LECTURE NOTES

DEPARTMENT OF COMPUTING

IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

Mathematical Methods (CO-145)

Instructors:
Marc Deisenroth, Mahdi Cheraghchi

Imperial College London

November 18, 2015

Contents

Chapter 1

Complex Numbers

This chapter is largely based on Jeremy Bradley's lecture notes from 2014.

1.1 Introduction

We can see need for complex numbers by looking at the shortcomings of all the simpler (more obvious) number systems that preceded them. In each case the next number system in some sense fixes a perceived problem or omission with the previous one:

- $\mathbb{N}$ Natural numbers, for counting, not closed under subtraction
- $\mathbb{Z}$ Integers, the natural numbers with 0 and negative numbers, not closed under division
- $\mathbb{Q}$ Rational numbers, closed under arithmetic operations but cannot represent the solution of all non-linear equations, e.g., $x^2 = 2$
- $\mathbb{R}$ Real numbers, solutions to some quadratic equations with real roots and some higher-order equations, but not all, e.g., $x^2 + 1 = 0$
- $\mathbb{C}$ Complex numbers, we require these to represent **all** the roots of all polynomial equations.[1]

Another important use of complex numbers is that often a real problem can be solved by mapping it into complex space, deriving a solution, and mapping back again: a direct solution may not be possible or would be much harder to derive in real space, e.g., finding solutions to integration or summation problems, such as

$$I = \int_0^x e^{a\theta} \cos b\theta \, \mathrm{d}\theta \qquad \text{or} \qquad S = \sum_{k=0}^{n} a^k \cos k\theta . \tag{1.1}$$

[1] Complex numbers form an algebraically closed field, where any polynomial equation has a root.

1.1.1 Applications

Complex numbers are important in many areas. Here are some:

- Signal analysis (e.g., Fourier transformation to analyze varying voltages and currents)
- Control theory (e.g., Laplace transformation from time to frequency domain)
- Quantum mechanics is founded on complex numbers (see Schrödinger equation and Heisenberg's matrix mechanics)
- Cryptography (e.g., finding prime numbers).
- Machine learning: Using a pair of uniformly distributed random numbers (x, y), we can generate random numbers in polar form $(r\cos(\theta), r\sin(\theta))$. This can lead to efficient sampling methods like the Box-Muller transform (Box and Muller, 1958).[2] The variant of the Box-Muller transform using complex numbers was proposed by Knop (1969).

1.1.2 Imaginary Number

An entity we cannot describe using real numbers are the roots to the equation

$$x^2 + 1 = 0, \tag{1.2}$$

which we will call i and define as

$$i := \sqrt{-1}. \tag{1.3}$$

There is no way of squeezing this into $\mathbb{R}$, it cannot be compared with a real number (in contrast to $\sqrt{2}$ or π, which we can compare with rationals and get arbitrarily accurate approximations in the rationals). We call i the **imaginary number/unit**, orthogonal to the reals.

Properties From the definition of i in (1.3) we get a number of properties for i.

1. $i^2 = -1, \quad i^3 = i^2 i = -i, \quad i^4 = (i^2)^2 = (-1)^2 = 1$ and so on
2. In general $i^{2n} = (i^2)^n = (-1)^n, \quad i^{2n+1} = i^{2n} i = (-1)^n i$ for all $n \in \mathbb{N}$
3. $i^{-1} = \frac{1}{i} = \frac{i}{i^2} = -i$
4. In general $i^{-2n} = \frac{1}{i^{2n}} = \frac{1}{(-1)^n} = (-1)^n, \quad i^{-(2n+1)} = i^{-2n} i^{-1} = (-1)^{n+1} i$ for all $n \in \mathbb{N}$
5. $i^0 = 1$

[2]This is a pseudo-random number sampling method, e.g., for generating pairs of independent, standard, normally distributed (zero mean, unit variance) random numbers, given a source of uniformly distributed random numbers.

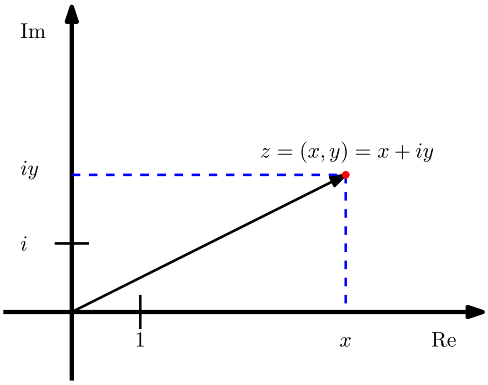


Figure 1.1: Complex plane (Argand diagram). A complex number can be represented in a two-dimensional Cartesian coordinate system with coordinates x and y. x is the real part and y is the imaginary part of a complex number $z = x + iy$.

1.1.3 Complex Numbers as Elements of $\mathbb{R}^2$

It is convenient (and correct[3]) to consider complex numbers

$$\mathbb{C} := \{a + ib : a, b \in \mathbb{R}, i^2 = -1\} \tag{1.4}$$

as the set of tuples $(a, b) \in \mathbb{R}^2$ with the following definition of addition and multiplication:

$$(a, b) + (c, d) = (a + c, b + d), \tag{1.5}$$
$$(a, b) \cdot (c, d) = (ac - bd, ad + bc). \tag{1.6}$$

In this context, the element $i := (0, 1)$ is the **imaginary number/unit**. With the complex multiplication defined in (1.6), we immediately obtain

$$i^2 = (0, 1)^2 = (0, 1)(0, 1) = -1, \tag{1.7}$$

which allows us to factorize the polynomial $z^2 + 1$ fully into $(z - i)(z + i)$.
Since elements of $\mathbb{R}^2$ can be drawn in a plane, we can do the same with complex numbers $z \in \mathbb{C}$. The plane is called **complex plane** or **Argand diagram**, see Figure 1.1.
The Argand diagram allows us to visualize addition and multiplication, which are defined in (1.5)–(1.6).

1.1.4 Closure under Arithmetic Operators

Closing $\mathbb{R} \cup \{i\}$ under the arithmetic operators $+, \cdot$ as defined in (1.5)–(1.6) gives the *complex numbers*, $\mathbb{C}$. To be more specific, if $z_1, z_2 \in \mathbb{C}$, then $z_1 + z_2 \in \mathbb{C}$, $z_1 - z_2 \in \mathbb{C}$, $z_1 \cdot z_2 \in \mathbb{C}$ and $z_1/z_2 \in \mathbb{C}$.

[3]There exists a bijective linear mapping (isomorphism) between $\mathbb{C}$ and $\mathbb{R}^2$. We will briefly discuss this in the Linear Algebra part of the course.

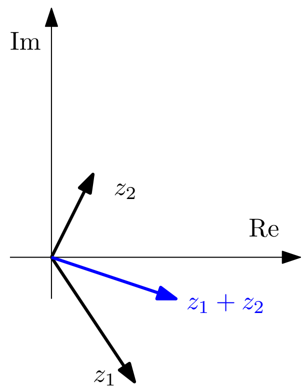


Figure 1.2: Visualization of complex addition. As known from geometry, we simply add the two vectors representing complex numbers.

1.2 Representations of Complex Numbers

In the following, we will discuss three important representations of complex numbers.

1.2.1 Cartesian Coordinates

Every element $z \in \mathbb{C}$ can be decomposed into

$$(x, y) = (x, 0) + (0, y) = (x, 0) + (0, 1)(y, 0) = \underbrace{(x, 0)}_{\in\mathbb{R}} + i \underbrace{(y, 0)}_{\in\mathbb{R}} = x + iy. \tag{1.8}$$

Therefore, every $z = x + iy \in \mathbb{C}$ has a **coordinate representation** (x, y), where x is called the **real part** and y is called the **imaginary part** of z, and we write $x = \mathfrak{Re}(z)$, $y = \mathfrak{Im}(z)$, respectively. $z = x + iy$ is the point (x, y) in the xy-plane (complex plane), which is uniquely determined by its Cartesian coordinates (x, y). An illustration is given in Figure 1.1.

1.2.2 Polar Coordinates

Equivalently, (x, y) can be represented by **polar coordinates**, r, ϕ, where r is the distance of z from the origin 0, and ϕ is the angle between the (positive) x-axis and the direction $\overrightarrow{0z}$. Then,

$$z = r(\cos\phi + i\sin\phi), \quad r \geq 0, \quad 0 \leq \phi < 2\pi \tag{1.9}$$

uniquely determines $z \in \mathbb{C}$. The polar coordinates of z are then

$$r = |z| = \sqrt{x^2 + y^2}, \tag{1.10}$$

$$\phi = \operatorname{Arg} z, \tag{1.11}$$

where r is the length of $\overrightarrow{0z}$ (the distance of z from the origin) and ϕ is the **argument** of z.

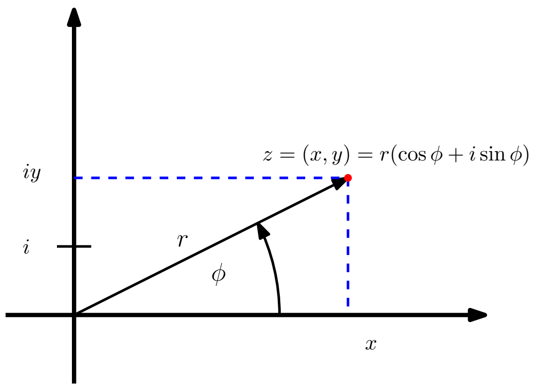


Figure 1.3: Polar coordinates.

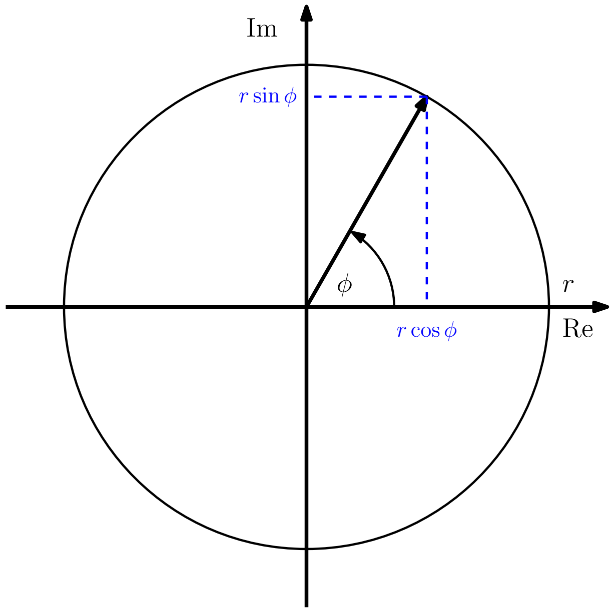


Figure 1.4: Euler representation. In the Euler representation, a complex number $z = r\exp(i\phi)$ "lives" on a circle with radius r around the origin. Therefore, $r\exp(i\phi) = r(\cos\phi + i\sin\phi)$.

1.2.3 Euler Representation

The third representation of complex numbers is the **Euler representation**

$$z = r\exp(i\phi) \tag{1.12}$$

where r and ϕ are the polar coordinates. We already know that $z = r(\cos\phi + i\sin\phi)$, i.e., it must also hold that $r\exp(i\phi) = r(\cos\phi + i\sin\phi)$. This can be proved by looking at the power series expansions of exp, sin, and cos:

$$\exp(i\phi) = \sum_{k=0}^{\infty}\frac{(i\phi)^k}{k!} = 1 + i\phi + \frac{(i\phi)^2}{2!} + \frac{(i\phi)^3}{3!} + \frac{(i\phi)^4}{4!} + \frac{(i\phi)^5}{5!} + \cdots \tag{1.13}$$

$$= 1 + i\phi - \frac{\phi^2}{2!} - \frac{i\phi^3}{3!} + \frac{\phi^4}{4!} + \frac{i\phi^5}{5!} \mp \cdots \tag{1.14}$$

$$= \left(1 - \frac{\phi^2}{2!} + \frac{\phi^4}{4!} \mp \cdots\right) + i\left(\phi - \frac{\phi^3}{3!} + \frac{\phi^5}{5!} \mp \cdots\right) \tag{1.15}$$

$$= \sum_{k=0}^{\infty} \frac{(-1)^k \phi^{2k}}{(2k)!} + i \sum_{k=0}^{\infty} \frac{(-1)^k \phi^{2k+1}}{(2k+1)!} = \cos\phi + i\sin\phi\,. \tag{1.16}$$

Therefore, $z = \exp(i\phi)$ is a complex number, which lives on the unit circle ($|z| = 1$) and traces out the unit circle in the complex plane as ϕ ranges through the real numbers.

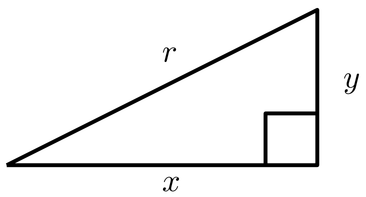
1.2.4 Transformation between Polar and Cartesian Coordinates

Cartesian coordinates ⟷ **Polar coordinates**

x, y $\qquad$ r, ϕ

$x = r\cos\phi$

$y = r\sin\phi$



$r = \sqrt{x^2 + y^2}$

$\tan\phi = \frac{y}{x}$ + quadrant

$$z = x + iy = r(\cos\phi + i\sin\phi)$$

Figure 1.5: Transformation between Cartesian and polar coordinate representations of complex numbers.

Figure 1.5 summarizes the transformation between Cartesian and polar coordinate representations of complex numbers z. We have to pay some attention when computing $\operatorname{Arg}(z)$ when transforming Cartesian coordinates into polar coordinates.

Example: Transformation from Polar to Cartesian Coordinates

Transform the polar representation $z = (r, \phi) = (2, \frac{2\pi}{3})$ into Cartesian coordinates (x, y).

It is always useful to draw the complex number. Figure 1.6(a) shows the setting. We are interested in the blue dots. With $x = r\cos\phi$ and $y = r\sin\phi$, we obtain

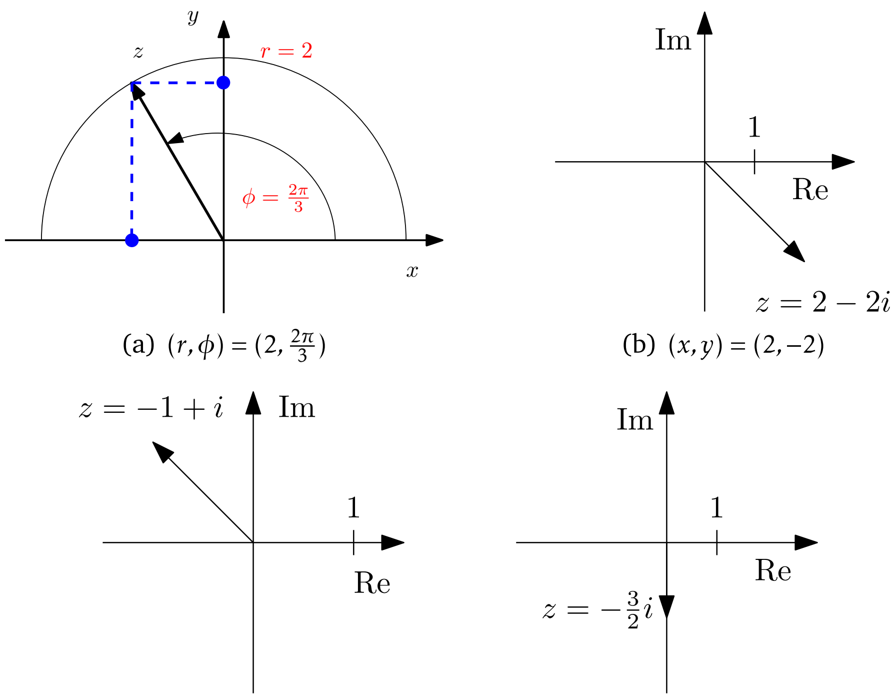
$$x = r\cos(\tfrac{2}{3}\pi) = -1 \tag{1.17}$$

$$y = r\sin(\tfrac{2}{3}\pi) = \sqrt{3}\,. \tag{1.18}$$

Therefore, $z = -1 + i\sqrt{3}$.

Example: Transformation from Cartesian to Polar Coordinates

Getting the Cartesian coordinates from polar coordinates is straightforward. The transformation from Cartesian to polar coordinates is somewhat more difficult because of the argument ϕ. The reason is that tan has a period of π, which means



(a) $(r, \phi) = (2, \frac{2\pi}{3})$ (b) $(x, y) = (2, -2)$

(c) $(x, y) = (-1, 1)$ (d) $(x, y) = (0, -\frac{3}{2})$

Figure 1.6: Coordinate transformations

that y/x has two possible angles, which differ by π, see Figure 1.7. By looking at the quadrant in which the complex number z lives we can resolve this ambiguity. Let us have a look at some examples:

1. $z = 2 - 2i$. We immediately obtain $r = \sqrt{2^2 + 2^2} = 2\sqrt{2}$. For the argument, we obtain $\tan\phi = -\frac{2}{2} = -1$. Therefore, $\phi \in \{\frac{3}{4}\pi, \frac{7}{4}\pi\}$. We identify the correct argument by plotting the complex number and identifying the quadrant. Figure 1.6(b) shows that z lies in the fourth quadrant. Therefore, $\phi = \frac{7}{4}\pi$.

2. $z = -1 + i$.

$$r = \sqrt{1 + 1} = \sqrt{2} \tag{1.19}$$

$$\tan\phi = \frac{-1}{1} = -1 \quad \Rightarrow \phi \in \{\tfrac{3}{4}\pi, \tfrac{7}{4}\pi\}. \tag{1.20}$$

Figure 1.6(c) shows that z lies in the second quadrant. Therefore, $\phi = \frac{3}{4}\pi$.

3. $z = -\frac{3}{2}i$.

$$r = \tfrac{3}{2} \tag{1.21}$$

$$\tan\phi = \frac{-\frac{3}{2}}{0} \quad \Rightarrow \phi \in \{\frac{\pi}{2}, \frac{3}{2}\pi\} \tag{1.22}$$

Figure 1.6(d) shows that z is between the third and fourth quadrant (and not between the first and second). Therefore, $\phi = \frac{3}{2}\pi$

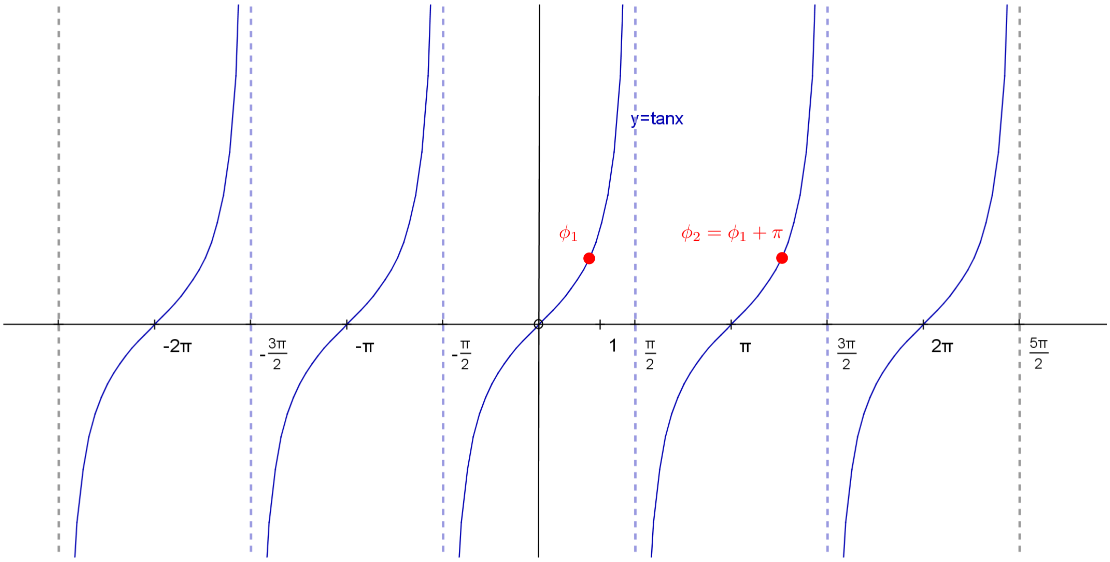


Figure 1.7: Tangens. Since the tangens possesses a period of π, there are two solutions for the argument $0 \le \phi < 2\pi$ of a complex number, which differ by π.

1.2.5 Geometric Interpretation of the Product of Complex Numbers

Let us now use the polar coordinate representation of complex numbers to geometrically interpret the product $z = z_1 z_2$ of two complex numbers z_1, z_2. For $z_1 = r_1(\cos\theta_1 + i\sin\theta_1)$ and $z_2 = r_2(\cos\theta_2 + i\sin\theta_2)$ we obtain

$$
\begin{aligned}
z_1 z_2 &= r_1 r_2(\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2 + i(\sin\theta_1\cos\theta_2 + \cos\theta_1\sin\theta_2)) \\
&= r_1 r_2(\cos(\theta_1+\theta_2) + i\sin(\theta_1+\theta_2)).
\end{aligned}
\tag{1.23}
$$

1. The length $r = |z| = |z_1||z_2|$ is the *product* of the lengths of z_1 and z_2.
2. The argument of z is the *sum* of the arguments of z_1 and z_2.

