 4 & 8 & -4 & -3 \end{pmatrix}$, $\vec{\nu}_3 = \begin{pmatrix} 0 & 1 & 3 & 4 \end{pmatrix}$ and $\vec{\nu}_4 = \begin{pmatrix} 2 & 5 & 1 & 4 \end{pmatrix}$.

We form the matrix $\mathbf{N}$ with the $\vec{\nu}_i$ as rows:

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

The RREF of $\mathbf{N}$ is

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

(Here we applied Gauss-Jordan elimination, but Gaussian elimination would be enough.) The non-zero rows of the RREF matrix are the vectors $\vec{\omega}_1 = \begin{pmatrix} 1 & 0 & -7 & 0 \end{pmatrix}$, $\vec{\omega}_2 = \begin{pmatrix} 0 & 1 & 3 & 0 \end{pmatrix}$, and $\vec{\omega}_3 = \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix}$.

So a basis of U is given by

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

where we use that

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

8.3 Change of basis

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

Definition 8.23 (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 β, β'. The *change of basis matrix from* $\mathbf{b}$ *to* $\mathbf{b}'$ is the matrix $\mathbf{C} \in M_{n,n}(\mathbb{K})$ satisfying

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

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

Remark 8.24 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 8.20 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 8.25 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 8.13, 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 8.26 *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 8.24 we have $\mathbf{C}^{-1} = \mathbf{C}(\mathbf{b}', \mathbf{b})$, hence applying Proposition 7.2 and Theorem 7.6, we need to show that

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

By Definition 8.10 we have

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

and by Definition 8.23 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 8.24. □

Example 8.27 (Example 8.15 and Example 8.25 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 8.26 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 8.28 *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

Lemma 5.15 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

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

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 8.1 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 8.2 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

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

Exercise 8.3 Consider the vector space $P_2(\mathbb{R})$ of real polynomials of degree $\leqslant 2$. Let $\mathbf{b}$ be the basis $(x^2 + x + 1, x + 1, 1)$, and let $\mathbf{e}$ be the obvious basis $(1, x, x^2)$

(i) Write down the coordinate vectors $\beta(f)$ and $\varepsilon(f)$ of $f = 1 + 2x + 3x^2$ in each of the bases $\mathbf{b}$ and $\mathbf{e}$.

(ii) Compute the change-of-basis matrix $C(\mathbf{e}, \mathbf{b})$ and verify that the formula $\beta(f) = C(\mathbf{e}, \mathbf{b})\varepsilon(f)$ holds.

(iii) Write down the matrix of the differentiation map $\frac{d}{dx} : P_2(\mathbb{R}) \to P_2(\mathbb{R})$ in the basis $\mathbf{e}$, and use the basis-change formula to compute its matrix with respect to $\mathbf{b}$.

CHAPTER 9

The determinant, I

Contents

9.1	**Axiomatic characterisation**	**84**
9.1.1	Multilinear maps	84
9.1.2	Existence and uniqueness	86
9.2	**Uniqueness of the determinant**	**88**
9.3	**Existence of the determinant**	**90**
Exercises		**93**

9.1 Axiomatic characterisation

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.

9.1.1 Multilinear maps

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 9.1 (Multilinear map) 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 $g_k : V \to W$, $v \mapsto f(v_1, \dots, v_{k-1}, v, v_{k+1}, \dots, v_\ell)$ is linear for all $1 \leqslant k \leqslant \ell$ and for all ℓ-tuples $(v_1, \dots, v_\ell) \in V^\ell$.

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 9.2 (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
$$f(v_1, \dots, v_\ell) = f(v_1, \dots, v_{i-1}, v_j, v_{i+1}, \dots, v_{j-1}, v_i, v_{j+1}, \dots, v_\ell)$$
for all $(v_1, \dots, v_\ell) \in V^\ell$.
- The map f is called *alternating* if $f(v_1, \dots, v_\ell) = 0_W$ 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 9.3 (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 9.4 (Multilinear map) Let $V = \mathbb{K}$ and consider $f : V^\ell \to \mathbb{K}, (x_1, \dots, x_\ell) \mapsto x_1 x_2 \cdots x_\ell$. Then f is ℓ-multilinear and symmetric.

Example 9.5 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 9.5 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 $\mathbb{K}_n$).

Remark 9.6 (Alternating and skew-symmetric maps) We say an ℓ- multilinear map $f : V^\ell \to W$ is said to be *antisymmetric* if interchanging any two of its inputs results in a minus sign, i.e.

$$f(v_1, \dots, v_\ell) = -f(v_1, \dots, v_{i-1}, v_j, v_{i+1}, \dots, v_{j-1}, v_i, v_{j+1}, \dots, v_\ell)$$

for all $v_1, \dots, v_n$ and $1 \leqslant i < j \leqslant n$.
One can check that any alternating map is antisymmetric. Let's show this for $n = 2$. Assuming f is alternating, for any $v_1, v_2 \in V$ we have

$$\begin{aligned} 0 &= f(v_1 + v_2, v_1 + v_2) \\ &= f(v_1, v_1) + f(v_1, v_2) + f(v_2, v_1) + f(v_2, v_2) \\ &= 0 + f(v_1, v_2) + f(v_2, v_1) + 0 \end{aligned}$$

so $f(v_2, v_1) + f(v_1, v_2) = 0$.
On the other hand, if f is antisymmetric, then we have $f(v, v) = -f(v, v)$ for all v (since we can swap v with itself); so $2f(v, v) = 0$. But this does *not* imply that f is alternating, since this '2' means $1_\mathbb{K} + 1_\mathbb{K}$, and there exist fields such that $1_\mathbb{K} + 1_\mathbb{K} = 0_\mathbb{K}$!
Of course, if $\mathbb{K}$ is one of the familiar fields like $\mathbb{R}$ or $\mathbb{C}$, where $2 \neq 0$, then "alternating" and "antisymmetric" are the same. But in general being alternating is a more restrictive condition.

9.1.2 Existence and uniqueness

Theorem 9.7 *Let $n \in \mathbb{N}$, and let $\{\vec{\varepsilon}_1, \dots, \vec{\varepsilon}_n\}$ denote the standard basis of $\mathbb{K}_n$. Then there exists a unique alternating n-multilinear map $f_n : (\mathbb{K}_n)^n \to \mathbb{K}$ satisfying $f_n(\vec{\varepsilon}_1, \dots, \vec{\varepsilon}_n) = 1$.*

It's helpful to rephrase this statement in terms of matrices. Let us write

$$\Omega : (\mathbb{K}_n)^n \to M_{n,n}(\mathbb{K})$$

for the map sending n row vectors of length n to the $n \times n$ matrix with those vectors as its rows. This map is clearly a bijection, so it makes sense to define a mapping $f : M_{n,n}(\mathbb{K}) \to \mathbb{K}$ to be "multilinear" if $f \circ \Omega$ is multilinear, i.e. if f is linear in each row of the matrix. Similarly, we define f to be "alternating" if $f \circ \Omega$ is, so $f(\mathbf{A}) = 0$ whenever two of the rows of $\mathbf{A}$ are equal.

