Linear Algebra 2

Understanding linear maps

$$A = S \begin{bmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & \mu \end{bmatrix} S^{-1}$$

Kathrin Hövelmanns

Last updated on November 9, 2023

Faculty of Mathematics and Computer Science

Eindhoven University of Technology

Preface

Linear Algebra 1 (2MBA20) introduced you to vector spaces and maps that respect (or 'play nicely with') their linear structures (hence called linear maps). LinA 2 is a sequel to LinA1 that takes a deeper dive into the area surrounding linear maps: We will discuss how linear maps can be described and dealt with by representing them via matrices, and how the nature of a linear map can be determined by simply looking at such a matrix representation. (For instance, how do you recognize that a matrix represents a reflection?)

This will involve techniques to

- switch between different vector bases; and to
- find a basis of particularly useful vectors (called eigenvectors) that depend on characteristic properties of the matrix (called eigenvalues).

Eigenvectors are so useful because they allow to describe linear maps in a simpler and more economical way. We will use this to study interesting special classes of maps like orthogonal maps (maps that 'preserve' shapes, like reflections and rotations) and symmetric maps (which can be used to analyse quadratic curves and surfaces like ellipses and hyperboloids).

These lecture notes aim to focus on the content of the lecture and not wander too far beyond what you are expected to learn during this course.

How does this relate to other courses, and why do we care about linear maps? Linear maps help to understand complicated structures. In many numerical methods, an important first step is to bring a matrix into a simpler form (e.g., diagonal or triangular), which often uses the techniques developed in this course. You might encounter this in the two courses *Introduction to Numerical Analysis* (2MBC10) or *Numerical Linear Algebra* (2MBC30).

Here are some other (hopefully interesting) examples and use cases:

- ***Geometric example - quadratic curves.*** Imagine you are handed the description of a geometric object (like a curve or a surface), and you are promised a cookie if you manage to determine the object's geometric properties. Unfortunately, however, the only information you got is a weird-looking equation. In lucky cases like, e.g., quadratic curves and surfaces, there exist recipes to get your cookie: we will see how certain tricks developed in chapter 2 can be applied to end up with a nice description of the object from which you can easily extract the most important geometric properties.
- ***Evolutions of physical systems - differential equations.*** Originally, the concept of Eigenvalues was devised in the context of dealing with systems of linear differential equations. It

hence comes as no surprise that the course *Theory and Application of Ordinary Differential Equations* (2MBC20) will draw heavily on the techniques and the language developed in this course. If time permits, we'll (briefly) introduce this topic in the lecture.

- ***Quantum mechanics.*** In quantum mechanics, linear maps play a mayor role: physical quantities (like velocity and momentum) are described by linear maps in vector spaces of functions. In this context, eigenvalues/vectors have a special physical interpretation.

- ***Analysis 2, tensor calculus and differential geometry.*** In Analysis 2, linear maps appear as first–order approximations of differentiable multi-variable functions: Say $f : \mathbb{R}^n \to \mathbb{R}^n$ is such a function and $\underline{p} \in \mathbb{R}^n$ is a vector. Then the matrix of this first–order approximation in $\underline{p}$ is the $n \times n$ matrix whose (i, j)-th element is $\dfrac{\partial f_i}{\partial x_j}$. This is also useful to study curved spaces.

- ***A more general view - algebra.*** Vector spaces are only one example for sets with mathematical structures. We already know other examples like fields, but there are many more. (E.g., rings, vector-space-like structures over such rings, ...) Studying maps that play nicely with the given structure will prove very useful also in these other examples. This theme will play a huge role in the algebra courses.

Changelog for students who are revisiting this course: These lecture are based on the ones by Hans Sterk who taught this course before I became responsible lecturer (and the Bachelor curriculum got redesigned). The main changes are that I

- assume the basic knowledge about linear maps that is now covered in *Linear Algebra I*;

- added some details about a certain way to simplify matrices, called the Jordan Normal Form;

- put less emphasis on how our techniques can be used to deal with linear differential equations (this will be covered in *Theory and Application of Ordinary Differential Equations*);

- do not cover dual spaces (these were previously covered in the lecture notes, but to my knowledge not in the lecture itself); and

- do not cover certain matrix decompositions (LU-decomposition, Singular Value Decomposition, this will be covered in Numerical courses).

For students that are curious about the removed/de-emphasised parts, I will also provide Hans' previous lecture notes on Canvas.

As I will try to improve my lecture notes with each year,

any constructive feedback is more than welcome!

Contents

1 Setting the stage: fundamental definitions/concepts for linear maps **2**

What happens in chapter 1? 2

1.1 Recap: Linear maps (as seen in *Linear Algebra 1*), plus another nice example 3

1.1.1 The new example for linear maps: the quotient map 8

1.2 Recap: Connecting matrices and linear maps (as seen in *Linear Algebra 1*) 10

1.2.1 Computing a matrix inverse via row reduction 13

1.3 Representation matrices for finite-dimensional spaces $\neq \mathbb{K}^n$, using coordinates 14

1.3.1 Coordinates 15

1.3.2 Coordinate transformations from a matrix POV: The basis transition matrix . . 16

1.3.3 Generalising the map-matrix connection for spaces that aren't $\mathbb{K}^n$ 19

1.3.4 How do base changes affect matrices of linear maps? 22

2 'Simpler descriptions' for maps/matrices **25**

What happens in chapter 2? 25

2.1 Diagonalisation via eigenvalues 26

2.1.1 Computing eigenvalues and -spaces 28

2.1.2 Linear independence of eigenvectors 36

2.1.3 Diagonalisation of a square matrix 37

2.2 Using subspaces that are stable under the mapping (invariant subspaces) 38

2.2.1 Nice results for combinations of invariant subspaces 42

2.2.2 How to get invariant subspaces from non-real roots 44

2.3 Diagonalisability and dealing with incomplete eigenvector bases 45

3 Neat tricks for special linear maps: Orthogonal and symmetric maps **46**

4 Another application example: linear differential equations **47**

A Notation used in this course **48**

A.1 Table of frequently used notation . . . 48

A.2 Sets and maps . . . 49

A.2.1 Sets . . . 49

A.2.2 Maps . . . 49

B Prerequisites: Vector spaces as seen in *Linear Algebra 1* **51**

B.1 Fields and complex numbers . . . 51

B.2 Vectors . . . 54

B.2.1 Quotient spaces . . . 56

B.3 Matrices . . . 57

C Additional helpers **58**

C.1 Identifying Greek letters . . . 58

C.2 Geometric shapes: ellipse, hyperbola, parabola . . . 59

C.3 Trigonometric formulas . . . 60

Chapter 1

Setting the stage: fundamental definitions/concepts for linear maps

What happens in chapter 1?

⚠ ***This course assumes familiarity with the knowledge on linear maps that was introduced in Linear Algebra I.***

To jog your memory (and simplify referencing), this chapter starts by summarising the most important concepts related to linear maps in section 1.1. Something that is very important (and useful!) is the close connection between matrices and linear maps, recapped in section 1.2.

If you're very comfortable with the material covered in *Linear Algebra 1*, you can skip sections 1.1 and 1.2 ***except for the following two parts***:

- a new example (example 1.1.1) that mixes linear maps with quotient spaces; and
- a new remark (remark 1.2.4) that translates our result on particular solutions for vector equations (theorem 1.1.14) to the setting where we deal with systems of linear equations.

After these recaps, we begin tackling new material in section 1.3: we dig deeper into the connection between matrices and linear maps.

Learning Goals of Chapter 1: You can

- explain and work with the concept of a linear map between vectors spaces
  - for concrete vector spaces like $\mathbb{R}^n$, $\mathbb{C}^n$, $\mathbb{K}^n$ as well as
  - for vector spaces that are handled as abstract objects;
- reason about the relation between linear maps and matrices;
- set up matrix representations of linear maps;
- analyse how the coordinates of a vector change when switching to another basis;

- perform basis transformations to switch between representations for different bases and compute matrices that represent such a basis transition.

Why do we care? The techniques developed in this chapter lay the groundwork for the rest of this course: in **??**, our main goal will be to develop tricks to transform matrices into a 'nicer' (simpler) shape. On a high level, the main trick will be to find a certain basis for which the matrix transforms into something nice. This will use the observations we made in this section.

1.1. Recap: Linear maps (as seen in *Linear Algebra 1*), plus another nice example

In this recap section, we recall the most fundamental concept of this course, *linear maps*. Besides recapping the notion itself, this section covers:

- that linear maps can be composed, added, multiplied (with each other and with scalars), and sometimes also inverted;
- the null space $\mathcal{N}$ and the range $\mathcal{R}$ of a linear map;
- how $\mathcal{N}$ and $\mathcal{R}$ relate to injectivity and surjectivity;
- how to specify a linear map on a given basis; and
- (very important) the dimension theorem.

This section requires familiarity with the following concepts which you can look up in the appendices when needed: fields (see definition B.1.1), vector spaces over fields (see definition B.2.1), bases (see definition B.2.5), inner product spaces (see definition B.2.7), matrix multiplication (see definition B.3.3), and basic notions concerning maps like injectivity/surjectivity (see the table in appendix A.2).

Let V and W be $\mathbb{K}$-vector spaces for some field $\mathbb{K}$. (E.g., think of real vector spaces, meaning $\mathbb{K}=\mathbb{R}$, or complex vector spaces, meaning $\mathbb{K}=\mathbb{C}$.) A map $\mathcal{A}: V \to W$ associates to *each* vector $\underline{v}$ in V exactly one vector $\mathcal{A}(\underline{v})$ (or short: $\mathcal{A}\underline{v}$) in W.

Definition 1.1.1 (Linear map). *Let V and W be two $\mathbb{K}$-vector spaces. A map $\mathcal{A}: V \to W$ is called* $\mathbb{K}$**-linear** *(or simply* **linear***) if for all vectors $\underline{x}, \underline{y} \in V$ and all field elements λ ('scalars') the following holds:*

i) $\mathcal{A}(\underline{0}) = \underline{0}$ — *0 is mapped to 0*

ii) $\mathcal{A}(\underline{x}+\underline{y}) = \mathcal{A}\underline{x} + \mathcal{A}\underline{y}$ — *addition works linearly*

iii) $\mathcal{A}(\lambda\underline{x}) = \lambda\mathcal{A}\underline{x}$ — *scalar multiplication works linearly*

A bijective linear map is also called an **isomorphism**.

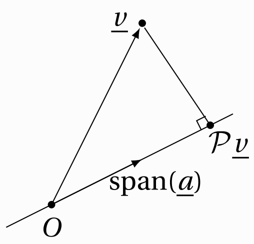
Equivalently, linearity can be defined through the following single requirement: $\mathcal{A}$ is linear if for all $\underline{x}, \underline{y} \in V$ and all scalars $\alpha, \beta \in \mathbb{K}$, one has

$$\mathcal{A}(\alpha\underline{x} + \beta\underline{y}) = \alpha\mathcal{A}\underline{x} + \beta\mathcal{A}\underline{y}\,.$$

This means that for linear maps, the image of a linear combination of two vectors is the same as the linear combination of their image vectors. In other words, ***we can switch the order of linearly combining and applying*** $\mathcal{A}$. (We will use this in this course very, very often.)

Example 1.1.2 ***(Orthogonal projections)*** Let V be a real inner product space, and let $l = \text{span}(\underline{a})$ be a line through the origin in V. The map that associates to each vector in V the orthogonal projection on l we will call $\mathcal{P}$. If $\underline{a}$ has length one, then $\mathcal{P}$ is given by the formula:

$$\mathcal{P}\underline{v} = (\underline{v}, \underline{a})\underline{a} \quad .$$



More generally, we saw that ***orthogonal projections on general linear subspaces of a real inner product space are linear***.

Example 1.1.3 ***(Trivial examples)*** For every vector space V, we have the so-called ***identity map*** (or just *identity*) $\mathcal{I} : V \to V$ given by $\mathcal{I}\underline{v} := \underline{v}$. It is very straightforward to verify that $\mathcal{I}$ is linear.
When V and W are two vector spaces that are either both real or both complex, then the so-called ***zero map*** $\mathcal{O} : V \to W$ given by $\mathcal{O}\underline{v} := \underline{0}$ is also linear.

Example 1.1.4 ***(Multiplication with a matrix)*** Let A be a real $m \times n$-matrix with entries in $\mathbb{R}$. In this example, we treat elements from $\mathbb{R}^n$ and $\mathbb{R}^m$ as columns. We defined the map $L_A : \mathbb{R}^n \to \mathbb{R}^m$ by

$$L_A(\underline{v}) := A\underline{v} \quad ,$$

and saw that L_A is linear.
Note that the argument you saw in LinA I would have worked for any other base field $\mathbb{K}$ just the same, meaning if the matrix A has entries in an arbitrary field $\mathbb{K}$ instead of $\mathbb{R}$, then the map

$$\mathbb{K}^n \xrightarrow{\underline{v} \mapsto A\underline{v}} \mathbb{K}^m \quad ,$$

is also linear.

Example 1.1.5 ***(Differentiation is a linear map.)*** The derivative D is defined by $f' = Df$. We saw that D behaves in a linear manner.
⚠ Remember, however, that we always need to be specific concerning the question on which vector space D is acting.

Now that we've recalled some examples, we recall another very useful observation: we learnt that switching the order of linearly combining and using $\mathcal{A}$ does work for any number of vectors:

Theorem 1.1.6. *$\mathcal{A}: V \to W$ is linear if and only if*

$$\mathcal{A}\left(\sum_{i=1}^{n} \alpha_i \underline{v}_i\right) = \sum_{i=1}^{n} \alpha_i \mathcal{A}\underline{v} \tag{1.1}$$

for all vectors $\underline{v}_1, \ldots, \underline{v}_n$ in V and all scalars $\alpha_1, \ldots, \alpha_n \in \mathbb{K}$.

We will also recall that compositions, inverses, sums and scalar multiples of linear maps are again linear maps.

Definition/theorem 1.1.7 (Composition or product). *Let $\mathcal{A}: V \to W$ and $\mathcal{B}: U \to V$ be linear maps. Take a vector $\underline{u} \in U$. First applying $\mathcal{B}$ to $\underline{u} \in U$, we get $\mathcal{B}\underline{u} \in V$. Now applying $\mathcal{A}$ to $\mathcal{B}\underline{u}$, we get $\mathcal{A}(\mathcal{B})\underline{u} \in W$. In a diagram:*

$$\begin{array}{ccccc} U & \xrightarrow{\mathcal{B}} & V & \xrightarrow{\mathcal{A}} & W \\ \underline{u} & \longrightarrow & \mathcal{B}\underline{u} & \longrightarrow & \mathcal{A}(\mathcal{B}\underline{u}) \end{array} \tag{1.2}$$

The **composition** *$\mathcal{A}\,\mathcal{B}: U \to W$ is defined by*

$$(\mathcal{A}\mathcal{B})\underline{u} := \mathcal{A}(\mathcal{B}\underline{u}) \text{ for all } \underline{u} \in U\,.$$

It is itself linear.

⚠ **Cave:** The notation $\mathcal{A}\mathcal{B}$ suggests that the composition can be understood as a 'product' of $\mathcal{A}$ and $\mathcal{B}$. Remember, however, that ***this is not a product in the sense as we know it, e.g., from fields***: Say U, the 'target space' of $\mathcal{B}$, also is the 'starting space' of $\mathcal{A}$, meaning we can indeed compose $\mathcal{A}$ and $\mathcal{B}$ into $\mathcal{A}\mathcal{B}$. This does not mean that $\mathcal{B}\mathcal{A}$ exists - we have no guarantee that the 'target space' of $\mathcal{A}$ also is the 'starting space' of $\mathcal{B}$. And even if $\mathcal{B}\mathcal{A}$ also exists, it is not necessarily equal to $\mathcal{A}\mathcal{B}$.

Definition/theorem 1.1.8 (Inverse map). *If the linear map $\mathcal{A}: V \to W$ is a bijection, then the* **inverse map**

$$\mathcal{A}^{-1}: W \xrightarrow{w \mapsto v \text{ s. th. } Av = w} V$$

is also linear.

Definition 1.1.9 (Powers of endomorphisms). *For $\mathcal{A}: V \to V$, we define $\mathcal{A}^2 := \mathcal{A}\mathcal{A}$. More generally, for $n = 2, 3, \ldots$, we define $\mathcal{A}^n := \mathcal{A}^{n-1}\mathcal{A}$. We take for $\mathcal{A}^0$ the identity map $\mathcal{I}$. If $\mathcal{A}$ is invertible (so if the map has an inverse $\mathcal{A}^{-1}$), then we define for positive integers n the map $\mathcal{A}^{-n}$ as the composite map $\left(\mathcal{A}^{-1}\right)^n$.*

Definition/theorem 1.1.10 (Sum, scalar multiple). *Let $\mathcal{A}: V \to W$ and $\mathcal{B}: V \to W$ be two linear maps. Then the* **sum** *$\mathcal{A} + \mathcal{B}: V \to W$ is defined by*

$$(\mathcal{A} + \mathcal{B})\underline{v} := \mathcal{A}\underline{v} + \mathcal{B}\underline{v}\,.$$

If α is a scalar, then the **scalar multiple** $\alpha\mathcal{A} : V \to W$ *is defined by*

$$(\alpha\mathcal{A})\underline{v} := \alpha(\mathcal{A}\underline{v}) \ .$$

Both constructions are linear.

Since linear maps are also ordinary maps, we can talk about the image of a vector or a subset of the vector space (notation: $\mathcal{A}(D)$ if $\mathcal{A}$ is the linear map and D the subset), and about the (complete) inverse image of a subset.

We recall two very important linear subspaces that are associated to linear maps: the first is a generalization of the solution space of a homogeneous system of linear equations, the second is a generalization of the column space of a matrix.

Definition/theorem 1.1.11 (Null space and range). *Let $\mathcal{A} : V \to W$ be a linear map. We define $\mathcal{N}$, the* **null space** *(or* **kernel***) of $\mathcal{A}$, by*

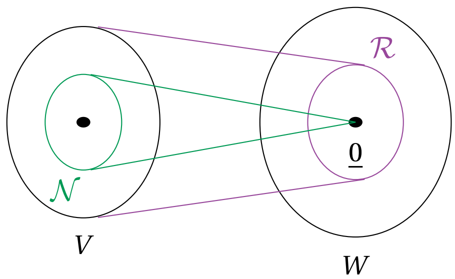
$$\mathcal{N}(\mathcal{A}) := \{\underline{v} \in V \mid \mathcal{A}\underline{v} = \underline{0}\} \quad ,$$

and $\mathcal{R}$, the **range** *of $\mathcal{A}$, by*

$$\mathcal{R}(\mathcal{A}) := \{\mathcal{A}\underline{v} \in W \mid \underline{v} \in V\} \ .$$

The range can also be denoted by $\mathcal{A}(V)$. If the context is clear, we will simply write $\mathcal{N}/\mathcal{R}$ instead of $\mathcal{N}(\mathcal{A})/\mathcal{R}(\mathcal{A})$.

$\mathcal{N}$ is a linear subspace of V and $\mathcal{R}$ is a linear subspace of W.



Notice that the null space is precisely the inverse image $\mathcal{A}^{-1}\{\underline{0}\}$ of the origin under $\mathcal{A}$.

Example 1.1.12 ***(Multiplication with a matrix)*** We first revisit example 1.1.4: The null space of the linear map L_A consists of the vectors $\underline{v}$ that satisfy $L_A(\underline{v}) = \underline{0}$, so all solutions of the homogeneous system $A\underline{v} = \underline{0}$.
The range of L_A consists of all vectors of the form $L_A(\underline{v})$. If the columns of A are $\underline{a}_1, \dots, \underline{a}_n$, then this is the set

$$\{x_1\underline{a}_1 + \cdots + x_n\underline{a}_n \mid x_1, \dots, x_n \ \text{arbitrary}\} \quad ,$$

which is the column space of A.
So ***null space and range generalise two notions from the matrix world.***

Example 1.1.13 ***(Orthogonal projections)*** We also revisit example 1.1.2, the orthogonal projection $\mathcal{P}$ on a line $\ell = \mathrm{span}(\underline{a})$: Geometrically, we see that the range $\mathcal{R}(\mathcal{P})$ (so the collection of vectors that occur as an image of $\mathcal{P}$) is ℓ itself. The null space $\mathcal{N}(\mathcal{P})$ consists of all vectors that are mapped to $\underline{0}$, so all vectors in $\ell^{\perp}$, the orthogonal complement.

We now recall how injectivity, surjectivity and the inverse image are connected to the null space and the range:

Theorem 1.1.14. *Consider a linear map $\mathcal{A} : V \to W$.*

1. $\mathcal{N} = \{\underline{0}\} \Leftrightarrow \mathcal{A}$ *is injective.*
2. $\mathcal{R} = W \Leftrightarrow \mathcal{A}$ *is surjective.*
3. *Let $\underline{b} \in \mathcal{R}$. Then there is a vector $\underline{p}$ satisfying $\mathcal{A}\underline{p} = \underline{b}$; we say that $\underline{p}$ is a* **particular solution** *of the vector equation $\mathcal{A}\underline{x} = \underline{b}$. Together all solutions of the vector equation $\mathcal{A}\underline{x} = \underline{b}$ are given by the set*
$$\underline{p} + \mathcal{N} := \{\underline{p} + \underline{w} \mid \underline{w} \in \mathcal{N}\} \quad ,$$
the **coset**. *In particular, the equation $\mathcal{A}\underline{x} = \underline{b}$ has exactly one solution if $\mathcal{N} = \{\underline{0}\}$.*

In particular, ***theorem 1.1.14 tells us how to we get all solutions of a vector equation $\mathcal{A}\underline{x} = \underline{b}$:*** find a single particular solution, and then add to that all solutions of the corresponding homogeneous equation $\mathcal{A}\underline{x} = \underline{0}$.