This means that when we multiply two complex numbers z_1, z_2, the corresponding distances r_1 and r_2 are multiplied while the corresponding arguments ϕ_1, ϕ_2 are summed up. This means, we are now ready to visualize complex multiplication, see Figure 1.8. Overall, multiplying z_1 with z_2 performs two (linear) transformations on z_1: a scaling by r_2 and a rotation by ϕ_2. Similarly, the transformations acting on z_2 are a scaling by r_1 and a rotation by ϕ_1.

1.2.6 Powers of Complex Numbers

We will encounter situations where we need to compute powers of complex numbers of the form z^n. For this, we can use some advantages of some representations of complex numbers. For instance, if we consider the representation using Cartesian coordinates computing $z^n = (x+iy)^n$ for large n will be rather laborious. However, the Euler representation makes our lives a bit easier since

$$
z^n = (r\exp(i\phi))^n = r^n \exp(in\phi) \tag{1.24}
$$

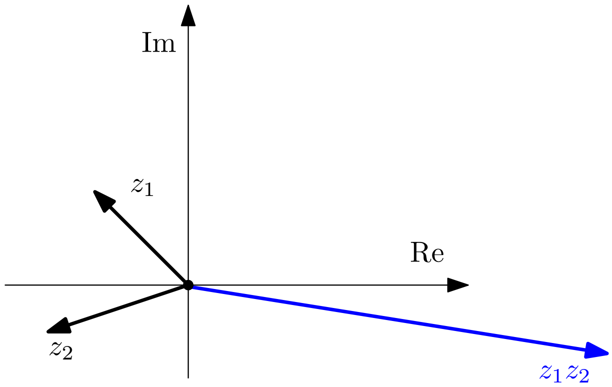


Figure 1.8: Complex multiplication. When we multiply two complex numbers z_1, z_2, the corresponding distances r_1 and r_2 are multiplied while the corresponding arguments ϕ_1, ϕ_2 are summed up.

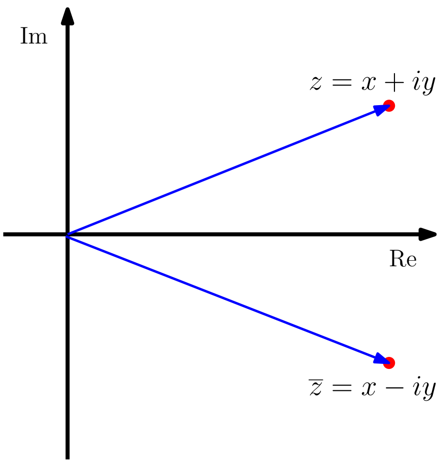


Figure 1.9: The complex conjugate $\overline{z}$ is a reflection of z about the real axis.

can be computed efficiently: The distance r to the origin is simply raised to the power of n and the argument is scaled/multiplied by n. This also immediately gives us the result

$$(r(\cos\phi + i\sin\phi))^n = r^n(\cos(n\phi) + i\sin(n\phi)) \tag{1.25}$$

which will later (Section 1.4) know as de Moivre's theorem.

1.3 Complex Conjugate

The **complex conjugate** of a complex number $z = x + iy$ is $\overline{z} = x - iy$. Some properties of complex conjugates include:

1. $\mathfrak{Re}(\overline{z}) = \mathfrak{Re}(z)$
2. $\mathfrak{Im}(\overline{z}) = -\mathfrak{Im}(z)$
3. $z + \overline{z} = 2x = 2\mathfrak{Re}(z) \in \mathbb{R}$

4. $z - \overline{z} = 2iy = 2i\,\mathrm{Im}(z)$ is purely imaginary
5. $\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$
6. $\overline{z_1 z_2} = \overline{z_1}\ \overline{z_2}$. This can be seen either by noting that the conjugate operation simply changes every occurrence of i to $-i$ or since

$$(x_1 + iy_1)(x_2 + iy_2) = (x_1x_2 - y_1y_2) + i(x_1y_2 + y_1x_2), \tag{1.26}$$
$$(x_1 - iy_1)(x_2 - iy_2) = (x_1x_2 - y_1y_2) - i(x_1y_2 + y_1x_2), \tag{1.27}$$

which are conjugates. Geometrically, the complex conjugate $\overline{z}$ is a reflection of z where the real axis serves as the axis of reflection. Figure 1.9 illustrates this relationship.

1.3.1 Absolute Value of a Complex Number

The **absolute value (length/modulus)** of $z \in \mathbb{C}$ is $|z| = \sqrt{z\overline{z}}$, where

$$z\overline{z} = (x + iy)(x - iy) = x^2 + y^2 \in \mathbb{R}. \tag{1.28}$$

Notice that the term 'absolute value' is the same as defined for real numbers when $\mathrm{Im}(z) = 0$. In this case, $|z| = |x|$.
The absolute value of the product has the following nice property that matches the product result for real numbers:

$$|z_1 z_2| = |z_1|\,|z_2|. \tag{1.29}$$

This holds since

$$|z_1 z_2|^2 = z_1 z_2 \overline{z_1 z_2} = z_1 z_2 \overline{z_1}\ \overline{z_2} = z_1 \overline{z_1} z_2 \overline{z_2} = |z_1|^2 |z_2|^2. \tag{1.30}$$

1.3.2 Inverse and Division

If $z = x + iy$, its **inverse (reciprocal)** is

$$\frac{1}{z} = \frac{\overline{z}}{z\overline{z}} = \frac{\overline{z}}{|z|^2} = \frac{x - iy}{x^2 + y^2}. \tag{1.31}$$

This can be written $z^{-1} = |z|^{-2}\overline{z}$, using only the complex operators multiply and add, see (1.5) and (1.6), but also real division, which we already know. Complex division is now defined by $z_1/z_2 = z_1 z_2^{-1}$. In practice, we compute the division z_1/z_2 by expanding the fraction by the complex conjugate of the denominator. This ensures that the denominator's imaginary part is 0 (only the real part remains), and the overall fraction can be written as

$$\frac{z_1}{z_2} = \frac{z_1\overline{z_2}}{z_2\overline{z_2}} = \frac{z_1\overline{z_2}}{|z_2|^2} \tag{1.32}$$

Geometric Interpretation of Division

When we use the Euler representations of two complex numbers $z_1, z_2 \in \mathbb{C}$, we can write the division as

$$\frac{z_1}{z_2} = z_1 z_2^{-1} = r_1 \exp(i\phi_1)\Big(r_2 \exp(i\phi_2)\Big) = \frac{r_1}{r_2}\exp(i(\phi_1 - \phi_2)). \tag{1.33}$$

Geometrically, we divide r_1 by r_2 (equivalently: scale r_1 by $\frac{1}{r_2}$) and rotate z_1 by $-\phi_2$. This is not overly surprising since the division by z_2 does exactly the opposite of a multiplication by r_2. Therefore, looking again at Figure 1.8, if we take $z_1 z_2$ and divide by z_2, we obtain z_1.

Example: Complex Division

Bring the following fraction into the form $x + iy$:

$$z = x + iy = \frac{3+2i}{7-3i} \tag{1.34}$$

Solution:

$$\frac{3+2i}{7-3i} = \frac{(3+2i)(7+3i)}{(7-3i)(7+3i)} = \frac{15+23i}{49+9} = \frac{15}{58} + i\frac{23}{58} \tag{1.35}$$

Now, the fraction can be written as $z = x + iy$ with $x = \frac{15}{58}$ and $y = \frac{23}{58}$.

1.4 De Moivre's Theorem

De Moivre's theorem (or formula) is a central result because it connects complex numbers and trigonometry.

Theorem 1 (De Moivre's Theorem)
For any $n \in \mathbb{N}$

$$(\cos\phi + i\sin\phi)^n = \cos n\phi + i\sin n\phi \tag{1.36}$$

The proof is done by induction (which you will see in detail in the course *Reasoning about Programs*). A proof by induction allows you to prove that a property is true for all values of a natural number n. To construct an induction proof, you have to prove that the property, $P(n)$, is true for some base value (say, $n = 1$). A further proof is required to show that if it is true for the parameter $n = k$, then that implies it is also true for the parameter $n = k + 1$: that is $P(k) \Rightarrow P(k+1)$ for all $k \geq 1$. The two proofs combined allow us to build an arbitrary chain of implication up to some value $n = m$:

$$P(1) \text{ and } (P(1) \Rightarrow P(2) \Rightarrow \cdots \Rightarrow P(m-1) \Rightarrow P(m)) \models P(m)$$

Proof 1
We start the induction proof by checking whether de Moivre's theorem holds for $n = 1$*:*

$$(\cos\phi + i\sin\phi)^1 = \cos\phi + i\sin\phi \tag{1.37}$$

is trivially true, and we can now make the induction step: We assume that (1.36) *is true for* k *and show that it also holds for* $k+1$.
Assuming

$$(\cos\phi + i\sin\phi)^k = \cos k\phi + i\sin k\phi \tag{1.38}$$

we can write

$$\begin{aligned}(\cos\phi + i\sin\phi)^{k+1} &= (\cos\phi + i\sin\phi)(\cos\phi + i\sin\phi)^k \\ &= (\cos\phi + i\sin\phi)(\cos k\phi + i\sin k\phi) && \textit{using assumption } (1.38) \\ &= (\cos(k+1)\phi + i\sin(k+1)\phi) && \textit{using complex product } (1.23)\end{aligned}$$

which concludes the proof.

1.4.1 Integer Extension to De Moivre's Theorem

We can extend de Moivre to include negative numbers, $n \in \mathbb{Z}$

$$(\cos\phi + i\sin\phi)^n = \cos n\phi + i\sin n\phi$$

We have tackled the case for $n > 0$ already, $n = 0$ can be shown individually. So we take the case $n < 0$. We let $n = -m$ for $m > 0$.

$$\begin{aligned}(\cos\phi + i\sin\phi)^n &= \frac{1}{(\cos\phi + i\sin\phi)^m} \\ &= \frac{1}{\cos m\phi + i\sin m\phi} && \text{by de Moivre's theorem} \\ &= \frac{\cos m\phi - i\sin m\phi}{\cos^2 m\phi + \sin^2 m\phi} \\ &= \cos(-m\phi) + i\sin(-m\phi) && \text{Trig. identity: } \cos^2 m\phi + \sin^2 m\phi = 1 \\ &= \cos n\phi + i\sin n\phi\end{aligned}$$

1.4.2 Rational Extension to De Moivre's Theorem

Finally, for our purposes, we will show that if $n \in \mathbb{Q}$, one value of $(\cos\phi + i\sin\phi)^n$ is $\cos n\phi + i\sin n\phi$. Take $n = p/q$ for $p, q \in \mathbb{Z}$ and $q \neq 0$. We will use both de Moivre's theorems in the following:

$$\left(\cos\frac{p}{q}\phi + i\sin\frac{p}{q}\phi\right)^q = \cos p\phi + i\sin p\phi \tag{1.39}$$

$$= (\cos\phi + i\sin\phi)^p \tag{1.40}$$

Hence $\cos\frac{p}{q}\phi + i\sin\frac{p}{q}\phi$ is *one* of the qth roots of $(cos\phi + i\sin\phi)^p$.
The qth roots of $\cos\phi + i\sin\phi$ are easily obtained. We need to use the fact that (repeatedly) adding 2π to the argument of a complex number does not change the complex number.

$$(\cos\phi + i\sin\phi)^{\frac{1}{q}} = (\cos(\phi + 2n\pi) + i\sin(\phi + 2n\pi))^{\frac{1}{q}} \tag{1.41}$$

$$= \cos\frac{\phi + 2n\pi}{q} + i\sin\frac{\phi + 2n\pi}{q} \quad \text{for } 0 \le n < q \tag{1.42}$$

We will use this later to calculate roots of complex numbers.
Finally, the full set of values for $(\cos + i\sin\phi)^n$ for $n = p/q \in \mathbb{Q}$ is:

$$\cos\frac{p\phi + 2n\pi}{q} + i\sin\frac{p\phi + 2n\pi}{q} \quad \text{for } 0 \le n < q \tag{1.43}$$

Example: Multiplication using Complex Products

We require the result of:

$$(3 + 3i)(1 + i)^3$$

We could expand $(1+i)^3$ and multiply by $3+3i$ using real and imaginary components. Alternatively, we could tackle this in polar form $(\cos\phi + i\sin\phi)$ using the complex product of (1.23) and de Moivre's theorem.

$$(1 + i)^3 = [2^{1/2}(\cos\pi/4 + i\sin\pi/4)]^3$$
$$= 2^{3/2}(\cos 3\pi/4 + i\sin 3\pi/4)$$

by de Moivre's theorem. $3 + 3i = 18^{1/2}(\cos\pi/4 + i\sin\pi/4)$ and so the result is

$$18^{1/2}2^{3/2}(\cos\pi + i\sin\pi) = -12$$

Geometrically, we just observe that the Arg of the second number is 3 times that of $1+i$, i.e., $3\pi/4$ (or $3\cdot 45^\circ$ in degrees). The first number has the same Arg, so the Arg of the result is π.
Similarly, the absolute values (lengths) of the numbers multiplied are $\sqrt{18}$ and $\sqrt{2^3}$, so the product has absolute value 12. The result is therefore -12.

1.5 Triangle Inequality for Complex Numbers

The triangle inequality for complex numbers is as follows:

$$\forall z_1, z_2 \in \mathbb{C} : |z_1 + z_2| \le |z_1| + |z_2| \tag{1.44}$$

An alternative form, with $w_1 = z_1$ and $w_2 = z_1 + z_2$ is $|w_2| - |w_1| \le |w_2 - w_1|$ and, switching w_1, w_2, $|w_1| - |w_2| \le |w_2 - w_1|$. Thus, relabelling back to z_1, z_2:

$$\forall z_1, z_2 \in \mathbb{C} : \big||z_1| - |z_2|\big| \le |z_2 - z_1| \tag{1.45}$$

In the Argand diagram, this just says that "In the triangle with vertices at $0, z_1, z_2$, the length of side z_1z_2 is not less than the difference between the lengths of the other two sides".

Proof 2
Let $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$. Squaring the left-hand side of (1.45) yields

$$(x_1 + x_2)^2 + (y_1 + y_2)^2 = |z_1|^2 + |z_2|^2 + 2(x_1x_2 + y_1y_2), \tag{1.46}$$

and the square of the right-hand side is

$$|z_1|^2 + |z_2|^2 + 2|z_1||z_2| \tag{1.47}$$

It is required to prove $x_1x_2 + y_1y_2 \le |z_1||z_2|$. We continue by squaring this inequality

$$x_1x_2 + y_1y_2 \le |z_1||z_2| \tag{1.48}$$
$$\Leftrightarrow (x_1x_2 + y_1y_2)^2 \le |z_1|^2|z_2|^2 \tag{1.49}$$
$$\Leftrightarrow x_1^2x_2^2 + y_1^2y_2^2 + 2x_1x_2y_1y_2 \le x_1^2x_2^2 + y_1^2y_2^2 + x_1^2y_2^2 + y_1^2x_2^2 \tag{1.50}$$
$$\Leftrightarrow 0 \le (x_1y_2 - y_1x_2)^2, \tag{1.51}$$

which concludes the proof.

The geometrical argument via the Argand diagram is a good way to understand the triangle inequality.

1.6 Fundamental Theorem of Algebra

Theorem 2 (Fundamental Theorem of Algebra)
Any polynomial of degree n of the form

$$p(z) = \sum_{k=0}^{n} a_k z^k, \quad a_k \in \mathbb{C}, \quad a_n \neq 0 \tag{1.52}$$

possesses, counted with multiplicity, exactly n roots in $\mathbb{C}$.

A root z_* of $p(z)$ satisfies $p(z_*) = 0$. Bear in mind that complex roots include all real roots as the real numbers are a subset of the complex numbers. Also some of the roots might be coincident, e.g., for $z^2 = 0$. Finally, we also know that if ω is a root and $\omega \in \mathbb{C}\backslash\mathbb{R}$, then $\overline{\omega}$ is also a root. So all truly complex roots occur in complex conjugate pairs.

1.6.1 *n*th Roots of Unity

In the following, we consider the equation

$$z^n = 1, \quad n \in \mathbb{N}, \tag{1.53}$$

for which we want to determine the roots. The fundamental theorem of algebra tells us that there exist exactly n roots, one of which is $z = 1$.

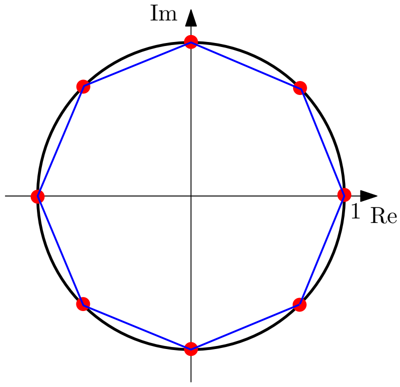


Figure 1.10: Then nth roots of $z^n = 1$ lie on the unit circle and form a regular polygon. Here, we show this for $n = 8$.

To find the other solutions, we write (1.53) in a slightly different form using the Euler representation:

$$z^n = 1 = e^{ik2\pi}, \quad \forall k \in \mathbb{Z}. \tag{1.54}$$

Then the solutions are $z = e^{i2k\pi/n}$ for $k = 0, 1, 2, \ldots, n-1$.[4]
Geometrically, all n roots lie on the unit circle, and they form a regular polygon with n corners where the roots are $360°/n$ apart, see an example in Figure 1.10. Therefore, if we know a single root and the total number of roots, we could even geometrically find all other roots.

Example: Cube Roots of Unity

The 3rd roots of 1 are $z = e^{2k\pi i/3}$ for $k = 0, 1, 2$, i.e., $1, e^{2\pi i/3}, e^{4\pi i/3}$. These are often referred to as ω_1 ω_1 and ω_3, and simplify to

$$\begin{aligned}
\omega_1 &= 1\\
\omega_2 &= \cos 2\pi/3 + i \sin 2\pi/3 = (-1 + i\sqrt{3})/2,\\
\omega_3 &= \cos 4\pi/3 + i \sin 4\pi/3 = (-1 - i\sqrt{3})/2.
\end{aligned}$$

Try cubing each solution directly to validate that they are indeed cubic roots.

1.6.2 Solution of $z^n = a + ib$

Finding the n roots of $z^n = a + ib$ is similar to the approach discussed above: Let $a + ib = re^{i\phi}$ in polar form. Then, for $k = 0, 1, \ldots, n-1$,

$$z^n = (a + ib)e^{2\pi ki} = re^{(\phi+2\pi k)i} \tag{1.55}$$

$$\Rightarrow z_k = r^{\frac{1}{n}} e^{\frac{(\phi+2\pi k)}{n} i}, \quad k = 0, \ldots, n-1. \tag{1.56}$$

[4] Note that the solutions repeat when $k = n, n+1, \ldots$

Example

Determine the cube roots of $1-i$.

1. The polar coordinates of $1-i$ are $r=\sqrt{2},\ \phi=\frac{7}{4}\pi$, and the corresponding Euler representation is

$$z=\sqrt{2}\exp(i\tfrac{7\pi}{4}). \tag{1.57}$$

2. Using (1.56), the cube roots of z are

$$z_1=2^{\frac{1}{6}}(\cos\tfrac{7\pi}{12}+i\sin\tfrac{7\pi}{12})=2^{\frac{1}{6}}\exp(i\tfrac{7\pi}{12}) \tag{1.58}$$
$$z_2=2^{\frac{1}{6}}(\cos\tfrac{15\pi}{12}+i\sin\tfrac{15\pi}{12})=2^{\frac{1}{6}}(\cos\tfrac{5\pi}{4}+i\sin\tfrac{5\pi}{4})=2^{\frac{1}{6}}\exp(i\tfrac{5\pi}{4}) \tag{1.59}$$
$$z_3=2^{\frac{1}{6}}(\cos\tfrac{23\pi}{12}+i\sin\tfrac{23\pi}{12})=2^{\frac{1}{6}}\exp(i\tfrac{23\pi}{12}). \tag{1.60}$$

1.7 Complex Sequences and Series*

A substantial part of the theory that we have developed for convergence of sequences and series of real numbers also applies to complex numbers. We will not reproduce all the results here, there is no need; we will highlight a couple of key concepts instead.