Since $\Omega(\vec{\varepsilon}_1, \dots, \vec{\varepsilon}_n) = \mathbf{1}_n$, we may phrase the above theorem equivalently as:

Theorem 9.8 (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 9.9 (Determinant) The mapping $f_n : M_{n,n}(\mathbb{K}) \to \mathbb{K}$ provided by Theorem 9.8 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 9.10 (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 9.11 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 9.8 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 9.12 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

$$f_2\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = ad - bc. \tag{9.1}$$

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 (9.1), 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

$$f_2\left(\begin{pmatrix} sa & sb \\ c & \mathrm{d} \end{pmatrix}\right) = s f_2\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right)$$

for all $a, b, c, d \in \mathbb{K}$ and $s \in \mathbb{K}$. Indeed, using the definition (9.1), we obtain

$$f_2\left(\begin{pmatrix} sa & sb \\ c & \mathrm{d} \end{pmatrix}\right) = sad - csb = s(ad - cb) = s f_2\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right),$$

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 9.8, hence by the uniqueness statement we must have $f_2 = \det$ and we obtain the formula

$$\det\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = ad - cb \tag{9.2}$$

for all $a, b, c, d \in \mathbb{K}$.

9.2 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 9.8 in general. We first show the uniqueness part. We start by deducing some consequences from the alternating property:

Lemma 9.13 *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

$$f(v_1, \dots, v_\ell) = -f(v_1, \dots, v_{i-1}, v_j, v_{i+1}, \dots, v_{j-1}, v_i, v_{j+1}, \dots, v_\ell)$$

for all $(v_1, \dots, v_\ell) \in V^\ell$;

(ii) if the vectors $(v_1, \dots, v_\ell) \in V^\ell$ are linearly dependent, then $f(v_1, \dots, v_\ell) = 0_W$;

(iii) for all $1 \leqslant i \leqslant \ell$, for all ℓ-tuples of vectors $(v_1, \dots, v_\ell) \in V^\ell$ and scalars $s_1, \dots, s_\ell \in \mathbb{K}$, we have

$$f(v_1, \dots, v_{i-1}, v_i + w, v_{i+1}, \dots, v_\ell) = f(v_1, \dots, v_\ell)$$

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 $(v_1, \dots, v_\ell) \in V^\ell$

$$f(v_1, \dots, v_{i-1}, v_i + v_j, v_{i+1}, \dots, v_{j-1}, v_i + v_j, v_{j+1}, \dots, v_\ell) = 0_W.$$

Using the linearity in the i-th argument, this gives

$$\begin{aligned} 0_W = {} & f(v_1, \dots, v_{i-1}, v_i, v_{i+1}, \dots, v_{j-1}, v_i + v_j, v_{j+1}, \dots, v_\ell) \\ & + f(v_1, \dots, v_{i-1}, v_j, v_{i+1}, \dots, v_{j-1}, v_i + v_j, v_{j+1}, \dots, v_\ell). \end{aligned}$$

Using the linearity in the j-th argument, we obtain

$$\begin{aligned} 0_W = {} & f(v_1, \dots, v_{i-1}, v_i, v_{i+1}, \dots, v_{j-1}, v_i, v_{j+1}, \dots, v_\ell) \\ & + f(v_1, \dots, v_{i-1}, v_i, v_{i+1}, \dots, v_{j-1}, v_j, v_{j+1}, \dots, v_\ell) \\ & + f(v_1, \dots, v_{i-1}, v_j, v_{i+1}, \dots, v_{j-1}, v_i, v_{j+1}, \dots, v_\ell) \\ & + f(v_1, \dots, v_{i-1}, v_j, v_{i+1}, \dots, v_{j-1}, v_j, v_{j+1}, \dots, v_\ell). \end{aligned}$$

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

$$f\left(v_1, \dots, v_{i-1}, -\sum_{j=1, j\neq i}^{\ell} \left(\frac{s_j}{s_i}\right) v_j, v_{i+1}, \dots, v_\ell\right) = -\sum_{j=1, j\neq i}^{\ell} \left(\frac{s_j}{s_i}\right) f\left(v_1, \dots, v_{i-1}, v_j, v_{i+1}, \dots, v_\ell\right) = 0_W,$$

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

$$f\Big(v_1, \dots, v_{i-1}, v_i + \sum_{j=1, j\neq i}^{\ell} s_j v_j, v_{i+1}, \dots, v_\ell\Big) = f(v_1, \dots, v_\ell) + \sum_{j=1, j\neq i}^{\ell} s_j f(v_1, \dots, v_{i-1} v_j, v_{i+1}, \dots, v_\ell) = f(v_1, \dots, v_\ell),$$

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 9.14 *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*

$$f_n(\mathbf{D}_k(s)) = s, \qquad f_n(\mathbf{L}_{k,l}(s)) = 1, \qquad f_n(\mathbf{P}_{k,l}) = -1. \tag{9.3}$$

Moreover, for $\mathbf{A} \in M_{n,n}(\mathbb{K})$ and an elementary matrix $\mathbf{B}$ of size n, we have

$$f_n(\mathbf{BA}) = f_n(\mathbf{B}) f_n(\mathbf{A}). \tag{9.4}$$

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 9.13 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 9.13 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 9.8.

Proposition 9.15 *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 3.18, the row vectors of $\mathbf{A}$ are linearly dependent and hence property (ii) of Lemma 9.13 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 $\mathbf{B}_1\mathbf{B}_2 \cdots \mathbf{B}_N = \mathbf{A}$.

Applying (9.4) repeatedly, we have

$$f_n(\mathbf{A}) = f_n(\mathbf{B}_1 \cdots \mathbf{B}_N) = f_n(\mathbf{B}_1) \dots f_n(\mathbf{B}_N)$$

and similarly

$$\hat{f}_n(\mathbf{A}) = \hat{f}_n(\mathbf{B}_1 \cdots \mathbf{B}_N) = \hat{f}_n(\mathbf{B}_1) \cdots \hat{f}_n(\mathbf{B}_N).$$

But (9.3) implies that $\hat{f}_n(\mathbf{B}_j) = f_n(\mathbf{B}_j)$ for all j, so these two products are equal. □

9.3 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 9.16 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 9.17

$$\text{If } \mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \text{ then } \quad \mathbf{A}^{(1,1)} = (d), \qquad \mathbf{A}^{(2,1)} = (b).$$

$$\text{If } \mathbf{A} = \begin{pmatrix} 1 & -2 & 0 & 4 \\ 3 & 1 & 1 & 0 \\ -1 & -5 & -1 & 8 \\ 3 & 8 & 2 & -12 \end{pmatrix}, \text{ then } \quad \mathbf{A}^{(3,2)} = \begin{pmatrix} 1 & 0 & 4 \\ 3 & 1 & 0 \\ 3 & 2 & -12 \end{pmatrix}.$$

We use induction to prove the existence of the determinant:

Lemma 9.18 *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 9.7 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 9.18 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 9.15 there is only one such mapping for each $n \in \mathbb{N}$. □

Proof of Lemma 9.18 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 $f_n(\hat{\mathbf{A}}) = s f_n(\mathbf{A})$. We compute

$$\begin{aligned} f_n(\hat{\mathbf{A}}) &= \sum_{k=1}^{n} (-1)^{l+k} \hat{A}_{kl} f_{n-1}(\hat{\mathbf{A}}^{(k,l)}) \\ &= (-1)^{l+i} s A_{il} f_{n-1}(\hat{\mathbf{A}}^{(i,l)}) + \sum_{k=1, k \neq i}^{n} (-1)^{l+k} A_{kl} f_{n-1}(\hat{\mathbf{A}}^{(k,l)}). \end{aligned}$$

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

$$f_n(\hat{\mathbf{A}}) = s(-1)^{l+i} A_{il} f_{n-1}(\mathbf{A}^{(i,l)}) + s \sum_{k=1, k\neq i}^{n} (-1)^{l+k} A_{kl} f_{n-1}(\mathbf{A}^{(k,l)})$$

$$= s \sum_{k=1}^{n} (-1)^{l+k} A_{kl} f_{n-1}\left(\mathbf{A}^{(k,l)}\right) = s f_n(\mathbf{A}).$$