If we know how $\mathcal{A}$ acts on a basis (meaning we know the images for all basis vectors), ***it is possible to determine the image of any vector simply from the basis vector images***, due to the linearity of $\mathcal{A}$,

We also learnt that ***any basis-images combination yields a unique linear map***:

Theorem 1.1.15. *Let V and W be vector spaces, let $\{\underline{a}_1, \dots, \underline{a}_n\}$ be a basis for V, and let $\underline{w}_1, \dots, \underline{w}_n$ be an n–tuple of vectors in W. Then there exists a unique linear map $\mathcal{A} : V \to W$ satisfying $\mathcal{A}\underline{a}_i = \underline{w}_i$ for $i = 1, \dots, n$.*

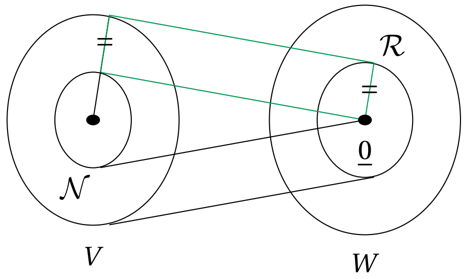
... and that ***the range of a map hence can also be written in terms of basis vector images***, by using the linear span (which was recalled in definition B.2.3):

Theorem 1.1.16. *Consider a linear map $\mathcal{A} : V \to W$ with $V = \mathrm{span}(\underline{a}_1, \dots, \underline{a}_n)$. Then*
$$\mathcal{R} = \mathrm{span}(\mathcal{A}\underline{a}_1, \dots, \mathcal{A}\underline{a}_n) \quad .$$

While it is not easy to give a comparably simple characterisation of $\mathcal{N}$, we have at least the following important result that says something about the dimension:

Theorem 1.1.17 (Dimension Theorem). *Let* $\mathcal{A}: V \to W$ *be a linear map with* $\dim(V) < \infty$. *Then*

$$\dim(V) = \dim(\mathcal{N}) + \dim(\mathcal{R})\ .$$



Example 1.1.18 ***(Orthogonal projections)*** Revisiting orthogonal projections $\mathcal{P}$ on a subspace W of the inner product space V (example 1.1.2) once more: We can easily convince ourselves that the range of $\mathcal{P}$ equals W, while the null space consists of all vectors perpendicular (orthogonal) to W, so $\mathcal{N} = W^{\perp}$. The dimension theorem above therefore implies that

$$\dim V = \dim W + \dim W^{\perp}\ .$$

Lastly, we recall a nice ***criterion that tells us when a linear map with finite-dimensional starting space is invertible***:

Theorem 1.1.19. *Let* V *be a vector space with* $\dim(V) < \infty$, *and let* $\mathcal{A}: V \to W$ *be linear map.* $\mathcal{A}$ *has an inverse if and only if* $\dim(V) = \dim(W)$ *and* $\mathcal{N} = \{\underline{0}\}$.

If we do not feel like checking the condition $\mathcal{N} = \{\underline{0}\}$ (for whatever reason), we can alternatively replace this condition with the condition that $\mathcal{R} = W$.

1.1.1 The new example for linear maps: the quotient map

In LinA 1, you learned about quotient spaces modulo a subspace U (recalled in appendix B.2.1), where two vectors $\underline{v}$ and $\underline{w}$ lie in the same residue class $[\underline{v}]$ if $\underline{v} - \underline{w} \in U$. ***We will now revisit the following example from LinA 1 and connect it to linear maps:***

Example 1.1.20 Let $V = R^3$ and let $U = \mathrm{span}((1,0,0))$. Then any two vectors (x, a, b) and (y, a, b) are equivalent modulo U, since their difference $(x-y, 0, 0) \in U$. So the residue class is $[(x, a, b)] = [(y, a, b)] = [(0, a, b)]$, meaning V/U ***looks a bit like the linear subspace spanned by*** $\underline{e}_2, \underline{e}_3$. This will be made precise in *Linear Algebra 2* using linear maps.

This section will make the meaning of the emphasised sentence in the example more precise by proving a fundamental theorem about linear maps (theorem 1.1.21 below). Originally, theorem 1.1.21 was devised by Emmy Noether, a last-century mathematician who made many important contributions to abstract algebra and mathematical physics, but wasn't allowed to officially teach for a long time due to her gender. (She did it anyways, under David Hilbert's name!)

We quickly introduce a concept needed for the theorem: Since for any linear map $\mathcal{A}: V \to W$, the null space $\mathcal{N}$ is a linear subspace of V, we can also define the respective quotient space $V/\mathcal{N}$.

> **Theorem 1.1.21 (Noether's fundamental theorem on homomorphisms, vector space edition).** *For any linear map $\mathcal{A}: V \to W$, there exists a linear bijection between its range $\mathcal{R}$ and the quotient space $V/\mathcal{N}$, where $\mathcal{N}$ is the null space of $\mathcal{A}$.*
>
> *In other words, $\mathcal{R}$ and $V/\mathcal{N}$ are isomorphic (or even shorter: $\mathcal{R} \cong V/\mathcal{N}$).*

How does this relate to example 1.1.20? To make the connection, we use as $\mathcal{A}$ the projection unto the subspace span$(\underline{e}_2, \underline{e}_3)$:

$$\mathcal{P}: \mathbb{R}^3 \xrightarrow{(x,a,b)\mapsto(0,a,b)} \mathbb{R}^3 \ .$$

As we reminded ourselves in example 1.1.18, the null space of a projection is the orthogonal complement of the subspace unto which we project. So $\mathcal{N}(\mathcal{P}) = \text{span}(\underline{e}_2, \underline{e}_3)^\perp = \text{span}((1,0,0)) = U$, and the range of $\mathcal{P}$ is $\mathcal{R}(\mathcal{P}) = \text{span}(\underline{e}_2, \underline{e}_3)$. Plugging this into theorem 1.1.21, we get that span$(\underline{e}_2, \underline{e}_3)$ is equivalent to V/U, up to an isomorphism.

To see that theorem 1.1.21 is true and how the isomorphism looks, we'll go through the proof:

(I interspersed the proof with our example to illustrate the main ideas.)

Proof of theorem 1.1.21. For any linear map $\mathcal{A}$ with null space $\mathcal{N}$, we can define a map π that maps vectors to their residue classes:

$$\pi: V \xrightarrow{\underline{v}\mapsto[\underline{v}]} V/\mathcal{N} \ .$$

(Applying this to example 1.1.20: π maps any vector (x,a,b) to $[(0,a,b)] \in V/U$.)

To practice a bit with the involved concepts, your homework will include convincing yourself that the map π always is linear, surjective (for the definition, see the table in appendix A.2.2); and that its null space is exactly the null space of $\mathcal{A}$.

We can also define a 'quotient version' of $\mathcal{A}$, by that we mean a map $\overline{\mathcal{A}}$ that maps residue classes to the image of an element of the residue class:

$$\overline{\mathcal{A}}: V/\mathcal{N} \xrightarrow{[\underline{v}]\mapsto\mathcal{A}\underline{v}} W \ .$$

(Applying this to example 1.1.20: $\overline{\mathcal{P}}$ maps any residue class $[(0,a,b)]$ to the projection $\mathcal{P}(0,a,b) = (0,a,b)$.)

But wait, is our mapping rule well-defined? There can be many different elements $\underline{v}_1, \underline{v}_2, \ldots$ in the same residue class $[\underline{v}]$ – which image $\mathcal{A}\underline{v}_1, \mathcal{A}\underline{v}_2, \ldots$ do we pick for $\overline{\mathcal{A}}[\underline{v}]$? Luckily, the definition still makes sense because all elements sharing the same residue class actually are mapped to the same image: If $\underline{v}$ and $\underline{w}$ are in the same residue class, then $\underline{v} - \underline{w} \in \mathcal{N}$, meaning $\mathcal{A}(\underline{v} - \underline{w}) = \underline{0}$. Due to the linearity of $\mathcal{A}$, we get $\mathcal{A}(\underline{v}) = \mathcal{A}(\underline{w})$.

So $\overline{\mathcal{A}}$ is a well-defined map from $V/\mathcal{N}$ to $\mathcal{R}(\mathcal{A})$ – we can restrict the 'target space' of $\overline{\mathcal{A}}$ to the actual range rather than using the whole space W – that is naturally surjective due to the restriction. (For example 1.1.20, this means restricting the target space to the subspace span$(\underline{e}_2, \underline{e}_3)$.)

We get the following diagram:

$$
\begin{tikzcd}
V \arrow[rr, two heads, "\mathcal{A}"] \arrow[dr, "\pi"'] & & \mathcal{R}(\mathcal{A}) \\
& V/\mathcal{N} \arrow[ur, hook, "\overline{\mathcal{A}}"'] &
\end{tikzcd}
$$

What's missing to verify that $\overline{\mathcal{A}}$ is an isomorphism? We still need to show that $\overline{\mathcal{A}}$ always is

- linear and
- injective (the definition of injectivity is also recalled in appendix A.2.2) ,

which you will do in your homework.

So $\overline{\mathcal{A}}$ is a linear bijection from $V/\mathcal{N}$ to $\mathcal{R}(\mathcal{A})$, which proves theorem 1.1.21. $\square$

Given that $\overline{\mathcal{A}}$ is a bijection, there also exists an inverse $\overline{\mathcal{A}}^{-1} : \mathcal{R}(\mathcal{A}) \to V/\mathcal{N}$. ***But how does the inverse look, concretely?*** Let $w \in \mathcal{R}(\mathcal{A})$. Since w is in the range of $\mathcal{A}$, there always exists a pre-image $v \in \mathcal{A}^{-1}(w)$, so we can set $\overline{\mathcal{A}}^{-1}(w) := [v]$.

1.2. Recap: Connecting matrices and linear maps (as seen in *Linear Algebra 1*)

In this recap section, we recall the close connection between matrices and linear maps that are of the form $\mathbb{K}^n \to \mathbb{K}^m$, where $\mathbb{K}$ is an arbitrary field. (E.g., you can think of $\mathbb{K} = \mathbb{R}$, or of $\mathbb{K} = \mathbb{C}$, but $\mathbb{K}$ can actually be any set that satisfies the rules for fields.)

This section covers:

- how every such matrix is connected to a linear map, and vice versa;
- how to compute the respective matrix for a given map;
- the connection with systems of linear equations;
- that the connection between maps and matrices behaves natural: we can easily switch between maps and matrices even when dealing, e.g., with compositions or inverses

This section requires familiarity with matrices and matrix multiplication which you can revisit in appendix B.3 when needed.

In example 1.1.4, ***we saw that any matrix*** A ***defines a linear map***

$$\mathcal{A} : \mathbb{K}^n \xrightarrow{v \mapsto Av} \mathbb{K}^m \; .$$

In a bit more detail, we also showed in LinA I by substitution that ***the columns of the matrix are the images of the standard base vectors:*** Let $\underline{e}_1, \ldots, \underline{e}_n$ be the standard basis of $\mathbb{K}^n$, and let $\underline{c}_1, \ldots, \underline{c}_n$

be the columns of the matrix A. Then

$$\mathcal{A}\underline{e}_1 = A\underline{e}_1 = \underline{c}_1,\ \mathcal{A}\underline{e}_2 = \underline{c}_2,\ \ldots, \mathcal{A}\underline{e}_n = \underline{c}_n\ .$$

We will now recall that ***vice versa, we can also associate to each linear map a matrix that 'represents' the map, meaning we found a new (and very important) use for matrices***:

Definition/theorem 1.2.1. *Every linear map $\mathcal{A} : \mathbb{K}^n \to \mathbb{K}^m$ is determined by an $m \times n$-matrix A, called* **the matrix of the linear map** *$\mathcal{A}$, whose columns are $\mathcal{A}\underline{e}_1, \ldots, \mathcal{A}\underline{e}_n$. The image of the vector $\underline{v}$ under $\mathcal{A}$ can be computed as the matrix product $A\underline{v}$.*

Proof. Let's take an arbitrary linear map $\mathcal{A} : \mathbb{K}^n \to \mathbb{K}^m$, and an arbitrary vector $\underline{v} \in \mathbb{K}^n$. We can always write a $\underline{v}$ as a linear combination of the standard basis vectors: $\underline{v} = x_1\underline{e}_1 + \cdots + x_n\underline{e}_n$. Then $\mathcal{A}\underline{v} = x_1\mathcal{A}\underline{e}_1 + \cdots + x_n\mathcal{A}\underline{e}_n$ due to the linearity of $\mathcal{A}$. Collecting all image vectors $\mathcal{A}\underline{e}_1, \ldots, \mathcal{A}\underline{e}_n$ as columns in an $m \times n$–matrix A, we obtain that $\mathcal{A}\underline{v} = x_1\mathcal{A}\underline{e}_1 + \cdots + x_n\mathcal{A}\underline{e}_n$ equals the matrix product $A\underline{v}$. □

Remark 1.2.2 ***(How to compute the matrix if we only know the images on some basis?)*** Assume we know how $\mathcal{A}$ acts on a basis $\alpha = \underline{a}_1, \ldots, \underline{a}_n$ for $\mathbb{K}^n$, meaning we are given a description of $\mathcal{A}$ that only tells us what $\mathcal{A}\underline{a}_1, \ldots, \mathcal{A}\underline{a}_n$ are. As mentioned in section 1.1, such a description is sufficient to uniquely determine $\mathcal{A}$, but how do we determine the corresponding matrix?

- If α is the standard basis, we can simply write down the matrix using the previous definition/theorem 1.2.1.
- If α is not the standard basis, we learnt in LinA how we can apply row reduction techniques to find the image vectors of the standard basis, which leads us back to the first case, as in the example below.

Example 1.2.3 The linear map $\mathcal{A} : \mathbb{R}^3 \to \mathbb{R}^3$ is given by

$$\mathcal{A}(-1,0,1) = (-4,2,4),\ \mathcal{A}(1,1,0) = (1,-1,-1),\ \mathcal{A}(0,1,2) = (-5,4,4)\ .$$

We put this data in three (vector, image)-rows (for clarity, we visually separate the vector and its image with a vertical bar):

$$\left(\begin{array}{rrr|rrr} -1 & 0 & 1 & -4 & 2 & 4 \\ 1 & 1 & 0 & 1 & -1 & -1 \\ 0 & 1 & 2 & -5 & 4 & 4 \end{array}\right).$$

Performing row-reduction, we obtain the (row-reduced) normal form

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

in which the left rows are the standard basis vectors $\underline{e}_1, \underline{e}_2, \underline{e}_3$ and the right rows are their images $\mathcal{A}\underline{e}_1, \mathcal{A}\underline{e}_2, \mathcal{A}\underline{e}_3$. Using definition/theorem 1.2.1 (and translating the rows on the right-hand side back into columns), we find that the matrix of $\mathcal{A}$ is

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

It is easy to verify correctness of our computations by using this matrix to compute the images of the vectors $(-1,0,1)$, $(1,1,0)$ and $(0,1,2)$.

Remark 1.2.4 ***(Connection with systems of linear equations)*** Consider the following system of linear equations

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

with $m \times n$–coefficient matrix A. Then A determines a linear map $\mathcal{A} : \mathbb{K}^n \to \mathbb{K}^m$, so ***systems of equations can be written as a vector equation***

$$\mathcal{A}\underline{x} = \underline{b} \quad .$$

Let's look how we can use what we learnt so far to ***find all solutions of the equation*** $\mathcal{A}\underline{x} = \underline{b}$ ***(if they exist):*** theorem 1.1.14 tells us that we find all solutions to $\mathcal{A}\underline{x} = \underline{b}$ by finding one particular solution $\underline{p}$, and adding to it all vectors from the null space, meaning the solution set is of the form

$$\underline{p} + \mathcal{N} := \{\underline{p} + \underline{w} \mid \underline{w} \in \mathcal{N}\} \quad .$$

But does such a particular solution $\underline{p}$ exist? To decide this, we first identify the range $\mathcal{R}$ of $\mathcal{A}$, using theorem 1.1.16:

$$\mathcal{R} = \operatorname{span}(\mathcal{A}\underline{e}_1, \dots, \mathcal{A}\underline{e}_n) \quad ,$$

which is exactly the column space of the matrix A. The equation $\mathcal{A}\underline{x} = \underline{b}$ has a solution (at least one) if and only if $\underline{b} \in \mathcal{R}$, so if and only if $\underline{b}$ is contained in the column space of the matrix.

Now that we've found a criterion to determine whether particular solutions exist, it only remains to quickly identify the null space $\mathcal{N}$: $\mathcal{N}$ consists of all vectors $\underline{v}$ with $\mathcal{A}\underline{v} = \underline{0}$, so all solutions of the homogeneous system $A\underline{x} = \underline{0}$.

The rest of this section briefly recalls that ***matrices of compositions/sums/inverses can be found in a very natural way***: e.g., the matrix of a sum $\mathcal{A} + \mathcal{B}$ of linear maps $\mathcal{A}$ and $\mathcal{B}$ is exactly the sum of their two representing matrices A and B. This makes it very easy to switch between matrix-POVs and map-POVs!

Theorem 1.2.5 (Matrix of sums and scalar multiples). *Let A and B be two matrices of the same size, and let $\mathcal{A}$ and $\mathcal{B}$ be the corresponding linear maps. Then $A+B$ is the matrix of the sum $\mathcal{A}+\mathcal{B}$.*

For every scalar $\alpha \in \mathbb{K}$, the matrix of the linear map $\alpha\mathcal{A}$ is αA.

Theorem 1.2.6 (Matrix of compositions). *Let A and B be two matrices such that the composition AB exists, and let $\mathcal{A}$ and $\mathcal{B}$ be the corresponding linear maps. Then AB is the matrix of the composition map $\mathcal{AB}$.*

Theorem 1.2.7 (Matrix of the inverse). *Let $\mathcal{A} : \mathbb{K}^n \to \mathbb{K}^n$ be a linear map with matrix A. Then the map $\mathcal{A}$ has an inverse if and only if its matrix A has an inverse, in which case the matrix for the map's inverse $\mathcal{A}^{-1}$ is exactly the inverse A^{-1} of the matrix A.*

1.2.1 Computing a matrix inverse via row reduction

Lastly, we recall how our accumulated knowledge was used to close a gap that had remained so far:

When matrix inverses were introduced, it was mentioned without proof that $AB = I$ implies $BA = I$ (for square matrices A and B). ***We will now revisit how we used the connection between linear maps and matrices to prove that*** $AB = I$ ***implies*** $BA = I$***:***

Theorem 1.2.8. *Let A and B be $n \times n$-matrices. If $AB = I$, then $BA = I$.*

Proof. Instead of proving the claim by dealing with matrices, we'll look at it through a linear maps lens: Let $\mathcal{A}, \mathcal{B} : \mathbb{K}^n \to \mathbb{K}^n$ be the linear maps corresponding to A and B. (Like at the beginning of 1.2.) Now, theorem 1.2.6 tells us that the matrix for the composition $\mathcal{AB}$ is the composition of the maps A and B (so the identity matrix), hence $\mathcal{AB} = \mathcal{I}$. If we could prove that also $\mathcal{BA} = \mathcal{I}$ is true, then we'd already be done, as we could translate this equation back into matrix terms and get that $BA = I$ as desired.

How do we prove that $\mathcal{BA} = \mathcal{I}$ is true? Since $\mathcal{AB} = \mathcal{I}$, we have in other words that $\mathcal{AB}\underline{v} = \underline{v}$ for all $\underline{v} \in \mathbb{K}^n$, meaning that $\mathcal{A}$ is the left-hand inverse of $\mathcal{B}$. For bijective maps, the right-hand inverse and the left-hand inverse is the same, meaning that also $\mathcal{BA} = \mathcal{I}$, and we're done. $\square$

The connection between maps and matrices also ***explains why/how we can do row reduction to compute the inverse of square matrices (if it exists)***, since we can compute the matrix of a map's inverse by working with base vector images (and doing row reduction). We saw the following two examples:

Example 1.2.9 ***(Computing the inverse via row reduction)*** Consider the matrix

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

For the corresponding linear map $\mathcal{A}$, we have $\mathcal{A}\underline{e}_1 = (1,0,4)$, $\mathcal{A}\underline{e}_2 = (0,1,-3)$ and $\mathcal{A}\underline{e}_3 = (3,2,8)$. For the inverse map $\mathcal{A}^{-1}$, it follows that $\mathcal{A}^{-1}(1,0,4) = \underline{e}_1$, $\mathcal{A}^{-1}(0,1,3) = \underline{e}_2$ and $\mathcal{A}^{-1}(3,2,8) = \underline{e}_3$. (Because, e.g., $\mathcal{A}^{-1}(1,0,4) = \mathcal{A}^{-1}\mathcal{A}\underline{e}_1 = \underline{e}_1$.) We can now determine the matrix of $\mathcal{A}^{-1}$ by computing $\mathcal{A}^{-1}\underline{e}_1$, $\mathcal{A}^{-1}\underline{e}_2$ and $\mathcal{A}^{-1}\underline{e}_3$, since we know that these vectors constitute the column vectors of $\mathcal{A}^{-1}$. We get $\mathcal{A}^{-1}\underline{e}_1$, $\mathcal{A}^{-1}\underline{e}_2$ and $\mathcal{A}^{-1}\underline{e}_3$ from

$\mathcal{A}^{-1}(1,0,4),\ldots,\mathcal{A}^{-1}(3,2,8)$ exactly as in example 1.2.3:

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

Row reduction yields the normal form

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 7 & 4 & -2 \\ 0 & 1 & 0 & -\frac{9}{2} & -2 & \frac{3}{2} \\ 0 & 0 & 1 & -\frac{3}{2} & -1 & \frac{1}{2} \end{array}\right)$$

Translating the rows on the right-hand side back into columns gives us the inverse:

$$A^{-1} = \frac{1}{2}\begin{pmatrix} 14 & -9 & -3 \\ 8 & -4 & -2 \\ -4 & 3 & 1 \end{pmatrix}.$$

Example 1.2.10 ***(Showing that there is no inverse via row reduction)*** We consider a similar matrix that differs only in the last row:

$$A = \begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & 2 \\ 4 & -3 & 6 \end{pmatrix}.$$

For the inverse $\mathcal{A}^{-1}$ of the corresponding linear map $\mathcal{A}$, (provided it exists), we must have $\mathcal{A}^{-1}(1,0,4) = \underline{e}_1$, $\mathcal{A}^{-1}(0,1,-3) = \underline{e}_2$, $\mathcal{A}^{-1}(3,2,6) = \underline{e}_3$, yielding the system

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

Partial row reduction gives

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

This shows that the columns of A are dependent, and that there hence is no inverse matrix.