1.7.1 Limits of a Complex Sequence

For a sequence of complex numbers $z_1, z_2, z_3, \ldots$, we can define limits of convergence, $z_n \to l$ as $n\to\infty$ where $z_n, l\in\mathbb{C}$. This means that for all $\epsilon>0$ we can find a natural number N, such that

$$\forall n>N : |z_n-l|<\epsilon. \tag{1.61}$$

The only distinction here is the meaning of $|z_n-l|$, which refers to the complex absolute value and not the absolute real value.

Example of complex sequence convergence Prove that the complex sequence $z_n=\frac{1}{n+i}$ converges to 0 as $n\to\infty$. Straight to the limit inequality:

$$\left|\frac{1}{n+i}\right|<\epsilon \tag{1.62}$$
$$\Leftrightarrow \frac{|n-i|}{n^2+1}<\epsilon \tag{1.63}$$
$$\Leftrightarrow \frac{\sqrt{(n-i)(n+i)}}{n^2+1}<\epsilon \tag{1.64}$$
$$\Leftrightarrow \frac{1}{\sqrt{n^2+1}}<\epsilon \tag{1.65}$$

$$\Rightarrow n > \sqrt{\frac{1}{\epsilon^2} - 1} \qquad \text{for } \epsilon \leq 1 \tag{1.66}$$

Thus, we can set

$$N(\epsilon) = \begin{cases} \left\lceil \sqrt{\frac{1}{\epsilon^2} - 1} \right\rceil & \epsilon \leq 1 \\ 1 & \text{otherwise} \end{cases} \tag{1.67}$$

We have to be a tiny bit careful as $N(\epsilon)$ needs to be defined *for all* $\epsilon > 0$ and the penultimate line of the limit inequality is true for all $n > 0$ if $\epsilon > 1$. In essence this was no different in structure from the normal sequence convergence proof. The only difference was how we treated the absolute value.

Absolute Convergence

Similarly, a complex series $\sum_{n=1}^{\infty} z_n$ is absolutely convergent if $\sum_{n=1}^{\infty} |z_n|$ converges. Again the $|z_n|$ refers to the complex absolute value.

Complex Ratio Test

A complex series $\sum_{n=1}^{\infty} z_n$ converges if

$$\lim_{n\to\infty} \left| \frac{z_{n+1}}{z_n} \right| < 1 \tag{1.68}$$

and diverges if

$$\lim_{n\to\infty} \left| \frac{z_{n+1}}{z_n} \right| > 1 \, . \tag{1.69}$$

Example of Complex Series Convergence

Let us take a general variant of the geometric series:

$$S = \sum_{n=1}^{\infty} az^{n-1} \tag{1.70}$$

We can prove that this will converge for some values of $z \in \mathbb{C}$ in the same way we could for the real-valued series. Applying the complex ratio test, we get $\lim_{n\to\infty} |\frac{az^n}{az^{n-1}}| = |z|$. We apply the standard condition and get that $|z| < 1$ for this series to converge. The radius of convergence is still 1 (and is an actual radius of a circle in the complex plane). What is different here is that now any z-point taken from within the circle centred on the origin with radius 1 will make the series converge, not just on the real interval $(-1, 1)$.
For your information, the limit of this series is $\frac{a}{1-z}$, which you can show using Maclaurin as usual, discussed below.

1.8 Complex Power Series

We can expand functions as power series in a complex variable, usually z, in the same way as we could with real-valued functions. The same expansions hold in $\mathbb{C}$ because the functions below (at any rate) are differentiable in the complex domain. Therefore, Maclaurin's series applies and yields

$$\exp(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!} = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \dots \tag{1.71}$$

$$\sin(z) = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n+1}}{(2n+1)!} = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \dots \tag{1.72}$$

$$\cos(z) = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n)!} = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \dots \tag{1.73}$$

1.8.1 A Generalized Euler Formula

A more general form of Euler's formula (1.12) is

$$\forall z \in \mathbb{C}, n \in \mathbb{Z} : z = re^{i(\phi + 2n\pi)} \tag{1.74}$$

since $e^{i2n\pi} = \cos 2n\pi + i \sin 2n\pi = 1$. This is the same general form we used in the rational extension to De Moivre's theorem to access the many roots of a complex number.
In terms of the Argand diagram, the points $e^{i(\phi+2n\pi)}$ for $i \geq 1$ lie on top of each other, each corresponding to one more revolution (through 2π).
The complex conjugate of $e^{i\phi}$ is $e^{-i\phi} = \cos\phi - i \sin\phi$. This allows us to get useful expressions for $\sin\phi$ and $\cos\phi$:

$$\cos\phi = (e^{i\phi} + e^{-i\phi})/2 \tag{1.75}$$

$$\sin\phi = (e^{i\phi} - e^{-i\phi})/2i. \tag{1.76}$$

We will be able to use these relationships to create trigonometric identities.

1.9 Applications of Complex Numbers*

1.9.1 Trigonometric Multiple Angle Formulae

How can we calculate $\cos n\phi$ in terms of $\cos\phi$ and $\sin\phi$? We can use de Moivre's theorem to expand $e^{in\phi}$ and equate real and imaginary parts: e.g., for $n = 5$, by the Binomial theorem,

$$\begin{aligned}(\cos\phi + i\sin\phi)^5 = {} & \cos^5\phi + i5\cos^4\phi\sin\phi - 10\cos^3\phi\sin^2\phi \\ & - i10\cos^2\phi\sin^3\phi + 5\cos\phi\sin^4\phi + i\sin^5\phi\end{aligned} \tag{1.77}$$

Comparing real and imaginary parts now gives

$$\cos 5\phi = \cos^5\phi - 10\cos^3\phi\sin^2\phi + 5\cos\phi\sin^4\phi \tag{1.78}$$

and

$$\sin 5\phi = 5\cos^4\phi\sin\phi - 10\cos^2\phi\sin^3\phi + \sin^5\phi \tag{1.79}$$

Trigonometric Power Formulae

We can also calculate $\cos^n\phi$ in terms of $\cos m\phi$ and $\sin m\phi$ for $m \in \mathbb{N}$: Let $z = e^{i\phi}$ so that $z + z^{-1} = z + \bar{z} = 2\cos\phi$. Similarly, $z^m + z^{-m} = 2\cos m\phi$ by de Moivre's theorem. Hence by the Binomial theorem, e.g., for $n = 5$,

$$(z + z^{-1})^5 = (z^5 + z^{-5}) + 5(z^3 + z^{-3}) + 10(z + z^{-1}) \tag{1.80}$$

$$2^5\cos^5\phi = 2(\cos 5\phi + 5\cos 3\phi + 10\cos\phi) \tag{1.81}$$

Similarly, $z - z^{-1} = 2i\sin\phi$ gives $\sin^n\phi$
When n is even, we get an extra term in the binomial expansion, which is *constant*. For example, for $n = 6$, we obtain

$$(z + z^{-1})^6 = (z^6 + z^{-6}) + 6(z^4 + z^{-4}) + 15(z^2 + z^{-2}) + 20 \tag{1.82}$$

$$2^6\cos^6\phi = 2(\cos 6\phi + 6\cos 4\phi + 15\cos 2\phi + 10) \tag{1.83}$$

and, therefore,

$$\cos^6\phi = \frac{1}{32}(\cos 6\phi + 6\cos 4\phi + 15\cos 2\phi + 10). \tag{1.84}$$

1.9.2 Summation of Series

Some series with sines and cosines can be summed similarly, e.g.,

$$C = \sum_{k=0}^{n} a^k \cos k\phi \tag{1.85}$$

Let $S = \sum\limits_{k=1}^{n} a^k \sin k\phi$. Then,

$$C + iS = \sum_{k=0}^{n} a^k e^{ik\phi} = \frac{1 - (ae^{i\phi})^{n+1}}{1 - ae^{i\phi}}. \tag{1.86}$$

Hence,

$$C + iS = \frac{(1 - (ae^{i\phi})^{n+1})(1 - ae^{-i\phi})}{(1 - ae^{i\phi})(1 - ae^{-i\phi})} \tag{1.87}$$

$$= \frac{1 - ae^{-i\phi} - a^{n+1}e^{i(n+1)\phi} + a^{n+2}e^{in\phi}}{1 - 2a\cos\phi + a^2}. \tag{1.88}$$

Equating real and imaginary parts, the cosine series is

$$C = \frac{1 - a\cos\phi - a^{n+1}\cos(n+1)\phi + a^{n+2}\cos n\phi}{1 - 2a\cos\phi + a^2}, \tag{1.89}$$

and the sine series is

$$S = \frac{a\sin\phi - a^{n+1}\sin(n+1)\phi + a^{n+2}\sin n\phi}{1 - 2a\cos\phi + a^2} \tag{1.90}$$

1.9.3 Integrals

We can determine integrals

$$C = \int_0^x e^{a\phi}\cos b\phi d\phi, \tag{1.91}$$

$$S = \int_0^x e^{a\phi}\sin b\phi d\phi \tag{1.92}$$

by looking at the sum[5]

$$C + iS = \int_0^x e^{(a+ib)\phi} d\phi \tag{1.93}$$

$$= \frac{e^{(a+ib)x} - 1}{a + ib} = \frac{(e^{ax}e^{ibx} - 1)(a - ib)}{a^2 + b^2} \tag{1.94}$$

$$= \frac{(e^{ax}\cos bx - 1 + ie^{ax}\sin bx)(a - ib)}{a^2 + b^2} \tag{1.95}$$

The result is therefore

$$C + iS = \frac{e^{ax}(a\cos bx + b\sin bx) - a + i(e^{ax}(a\sin bx - b\cos bx) + b)}{a^2 + b^2} \tag{1.96}$$

and so we get

$$C = \frac{e^{ax}(a\cos bx + b\sin bx - a)}{a^2 + b^2}, \tag{1.97}$$

$$S = \frac{e^{ax}(a\sin bx - b\cos bx) + b}{a^2 + b^2} \tag{1.98}$$

as the solutions to the integrals we were seeking.

[5] The reduction formula would require a and b to be integers.

Chapter 2

Linear Algebra

This chapter is largely based on the lecture notes and books by Drumm and Weil (2001); Strang (2003); Hogben (2013) as well as Pavel Grinfeld's Linear Algebra series[1]. Another excellent source is Gilbert Strang's Linear Algebra lecture at MIT[2].
Linear algebra is the study of vectors. Generally, vectors are special objects that can be added together and multiplied by scalars to produce another object of the same kind. Any object that satisfies these two properties can be considered a vector. Here are three examples of such vectors:

1. Geometric vectors. This example of a vector may be familiar from High School. Geometric vectors are directed segments, which can be drawn, see Fig. 2.1. Two vectors $\vec{x}, \vec{y}$ can be added, such that $\vec{x} + \vec{y} = \vec{z}$ is another geometric vector. Furthermore, $\lambda\vec{x}, \lambda \in \mathbb{R}$ is also a geometric vector. In fact, it is the original vector scaled by λ. Therefore, geometric vectors are instances of the vector concepts introduced above.

2. Polynomials are also vectors: Two polynomials can be added together, which results in another polynomial; and they can be multiplied by a scalar $\lambda \in \mathbb{R}$, and the result is a polynomial as well. Therefore, polynomial are (rather unusual) instances of vectors. Note that polynomials are very different from geometric vectors. While geometric vectors are concrete "drawings", polynomials are abstract concepts. However, they are both vectors.

3. $\mathbb{R}^n$ is a set of numbers, and its elements are n-tuples. $\mathbb{R}^n$ is even more abstract than polynomials, and the most general concept we consider in this course. For example,

$$\boldsymbol{a} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \in \mathbb{R}^3 \tag{2.1}$$

   is an example of a triplet of numbers. Adding two vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$ componentwise results in another vector: $\boldsymbol{a} + \boldsymbol{b} = \boldsymbol{c} \in \mathbb{R}^n$. Moreover, multiplying $\boldsymbol{a} \in \mathbb{R}^n$ by $\lambda \in \mathbb{R}$ results in a scaled vector $\lambda\boldsymbol{a} \in \mathbb{R}^n$.

[1]http://tinyurl.com/nahclwm
[2]http://tinyurl.com/29p5q8j

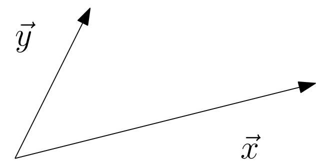


Figure 2.1: Example of two geometric vectors in two dimensions.

Linear algebra focuses on the similarities between these vector concepts: We can add them together and multiply them by scalars. We will largely focus on the third kind of vectors since most algorithms in linear algebra are formulated in $\mathbb{R}^n$. There is a 1:1 correspondence between any kind of vector and $\mathbb{R}^n$. By studying $\mathbb{R}^n$, we implicitly study all other vectors. Although $\mathbb{R}^n$ is rather abstract, it is most useful.

Practical Applications of Linear Algebra

Linear algebra centers around solving linear equation systems and is at the core of many computer science applications. Here is a selection:[3]

- Ranking of web pages (web search)
- Linear programming (optimziation)
- Error correcting codes (e.g., in DVDs)
- Decomposition of sounds into different sources
- Projections, rotations, scaling (computer graphics)
- Data visualization
- En/Decryption algorithms (cryptography)
- State estimation and optimal control (e.g., in robotics and dynamical systems)

2.1 Linear Equation Systems

2.1.1 Example

A company produces products $N_1, \dots, N_n$ for which resources $R_1, \dots, R_m$ are required. To produce a unit of product N_j, a_{ij} units of resource R_i are needed, where $i = 1, \dots, m$ and $j = 1, \dots, n$.
The objective is to find an optimal production plan, i.e., a plan how many units x_j of product N_j should be produced if a total of b_i units of resource R_i are available and (ideally) no resources are left over.
If we produce $x_1, \dots, x_n$ units of the corresponding products, we need a total of

$$a_{i1}x_1 + \cdots + a_{in}x_n \tag{2.2}$$

[3] More details can be found on Jeremy Kun's blog: `http://tinyurl.com/olkbkct`

many units of resource R_i. The desired optimal production plan $(x_1,\ldots,x_n) \in \mathbb{R}^n$, therefore, has to satisfy the following system of equations:

$$\begin{matrix} a_{11}x_1 + \cdots + a_{1n}x_n & & b_1 \\ \vdots & = & \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n & & b_m \end{matrix}, \tag{2.3}$$

where $a_{ij} \in \mathbb{R}$ and $b_i \in \mathbb{R}$. Equation (2.3) is the general form of a **linear equation system**, and $x_1,\ldots,x_n$ are the **unknowns** of this linear equation system. Every n-tuple $(x_1,\ldots,x_n) \in \mathbb{R}^n$ that satisfies (2.3) is a **solution** of the linear equation system.
The linear equation system

$$\begin{array}{rcrcrcrl} x_1 & + & x_2 & + & x_3 & = & 3 & (1) \\ x_1 & - & x_2 & + & 2x_3 & = & 2 & (2) \\ 2x_1 & & & + & 3x_3 & = & 1 & (3) \end{array} \tag{2.4}$$

has **no solution:** Adding the first two equations yields $(1)+(2) = 2x_1 + 3x_3 = 5$, which contradicts the third equation (3).
Let us have a look at the linear equation system

$$\begin{array}{rcrcrcrl} x_1 & + & x_2 & + & x_3 & = & 3 & (1) \\ x_1 & - & x_2 & + & 2x_3 & = & 2 & (2) \\ & & x_2 & + & x_3 & = & 2 & (3) \end{array}. \tag{2.5}$$

From the first and third equation it follows that $x_1 = 1$. From $(1)+(2)$ we get $2+3x_3 = 5$, i.e., $x_3 = 1$. From (3), we then get that $x_2 = 1$. Therefore, $(1,1,1)$ is the only possible and **unique solution** (verify by plugging in).
As a third example, we consider

$$\begin{array}{rcrcrcrl} x_1 & + & x_2 & + & x_3 & = & 3 & (1) \\ x_1 & - & x_2 & + & 2x_3 & = & 2 & (2) \\ 2x_1 & & & + & 3x_3 & = & 5 & (3) \end{array}. \tag{2.6}$$

Since (1)+(2)=(3), we can omit the third equation (redundancy). From (1) and (2), we get $2x_1 = 5 - 3x_3$ and $2x_2 = 1 + x_3$. We define $x_3 = a \in \mathbb{R}$ as a free variable, such that any triplet

$$\left(\frac{5}{2} - \frac{3}{2}a, \frac{1}{2} + \frac{1}{2}a, a\right), \quad a \in \mathbb{R} \tag{2.7}$$

is a solution to the linear equation system, i.e., we obtain a solution set that contains **infinitely many** solutions.
In general, for a real-valued linear equation system we obtain either no, exactly one or infinitely many solutions.
For a systematic approach to solving linear equation systems, we will introduce a useful compact notation. We will write the linear equation system from (2.3) in the following form:

$$x_1 \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \Leftrightarrow \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}. \tag{2.8}$$

In order to work with these **matrices**, we need to have a close look at the underlying algebraic structures and define computation rules.

2.2 Groups

Groups play an important role in computer science. Besides providing a fundamental framework for operations on sets, they are heavily used in cryptography, coding theory and graphics.

2.2.1 Definitions

Consider a set G and an operation $\otimes : G \to G$ defined on G. For example, $\otimes$ could be $+, \cdot$ defined on $\mathbb{R}, \mathbb{N}, \mathbb{Z}$ or $\cup, \cap, \setminus$ defined on $\mathcal{P}(B)$, the power set of B.
Then $(G, \otimes)$ is called a **group** if

- **Closure** of G under $\otimes$: $\forall x, y \in G : x \otimes y \in \mathcal{G}$
- **Associativity:** $\forall x, y, z \in G : (x \otimes y) \otimes z = x \otimes (y \otimes z)$
- **Neutral element:** $\exists e \in G \forall x \in G : x \otimes e = x$ and $e \otimes x = x$
- **Inverse element:** $\forall x \in G \exists y \in G : x \otimes y = e$ and $y \otimes x = e$. We often write x^{-1} to denote the inverse element of x.[4]

If additionally $\forall x, y \in G : x \otimes y = y \otimes x$ then $(G, \otimes)$ is **Abelian group** (commutative).

2.2.2 Examples

$(\mathbb{Z}, +)$ is a group, whereas $(\mathbb{N}_0, +)$[5] is not: Although $(\mathbb{N}_0, +)$ possesses a neutral element (0), the inverse elements are missing.
$(\mathbb{Z}, \cdot)$ is not a group: Although $(\mathbb{Z}, \cdot)$ contains a neutral element (1), the inverse elements for any $z \in \mathbb{Z}, z \neq \pm 1$, are missing.
$(\mathbb{R}, \cdot)$ is not a group since 0 does not possess an inverse element. However, $(\mathbb{R} \setminus \{0\})$ is Abelian.
$(\mathbb{R}^n, +), (\mathbb{Z}^n, +), n \in \mathbb{N}$ are Abelian if $+$ is defined componentwise, i.e.,

$$(x_1, \cdots, x_n) + (y_1, \cdots, y_n) = (x_1 + y_1, \cdots, x_n + y_n). \tag{2.9}$$

Then, $e = (0, \cdots, 0)$ is the neutral element and $(x_1, \cdots, x_n)^{-1} := (-x_1, \cdots, -x_n)$ is the inverse element.

2.3 Matrices

Definition 1 (Matrix)
With $m, n \in \mathbb{N}$ a real-valued (m, n) ***matrix*** *A is an $m \cdot n$-tuple of elements a_{ij}, $i = 1, \ldots, m$, $j = 1, \ldots, n$, which is ordered according to a rectangular scheme consisting of m rows*

[4] The inverse element is defined with respect to the operation $\otimes$ and does not necessarily mean $\frac{1}{x}$.
[5] $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$

and n columns:

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad a_{ij} \in \mathbb{R}. \tag{2.10}$$

$(1,n)$-matrices are called **rows***, $(m,1)$-matrices are called* **columns***. These special matrices are also called* **row/column vectors***.*

$\mathbb{R}^{m\times n}$ is the set of all real-valued (m,n)-matrices. $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ can be equivalently represented as $\boldsymbol{A} \in \mathbb{R}^{mn}$. Therefore, $(\mathbb{R}^{m\times n},+)$ is Abelian group (with componentwise addition as defined in (2.9)).

2.3.1 Matrix Multiplication

For $\boldsymbol{A} \in \mathbb{R}^{m\times n}, \boldsymbol{B} \in \mathbb{R}^{n\times k}$ (note the size of the matrices!) the elements c_{ij} of the product $\boldsymbol{C} = \boldsymbol{AB} \in \mathbb{R}^{m\times k}$ are defined as

$$c_{ij} = \sum_{l=1}^{n} a_{il} b_{lj}, \qquad i = 1,\dots,m, \quad j = 1,\dots,k. \tag{2.11}$$

This means, to compute element c_{ij} we multiply the elements of the ith row of $\boldsymbol{A}$ with the jth column of $\boldsymbol{B}$[6] and sum them up.[7]

Remark 1

Matrices can only be multiplied if their "neighboring" dimensions match. For instance, an $n\times k$-matrix $\boldsymbol{A}$ can be multiplied with a $k\times m$-matrix $\boldsymbol{B}$, but only from the left side:

$$\underbrace{\boldsymbol{A}}_{n\times k}\underbrace{\boldsymbol{B}}_{k\times m} = \underbrace{\boldsymbol{C}}_{n\times m} \tag{2.12}$$

The product $\boldsymbol{BA}$ is not defined if $m \neq n$ since the neighboring dimensions do not match.