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

$$f_n(\mathbf{C}) = (-1)^{l+i} B_l f_{n-1}(\mathbf{B}^{(i,l)}) + \sum_{k=1, k\neq i}^{n} (-1)^{l+k} A_{kl} f_{n-1}(\mathbf{B}^{(k,l)})$$

$$+ (-1)^{l+i} A_{il} f_{n-1}(\mathbf{A}^{(i,l)}) + \sum_{k=1, k\neq i}^{n} (-1)^{l+k} A_{kl} f_{n-1}(\mathbf{A}^{(k,l)}) = f_n(\mathbf{A}) + f_n(\mathbf{B}).$$

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

$$f_n(\mathbf{A}) = (-1)^{i+l} A_{il} f_{n-1}(\mathbf{A}^{(i,l)}) + (-1)^{j+l} A_{jl} f_{n-1}(\mathbf{A}^{(j,l)})$$

$$= A_{il}(-1)^l \left((-1)^i f_{n-1}(\mathbf{A}^{(i,l)}) + (-1)^j f_{n-1}(\mathbf{A}^{(j,l)})\right)$$

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 9.13, 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 9.19 (Laplace expansion) As a by-product of the proof of Lemma 9.18 we obtain the formula

$$\det(\mathbf{A}) = \sum_{k=1}^{n} (-1)^{l+k} [\mathbf{A}]_{kl} \det\left(\mathbf{A}^{(k,l)}\right), \tag{9.5}$$

known as the *Laplace expansion* of the determinant (along the l-th column). The uniqueness statement of Theorem 9.8 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 9.20 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 (9.1). 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 9.1 (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.

Exercise 9.2 Define the matrix

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

Compute the determinant of $\mathbf{A}$ by Laplace expansion with respect to column ℓ for each $\ell \in \{1, 2, 3\}$. Conclude that all choices for ℓ give the same answer.

Exercise 9.3 Use the explicit formulae in Example 9.20 to show that we have $\det(\mathbf{A}^T) = \det(\mathbf{A})$ for all $n \times n$ square matrices with $n \leqslant 3$.

Exercise 9.4 Let $\mathbf{A} \in M_{m,m}(\mathbb{K})$ and $\mathbf{B} \in M_{n,n}(\mathbb{K})$. Show that the determinant of the $(m+n) \times (m+n)$ matrix $\begin{pmatrix} \mathbf{A} & 0 \\ 0 & \mathbf{B} \end{pmatrix}$ is given by $\det(\mathbf{A}) \det(\mathbf{B})$.
(Hint: Use induction on m, and do a Laplace expansion on the first column.)

CHAPTER 10

The determinant, II

Contents

10.1 Properties of the determinant 95
10.2 Permutations 96
10.3 The Leibniz formula 99
10.4 Cramer's rule 102
Exercises 104

10.1 Properties of the determinant

Proposition 10.1 (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 9.15). 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 the matrix product is associative, this also gives $\mathbf{A}(\mathbf{BC}) = \mathbf{1}_n$, so that $\mathbf{BC}$ is a right inverse of $\mathbf{A}$. By Proposition 7.7, $\mathbf{A}$ is invertible, 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 9.14, we conclude that $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$. □

Corollary 10.2 *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 10.1 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 10.3 (Product symbol) Recall that for scalars $x_1, \dots, x_n \in \mathbb{K}$, we write

$$\prod_{i=1}^{n} x_i = x_1 x_2 \cdots x_n.$$

Proposition 10.4 *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*

$$\det(\mathbf{A}) = \prod_{i=1}^{n} A_{ii} = A_{11} A_{22} \cdots A_{nn}. \tag{10.1}$$

Proof We use induction. For $n = 1$ the condition $A_{ij} = 0$ for $i > j$ is vacuous and (10.1) 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 (10.1) 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

$$\begin{aligned} \det(\mathbf{A}) &= \sum_{k=1}^{n} (-1)^{k+1} A_{k1} \det\left(\mathbf{A}^{(k,1)}\right) = A_{11} \det\left(\mathbf{A}^{(1,1)}\right) + \sum_{k=2}^{n} A_{k1} \det\left(\mathbf{A}^{(k,1)}\right) \\ &= A_{11} \det\left(\mathbf{A}^{(1,1)}\right), \end{aligned}$$

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. □

10.2 Permutations

A rearrangement of the natural numbers from 1 up to n is called a permutation:

Definition 10.5 (Permutation) 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 10.6 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 10.7 (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 10.8 (Permutation matrix) We call $\mathbf{P}_\sigma \in M_{n,n}(\mathbb{K})$ the *permutation matrix* associated to $\sigma \in S_n$.

Example 10.9 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 10.10 Notice that $\mathbf{P}_{\tau_{k,l}} = \mathbf{P}_{k,l}$, where $\mathbf{P}_{k,l}$ is one of the elementary matrices of size n (see M01 Algorithmics).

Assigning to a permutation its permutation matrix turns composition of permutations into matrix multiplication:

Proposition 10.11 *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 10.12 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 10.11.

Proof of Proposition 10.11 The matrix $\mathbf{P}_1$ has column vectors given by $\vec{e}_1, \dots, \vec{e}_n$, hence $\mathbf{P}_1 = \mathbf{1}_n$.

Using Proposition 7.2 and Theorem 7.6 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 $\mathbf{e} = (\vec{e}_1, \dots, \vec{e}_n)$ of $\mathbb{K}^n$, so that the second claim follows by applying Lemma 8.6.

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 10.13 (Signature of a permutation) For $\sigma \in S_n$ we call $\operatorname{sgn}(\sigma) = \det(\mathbf{P}_\sigma)$ its *signature*.

Remark 10.14

(i) Combining Proposition 10.1 and Proposition 10.11, 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 9.14, 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 10.15 *Let $n \in \mathbb{N}$ and $\sigma \in S_n$. Then there exists $m \geqslant 0$ and m transpositions $\tau_{k_1,l_1}, \dots, \tau_{k_m,l_m} \in S_n$ 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 10.16 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 10.15 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 10.13, Remark 10.14 and Proposition 10.15, we conclude:

Proposition 10.17 *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 10.18 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.

10.3 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 $\{\vec{\alpha}_1, \dots, \vec{\alpha}_n\}$. 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

$$\det \mathbf{A} = \sum_{k=1}^{n^n} \det \mathbf{M}_k \tag{10.2}$$

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 $\{1, \dots, n\}$ occurs precisely once in the expansion (10.2), 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 $\{1, \dots, n\}$. 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, 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

$$\det\mathbf{A} = \sum_{\sigma\in S_n}\det\left(\mathbf{P}_\sigma\mathbf{D}_\sigma\right) = \sum_{\sigma\in S_n}\operatorname{sgn}(\sigma)\det(\mathbf{D}_\sigma),$$

where we use the product rule Proposition 10.1 and the definition of the signature of a permutation. By Proposition 10.4, 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 10.19 (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*

$$\det(\mathbf{A}) = \sum_{\sigma\in S_n}\operatorname{sgn}(\sigma)\prod_{i=1}^n A_{\sigma(i)i} = \sum_{\pi\in S_n}\operatorname{sgn}(\pi)\prod_{j=1}^n A_{j\pi(j)}. \tag{10.3}$$

Example 10.20 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 9.20, 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 10.21 It follows from Leibniz' formula that $\det(\mathbf{A}) = \det(\mathbf{A}^T)$ (see Exercise 10.3 below). This has the following important consequences:

(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 10.22 (♡ – 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 $q(\vec{x})$.