1.3. Representation matrices for finite-dimensional spaces $\neq \mathbb{K}^n$, using coordinates

What will we do? So far, we saw that linear maps from $\mathbb{K}^n \to \mathbb{K}^m$ can be described using matrices. ($\mathbb{K}$ was an arbitrary field – e.g., you can think of $\mathbb{K} = \mathbb{R}$, or of $\mathbb{K} = \mathbb{C}$, but $\mathbb{K}$ can actually be any set that satisfies the rules for fields.) What helped us there was that these spaces have standard bases, which allows us to identify the base images with the column vectors of the matrix. What happens if we have an abstract $\mathbb{K}$-vector space instead of $\mathbb{K}^n$, meaning we might not have such a nice standard basis? ***The result of section 1.3 is that we still can come up with matrices representing the map!*** Our ***main idea*** will be: Pick a basis and ***translate the whole setting back to the $\mathbb{K}^n \to \mathbb{K}^m$ case by working with coordinates***.

In more detail, we'll

- analyse how the coordinates of a vector change when switching to another basis;
- connect linear maps to matrices for arbitrary vector spaces (depending on the choice of basis);
- use the results to analyse how these matrices transform when switching to another basis.

Why do we care? As mentioned in the 'Why do we care?' paragraph at the beginning of this chapter, we will use these techniques in the later chapters ***a lot***, essentially whenever we find and use 'nicer' bases to bring matrices into a 'nicer' (simpler) shape.

1.3.1 Coordinates

We start by recalling coordinates from LinA I:

Definition 1.3.1 (Coordinates). *Let V be an n-dimensional vector space with basis $\alpha = \{\underline{a}_1, \dots, \underline{a}_n\}$. Every vector $\underline{v} \in V$ can be expressed as a linear combination of the basis vectors in exactly one way:*

$$\underline{v} = x_1 \underline{a}_1 + x_2 \underline{a}_2 + \cdots + x_n \underline{a}_n \,.$$

The numbers $x_1, \dots, x_n$ are the **coordinates of $\underline{v}$ with respect to the basis** *α and $(x_1, \dots, x_n)$ is the* **coordinate vector of $\underline{v}$ with respect to the basis** *α.*

⚠ Clearly, the coordinates depend on the choice of the basis α.

Example 1.3.2 ***(Coordinates for the real-polynomial vector space)*** Consider the vector space V of real polynomials of degree at most 2, and take the polynomial $p := 1 + 2x + 3x^2$ (which is a vector in V). We pick the basis $\alpha := \{1, x, x^2\}$.
The α-coordinates of p are 1, 2, 3, so the α-coordinate vector of p is $(1, 2, 3)$.

Convince yourself that the map associating to each vector $\underline{v}$ the corresponding coordinate vector is a bijection. We make an even stronger observation: The following theorem tells us that ***mapping vectors to their coordinate vectors yields an invertible linear map***.

Definition/theorem 1.3.3. *Let V be an n-dimensional vector space with basis $\alpha = \{\underline{a}_1, \dots, \underline{a}_n\}$. We will* **denote the map sending each vector $\underline{v}$ to its coordinates with respect to basis α also by α***.*

Then α is an invertible linear map from V to $\mathbb{K}^n$.

With this notation, $\alpha(\underline{v})$ is the coordinate vector of the vector $\underline{v} \in V$ with respect to the basis α.

Proof of Theorem 1.3.3. As we learnt in LinAI, the coordinates of the sum $\underline{v}_1 + \underline{v}_2$ are the sum of the coordinates of $\underline{v}_1$ and of $\underline{v}_2$, and the coordinates of $\alpha \underline{v}$ are precisely α times the coordinates of $\underline{v}$. □

What happens if we switch from one basis α to another basis β? Then any vector $\underline{v} \in V$ corresponds to two different sets of coordinates: $\alpha(\underline{v})$ with respect to the basis α, and $\beta(\underline{v})$ with respect to the basis β. ***How are the α-coordinates and the β-coordinates of $\underline{v}$ related?*** Starting with the α-coordinates (viewed as a vector in $\mathbb{K}^n$), we first apply the map α^{-1} which gives us $\underline{v} \in V$, and then we apply the map β which gives us $\beta(\underline{v}) \in \mathbb{K}^n$.

$$\begin{tikzcd}
 & V \arrow[dl, "\alpha"'] \arrow[dr, "\beta"] & \\
\mathbb{K}^n \arrow[rr, "\beta\alpha^{-1}"] & & \mathbb{K}^n
\end{tikzcd}$$

'abstract vectors' level

coordinate level

Definition 1.3.4 (Coordinate transformation (map)). *Let α and β be bases of an n-dimensional vector space V. The map $\beta\alpha^{-1} : \mathbb{K}^n \to \mathbb{K}^n$ is called the* **coordinate transformation (map)** *from α to β.*

The coordinate transformation $\beta\alpha^{-1}$ is itself a linear map: We already know that α and β are linear, hence linearity of $\beta\alpha^{-1}$ follows directly from the linearity of compositions/inverses. ***Like for any linear map, we can hence associate $\beta\alpha^{-1}$ with a corresponding matrix.***

1.3.2 Coordinate transformations from a matrix POV: The basis transition matrix

Given that the coordinate transformation is a linear map from $\mathbb{K}^n$ to itself, it corresponds to an $n \times n$–matrix. ***We will use this matrix a lot in the following chapters, therefore we fix notation***:

Definition 1.3.5 (Transition matrix). *Let α and β be bases of an n-dimensional vector space V. We call the $n \times n$–matrix associated to the linear map $\beta\alpha^{-1}$ the* **transition matrix** *from basis α to basis β, and* **denote it by** ${}_\beta S_\alpha$.

The following theorem states that - as one might expect - ***multiplication with the matrix ${}_\beta S_\alpha$ translates α- into β-coordinates***, and gives a ***direct description of how the matrix looks, entry-wise:***

> **Theorem 1.3.6.** *Let α and β be bases of an n-dimensional vector space V, and let ${}_\beta S_\alpha$ be the basis transition matrix, i.e., the matrix of $\beta\alpha^{-1}$. Let $\underline{x} := \alpha(\underline{v})$ be the α–coordinate vector of a vector $\underline{v} \in V$. Then the β–coordinate vector of $\underline{v}$ is equal to the product ${}_\beta S_\alpha\, \underline{x}$. Furthermore, the columns of matrix ${}_\beta S_\alpha$ are the β–coordinate vectors of the α–basis vectors.*

Proof. To show that the product ${}_\beta S_\alpha\, \underline{x}$ yields the β–coordinate vector of $\underline{v}$, we can simply use that applying the map and taking the matrix product is equivalent: ${}_\beta S_\alpha\, \underline{x} = (\beta\alpha^{-1})(\underline{x}) = (\beta\alpha^{-1})(\alpha(\underline{v})) \overset{\star}{=} \beta(\alpha^{-1}(\alpha(\underline{v}))) = \beta(\underline{v})$, where $\star$ used associativity of map compositions.

To show that the columns of ${}_\beta S_\alpha$ indeed are the β–coordinate vectors of the α–basis vectors, we recall that the columns of a matrix are the images of the standard base vectors $\underline{e}_1, \ldots, \underline{e}_n$ under the corresponding linear map. The i-th column of ${}_\beta S_\alpha$ hence is equal to ${}_\beta S_\alpha \underline{e}_i = (\beta\alpha^{-1})(\underline{e}_i)$. Since α^{-1} (per definition) maps coordinate vectors back into the respective linear combination of the base vectors in α, we have that $\alpha^{-1}(\underline{e}_i) = \underline{a}_i$, where $\underline{a}_i$ is the i-th vector in the base α. Plugging this in, we get that $\beta(\alpha^{-1}(\underline{e}_i)) = \beta(\underline{a}_i)$, the β–coordinate vector of $\underline{a}_i$. $\square$

Example 1.3.7 ***(Basis transition matrix for real-polynomial vector space)*** We again consider the vector space V of real polynomials of degree at most 2. We pick the two bases $\alpha := \{x-1, x^2-1, x^2+1\}$ and $\beta := \{1, x, x^2\}$. We can easily express the basis vectors of α in the basis vectors of β:

$$\begin{aligned} x-1 &= \quad (-1)\cdot 1 + 1\cdot x + 0\cdot x^2 \\ x^2-1 &= \quad (-1)\cdot 1 + 0\cdot x + 1\cdot x^2 \\ x^2+1 &= \qquad 1\cdot 1 + 0\cdot x + 1\cdot x^2 \quad . \end{aligned}$$

We therefore know the β–coordinates of the vectors of α. We can use them to give the basis transition matrix, since its columns are the β–coordinate vectors of the α–basis vectors:

$${}_\beta S_\alpha = \begin{pmatrix} -1 & -1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}.$$

It is of course also possible to switch from β–coordinates to α–coordinates, which is done with the matrix ${}_\alpha S_\beta$. When looking at the product matrix ${}_\alpha S_\beta\, {}_\beta S_\alpha$ of the two matrices, we notice that it first transforms α–coordinates to β–coordinates, and then the β–coordinates back to α–coordinates, so it doesn't do anything:

$${}_\alpha S_\beta\, {}_\beta S_\alpha = I\text{, so } {}_\alpha S_\beta = {}_\beta S_\alpha{}^{-1} \quad .$$

Example 1.3.8 ***(Example 1.3.7, continued)*** As we just noticed, the transition matrix ${}_\alpha S_\beta$ is the inverse of matrix ${}_\beta S_\alpha$. Computing the inverse, we find

$${}_\alpha S_\beta = {}_\beta S_\alpha{}^{-1} = \frac{1}{2}\begin{pmatrix} 0 & 2 & 0 \\ -1 & -1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

The columns of this matrix should consist of the α–coordinates of the base vectors of β. Let's quickly verify this:

- The first column vector is $(0, -\frac{1}{2}, \frac{1}{2})$, translating this back via α^{-1} corresponds to vector $0\,(x-1) - \frac{1}{2}(x^2-1) + \frac{1}{2}(x^2+1) = 1$. ✓
- In the same way, we can verify that the second column consists of the α–coordinates of x and that the third column consists of the α–coordinates of x^2. ✓

Computing the α–coordinates of some vector: As an example, let's pick the vector $v := 2x^2 - 3x + 4$. The β-coordinates of this vector are $(4, -3, 2)$. We transform them to α–coordinates using the transition matrix ${}_\alpha S_\beta$:

$$ {}_\alpha S_\beta \begin{pmatrix} 4 \\ -3 \\ 2 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 0 & 2 & 0 \\ -1 & -1 & 1 \\ 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 4 \\ -3 \\ 2 \end{pmatrix} = \begin{pmatrix} -3 \\ \frac{1}{2} \\ \frac{3}{2} \end{pmatrix} . $$

We verify the result:

$$ -3(x-1) + \frac{1}{2}(x^2-1) + \frac{3}{2}(x^2+1) = 2x^2 - 3x + 4 \quad . \checkmark $$

If a third basis γ enters the picture, we could transform α– to β–coordinates, and these subsequently to γ–coordinates. The following theorem tells us that ***doing the β-detour is unnecessary and we can directly switch from α to γ***:

Theorem 1.3.9. *Let α, β and γ be bases of an n-dimensional vector space V, with respective basis transition matrices ${}_\beta S_\alpha$, ${}_\gamma S_\alpha$ and ${}_\gamma S_\beta$. Then*

$$ {}_\gamma S_\beta \, {}_\beta S_\alpha = {}_\gamma S_\alpha $$

Proof. The product of the two matrices is the matrix of their composition (remember theorem 1.2.6). For the composition under consideration, we have $(\gamma\beta^{-1})(\beta\alpha^{-1}) = \gamma(\beta^{-1}\beta)\alpha^{-1} = \gamma\alpha^{-1}$. □

⚠ ***It is important to distinguish between calculating with vectors*** (so elements of the vector space V) ***and calculating with coordinates*** (so sequences of elements from $\mathbb{K}^n$). In example 1.3.7, the separation between the two concepts was clear: the vectors were polynomials and the coordinate vectors were sequences of real numbers.

We need to be more careful with this distinction if our vector space is $\mathbb{K}^n$ itself, as a sequence could either represent a vector from the vector space $\mathbb{K}^n$ or a sequence of coordinates. We note that this strict separation is less crucial if we talk about coordinates w.r.t. the standard basis $\varepsilon = \{\underline{e}_1, \ldots, \underline{e}_n\}$:

$$ (1,2,3) = 1\underline{e}_1 + 2\underline{e}_2 + 3\underline{e}_3 \, . $$

While having $\mathbb{K}^n$ as the vector space itself might make it more complicated to separate between vectors and their coordinate vectors, the following example shows that $\mathbb{K}^n$ ***allows for a nice shortcut when you need to compute basis transition matrices***.

Example 1.3.10 ***(Computing a basis transition matrix in $\mathbb{R}^3$)*** Let's consider the following two bases:

$$ \alpha = \{(1,0,2),\ (-1,1,0),\ (0,-2,1)\} \, ; $$
$$ \beta = \{(0,1,1),\ (1,2,-1),\ (1,0,1)\} \quad . $$

We are looking for the transition matrix that transforms α– into β–coordinates, so ${}_\beta S_\alpha$.
Method like in example 1.3.7. We'll first solve the problem directly, by finding a way to write each of the α-vectors as a linear combination of the β-vectors. This means having to solve

three systems of equations in three unknowns:

$$\left(\begin{array}{ccc|ccc} 0 & 1 & 1 & 1 & -1 & 0 \\ 1 & 2 & 0 & 0 & 1 & -2 \\ 1 & -1 & 1 & 2 & 0 & 1 \end{array}\right) \xrightarrow{\text{row reduction}} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{1}{2} & 1 & -\frac{1}{2} \\ 0 & 1 & 0 & -\frac{1}{4} & 0 & -\frac{3}{4} \\ 0 & 0 & 1 & \frac{5}{4} & -1 & \frac{3}{4} \end{array}\right).$$

Now the last three columns are the β-coordinates of the three vectors of α, hence

$$_\beta S_\alpha = \frac{1}{4}\begin{pmatrix} 2 & 4 & -2 \\ -1 & 0 & -3 \\ 5 & -4 & 3 \end{pmatrix}.$$

Method with shortcut. Our alternative method uses that ***in*** $\mathbb{K}^n$***, the matrix for switching to the standard basis*** ε ***is super-easy to compute***: We know the standard-basis coordinates of both bases. E.g., the first vector of α is $(1,0,2) = 1\underline{e}_1 + 0\underline{e}_2 + 2\underline{e}_3$, meaning the first column of $_\varepsilon S_\alpha$ must be (1,0,2), so the first base vector from α itself, and so on. So, ***without any computational work we already know the transition matrices*** $_\varepsilon S_\alpha$ ***and*** $_\varepsilon S_\beta$:

$$_\varepsilon S_\alpha = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -2 \\ 2 & 0 & 1 \end{pmatrix}, \quad _\varepsilon S_\beta = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & -1 & 1 \end{pmatrix}.$$

To be as lazy as possible on our way to finding the α–to–β transition matrix, we now use that we can do a detour via basis ε: $_\beta S_\alpha = {}_\beta S_\varepsilon\, {}_\varepsilon S_\alpha$ (due to theorem 1.3.9). We'll also use that switching the direction of the base transition is the same as taking the inverse: $_\beta S_\varepsilon = {}_\varepsilon S_\beta{}^{-1}$. Combining both observations, we get the identity $_\beta S_\alpha = {}_\varepsilon S_\beta{}^{-1}\, {}_\varepsilon S_\alpha$.
So, we first determine the inverse of $_\varepsilon S_\beta$ and find

$$_\beta S_\varepsilon = {}_\varepsilon S_\beta{}^{-1} = \frac{1}{4}\begin{pmatrix} -2 & 2 & 2 \\ 1 & 1 & -1 \\ 3 & -1 & 1 \end{pmatrix}.$$

We can now multiply this with $_\varepsilon S_\alpha$ from the right and get

$$_\beta S_\alpha = \frac{1}{4}\begin{pmatrix} -2 & 2 & 2 \\ 1 & 1 & -1 \\ 3 & -1 & 1 \end{pmatrix}\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -2 \\ 2 & 0 & 1 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 2 & 4 & -2 \\ -1 & 0 & -3 \\ 5 & -4 & 3 \end{pmatrix}.$$

1.3.3 Generalising the map-matrix connection for spaces that aren't $\mathbb{K}^n$

So far, we saw that linear maps from $\mathcal{A}: \mathbb{K}^n \to \mathbb{K}^m$ can be described using matrices. ***We will now generalise the connection***: we will replace $\mathbb{K}^n/\mathbb{K}^m$ with arbitrary finite-dimensional $\mathbb{K}$-vector spaces V/W. ***Our main idea will be to translate the whole thing back to the*** $\mathbb{K}^n \to \mathbb{K}^m$ ***case by picking a basis and working with coordinates.***

So, let V be a finite-dimensional vector space with basis α, let W be a finite-dimensional vector space with basis β, and let $\mathcal{A}: V \to W$ be a linear map.

Consider the following diagram:

$$
\begin{tikzcd}
V \arrow[r, "\mathcal{A}"] \arrow[d, "\alpha"'] & W \arrow[d, "\beta"] \\
\mathbb{K}^n \arrow[r, "\beta\mathcal{A}\alpha^{-1}"'] & \mathbb{K}^m
\end{tikzcd}
$$

- To every vector $\underline{v} \in V$, there corresponds a unique coordinate vector $\alpha(\underline{v})$.
- To $\mathcal{A}\underline{v}$, the image of $\underline{v}$ under map $\mathcal{A}$, there corresponds a unique coordinate vector $\beta(\mathcal{A}\underline{v})$.

This diagram might look complicated at a first glance, but in fact, I already tricked you into using the coordinate concept in the last chapter! We did this in examples 1.3.7 and 1.3.8, where the bases were polynomials instead of vectors in $\mathbb{R}^n$.

The composite linear map $\beta\mathcal{A}\alpha^{-1} : \mathbb{K}^n \to \mathbb{K}^m$ maps the coordinate vector $\alpha(\underline{v})$ to the coordinate vector $\beta(\mathcal{A}\underline{v})$. Since $\beta\mathcal{A}\alpha^{-1}$ is a linear map between $\mathbb{K}^n$ and $\mathbb{K}^m$, we already know how to represent it with a matrix (remember definition/theorem 1.2.1). ***This gives us a representation matrix for the linear map $\mathcal{A}$ on the coordinate level:***

Definition 1.3.11 (Matrix of a linear map). *Let V, W, α, β, $\mathcal{A}$ be as above. We denote the matrix of the linear map $\beta\mathcal{A}\alpha^{-1}$ by ${}_\beta A_\alpha$ and call it the* **matrix of $\mathcal{A}$ w.r.t. the bases α and β** *.*

If $V = W$ and $\beta = \alpha$, then we simplify notation by denoting the corresponding matrix by A_α; we call it the **matrix of $\mathcal{A}$ w.r.t. basis α***.*

In other words: The matrix A_α maps the α–coordinates of a vector $\underline{v}$ to the α–coordinates of $\mathcal{A}\underline{v}$.

We will spend a lot of time with the special case $V = W$ and $\beta = \alpha$ during this course, but let us first say a few things about the general case.

How does the matrix look? According to definition/theorem 1.2.1, the columns of ${}_\beta A_\alpha$ are

$$(\beta\mathcal{A}\alpha^{-1})(\underline{e}_i) = \beta(\mathcal{A}\underline{a}_i)\ , \qquad i = 1, \ldots, n\ ,$$

meaning the i–th column consists of the β–coordinates of the image $\mathcal{A}\underline{a}_i$ of the i–th base vector in α.

This indeed gives us a description of $\mathcal{A}$ on the coordinate level: for example, to find the image of a vector $\underline{v} \in V$, we can

- determine the coordinate vector $\alpha(\underline{v})$ of $\underline{v}$;
- multiply $\alpha(\underline{v})$ with the representation matrix ${}_\beta A_\alpha$, yielding the coordinate vector of $\mathcal{A}\underline{v}$; and
- translate the coordinate vector of $\mathcal{A}\underline{v}$ back to the corresponding vector in W.

The last thing we'll do before we look at some examples is to make sure that ***our new, more general view does not clash with our definition in the old, less general case where*** $V = \mathbb{K}^n$ ***and*** $W = \mathbb{K}^m$ with their respective standard basis:

Observation 1.3.12 (Connection with *the* matrix). *If $\mathcal{A} : \mathbb{K}^n \to \mathbb{K}^m$ is a linear map, and α and β are the standard bases for these spaces, then ${}_\beta A_\alpha$, the matrix of $\mathcal{A}$ w.r.t. these bases as defined in definition 1.3.11, is just the matrix of $\mathcal{A}$ as defined in section 1.2. In this case, the coordinate maps α and β both are the identity maps.*

Example 1.3.13 ***(Determining a vector image from the representation matrix)*** Say V is two-dimensional with basis $\alpha = \{\underline{a}, \underline{b}\}$, and the linear map $\mathcal{A} : V \to V$ has the matrix

$$A_\alpha = \begin{pmatrix} 1 & 4 \\ -2 & 3 \end{pmatrix}$$

w.r.t. α. The matrix tells us that $\mathcal{A}\underline{a} = 1 \cdot \underline{a} - 2 \cdot \underline{b}$ (first column) and $\mathcal{A}\underline{b} = 4 \cdot \underline{a} + 3 \cdot \underline{b}$ (second column). To compute the image of some vector $\underline{v} = \lambda\underline{a} + \mu\underline{b}$, we multiply A_α with $(\lambda, \mu)^\top$:

$$\begin{pmatrix} 1 & 4 \\ -2 & 3 \end{pmatrix} \begin{pmatrix} \lambda \\ \mu \end{pmatrix} = \begin{pmatrix} \lambda + 4\mu \\ -2\lambda + 3\mu \end{pmatrix}$$

Hence the image of $\underline{v}$ is $(\lambda + 4\mu)\underline{a} + (-2\lambda + 3\mu)\underline{b}$.