Remark 2

Note that matrix multiplication is **not** *defined as an element-wise operation on matrix elements, i.e., $c_{ij} \neq a_{ij} b_{ij}$ (even if the size of $\boldsymbol{A},\boldsymbol{B}$ was chosen appropriately).*[8]

Example

For $\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \in \mathbb{R}^{2\times 3}$, $\boldsymbol{B} = \begin{bmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 2}$, we obtain

$$\boldsymbol{AB} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 2 & 5 \end{bmatrix} \in \mathbb{R}^{2\times 2}, \tag{2.13}$$

[6] They are both of length k, such that we can compute $a_{il} b_{lj}$ for $l = 1,\dots,n$.

[7] Later, we will call this the **scalar product** or **dot product** of the corresponding row and column.

[8] This kind of element-wise multiplication appears often in computer science where we multiply (multi-dimensional) arrays with each other.

$$BA = \begin{bmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 4 & 2 \\ -2 & 0 & 2 \\ 3 & 2 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 3}. \tag{2.14}$$

From this example, we can already see that matrix multiplication is not commutative, i.e., $AB \neq BA$.

Definition 2 (Identity Matrix)
In $\mathbb{R}^{n\times n}$, we define the **identity matrix** *as*

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix} \in \mathbb{R}^{n\times n}. \tag{2.15}$$

With this, $A \cdot I_n = A = I_n A$ for all $A \in \mathbb{R}^{n\times n}$. Therefore, the identity matrix is the neutral element with respect to matrix multiplication "$\cdot$" in $(\mathbb{R}^{n\times n}, \cdot)$.[9]

Properties

- Associativity: $\forall A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times p}, C \in \mathbb{R}^{p\times q} : (AB)C = A(BC)$
- Distributivity: $\forall A_1, A_2 \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times p} : (A_1 + A_2)B = A_1B + A_2B$
  $A(B + C) = AB + AC$
- $\forall A \in \mathbb{R}^{m\times n} : I_m A = A I_n = A$. Note that $I_m \neq I_n$ for $m \neq n$.

2.3.2 Inverse and Transpose

Definition 3 (Inverse)
For a square matrix[10] *$A \in \mathbb{R}^{n\times n}$ a matrix $B \in \mathbb{R}^{n\times n}$ with $AB = I_n = BA$ is called* **inverse** *and denoted by A^{-1}.*

Not every matrix A possesses an inverse A^{-1}. If this inverse does exist, A is called **regular/invertible**, otherwise **singular**. We will discuss these properties much more later on in the course.

Remark 3
The set of regular (invertible) matrices $A \in \mathbb{R}^{n\times n}$ is a group with respect to matrix multiplication as defined in (2.11) *and is called* **general linear group** *$GL(n, \mathbb{R})$.*

Definition 4 (Transpose)
For $A \in \mathbb{R}^{m\times n}$ the matrix $B \in \mathbb{R}^{n\times m}$ with $b_{ij} = a_{ji}$ is called the **transpose** *of A. We write $B = A^\top$.*

[9] If $A \in \mathbb{R}^{m\times n}$ then I_n is only a right neutral element, succh that $AI_n = A$. The corresponding left-neutral element would be I_m since $I_m A = A$.

[10] The number columns equals the number of rows.

For a square matrix $\boldsymbol{A}^\top$ is the matrix we obtain when we "mirror" $\boldsymbol{A}$ on its main diagonal.[11] In general, $\boldsymbol{A}^\top$ can be obtained by writing the columns of $\boldsymbol{A}$ as the rows of $\boldsymbol{A}^\top$.

Remark 4

- $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I} = \boldsymbol{A}^{-1}\boldsymbol{A}$
- $(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$
- $(\boldsymbol{A}^\top)^\top = \boldsymbol{A}$
- $(\boldsymbol{A}+\boldsymbol{B})^\top = \boldsymbol{A}^\top + \boldsymbol{B}^\top$
- $(\boldsymbol{A}\boldsymbol{B})^\top = \boldsymbol{B}^\top\boldsymbol{A}^\top$
- *If* $\boldsymbol{A}$ *is invertible,* $(\boldsymbol{A}^{-1})^\top = (\boldsymbol{A}^\top)^{-1}$
- *Note:* $(\boldsymbol{A}+\boldsymbol{B})^{-1} \neq \boldsymbol{A}^{-1} + \boldsymbol{B}^{-1}$. *Example: in the scalar case* $\frac{1}{2+4} = \frac{1}{6} \neq \frac{1}{2} + \frac{1}{4}$.

$\boldsymbol{A}$ is **symmetric** if $\boldsymbol{A} = \boldsymbol{A}^\top$. Note that this can only hold for (n,n)-matrices (**quadratic matrices**). The sum of symmetric matrices is symmetric, but this does not hold for the product in general (although it is always defined). A counterexample is

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}. \tag{2.16}$$

2.3.3 Multiplication by a Scalar

Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\lambda \in \mathbb{R}$. Then $\lambda\boldsymbol{A} = \boldsymbol{K}$, $K_{ij} = \lambda\, a_{ij}$. Practically, λ scales each element of $\boldsymbol{A}$. For $\lambda, \psi \in \mathbb{R}$ it holds:

- Distributivity:
  $(\lambda+\psi)\boldsymbol{C} = \lambda\boldsymbol{C} + \psi\boldsymbol{C}, \quad \boldsymbol{C} \in \mathbb{R}^{m\times n}$
  $\lambda(\boldsymbol{B}+\boldsymbol{C}) = \lambda\boldsymbol{B} + \lambda\boldsymbol{C}, \quad \boldsymbol{B},\boldsymbol{C} \in \mathbb{R}^{m\times n}$
- Associativity:
  $(\lambda\psi)\boldsymbol{C} = \lambda(\psi\boldsymbol{C}), \quad \boldsymbol{C} \in \mathbb{R}^{m\times n}$
  $\lambda(\boldsymbol{B}\boldsymbol{C}) = (\lambda\boldsymbol{B})\boldsymbol{C} = \boldsymbol{B}(\lambda\boldsymbol{C}), \quad \boldsymbol{B} \in \mathbb{R}^{m\times n}, \boldsymbol{C} \in \mathbb{R}^{n\times k}$.
  Note that this allows us to move scalar values around.
- $(\lambda\boldsymbol{C})^\top = \boldsymbol{C}^\top\lambda^\top = \boldsymbol{C}^\top\lambda = \lambda\boldsymbol{C}^\top$ since $\lambda = \lambda^\top$ for all $\lambda \in \mathbb{R}$.

[11] The main diagonal (sometimes principal diagonal, primary diagonal, leading diagonal, or major diagonal) of a matrix $\boldsymbol{A}$ is the collection of entries A_{ij} where $i = j$.

2.3.4 Compact Representations of Linear Equation Systems

If we consider a linear equation system

$$\begin{aligned} 2x_1 + 3x_2 + 5x_3 &= 1 \\ 4x_1 - 2x_2 - 7x_3 &= 8 \\ 9x_1 + 5x_2 - 3x_3 &= 2 \end{aligned}$$

and use the rules for matrix multiplication, we can write this equation system in a more compact form as

$$\begin{bmatrix} 2 & 3 & 5 \\ 4 & -2 & -7 \\ 9 & 5 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 8 \\ 2 \end{bmatrix}. \tag{2.17}$$

Note that x_1 scales the first column, x_2 the second one, and x_3 the third one. Generally, linear equation systems can be compactly represented in their matrix form as $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, see (2.3), and the product $\boldsymbol{A}\boldsymbol{x}$ is a (linear) combination of the columns of $\boldsymbol{A}$.[12]

2.4 Solving Linear Equation Systems via Gaussian Elimination

In (2.3), we have introduced the general form of an equation system, i.e.,

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ \vdots \qquad\qquad & \ \ \vdots \ , \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned} \tag{2.18}$$

where $a_{ij} \in \mathbb{R}$ and $b_i \in \mathbb{R}$ are known constants and x_j are unknowns, $i = 1, \ldots, m$, $j = 1, \ldots, n$. Thus far, we have introduced matrices as a compact way of formulating linear equation systems, i.e., such that we can write $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, see (2.8). Moreover, we defined basic matrix operations, such as addition and multiplication of matrices. In the following, we will introduce a constructive and systematic way of solving linear equation systems. Before doing this, we introduce the **augmented matrix** $\left[\boldsymbol{A} \,|\, \boldsymbol{b}\right]$ of the linear equation system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$. This augmented matrix will turn out to be useful when solving linear equation systems.

2.4.1 Example: Solving a Simple Linear Equation System

Now we are turning towards solving linear equation systems. Before doing this in a systematic way using Gaussian elimination, let us have a look at an example.

[12]We will discuss linear combinations in Section 2.5.

Consider the following linear equation system:

$$\begin{bmatrix} 1 & 0 & 8 & -4 \\ 0 & 1 & 2 & 12 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 42 \\ 8 \end{bmatrix}. \tag{2.19}$$

This equation system is in a particularly easy form, where the first two columns consist of a 1 and a 0.[13] Remember that we want to find scalars $x_1, \dots, x_4$, such that $\sum_{i=1}^{4} x_i \boldsymbol{c}_i = \boldsymbol{b}$, where we define $\boldsymbol{c}_i$ to be the ith column of the matrix and $\boldsymbol{b}$ the right-hand-side of (2.19). A solution to the problem in (2.19) can be found immediately by taking 42 times the first column and 8 times the second column, i.e.,

$$\boldsymbol{b} = \begin{bmatrix} 42 \\ 8 \end{bmatrix} = 42 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 8 \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{2.20}$$

Therefore, one solution vector is $[42, 8, 0, 0]^\top$. This solution is called a **particular solution** or **special solution**. However, this is not the only solution of this linear equation system. To capture all the other solutions, we need to be creative of generating $\boldsymbol{0}$ in a non-trivial way using the columns of the matrix: Adding a couple of $\boldsymbol{0}$s to our special solution does not change the special solution. To do so, we express the third column using the first two columns (which are of this very simple form):

$$\begin{bmatrix} 8 \\ 2 \end{bmatrix} = 8 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \tag{2.21}$$

such that $\boldsymbol{0} = 8\boldsymbol{c}_1 + 2\boldsymbol{c}_2 - 1\boldsymbol{c}_3 + 0\boldsymbol{c}_4$. In fact, any scaling of this solution produces the $\boldsymbol{0}$ vector:

$$\lambda_1 \begin{bmatrix} 8 \\ 2 \\ -1 \\ 0 \end{bmatrix} = \boldsymbol{0}, \qquad \lambda_1 \in \mathbb{R}. \tag{2.22}$$

Following the same line of reasoning, we express the fourth column of the matrix in (2.19) using the first two columns and generate another set of non-trivial versions of $\boldsymbol{0}$ as

$$\lambda_2 \begin{bmatrix} -4 \\ 12 \\ 0 \\ -1 \end{bmatrix} = \boldsymbol{0} \qquad \lambda_2 \in \mathbb{R}. \tag{2.23}$$

Putting everything together, we obtain all solutions of the linear equation system in (2.19), which is called the **general solution**, as

$$\begin{bmatrix} 42 \\ 8 \\ 0 \\ 0 \end{bmatrix} + \lambda_1 \begin{bmatrix} 8 \\ 2 \\ -1 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} -4 \\ 12 \\ 0 \\ -1 \end{bmatrix}. \tag{2.24}$$

[13]Later, we will say that this matrix is in reduced row echelon form.

Remark 5

- *The general approach we followed consisted of the following three steps:*
  1. *Find a particular solution to* $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$
  2. *Find all solutions to* $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$
  3. *Combine the solutions from 1. and 2. to the general solution.*
- *Neither the general nor the particular solution is unique.*

The linear equation system in the example above was easy to solve because the matrix in (2.19) has this particularly convenient form, which allowed us to find the particular and the general solution by inspection. However, general equation systems are not of this simple form. Fortunately, there exists a constructive way of transforming any linear equation system into this particularly simple form: **Gaussian elimination**.
The rest of this section will introduce Gaussian elimination, which will allow us to solve all kinds of linear equation systems by first bringing them into a simple form and then applying the three steps to the simple form that we just discussed in the context of the example in (2.19), see Remark 5.

2.4.2 Elementary Transformations

Key to solving linear equation systems are **elementary transformations** that keep the solution set the same[14], but that transform the equation system into a simpler form:

- Exchange of two equations
- Multiplication of an equation with a constant $\lambda \in \mathbb{R}\backslash\{0\}$
- Addition of an equation to another equation

Example

$$\begin{array}{rcrcrcrcrcr} -2x_1 & + & 4x_2 & - & 2x_3 & - & x_4 & + & 4x_5 & = & -3 \\ 4x_1 & - & 8x_2 & + & 3x_3 & - & 3x_4 & + & x_5 & = & 2 \\ x_1 & - & 2x_2 & + & x_3 & - & x_4 & + & x_5 & = & 0 \\ x_1 & - & 2x_2 & & & - & 3x_4 & + & 4x_5 & = & a \end{array}, \quad a \in \mathbb{R} \tag{2.25}$$

Swapping rows 1 and 3 leads to

$$\begin{array}{rcrcrcrcrcrl} x_1 & - & 2x_2 & + & x_3 & - & x_4 & + & x_5 & = & 0 & \\ 4x_1 & - & 8x_2 & + & 3x_3 & - & 3x_4 & + & x_5 & = & 2 & |-4R_1 \\ -2x_1 & + & 4x_2 & - & 2x_3 & - & x_4 & + & 4x_5 & = & -3 & |+2R_1 \\ x_1 & - & 2x_2 & & & - & 3x_4 & + & 4x_5 & = & a & |-R_1 \end{array} \tag{2.26}$$

[14] Therefore, the original and the modified equation system are **equivalent**.

When we now apply the indicated transformations (e.g., subtract Row 1 4 times from Row 2), we obtain

$$\begin{array}{rcrcrcrcrcrl} x_1 & - & 2x_2 & + & x_3 & - & x_4 & + & x_5 & = & 0 & \\ & & & - & x_3 & + & x_4 & - & 3x_5 & = & 2 & \\ & & & & & - & 3x_4 & + & 6x_5 & = & -3 & \\ & & & - & x_3 & - & 2x_4 & + & 3x_5 & = & a & \quad |-R_2 \end{array} \tag{2.27}$$

then

$$\begin{array}{rcrcrcrcrcrl} x_1 & - & 2x_2 & + & x_3 & - & x_4 & + & x_5 & = & 0 & \\ & & & - & x_3 & + & x_4 & - & 3x_5 & = & 2 & \\ & & & & & - & 3x_4 & + & 6x_5 & = & -3 & \\ & & & & & - & 3x_4 & + & 6x_5 & = & a-2 & \quad |-R_3 \end{array} \tag{2.28}$$

and finally

$$\begin{array}{rcrcrcrcrcrl} x_1 & - & 2x_2 & + & x_3 & - & x_4 & + & x_5 & = & 0 & \\ & & & - & x_3 & + & x_4 & - & 3x_5 & = & 2 & \quad |\cdot(-1) \\ & & & & & - & 3x_4 & + & 6x_5 & = & -3 & \quad |\cdot(-\frac{1}{3}) \\ & & & & & & & & 0 & = & a+1 & \end{array} \tag{2.29}$$

If we now multiply the second equation with (-1) and the third equation with $-\frac{1}{3}$, we obtain the **row echelon form**

$$\begin{array}{rcrcrcrcrcr} x_1 & - & 2x_2 & + & x_3 & - & x_4 & + & x_5 & = & 0 \\ & & & & x_3 & - & x_4 & + & 3x_5 & = & -2 \\ & & & & & & x_4 & - & 2x_5 & = & 1 \\ & & & & & & & & 0 & = & a+1 \end{array} \tag{2.30}$$

Only for $a = -1$, this equation system can be solved. A **particular solution** is given by

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} \tag{2.31}$$

and the **general solution**, which captures the set of all possible solutions, is given as

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} + \lambda_1 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 2 \\ 0 \\ -1 \\ 2 \\ 1 \end{bmatrix}, \quad \lambda_1, \lambda_2 \in \mathbb{R} \tag{2.32}$$

Remark 6 (Pivots and Staircase Structure)
*The leading coefficient (**pivot**) of a row (the first nonzero number from the left) is always strictly to the right of the leading coefficient of the row above it. This ensures that an equation system in row echelon form always has a "staircase" structure.*

Remark 7 (Obtaining a Particular Solution)
The row echelon form makes our lives easier when we need to determine a particular solution. To do this, we express the right-hand side of the equation system using the pivot columns, such that $\boldsymbol{b} = \sum_{i=1}^{P} \lambda_i \boldsymbol{p}_i$, *where* $\boldsymbol{p}_i$, $i = 1, \ldots, P$ *are the pivot columns. The* λ_i *are determined easiest if we start with the most-right pivot column and work our way to the left.*
In the above example, we would try to find $\lambda_1, \lambda_2, \lambda_3$ *such that*

$$\lambda_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \lambda_3 \begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} \tag{2.33}$$

From here, we find relatively directly that $\lambda_3 = 1, \lambda_2 = -1, \lambda_1 = 2$. *When we put everything together, we must not forget the non-pivot columns for which we set the coefficients implicitly to* 0. *Therefore, we get the particular solution*

$$\boldsymbol{x} = \begin{bmatrix} 2 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix}. \tag{2.34}$$

Example 2

In the following, we will go through solving a linear equation system in matrix form. Consider the problem of finding $\boldsymbol{x} = [x_1, x_2, x_3]^\top$, such that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, where

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 4 \\ 6 \\ 8 \end{bmatrix} \tag{2.35}$$

First, we write down the augmented matrix $[\boldsymbol{A} \,|\, \boldsymbol{b}]$, which is given by

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 4 & 5 & 6 & 6 \\ 7 & 8 & 9 & 8 \end{array}\right],$$

which we now transform into row echelon form using the elementary row operations:

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 4 & 5 & 6 & 6 \\ 7 & 8 & 9 & 8 \end{array}\right] \begin{array}{l} \\ -4R_1 \\ -7R_1 \end{array} \rightsquigarrow \left[\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 0 & -3 & -6 & -10 \\ 0 & -6 & -12 & -20 \end{array}\right] \begin{array}{l} \\ \cdot(-\frac{1}{3}) \\ -2R_2 \end{array}$$

$$\rightsquigarrow \left[\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & \frac{10}{3} \\ 0 & 0 & 0 & 0 \end{array}\right]$$

From the row echelon form, we see that x_3 is a free variable. To find a particular solution, we can set x_3 to any real number. For convenience, we choose $x_3 = 0$, but any other number would have worked. With $x_3 = 0$, we obtain a particular solution

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -\frac{8}{3} \\ \frac{10}{3} \\ 0 \end{bmatrix}. \tag{2.36}$$

To find the general solution, we combine the particular solution with the solution of the homogeneous equation system $\boldsymbol{A}\boldsymbol{x} = \mathbf{0}$. There are two ways of getting there: the matrix view and the equation system view. Looking at it from the matrix perspective, we need to express the third column of the row-echelon form in terms of the first two columns. This can be done by seeing that

$$\begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix} = -\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 2\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \Leftrightarrow -\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 2\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix} = \mathbf{0}. \tag{2.37}$$

We now take the coefficients $[-1, 2, -1]^\top$ of these colums that are a non-trivial representation of $\mathbf{0}$ as the solution (and any multiple of it) to the homogeneous equation system $\boldsymbol{A}\boldsymbol{x} = 0$.
An alternative and equivalent way is to remember that we wanted to solve a linear equation system, we find the solution to the homogeneous equation system by expressing x_3 in terms of x_1, x_2. From the row echelon form, we see that $x_2 + 2x_3 = 0 \Rightarrow x_3 = -\frac{1}{2}x_2$. With this, we now look at the first set of equations and obtain $x_1 + 2x_2 + 3x_3 = 0 \Rightarrow x_1 - x_3 = 0 \Rightarrow x_3 = x_1$.
Independent of whether we use the matrix or equation system view, we arrive at the general solution

$$\begin{bmatrix} -\frac{8}{3} \\ \frac{10}{3} \\ 0 \end{bmatrix} + \lambda \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \lambda \in \mathbb{R}. \tag{2.38}$$

Remark 8 (Reduced Row Echelon Form)
An equation system is in **reduced row echelon form**[15] *if*

- *It is in row echelon form*
- *Every pivot must be* 1 *and is the only non-zero entry in its column.*

The reduced row echelon form will play an important role in later sections because it allows us to determine the general solution of a linear equation system in a straightforward way.