On the other hand, if we multiply all the x_i by a constant $s \in \mathbb{K}$, the determinant multiplies by $s^{1+2+\cdots+(n-1)} = s^{n(n-1)/2}$. The product $\prod_{1 \leqslant i < j \leqslant n} (x_j - x_i)$ has the same scaling behaviour (since it has $n(n-1)/2$ linear factors); so q must be invariant under scaling the variables. This implies that q has to be constant.

Using the Leibniz formula, 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, $x_2 (x_3)^2 \cdots (x_n)^{n-1}$ (and no other term in the sum has this combination of exponents). Taking the first term in all factors of $\prod_{1 \leqslant i < j \leqslant n} (x_j - x_i)$, we also obtain $x_2 (x_3)^2 \cdots (x_n)^{n-1}$ (and no other term in the product has these exponents). Hence the constant q must be 1, and so $\det(\mathbf{V}_{\vec{x}}) = \prod_{1 \leqslant i < j \leqslant n} (x_j - x_i)$, as claimed. □

10.4 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 10.23 (Adjugate matrix) 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 $\operatorname{Adj}(\mathbf{A})$ whose entries are given by (notice the reverse order of i and j on the right hand side)

$$[\operatorname{Adj}(\mathbf{A})]_{ij} = (-1)^{i+j} \det\left(\mathbf{A}^{(j,i)}\right), \qquad 1 \leqslant i, j \leqslant n.$$

Example 10.24

$$\operatorname{Adj}\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}, \quad \operatorname{Adj}\left(\begin{pmatrix} 1 & 1 & 2 \\ 0 & 2 & 1 \\ 1 & 0 & 2 \end{pmatrix}\right) = \begin{pmatrix} 4 & -2 & -3 \\ 1 & 0 & -1 \\ -2 & 1 & 2 \end{pmatrix}$$

The determinant and the adjugate matrix provide a formula for the inverse of a matrix:

Theorem 10.25 *Let $n \in \mathbb{N}$ and $\mathbf{A} \in M_{n,n}(\mathbb{K})$. Then we have*

$$\operatorname{Adj}(\mathbf{A})\mathbf{A} = \mathbf{A}\operatorname{Adj}(\mathbf{A}) = \det(\mathbf{A})\mathbf{1}_n.$$

In particular, if $\mathbf{A}$ is invertible then

$$\mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}} \operatorname{Adj}(\mathbf{A}).$$

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

$$[\operatorname{Adj}(\mathbf{A})\mathbf{A}]_{ii} = \sum_{k=1}^{n} (-1)^{i+k} \det\left(\mathbf{A}^{(k,i)}\right) A_{ki} = \det(\mathbf{A}),$$

where we use the Laplace expansion (9.5) of the determinant. The diagonal entries of $\operatorname{Adj}(\mathbf{A})\mathbf{A}$ are thus all equal to $\det \mathbf{A}$. For $1 \leqslant i, j \leqslant n$ with $i \neq j$ we have

$$[\operatorname{Adj}(\mathbf{A})\mathbf{A}]_{ij} = \sum_{k=1}^{n} (-1)^{i+k} \left(\det \mathbf{A}^{(k,i)}\right) A_{kj}.$$

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

$$\hat{A}_{kl} = \begin{cases} A_{kj}, & l = i, \\ A_{kl}, & l \neq i. \end{cases} \tag{10.4}$$

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 (10.4), the Laplace expansion of $\hat{\mathbf{A}}$ with respect to the i-th column gives

$$\begin{aligned} \det \hat{\mathbf{A}} &= \sum_{k=1}^{n} (-1)^{(i+k)} \hat{A}_{ki} \det\left(\hat{\mathbf{A}}^{(k,i)}\right) = \sum_{k=1}^{n} (-1)^{i+k} \left(\det \mathbf{A}^{(k,i)}\right) A_{kj} \\ &= [\operatorname{Adj}(\mathbf{A})\mathbf{A}]_{ij} \end{aligned}$$

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 3.18 and so $\det \hat{\mathbf{A}} = [\operatorname{Adj}(\mathbf{A})\mathbf{A}]_{ij} = 0$ by Corollary 10.2. The off-diagonal entries of $\operatorname{Adj}(\mathbf{A})\mathbf{A}$ are thus all zero and we conclude $\operatorname{Adj}(\mathbf{A})\mathbf{A} = \det(\mathbf{A})\mathbf{1}_n$. Using the row version of the Laplace expansion we can conclude analogously that $\mathbf{A}\operatorname{Adj}(\mathbf{A}) = \det(\mathbf{A})\mathbf{1}_n$.

Finally, if $\mathbf{A}$ is invertible, then $\det \mathbf{A} \neq 0$ by Corollary 10.2, so that $\mathbf{A}^{-1} = \operatorname{Adj}(\mathbf{A})/\det(\mathbf{A})$, as claimed. □

We now use Theorem 10.25 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 10.26 (Cramer's rule) We consider the system $\mathbf{A}\vec{x} = \vec{b}$ for

$$\mathbf{A} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} \qquad \text{and} \qquad \vec{b} = \begin{pmatrix} -2 \\ 2 \\ 4 \end{pmatrix}.$$

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 10.1 Let

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 3 & 2 & 4 & 1 \end{pmatrix}, \qquad \tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 5 & 4 & 3 & 1 \end{pmatrix}.$$

(i) Compute $\sigma\tau$.
(ii) Express σ and τ products of transpositions.

Exercise 10.2 (Sign of an n-cycle) Show that the permutation $\sigma_n \in S_n$ defined by $\sigma_n(i) = i + 1$ for $1 \leqslant i \leqslant n - 1$, and $\sigma_n(n) = 1$, has $\operatorname{sgn}(\sigma_n) = (-1)^{n+1}$.

Exercise 10.3 Use the Leibniz formula to show that

$$\det(\mathbf{A}) = \det(\mathbf{A}^T)$$

for all n and all $\mathbf{A} \in M_{n,n}(\mathbb{K})$.

Exercise 10.4 Let $\mathbf{A} = \begin{pmatrix} -1 & -2 & 11 \\ 1 & -1 & -1 \\ 6 & 0 & 0 \end{pmatrix}$. Compute $\operatorname{Adj} \mathbf{A}$, and hence $\det \mathbf{A}$.

CHAPTER 11

Endomorphisms, I

Contents

11.1	**Matrices of endomorphisms**	**106**
	Similarity	106
	Invariants of similarity classes	107
11.2	**Detour: More on Subspaces**	**110**
	Sums of subspaces	110
	Direct sums	110
	Complements	111
11.3	**Eigenvectors and eigenvalues**	**113**
11.4	**The characteristic polynomial**	**117**
Exercises		**119**

In this chapter we study linear mappings from a vector space to itself.

Definition 11.1 (Endomorphism) 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*.

Working with endomorphisms has a different flavour than working with linear maps between two different spaces. In particular, if we want to write g as a matrix, it clearly makes sense to work with just one coordinate system for V – that is, we want to consider the matrices $M(g, \mathbf{b}, \mathbf{b})$ for $\mathbf{b}$ an ordered basis of V.

11.1 Matrices of endomorphisms

Similarity

Let V be a finite dimensional vector space equipped with an ordered basis $\mathbf{b}$ and $g : V \to V$ an endomorphism. As a special case of Theorem 8.26, we see 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 11.2 (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 11.3 *Let $n \in \mathbb{N}$ and $\mathbf{A}, \mathbf{B}, \mathbf{X} \in M_{n,n}(\mathbb{K})$. Then we have*
(i) **A** *is similar to itself;*
(ii) **A** *is similar to* **B** *then* **B** *is similar to* **A***;*
(iii) If **A** *is similar to* **B** *and* **B** *is similar to* **X***, then* **A** *is also similar to* **X***.*

Proof (i) We take $\mathbf{C} = \mathbf{1}_n$.