Example 1.3.14 ***(Representation matrix for differentiation.)*** Consider the vector space V of real polynomials of degree at most 2, and the linear map $\mathcal{D} : V \to V$ defined by $\mathcal{D}p = p'$. We already saw that this is a linear map (remember example 1.1.5). Take for V the basis $\alpha = \{1, x, x^2\}$. The derivatives of the basis vectors are:

$$\begin{aligned} \mathcal{D}1 \quad &= 0 = 0{\cdot}1 + 0{\cdot}x + 0{\cdot}x^2\,, \\ \mathcal{D}x \quad &= 1 = 1{\cdot}1 + 0{\cdot}x + 0{\cdot}x^2\,, \\ \mathcal{D}x^2 \quad &= 2x = 0{\cdot}1 + 2{\cdot}x + 0{\cdot}x^2\,. \end{aligned}$$

The matrix D_α is therefore

$$D_\alpha = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

To illustrate how we can now differentiate using the matrix D_α, take the polynomial $p(x) = 2x^2 - 3x + 5$: The coordinate-vector of p w.r.t. α is $(5, -3, 2)$ and

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

$(-3, 4, 0)$ is the coordinate vector of $4x - 3$ and this is indeed the derivative of p.

1.3.4 How do base changes affect matrices of linear maps?

The rules for coordinate transformations allow us to switch between bases. ***The following observation concerns how the different representation matrices are related to each other, and will be used in the following sections a lot***:

Theorem 1.3.15 (Effect of change of basis). *Choose in a finite-dimensional space V two bases α and β, and suppose $\mathcal{A}: V \to V$ is linear. Then*

$$A_\beta = {}_\beta S_\alpha \, A_\alpha \, {}_\alpha S_\beta \, .$$

In other words: To compute how the β–coordinates of a vector $\underline{v}$ are mapped to the β–coordinates of $\mathcal{A}\underline{v}$, we can

- first transform the β–coordinates of $\underline{v}$ into α–coordinates;
- compute the α–coordinates of $\mathcal{A}\underline{v}$ using the matrix A_α (we'll see in example 1.3.16 below that this can sometimes be easier than computing A_β directly); and
- lastly, transform the result back to β–coordinates.

Proof of theorem 1.3.15. Per definition, A_β is the matrix for the composed map $\beta\mathcal{A}\beta^{-1}$ and A_α is the matrix for the composed map $\alpha\mathcal{A}\alpha^{-1}$. We now use an 'insert-1'-trick to relate the two composed maps to each other: We have

$$\beta\mathcal{A}\beta^{-1} \overset{(\star)}{=} \beta(\alpha^{-1}\alpha)\mathcal{A}(\alpha^{-1}\alpha)\beta^{-1} = (\beta\alpha^{-1})(\alpha\mathcal{A}\alpha^{-1})(\alpha\beta^{-1}) \ ,$$

where $(\star)$ used that the factors α^{-1} and α cancel (so we inserted a 'matrix-1'), and we afterwards reordered the brackets. We now get our claim by identifying $\beta\alpha^{-1}$ and $\alpha\beta^{-1}$ with their matrices ${}_\beta S_\alpha$ and ${}_\alpha S_\beta$ and using that the matrix of a composition is the matrix product of its factors' matrices (theorem 1.2.6). □

Example 1.3.16 ***(Representation matrix for an orthogonal projection)*** We once more revisit the projection example 1.1.2, but this time we make it less abstract by picking $V := \mathbb{R}^2$ with the standard inner product, and letting $\mathcal{P}$ be the orthogonal projection unto the line ℓ with equation $2x - 3y = 0$.
Our goal is to determine the matrix P_ε of $\mathcal{P}$ w.r.t. the standard basis ε. Determining the coordinates of $\mathcal{P}\underline{e}_1$ and $\mathcal{P}\underline{e}_2$ directly, however, is a bit unwieldy. Therefore, we ***start by picking a basis $\alpha = \{\underline{a}_1, \underline{a}_2\}$ such that $\mathcal{P}$ has an easy description*** in terms of α-coordinates. We take $\underline{a}_1 = (3,2) \in \ell$ and $\underline{a}_2 = (2,-3) \perp \ell$ and write the projection images in terms of the α-vectors: $\mathcal{P}\underline{a}_1 = \underline{a}_1 = 1 \cdot \underline{a}_1 + 0 \cdot \underline{a}_2$ and $\mathcal{P}\underline{a}_2 = \underline{0} = 0 \cdot \underline{a}_1 + 0 \cdot \underline{a}_2$. So the matrix P_α is

$$P_\alpha = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} .$$

Second, we 'translate P_α into ε-coordinates': Per definition, α maps vectors in V to their coordinate vectors w.r.t. α. In our special case where $V = \mathbb{R}^2$, α is the map from $\mathbb{R}^2$ to $\mathbb{R}^2$ that simply translates ε-coordinates to α-coordinates (so, e.g., $(3,2)$ is mapped to $(1,0)$), and we can identify its matrix with the basis transition matrix ${}_{\alpha}S_{\varepsilon}$. Accordingly, the matrix for α^{-1} is ${}_{\varepsilon}S_{\alpha}$. Like we already observed in example 1.3.10, computing these transition matrices is quite easy:

$$ {}_{\varepsilon}S_{\alpha} = \begin{pmatrix} 3 & 2 \\ 2 & -3 \end{pmatrix} \quad \text{and} \quad {}_{\alpha}S_{\varepsilon} = {}_{\varepsilon}S_{\alpha}{}^{-1} = \frac{1}{13}\begin{pmatrix} 3 & 2 \\ 2 & -3 \end{pmatrix}. $$

Using theorem 1.3.15, we get

$$ P_{\varepsilon} = {}_{\varepsilon}S_{\alpha}\, P_{\alpha}\ {}_{\alpha}S_{\varepsilon} = \frac{1}{13}\begin{pmatrix} 9 & 6 \\ 6 & 4 \end{pmatrix}. $$

Alternatively, the result could be obtained with the technique in remark 1.2.2: Put the information $\mathcal{P}\underline{a}_1 = \underline{a}_1$ and $\mathcal{P}\underline{a}_2 = \underline{0}$ in ε-coordinates in the rows of a matrix

$$ \left(\begin{array}{cc|cc} 3 & 2 & 3 & 2 \\ 2 & -3 & 0 & 0 \end{array}\right) \xrightarrow{\text{row reduction}} \left(\begin{array}{cc|cc} 1 & 0 & \frac{9}{13} & \frac{6}{13} \\ 0 & 1 & \frac{6}{13} & \frac{4}{13} \end{array}\right) $$

yielding

$$ P_{\varepsilon} = \frac{1}{13}\begin{pmatrix} 9 & 6 \\ 6 & 4 \end{pmatrix}. $$

Example 1.3.17 ***(Representation matrix for differentiation.)*** We look back to example 1.3.14, where we gave the representation matrix D_{α} w.r.t. basis $\alpha = \{1, x, x^2\}$. We will take a different basis, $\beta = \{x^2 - x, x^2 + 3, x^2 - 1\}$, and determine the matrix D_{β} of $\mathcal{D}$. We can now use:

$$ D_{\beta} = {}_{\beta}S_{\alpha}\, D_{\alpha}\, {}_{\alpha}S_{\beta}\,. $$

Using $x^2 - x = 0 \cdot 1 - 1 \cdot x + 1 \cdot x^2$ and so on, we get

$$ {}_{\alpha}S_{\beta} = \begin{pmatrix} 0 & 3 & -1 \\ -1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}, $$

so

$$ {}_{\beta}S_{\alpha} = {}_{\alpha}S_{\beta}{}^{-1} = \frac{1}{4}\begin{pmatrix} 0 & -4 & 0 \\ 1 & 1 & 1 \\ -1 & 3 & 3 \end{pmatrix} $$

and

$$ D_{\beta} = {}_{\beta}S_{\alpha}\, D_{\alpha}\, {}_{\alpha}S_{\beta} = \frac{1}{4}\begin{pmatrix} -8 & -8 & -8 \\ 1 & 2 & 2 \\ 7 & 6 & 6 \end{pmatrix}. $$

As an exemplary sanity check, we look at $p(x) = 2x^2 - 3x + 5$: Since $2x^2 - 3x + 5 = 3(x^2 - x) + (x^2 + 3) - 2(x^2 - 1)$, p has β–coordinates $(3, 1, -2)$, and

$$ D_{\beta}\begin{pmatrix} 3 \\ 1 \\ -2 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} -8 & -8 & -8 \\ 1 & 2 & 2 \\ 7 & 6 & 6 \end{pmatrix}\begin{pmatrix} 3 \\ 1 \\ -2 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} -16 \\ 1 \\ 15 \end{pmatrix}. $$

These now should be the β–coordinates of the derivative of $p(x)$: $-4(x^2 - x) + \frac{1}{4}(x^2 + 3) + \frac{15}{4}(x^2 + 1) = 4x - 3 = p'(x)$. ✓

Chapter 2

'Simpler descriptions' for maps/matrices

What happens in chapter 2?

A major reason why linear maps between finite-dimensional vector spaces are well-understood is because they can be analysed using specific values, called eigenvalues. The theory of eigenvalues allows to describe such maps via matrices that have a pleasantly simple form. Eigenvalue theory was first developed to study the rotational motion of rigid bodies, but as it turns out, they have much wider range of applications: you can find them, e.g., in vibration analysis to detect equipment faults, stability analysis, atomic theory, quantum mechanics, and facial recognition.

In section 2.1, we introduce the theory of eigenvalues which is central to the course and will accompany us throughout. We proceed by slightly generalising the theory in section 2.2, where we look at subspaces that are mapped into itself (so-called 'invariant' subspaces) and see how they also help simplify matrices. Based on the knowledge we gathered in these two sections, we'll develop some criteria in section 2.3 that let us decide whether a quadratic matrix is diagonalisable. We finish this chapter with **??** that discusses the 'next best thing' to being diagonal, to deal with non-diagonalisable matrices.

Learning Goals of Chapter 2: When we are finished with chapter 2, you can

- work with eigenvalues: you can
  - restate the definition of eigenvalues/vectors/spaces;
  - state their fundamental properties;
  - demonstrate your understanding of these concepts, e.g., by giving (counter)examples and deciding whether a particular value/vector is an eigenvalue/vector;
  - apply the discussed 'algorithm' to compute eigenvalues/vectors/spaces;
- work with invariant subspaces: you can
  - restate the definition of invariant subspaces;
  - demonstrate your understanding of the definition, e.g., by giving (counter)examples and deciding whether a particular subspace is invariant;

- compute the restriction of a map unto such a subspace;
  - explain how invariant subspaces help to simplify the matrix of a map;
  - compute invariant subspaces for both linear maps and square matrices over real vector spaces;
- decide if a quadratic matrix is diagonisable;
- deal with the Jordan Normal Form of a matrix; and
- prove (simple) statements about all involved definitions/concepts.

Why do we care? In conclusion, all sections help us bring matrices into simpler form. As mentioned before, you might encounter the respective techniques throughout your studies, e.g., when studying numerical methods or differential equations. Chapters chapter 3 and chapter 4 serve as examples that illustrate how what we can use the techniques to simplify several mathematical problems.

2.1. Diagonalisation via eigenvalues

What will we do? Let $\mathcal{A} : V \to V$ be a linear map, mapping a finite-dimensional $\mathbb{K}$- vector space V into itself. So far, we saw that for every choice of a basis α for V, the map $\mathcal{A}$ is determined by a matrix A_α. We also saw the connection between the two matrices A_α and A_β for different bases α and β (in theorem 1.3.15). In the two examples 1.3.14 and 1.3.16, we have already seen that the matrix A_α sometimes has a pleasantly simple form for a suitably chosen basis α. It turns out that there exist tricks to find such simple forms systematically, and this section develops these tricks/techniques. In more detail, we'll learn

- the most central concepts of the whole course, called eigenvalues, eigenvectors and eigenspaces;
- how these concepts help us bring maps into a 'pleasantly simple' form; and
- how to compute eigenvalues/vectors/spaces for any given linear map via an 'algorithm'.

We start by ***making precise what we mean by 'simple form'***:

Definition 2.1.1. *A square matrix A has* **diagonal form** *(or shorter,* **is diagonal***) if all elements a_{ij} with $i \neq j$ are equal to zero.*

Diagonality of a matrix can also be expressed through how it acts on bases:

Theorem 2.1.2. *Let $\mathcal{A} : V \to V$ be a linear map and let $\alpha = \{\underline{a}_1, \ldots, \underline{a}_n\}$ be a basis for V. The matrix A_α has the diagonal form*

$$A_\alpha = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \lambda_n \end{pmatrix}$$

if and only if $\mathcal{A}\underline{a}_i = \lambda_i \underline{a}_i$ for $i = 1, \ldots, n$.

Proof. We consider both directions:

⇒: If A_α has the diagonal form above, this means that $\mathcal{A}$ maps the α–coordinates of $\underline{a}_i$ (so $\underline{e}_i$) to $\lambda_i \underline{e}_i$, the α–coordinates of $\lambda_i \underline{a}_i$. In other words, $\mathcal{A}$ maps $\underline{a}_i$ to $\lambda_i \underline{a}_i$.

⇐: The argument is quite similar to the ⇒ direction, just the other way around: the i-th column of A_α is the image of $\underline{a}_i$, in α-coordinates. If $\mathcal{A}$ maps $\underline{a}_i$ to $\lambda_i \underline{a}_i$, then the i-th column of A_α is $\lambda_i \underline{e}_i$.

□

We are now ready to give the ***central definition of this course: eigenvalues and eigenvectors***.

Definition 2.1.3 (Eigenvector and eigenvalue). *Let $\mathcal{A} : V \to V$ be a linear map from a $\mathbb{K}$-vector space V to itself. A vector $\underline{v} \neq \underline{0}$ is called* **eigenvector** *of $\mathcal{A}$ with* **eigenvalue** *$\lambda \in \mathbb{K}$ if $\mathcal{A}\underline{v} = \lambda \underline{v}$. We denote the set of all eigenvalues of $\mathcal{A}$ by* $\mathrm{spec}(\mathcal{A})$ *and call it the* **spectrum** *of $\mathcal{A}$.*

The prefix 'eigen' is adopted from the German word for 'characteristic'/'own'. So 'eigenvalues/vectors' roughly means 'own' values/vectors. To explain this naming convention: the definition expresses that non-zero vectors are mapped unto a multiple of themselves (with the multiple being the eigenvalue), meaning they do not change direction. Or, in other words: ***they are mapped into their own linear span***.

You will re-encounter the spectrum in a more general setting, e.g., in the course on ordinary differential equations.

To get used to this definition, we first reformulate theorem 2.1.2 in eigenvalue/-vector terms:

Theorem 2.1.4. *Let $\mathcal{A} : V \to V$ be a linear map with representation matrix A_α for basis α. A_α is in diagonal form if and only if α is a basis of eigenvectors. In that case, the diagonal elements are the eigenvalues.*

Example 2.1.5 ***(Example 1.3.16, cont'd)*** In example 1.3.16, we looked at the projection unto a line. The vectors $\underline{a}_1$ and $\underline{a}_2$ we defined there are eigenvectors: $\underline{a}_1$ has eigenvalue 1 (since line elements are not changed at all by projections) $\underline{a}_2$ has eigenvalue 0 (since the orthogonal complement is mapped to 0). The matrix w.r.t. the basis $\{\underline{a}_1, \underline{a}_2\}$ is therefore

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

a diagonal matrix with the eigenvalues on the diagonal. ***Notice the order of the two eigenvalues 1 and 0 on the diagonal***: this corresponds to the order of the eigenvectors.

Example 2.1.6 Consider in the euclidean plane E^2 a rotation around the origin by 90° . No vector different from $\underline{0}$ is mapped to a multiple of itself. So this linear map has no eigenvectors (let alone a basis of such). There is hence no choice of basis for which the rotation matrix could possibly be diagonal.

We give another ***central definition of this course: eigenspaces***, i.e., the spaces of eigenvectors for an eigenvalue.

Definition 2.1.7 (Eigenspace). *Let $\mathcal{A} : V \to V$ be a linear map from a $\mathbb{K}$-vector space V to itself. For any scalar $\lambda \in \mathbb{K}$, we denote*

$$E_\lambda := \mathcal{N}(\mathcal{A} - \lambda\mathcal{I}) \ .$$

Since null spaces are subspaces, E_λ is a subspace, called the **eigenspace** *of $\mathcal{A}$ for λ.*

Eigenspaces ***indeed are spaces of eigenvectors for a given eigenvalue***: E_λ is the null space of the linear map $\mathcal{A} - \lambda\mathcal{I}$. So any vector $\underline{v}$ lies in E_λ if and only if $(\mathcal{A} - \lambda\mathcal{I})\underline{v} = \underline{0}$, which is equivalent to $\mathcal{A}v - \lambda\underline{v} = \underline{0}$. Hence $\underline{v} \in E_\lambda$ if and only if $\mathcal{A}\underline{v} = \lambda\underline{v}$, so if and only if $\underline{v}$ is eigenvector for eigenvalue λ or $\underline{v} = \underline{0}$. (Note that we ruled out $\underline{0}$ as a possible eigenvector by definition).

A few more remarks on the definition. E_λ only is interesting if λ is an eigenvalue in the other case, E_λ only contains $\underline{0}$. The other way around, we say that λ is an eigenvalue if and only if E_λ contains a non-trivial vector, that is, a vector $\underline{v} \neq \underline{0}$. (So if and only if $\dim(E_\lambda) > 0$.) Some authors only use the word eigenspace if λ is an eigenvalue.

We can also write the null space of $\mathcal{A}$ as an eigenspace: E_0 consists of vectors that are mapped to 0 times itself, so on $\underline{0}$. So E_0 is the null space of $\mathcal{A}$.

2.1.1 Computing eigenvalues and -spaces

We will need to compute eigenvalues, eigenvectors and eigenspaces a lot during this course. So how do we do this? Assume you're given a map $\mathcal{A}$, and you're tasked with the following two things:

- determine the eigenvalues, i.e., all values λ such that $\dim(E_\lambda) > 0$; and
- find the eigenvectors for such a λ, i.e., the vectors in E_λ different from $\underline{0}$.

To tackle this problem, we'll use the following theorem:

Theorem 2.1.8. *λ is an eigenvalue if and only if $\det(A - \lambda\mathcal{I}) = 0$. Let $\alpha = \{\underline{a}_1, \ldots, \underline{a}_n\}$ be a basis for V, and let*

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{n1} & \ldots & \ldots & a_{nn} \end{pmatrix}$$

be the matrix of $\mathcal{A}$ w.r.t. this basis. Then the eigenvectors for eigenvalue λ, in α-coordinates, are *the non-zero solutions of the system*

$$\begin{pmatrix} a_{11}-\lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22}-\lambda & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{n1} & \dots & \dots & a_{nn}-\lambda \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} . \tag{2.1}$$

Proof. We fix a value $\lambda \in \mathbb{K}$. When looking for eigenvectors for λ, we are concerned with solutions to the system $(\mathcal{A} - \lambda\mathcal{I})\underline{v} = \underline{0}$. The matrix of $\mathcal{A} - \lambda\mathcal{I}$ w.r.t. basis α is

$$A - \lambda\mathcal{I} = \begin{pmatrix} a_{11}-\lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22}-\lambda & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{n1} & \dots & \dots & a_{nn}-\lambda \end{pmatrix} .$$

et $\underline{v}$ be a vector in V, written in α-coordinates as $\underline{v} = v_1\underline{a}_1 + \cdots + v_n\underline{a}_n$. Per definition, $\underline{v}$ is eigenvector for eigenvalue λ if and only if $\underline{v} \neq \underline{0}$ and $(\mathcal{A} - \lambda\mathcal{I})\underline{v} = \underline{0}$, so if and only if $(v_1, \dots, v_n) \neq (0, \dots, 0)$ and

$$\begin{pmatrix} a_{11}-\lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22}-\lambda & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{n1} & \dots & \dots & a_{nn}-\lambda \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

This means that there exist eigenvectors for λ if and only the dimension of the solution space of the homogeneous system $(A - \lambda\mathcal{I})\underline{x} = \underline{0}$ is bigger than 0. This is the case if and only if the rank of $A - \lambda\mathcal{I}$ is less than n, which happens if and only if $\det(A - \lambda\mathcal{I}) = 0$.

If λ is an eigenvalue, then the solutions of the system eq. (2.1) give us the coordinate vectors of the elements of E_λ. □

Example 2.1.9 ***(Example 1.3.16, cont'd)*** In example 2.1.5, the continuation of 1.3.16, we saw that the vectors $\underline{a}_1$ and $\underline{a}_2$ are eigenvectors of the considered projection $\mathcal{P}$ belonging to eigenvalues 1 and 0, and that the matrix for $\mathcal{P}$ w.r.t. the basis $\{\underline{a}_1, \underline{a}_2\}$ is

$$P := \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} .$$

We'll now use theorem 2.1.8 to determine all eigenvalues of the projection and their respective eigenspaces. We first use the determinant criterion to find the eigenvalue candidates:

$$\det(P - \lambda\mathcal{I}) = \begin{vmatrix} 1-\lambda & 0 \\ 0 & 0-\lambda \end{vmatrix} = -\lambda(1-\lambda) ,$$

which is 0 if and only if λ is 0 or 1, meaning the only possible eigenvalues are the ones we already found. To determine E_1 in α-coordinates, we solve the corresponding homogeneous solution, plugging in 1 for λ:

$$\begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Leftrightarrow \begin{pmatrix} 0 \\ -v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

So the eigenvectors for eigenvalue 1 are the vectors of the form $\underline{v} = v_1 \cdot \underline{a}_1 + 0 \cdot \underline{a}_2$ for arbitrary scalars v_1. We can put this shorter (and get rid of the coordinate description): $E_1 = \operatorname{span}(\underline{a}_1)$. To determine E_0 in α-coordinates, we do essentially the same thing, plugging in 0 for λ:

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Leftrightarrow \begin{pmatrix} v_1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

yields $\underline{v} = 0 \cdot \underline{a}_1 + v_2 \cdot \underline{a}_2$ for arbitrary scalars v_2, so $E_0 = \operatorname{span}(\underline{a}_2)$.

Example 2.1.10 ***(Example 2.1.9, generalised)*** If you feel nerdy, you can try to generalise the result of example 2.1.9 by proving the following claim: If $\mathcal{P}$ is a projection unto a line ℓ in $\mathbb{R}^2$, then $\mathcal{P}$ has exactly the two eigenvalues 0 and 1, and the eigenspaces are $E_1 = \ell$ and $E_0 = \ell^\perp$. (The general case works because you always can get a basis like in example 2.1.9 by picking one vector from ℓ and one from its complement.)