[15] also: **row reduced echelon form** or **row canonical form**

Example: Reduced Row Echelon Form Verify that the following matrix is in reduced row echelon form:

$$A = \begin{bmatrix} 1 & 3 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 & 9 \\ 0 & 0 & 0 & 1 & -4 \end{bmatrix} \tag{2.39}$$

The pivots are colored red.
To read out the solutions of $Ax = 0$, we are mainly interested in the non-pivot columns, which we will need to express as a sum of the pivot columns. The reduced row echelon form makes this relatively straightforward, and we express the non-pivot columns in terms of sums and multiples of the pivot columns that are on their left: The second column is three times the first column (we can ignore the pivot columns on the right of the second column). Therefore, to obtain **0**, we need to subtract the second column from three times the first column. Now, we look at the fifth column, which is our second non-pivot column. The fifth column is given by 3 times the first pivot column, 9 times the second pivot column, and -4 times the third pivot column. We need to keep track of the indices of the pivot columns and translate this into 3 times the first column, 9 times the third pivot column (which is our second pivot column) and -4 times the fourth column (which is the third pivot column). Then we need to subtract the fifth column to obtain **0**—in the end, we are still solving a homogeneous equation system.
To summarize, all solutions of $Ax = 0, x \in \mathbb{R}^5$ are given by

$$\lambda_1 \begin{bmatrix} 3 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 3 \\ 0 \\ 9 \\ -4 \\ -1 \end{bmatrix}, \quad \lambda_1, \lambda_2 \in \mathbb{R}. \tag{2.40}$$

2.4.3 The Minus-1 Trick for Solving Homogeneous Equation Systems

In the following, we introduce a practical trick for reading out the solutions x of a homogeneous linear equation system $Ax = 0$, where $A \in \mathbb{R}^{k \times n}, x \in \mathbb{R}^n$.
To start, we assume that A is in reduced row echelon form without any rows that just contain zeros (e.g., after applying Gaussian elimination), i.e.,

$$A = \begin{bmatrix} 0 & \cdots & 0 & 1 & * & \cdots & * & 0 & * & \cdots & * & 0 & * & \cdots & * \\ \vdots & & \vdots & 0 & 0 & \cdots & 0 & 1 & * & \cdots & * & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & 0 & * & \ddots & * & 0 & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & * & \ddots & * & 0 & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & * & \cdots & * & 1 & * & \cdots & * \end{bmatrix} \tag{2.41}$$

Note that the columns $j_1, \dots, j_k$ with the pivots (marked red) are the standard unit vectors $e_1, \dots, e_k \in \mathbb{R}^k$.

We now extend this matrix to an $n \times n$-matrix $\tilde{\boldsymbol{A}}$ by adding $n-k$ rows of the form

$$\begin{bmatrix} 0 & \cdots & 0 & -1 & 0 & \cdots & 0 \end{bmatrix}, \tag{2.42}$$

such that the diagonal of the augmented matrix $\tilde{\boldsymbol{A}}$ contains only 1 or -1. Then, the columns of $\tilde{\boldsymbol{A}}$, which contain the -1 as pivots are solutions of the homogeneous equation system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$.[16] To be more precise, these columns form a basis (Section 2.7) of the solution space of $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$, which we will later call the **kernel** or **null space** (Section 2.9.1).

Example

Let us revisit the matrix in (2.39), which is already in reduced row echelon form:

$$\boldsymbol{A} = \begin{bmatrix} 1 & 3 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 & 9 \\ 0 & 0 & 0 & 1 & -4 \end{bmatrix}. \tag{2.43}$$

We now augment this matrix to a 5×5 matrix by adding rows of the form (2.42) at the places where the pivots on the diagonal are missing and obtain

$$\tilde{\boldsymbol{A}} = \begin{bmatrix} 1 & 3 & 0 & 0 & 3 \\ 0 & \mathbf{-1} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 9 \\ 0 & 0 & 0 & 1 & -4 \\ 0 & 0 & 0 & 0 & \mathbf{-1} \end{bmatrix} \tag{2.44}$$

From this form, we can immediately read out the solutions of $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ by taking the columns of $\tilde{\boldsymbol{A}}$, which contain -1 on the diagonal:

$$\lambda_1 \begin{bmatrix} 3 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 3 \\ 0 \\ 9 \\ -4 \\ -1 \end{bmatrix}, \quad \lambda_1, \lambda_2 \in \mathbb{R}, \tag{2.45}$$

which is identical to the solution in (2.40) that we obtained by "insight".

2.4.4 Applications of Gaussian Elimination in Linear Algebra

Gaussian elimination can also be used to find the rank of a matrix (Chapter 2.7), to calculate the determinant of a matrix (Chapter 2.10), the null space, and the inverse of an invertible square matrix. Because of its relevance to central concepts in Linear Algebra, Gaussian elimination is the most important algorithm we will cover.

[16] The proof of this trick is out of the scope of this course.

Calculating the Inverse

To compute the inverse A^{-1} of $A \in \mathbb{R}^{n\times n}$, we need to satisfy $AA^{-1} = I_n$. We can write this down as a set of simultaneous linear equations $AX = I_n$, where we solve for $X = [x_1|\cdots|x_n]$. We use the augmented matrix notation for a compact representation of this set of linear equation systems and obtain

$$\left[\begin{array}{c|c} A & I \end{array}\right] \rightsquigarrow \left[\begin{array}{c|c} I & A^{-1} \end{array}\right].$$

This means that if we bring the augmented equation system into reduced row echelon form, we can read off the inverse on the right-hand side of the equation system.

Example 1 For $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, we determine its inverse by solving the following linear equation system:

$$\left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{array}\right]$$

We bring this system now into reduced row echelon form

$$\left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{array}\right] \begin{array}{l} \\ -3R_1 \end{array} \rightsquigarrow \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & -2 & -3 & 1 \end{array}\right] \begin{array}{l} +R_2 \\ \cdot(-\frac{1}{2}) \end{array}$$
$$\rightsquigarrow \left[\begin{array}{cc|cc} 1 & 0 & -2 & 1 \\ 0 & 1 & \frac{3}{2} & -\frac{1}{2} \end{array}\right].$$

The right-hand side of this augmented equation system contains the inverse

$$A^{-1} = \begin{bmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix}. \tag{2.46}$$

Example 2 To determine the inverse of

$$A = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \tag{2.47}$$

we write down the augmented matrix

$$\left[\begin{array}{cccc|cccc} 1 & 0 & 2 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{array}\right]$$

and transform it into reduced row echelon form

$$\left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & -1 & 2 & -2 & 2 \\ 0 & 1 & 0 & 0 & 1 & -1 & 2 & -2 \\ 0 & 0 & 1 & 0 & 1 & -1 & 1 & -1 \\ 0 & 0 & 0 & 1 & -1 & 0 & -1 & 2 \end{array}\right],$$

such that the desired inverse is given as its right-hand side:

$$A^{-1} = \begin{bmatrix} -1 & 2 & -2 & 2 \\ 1 & -1 & 2 & -2 \\ 1 & -1 & 1 & -1 \\ -1 & 0 & -1 & 2 \end{bmatrix}. \tag{2.48}$$

Remark 9
You may have encountered a way of computing the inverse of a matrix using co-factors and/or cross-products. This approach only works in three dimensions and is not used in practice.

2.5 Vector Spaces

When we discussed group theory, we were looking at sets G and inner operations on G, i.e., mappings $G \times G \to G$. In the following, we will consider sets that in addition to an inner operation $+$ also contain an outer operation $\cdot$, the multiplication by a scalar $\lambda \in \mathbb{R}$.

Definition 5 (Vector space)
A real-valued **vector space** *(also called an $\mathbb{R}$-vector space) is a set V with two operations*

$$+: \quad V \times V \to V \tag{2.49}$$
$$\cdot: \quad \mathbb{R} \times V \to V \tag{2.50}$$

where

1. *$(V,+)$ is an Abelian group*
2. *Distributivity:*
   (a) $\lambda \cdot (\boldsymbol{x}+\boldsymbol{y}) = \lambda \cdot \boldsymbol{x} + \lambda \cdot \boldsymbol{y} \quad \forall \lambda \in \mathbb{R}, \boldsymbol{x}, \boldsymbol{y} \in V$
   (b) $(\lambda + \psi) \cdot \boldsymbol{x} = \lambda \cdot \boldsymbol{x} + \psi \cdot \boldsymbol{x} \quad \forall \lambda, \psi \in \mathbb{R}, \boldsymbol{x} \in V$
3. *Associativity (outer operation):* $\lambda \cdot (\psi \cdot \boldsymbol{x}) = (\lambda\psi) \cdot \boldsymbol{x} \quad \forall \lambda, \psi \in \mathbb{R}, \boldsymbol{x} \in V$
4. *Neutral element with respect to the outer operation:* $1 \cdot \boldsymbol{x} = \boldsymbol{x}, \quad \forall \boldsymbol{x} \in V$

The elements $\boldsymbol{x} \in V$ are called **vectors***. The neutral element of $(V,+)$ is the zero vector $\boldsymbol{0} = [0,\dots,0]^\top$, and the inner operation $+$ is called* **vector addition***. The elements $\lambda \in \mathbb{R}$ are called* **scalars** *and the outer operation $\cdot$ is a* **multiplication by scalars**.[17]

[17] Note: A **scalar product** is something different, and we will get to this in Section 2.13.

Remark 10
When we started the course, we defined vectors as special objects that can be added together and multiplied by scalars to yield another element of the same kind (see p. 21). Examples were geometric vectors, polynomials and $\mathbb{R}^n$. Definition 5 gives now the corresponding formal definition and applies to all kinds of vectors. We will continue focusing on vectors as elements of $\mathbb{R}^n$ because it is the most general formulation, and most algorithms are formulated in $\mathbb{R}^n$.

Remark 11
Note that a "vector multiplication" $\boldsymbol{ab}$, $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$, is not defined. Theoretically, we could define it in two ways: (a) We could define an element-wise multiplication, such that $\boldsymbol{c} = \boldsymbol{a} \cdot \boldsymbol{b}$ with $c_j = a_j b_j$. This "array multiplication" is common to many programming languages but makes mathematically only limited sense; (b) By treating vectors as $n \times 1$ matrices (which we usually do), we can use the matrix multiplication as defined in (2.11). *However, then the dimensions of the vectors do not match. Only the following multiplications for vectors are defined: $\boldsymbol{ab}^\top$ (outer product), $\boldsymbol{a}^\top\boldsymbol{b}$ (inner/scalar product).*

2.5.1 Examples

- $V = \mathbb{R}^n, n \in \mathbb{N}$ is a vector space with operations defined as follows:
  - Addition: $\boldsymbol{x}+\boldsymbol{y} = (x_1, \dots, x_n)+(y_1, \dots, y_n) = (x_1+y_1, \dots, x_n+y_n)$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$
  - Multiplication by scalars: $\lambda \boldsymbol{x} = \lambda(x_1, \dots, x_n) = (\lambda x_1, \dots, \lambda x_n)$ for all $\lambda \in \mathbb{R}, \boldsymbol{x} \in \mathbb{R}^n$
- $V = \mathbb{R}^{m \times n}, m, n \in \mathbb{N}$ is a vector space with
  - Addition: $\boldsymbol{A} + \boldsymbol{B} = \begin{bmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{bmatrix}$ is defined elementwise for all $\boldsymbol{A}, \boldsymbol{B} \in V$
  - Multiplication by scalars: $\lambda \boldsymbol{A} = \begin{bmatrix} \lambda a_{11} & \cdots & \lambda a_{1n} \\ \vdots & & \vdots \\ \lambda a_{m1} & \cdots & \lambda a_{mn} \end{bmatrix}$ as defined in Section 2.3. Remember that $\mathbb{R}^{m \times n}$ is equivalent to $\mathbb{R}^{mn}$.
- $V = \mathbb{C}$, where the addition is defined in (1.5).

Remark 12 (Notation)
The three vector spaces $\mathbb{R}^n, \mathbb{R}^{n \times 1}, \mathbb{R}^{1 \times n}$ are only different with respect to the way of writing. In the following, we will not make a distinction between $\mathbb{R}^n$ and $\mathbb{R}^{n \times 1}$, which allows us to write n-tuples as **column vectors**

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}. \tag{2.51}$$

This will simplify the notation regarding vector space operations. However, we will distinguish between $\mathbb{R}^{n\times 1}$ and $\mathbb{R}^{1\times n}$ (the **row vectors***) to avoid confusion with matrix multiplication. By default we write $\boldsymbol{x}$ to denote a column vector, and a row vector is denoted by $\boldsymbol{x}^\top$, the* **transpose** *of $\boldsymbol{x}$.*

2.5.2 Generating Set and Vector Subspaces

Definition 6 (Linear Combination)
Consider a vector space V and a finite number of vectors $\boldsymbol{x}_1,\dots,\boldsymbol{x}_k \in V$. Then, every vector $\boldsymbol{v}\in V$ of the form

$$\boldsymbol{v} = \lambda_1\boldsymbol{x}_1 + \cdots + \lambda_k\boldsymbol{x}_k = \sum_{i=1}^{k}\lambda_i\boldsymbol{x}_i \in V \tag{2.52}$$

with $\lambda_1,\dots,\lambda_k\in\mathbb{R}$ is a **linear combination** *of the vectors $\boldsymbol{x}_1,\dots,\boldsymbol{x}_k$.*

Definition 7 (Generating Set/Span)
Consider an $\mathbb{R}$-vector space V and $A = \{\boldsymbol{x}_1,\dots,\boldsymbol{x}_k\}\subset V$. If every vector $\boldsymbol{v}\in V$ can be expressed as a linear combination of $\boldsymbol{x}_1,\dots,\boldsymbol{x}_k$, A is called a **generating set** *or* **span***, which spans the vector space V. In this case, we write $V = [A]$ or $V = [\boldsymbol{x}_1,\dots,\boldsymbol{x}_k]$.*

Definition 8 (Vector Subspace)
Let V be an $\mathbb{R}$-vector space and $U\subset V$, $U\neq\emptyset$. U is called **vector subspace** *of V (or* **linear subspace***) if U is a vector space with the vector space operations $+$ and $\cdot$ restricted to $U\times U$ and $\mathbb{R}\times U$.*

Examples

- For every vector space V the trivial subspaces are V itself and $\{\boldsymbol{0}\}$.
- The solution set of a homogeneous linear equation system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ with n unknowns $\boldsymbol{x} = [x_1,\dots,x_n]^\top$ is a subspace of $\mathbb{R}^n$.
- However, the solution of an inhomogeneous equation system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, $\boldsymbol{b}\neq\boldsymbol{0}$ is not a subspace of $\mathbb{R}^n$.
- The intersection of arbitrarily many subspaces is a subspace itself.
- The intersection of all subspaces $U_i\subset V$ is called **linear hull** of V.

Remark 13
Every subspace $U\subset\mathbb{R}^n$ is the solution space of a homogeneous linear equation system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$.

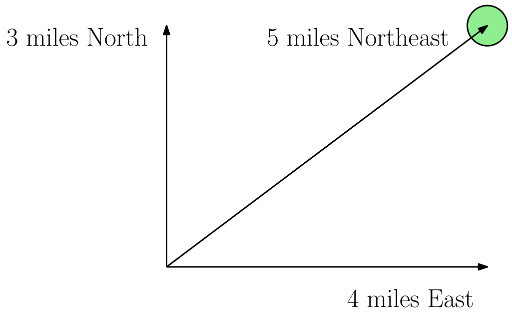


Figure 2.2: Linear dependence of three vectors in a two-dimensional space (plane).

2.6 Linear (In)Dependence

In Section 2.5, we learned about linear combinations of vectors, see (2.52). The $\mathbf{0}$ vector can always be written as the linear combination of k vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ because $\mathbf{0} = \sum_{i=1}^{k} 0\boldsymbol{x}_i$ is always true. In the following, we are interested in non-trivial linear combinations of a set of vectors to represent $\mathbf{0}$.

Definition 9 (Linear (In)dependence)
Let us consider a vector space V with $k \in \mathbb{N}$ and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in V$. If there is a non-trivial linear combination, such that $\mathbf{0} = \sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i$ with at least one $\lambda_i \neq 0$, the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ are **linearly dependent***. If only the trivial solution exists, i.e., $\lambda_1 = \ldots = \lambda_k = 0$ the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ are* **linearly independent***.*

Intuitively, a set of linearly dependent vectors contains some redundancy, whereas linearly independent vectors are all essential. Throughout this chapter, we will formalize this intuition more.

Remark 14 (From Wikipedia (2015))
A geographic example may help to clarify the concept of linear independence. A person describing the location of a certain place might say, "It is 3 miles North and 4 miles East of here." This is sufficient information to describe the location, because the geographic coordinate system may be considered as a 2-dimensional vector space (ignoring altitude and the curvature of the Earth's surface). The person might add, "The place is 5 miles Northeast of here." Although this last statement is true, it is not necessary to find this place (see Fig. 2.2 for an illustration).
In this example, the "3 miles North" vector and the "4 miles East" vector are linearly independent. That is to say, the north vector cannot be described in terms of the east vector, and vice versa. The third "5 miles Northeast" vector is a linear combination of the other two vectors, and it makes the set of vectors linearly dependent, that is, one of the three vectors is unnecessary.

Remark 15
The following properties are useful to find out whether vectors are linearly independent.

- *k vectors are either linearly dependent or linearly independent. There is no third option.*

- *If at least one of the vectors* $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k$ *is* $\boldsymbol{0}$ *then they are linearly dependent. The same holds if two vectors are identical.*
- *The vectors* $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k$, $k \geq 2$, *are linearly dependent if and only if (at least) one of them is a linear combination of the others.*
- *In a vector space* V m *linear combinations of* k *vectors* $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k$ *are linearly dependent if* $m > k$.
- *Consider an* $\mathbb{R}$*-vector space* V *with* k *vectors* $\boldsymbol{b}_1, \dots, \boldsymbol{b}_k$ *and* m *linear combinations*

$$\boldsymbol{x}_1 = \sum_{i=1}^{k} \lambda_{i1} \boldsymbol{b}_i, \tag{2.53}$$

$$\vdots \tag{2.54}$$

$$\boldsymbol{x}_m = \sum_{i=1}^{k} \lambda_{im} \boldsymbol{b}_i. \tag{2.55}$$

We want to test whether $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m$ *are linearly independent. For this purpose, we follow the general approach of testing when* $\sum_{i=1}^{m} \psi_i \boldsymbol{x}_i = \boldsymbol{0}$ *and obtain*

$$\boldsymbol{0} = \sum_{j=1}^{m} \psi_j \boldsymbol{x}_j = \sum_{j=1}^{m} \psi_j \left(\sum_{i=1}^{k} \lambda_{ij} \boldsymbol{b}_i \right) = \sum_{j=1}^{m} \sum_{i=1}^{k} \psi_j \lambda_{ij} \boldsymbol{b}_i = \sum_{i=1}^{k} \left(\sum_{j=1}^{m} \psi_j \lambda_{ij} \right) \boldsymbol{b}_i. \tag{2.56}$$

Therefore, $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m$ *are linearly independent* **if and only if** *the column vectors*

$$\hat{\boldsymbol{x}}_1 = \begin{bmatrix} \lambda_{11} \\ \vdots \\ \lambda_{k1} \end{bmatrix} \in \mathbb{R}^k, \dots, \hat{\boldsymbol{x}}_m = \begin{bmatrix} \lambda_{1m} \\ \vdots \\ \lambda_{km} \end{bmatrix} \in \mathbb{R}^k \tag{2.57}$$

are linearly independent.

Proof 3
Since $\boldsymbol{b}_1, \dots, \boldsymbol{b}_k$ *are linearly independent it follows that for all* $j = 1, \dots, m$ *we get* $\sum_{i=1}^{k} \lambda_{ij} \boldsymbol{b}_i = \boldsymbol{0}$ *with* $\lambda_{ij} = 0, i = 1, \dots, k$. *Therefore,* $\sum_{j=1}^{m} \psi_j \lambda_{ij} = 0$ *for* $i = 1, \dots, k$. *This implies* $\sum_{i=1}^{k} (\sum_{j=1}^{m} \psi_j \lambda_{ij}) \boldsymbol{b}_i = \boldsymbol{0}$. *Hence,* $\sum_{j=1}^{m} \psi_j \boldsymbol{x}_j = \boldsymbol{0}$ *is equivalent to* $\sum_{j=1}^{m} \psi_j \hat{\boldsymbol{x}}_j$.