(ii) Suppose **A** is similar to **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 **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 **C**, **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 11.4
- Because of (ii) in particular, one can say that two matrices **A** and **B** are similar without ambiguity.
- Theorem 8.26 shows that **A** and **B** are similar if and only if there exists an endomorphism g of $\mathbb{K}^n$ such that **A** and **B** represent g with respect to two ordered bases of $\mathbb{K}^n$.

The main goal of this chapter (and the next) is: *given an endomorphism g, how can we choose a basis* **b** *which makes the matrix of g as nice as possible?* Equivalently, given a square matrix **A**, is there a "nicest" matrix among all the matrices similar to **A**?

Remark 11.5 This should remind you a lot of row echelon form. The RREF of **A** is a unique "best" representative among all the matrices which are *left-equivalent* to **A**. We're now looking for a unique "best" representative among all matrices *similar* to **A**.
(This sort of classification problem – define some equivalence relation, and then look for a nicest representative of each equivalence class – comes up a great deal in many areas of mathematics.)

Invariants of similarity classes

As a first step, we want to study 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 10.1 and Corollary 10.2, we compute

$$\begin{aligned} \det\left(\mathbf{C}\mathbf{A}\mathbf{C}^{-1}\right) &= \det(\mathbf{C}\mathbf{A})\det\left(\mathbf{C}^{-1}\right) = \det(\mathbf{C})\det(\mathbf{A})\det\left(\mathbf{C}^{-1}\right) \\ &= \det(\mathbf{A}). \end{aligned} \tag{11.1}$$

Because of this fact, the following definition makes sense:

Definition 11.6 (Determinant of an endomorphism) Let V be a finite dimensional $\mathbb{K}$-vector space and $g : V \to V$ an endomorphism. We define

$$\det(g) = \det\left(\mathbf{M}(g, \mathbf{b}, \mathbf{b})\right)$$

where $\mathbf{b}$ is any ordered basis of V. By Theorem 8.26 and (11.1), 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 11.7 (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 $\mathrm{Tr}(\mathbf{A})$ or $\mathrm{Tr}\,\mathbf{A}$.

Example 11.8 For all $n \in \mathbb{N}$ we have $\mathrm{Tr}(\mathbf{1}_n) = n$. For

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

we have $\mathrm{Tr}(\mathbf{A}) = 2 + 2 + 3 = 7$.

The trace of a product of square matrices is independent of the order of multiplication:

Proposition 11.9 *Let $n \in \mathbb{N}$ and $\mathbf{A}, \mathbf{B} \in M_{n,n}(\mathbb{K})$. Then we have*

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

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

$$\mathrm{Tr}(\mathbf{AB}) = \sum_{i=1}^{n} \sum_{k=1}^{n} A_{ik} B_{ki} = \sum_{k=1}^{n} \sum_{i=1}^{n} B_{ki} A_{ik} = \mathrm{Tr}(\mathbf{BA}).$$

□

Using the previous proposition, we obtain

$$\mathrm{Tr}\left(\mathbf{CAC}^{-1}\right) = \mathrm{Tr}\left(\mathbf{AC}^{-1}\mathbf{C}\right) = \mathrm{Tr}(\mathbf{A}). \tag{11.2}$$

As for the determinant, the following definition thus makes sense:

Definition 11.10 (Trace of an endomorphism) Let V be a finite dimensional $\mathbb{K}$-vector space and $g : V \to V$ an endomorphism. We define

$$\mathrm{Tr}(g) = \mathrm{Tr}\left(\mathbf{M}(g, \mathbf{b}, \mathbf{b})\right)$$

where $\mathbf{b}$ is any ordered basis of V. By Theorem 8.26 and (11.2), 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 11.11 *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 $\mathbf{b}$ of V. Then, using Corollary 8.19 and Proposition 11.9, 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 10.1 instead of Proposition 11.9. □

We also have:

Proposition 11.12 *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 6.21. We fix an ordered basis $\mathbf{b}$ 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 10.2 and Proposition 8.20 implies that g is bijective. □

Remark 11.13 Notice that assertions (i)–(iii) of Proposition 11.12 are not equivalent for infinite-dimensional vector spaces (where the determinant doesn't make sense). For instance, consider $V = \mathbb{K}^\infty$, the $\mathbb{K}$-vector space of sequences from Example 4.7; then 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.

11.2 Detour: More on Subspaces

Before we develop the theory of endomorphisms further, we need to make a detour, by developing a bit more theory about *subspaces* of vector spaces.

Sums of subspaces

Definition 11.14 (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 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 4.21, 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 11.15 *The sum of the subspaces $U_i \subset V$, $i = 1 \dots, n$ 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 $v_i, v_i' \in U_i$, $i = 1, \dots, n$. 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, $i = 1, \dots, n$. Thus

$$sv + tv' = sv_1 + tv_1' + \cdots + sv_n + tv_n'$$

is an element of $U_1 + \cdots + U_n$. By Definition 4.16, it follows that $U_1 + \cdots + U_n$ is a vector subspace of V. □

Remark 11.16 Notice that $U_1 + \cdots + U_n$ is the smallest vector subspace of V containing all vector subspaces U_i, $i = 1, \dots, n$.

Direct sums

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 11.17 (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 11.18 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 11.19

(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 11.20 *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. □

Complements

Definition 11.21 (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 11.22 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 11.23 (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 5.10, 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 11.20 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 11.24 *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 5.10, 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. □

Remark 11.25 If you've seen the *Inclusion-Exclusion Principle* for counting the sizes of finite sets, it's tempting to guess that the previous lemma generalises to three or more subspaces as follows:

$$\begin{aligned}\dim(U_1 + U_2 + U_3) &\stackrel{?}{=} \dim(U_1) + \dim(U_2) + \dim(U_3)\\ &\quad - \dim(U_1 \cap U_2) - \dim(U_2 \cap U_3) - \dim(U_3 \cap U_1) + \dim(U_1 \cap U_2 \cap U_3).\end{aligned}$$

This is, surprisingly, *false* – taking U_i to be any three distinct lines through the origin in $\mathbb{R}^2$ gives a counterexample.

11.3 Eigenvectors and eigenvalues

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 11.26 (Eigenvector, eigenspace, eigenvalue) 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* $\mathrm{Eig}_g(\lambda)$ is the subspace of vectors $v \in V$ satisfying $g(v) = \lambda v$.
- The dimension of $\mathrm{Eig}_g(\lambda)$ 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 11.27 By definition, the zero vector 0_V is not an eigenvector, it is however an element of the eigenspace $\mathrm{Eig}_g(\lambda)$ for every eigenvalue λ.

Example 11.28

(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 $\mathrm{P}(\mathbb{R})$ of polynomials and $f = \frac{\mathrm{d}}{\mathrm{d}x} : \mathrm{P}(\mathbb{R}) \to \mathrm{P}(\mathbb{R})$ 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 11.29 *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 β. Then $v \in V$ is an eigenvector of g with eigenvalue $\lambda \in \mathbb{K}$ if and only if $\beta(v) \in \mathbb{K}^n$ 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 8.10, we have $f_{\mathbf{A}} = \beta \circ g \circ \beta^{-1}$. 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 $\vec{x} = \beta(v) \in \mathbb{K}^n$ which is non-zero, since $\beta : V \to \mathbb{K}^n$ is an isomorphism. Then

$$f_{\mathbf{A}}(\vec{x}) = \beta(g(\beta^{-1}(\vec{x}))) = \beta(g(v)) = \beta(\lambda v) = \lambda\beta(v) = \lambda\vec{x},$$

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 $v = \beta^{-1}(\vec{x}) \in V$ is an eigenvector of g with eigenvalue λ.