We translate theorem 2.1.8 into a method to compute eigenspaces (in coordinate form), by means of an 'algorithm':

Algorithm 2.1.11 [Computing the eigenspace for a given eigenvalue, in coordinates]

Input:	`Map` $\mathcal{A}$, `eigenvalue` λ, `basis` α
Output:	`Eigenspace` $E_\lambda(A)$
Step 1:	`Compute matrix` A_α `of` $\mathcal{A}$ `w.r.t` α
Step 2:	$E_\lambda(\mathcal{A}) :=$ `space of solutions to the equation` $(A_\alpha - \lambda\mathcal{I})\underline{v} = 0$:
	`If necessary, row-reduce matrix` $(A_\alpha - \lambda\mathcal{I})$
	`// solution to row-reduced matrix = coordinates of vector in` E_λ.
	`// write set of solutions as linear span like in example 2.1.9.`
	`Find` $\operatorname{span}(\underline{v}_1, \dots, \underline{v}_n)$ `description of solutions set`
Step 3:	`output` $\operatorname{span}(\underline{v}_1, \dots, \underline{v}_n) = E_\lambda(\mathcal{A})$ `in` α`-coordinates`

In theorem 2.1.8, the polynomial $\det(A - \lambda\mathcal{I})$ played a central role in determining possible eigenvalues. As we will develop and use many nice results about this function, it gets its own name:

Definition 2.1.12 (Characteristic polynomial). *Let $\mathcal{A} : V \to V$ be a linear map and let A_α be the matrix of $\mathcal{A}$ w.r.t. a basis α. We call the equation* $\det(A_\alpha - \lambda\mathcal{I}) = 0$ *the* **characteristic equation** *of A_α, and the left-hand side of this equation,* $\det(A_\alpha - \lambda\mathcal{I})$*, the* **characteristic polynomial** *of A_α.*

We will also call these objects the characteristic equation/polynomial of $\mathcal{A}$, and denote the characteristic polynomial by $\chi_{\mathcal{A}}$.

Looking at this definition, you might ask yourself whether the term $\chi_{\mathcal{A}}$ is well-defined - the way we defined it is relative to a basis. ***The following theorem states that our naming convention is okay because $\chi_{\mathcal{A}}$ actually doesn't depend on the choice of basis.***

Theorem 2.1.13. *Let $\mathcal{A} : V \to V$ be a linear map, let α and β be two bases for V, and let A_α/A_β be the matrix of $\mathcal{A}$ w.r.t. α/β. Then* $\det(A_\alpha - \lambda\mathcal{I}) = \det(A_\beta - \lambda\mathcal{I})$.

Proof. We prove the equation by using the 'coordinate-switching relation' between A_α and A_β, rewriting $\mathcal{I}$, and using that the determinant is multiplicative:

$$\begin{aligned}\det(A_\beta - \lambda\mathcal{I}) &= \det({}_\beta S_\alpha A_\alpha {}_\alpha S_\beta - \lambda {}_\beta S_\alpha {}_\alpha S_\beta) \\ &= \det({}_\beta S_\alpha (A_\alpha - \lambda\mathcal{I}) {}_\alpha S_\beta) \qquad = \det({}_\beta S_\alpha)\det(A_\alpha - \lambda\mathcal{I})\det({}_\alpha S_\beta) \\ &= \det(A_\alpha - \lambda\mathcal{I})\det({}_\beta S_\alpha {}_\alpha S_\beta) \quad = \det(A_\alpha - \lambda\mathcal{I})\det(\mathcal{I}) \\ &= \det(A_\alpha - \lambda\mathcal{I}) \ .\end{aligned}$$

□

We now make some useful observations about how the characteristic polynomial looks:

Theorem 2.1.14. *Let $\mathcal{A} : V \to V$ be a linear map on a vector space V of dimension n, and let*

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{n1} & \dots & \dots & a_{nn} \end{pmatrix}$$

be the matrix of $\mathcal{A}$ w.r.t. some basis. Then the characteristic polynomial $\chi_{\mathcal{A}}$ is a polynomial of degree (exactly) n, and of the following shape:

$$\chi_{\mathcal{A}} = (-1)^n \lambda^n + (-1)^{n-1}(a_{11} + a_{22} + \dots + a_{nn})\lambda^{n-1} + \dots + c_1\lambda + c_0$$

for some coefficients $c_0, c_1, \ldots \in \mathbb{K}$.

Proof. We know that the characteristic polynomial $\chi_{\mathcal{A}}$ is the determinant

$$\begin{vmatrix} a_{11} - \lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - \lambda & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{n1} & \dots & \dots & a_{nn} - \lambda \end{vmatrix} .$$

This determinant is the sum of $n!$ many terms, with every summand being a product of n many matrix elements (one from each row and one from each column). We first reason that the degree of $\chi_{\mathcal{A}}$ is at most n: In each summand, λ can pop up at most n many times. Therefore, each summand in this sum is a polynomial in λ of degree at most n.

Now, we look at the terms in $\chi_{\mathcal{A}}$ that have the highest degree. When we look at the product of the elements on the main diagonal, λ pops up in every single factor :

$$(a_{11} - \lambda)(a_{22} - \lambda)\cdots(a_{nn} - \lambda) = (-1)^n\lambda^n + (-1)^{n-1}(a_{11} + a_{22} + \dots + a_{nn})\lambda^{n-1} + \cdots \ .$$

The remaining summands in $\chi_{\mathcal{A}}$ all contain an element a_{ij} with $i \neq j$. In these summands, the diagonal elements $(a_{ii} - \lambda)$ and $(a_{jj} - \lambda)$ therefore do not occur, so what is left are polynomials of degree at most $n-2$.The terms in $\chi_{\mathcal{A}}$ that have degree n and $n-1$ hence are the terms from the main diagonal product, and we see that $\chi_{\mathcal{A}}$ has the claimed shape. □

Definition 2.1.15 (Trace). *The sum of the diagonal elements of a square matrix A is called the* **trace** *of the matrix A. We denote it by* $\operatorname{tr}(A)$.

Using theorem 2.1.14, we now show that trace and determinant are invariant under the choice of basis, and that we can read out trace and determinant of $\mathcal{A}$ from the characteristic polynomial $\chi_{\mathcal{A}}$:

Definition/theorem 2.1.16. *Let $\mathcal{A} : V \to V$ be a linear map with* $\dim(V) < \infty$. *For every basis α, the matrix A_α*

1. *has the same trace, which we therefore also call the* **trace of** $\mathcal{A}$ *and denote by* $\operatorname{tr}(\mathcal{A})$. *We have the identity* $\operatorname{tr}(\mathcal{A}) = c_{n-1}$, *where c_{n-1} is the second-highest coefficient in the characteristic polynomial $\chi_{\mathcal{A}}$.*
2. *has the same determinant, which we therefore also call the* **determinant of** $\mathcal{A}$. *We have the identity* $\det(\mathcal{A}) = c_0$, *where c_0 is the constant coefficient in the characteristic polynomial $\chi_{\mathcal{A}}$.*

Proof. To prove 1., we recall that the trace of a map's matrix A_α is exactly c_{n-1}, the second-highest coefficient of $\chi_{\mathcal{A}}$ (according to theorem 2.1.14). But theorem 2.1.13 already told us that $\chi_{\mathcal{A}}$ does not depend on the choice of basis, so this applies in particular to c_{n-1}.

To prove 2., we plug 0 into $\chi_{\mathcal{A}}$ and get

$$\chi_{\mathcal{A}}(0) = c_0 \ .$$

At the same time,

$$\chi_{\mathcal{A}}(0) = \det(\mathcal{A} - 0\mathcal{I}) = \det(\mathcal{A}) \ .$$

□

To make the previous observation more useful, let's look at the special case where V is a $\mathbb{C}$- or $\mathbb{R}$-vector space. In that case, the characteristic polynomial $\chi_{\mathcal{A}}$ is a polynomial with coefficients in $\mathbb{C}$. Due to the fundamental theorem of algebra, we know that the characteristic polynomial $\chi_{\mathcal{A}}(\lambda)$ then – like any other polynomial – completely factors into linear terms $\lambda - \lambda_i$, where the constant terms λ_i are the zeros (or 'roots') of $\chi_{\mathcal{A}}$. (See theorem B.1.6 and the subsequent discussion if you do not recall this.) We can relate these roots to the trace and the determinant of matrices:

Theorem 2.1.17. *Let A be a square matrix with entries in $\mathbb{K}$, where $\mathbb{K} \in \{\mathbb{C}, \mathbb{R}\}$, with characteristic polynomial χ_A. Then the*

- *trace of the matrix is the sum of the roots of χ_A; and the*
- *determinant of the matrix is the product of the roots of χ_A.*

Proof. Let $\lambda_1,\ldots,\lambda_n$ be the n roots (in $\mathbb{C}$, with multiplicity) of the characteristic polynomial χ_A. Then we have the following identity:

$$\begin{aligned}\chi_A &= (-1)^n(\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_n)\\ &= (-1)^n\lambda^n + (-1)^{n-1}(\lambda_1+\lambda_2+\ldots+\lambda_n)\lambda^{n-1}+\cdots+\lambda_1\lambda_2\cdots\lambda_n\ .\end{aligned}$$

□

Why is this useful? Because it tells us that we can compute eigenvalues of a map $\mathcal{A}$ by simply looking for the roots of $\chi_{\mathcal{A}}$, and taking the roots that are in the right field. We summarise the techniques we saw so far in the 'algorithm' below.

Algorithm 2.1.18 [Computing the eigenvalues and -spaces of a map when $\mathbb{K}=\mathbb{R}$ or $\mathbb{K}=\mathbb{C}$]

Input:	Map $\mathcal{A}$
Output:	Eigenvalues $\lambda_1,\ldots,\lambda_n$, eigenspaces $E_{\lambda_1}(A),\ldots,E_{\lambda_n}(A)$
Step 1:	Pick a basis α
Step 2:	Compute matrix A_α of $\mathcal{A}$ w.r.t α
Step 3:	Determine eigenvalues:
	Compute characteristic polynomial $\chi_{\mathcal{A}}=\det(A_\alpha-\lambda\mathcal{I})$.
	$\lambda_1,\ldots,\lambda_n :=$ the zeros of polynomial $\chi_{\mathcal{A}}$ that are in $\mathbb{K}$.
	// In a real vector space, only the real roots are eigenvalues.
	// In a complex space, every root is an eigenvalue.
Step 4:	for $1\le i\le n$, compute eigenspace $E_{\lambda_i}(\mathcal{A})$:
	Algorithm 2.1.11 yields description of $E_{\lambda_i}(\mathcal{A})$ α-coordinates.
	If necessary, transform coordinates back to vectors in V.
Step 5:	output $\lambda_1,\ldots,\lambda_n$ and $E_{\lambda_1}(\mathcal{A}),\ldots,E_{\lambda_n}(\mathcal{A})$

⚠ There will sometimes be smarter ways to compute eigenvalues/-spaces. Algorithm 2.1.18 simply is a 'sledgehammer' solution that always works.

We will proceed with some application examples for algorithm 2.1.18. But first, we briefly recall why we care about finding eigenvalues/spaces: as seen in theorem 2.1.2, a basis of eigenvectors allows to describe the map via a matrix that is pleasantly simple – the respective matrix is a diagonal matrix that has the eigenvalues on the diagonal. (In the same order as the corresponding eigenvectors are ordered in the basis). This will be reflected in examples 2.1.19, 2.1.21 and 2.1.22 below.

Computing eigenvalues and -spaces: examples

Example 2.1.19 ***(Revisiting projection example 1.3.16)*** We once more consider the projection unto the line $\ell: 2x-3y=0$ in $\mathbb{R}^2$ with the standard inner product. In example 2.1.9, we already determined the eigenvalues, as well as the eigenspaces in coordinates relative to an eigenvector basis. We could now simply translate the coordinates back into standard basis terms using the basis transition matrix ${}_{\varepsilon}S_{\alpha}$.
But this approach cheated in a way - we already started from an eigenvector basis to get the

eigenspaces! We will now practice with algorithm 2.1.18 to practice determining eigenvalues/-spaces without knowing a nice eigenvector basis. Let's pick the standard basis.
In example 1.3.16, we saw that the matrix of this projection (in standard coordinates) is

$$P_\varepsilon = \frac{1}{13}\begin{pmatrix} 9 & 6 \\ 6 & 4 \end{pmatrix} ,$$

the characteristic equation therefore is

$$\begin{vmatrix} \frac{9}{13}-\lambda & \frac{6}{13} \\ \frac{6}{13} & \frac{4}{13}-\lambda \end{vmatrix} = \frac{1}{169}\begin{vmatrix} 9-13\lambda & 6 \\ 6 & 4-13\lambda \end{vmatrix}$$

$$= \frac{1}{169}\Big((9-13\lambda)(4-13\lambda)-36\Big) = \lambda^2-\lambda = 0 .$$

The characteristic equation has two roots: $\lambda_1 = 1$ and $\lambda_2 = 0$. (Nerdy as we are, we make the side observation that the trace of the matrix is $1 = \lambda_1 + \lambda_2$ and that the determinant is $0 = \lambda_1 \cdot \lambda_2$, exactly as promised by theorem 2.1.17.)
We find the eigenspace E_1 for $\lambda_1 = 1$ by analysing the system of equations with coefficient matrix $P_\varepsilon - 1\cdot\mathcal{I}$:

$$\left(\begin{array}{cc|c} \frac{9}{13}-1 & \frac{6}{13} & 0 \\ \frac{6}{13} & \frac{4}{13}-1 & 0 \end{array}\right) \sim \begin{pmatrix} -4 & 6 \\ 6 & -9 \end{pmatrix} \sim \begin{pmatrix} 2 & -3 \\ 0 & 0 \end{pmatrix} .$$

So, the eigenspace E_1 satisfies the equation $2x-3y=0$ (the equation of ℓ, no big surprise). All solutions (x,y) are therefore multiples of $(3,2)$, so

$$E_1 = \text{span}(3\underline{e}_1 + 2\underline{e}_2) \quad (=\ell) .$$

The eigenspace E_0 for $\lambda = 0$ can be found from the system

$$\begin{pmatrix} \frac{9}{13}-0 & \frac{6}{13} \\ \frac{6}{13} & \frac{4}{13}-0 \end{pmatrix} \sim \begin{pmatrix} 9 & 6 \\ 6 & 4 \end{pmatrix} \sim \begin{pmatrix} 3 & 2 \\ 0 & 0 \end{pmatrix}.$$

So, the eigenspace E_0 satisfies the equation $3x+2y=0$, meaning all solutions (x,y) are multiples of $(2,-3)$, so

$$E_0 = \text{span}(2\underline{e}_1 - 3\underline{e}_2) \quad (=\ell^\perp) .$$

We can now build an eigenvector basis α by taking $\underline{a}_1 = 3\underline{e}_1 + 2\underline{e}_2$ and $\underline{a}_2 = 2\underline{e}_1 - 3\underline{e}_2$, and setting $\alpha := \{\underline{a}_1, \underline{a}_2\}$. W.r.t. the eigenvector basis α, the projection's matrix has the diagonal form

$$P_\alpha = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

that has the eigenvalues on the diagonal. Just like the matrix P_ε, matrix P_α has trace 1 and determinant 0 (in accordance with theorems 2.1.16 and 2.1.17).

Example 2.1.20 ***(Rotation)*** We choose in the euclidean plane E^2 an orthonormal basis $\{\underline{e}_1, \underline{e}_2\}$ and consider a rotation by 90°, sending $\underline{e}_1$ to $\underline{e}_2$ and $\underline{e}_2$ to $-\underline{e}_1$ (so $\mathcal{A}\underline{e}_1 = \underline{e}_2$ and $\mathcal{A}\underline{e}_2 = -\underline{e}_1$). W.r.t. basis $\{\underline{e}_1, \underline{e}_2\}$, the respective matrix is

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

The characteristic equation is

$$\begin{vmatrix} -\lambda & -1 \\ 1 & -\lambda \end{vmatrix} = \lambda^2 + 1 = 0 \ ,$$

which has the two roots $\lambda = i$ and $\lambda = -i$. (We furthermore note that the trace is 0 and the determinant is 1.) Since E^2 is a real vector space, neither i nor $-i$ can be an eigenvalue. So we have shown that this rotation has no eigenvectors (with an alternative method to example 2.1.6).

The next example covers the case where we the map does have an eigenvalue, but not enough linearly independent eigenvectors to be brought into diagonal shape.

Example 2.1.21 ***(Polynomial differentiation)*** Let V be the vector space of real polynomials of degree at most two, and let $\mathcal{D}$ be the differentiation map, so $\mathcal{D} : V \to V$ is defined by $\mathcal{A}p = p'$. We pick the basis $\alpha = \{1, x, x^2\}$. We saw in example 1.3.14 that w.r.t. α, $\mathcal{D}$ has the matrix

$$D_\alpha = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix} .$$

The characteristic equation is therefore

$$\begin{vmatrix} -\lambda & 1 & 0 \\ 0 & -\lambda & 2 \\ 0 & 0 & -\lambda \end{vmatrix} = -\lambda^3 = 0 \, .$$

The only root of the characteristic polynomial is $\lambda = 0$ with multiplicity 3; the trace is 0 and the determinant is 0. We find the eigenspace E_0 by solving the homogeneous system with coefficient matrix

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

All solutions to this system of equations (in α-coordinates) are multiples of $(1, 0, 0)$, so

$$E_0 = \text{span}(1 \cdot 1 + 0 \cdot x + 0 \cdot x^2) = \text{span}(1).$$

E_0 therefore consists of the constant polynomials, and indeed, constant polynomials have the zero polynomial as derivative. We note that $\dim(E_0) = 1$, even though $\lambda = 0$ has multiplicity 3. Since $\lambda = 0$ is the only root of the characteristic polynomial (and therefore the only eigenvalue candidate), we cannot even find two linearly independent eigenvectors (let alone 3). Since we do not get an eigenvector basis, the differentiation has no diagonal form.

Example 2.1.22 We define a linear map $\mathcal{A}:\mathbb{R}^3\to\mathbb{R}^3$ via the matrix

$$A=\begin{pmatrix}4&-1&6\\2&1&6\\2&-1&8\end{pmatrix}.$$

In this example, we work w.r.t. the standard basis $\varepsilon=\{\underline{e}_1,\underline{e}_2,\underline{e}_3\}$. In this case, we can identify vectors and coordinates. The characteristic equation is:

$$\begin{vmatrix}4-\lambda&-1&6\\2&1-\lambda&6\\2&-1&8-\lambda\end{vmatrix}=0,\quad\text{so}\quad-\lambda^3+13\lambda^2-40\lambda+36=0.$$

This polyomial factors as $-(\lambda-9)(\lambda-2)^2$. So there are two eigenvalues, $\lambda=9$ and $\lambda=2$, the last one with multiplicity 2. (The trace is indeed $9+2+2=13$ and the determinant is 36. Check this.)
We find the eigenspace E_9 from the system of equations with matrix

$$\begin{pmatrix}-5&-1&6\\2&-8&6\\2&-1&-1\end{pmatrix}\sim\begin{pmatrix}1&0&-1\\0&1&-1\\0&0&0\end{pmatrix}$$

which has as solutions the multiples of $(1,1,1)$. So $E_9=\mathrm{span}((1,1,1))$.
The eigenspace E_2 we find from the equations with matrix

$$\begin{pmatrix}2&-1&6\\2&-1&6\\2&-1&6\end{pmatrix},$$

so E_2 is the plane $2x-y+6z=0$, i.e., $E_2=\mathrm{span}((1,2,0),(0,6,1))$.
We can collect all base vectors into one basis α: $\alpha=\{(1,1,1),(1,2,0),(0,6,1)\}$ is a basis of eigenvectors. W.r.t. eigenvector basis α, we find that the respective matrix is

$$A_\alpha=\begin{pmatrix}9&0&0\\0&2&0\\0&0&2\end{pmatrix},$$

so a diagonal matrix with the eigenvalues on the diagonal (and of course with trace 13 and determinant 36).

2.1.2 Linear independence of eigenvectors

In examples 2.1.19 and 2.1.22, the maps' matrices were diagonalisable because – going over the different eigenvalues – we found a complete basis of eigenvectors. We will now answer the question whether this always works, assuming the eigenspaces gift us with enough vectors (unlike in example 2.1.21).

More concretely, say the dimensions d_λ of the different eigenspaces E_λ add up to the dimension

d of the vector space. Then collecting d_λ many linearly independent vectors per eigenspace E_λ would yield a generating system of size d. This is the right size for us to hope we found a complete eigenvector basis. But maybe eigenvectors belonging to different eigenvalues could be linearly dependent, meaning the generating system has a span of dimension smaller than d and thus does not span the whole vector space?

Luckily, the following theorem tells us that we do not have to worry about this: it shows that ***systems of eigenvectors for mutually different eigenvalues*** $\lambda_1, \ldots, \lambda_n$ ***are always linearly independent, meaning they always form an eigenvector basis (for a subspace of dimension*** n***).*** We will use this result in **??** to give a diagonalisability criterion which allows us to decide when certain maps are diagonalisable.

> **Theorem 2.1.23.** *Let* $\mathcal{A}: V \to V$ *be a linear map and let* $\underline{v}_1, \ldots, \underline{v}_n$ *be eigenvectors of* $\mathcal{A}$ *for* **mutually different** *eigenvalues* $\lambda_1, \ldots, \lambda_n$*. Then* $\underline{v}_1, \ldots, \underline{v}_n$ *are independent.*

Proof. Suppose $\underline{v}_1, \ldots, \underline{v}_n$ are not independent, meaning there is a vector depending on the others. We may assume that $\underline{v}_1$ is dependent of $\underline{v}_2, \ldots, \underline{v}_n$ (if necessary, we simply renumber the vectors).

We will now use the eigenvector properties to show that $\underline{v}_1$ must be $\underline{0}$. This yields a contradiction (because $\underline{v}_1$ is an eigenvector and thus cannot be $\underline{0}$), hence there cannot be any dependent vector and we're done with the proof (as soon as we have shown that $\underline{v}_1$ indeed must be $\underline{0}$).