- *A practical way of checking whether the column vectors are linearly independent is to use Gaussian elimination: Write all vectors as columns of a matrix* $\boldsymbol{A}$. *Gaussian elimination yields a matrix in (reduced) row echelon form. The pivot columns indicate the vectors, which are linearly independent of the previous*[18] *vectors (note that there is an ordering of vectors when the matrix is built). If all columns are pivot columns, the column vectors are linearly independent.*

[18] the vectors on the left

- *The non-pivot columns can be expressed as linear combinations of the columns that were before (left of) them. If the matrix is in reduced row echelon form, we can immediately see how the columns relate to each other. For instance, in*

$$\begin{bmatrix} 1 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.58}$$

first and third column are pivot columns. The second column is a non-pivot column because it is 3 times the first column. If there is at least one non-pivot column, the columns are linearly dependent.

2.6.1 Examples

- Consider $\mathbb{R}^4$ with

$$\boldsymbol{x}_1 = \begin{bmatrix} 2 \\ -3 \\ 1 \\ 4 \end{bmatrix}, \quad \boldsymbol{x}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix}, \quad \boldsymbol{x}_3 = \begin{bmatrix} -2 \\ 1 \\ -1 \\ 1 \end{bmatrix}. \tag{2.59}$$

To check whether they are linearly dependent, we follow the general approach and solve

$$\lambda_1\boldsymbol{x}_1 + \lambda_2\boldsymbol{x}_2 + \lambda_3\boldsymbol{x}_3 = \lambda_1 \begin{bmatrix} 2 \\ -3 \\ 1 \\ 4 \end{bmatrix} + \lambda_2 \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix} + \lambda_3 \begin{bmatrix} -2 \\ 1 \\ -1 \\ 1 \end{bmatrix} = \boldsymbol{0} \tag{2.60}$$

for $\lambda_1, \ldots, \lambda_3$. We write the vectors $\boldsymbol{x}_i$, $i = 1, 2, 3$, as the columns of a matrix and apply Gaussian elimination.

$$\left[\begin{array}{rrr|r} 1 & 1 & -1 & 0 \\ 2 & 1 & -2 & 0 \\ -3 & 0 & 1 & 0 \\ 4 & 2 & 1 & 0 \end{array}\right] \begin{array}{l} \\ -2R_1 \\ +3R_1 \\ -4R_1 \end{array} \rightsquigarrow \left[\begin{array}{rrr|r} 1 & 1 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 3 & -2 & 0 \\ 0 & -2 & 5 & 0 \end{array}\right] \begin{array}{l} \\ \\ +3R_2 \\ -2R_2 \end{array}$$

$$\rightsquigarrow \left[\begin{array}{rrr|r} 1 & 1 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 5 & 0 \end{array}\right] \begin{array}{l} \\ \cdot(-1) \\ \\ +\frac{5}{2}R_3 \end{array} \rightsquigarrow \left[\begin{array}{rrr|r} 1 & 1 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Here, every column of the matrix is a pivot column[19], i.e., every column is linearly independent of the columns on its left. Therefore, there is no non-trivial solution, and we require $\lambda_1 = 0, \lambda_2 = 0, \lambda_3 = 0$. Hence, the vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$ are linearly independent.

[19]Note that the matrix is not in reduced row echelon form; it also does not need to be.

- Consider a set of linearly independent vectors $\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3, \boldsymbol{b}_4 \in \mathbb{R}^n$ and

$$\begin{array}{rcrcrcrcr} \boldsymbol{x}_1 &=& \boldsymbol{b}_1 &-& 2\boldsymbol{b}_2 &+& \boldsymbol{b}_3 &-& \boldsymbol{b}_4 \\ \boldsymbol{x}_2 &=& -4\boldsymbol{b}_1 &-& 2\boldsymbol{b}_2 && &+& 4\boldsymbol{b}_4 \\ \boldsymbol{x}_3 &=& 2\boldsymbol{b}_1 &+& 3\boldsymbol{b}_2 &-& \boldsymbol{b}_3 &-& 3\boldsymbol{b}_4 \\ \boldsymbol{x}_4 &=& 17\boldsymbol{b}_1 &-& 10\boldsymbol{b}_2 &+& 11\boldsymbol{b}_3 &+& \boldsymbol{b}_4 \end{array} \tag{2.61}$$

Are the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_4 \in \mathbb{R}^n$ linearly independent? To answer this question, we investigate whether the column vectors

$$\left\{ \begin{bmatrix} 1 \\ -2 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} -4 \\ -2 \\ 0 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ -1 \\ -3 \end{bmatrix}, \begin{bmatrix} 17 \\ -10 \\ 11 \\ 1 \end{bmatrix} \right\} \tag{2.62}$$

are linearly independent. The reduced row echelon form of the corresponding linear equation system with coefficient matrix

$$\boldsymbol{A} = \begin{bmatrix} 1 & -4 & 2 & 17 \\ -2 & -2 & 3 & -10 \\ 1 & 0 & -1 & 11 \\ -1 & 4 & -3 & 1 \end{bmatrix} \tag{2.63}$$

is given as

$$\begin{bmatrix} 1 & 0 & 0 & -7 \\ 0 & 1 & 0 & -15 \\ 0 & 0 & 1 & -18 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{2.64}$$

From the reduced row echelon form, we see that the corresponding linear equation system is non-trivially solvable: The last column is not a pivot column, and $\boldsymbol{x}_4 = -7\boldsymbol{x}_1 - 15\boldsymbol{x}_2 - 18\boldsymbol{x}_3$. Therefore, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_4$ are linearly dependent as $\boldsymbol{x}_4$ lies in the span of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_3$.

2.7 Basis and Dimension

In a vector space V, we are particularly interested in the set of linearly independent vectors A that possesses the property that any vector $\boldsymbol{v} \in V$ can be obtained by a linear combination of vectors in A.

Definition 10 (Basis)
Consider a real vector space V and $A \subset V$

- *A generating set A of V is called* **minimal** *if there exists no smaller set $\tilde{A} \subset A \subset V$, which spans V.*
- *Every linearly independent generating set of V is minimal and is called* **basis** *of V.*

Let V be a real vector space and $B \subset V, B \neq \emptyset$. Then, the following statements are equivalent:

- B is basis of V
- B is a minimal generating set
- B is a maximal linearly independent subset of V.
- Every vector $\boldsymbol{x} \in V$ is a linear combination of vectors from B, and every linear combination is unique, i.e., with

$$\boldsymbol{x} = \sum_{i=1}^{k} \lambda_i \boldsymbol{b}_i = \sum_{i=1}^{k} \psi_i \boldsymbol{b}_i \tag{2.65}$$

and $\lambda_i, \psi_i \in \mathbb{R}$, $\boldsymbol{b}_i \in B$ it follows that $a_i = b_i, i = 1, \ldots, k$.

2.7.1 Examples

- In $\mathbb{R}^3$, the canonical/standard basis is

$$B = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}. \tag{2.66}$$

- Different bases in $\mathbb{R}^3$ are

$$B_1 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}, \qquad B_2 = \left\{ \begin{bmatrix} 0.53 \\ 0.86 \\ -0.43 \end{bmatrix}, \begin{bmatrix} 1.83 \\ 0.31 \\ 0.34 \end{bmatrix}, \begin{bmatrix} -2.25 \\ -1.30 \\ 3.57 \end{bmatrix} \right\} \tag{2.67}$$

- The set

$$A = \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ -4 \end{bmatrix} \right\} \tag{2.68}$$

is linearly independent, but not a generating set (and no basis): For instance, the vector $[1, 0, 0, 0]^\top$ cannot be obtained by a linear combination of elements in A.

Remark 16

- *Every vector space V possesses a basis B.*
- *The examples above show that there can be many bases of a vector space V, i.e., there is no unique basis. However, all bases possess the same number of elements, the* **basis vectors**.
- *We only consider finite-dimensional vector spaces V. In this case, the* **dimension of V** *is the number of basis vectors, and we write $\dim(V)$.*
- *If $U \subset V$ is a subspace of V then $\dim(U) \leq \dim(V)$ and $\dim(U) = \dim(V)$ if and only if $U = V$.*

2.7.2 Example: Determining a Basis

- For a vector subspace $U \subset \mathbb{R}^5$, spanned by the vectors

$$x_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \\ -1 \\ -1 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 2 \\ -1 \\ 1 \\ 2 \\ -2 \end{bmatrix}, \quad x_3 = \begin{bmatrix} 3 \\ -4 \\ 3 \\ 5 \\ -3 \end{bmatrix}, \quad x_4 = \begin{bmatrix} -1 \\ 8 \\ -5 \\ -6 \\ 1 \end{bmatrix} \in \mathbb{R}^5, \tag{2.69}$$

we are interested in finding out which vectors $x_1, \ldots, x_4$ are a basis for U. For this, we need to check whether $x_1, \ldots, x_4$ are linearly independent. Therefore, we need to solve

$$\sum_{i=1}^{4} \lambda_i x_i = \mathbf{0}, \tag{2.70}$$

which leads to a homogeneous equation system with the corresponding matrix

$$\left[x_1|x_2|x_3|x_4\right] = \begin{bmatrix} 1 & 2 & 3 & -1 \\ 2 & -1 & -4 & 8 \\ -1 & 1 & 3 & -5 \\ -1 & 2 & 5 & -6 \\ -1 & -2 & -3 & 1 \end{bmatrix}. \tag{2.71}$$

With the basic transformation of linear equation systems, we obtain

$$\left[\begin{array}{cccc|l} 1 & 2 & 3 & -1 & \\ 2 & -1 & -4 & 8 & -2R_1 \\ -1 & 1 & 3 & -5 & +R_1 \\ -1 & 2 & 5 & -6 & +R_1 \\ -1 & -2 & -3 & 1 & +R_1 \end{array}\right] \rightsquigarrow \left[\begin{array}{cccc|l} 1 & 2 & 3 & -1 & \\ 0 & -5 & -10 & 10 & \cdot(-\frac{1}{5}) \\ 0 & 3 & 6 & -6 & \cdot\frac{1}{3}| - R_2 \\ 0 & 4 & 8 & -7 & -4R_2 \\ 0 & 0 & 0 & 0 & \end{array}\right]$$

$$\rightsquigarrow \left[\begin{array}{cccc|l} 1 & 2 & 3 & -1 & \\ 0 & 1 & 2 & -2 & +R_4 \\ 0 & 0 & 0 & 0 & +2R_4 \\ 0 & 0 & 0 & 1 & \text{swap with } R_3 \\ 0 & 0 & 0 & 0 & \end{array}\right] \rightsquigarrow \left[\begin{array}{cccc|l} 1 & 2 & 3 & 0 & -2R_2 \\ 0 & 1 & 2 & 0 & \\ 0 & 0 & 0 & 1 & \\ 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & \end{array}\right]$$

$$\rightsquigarrow \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

From this reduced-row echelon form we see that x_1, x_2, x_4 are linearly independent (because the linear equation system $\lambda_1 x_1 + \lambda_2 x_2 + \lambda_4 x_4 = \mathbf{0}$ can only be solved with $\lambda_1 = \lambda_2 = \lambda_4 = 0$). Therefore, $\{x_1, x_2, x_4\}$ is a basis of U.

- Let us now consider a slightly different problem: Instead of finding out which vectors $\boldsymbol{x}_1,\dots,\boldsymbol{x}_4$ of the span of U form a basis, we are interested in finding a "simple" basis for U. Here, "simple" means that we are interested in basis vectors with many coordinates equal to 0.

  To solve this problem we replace the vectors $\boldsymbol{x}_1,\dots,\boldsymbol{x}_4$ with suitable linear combinations. **In practice,** we write $\boldsymbol{x}_1,\dots,\boldsymbol{x}_4$ as *row vectors* in a matrix and perform Gaussian elimination:

$$\begin{bmatrix} 1 & 2 & -1 & -1 & -1 \\ 2 & -1 & 1 & 2 & -2 \\ 3 & -4 & 3 & 5 & -3 \\ -1 & 8 & -6 & -6 & 1 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 2 & -1 & -1 & -1 \\ 0 & -5 & 3 & 4 & 0 \\ 0 & -10 & 6 & 8 & 0 \\ 0 & 10 & -6 & -7 & 0 \end{bmatrix}$$

$$\rightsquigarrow \begin{bmatrix} 1 & 2 & -1 & -1 & -1 \\ 0 & 1 & -\frac{3}{5} & -\frac{4}{5} & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 2 & -1 & 0 & -1 \\ 0 & 1 & -\frac{3}{5} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\rightsquigarrow \begin{bmatrix} 1 & 0 & \frac{1}{5} & 0 & -1 \\ 0 & 1 & -\frac{3}{5} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

  From the reduced row echelon form, the simple basis vectors are the rows with the leading 1s (the "steps").

$$U = [\underbrace{\begin{bmatrix} 1 \\ 0 \\ \frac{1}{5} \\ 0 \\ -1 \end{bmatrix}}_{\boldsymbol{b}_1}, \underbrace{\begin{bmatrix} 0 \\ 1 \\ -\frac{3}{5} \\ 0 \\ 0 \end{bmatrix}}_{\boldsymbol{b}_2}, \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}}_{\boldsymbol{b}_3}] \tag{2.72}$$

  and $B = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3\}$ is a (simple) basis of U (check that they are linearly independent!).

2.7.3 Rank

- The number of linearly independent columns of a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ equals the number of linearly independent rows and is called **rank** of $\boldsymbol{A}$ and is denoted by $\text{rk}(\boldsymbol{A})$.
- $\text{rk}(\boldsymbol{A}) = \text{rk}(\boldsymbol{A}^\top)$, i.e., the column rank equals the row rank.
- The columns of $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ span a subspace $U \subset \mathbb{R}^m$ with $\dim(U) = \text{rk}(A)$
- A basis of a subspace $U = [\boldsymbol{x}_1,\dots,\boldsymbol{x}_m] \subset \mathbb{R}^n$ can be found by executing the following steps:

1. Write the spanning vectors as columns of a matrix $\boldsymbol{A}$
2. Apply Gaussian elimination algorithm to $\boldsymbol{A}$.
3. The spanning vectors associated with the pivot columns form a basis of U.

- The rows of $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ span a subspace $W \subset \mathbb{R}^n$ with $\dim(W) = \text{rk}(\boldsymbol{A})$. A basis of W can be found by applying the Gaussian elimination algorithm to the rows of $\boldsymbol{A}$ (or the columns of $\boldsymbol{A}^\top$).
- For all $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ holds: $\boldsymbol{A}$ is regular (invertible) if and only if $\text{rk}(\boldsymbol{A}) = n$.
- For all $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and all $\boldsymbol{b} \in \mathbb{R}^m$: The linear equation system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ can be solved if and only if $\text{rk}(\boldsymbol{A}) = \text{rk}(\boldsymbol{A}|\boldsymbol{b})$, where $\boldsymbol{A}|\boldsymbol{b}$ denotes the "extended" system.
- For $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ the space of solutions for $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ possesses dimension $n - \text{rk}(A)$.
- A matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ has **full rank** if its rank equals the largest possible for a matrix of the same dimensions, which is the lesser of the number of rows and columns, i.e., $\text{rk}(\boldsymbol{A}) = \min(m, n)$. A matrix is said to be **rank deficient** if it does not have full rank.

Examples

- $\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$. $\boldsymbol{A}$ possesses two linearly independent rows (and columns). Therefore, $\text{rk}(\boldsymbol{A}) = 2$.
- $\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 8 & 12 \end{bmatrix}$. We see that the second row is a multiple of the first row, such that the row-echelon form of $\boldsymbol{A}$ is $\begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \end{bmatrix}$, and $\text{rk}(\boldsymbol{A}) = 1$.
- $\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix}$ We use Gaussian elimination to determine the rank:

$$\begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix} \begin{matrix} \\ \\ +R_1 - R_2 \end{matrix} \rightsquigarrow \begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{matrix} \\ +2R_1 \\ \\ \end{matrix}$$

$$\rightsquigarrow \begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & 3 \\ 0 & 0 & 0 \end{bmatrix}$$

Here, we see that the number of linearly independent rows and columns is 2, such that $\text{rk}(\boldsymbol{A}) = 2$.

2.8 Intersection of Subspaces

In the following, we consider two approaches to determining a basis of the intersection $U_1 \cap U_2$ of two subspaces $U_1, U_2 \subset V$. This means, we are interested in finding all $\boldsymbol{x} \in V$, such that $\boldsymbol{x} \in U_1$ and $\boldsymbol{x} \in U_2$.

2.8.1 Approach 1

Consider $U_1 = [\boldsymbol{b}_1, \dots, \boldsymbol{b}_k] \subset V$ and $U_2 = [\boldsymbol{c}_1, \dots, \boldsymbol{c}_l] \subset V$. We know that and $\boldsymbol{x} \in U_1$ can be represented as a linear combination $\sum_{i=1}^{k} \lambda_i \boldsymbol{b}_i$ of the basis vectors (or spanning vectors) $\boldsymbol{b}_1, \dots, \boldsymbol{b}_k$. Equivalently $\boldsymbol{x} = \sum_{j=1}^{l} \psi_j \boldsymbol{c}_j$. Therefore, the approach is to find $\lambda_1, \dots, \lambda_k$ and/or $\psi_1, \dots, \psi_l$, such that

$$\sum_{i=1}^{k} \lambda_i \boldsymbol{b}_i = \boldsymbol{x} = \sum_{j=1}^{l} \psi_j \boldsymbol{c}_j \tag{2.73}$$

$$\Leftrightarrow \sum_{i=1}^{k} \lambda_i \boldsymbol{b}_i - \sum_{j=1}^{l} \psi_j \boldsymbol{c}_j = \boldsymbol{0}. \tag{2.74}$$

For this, we write the basis vectors into a matrix

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{b}_1 & \cdots & \boldsymbol{b}_k & -\boldsymbol{c}_1 & \cdots & -\boldsymbol{c}_l \end{bmatrix} \tag{2.75}$$

and solve the linear equation system

$$\boldsymbol{A} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_k \\ \psi_1 \\ \vdots \\ \psi_l \end{bmatrix} = \boldsymbol{0} \tag{2.76}$$

to find either $\lambda_1, \dots, \lambda_k$ or $\psi_1, \dots, \psi_l$, which we can then use to determine $U_1 \cap U_2$.

Example

We consider

$$U_1 = [\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}] \subset \mathbb{R}^4, \quad U_2 = [\begin{bmatrix} -1 \\ 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}] \subset \mathbb{R}^4. \tag{2.77}$$

To find a basis of $U_1 \cap U_2$, we need to find all $\boldsymbol{x} \in V$ that can be represented as linear combinations of the basis vectors of U_1 and U_2, i.e.,

$$\sum_{i=1}^{3} \lambda_i \boldsymbol{b}_i = \boldsymbol{x} = \sum_{j=1}^{2} \psi_j \boldsymbol{c}_j, \tag{2.78}$$

where $\boldsymbol{b}_i$ and $\boldsymbol{c}_j$ are the basis vectors of U_1 and U_2, respectively. The matrix $\boldsymbol{A} = [\boldsymbol{b}_1|\boldsymbol{b}_2|\boldsymbol{b}_3|-\boldsymbol{c}_1|-\boldsymbol{c}_2]$ from (2.75) is given as

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & -1 & -1 \\ 0 & 1 & 1 & -2 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}. \tag{2.79}$$

By using Gaussian elimination, we determine the corresponding reduced row echelon form

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -2 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{2.80}$$

We keep in mind that we are interested in finding $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$ and/or $\psi_1, \psi_2 \in \mathbb{R}$ with

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -2 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \psi_1 \\ \psi_2 \end{bmatrix} = \boldsymbol{0}. \tag{2.81}$$

From here, we can immediately see that $\psi_2 = 0$ and $\psi_1 \in \mathbb{R}$ is a free variable since it corresponds to a non-pivot column, and our solution is

$$U_1 \cap U_2 = \psi_1 \boldsymbol{c}_1 = [\begin{bmatrix} -1 \\ 1 \\ 2 \\ 0 \end{bmatrix}], \quad \psi_1 \in \mathbb{R}. \tag{2.82}$$

Remark 17
Alternatively, we could have used $\lambda_1 = -\psi_1, \lambda_2 = \psi_1, \lambda_3 = 0$ *and determined the (same) solution via the basis vectors of* U_1 *as*

$$\psi_1 \left(-\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + 2\begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \right) = \psi_1 \begin{bmatrix} -1 \\ 1 \\ 2 \\ 0 \end{bmatrix} = [\begin{bmatrix} -1 \\ 1 \\ 2 \\ 0 \end{bmatrix}], \quad \psi_1 \in \mathbb{R}. \tag{2.83}$$

2.8.2 Approach 2

In the second approach, we exploit Remark 13, which says that any subspace is the solution of a homogeneous linear equation system, to determine the intersection $U_1 \cap U_2$ of two subspaces $U_1, U_2 \subset \mathbb{R}^n$.
First, we show how to determine the linear equation system that generates a subspace; second, we exploit these insights to find $U_1 \cap U_2$.