By Remark 11.4, 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 11.30 (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 11.31

(i) We consider $V = \mathrm{P}(\mathbb{R})$ 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 $\mathrm{P}(\mathbb{R})$ 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 counterclockwise 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 11.32 Applying Proposition 11.29, 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 $U = \mathrm{span}\{v\}$ 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 11.33 (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 11.34 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 11.35

(i) Every eigenspace of an endomorphism $g : V \to V$ is a stable subspace. By definition $g|_{\mathrm{Eig}_g(\lambda)} : \mathrm{Eig}_g(\lambda) \to \mathrm{Eig}_g(\lambda)$ 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 $\mathrm{P}(\mathbb{R})$ of polynomials and $f = \frac{\mathrm{d}}{\mathrm{d}x} : \mathrm{P}(\mathbb{R}) \to \mathrm{P}(\mathbb{R})$ 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 11.36 *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 $\mathbf{b} = (v_1, \dots, v_m)$ for vectors $v_i \in U$ and $\mathbf{b}' = (w_1, \dots, w_{n-m})$ 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 8.11, 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 8.11, $g(v_j) = \sum_{i=1}^{m} \hat{A}_{ij} v_i \in U$ for all $1 \leqslant j \leqslant m$, hence U is stable under g, by Remark 11.34. □

From Proposition 11.36 we can conclude:

Remark 11.37 Suppose V is the direct sum of subspaces $U_1, U_2, \dots, U_m$, all of which are stable under the endomorphism $g : V \to V$. If $\mathbf{b}_i$ is an ordered basis of U_i for $i = 1, \dots, m$, 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 $i = 1, \dots, m$.

11.4 The characteristic polynomial

The eigenvalues of an endomorphism are the solutions of a polynomial equation:

Lemma 11.38 *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

$$g(v) = \lambda v \iff 0_V = (\lambda \mathrm{Id}_V - g)(v) \iff v \in \mathrm{Ker}(\lambda \mathrm{Id}_V - g)$$

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 11.12 implies that $\lambda \in \mathbb{K}$ is an eigenvalue of g if and only if $\det(\lambda \mathrm{Id}_V - g) = 0$. □

Definition 11.39 (Characteristic polynomial) 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 11.40 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 11.41 (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 11.42

(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 $\mathbf{A}$ and hence $\mathbf{A}$ is diagonalisable, c.f. Example 8.15.

(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 $\mathbf{A}$ is not diagonalisable.

The determinant and trace of an endomorphism do appear as coefficients in its characteristic polynomial:

Lemma 11.43 *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 Fix an ordered basis $\mathbf{b}$ of V and let $\mathbf{A} = \mathbf{M}(g, \mathbf{b}, \mathbf{b}) = (A_{ij})_{1 \leqslant i,j \leqslant n}$, so we need to compute $\det(\mathbf{B})$ where $\mathbf{B} = x\mathbf{1}_n - \mathbf{A}$. The result is obvious if $n = 1$, so by induction we may assume it holds for $n - 1$.

Performing a Laplace expansion on the first column of $\mathbf{B}$, we get the following terms:

- A term $(x - A_{11}) \det(\mathbf{B}^{(1,1)})$. Using the induction hypothesis we have $\det \mathbf{B}^{(1,1)} = \mathrm{char}(\mathbf{A}^{(1,1)}) = x^{n-1} - x^{n-2}\,\mathrm{Tr}(\mathbf{A}^{(1,1)}) + \dots$, where the dots denote terms of degree $\leqslant (n-2)$. So we have

  $$\begin{aligned}(x - A_{11}) \det(\mathbf{B}^{(1,1)}) &= (x - A_{11})(x^{n-1} - x^{n-2}\,\mathrm{Tr}(\mathbf{A}^{(1,1)}) + \dots) \\ &= x^n - (A_{11} + \mathrm{Tr}(\mathbf{A}^{(1,1)}))x^{n-1} + \dots .\end{aligned}$$

  Since the diagonal entries of $\mathbf{A}^{(1,1)}$ are precisely the diagonal entries of A with $A_{1,1}$ removed, we have $A_{11} + \mathrm{Tr}(\mathbf{A}^{(1,1)}) = \mathrm{Tr}(\mathbf{A})$, so this is just $x^n - \mathrm{Tr}(\mathbf{A})x^{n-1} + \dots$.
- Another $(n-1)$ terms of the form $(-1)^{k+1} \cdot (-A_{k1}) \cdot \det(\mathbf{B}^{(k,1)})$ for $2 \leqslant k \leqslant n$. We claim that each of these is a polynomial of degree $\leqslant n - 2$. This is because crossing out the k-th row and first column, for $k \neq 1$, removes 2 of the terms with x in them from the diagonal of $\mathbf{B}$. Hence $\mathbf{B}^{(k,1)}$ has $n-2$ entries which are linear in x and all the rest are constant; so each of the terms in the Leibniz formula for $\det \mathbf{B}^{(k,1)}$ has degree at most $n - 2$.

So the Laplace-expansion terms with $2 \leqslant k \leqslant n$ don't contribute anything to the x^n and x^{n-1} coefficients of $\det(\mathbf{B})$ and we conclude

$$\mathrm{char}_g(x) = x^n - \mathrm{Tr}(g)x^{n-1} + c_{n-2}x^{n-2} + \cdots + c_1 x + c_0$$

for coefficients $c_{n-2}, \dots, c_0 \in \mathbb{K}$. It remains to show that $c_0 = (-1)^n \det(g)$. We have $c_0 = \mathrm{char}_g(0) = \det(-g) = \det(-\mathbf{A})$. Since the determinant is linear in each row of $\mathbf{A}$, this gives $\det(-\mathbf{A}) = (-1)^n \det(\mathbf{A})$, as claimed. □

Remark 11.44 In particular, for $n = 2$ we have $\mathrm{char}_g(x) = x^2 - \mathrm{Tr}(g)x + \det(g)$. Compare with Example 11.42.

Exercises

Exercise 11.1 Let U and U' be the subspaces of $\mathbb{R}^3$ with bases

$$\left\{ \begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}1\\2\\2\end{pmatrix} \right\} \quad \text{and} \quad \left\{ \begin{pmatrix}1\\2\\3\end{pmatrix} \right\}$$

respectively. Show that U and U' are complementary subspaces of $\mathbb{R}^3$. Find another subspace U'', with $U'' \neq U'$, such that U'' is also complementary to U.

Exercise 11.2 Let $\mathbf{A} \in M_{n,n}(\mathbb{K})$. Show that the map $t_{\mathbf{A}} : M_{n,n}(\mathbb{K}) \to \mathbb{K}$ defined by $t_{\mathbf{A}}(\mathbf{B}) = \mathrm{tr}(\mathbf{AB})$ is linear. Show, conversely, that given any linear map $t : M_{n,n}(\mathbb{K}) \to \mathbb{K}$, we can find a unique $\mathbf{A}$ such that $t = t_{\mathbf{A}}$.

Exercise 11.3 Compute the characteristic polynomials and eigenvalues of the following matrices, and bases of their eigenspaces:

$$(i)\ \begin{pmatrix} 7 & -4 \\ -8 & -7 \end{pmatrix}, \quad (ii)\ \begin{pmatrix} 5 & 2 & 3 \\ -13 & -6 & -11 \\ 4 & 2 & 4 \end{pmatrix}, \quad (iii)\ \begin{pmatrix} 3 & 1 & 1 \\ -15 & -5 & -5 \\ 6 & 2 & 2 \end{pmatrix}.$$

Which of them are diagonalisable?