We first prune $\underline{v}_2, \ldots, \underline{v}_n$ to an independent system $\underline{v}_2, \ldots, \underline{v}_p$ (to accomplish this, we might have to renumber the vectors again). After this, we have $\operatorname{span}(\underline{v}_2, \ldots, \underline{v}_p) = \operatorname{span}(\underline{v}_2, \ldots, \underline{v}_n)$.

Since $\underline{v}_1$ is dependent of $\underline{v}_2, \ldots, \underline{v}_n$ and since $\operatorname{span}(\underline{v}_2, \ldots, \underline{v}_p) = \operatorname{span}(\underline{v}_2, \ldots, \underline{v}_n)$, we know that there exists some linear combination such that

$$\underline{v}_1 = \sum_{i=2}^{p} \alpha_i \underline{v}_i \quad .$$

We now multiply by eigenvalue λ_1:

$$\lambda_1 \underline{v}_1 = \sum_{i=2}^{p} \alpha_i \lambda_1 \underline{v}_i \quad \text{,but at the same time,}$$
$$\lambda_1 \underline{v}_1 = \mathcal{A}\underline{v}_1 = \mathcal{A}(\sum_{i=2}^{p} \alpha_i \underline{v}_i) = \sum_{i=2}^{p} \alpha_i \mathcal{A}\underline{v}_i = \sum_{i=2}^{p} \alpha_i \lambda_i \underline{v}_i \quad ,$$

where the last line used the eigenvector properties of the different vectors. Subtracting the two equations from another, we see that

$$\sum_{i=2}^{p} \alpha_i (\lambda_i - \lambda_1) \underline{v}_i = \underline{0}.$$

Since $\underline{v}_2, \ldots, \underline{v}_p$ are independent and $\lambda_i - \lambda_1 \neq 0$ for $i = 2, \ldots, p$, it follows that $\alpha_i = 0$ for $i = 2, \ldots, p$, meaning $\underline{v}_1 = \underline{0}$. □

2.1.3 Diagonalisation of a square matrix

So far, we have developed methods to represent a linear map by a diagonal matrix, with the starting point being a linear map. But we can also start from a square matrix A, by viewing A as the matrix of a linear map $\mathbb{K}^n \to \mathbb{K}^n$ (using the standard basis ε) and looking for a basis α of eigenvectors (provided such a basis exists). Collecting the eigenvalues in a diagonal matrix D, we obtain a diagonal form of the matrix A: We have

$$\begin{aligned} D &= {}_{\alpha}S_{\varepsilon}\, A\, {}_{\varepsilon}S_{\alpha} \quad \text{, or, equivalently,} \\ A &= {}_{\varepsilon}S_{\alpha}\, D\, {}_{\alpha}S_{\varepsilon} \quad . \end{aligned}$$

This procedure is called ***diagonalising the matrix*** A .

2.2. Using subspaces that are stable under the mapping (invariant subspaces)

What will we do? In the previous section, we saw that if $\alpha = \{\underline{v}_1, \dots, \underline{v}_n\}$ is an eigenvector basis for map $\mathcal{A}$, then A_α has the nice shape

$$A_\alpha = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix} ,$$

where λ_i is the eigenvalue for $\underline{v}_i$, and we learnt how we can find such an eigenvector basis if it exists. But we also learnt from the examples in section 2.1.1 that finding an eigenvector basis is not always possible, meaning we cannot always represent the map by a diagonal matrix. Nonetheless, it is often possible to at least break down the matrix into simpler blocks, which will be the subject of this section.

Intuitively, this section generalises the concept of an eigenspace by making its defining feature a bit more general: per definition, eigenvectors are mapped to a multiple of themselves. (Taking a geometrical perspective, the line spanned by an eigenvector is mapped into itself.) We will generalise this by looking at subspaces whose vectors might not necessarily be mapped unto multiples of ***themselves***, but rather stay in the subspace. We call such subspaces 'invariant' because the map will not move them outside of themselves.

As an example, imagine a rotation about an axis in $\mathbb{R}^3$ – neither will the axis be changed by the rotation, nor will the plane which is perpendicular to the rotation axis. Each of them is hence an invariant subspace. Looking ahead, we will get back to this important example in more detail in chapter 3, and you will re-encounter invariant subspaces in the course on ordinary differential equations.

In more detail, we'll learn

- what invariant subspaces are;
- how we can use them to bring matrices into a simpler form; and
- for linear maps $\mathcal{A} : V \to V$ where V is real and the characteristic equation has a non–real root: how to get an invariant subspace from the root.

Why do we care? If a map has no (or not enough) eigenvectors, it might still have invariant subspaces, which will lead to simpler matrix forms. More concretely, we will see later in the course that we can use this section's techniques to bring certain types of linear maps into diagonal block shape, which will allow us to determine their geometric meaning even if the initial matrix looked super unwieldy.

Like in the section before, ***all maps considered in this section are maps from a finite-dimensional vector space into itself***, i.e., they are maps $\mathcal{A}: V \to V$ with $\dim(V) < \infty$.

We start with the ***central definition of this section: invariant subspaces***.

Definition 2.2.1 (Invariant subspace). *Let W be a subspace of V. W is called* **invariant under linear map** $\mathcal{A}: V \to V$ *if* $\mathcal{A}\underline{w} \in W$ *for all* $\underline{w} \in W$.

Example 2.2.2 ***(Null space and range)***

- The null space $\mathcal{N}$ of a linear map $\mathcal{A}$ is always invariant: if $\underline{x} \in \mathcal{N}$, then $\mathcal{A}\underline{x} = \underline{0}$, and $\underline{0}$ again belongs to $\mathcal{N}$.
- The range $\mathcal{R}$ is also always invariant: if $\underline{y} \in \mathcal{R}$, then $\mathcal{A}\underline{y}$ is obviously again contained in $\mathcal{R}$.

Example 2.2.3 ***(Counterexample: rotation in two-dimension space)*** Like in example 2.1.20, we take as map $\mathcal{A}$ a rotation by 90° in the euclidean plane E^2, so $\mathcal{A}\underline{e}_1 = \underline{e}_2$ and $\mathcal{A}\underline{e}_2 = -\underline{e}_1$. We set $W = \text{span}(\underline{e}_1)$. Since $\mathcal{A}\underline{e}_1 = \underline{e}_2 \notin W$, W is not invariant.

Example 2.2.4 ***(Eigenspaces)*** We will now show eigenspaces are invariant. Say λ is an eigenvalue of a linear map $\mathcal{A}$, and let $\underline{v} \in E_\lambda$. We need to show that $A\underline{v}$ is also in E_λ, which is true because $A(A\underline{v}) = A(\lambda\underline{v}) = \lambda A\underline{v}$.

Fortunately, ***we do not need to check the invariance criterion for each single vector of*** W – it is enough if we check it on a basis of W:

Theorem 2.2.5. *Let $\mathcal{A}: V \to V$ be linear and let $W = \text{span}(\underline{a}_1, \dots, \underline{a}_n)$. W is invariant under $\mathcal{A}$ if and only if $\mathcal{A}\underline{a}_i \in \text{span}(\underline{a}_1, \dots, \underline{a}_n)$ for $i = 1, \dots, n$.*

Proof. If W is invariant, then $\mathcal{A}\underline{w} \in W = \text{span}(\underline{a}_1, \dots, \underline{a}_n)$ for all $\underline{w} \in W$, so this is true in particular for $\underline{w} = \underline{a}_1, \dots, \underline{a}_n$.

Conversely, suppose $\mathcal{A}\underline{a}_i \in W$ for $i = 1, \dots, n$. Now take an arbitrary $\underline{w} \in W$, which we can write as $\underline{w} = w_1\underline{a}_1 + \dots + w_n\underline{a}_n$. Then $\mathcal{A}\underline{w} = w_1\mathcal{A}\underline{a}_1 + \dots + w_n\mathcal{A}\underline{a}_n$. Since every $\mathcal{A}\underline{a}_i \in W$ and W is a linear subspace, $\mathcal{A}\underline{w} \in W$. □

The next definition reflects that for invariant subspaces, we can essentially 'forget' about the rest of the vector space. We call this forgetting about the rest 'restricting':

Definition 2.2.6 (Restriction unto an invariant subspace). *If W is invariant under $\mathcal{A}$, then all image vectors $\mathcal{A}\underline{w}$ with $\underline{w} \in W$ are again in W. So if we restrict $\mathcal{A}$ to W, we obtain a well-defined linear map $W \to W$, the* **restriction of the map** $\mathcal{A}$ *unto W, which we sometimes denote by $\mathcal{A}_{|W}$.*

Invariant spaces give us a simpler matrix shape, because the matrix contains a block for the restriction:

Theorem 2.2.7. *Suppose $\alpha = \{\underline{a}_1, \ldots, \underline{a}_n\}$ is a basis for V such that $W = \operatorname{span}(\underline{a}_1, \ldots, \underline{a}_m)$ is invariant under $\mathcal{A}$. Then the matrix A_α has the following form*

$$A_\alpha = \begin{pmatrix} & * & \ldots & * \\ M_1 & \vdots & & \vdots \\ 0 \ldots 0 & \vdots & & \vdots \\ \vdots\vdots & \vdots & & \vdots \\ 0 \ldots 0 & * & \ldots & * \end{pmatrix},$$

The $m \times m$–matrix M_1 is the matrix of the restriction $\mathcal{A}_{|W} : W \to W$ w.r.t. the basis $\{\underline{a}_1, \ldots, \underline{a}_m\}$.

Proof. For each $i = 1, \ldots, m$, we have that $\mathcal{A}\underline{a}_i \in W = \operatorname{span}(\underline{a}_1, \ldots, \underline{a}_m)$, so for each $i = 1, \ldots, m$, $\mathcal{A}\underline{a}_i$ can be written as the linear combination

$$\mathcal{A}\underline{a}_i = a_{i1}\underline{a}_1 + \cdots + a_{im}\underline{a}_m + 0\underline{a}_{m+1} + \cdots + 0\underline{a}_n \quad .$$

Thinking back to what this tells us about the matrix A_α (see section 1.3.3, the i–th column of A_α consists of the α–coordinates of $\mathcal{A}\underline{a}_i$), we know that the i–th column of A_α is the (vertical) coordinate vector $(a_{i1}, \cdots a_{im}, 0, \cdots, 0)$, so the first m many columns of A_α have only 0s below the m–th row.

The restriction matrix M_1 is exactly the m-by-m-matrix that only concerns the m-dimensional subspace W. It therefore consists of the aforementioned columns, when removing the 0s below the m–th entry (because the vectors $\underline{a}_{m+1}, \ldots, \underline{a}_n$ are not in W and hence do not show up). □

Example 2.2.8 ***(Proving invariance and analysing a map without even knowing its full map description)***
Consider in $\mathbb{R}^4$ the (independent) vectors

$$\underline{a} = (1, -1, 1, -1) \text{ and } \underline{b} = (1, 1, 1, 1).$$

Say we have a linear map $\mathcal{A} : \mathbb{R}^4 \to \mathbb{R}^4$ of which we only know that

$$\mathcal{A}\underline{a} = (4, -6, 4, -6) \text{ and } \mathcal{A}\underline{b} = (4, 6, 4, 6) \quad .$$

Even without knowing the full description of $\mathcal{A}$, we will now show that $W = \operatorname{span}(\underline{a}, \underline{b})$ is invariant and determine a matrix of the restriction unto W $\mathcal{A}_{|W} : W \to W$.
To show the invariance of $\operatorname{span}(\underline{a}, \underline{b})$, we must verify that $\mathcal{A}\underline{a}$ and $\mathcal{A}\underline{b}$ are linear combinations of $\underline{a}$ and $\underline{b}$. We do this by simultaneously solving the systems of equations with columns $\underline{a}, \underline{b}, \mathcal{A}\underline{a}, \mathcal{A}\underline{b}$:

$$\left(\begin{array}{rr|rr} 1 & 1 & 4 & 4 \\ -1 & 1 & -6 & 6 \\ 1 & 1 & 4 & 4 \\ -1 & 1 & -6 & 6 \end{array}\right).$$

After row reduction and deleting zero rows, the system reduces to

$$\left(\begin{array}{cc|cc} 1 & 0 & 5 & -1 \\ 0 & 1 & -1 & 5 \end{array}\right),$$

which tells us that $\mathcal{A}\underline{a} = 5\underline{a} - \underline{b}$ and $\mathcal{A}\underline{b} = -\underline{a} + 5\underline{b}$. So $W = \operatorname{span}(\underline{a}, \underline{b})$ is invariant under $\mathcal{A}$. This also tells us how the matrix of the restriction $\mathcal{A}_{|W} : W \to W$ w.r.t. the basis $\{\underline{a}, \underline{b}\}$ looks (we can simply read it out from the system): the matrix is

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

At a first glance, this might look weird to you – the vector space has 4 dimensions, so how did we end up with a 2-dimensional matrix? This is because the subspace W has only two dimensions, and because the restriction matrix is given with respect to ***the coordinates*** of the two basis vectors $\underline{a}$ and $\underline{b}$. (Remember that the coordinate vector of $\underline{a}$ with respect to the basis $\{\underline{a}, \underline{b}\}$ simply is $\underline{e}_1$, the first unity vector in $\mathbb{R}^2$.)
Using the restriction matrix, we can now even ***determine some eigenvectors without knowing the full map***: Using algorithm 2.1.18, we find that the matrix has eigenvalues 4 and 6. In coordinates, we compute the respective eigenspaces as $\operatorname{span}((1,1))$ (for eigenvalue 4) and $\operatorname{span}((1,-1))$ (for eigenvalue 6). In this basis, the restriction map is simply the diagonal map with the eigenvalues on the diagonal (as usual): it is

$$\begin{pmatrix} 4 & 0 \\ 0 & 6 \end{pmatrix}.$$

We transform the coordinate vectors back into elements of $\mathbb{R}^4$: $\underline{a} + \underline{b} = (2,0,2,0)$ and $\underline{a} - \underline{b} = (0,2,0,2)$. So the eigenvector basis of W is $\{(2,0,2,0), (0,2,0,2)\}$.
We now can simplify the representation of the full map: If we pick any basis α of $\mathbb{R}^4$ such that the first two basis vectors are the eigenvectors $(2,0,2,0)$ and $(0,2,0,2)$, then the full matrix has the shape

$$A_\alpha = \left(\begin{array}{ccc} & * & \dots & * \\ \begin{pmatrix} 4 & 0 \\ 0 & 6 \end{pmatrix} & \vdots & & \vdots \\ 0 \dots 0 & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ 0 \dots 0 & * & \dots & * \end{array}\right),$$

Using theorem 2.2.7, we see that ***the characteristic polynomial of a restriction always divides the characteristic polynomial of the larger map***:

Theorem 2.2.9. *If W is an invariant subspace for the linear map $\mathcal{A} : V \to V$, then $\chi_{\mathcal{A}_{|W}}$, the characteristic polynomial of $\mathcal{A}$'s restriction unto W, $\mathcal{A}_{|W} : W \to W$, is a factor of $\chi_{\mathcal{A}}$, the characteristic polynomial of the map $\mathcal{A} : V \to V$.*

Proof. Let W be an invariant subspace for linear map $\mathcal{A} : V \to V$. Pick a basis $\alpha = \{\underline{a}_1, \dots, \underline{a}_n\}$ for

V in a way such that $\{\underline{a}_1, \dots, \underline{a}_m\}$ is a basis for W. Then theorem 2.2.7 tells us how A_α looks, and subtracting $\lambda\mathcal{I}$ from A_α yields

$$A_\alpha - \lambda\mathcal{I} = \begin{pmatrix} & * & \dots & * \\ M_1 - \lambda\mathcal{I}_m & \vdots & & \vdots \\ 0 \dots 0 & \vdots & & \vdots \\ \vdots & \vdots & M_2 & \vdots \\ 0 \dots 0 & * & \dots & * \end{pmatrix},$$

where M_1 is the matrix of the restriction $\mathcal{A}_{|W}$ and $\mathcal{I}_m$ is the $m \times m$ identity matrix.

I now claim that

$$\det(A_\alpha - \lambda\mathcal{I}) = \det(M_1 - \lambda\mathcal{I}_m)\det(M_2). \tag{2.2}$$

On the left, we now have $\chi_{\mathcal{A}}$, the characteristic polynomial of $\mathcal{A} : V \to V$. The first factor on the right is $\chi_{\mathcal{A}_{|W}}$, the characteristic polynomial of the restriction $\mathcal{A}_{|W} : W \to W$.

The only thing left to do is to prove claim 2.2, which we do via the more general helper lemma 2.2.10 below. □

Lemma 2.2.10. *Let A be a $p \times p$–matrix, and let B be a $q \times q$–matrix B. Then*

$$\det\begin{pmatrix} A & * \\ 0 & B \end{pmatrix} = \det(A) \cdot \det(B).$$

Here, $*$ stands for an arbitrary $p \times q$–matrix, and O stands for the $q \times p$–zero matrix.

Proof. Let us first assume that A is not invertible. Since A does not have full rank (recall Theorem 4.2.25 from LinA 1), neither does the matrix given in the claim, meaning both sides are 0 and the claim is true.

Now let us assume that A is invertible. We will rewrite the matrix given in the claim as a matrix product, and then use that the determinant is multiplicative (recall Theorem 4.2.18 from LinA 1):

$$\begin{pmatrix} A & * \\ 0 & B \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \cdot \begin{pmatrix} \mathcal{I} & A^{-1} \cdot * \\ 0 & \mathcal{I} \end{pmatrix} \text{ and } \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} = \begin{pmatrix} \mathcal{I} & 0 \\ 0 & B \end{pmatrix} \cdot \begin{pmatrix} A & 0 \\ 0 & \mathcal{I} \end{pmatrix},$$

so

$$\det\begin{pmatrix} A & * \\ 0 & B \end{pmatrix} = \det\begin{pmatrix} \mathcal{I} & 0 \\ 0 & B \end{pmatrix} \cdot \det\begin{pmatrix} A & 0 \\ 0 & \mathcal{I} \end{pmatrix} \cdot \det\begin{pmatrix} \mathcal{I} & A^{-1} \cdot * \\ 0 & \mathcal{I} \end{pmatrix} = \det(B) \cdot \det(A) \cdot 1,$$

where the last step used that we can expand across the rows belonging to identity matrices. □

2.2.1 Nice results for combinations of invariant subspaces

The following theorem is a slight extension of theorem 2.2.7. It tells us that ***if V can be broken down into two invariant subspaces for $\mathcal{A}$, we can find a representation matrix with a shape that is even nicer (block diagonal)***, and that ***determinants/characteristic polynomials become products of the respective restriction terms***:

Theorem 2.2.11. *Let $\alpha = \{\underline{a}_1, \ldots, \underline{a}_n\}$ be a basis for V such that $W_1 = \operatorname{span}(\underline{a}_1, \ldots, \underline{a}_m)$ and $W_2 = \operatorname{span}(\underline{a}_{m+1}, \ldots, \underline{a}_n)$ are invariant under $\mathcal{A} : V \to V$. Then the matrix A_α has the form*

$$A_\alpha = \begin{pmatrix} & & 0 & \ldots & 0 \\ & M_1 & \vdots & & \vdots \\ & & 0 & \ldots & 0 \\ 0 \ldots 0 & & & & \\ \vdots \vdots & & & M_2 & \\ 0 \ldots 0 & & & & \end{pmatrix} .$$

Here M_1 and M_2 are the $m \times m$ and $(n-m) \times (n-m)$ matrices of the two restrictions $\mathcal{A}_{|W_1} : W_1 \xrightarrow{A} W_1$ and $\mathcal{A}_{|W_2} : W_2 \xrightarrow{A} W_2$. (With respect to their respective bases $\{\underline{a}_1, \ldots, \underline{a}_m\}$ and $\{\underline{a}_{m+1}, \ldots, \underline{a}_n\}$.)

In addition we have that

$$\det(A_\alpha) = \det(M_1)\det(M_2) \ ,$$

and that the characteristic polynomial of $\mathcal{A}$ is the product of the characteristic polynomials of the two restrictions:

$$\chi_{\mathcal{A}} = \chi_{\mathcal{A}_{|W_1}} \cdot \chi_{\mathcal{A}_{|W_2}} \ .$$

Proof. The claim about the matrix shape can be proven by adapting the proof of theorem 2.2.7. We can now apply lemma 2.2.10 to prove the determinant equation and adapt the proof of theorem 2.2.9 to prove the claim about the characteristic polynomials. □

Remark 2.2.12 We remark that this result can be generalised further such that it holds for an arbitrary number of invariant subspaces: if V can be broken down into invariant subspaces $W_1, \ldots W_p$, we can pick a basis α whose i-th section is a basis for W_i. Let $\mathcal{A}_i : W_i \xrightarrow{A} W_i$ denote the restriction of $\mathcal{A}$ unto the subspace W_i.

Then the matrix A_α has the form

$$A_\alpha = \begin{bmatrix} M_1 & & \\ & \ddots & \\ & & M_p \end{bmatrix} .$$

where M_i is the matrix of the respective restriction $\mathcal{A}_i$ (with respect to the respective basis.)

In that case, the determinant of A_α is the product of all of the restriction matrix determinants, the same goes for the characteristic polynomial.

With this, we also get a result about $\operatorname{spec}(\mathcal{A})$, the set of all eigenvalues of $\mathcal{A}$: you get the eigenvalues of $\mathcal{A}$ by picking all roots of the characteristic polynomial that actually lie in the field. Since the characteristic polynomial of $\mathcal{A}$ is the product of the characteristic polynomials of the restrictions $\mathcal{A}_i$, you get $\mathcal{A}$'s spectrum by collecting the eigenvalues of the restrictions. In short:

$$\operatorname{spec}(\mathcal{A}) = \bigcup_{i=1}^{p} \operatorname{spec}(\mathcal{A}_i) \ .$$

You will re-encounter this in, e.g., Theory and Application of Ordinary Differential Equations. There, the family of restrictions $\{\mathcal{A}_1, \ldots \mathcal{A}_p\}$ will be called a ***decomposition of*** $\mathcal{A}$ ***with respect to the invariant, independent subspaces*** $W_1, \ldots W_p$.