Lemma 1
Consider $U = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_m] \subset \mathbb{R}^n$ *and* $\dim(U) = r$. *We write the vectors* $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m$ *as rows of a matrix*

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{x}_1^\top \\ \vdots \\ \boldsymbol{x}_m^\top \end{bmatrix} \in \mathbb{R}^{m \times n} \tag{2.84}$$

and investigate the homogeneous linear equation system $\boldsymbol{A}\boldsymbol{y} = \boldsymbol{0}$. *First, the solution space* V *possesses dimension* $k = n - \mathrm{rk}(\boldsymbol{A}) = n - \dim(U) = n - r$. *Second, we choose a basis* $(\boldsymbol{b}_1, \dots, \boldsymbol{b}_k)$ *in this solution space and again write these basis vectors as the rows of a matrix*

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{b}_1^\top \\ \vdots \\ \boldsymbol{b}_k^\top \end{bmatrix} \in \mathbb{R}^{k \times n} \tag{2.85}$$

with $\mathrm{rk}(\boldsymbol{B}) = k$. *Then* U *is the solution space of* $\boldsymbol{B}\boldsymbol{y} = \boldsymbol{0}$.

Proof 4
Define S_h *as the solution space of* $\boldsymbol{B}\boldsymbol{y} = \boldsymbol{0}$. *It holds that* $\dim(S_h) = n - \mathrm{rk}(\boldsymbol{B}) = n - k = r$. *Therefore,* $\dim(S_h) = \dim(U)$. *From* $\boldsymbol{A}\boldsymbol{b}_j = \boldsymbol{0}$, $j = 1, \dots, k$ *it follows that* $\boldsymbol{x}_i^\top \boldsymbol{b}_j = 0$ *for* $i = 1, \dots, m$ *and* $j = 1, \dots, k$ *(remember how matrix-vector multiplication works), and at the same time* $\boldsymbol{b}_j^\top \boldsymbol{x}_i = 0$. *Therefore,* $\boldsymbol{B}\boldsymbol{x}_i = \boldsymbol{0}$, $i = 1, \dots, m$ *and, hence,* $U \subset S_h$. *However, since* $\dim(S_h) = \dim(U)$ *it follows that* $S_h = U$.

Practical Algorithm

Let us summarize the main steps to determine $U_1 \cap U_2$:

1. Write U_1, U_2 as solution spaces of two linear equation systems $\boldsymbol{B}_1\boldsymbol{x} = \boldsymbol{0}$ and $\boldsymbol{B}_2\boldsymbol{x} = \boldsymbol{0}$:

   (a) Write spanning vectors of U_1, U_2 as the rows of two matrices $\boldsymbol{A}_1, \boldsymbol{A}_2$, respectively.

   (b) Determine S_1 as the solution of $\boldsymbol{A}_1\boldsymbol{x} = \boldsymbol{0}$ and S_2 as the solution of $\boldsymbol{A}_2\boldsymbol{x} = \boldsymbol{0}$

   (c) Write spanning vectors of S_1 and S_2 as the rows of the matrices $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$, respectively.

2. $U_1 \cap U_2$ is the solution space of $\boldsymbol{C}\boldsymbol{x} = \boldsymbol{0}$, where $\boldsymbol{C} = \begin{bmatrix} \boldsymbol{B}_1 \\ \boldsymbol{B}_2 \end{bmatrix}$, which we find by means of Gaussian elimination.

Example 1

To determine the intersection of two subspaces $U_1, U_2 \subset \mathbb{R}^n$, we use the above method. We consider again the subspaces $U_1, U_2 \subset \mathbb{R}^4$ from the example above (and hopefully, we end up with the same solution):

$$U_1 = [\begin{bmatrix}1\\1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\1\end{bmatrix}] \subset \mathbb{R}^4, \quad U_2 = [\begin{bmatrix}-1\\1\\2\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\0\end{bmatrix}] \subset \mathbb{R}^4. \tag{2.86}$$

1. To determine the intersection $U_1 \cap U_2$, we first write U_1, U_2 as solution spaces of linear equation systems.

   (a) We write the spanning vectors of U_1, U_2 as the rows of the matrices

   $$\boldsymbol{A}_1 = \begin{bmatrix}1 & 1 & 0 & 0\\0 & 1 & 1 & 0\\0 & 0 & 1 & 1\end{bmatrix}, \quad \boldsymbol{A}_2 = \begin{bmatrix}-1 & 1 & 2 & 0\\0 & 1 & 0 & 0\end{bmatrix}, \tag{2.87}$$

   respectively.

   (b) We use Gaussian elimination to determine the corrsponding reduced row echelon forms

   $$\tilde{\boldsymbol{A}}_1 = \begin{bmatrix}1 & 0 & 0 & 1\\0 & 1 & 0 & -1\\0 & 0 & 1 & 1\end{bmatrix}, \quad \tilde{\boldsymbol{A}}_2 = \begin{bmatrix}1 & 0 & -2 & 0\\0 & 1 & 0 & 0\end{bmatrix}. \tag{2.88}$$

   Third, we determine the solution spaces of $\boldsymbol{A}_1\boldsymbol{x} = \boldsymbol{0}$ and $\boldsymbol{A}_2\boldsymbol{x} = \boldsymbol{0}$, e.g., using the Minus-1 Trick from Section 2.4.3, as

   $$S_1 = [\begin{bmatrix}1\\-1\\1\\-1\end{bmatrix}], \quad S_2 = [\begin{bmatrix}2\\0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\1\end{bmatrix}]. \tag{2.89}$$

   (c) U_1 is now the solution space of the linear equation system $\boldsymbol{B}_1\boldsymbol{x} = \boldsymbol{0}$ with

   $$\boldsymbol{B}_1 = \begin{bmatrix}1 & -1 & 1 & -1\end{bmatrix}, \tag{2.90}$$

   and U_2 is the solution space of the linear equation system $\boldsymbol{B}_2\boldsymbol{x} = \boldsymbol{0}$ with

   $$\boldsymbol{B}_2 = \begin{bmatrix}2 & 0 & 1 & 0\\0 & 0 & 0 & 1\end{bmatrix}. \tag{2.91}$$

2. $U_1 \cap U_2$ is the solution space of the linear equation system $\boldsymbol{C}\boldsymbol{x} = \boldsymbol{0}$ with

$$\boldsymbol{C} = \begin{bmatrix}\boldsymbol{B}_1\\\boldsymbol{B}_2\end{bmatrix} = \begin{bmatrix}1 & -1 & 1 & -1\\2 & 0 & 1 & 0\\0 & 0 & 0 & 1\end{bmatrix}. \tag{2.92}$$

To determine this solution space, we follow the standard procedure of (a) computing the reduced row echeolon form

$$\begin{bmatrix}1 & 0 & 0 & 0 & -1\\ 0 & 1 & 0 & -1 & -1\\ 0 & 0 & 1 & -1 & -1\end{bmatrix} \tag{2.93}$$

using Gaussian elimination and (b) finding the (general) solution using the Minus-1 Trick from Section 2.4.3 as

$$U_1 \cap U_2 = [\begin{bmatrix}-1\\1\\2\\0\end{bmatrix}], \tag{2.94}$$

which is identical to the solution in (2.83) found by using Approach 1.

Example 2

We apply again Approach 2 and consider the two subspaces $U_1, U_2 \subset \mathbb{R}^5$, where

$$U_1 = [\begin{bmatrix}1\\-1\\-1\\-2\\1\end{bmatrix}, \begin{bmatrix}0\\3\\3\\3\\0\end{bmatrix}, \begin{bmatrix}1\\-3\\1\\-2\\4\end{bmatrix}], \quad U_2 = [\begin{bmatrix}-1\\0\\-4\\-5\\1\end{bmatrix}, \begin{bmatrix}-5\\-1\\2\\2\\-6\end{bmatrix}, \begin{bmatrix}1\\2\\-1\\3\\2\end{bmatrix}, \begin{bmatrix}3\\1\\0\\3\\3\end{bmatrix}]. \tag{2.95}$$

1. To determine the intersection $U_1 \cap U_2$, we first write U_1, U_2 as solution spaces of linear equation systems.

   (a) We write the spanning vectors of U_1, U_2 as the rows of the matrices

$$\boldsymbol{A}_1 = \begin{bmatrix}1 & -1 & -1 & -2 & 1\\ 0 & 3 & 3 & 3 & 0\\ 1 & -3 & 1 & -2 & 4\end{bmatrix}, \quad \boldsymbol{A}_2 = \begin{bmatrix}-1 & 0 & -4 & -5 & 1\\ -5 & -1 & 2 & 2 & -6\\ 1 & 2 & -1 & 3 & 2\\ 3 & 1 & 0 & 3 & 3\end{bmatrix}, \tag{2.96}$$

   respectively.

   (b) We use Gaussian elimination to determine the corrsponding reduced row echelon forms

$$\tilde{\boldsymbol{A}}_1 = \begin{bmatrix}1 & 0 & 0 & -1 & 1\\ 0 & 1 & 0 & \frac{1}{2} & -\frac{3}{4}\\ 0 & 0 & 1 & \frac{1}{2} & \frac{3}{4}\end{bmatrix}, \quad \tilde{\boldsymbol{A}}_2 = \begin{bmatrix}1 & 0 & 0 & 0 & \frac{12}{13}\\ 0 & 1 & 0 & 0 & \frac{6}{13}\\ 0 & 0 & 1 & 0 & -\frac{5}{13}\\ 0 & 0 & 0 & 1 & -\frac{1}{13}\end{bmatrix}. \tag{2.97}$$

(c) We determine the solution spaces of $A_1x = 0$ and $A_2x = 0$, e.g., using the Minus-1 Trick from Section 2.4.3, as

$$S_1 = [\begin{bmatrix}-2\\1\\1\\-2\\0\end{bmatrix}, \begin{bmatrix}4\\-3\\3\\0\\-4\end{bmatrix}], \quad S_2 = [\begin{bmatrix}12\\6\\-5\\-1\\-13\end{bmatrix}]. \tag{2.98}$$

U_1 is the solution space of the linear equation system $B_1x = 0$ with

$$B_1 = \begin{bmatrix}-2 & 1 & 1 & -2 & 0\\ 4 & -3 & 3 & 0 & -4\end{bmatrix} \tag{2.99}$$

and U_2 is the solution space of the linear equation system $B_2x = 0$ with

$$B_2 = \begin{bmatrix}12 & 6 & -5 & -1 & -13\end{bmatrix}. \tag{2.100}$$

2. $U_1 \cap U_2$ is the solution space of the linear equation system $Cx = 0$ with

$$C = \begin{bmatrix}B_1\\B_2\end{bmatrix} = \begin{bmatrix}-2 & 1 & 1 & -2 & 0\\ 4 & -3 & 3 & 0 & -4\\ 12 & 6 & -5 & -1 & -13\end{bmatrix}. \tag{2.101}$$

To determine this solution space, we follow the standard procedure of (a) computing the reduced row echeolon form

$$\begin{bmatrix}1 & 0 & 0 & 0 & -1\\ 0 & 1 & 0 & -1 & -1\\ 0 & 0 & 1 & -1 & -1\end{bmatrix} \tag{2.102}$$

using Gaussian elimination and (b) finding the (general) solution using the Minus-1 Trick from Section 2.4.3 as

$$U_1 \cap U_2 = [\begin{bmatrix}0\\1\\1\\1\\0\end{bmatrix}, \begin{bmatrix}1\\1\\1\\0\\1\end{bmatrix}]. \tag{2.103}$$

2.9 Linear Mappings

In the following, we will study mappings on vector spaces that preserve their structure: Consider two real vector spaces V, W. A mapping $\Phi : V \to W$ preserves the structure of the vector space if

$$\Phi(x + y) = \Phi(x) + \Phi(y) \tag{2.104}$$

$$\Phi(\lambda x) = \lambda \Phi(x) \tag{2.105}$$

for all $x, y \in V$ and $\lambda \in \mathbb{R}$. We can summarize this in the following definition:

Definition 11 (Linear Mapping)
For real vector spaces V, W, a mapping $\Phi : V \to W$ is called **linear** *(or* **vector space homomorphism***) if*

$$\Phi(\lambda \boldsymbol{x} + \psi \boldsymbol{y}) = \lambda \Phi(\boldsymbol{x}) + \psi \Phi(\boldsymbol{y}) \tag{2.106}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in V$ and $\lambda, \psi \in \mathbb{R}$.

Important special cases:

- **Isomorphism:** $\Phi : V \to W$ linear and bijective
- **Endomorphism:** $\Phi : V \to V$ linear
- **Automorphism:** $\Phi : V \to V$ linear and bijective
- We define $\mathrm{id}_V : V \to V$, $\boldsymbol{x} \mapsto \boldsymbol{x}$ as the **identity mapping** in V.

Example: Homomorphism

The mapping $\Phi : \mathbb{R}^2 \to \mathbb{C}$, $\Phi(\boldsymbol{x}) = x_1 + ix_2$, is a homomorphism:

$$\begin{aligned}
&\Phi\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} y_1 \\ x_2 \end{bmatrix}\right) = (x_1 + y_1) + i(x_2 + y_2) = x_1 + ix_2 + y_1 + iy_2 = \Phi\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) + \Phi\left(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right) \\
&\Phi\left(\lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \lambda x_1 + \lambda i x_2 = \lambda(x_1 + ix_2) = \lambda \Phi\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)
\end{aligned} \tag{2.107}$$

We have already discussed the representation of complex numbers as tuples in $\mathbb{R}^2$, but now we know why we can do this: There is a bijective linear mapping (we only showed linearity, but not the bijection) that converts the elementwise addition of tuples in $\mathbb{R}^2$ the set of complex numbers with the corresponding addition.

2.9.1 Image and Kernel (Null Space)

Definition 12
Image and Kernel

For $\Phi : V \to W$, we define the **kernel/null space**

$$\mathrm{Ker}(\Phi) := \Phi^{-1}(\{\boldsymbol{0}\}) = \{\boldsymbol{v} \in V : \Phi(\boldsymbol{v}) = \boldsymbol{0}\} \tag{2.108}$$

and the **image**

$$\mathrm{Im}(\Phi) := \Phi(V) = \{\boldsymbol{w} \in W | \exists \boldsymbol{v} \in V : \Phi(\boldsymbol{v}) = \boldsymbol{w}\}. \tag{2.109}$$

An illustration is given in Figure 2.3.

Remark 18
Consider a linear mapping $\Phi : V \to W$, where V, W are vector spaces.

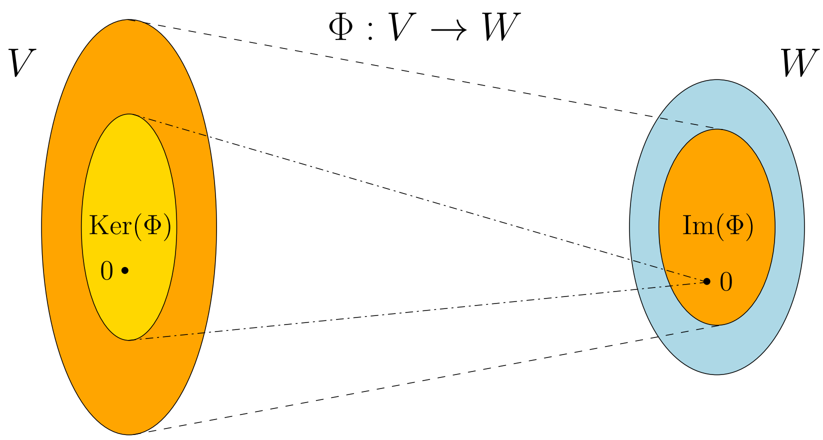


Figure 2.3: Kernel and Image of a linear mapping $\Phi : V \to W$.

- *It always holds that* $\Phi(\{\mathbf{0}_V\}) = \mathbf{0}_W$ *and, therefore,* $\mathbf{0}_v \in \ker(\Phi)$. *In particular, the null space is never empty.*
- $\mathrm{Im}(\Phi) \subset W$ *is a subspace of* W*, and* $\ker(\Phi) \subset V$ *is a subspace of* V.
- Φ *is injective (one-to-one) if and only if* $\ker(\Phi) = \{\mathbf{0}\}$

Remark 19
For $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ *the mapping* $\Phi : \mathbb{R}^n \to \mathbb{R}^m$, $\boldsymbol{x} \mapsto \boldsymbol{Ax}$ *is linear. For* $\boldsymbol{A} = (\boldsymbol{a}_1|...|\boldsymbol{a}_n)$ *we obtain*

$$\mathrm{Im}(\Phi) = \{\boldsymbol{Ax}|\boldsymbol{x} \in \mathbb{R}^n\} = \{\lambda_1\boldsymbol{a}_1 + ... + \lambda_n\boldsymbol{a}_n|\lambda_1,\ldots,\lambda_n \in \mathbb{R}\} = [\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n] \subset \mathbb{R}^m, \quad (2.110)$$

i.e., the image is the span of the columns of $\boldsymbol{A}$*, also called the* **column space**.
The kernel/null space $\ker(\Phi)$ *is the general solution to the linear homogeneous equation system* $\boldsymbol{Ax} = \mathbf{0}$.

Example: Image and Kernel of a Linear Mapping

The mapping

$$\Phi : \mathbb{R}^4 \to \mathbb{R}^2, \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \mapsto \begin{bmatrix} 1 & 2 & -1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_1 + 2x_2 - x_3 \\ x_1 + x_4 \end{bmatrix} \quad (2.111)$$

$$= x_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (2.112)$$

is linear. To determine $\mathrm{Im}(\Phi)$ we can simply take the span of the columns of the transformation matrix and obtain

$$\mathrm{Im}(\Phi) = [\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}]. \quad (2.113)$$

To compute the kernel (null space) of Φ, we need to solve $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$, i.e., we need to solve a homogeneous equation system. To do this, we use Gaussian elimination to transform $\boldsymbol{A}$ into reduced row echelon form:

$$\begin{bmatrix} 1 & 2 & -1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 2 & -1 & 0 \end{bmatrix} \begin{matrix} \\ -R_1 | \cdot (\frac{1}{2}) \end{matrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix}$$

This matrix is now in reduced row echelon form, and we can now use the Minus-1 Trick to compute a basis of the kernel (see Section 2.4.3). Alternatively, we can express the non-pivot columns (columns 3 an 4) as linear combinations of the pivot-columns (columns 1 and 2). The third column $\boldsymbol{a}_3$ is equivalent to $-\frac{1}{2}$ times the second column $\boldsymbol{a}_2$. Therefore, $\boldsymbol{0} = \boldsymbol{a}_3 + \frac{1}{2}\boldsymbol{a}_2$. In the same way, we see that $\boldsymbol{a}_4 = \boldsymbol{a}_1 - \frac{1}{2}\boldsymbol{a}_2$ and, therefore, $\boldsymbol{0} = \boldsymbol{a}_1 - \frac{1}{2}\boldsymbol{a}_2 - \boldsymbol{a}_4$. This gives us now the kernel (null space) as

$$\ker(\Phi) = [\begin{bmatrix} 0 \\ \frac{1}{2} \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ \frac{1}{2} \\ 0 \\ 1 \end{bmatrix}]. \tag{2.114}$$

Theorem 3 (Rank-Nullity Theorem)
For vector spaces V, W and a linear mapping $\Phi : V \to W$ it holds that

$$\dim(\ker(\Phi)) + \dim(\mathrm{Im}(\Phi)) = \dim(V) \tag{2.115}$$

Remark 20
Consider $\mathbb{R}$-vector spaces V, W, X. Then:

- *For linear mappings $\Phi : V \to W$ and $\Psi : W \to X$ the mapping $\Psi \circ \Phi : V \to X$ is also linear.*
- *If $\Phi : V \to W$ is an isomorphism then $\Phi^{-1} : W \to V$ is an isomorphism as well.*
- *If $\Phi, \Psi : V \to W$ are linear then $\Phi + \Psi$ and $\lambda\Phi, \lambda \in \mathbb{R}$ are linear, too.*
- *For a linear mapping $\Phi : V \to W$ the null space (kernel) captures all possible linear combinations of the elements in V that produce $\boldsymbol{0} \in W$.*

Theorem 4
Finite-dimensional $\mathbb{R}$-vector spaces V and W are isomorph if and only if $\dim(V) = \dim(W)$.

2.9.2 Matrices to Represent Linear Mappings

Any n-dimensional $\mathbb{R}$-vector space is isomorph to $\mathbb{R}^n$ (Theorem 4). If we define a basis $\{\boldsymbol{b}_1, \dots, \boldsymbol{b}_n\}$ of V we can construct an isomorphism concretely. In the following, the order of the basis vectors will be important. Therefore, we write

$$B = (\boldsymbol{b}_1, \dots, \boldsymbol{b}_n) \tag{2.116}$$

and call this n-tuple an **ordered basis** of V.