Exercise 11.4 *(hard!)*
Show that the coefficient of x^{n-2} in $\mathrm{char}_{\mathbf{A}}(x)$, for $\mathbf{A} \in M_{n,n}(\mathbb{K})$, is given by the sum

$$\sum_{1 \leqslant i < j \leqslant n} \det \begin{pmatrix} A_{ii} & A_{ij} \\ A_{ji} & A_{jj} \end{pmatrix}$$

(the sum over all "second-order diagonal minors" of A). Can you spot a generalisation to the x^{n-r} coefficient for an arbitrary r?

CHAPTER 12

Endomorphisms, II

Contents

12.1 Properties of eigenvalues 121
12.2 Special endomorphisms 126
Involutions 126
Projections 127
Exercises 127

12.1 Properties of eigenvalues

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 12.1 (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 $(a_0, \dots, a_n) \in \mathbb{K}^{n+1}$. For $0 \leqslant j \leqslant n-1$ consider

$$b_j = \sum_{k=0}^{n-j-1} a_{k+j+1} x_0^k \tag{12.1}$$

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 12.2 *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 $\lambda_1, \dots, \lambda_{n+1}$. Since $f(\lambda_{n+1}) = 0$, Theorem 12.1 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 12.3 *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 11.38 and Lemma 11.43, 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 12.2. □

Proposition 12.4 (Linear independence of eigenvectors) *Let V be a finite dimensional $\mathbb{K}$-vector space and $g : V \to V$ an endomorphism. Then the eigenspaces $\mathrm{Eig}_g(\lambda)$ 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 $v_i, v_i' \in \mathrm{Eig}_g(\lambda_i)$ be eigenvectors such that

$$v_1 + v_2 + \cdots + v_m = v_1' + v_2' + \cdots + v_{\tilde{m}}. \tag{12.2}$$

Applying g to this last equation gives

$$\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_m v_m = \lambda_1 v_1' + \lambda_2 v_2' + \cdots + \lambda_m v_{\tilde{m}}. \tag{12.3}$$

Subtracting λ_m times (12.2) from (12.3) 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 (12.3) 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 11.19 (ii). □

In the case where all the eigenvalues are distinct, we conclude that g is diagonalisable.

Proposition 12.5 *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 $0_V \neq v_i \in \mathrm{Eig}_g(\lambda_i)$ for $i = 1, \dots, n$. Then, by Proposition 12.4, 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 12.6 Proposition 12.5 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 $\mathrm{P}(\mathbb{C})$ 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 12.7 (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 11.38 and Lemma 11.43, the eigenvalues of g are the zeros of the characteristic polynomial and this is an element of $\mathrm{P}(\mathbb{C})$. The first statement thus follows by applying the fundamental theorem of algebra to the characteristic polynomial of g.

Applying Theorem 12.1 and the fundamental theorem of algebra repeatedly, we find $k \in \mathbb{N}$ and multiplicities $m_1, \dots, m_k \in \mathbb{N}$ such that

$$\mathrm{char}_g(x) = (x - \lambda_1)^{m_1}(x - \lambda_2)^{m_2} \cdots (x - \lambda_k)^{m_k}$$

where $\lambda_1, \dots, \lambda_k$ are zeros of char_g. Since char_g has degree n, it follows that $\sum_{i=1}^{k} m_i = n$. □

Example 12.8

(i) Recall that the *discriminant* of a quadratic polynomial $x \mapsto ax^2 + bx + c \in \mathrm{P}(\mathbb{K})$ 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 10.4, 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

$$\operatorname{char}_{\mathbf{A}}(x) = \prod_{i=1}^{n} (x - A_{ii}) = (x - A_{11})(x - A_{22}) \cdots (x - A_{nn}).$$

Consequently, an upper triangular matrix has spectrum $\{A_{11}, A_{22}, \dots, A_{nn}\}$ and is diagonalisable if all its diagonal entries are distinct. Notice that by Example 11.42 (ii) not every upper triangular matrix is diagonalisable.

Example 12.9 (Fibonacci sequences) The Fibonacci sequence is the sequence $(F_n)_{n \in \mathbb{N}}$ defined by the relations

$$F_0 = 0, \quad F_1 = 1, \quad F_n = F_{n-1} + F_{n-2} \text{ for } n \geqslant 2.$$

Consider the matrix

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} F_0 & F_1 \\ F_1 & F_2 \end{pmatrix}.$$

Then, using induction, we can show that

$$\mathbf{A}^n = \begin{pmatrix} F_{n-1} & F_n \\ F_n & F_{n+1} \end{pmatrix}$$

for all $n \in \mathbb{N}$. We would like to give a simple formula for computing $\mathbf{A}^n$, and hence the Fibonacci numbers F_n.
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

$$\mathbf{A}^n = \frac{1}{\lambda_2 - \lambda_1}\begin{pmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{pmatrix}\begin{pmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{pmatrix}\begin{pmatrix} \lambda_2 & -1 \\ -\lambda_1 & 1 \end{pmatrix} = \begin{pmatrix} F_{n-1} & F_n \\ F_n & F_{n+1} \end{pmatrix}.$$

This yields the formula

$$F_n = \frac{\lambda_1^n - \lambda_2^n}{\lambda_1 - \lambda_2} = \frac{((1+\sqrt{5})/2)^n - ((1-\sqrt{5})/2)^n}{\sqrt{5}}.$$

Proposition 12.10 *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 $\dim \mathrm{Eig}_g(\lambda) = m$ and $\mathbf{b}$ be an ordered basis of $\mathrm{Eig}_g(\lambda)$. 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 $\mathrm{Eig}_g(\lambda)$ is stable under g and

$$\mathbf{M}(g|_{\mathrm{Eig}_g(\lambda)}, \mathbf{b}, \mathbf{b}) = \lambda \mathbf{1}_m.$$

By Proposition 11.36, the matrix representation of g with respect to the basis $\mathbf{c}$ takes the form

$$\mathbf{M}(g, \mathbf{c}, \mathbf{c}) = \begin{pmatrix} \lambda \mathbf{1}_m & * \\ \mathbf{0}_{n-m,m} & \mathbf{B} \end{pmatrix}$$

for some matrix $\mathbf{B} \in M_{n-m,n-m}(\mathbb{K})$. We thus obtain

$$\mathrm{char}_g(x) = \det\begin{pmatrix} (x-\lambda)\mathbf{1}_m & * \\ \mathbf{0}_{n-m,m} & x\mathbf{1}_{n-m} - \mathbf{B} \end{pmatrix}$$

Applying the Laplace expansion (9.5) with respect to the first column, we have

$$\mathrm{char}_g(x) = (x-\lambda)\det\begin{pmatrix} (x-\lambda)\mathbf{1}_{m-1} & * \\ \mathbf{0}_{n-m,m-1} & x\mathbf{1}_{n-m} - \mathbf{B} \end{pmatrix}$$

Applying the Laplace expansion again with respect to the first column, m-times in total, we get

$$\mathrm{char}_g(x) = (x-\lambda)^m \det(x\mathbf{1}_{n-m} - \mathbf{B}) = (x-\lambda)^m \mathrm{char}_{\mathbf{B}}(x).$$

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 $(v_1, \dots, v_n)$ of V consisting of eigenvectors of g. Therefore,

$$\mathrm{char}_g(x) = \prod_{i=1}^{n}(x - \lambda_i)$$

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 $\mathrm{Eig}_g(\lambda_j)$. 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 12.7, the sum of the algebraic multiplicities is n. The sum of the geometric multiplicities is by assumption also n. Since, by Proposition 12.4, the eigenspaces with respect to different eigenvalues are in direct sum, we obtain a basis of V consisting of eigenvectors of g. □

12.2 Special endomorphisms

Involutions

A mapping $\iota : \mathcal{X} \to \mathcal{X}$ from a set $\mathcal{X}$ into itself is called an *involution*, if $\iota \circ \iota = \mathrm{Id}_{\mathcal{X}}$. In the case where $\mathcal{X}$ is a vector space and ι is linear, then ι is called a *linear involution*.