2.2.2 How to get invariant subspaces from non-real roots

For real vector spaces, only the real roots of a characteristic polynomial $\chi_{\mathcal{A}}$ are eigenvalues. In consequence, its non-real roots have been useless so far, as they do not provide us with eigenvectors. We will now see that they are useful nonetheless: ***non-real roots give rise to a two-dimensional invariant subspace, which still helps simplifying the matrix.*** (Because we can at split the matrix into a block belonging to this subspace, and another block.) We will use this again in chapter 3 to prove that linear maps belonging to a certain class will always be a combination of simple geometric building blocks (rotations and reflections).

Theorem 2.2.13 (Two-dimensional invariant subspaces). *Let* $\mathcal{A}$ *be a linear map* $\mathcal{A} : V \to V$ *of a* **real** *finite-dimensional vector space* V *unto itself, and let* μ *be a non-real root of the map's characteristic polynomial* $\chi_{\mathcal{A}}$.

Then μ *gives rise to a two-dimensional invariant subspace* W *of* V*. We find a natural basis of* W *such that the matrix of the respective restriction is*

$$\begin{pmatrix} \mathrm{Re}(\mu) & \mathrm{Im}(\mu) \\ -\mathrm{Im}(\mu) & \mathrm{Re}(\mu) \end{pmatrix} ,$$

The characteristic polynomial of the respective restriction, $\chi_{\mathcal{A}_{|W}}$*, is*

$$\chi_{\mathcal{A}_{|W}} = (\lambda - \mu) * (\lambda - \overline{\mu}) \ .$$

Proof. Fix any basis $\alpha = \{\underline{a}_1, \ldots, \underline{a}_n\}$ for V. We will now view the representation matrix A_α as the matrix of a map $\mathbb{C}^n \to \mathbb{C}^n$. (We can do this since A_α has real coefficients, meaning it can also be viewed as a complex matrix.) Then μ is an eigenvalue of that complex map, meaning we find a complex eigenvector $\underline{z}$, i.e., a vector $\underline{z} \neq \underline{0}$ such that

$$A_\alpha \underline{z} = \mu \underline{z} \ .$$

We now split $\underline{z}$ into a vector $\underline{x}$ for the real part of $\underline{z}$ (so '$\underline{x} = \mathrm{Re}\underline{z}$') and a vector $\underline{y}$ for the complex part of $\underline{z}$ (so '$\underline{y} = \mathrm{Im}\underline{z}$'):

$$\begin{aligned}(x_1, \ldots, x_n) &:= (\mathrm{Re} z_1, \ldots, \mathrm{Re} z_n) \\ (y_1, \ldots, y_n) &:= (\mathrm{Im} z_1, \ldots, \mathrm{Im} z_n) \ .\end{aligned}$$

Since $\underline{x}$ and $\underline{y}$ both only have real entries, we can view them as vectors in V. We will just write $\underline{x}$ and $\underline{y}$ for those, though strictly speaking, we translated α-coordinate vectors back into vectors in V. (So more formally, we are talking about $\alpha^{-1}(\underline{x}) = x_1\underline{a}_1 + \cdots + x_n\underline{a}_n \in V$ and $\alpha^{-1}\underline{y} = y_1\underline{a}_1 + \cdots + y_n\underline{a}_n \in V$.)

We set $W := \mathrm{span}(\underline{x}, \underline{y})$ (so '$W = \mathrm{span}(\mathrm{Re}\underline{z}, \mathrm{Im}\underline{z})$'), and show that W is invariant: We only need to argue that $\mathcal{A}\underline{x}, \mathcal{A}\underline{y}$ can both be written as a (real) linear combination of $\underline{x}$ and $\underline{y}$ (due to theorem 2.2.5).

We achieve this by looking at $A_\alpha \underline{z}$ in the complex space. We compute $A_\alpha \underline{z}$ in two different ways and compare the results: Per definition, $\underline{z} = \underline{x} + i\,\underline{y}$.

- The rules for matrix multiplication give
$$A_\alpha \underline{z} = A_\alpha (\underline{x} + i\,\underline{y}) = A_\alpha \underline{x} + i\,A_\alpha \underline{y} \ .$$
- At the same time, $\underline{z}$ is an eigenvector for eigenvalue μ, hence
$$A_\alpha \underline{z} = \mu\,\underline{z} = \mathrm{Re}(\mu)\ \underline{x} - \ \mathrm{Im}(\mu)\ \underline{y} + i\,\mathrm{Im}(\mu)\ \underline{x} + i\,\mathrm{Re}(\mu)\ \underline{y}\ ,$$
where the last step split μ into real and imaginary and dissolved the brackets.

Comparing the two results (and splitting them into real/imaginary), we get
$$A_\alpha \underline{x} = \ \mathrm{Re}(\mu)\ \underline{x} - \ \mathrm{Im}(\mu)\ \underline{y}, \qquad A_\alpha \underline{y} = \ \mathrm{Im}(\mu)\ \underline{x} + \ \mathrm{Re}(\mu)\ \underline{y}.$$

We have just shown that $\mathcal{A}\underline{x}, \mathcal{A}\underline{y}$ can both be written as a (real) linear combination of $\underline{x}$ and $\underline{y}$:
$$\begin{aligned} \mathcal{A}\underline{x} &= \mathrm{Re}(\mu)\ \underline{x}\ -\ \mathrm{Im}(\mu)\ \underline{y}, \text{ and} \\ \mathcal{A}\underline{y} &= \mathrm{Im}(\mu)\ \underline{x}\ +\ \mathrm{Re}(\mu)\ \underline{y}. \end{aligned}$$

So W is an invariant subspace that has dimension 2 over $\mathbb{R}$ (check this!). With respect to the basis $\{\underline{x}, \underline{y}\}$, the restriction of $\mathcal{A}$ unto W has the following matrix M_W:
$$M_W = \begin{pmatrix} \mathrm{Re}(\mu) & \mathrm{Im}(\mu) \\ -\mathrm{Im}(\mu) & \mathrm{Re}(\mu) \end{pmatrix} \ .$$
The restriction $\mathcal{A}_{|W}$ unto W has μ and $\overline{\mu}$ as roots of the characteristic polynomial $\chi_{\mathcal{A}_{|W}}$: We have
$$\chi_{\mathcal{A}_{|W}}(\lambda) = \det(M_W - \lambda \cdot \mathcal{I}) = \left(\mathrm{Re}(\mu) - \lambda\right)^2 + \mathrm{Im}(\mu)^2 \ .$$
For this to be 0, we need $\mathrm{Re}(\mu) - \lambda \in \{\pm i \cdot \mathrm{Im}(\mu)\}$. In other words, we need $\lambda \in \{\mathrm{Re}(\mu) \pm i \cdot \mathrm{Re}(\mu)\}$, so $\lambda \in \{\mu, \overline{\mu}\}$. $\square$

2.3. Diagonalisability and dealing with incomplete eigenvector bases

Chapter 3

Neat tricks for special linear maps: Orthogonal and symmetric maps

Chapter 4

Another application example: linear differential equations

Appendix A

Notation used in this course

A.1. Table of frequently used notation

Since we use a fair amount of notation, I thought it would be handy to include a reference table where you can look up the most prominently used symbols. Additionally, you can find set- and map- related notions and notation in appendix A.2.

Notation	What does it denote?	Definition in...
$\mathbb{K}$	an arbitrary field	B.1.1 on page 51
$\mathbb{C}$	field of complex numbers	B.1.2 on page 52
$\exp(x+yi)$ or e^{x+yi}	exponentiation by a complex number	B.1.5 on page 53
V	usually a vector space	B.2.1 on page 54
$\mathbb{K}^n$	vector space: n copies of $\mathbb{K}$	B.2.2 on page 55
$\underline{v}$	underlining indicates a vector	
$\operatorname{span}(\underline{v}_1,\dots,\underline{v}_m)$	subspace spanned by $\underline{v}_1,\dots,\underline{v}_m$	B.2.3 on page 55
$(\underline{v},\underline{w})$	the inner product of two vectors $\underline{v}$,$\underline{w}$	B.2.7 on page 55 (standard inn. prod., $\mathbb{K}^n$) B.2.8 on page 55 (general inn. prod.)
$\underline{v}\perp\underline{w}$	$\underline{v}$, $\underline{w}$ are orthogonal	B.2.8 on page 55
V/U	quotient space of V modulo subspace U	B.2.12 on page 57
$\operatorname{Mat}_{\mathbb{K}}(m,n)$	vector space of n-by$m$-matrices	B.3.2 on page 57
${}_\beta S_\alpha$	matrix that transforms α-coords into β-coords	1.3.5 on page 16
$\mathcal{A}_{\mid W}$	restriction of map $\mathcal{A}: V\to V$ unto a subspace W of V so $\mathcal{A}_{\mid W}$ is defined by $w\mapsto Aw$ for $w\in W$	2.2.6 on page 39

A.2. Sets and maps

A.2.1 Sets

Sets are usually specified in one of the following ways:

- ***Listing the elements between curly braces***. For example:

$$\{1,2,3,5\},\quad \{1,2,3,\ldots\},\quad \{1,2,3,5,3\},\quad \{2,\sqrt{3},x^2-1\}.$$

  The dots in the second example mean that we expect the reader to recognize the pattern and complete it: so 4, 5 etc. also belong to this set. Two sets are equal if they contain the same elements, so the first and the third set are equal.

- ***Using a defining property***. Examples:

$$\{x \mid x \text{ is an even integer}\},\quad \{y \mid y \text{ is real and } y<0\}.$$

  To highlight in which set (universe) our elements live, we can alternatively write

$$\{x\in\mathbf{Z} \mid x \text{ even}\},\quad \{y\in\mathbf{R} \mid y<0\}\ .$$

As a reminder, we list some frequently used no(ta)tions.

$\emptyset$	the empty set
$a \in A$	a is an element of the set A
$a \notin A$	a is ***not an element*** of A
$A \subset B$ (sometimes: $B \supset A$ or $A \subseteq B$)	A is a subset of B (or: A is contained in B) i.e. if $a\in A$ then $a\in B$
$A \not\subset B$	A is not a subset of B
$A\cap B := \{x \mid x\in A \text{ and } x\in B\}$	the ***intersection*** of A and B
$A\cup B := \{x \mid x\in A \text{ or } x\in B\}$	the ***union*** of A and B
$A - B := \{x \mid x\in A \text{ and } x\notin B\}$ (sometimes: $A\setminus B$)	the ***complement*** or ***set-theoretic difference*** of A and B
$A\times B := \{(a,b) \mid a\in A,\, b\in B\}$	the ***(Cartesian) product*** of A and B
$A_1\times A_2\times\cdots\times A_n :=$ $\{(a_1,a_2,\ldots,a_n) \mid a_1\in A_1, a_2\in A_2,\ldots,a_n\in A_n\}$	the ***product*** of $A_1, A_2,\ldots,A_n$
$A^n := \{(a_1,a_2,\ldots,a_n) \mid a_1,\ldots,a_n\in A\}$	$A^n = A\times A\times\cdots\times A$

A.2.2 Maps

For sets A and B, a ***map*** from A to B is a rule associating to each element of A exactly one element of B, denoted by $f: A\to B$. The set A is called the ***domain*** of the map, B the ***codomain***. If the elements of B are numbers, then one often uses the more common word ***function*** instead of map. Two functions are equal if they have the same domain, the same codomain and the same value for every element of the domain.

We again list some frequently used notation and definitions:

$f : A \to B$	map with domain A and codomain B
	(other letters are also allowed of course!)
$f(a)$	the element in B that is associated to a
	(called value of f in a or ***image*** of a (under f)
$f : a \mapsto b$	f maps a to b
$f(D) := \{f(d) \mid d \in D\}$	the ***image*** of D, for a subset D of A
$f(A)$	***image*** or ***range*** of f
$f^{-1}(E) := \{a \in A \mid f(a) \in E\}$	the (complete) ***inverse image*** or ***pre-image*** of a subset E of B
(sometimes: $f^{\leftarrow}(E)$)	
$f^{-1}(b) := f^{-1}(\{b\})$	set of all elements that are mapped to b
$f : A \to B$ ***injective***	for all $a, a' \in A$: $f(a) = f(a') \Rightarrow a = a'$;
	equivalently: for all $a, a' \in A$: if $a \neq a'$ then $f(a) \neq f(a')$
$f : A \to B$ ***surjective***	for all $b \in B$ there is a pre-image $a \in A$, i.e. an $a \in A$ with $f(a) = b$
	in other words: $f(A) = B$
$f : A \to B$ ***bijective***	f is injective and surjective
	So it's one-to-one: for all $b \in B$ there is a unique $a \in A$ with $f(a) = b$)

If $f : A \to B$ is a bijection, then for all $b \in B$ there is a unique $a \in A$ with $f(a) = b$. In this case, we can define a map from B to A by the rule: $b \mapsto a$ if $f(a) = b$. This map (which only exists if f is a bijection) is called the ***inverse*** of f and is denoted by f^{-1}. Be aware that the same symbol sometimes also refers to the inverse image (see the table above), but it will always be clear from context how f^{-1} is used.

Appendix B

Prerequisites: Vector spaces as seen in *Linear Algebra 1*

This chapter recaps the most important definitions you saw in *Linear Algebra 1*.

B.1. Fields and complex numbers

Definition B.1.1 (fields). *A field is a set* $\mathbb{K}$ *that contains two special elements,* 0 *and* 1*, together with two binary operations, addition and multiplication* $+, \cdot : \mathbb{K} \times \mathbb{K} \to \mathbb{K}$*, such that addition and multiplication obey the following rules (we omit most of the* $\forall$*-quantifiers, and mostly write* ab *for* $a \cdot b$*):*

i) $a + b = b + a$ — *addition is commutative*

ii) $(a+b)+c = a+(b+c)$ — *addition is associative*

iii) $a + 0 = a$ — 0 *acts as the neutral element for addition*

iv) $\forall a : \exists(-a) : a + (-a) = 0$ — *existence of an inverse for addition*

In short, i-iv express that $(\mathbb{K}, +, 0)$ *is an abelian group.*

v) $ab = ba$ — *multiplication is commutative*

vi) $(ab)c = a(bc)$ — *multiplication is associative*

vii) $1 \cdot a = a$ — 1 *acts as the neutral element for multiplication*

viii) $\forall a \neq 0 : \exists a^{-1} : a \cdot a^{-1} = 1$ — *existence of an inverse for multiplication*

In short, v-viii express that $(\mathbb{K} \setminus \{0\}, \cdot, 1)$ *is an abelian group.*

ix) $(a+b)c = ac + bc$ — *multiplication is distributive over addition*

Examples. The fields we know from high school are $\mathbb{Q}$ (the rational numbers) and $\mathbb{R}$ (the real numbers). In *Linear Algebra 1*, you also encountered $\mathbb{C}$, the complex numbers, that were constructed as follows:

Definition B.1.2 (complex numbers). *We turn the vector space* $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ *into a field, called* $\mathbb{C}$*, by defining* $0 := (0,0)$ *and* $1 := (1,0)$ *and using the following addition and multiplication:*

- $(a,b) + (c,d) := (a+c, b+d)$*, the usual addition in* $\mathbb{R}^2$.
- $(a,b) \cdot (c,d) := (ac - bd, ad + bc)$.

You saw in *Linear Algebra 1* that $\mathbb{C}$ indeed is a field. To briefly recap, it's easy to check that we get additive inverses via $-(a,b) := (-a,-b)$. Less obviously, we also get multiplicative inverses via $(a,b)^{-1} := (\frac{a}{a^2+b^2}, \frac{-b}{a^2+b^2})$.

Setting $i := (0,1)$, we can use that $1 = (1,0)$ to rewrite $(a,b) = a + bi$. We briefly verify that this alternative notation is compatible with how we defined addition and multiplication:

- $(a+bi) + (c+di) = (a+c) + (b+d)i$ and
- $(a+bi)(c+di) = ac + adi + bic + bidi = ac - bd + (ad+bc)i$, where we used that $i^2 = -1$.

We recall some notation that was introduced for complex numbers:

Definition B.1.3. *Let* $z = x + yi$*,* $x, y \in \mathbb{R}$ *be a complex number. We denote*

- *Re* $z := x$ — *the real part of* z
- *Im* $z := y$ — *the imaginary part of* z
- $\bar{z} := x - yi$ — *the complex conjugate of* z
- $|z| := \|(x,y)\| := \sqrt{x^2+y^2}$ — *the absolute value (also 'length' or 'modulus') of* z

It can be easily checked that $\bar{0} = 0$, $\bar{1} = 1$, $\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$ and $\overline{z_1 z_2} = \overline{z_1}\,\overline{z_2}$.

Conjugation does not do anything to a complex number z exactly when z itself is a real number, i.e., when its imaginary part is 0: $z = \bar{z} \Leftrightarrow z \in \mathbb{R}$. Using the last lines, it can also be argued that the conjugation map $z \mapsto \bar{z}$ is an $\mathbb{R}$-automorphism of $\mathbb{C}$.

We furthermore have: $z + \bar{z} = 2 \operatorname{Re} z$, $z - \bar{z} = 2i \operatorname{Im} z$, $z\bar{z} = |z|^2$, and $z^{-1} = \bar{z}/|z|^2$.

We saw that complex numbers can be expressed in polar coordinates:

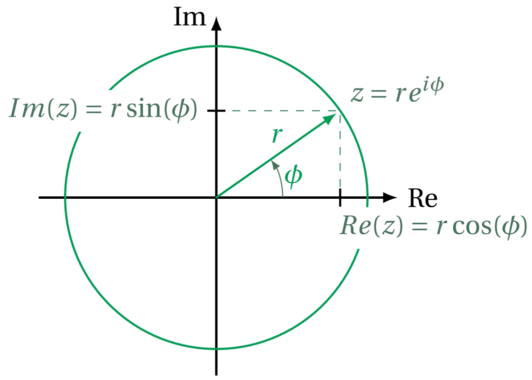
Definition B.1.4 (polar representation). *Every* $z \in \mathbb{C}$ *has a representation of the form*

$$z = r(\cos\phi + i \sin\phi) =: r \ cis\,\phi$$

for some $\phi \in \mathbb{R}$ *and* r *being the absolute value of* z*, i.e.,* $r := |z| \geq 0$.

For $z = 0$*,* ϕ *can be arbitrary. If* $z \neq 0$*, then* ϕ *represents the angle between the positive* x*-axis and the vector* $(x,y) \in \mathbb{R}^2$*. In that case,* ϕ *hence is determined up to a multiple of* 2π.

We call ϕ *the* argument *of* z *and denote it by* $\arg z$*. When we pick a fixed interval of length* 2π *(usually* $[0,2\pi)$ *or* $(-\pi,\pi]$*), exactly one argument* ϕ *lies in this interval. We call this uniquely determined argument the* principal argument.



The polar representation allows us to derive two important rules:

- $|z_1 z_2| = |z_1||z_2|$ — the modulus of the product is the product of the moduli
- $\arg\, z_1 z_2 = \arg\, z_1 + \arg\, z_2$ — the argument of the product is the sum of the arguments

(Proof: Say that $z_1 = r_1 \text{ cis } \alpha$ and $z_2 = r_2 \text{ cis } \beta$. Then

$$\begin{aligned} z_1 z_2 &= r_1 r_2\,((\cos\alpha\cos\beta - \sin\alpha\sin\beta) + i(\cos\alpha\sin\beta + \sin\alpha\cos\beta)) = \\ &= r_1 r_2\,(\cos(\alpha+\beta) + i\sin(\alpha+\beta)) = r_1 r_2 \text{ cis}(\alpha+\beta)\ , \end{aligned}$$

where we used the addition formulas for cosine and sine.)

The function cis behaves very much like the function exp: $\text{cis}\,\alpha \cdot \text{cis}\,\beta = \text{cis}\,(\alpha+\beta)$ and $\exp x \cdot \exp y = \exp(x+y)$. This property allows us to extend the definition of the exponential function to the field $\mathbb{C}$ (and forget about cis altogether).

Definition B.1.5 (exponential representation). *Let $\alpha \in \mathbb{R}$, and let $z = x + yi$ for $x, y \in \mathbb{R}$. We define:*

$$\begin{aligned} \exp(i\alpha) &:= e^{i\alpha} := cis\ \alpha = \cos\alpha + i\sin\alpha \\ \exp(z) &:= e^z = e^{x+yi} := e^x \cdot e^{iy} = e^x \cdot cis\ y = e^x(\cos y + i\sin y). \end{aligned}$$

Since cis y is of length one, we have that $|e^z| = e^x = e^{\text{Re}\ z}$. We also have that $\arg(e^z) = y = \text{Im}\ z$. If z has modulus r and argument ϕ, we can identify $z = r\ e^{i\phi}$.

An important reason why complex numbers play a useful role in the theory of real vector spaces is the following theorem.

Theorem B.1.6 (fundamental theorem of algebra). *Every non-constant polynomial over $\mathbb{C}$ has a zero (or 'root') in $\mathbb{C}$.*

As the name suggests, this theorem is of fundamental importance. One of its consequences is that any polynomial over $\mathbb{C}$ can be neatly written as a product of linear factors: If f is a non-constant polynomial over $\mathbb{C}$, it has a zero $\alpha \in \mathbb{C}$. Then the linear polynomial $z - \alpha$ divides f, i.e. $f(z) = (z-\alpha)g(z)$ for some polynomial g. If g is neither a linear polynomial nor a constant, we can apply the theorem to g, and so forth. Since the degree of the remaining polynomial decreases by one with each repetition, we will at some point end up with a complete factorization into linear factors and one constant factor $\beta \in \mathbb{C}$: If f has degree n, we will end up with a factorisation $f(z) = \beta(z-\alpha_1)(z-\alpha_2)\cdots(z-\alpha_n)$.

Definition/theorem B.1.7 (Conjugates of polynomials). *If $p(z) = p_n z^n + p_{n-1}z^{n-1} + \cdots + p_1 z + p_0$ is polynomial with real coefficients, i.e., if $p_i \in \mathbb{R}$ for all i, then we let*

$$\begin{aligned}\overline{p(z)} &:= \overline{p_n z^n + p_{n-1}z^{n-1} + \cdots + p_1 z + p_0} \\ &\overset{(\star)}{=} p_n \bar{z}^n + p_{n-1}\bar{z}^{n-1} + \cdots + p_1 \bar{z} + p_0 = p(\bar{z}) \ ,\end{aligned}$$

where $(\star)$ used that conjugation is an automorphism of $\mathbb{C}$ that leaves elements of $\mathbb{R}$ unchanged.