Example 12.11 (Involutions)

(i) Let V be a $\mathbb{K}$-vector space. Then the identity mapping $\mathrm{Id}_V : V \to V$ 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 $\mathrm{F}(I, \mathbb{K})$, the $\mathbb{K}$-vector space of functions on the interval $I \subset \mathbb{R}$, we obtain a linear involution of $\mathrm{F}(V, \mathbb{K})$ 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.

In the rest of this section we suppose $2 \neq 0$ in $\mathbb{K}$.

Proposition 12.12 *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 $V = \mathrm{Eig}_\iota(1) \oplus \mathrm{Eig}_\iota(-1)$ 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 12.4, the eigenspaces $\mathrm{Eig}_\iota(1)$ and $\mathrm{Eig}_\iota(-1)$ are in direct sum (interpreting $\mathrm{Eig}_\iota(1)$ as $\{0\}$ if 1 is not an eigenvalue, and similarly for -1). What we need to show is that their sum is all of V. For $v \in V$ we write

$$v = \frac{1}{2}(v + \iota(v)) + \frac{1}{2}(v - \iota(v)).$$

We claim that $\frac{1}{2}(v + \iota(v)) \in \mathrm{Eig}_\iota(1)$, and $\frac{1}{2}(v - \iota(v)) \in \mathrm{Eig}_\iota(-1)$.

For the first half of the claim, we compute

$$\iota\left(\frac{1}{2}(v + \iota(v))\right) = \frac{1}{2}(\iota(v) + \iota(\iota(v))) = \frac{1}{2}(\iota(v) + v).$$

The second half of the claim is similar.

It follows that the sum of $\mathrm{Eig}_\iota(1)$ and $\mathrm{Eig}_\iota(-1)$ is V, so ι is diagonalisable as required. □

Projections

A linear mapping $\Pi : V \to V$ satisfying $\Pi \circ \Pi = \Pi$ is called a *projection*.

Example 12.13 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 $\{\vec{x} \in \mathbb{R}^3 | x_3 = 0\}$.

Similar to Proposition 12.12, we obtain:

Proposition 12.14 *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 $V = \mathrm{Eig}_\Pi(0) \oplus \mathrm{Eig}_\Pi(1)$, Π is diagonalisable and $\mathrm{Im}\,\Pi = \mathrm{Eig}_\Pi(1)$.*

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 $\lambda = 1$.

Using an argument similar to Proposition 12.12, one can show that $V = \mathrm{Ker}\,\Pi + \mathrm{Im}\,\Pi$. Since $\mathrm{Ker}\,\Pi = \mathrm{Eig}_\Pi(0)$, the theorem will follow if we can show that $\mathrm{Im}\,\Pi = \mathrm{Eig}_\Pi(1)$. Let $v \in \mathrm{Im}\,\Pi$ 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 $v \in \mathrm{Im}\,\Pi$. We thus conclude that $\mathrm{Im}\,\Pi = \mathrm{Eig}_\Pi(1)$. Choosing an ordered basis of $\mathrm{Ker}\,\Pi$ and an ordered basis of $\mathrm{Im}\,\Pi$ gives a basis of V consisting of eigenvectors, hence Π is diagonalisable. □

Remark 12.15 In a sense there is only one kind of projection. It follows from Proposition 12.14 that for a projection $\Pi : V \to V$, we have $V = \mathrm{Ker}\,\Pi \oplus \mathrm{Im}\,\Pi$. Conversely, 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 $f \mapsto (\mathrm{Ker}\, f, \mathrm{Im}\, f)$, and one can check that this is a bijection.

Exercises

Exercise 12.1 Derive the formula (12.1) for the coefficients b_j.

Exercise 12.2 Consider the matrix

$$\mathbf{R}_\alpha = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$$

for $\alpha \in (0, \pi)$, as in Example 11.31. Show that this matrix *is* diagonalisable over $\mathbb{C}$, and find its eigenvalues and eigenvectors.

Exercise 12.3 Show that the map Λ of Remark 12.15 is a bijection.

Exercise 12.4 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.

Exercise 12.5 Let $\mathbf{A}, \mathbf{B} \in M_{n,n}(\mathbb{K})$ be matrices which commute with each other (i.e. $\mathbf{AB} = \mathbf{BA}$). Show that each eigenspace of $\mathbf{A}$ is stable under $\mathbf{B}$, and vice versa. Hence show that if $\mathbf{A}$ is diagonalisable with distinct eigenvalues, any matrix which commutes with $\mathbf{A}$ is also diagonalisable.

CHAPTER 13

Affine spaces and quotient vector spaces

Contents

13.1 Affine mappings and affine spaces 129
13.2 Quotient vector spaces 130
Exercises 132

13.1 Affine mappings and affine spaces

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 13.1 (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 13.2 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 (6.1), 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 13.3 (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 13.4 *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 6.6. 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 13.5 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 7.4 and Proposition 13.4, 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 13.6 (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 13.7 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 $y = x$.

13.2 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

$$(U + v_1) +_{V/U} (U + v_2) = U + (v_1 + v_2) \tag{13.1}$$

for all $v_1, v_2 \in V$ as well as

$$s \cdot_{V/U} (U + v) = U + (sv) \tag{13.2}$$

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 13.8 *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 (13.1) and (13.2) are well defined. We start with (13.1). 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 13.8 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 (13.2) 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 13.8 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 (13.1) and scalar multiplication $\cdot_{V/U}$ defined by (13.2), 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 4.2 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 13.9 (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 13.10 *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 6.20 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 13.11 (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 13.1 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 either an affine subspace, or is empty (cf. Proposition 6.10).

Exercise 13.2 Show that in $P_2(\mathbb{R})$, the set of polynomials f with $f(1) = 1$ is an affine subspace. What is the associated vector subspace?

Exercise 13.3 Let A be an affine subspace of a $\mathbb{K}$-vector space V. Show that for all integers $n \geqslant 1$, all $v_1, \dots, v_n \in A$, and all scalars $s_1, \dots, s_n \in \mathbb{K}$ with $\sum_i s_i = 1$, we have $\sum_i s_i v_i \in A$.
Show, conversely, that if A is a non-empty subset of V with this property, then A is an affine subspace.
(Hint: Choose an $a_0 \in A$ and show that $U = \{a - a_0 : a \in A\}$ is a vector subspace.)

Exercise 13.4 *(hard!)* Let $f : \mathbb{Q}^2 \to \mathbb{Q}^2$ be a map which preserves *co-linearity*, i.e. if P, Q, R lie on a straight line then $f(P), f(Q), f(R)$ lie on a straight line. Show that f is an affine map.

Exercise 13.5 Let P be the space of all polynomial functions $\mathbb{R} \to \mathbb{R}$. Define $W = \{p \in P : p(0) = p(1)\}$ and $V = \{p \in P : p(0) = p(1) = 0\}$. What is the dimension of the quotient spaces P/W and P/V?

Exercise 13.6 Let $f : V \to V'$ be a linear map, and $W \subset V$, $W' \subset V'$ vector subspaces such that $f(W) \subset W'$.

(i) Show that there is a uniquely determined linear map $\bar{f} : V/W \to V'/W'$ satisfying
$$f(v + W) = f(v) + W'$$
for all $v \in V$.

(ii) Show that if $V = V'$ and $W = W'$, then we have
$$\det f = \det(\bar{f}) \det(f|_W),$$
where $f|_W : W \to W$ denotes the restriction of f to W.