As a consequence, we get that any polynomial p with real coefficients can be written as a product of real polynomials that are either linear or quadratic, and a real factor $\beta \in \mathbb{R}$: We know that p splits into the product of linear complex polynomials $(z-\alpha_i)$, where $\alpha_1, \cdots \alpha_n \in \mathbb{C}$ are the zeros of p, and some factor $\beta \in \mathbb{C}$. If a particular zero α is real, the corresponding factor $(z-\alpha)$ is a linear real polynomial, which matches our claim. Let's also take a closer look at the zeros α that are complex, but not real: Given that $p(\alpha) = 0$, we can defer that $\bar{\alpha}$ also is a zero of p since $p(\bar{\alpha}) \overset{B.1.7}{=} \overline{p(\alpha)} = 0$. In other words, we can sort the non-real zeros of p into tuplets $(\alpha, \bar{\alpha})$ of conjugates. For each such tuplet, we obtain that the product $(z-a)(z-\bar{a}) = z^2 - (a+\bar{a})z + a\bar{a} = z^2 - 2\,\mathrm{Re}\, az + |a|^2$ is a quadratic real polynomial, thus matching our claim. We finish by observing that β is real since the highest coefficient of p is real.

B.2. Vectors

Definition B.2.1 (vector space). *Let $\mathbb{K}$ be a field. A $\mathbb{K}$-vector space is a set V that contains a special element, $\underline{0}$, called the zero vector, together with two operations, addition $+ : V \times V \to V$, $(\underline{v}, \underline{w}) \mapsto \underline{v} + \underline{w}$ and scalar multiplication $\cdot : \mathbb{K} \times V \to V$, $(\lambda, \underline{v}) \mapsto \lambda \cdot \underline{v}$ (or simply $\lambda \underline{v}$), such that addition and multiplication obey the following rules (we omit the $\forall$-quantifiers):*

i) $\underline{a} + \underline{b} = \underline{b} + \underline{a}$ — *addition is commutative*

ii) $(\underline{a} + \underline{b}) + \underline{c} = \underline{a} + (\underline{b} + \underline{c})$ — *addition is associative*

iii) $\underline{a} + \underline{0} = \underline{a}$ — *$\underline{0}$ acts as the neutral element for addition*

iv) $\forall \underline{a} : \exists(-\underline{a}) : \underline{a} + (-\underline{a}) = \underline{0}$ — *existence of an inverse for addition*

In short, i-iv express that $(V, +, \underline{0})$ is an abelian group.

v) $\lambda(\underline{a} + \underline{b}) = \lambda\underline{a} + \lambda\underline{b}$ — *multiplication is distributive over vector addition*

vi) $(\lambda + \mu)\underline{a} = \lambda\underline{a} + \mu\underline{a}$ — *multiplication is distributive over scalar addition*

vii) $(\lambda\mu)\underline{a} = \lambda(\mu\underline{a})$ — *associativity of combining field multiplication with scalar multiplication*

In $\mathbb{K}$-vector spaces, we will always have (due to the rules) that $1 \cdot \underline{a} = \underline{a}$, $(-1) \cdot \underline{a} = -\underline{a}$, and $0 \cdot \underline{a} = \underline{0}$.

Definition B.2.2 (the vector space $\mathbb{K}^n$). *By $\mathbb{K}^n$ we denote the set of (ordered) n-tuples of elements of a field $\mathbb{K}$, so vectors $\underline{v} \in \mathbb{K}^n$ are tuples $\underline{v} = (v_1, \ldots, v_n)$ with $v_1, \ldots, v_n \in \mathbb{K}$. We define $\mathbb{K}^n$ as a $\mathbb{K}$-vector space by defining the following rules for addition and scalar multiplication: $\underline{v} + \underline{w} := (v_1 + w_1, \ldots, v_n + w_n)$ and $\lambda \underline{v} := (\lambda v_1, \ldots, \lambda v_n)$. The zero vector in this vector space is $\underline{0} = (0, \ldots, 0)$.*

Definition B.2.3 (subspace and span). *A subset W of V is a (linear) subspace if (i) $\underline{0} \in W$ and (ii) W is closed under addition and scalar multiplication.*

If $\{\underline{v}_1, \ldots, \underline{v}_m\}$ is a set of vectors in V, then we define as the span of (or also the subspace spanned by) $\{\underline{v}_1, \ldots, \underline{v}_m\}$ the set of all linear combinations $\lambda_1 \underline{v}_1 + \cdots + \lambda_m \underline{v}_m$. We denote it by $\operatorname{span}(\underline{v}_1, \ldots, \underline{v}_m)$. *(Another common notation is $< \underline{v}_1, \ldots, \underline{v}_m >$.)*

We define the span of the empty set to be $\{\underline{0}\}$, and the span of an infinite set as the collection of all finite *linear combinations, meaning we allow only a finite number of non-zero coefficients λ.*

Definition B.2.4 (linear (in)dependence). *A linear combination is called* non-trivial *if at least one of its coefficients is non-zero. A set A of vectors in a vector space V is called* dependent *if there is a non-trivial linear combination of vectors in A that equals the zero vector, otherwise the set is called* independent.

On a high level, the usual way to prove independence is this: Start by supposing that $\lambda_1 \underline{v}_1 + \cdots + \lambda_m \underline{v}_m = \underline{0}$, then do some reasoning to show that $\lambda_1 = \cdots = \lambda_m = 0$.

Definition B.2.5 (basis). *If a set of vectors α spans V and is linearly independent, we say that α is a basis for V.*

An important property of vector spaces is that every vector space has a basis (if the basis is infinite, we need some version of the axiom of choice to prove this).

> **Definition/theorem B.2.6 (dimension).** *All bases of a vector space have the same size (cardinality), called the* dimension *of V and denoted by* $\dim(V)$.

Definition B.2.7 (standard inner product of $\mathbb{K}^n$). *Let $\mathbb{K}$ be a field. For two vectors $\underline{a}, \underline{b} \in \mathbb{K}^n$, we denote the standard inner product by $(\underline{a}, \underline{b})$:*

$$(\underline{a}, \underline{b}) = a_1 b_1 + \cdots + a_n b_n.$$

(Other popular notations are $\underline{a} \bullet \underline{b}$ or $\langle \underline{a}, \underline{b} \rangle$.)

Definition B.2.8. *[inner product, orthogonality, length of a vector, angle between vectors] More generally, let V be a real vector space. We say that V together with a map $(.,.) : V \times V \to \mathbb{R}$ is a real inner product space if the following properties are satisfied:*

i) $(\underline{a}, \underline{b}) = (\underline{b}, \underline{a})$ *inner product is symmetric*

ii) inner product is linear:

   1. $(\underline{a} + \underline{b}, \underline{c}) = (\underline{a}, \underline{c}) + (\underline{b}, \underline{c})$, *and*

2. $(\lambda \cdot \underline{a}, \underline{b}) = \lambda \cdot (\underline{a}, \underline{b})$

iii) $(\underline{a}, \underline{a}) \geq 0$ *with equality iff* $\underline{a} = \underline{0}$ *inner product is positive definite*

We say that $\underline{a}$ *and* $\underline{b}$ *are* **orthogonal** *(or* **perpendicular***) and write* $\underline{a} \perp \underline{b}$ *if* $(\underline{a}, \underline{b}) = 0$.

The **length** *of* $\underline{a}$ *is given by* $\|\underline{a}\| = \sqrt{(\underline{a}, \underline{a})}$*, and the angle between non-zero vectors* $\angle(\underline{a}, \underline{b})$ *is given by* $\cos\angle(\underline{a}, \underline{b}) := \frac{(\underline{a}, \underline{b})}{\|\underline{a}\|\|\underline{b}\|}$.

Theorem B.2.9 (theorem of Cauchy-Schwarz).

$$(\underline{a}, \underline{b})^2 \leq (\underline{a}, \underline{a})(\underline{b}, \underline{b}),$$

with equality if and only if the set $\{\underline{a}, \underline{b}\}$ *is dependent (so* $\underline{a} = \underline{0}$*, or* $\underline{b} = \lambda\underline{a}$ *for some* λ*).*

Definition B.2.10 (orthonormal basis). *A basis* α *for a real inner product space* V *is called* orthonormal *if the basis vectors are mutually orthogonal unit vectors, i.e., if*

$$(\underline{a}_i, \underline{a}_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} .$$

If $\{\underline{a}_1, \ldots, \underline{a}_m\}$ is an orthonormal basis of a subspace W, then the orthogonal projection of a vector $\underline{v} \in V$ on the subspace W is given by $\mathcal{P}_W(\underline{v}) = (\underline{v}, \underline{a}_1)\underline{a}_1 + \cdots + (\underline{v}, \underline{a}_m)\underline{a}_m$.

Definition B.2.11 (Gram–Schmidt). *Given a basis* $\alpha = \{\underline{a}_1, \ldots, \underline{a}_n\}$ *for a real inner product space* V*, an orthonormal basis* β *can be computed efficiently as follows: Let* $B_i = \operatorname{span}(\underline{b}_1, \ldots, \underline{b}_i)$ *(so we start with* $B_0 = \{\underline{0}\}$*). For* $i = 1, \ldots, n$*, define*

- $\underline{v}_i := \underline{a}_i - \mathcal{P}_{B_i}(\underline{a}_i)$*, then*
- $\underline{b}_i := \underline{v}_i / \|\underline{v}_i\|$.

In practice, it is more reasonable to work with the projection formula for mutually orthogonal basis vectors and only normalise in the end, so: $\underline{v}_1 = \underline{a}_1$, $\underline{v}_2 = \underline{a}_2 - \dfrac{(\underline{a}_2, \underline{v}_1)}{(\underline{v}_1, \underline{v}_1)}\underline{v}_1$, $\underline{v}_3 = \underline{a}_3 - (\cdots)\underline{v}_1 - (\cdots)\underline{v}_2$, *etc..*

B.2.1 Quotient spaces

From set theory we know that, given an equivalence relation $\sim$ on a set S, the quotient set $S/\sim$ can be formed consisting of the equivalence classes with respect to $\sim$. In the setting of vector spaces, we saw that a linear subspace U of a vector space V can be used to define an equivalence relation and construct a quotient set: We defined the equivalence relation $\sim$ on V as follows:

$$\underline{v}_1 \sim \underline{v}_2 \Leftrightarrow \underline{v}_1 - \underline{v}_2 \in U \ ,$$

and denoted the equivalence class of a vector $v \in V$ by $[v]$.

We saw that with this relation, the quotient set $V/\sim$ turns out to be a vector space:

Definition B.2.12 (Quotient space modulo a subspace). *Let V be a vector space with subspace U. We denote the quotient set $V/\sim$ by V/U.*

V/U becomes a vector space by introducing the following vector space operations: For vectors $\underline{v}_1, \underline{v}_2, \underline{v} \in V/U$ and scalars $\lambda \in \mathbb{K}$, we define

$$+ : [\underline{v}_1] + [\underline{v}_2] := [\underline{v}_1 + \underline{v}_2] \ ,$$
$$\cdot : \lambda \cdot [\underline{v}] := [\lambda \underline{v}] \ .$$

V/U is called it the **quotient space**.

B.3. Matrices

Definition B.3.1 (matrices). *An $m \times n$ matrix A is a rectangular block with m many rows and n many columns, consisting of numbers $a_{ij} \in \mathbb{K}$.*

Abstract and more general: if R and C are sets, then an $R \times C$ matrix A is a map $R \times C \to \mathbb{K}$. (Note the little difference: in the first definition, rows and columns are ordered.)

Definition B.3.2 (addition and scalar multiplication, vector space of matrices). *If A and B are $m \times n$ matrices, then we can define a matrix $S = A + B$ by setting $s_{ij} = a_{ij} + b_{ij}$. For scalars $\lambda \in \mathbb{K}$, we can also define λA as the matrix with ij-entry λa_{ij}.*

For the more abstract definition, where A and B are $R \times C$ matrices, we define $S = A + B$ by setting $S(r, c) = A(r, c) + B(r, c)$ for all $r \in R$ and $c \in C$ and λA as the matrix with (r, c)-entry $\lambda A(r, c)$.

Together with the zero matrix, these operations turn the set of $m \times n$ matrices into a vector space, which we denote by $Mat_{m,n}(\mathbb{K})$ or $Mat_{\mathbb{K}}(m, n)$ or also $\mathbb{K}^{m \times n}$.

Definition B.3.3 (matrix multiplication). *If A is an $m \times k$ matrix, and B is a $k \times n$ matrix, then the product $P = A \cdot B$ is the $m \times n$ matrix defined by*

$$p_{ij} = \sum_{*=1}^{k} a_{i*} b_{*j},$$

More general: if A is an $R \times K$ matrix and B is a $K \times C$ matrix, then $P = A \cdot B$ is the $R \times C$ matrix defined by

$$p(r, c) = \sum_{* \in K} a(r, *) b(*, c).$$

We also saw that matrix multiplication is *associative*: $(AB)C = A(BC)$.

(Proof: $ABC(i, j) = \sum_{kl} a_{ik} b_{kl} c_{li}$.)

Products of vectors and matrices. We usually consider a vector $\underline{v} \in \mathbb{K}^n$ as a column vector, so we identify it with an $n \times 1$ matrix, again denoted $\underline{v}$. In case we want to view it as a row vector (a $1 \times n$ matrix), we write $\underline{v}^\top$. For example, $\underline{v} = (x, y) \in \mathbb{K}^2$ corresponds to the matrix $\underline{v} = \begin{pmatrix} x \\ y \end{pmatrix}$, with $\underline{v}^\top = (x \ \ y)$.

This way, we can also define products of vectors and an $m \times n$ matrix A: we can define $\underline{u}^\top A$ and $A\underline{v}$ for $\underline{u} \in \mathbb{K}^m$, $\underline{v} \in \mathbb{K}^n$.

Appendix C

Additional helpers

C.1. Identifying Greek letters

In mathematics we use many letters from the Greek alphabet. If you feel uncertain which one is which, you can look up the letter in the table below that lists the Greek alphabet. (Letters that are frequently used in Linear Algebra are highlighted in green and indicated with a *.)

name	**m**inuscule (= lowercase letter)	**c**apital (= uppercase letter)
alpha	α *	A
beta	β *	B
gamma	γ *	Γ
delta	δ *	Δ
epsilon	ε	E
zeta	ζ	Z
eta	η	H
theta	θ or ϑ	Θ
iota	ι	I
kappa	κ	K
lambda	λ *	Λ
mu	μ *	M
nu	ν	N
xi	ξ	Ξ
omikron	o	O
pi	π	Π
rho	ρ *	R
sigma	σ *	Σ
tau	τ *	T
upsilon	υ	Υ
phi	ϕ or φ *	Φ
chi	χ	X
psi	ψ *	Ψ
omega	ω *	Ω

C.2. Geometric shapes: ellipse, hyperbola, parabola

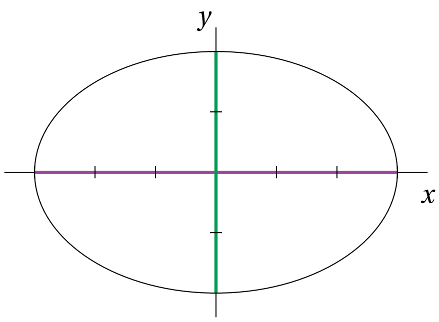
Ellipse

The standard equation for an ***ellipse*** with center (c_1, c_2) is $\frac{(x-c_1)^2}{a^2} + \frac{(y-c_2)^2}{b^2} = 1$, where the denominators a, b are real (positive) numbers. Every ellipse has two axes of symmetry, the longer axis one is called ***major axis*** and the shorter one is called ***minor axis***. The center is the midpoint of both the major and minor axes. The denominators a, b express the distance between the center and the outmost points of the ellipse on the two axes.

Reading out the axis lengths from the standard equation: Since a, b express the distance between the center and the outmost points of the ellipse on the two axes, the length of the major axis is $2 \cdot \max\{a, b\}$ and the length of the minor axis is $2 \cdot \min\{a, b\}$.

Determining major and minor axis, origin center case: If the ellipsis' center is the origin, the two axes of symmetry simply are the x-axis and the y-axis. We can determine which one is major/minor by intersecting the two axes with the ellipse, i.e., by plugging either $x = 0$ or $y = 0$ into the left-hand side of the standard equation, and seeing in which case the remaining denominator is bigger/smaller, thus indicating bigger/smaller axis length.

E.g., take the ellipse $\frac{1}{9}x^2 + \frac{1}{4}y^2 = 1$: Plugging 0 in for x yields $\frac{1}{4}y^2$ and plugging in 0 for y yields $\frac{1}{9}x^2$. This means that the outmost points on the y-axis are $(0, -2)$ and $(0, 2)$, and that the outmost points on the x-axis are $(-3, 0)$ and $(3, 0)$. Thus, the major axis is $y = 0$ (the x-axis) and the minor axis is $x = 0$ (the y-axis).



Ellipse $\frac{1}{9}x^2 + \frac{1}{4}y^2 = 1$ with center $(0, 0)$, major axis $y = 0$ and minor axis $x = 0$.

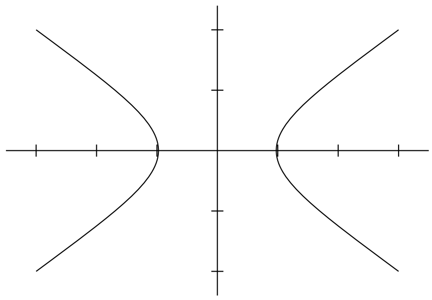
Determining major and minor axis, general case: The case where the center isn't the origin back can be brought back to the origin center case by substituting x and y by $x' := x - c_1$ and $y' := y - c_2$. The minor/major axis can then be determined in (x', y')-coordinates using the reasoning above, and afterwards can be translated to (x, y)-coordinates by substituting back.

E.g., assume you have an ellipse $\frac{(x-1)^2}{a^2} + \frac{(y-2)^2}{b^2} = 1$, so the center is $(1, 2)$. Say you substituted $x' := x - 1$ and $y' := y - 2$ and determined that the major axis in (x', y')-coordinates is $y' = 0$ since a is bigger than b. Then re-substituting yields the major axis in the orginal coordinates: $y - 2 = y' = 0$, so the major axis is $y = 2$.

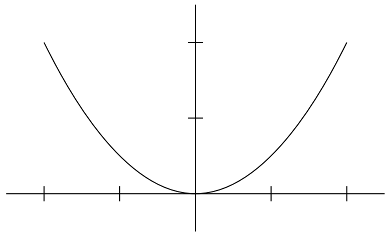
Hyperbola and parabola

The standard equation for a *h*yperbola is $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$, where a, b are real (positive) numbers. For hyperbolas in standard shape, the axes symmetry are the x- and the y-axis.

The standard equation for a *p*arabola is $y = ax^2$, where a is a non-zero real number. For parabolas in standard shape, the axis of symmetry is the y-axis.



The hyperbola $x^2 - 2y^2 = 1$.



The parabola with equation $y = \frac{1}{2}x^2$.

C.3. Trigonometric formulas

This section recalls some useful identities concerning the trigonometric functions:

- $\cos^2(x) + \sin^2(x) = 1$;
- $\sin(x + 2\pi) = \sin(x)$ and $\cos(x + 2\pi) = \cos(x)$;
- $\sin(\pi - x) = \sin(x)$ and $\cos(\pi - x) = -\cos(x)$;
- $\sin(\pi + x) = -\sin(x)$ and $\cos(\pi + x) = -\cos(x)$;
- $\sin(\pi/2 - x) = \cos(x)$ and $\cos(\pi/2 - x) = \sin(x)$;
- $\sin(2x) = 2\sin(x)\cos(x)$ and $\cos(2x) = \cos^2(x) - \sin^2(x)$;
- $\sin(x + y) = \sin(x)\cos(y) + \cos(x)\sin(y)$ and $\cos(x + y) = \cos(x)\cos(y) - \sin(x)\sin(y)$.

Index

axis of rotation, 54

Cauchy-Schwarz inequality, 80
characteristic equation, 30
characteristic polynomial, 30
complex numbers, 76
    exponential function, 77
    polar representation, 76
composition, 5
coordinate transformation, 16
coordinates, 15
coset, 7

decoupled system, 64
diagonal form, 26

eigenspace, 27
eigenvalue, 27
    algebraic and geometric multiplicity, 43
eigenvector, 27
    generalised eigenvector, 45
ellipse, 83

field, 75
    $\mathbb{C}$, 76
fundamental theorem of algebra, 77

Gram-Schmidt, 80

hyperbola, 84

identity, 4
identity map, 4
imaginary numbers, 76
isomorphism, 3

kernel, 6

linear (in)dependence, 79
linear map, 3
    classification of orthogonal, 52
    determinant, 32
    matrix of, 11, 20
    orthogonal, 48
    restriction unto invariant subspace, 39
    symmetric, 55
    trace, 31
linear subspace
    invariant, 38
    invariant subspace from non-real root, 42

maps
    domain, codomain, 73
    image, range, inverse image (or pre-image), injectivity, surjectivity, bijectivity, 74
    inverse of a bijection (one-to-one map), 74
matrices, 81
    addition, scalar multiplication, 81
    multiplication, 81
matrix
    diagonal form, 26
    diagonalise, 37
    orthogonal, 50
    trace, 31
    transition matrix, 16
matrix map, 4
mirrorplane, 54

null space of a linear map, 6

orthogonal map
    direct, 52
    indirect, 52
orthogonal projection, 3
orthonormal basis, 80

parabola, 84
polarisation formula, 49
power of a map, 5

quotient space, 81

range of a linear map, 6
rotation, 52, 54
rotoreflection, 54

scalar multiple of a map, 5
sets
    intersection, union, complement/set-theoretic difference, (Cartesian) product, 73
stelsel differentiaalvergelijkingen, 63
sum of two maps, 5
superposition principle, 68
system differential equations
    homogeneous, 65
    inhomogeneous, 65

vector space, 78
    $\mathbb{K}^n$, 79
    $\mathrm{Mat}_{\mathbb{K}}(m,n)$, 81
    angle between vectors, 80
    dimension, 79
    inner product, 79
    length of a vector, 80
    linar span, 79
    orthogonality/perpendicularity, 80
    standard inner product of $\mathbb{K}^n$, 79
    subspace, 79

zero map, 4