Definition 13 (Coordinates)
Consider an $\mathbb{R}$-vector space V and an ordered basis $B = (\boldsymbol{b}_1, \dots, \boldsymbol{b}_n)$. For $\boldsymbol{x} \in V$ we obtain a unique representation (linear combination)

$$\boldsymbol{x} = \lambda_1 \boldsymbol{b}_1 + \ldots + \lambda_n \boldsymbol{b}_n \tag{2.117}$$

of $\boldsymbol{x}$ with respect to B. Then $\lambda_1, \dots, \lambda_n$ are the **coordinates** *of $\boldsymbol{x}$ with respect to B and the vector*

$$\begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix} \in \mathbb{R}^n \tag{2.118}$$

is the **coordinate vector/coordinate representation** *of $\boldsymbol{x}$ with respect to B.*

Now we are ready to make a connection between linear mappings between finite-dimensional vector spaces and matrices.

Definition 14 (Transformation matrix)
Consider $\mathbb{R}$-vector spaces V, W with corresponding (ordered) bases $= (\boldsymbol{b}_1, \dots, \boldsymbol{b}_n)$ and $C = (\boldsymbol{c}_1, \dots, \boldsymbol{c}_m)$. Moreover, we consider a linear mapping $\Phi : V \to W$. For $j \in \{1, \dots n\}$

$$\Phi(\boldsymbol{b}_j) = a_{1j}\boldsymbol{c}_1 + \dots + a_{mj}\boldsymbol{c}_m \tag{2.119}$$

is the unique representation of $\Phi(\boldsymbol{b}_j)$ with respect to $\boldsymbol{C}$. Then, we call the $m \times n$-matrix

$$\boldsymbol{A}_\Phi := ((a_{ij})) \tag{2.120}$$

the **transformation matrix** *of Φ (with respect to the bases B of V and C of W).*

Remark 21

- *The coordinates of $\Phi(\boldsymbol{b}_j)$ are the j-th column of $\boldsymbol{A}_\Phi$.*
- $\mathrm{rk}(\boldsymbol{A}_\Phi) = \dim(\mathrm{Im}(\Phi))$
- *Consider (finite-dimensional) $\mathbb{R}$-vector spaces V, W with ordered bases B, C, $\Phi : V \to W$ linear and transformation matrix $\boldsymbol{A}_\Phi$. If $\hat{\boldsymbol{x}}$ is the coordinate vector of $\boldsymbol{x} \in V$ and $\hat{\boldsymbol{y}}$ the coordinate vector of $\boldsymbol{y} = \Phi(\boldsymbol{x}) \in W$, then*

  $$\hat{\boldsymbol{y}} = \boldsymbol{A}_\Phi \hat{\boldsymbol{x}}. \tag{2.121}$$

  This means that the transformation matrix can be used to map coordinates with respect to an ordered basis in V to coordinates with respect to an ordered basis in W.

Remark 22 (Null Space and Column Space)

- *The null space describes all linear combinations of columns to get* $\boldsymbol{0}$.
- *The image (column space) of a transformation matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is the span of the columns of $\boldsymbol{A}$ and, therefore, has the same dimension as the "height" of the matrix, whereas the kernel (null space) has the same dimension as the "width" of the matrix.*

- *The purpose of the null space is to determine whether a solution of the linear equation system is unique and, if not, to capture all possible solutions.*
- *The null space focuses on the coefficients of the linear combinations of the columns, whereas the column space focuses on the values of the linear combinations of columns.*

2.9.3 Basis Change

In the following, we will have a closer look at how transformation matrices of a linear mapping $\Phi : V \to W$ change if we change the bases in V and W. Consider ordered bases

$$B = (\boldsymbol{b}_1, \dots, \boldsymbol{b}_n), \quad \tilde{B} = (\tilde{\boldsymbol{b}}_1, \dots, \tilde{\boldsymbol{b}}_n) \tag{2.122}$$

ordered bases of V and

$$C = (\boldsymbol{c}_1, \dots, \boldsymbol{c}_m), \quad \tilde{C} = (\tilde{\boldsymbol{c}}_1, \dots, \tilde{\boldsymbol{c}}_n) \tag{2.123}$$

ordered bases of W. Moreover, $\boldsymbol{A}_\Phi$ is the transformation matrix of the linear mapping $\Phi : V \to W$ with respect to the bases B and C, and $\tilde{\boldsymbol{A}}_\Phi$ is the corresponding transformation mapping with respect to $\tilde{B}$ and $\tilde{C}$. We will now investigate how $\boldsymbol{A}$ and $\tilde{\boldsymbol{A}}$ are related, i.e., how/whether we can transform $\boldsymbol{A}$ into $\tilde{\boldsymbol{A}}$ if we choose to perform a basis change from B, C to $\tilde{B}, \tilde{C}$.
We can write the vectors of the new basis $\tilde{B}$ of V as a linear combination of the basis vectors of B, such that

$$\tilde{\boldsymbol{b}}_j = s_{1j}\boldsymbol{b}_1 + \dots + s_{nj}\boldsymbol{b}_n, \quad j = 1, \dots, n. \tag{2.124}$$

Similarly, we write the new basis vectors $\tilde{C}$ of W as a linear combination of the basis vectors of C, which yields

$$\tilde{\boldsymbol{c}}_k = t_{1k}\boldsymbol{c}_1 + \cdots t_{mk}\boldsymbol{c}_m. \tag{2.125}$$

Note that both $\boldsymbol{S} = ((s_{ij})) \in \mathbb{R}^{n \times n}$ and $\boldsymbol{T} = ((t_{ij})) \in \mathbb{R}^{m \times m}$ are regular.
For all $j = 1, \dots, n$, we get

$$\Phi(\tilde{\boldsymbol{b}}_j) = \sum_{k=1}^{m} \underbrace{\tilde{a}_{kj}\tilde{\boldsymbol{c}}_k}_{\in W} = \sum_{k=1}^{m} \tilde{a}_{kj} \sum_{i=1}^{m} t_{ik}\boldsymbol{c}_i = \sum_{i=1}^{m} \left(\sum_{k=1}^{m} t_{ik}\tilde{a}_{kj} \right) \boldsymbol{c}_i \tag{2.126}$$

where we expressed the new basis vectors $\tilde{\boldsymbol{c}}_k \in W$ as linear combinations of the basis vectors $\boldsymbol{c}_i \in W$. When we express the $\tilde{\boldsymbol{b}}_k \in V$ as linear combinations of $\boldsymbol{b}_i \in V$, we arrive at

$$\Phi(\tilde{\boldsymbol{b}}_j) = \Phi\left(\sum_{k=1}^{n} s_{kj}\boldsymbol{b}_k \right) = \sum_{k=1}^{n} s_{kj}\Phi(\boldsymbol{b}_k) = \sum_{k=1}^{n} s_{kj} \sum_{i=1}^{m} a_{ik}\boldsymbol{c}_i = \sum_{i=1}^{m} \left(\sum_{k=1}^{n} a_{ik}s_{kj} \right) \boldsymbol{c}_i \tag{2.127}$$

Comparing (2.126) and (2.127), it follows for all $j = 1,\dots,n$ and $i = 1,\dots,m$ that

$$\sum_{k=1}^{m} t_{ik}\tilde{a}_{kj} = \sum_{k=1}^{n} a_{ik}s_{kj} \tag{2.128}$$

and, therefore,

$$\boldsymbol{T}\tilde{\boldsymbol{A}} = \boldsymbol{A}\boldsymbol{S}, \tag{2.129}$$

such that

$$\tilde{\boldsymbol{A}} = \boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{S}. \tag{2.130}$$

Hence, with a basis change in V (B is replaced with $\tilde{B}$) and W (C is replaced with $\tilde{C}$) the transformation matrix $\boldsymbol{A}_\Phi$ of a linear mapping $\Phi : V \to W$ is expressed by an equivalent matrix $\tilde{\boldsymbol{A}}_\Phi$ with

$$\tilde{\boldsymbol{A}}_\Phi = \boldsymbol{T}^{-1}\boldsymbol{A}_\Phi\boldsymbol{S}. \tag{2.131}$$

Definition 15 (Equivalence)
Two matrices $\boldsymbol{A}, \tilde{\boldsymbol{A}}$ are **equivalent** *if there exist regular matrices $\boldsymbol{S} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{T} \in \mathbb{R}^{m\times m}$, such that $\tilde{\boldsymbol{A}} = \boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{S}$.*

Definition 16 (Similarity)
Two matrices $\boldsymbol{A}, \tilde{\boldsymbol{A}}$ are **similar** *if there exists a regular matrix $\boldsymbol{S} \in \mathbb{R}^{n\times n}$ with $\tilde{\boldsymbol{A}} = \boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}$*

Remark 23
Similar matrices are always equivalent. However, equivalent matrices are not necessarily similar.

Example

Consider a linear mapping $\Phi : \mathbb{R}^3 \to \mathbb{R}^4$ whose transformation matrix is

$$\boldsymbol{A}_\Phi = \begin{bmatrix} 1 & 2 & 0 \\ -1 & 1 & 3 \\ 3 & 7 & 1 \\ -1 & 2 & 4 \end{bmatrix} \tag{2.132}$$

with respect to the standard bases

$$B = (\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}), \quad C = (\begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\1\end{bmatrix}). \tag{2.133}$$

We now want to perform a basis change toward the new bases

$$\tilde{B} = (\begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}0\\1\\1\end{bmatrix}, \begin{bmatrix}1\\0\\1\end{bmatrix},) \in \mathbb{R}^3, \quad \tilde{C} = (\begin{bmatrix}1\\1\\0\\0\end{bmatrix}, \begin{bmatrix}1\\0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\1\\1\\0\end{bmatrix}, \begin{bmatrix}1\\0\\0\\1\end{bmatrix}). \tag{2.134}$$

Then,

$$S = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \quad T = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.135}$$

and, therefore,

$$\tilde{A}_\Phi = T^{-1} A_\Phi S = \begin{bmatrix} 1/2 & 1/2 & -1/2 & -1/2 \\ 1/2 & -1/2 & 1/2 & -1/2 \\ -1/2 & 1/2 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 2 & 1 \\ 0 & 4 & 2 \\ 10 & 8 & 4 \\ 1 & 6 & 3 \end{bmatrix} = \begin{bmatrix} -4 & -4 & -2 \\ 6 & 0 & 0 \\ 4 & 8 & 4 \\ 1 & 6 & 3 \end{bmatrix}. \tag{2.136}$$

Remark 24
Consider $\mathbb{R}$-vector spaces V, W, X. From Remark 20 we already know that for linear mappings $\Phi : V \to W$ and $\Psi : W \to X$ the mapping $\Psi \circ \Phi : V \to X$ is also linear. With transformation matrices A_Φ and A_Ψ of the corresponding mappings, the overall transformation matrix $A_{\Psi\circ\Phi}$ is given by $A_{\Psi\circ\Phi} = A_\Psi A_\Phi$.

In light of this remark, we can look at basis changes from the perspective of concatenating linear mappings:

- $\tilde{A}_\Phi$ implements a linear mapping $\Phi : V \to W$ with respect to the bases $\tilde{B}, \tilde{C}$.
- S is the transformation matrix of a linear mapping $V \to V$ (automorphism) that represents $\tilde{B}$ in terms of B.
- T is the transformation matrix of a linear mapping $W \to W$ (automorphism) that represents $\tilde{C}$ in terms of C.

If we (informally) write down the transformations just in terms of bases then

- $A_\Phi : B \to C$
- $\tilde{A}_\Phi : \tilde{B} \to \tilde{C}$
- $S : \tilde{B} \to B$
- $T : \tilde{C} \to C$ and $T^{-1} : C \to \tilde{C}$

and

$$\tilde{B} \to \tilde{C} = \tilde{B} \to B \to C \to \tilde{C} \tag{2.137}$$

$$\tilde{A}_\Phi = T^{-1} A_\Phi S\,. \tag{2.138}$$

Note that the execution order in (2.138) is from right to left because vectors are multiplied at the right-hand side.

2.10 Determinants

Determinants are important concepts in linear algebra. For instance, they indicate whether a matrix can be inverted or we can use them to check for linear independence. A geometric intuition is that the absolute value of the determinant of real vectors is equal to the volume of the parallelepiped spanned by those vectors. Determinants will play a very important role for determining eigenvalues and eigenvectors (Section 2.11).

Determinants are only defined for square matrices $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, and we write $\det(\boldsymbol{A})$ or $|\boldsymbol{A}|$.

Remark 25

- *For* $n = 1$, $\det(\boldsymbol{A}) = \det(a_{11}) = a_{11}$
- *For* $n = 2$,

$$\det(\boldsymbol{A}) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21} \tag{2.139}$$

- *For* $n = 3$ *(Sarrus rule):*

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23} \tag{2.140}$$

$$- a_{31}a_{22}a_{13} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33}$$

- *For an upper/lower triangular matrix* $\boldsymbol{A}$, *the determinant is the product of the diagonal elements:* $\det(\boldsymbol{A}) = \prod_{i=1}^{n} a_{ii}$

Remark 26 (Properties of Determinants)

- $\det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det(\boldsymbol{B})$
- $\det(\boldsymbol{A}) = 0 \Leftrightarrow \boldsymbol{A}$ *is singular (not invertible)*
- *Alternatively:* $\boldsymbol{A}$ *is regular* $\Leftrightarrow \det(\boldsymbol{A}) \neq 0$.
- $\det(\boldsymbol{A}) = \det(\boldsymbol{A}^\top)$
- *If* $\boldsymbol{A}$ *is regular then* $\det(\boldsymbol{A}^{-1}) = 1/\det(\boldsymbol{A})$
- *Similar matrices possess the same determinant. Therefore, for a linear mapping* $\phi : V \to V$ *all transformation matrices* $\boldsymbol{A}_\phi$ *of* ϕ *have the same determinant.*

Theorem 5

For $\boldsymbol{A} \in \mathbb{R}^{n\times n}$:

1. *Adding a multiple of a column/row to another one does not change* $\det(\boldsymbol{A})$.
2. *Multiplication of a column/row with* $\lambda \in \mathbb{R}$ *scales* $\det(\boldsymbol{A})$ *by* λ. *In particular,* $\det(\lambda\boldsymbol{A}) = \lambda^n \det(\boldsymbol{A})$.
3. *Swapping two columns/rows changes the sign of* $\det(\boldsymbol{A})$.

Because of this theorem, we can use Gaussian elimination to compute $\det(\boldsymbol{A})$. However, we need to pay attention to swapping the sign when swapping rows.

Example

$$\begin{vmatrix} 2 & 0 & 1 & 2 & 0 \\ 2 & -1 & 0 & 1 & 1 \\ 0 & 1 & 2 & 1 & 2 \\ -2 & 0 & 2 & -1 & 2 \\ 2 & 0 & 0 & 1 & 1 \end{vmatrix} = \begin{vmatrix} 2 & 0 & 1 & 2 & 0 \\ 0 & -1 & -1 & -1 & 1 \\ 0 & 1 & 2 & 1 & 2 \\ 0 & 0 & 3 & 1 & 2 \\ 0 & 0 & -1 & -1 & 1 \end{vmatrix} = \begin{vmatrix} 2 & 0 & 1 & 2 & 0 \\ 0 & -1 & -1 & -1 & 1 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 3 & 1 & 2 \\ 0 & 0 & -1 & -1 & 1 \end{vmatrix} \tag{2.141}$$

$$= \begin{vmatrix} 2 & 0 & 1 & 2 & 0 \\ 0 & -1 & -1 & -1 & 1 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & -7 \\ 0 & 0 & 0 & -1 & 4 \end{vmatrix} = \begin{vmatrix} 2 & 0 & 1 & 2 & 0 \\ 0 & -1 & -1 & -1 & 1 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & -7 \\ 0 & 0 & 0 & 0 & -3 \end{vmatrix} = 6 \tag{2.142}$$

We first used Gaussian elimination to bring $\boldsymbol{A}$ into triangular form, and then exploited the fact that the determinant of a triangular matrix is the product of its diagonal elements.

Theorem 6

Consider a matrix $\boldsymbol{A} = ((a_{ij})) \in \mathbb{R}^{n\times n}$. We define $\boldsymbol{A}_{i,j}$ to be the matrix that remains if we delete the ith row and the jth column from $\boldsymbol{A}$. Then, for $j = 1,\dots,n$:

1. $\det(\boldsymbol{A}) = \sum_{k=1}^{n} (-1)^{k+j} a_{kj} \det(\boldsymbol{A}_{k,j})$
2. $\det(\boldsymbol{A}) = \sum_{k=1}^{n} (-1)^{k+j} a_{jk} \det(\boldsymbol{A}_{j,k})$

Example

Let us re-compute the above example:

$$\begin{vmatrix} 2 & 0 & 1 & 2 & 0 \\ 2 & -1 & 0 & 1 & 1 \\ 0 & 1 & 2 & 1 & 2 \\ -2 & 0 & 2 & -1 & 2 \\ 2 & 0 & 0 & 1 & 1 \end{vmatrix} = \begin{vmatrix} 2 & 0 & 1 & 2 & 0 \\ 0 & -1 & -1 & -1 & 1 \\ 0 & 1 & 2 & 1 & 2 \\ 0 & 0 & 3 & 1 & 2 \\ 0 & 0 & -1 & -1 & 1 \end{vmatrix} \overset{\text{1st col.}}{=} (-1)^{1+1} 2 \cdot \begin{vmatrix} -1 & -1 & -1 & 1 \\ 1 & 2 & 1 & 2 \\ 0 & 3 & 1 & 2 \\ 0 & -1 & -1 & 1 \end{vmatrix} \tag{2.143}$$

If we now subtract the fourth row from the first row and multiply (-2) times the third column to the fourth column we obtain

$$2\begin{vmatrix} -1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 3 & 1 & 0 \\ 0 & -1 & -1 & 3 \end{vmatrix} \overset{\text{1st row}}{=} -2 \begin{vmatrix} 2 & 1 & 0 \\ 3 & 1 & 0 \\ -1 & -1 & 3 \end{vmatrix} \overset{\text{3rd col.}}{=} (-2)\cdot 3 \cdot \begin{vmatrix} 2 & 1 \\ 3 & 1 \end{vmatrix} = 6 \tag{2.144}$$

2.11 Eigenvalues

Definition 17 (Eigenvalue, eigenvector)

For an $\mathbb{R}$-vector space V and a linear map $\Phi : V \to V$, the scalar $\lambda \in \mathbb{R}$ is called **eigenvalue** *if there exists a $\boldsymbol{x} \in V, \boldsymbol{x} \neq \boldsymbol{0}$ with*

$$\Phi(\boldsymbol{x}) = \lambda \boldsymbol{x}. \tag{2.145}$$

The corresponding vector $\boldsymbol{x}$ is called **eigenvector** *of Φ associated with eigenvalue λ.*

Definition 18 (Eigenspace and Spectrum)

- *The set of all eigenvectors of Φ associated with an eigenvalue λ forms (together with $\mathbf{0}$) a subspace of V, which is called* **eigenspace** *of Φ and denoted by E_λ.*

- *The set of all eigenvalues of Φ is called* **spectrum** *of Φ.*

Remark 27

- *Apparently, $E_\lambda = \mathrm{Ker}(\Phi - \lambda \mathrm{id}_V)$ since*

$$\Phi(\boldsymbol{x}) = \lambda\boldsymbol{x} \Leftrightarrow \Phi(\boldsymbol{x}) - \lambda\boldsymbol{x} = \mathbf{0} \Leftrightarrow (\Phi - \lambda\mathrm{id}_V)\boldsymbol{x} = \mathbf{0} \Leftrightarrow \boldsymbol{x} \in \mathrm{Ker}(\Phi - \lambda\mathrm{id}_V). \quad (2.146)$$

- *A matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ uniquely determines the linear mapping $\Phi : \mathbb{R}^n \to \mathbb{R}^n, \boldsymbol{x} \mapsto \boldsymbol{A}\boldsymbol{x}$. Therefore, we can also talk of eigenvalues, eigenvectors and eigenspaces of square matrices.*

- *Similar matrices possess the same eigenvalues*

- *If $\boldsymbol{x}$ is an eigenvector of Φ with eigenvalue λ, then $\alpha\boldsymbol{x}$, $\alpha \in \mathbb{R}$ is an eigenvector with the same eigenvalue. Therefore, there exist an infinite number of eigenvectors for every eigenvalue λ, i.e., the eigenvectors are not unique.*

Theorem 7

Consider an $\mathbb{R}$-vector space V and a linear map $\Phi : V \to V$ with pairwise different eigenvalues $\lambda_1, \dots, \lambda_k$ and corresponding eigenvectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k$. Then the vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k$ are linearly independent.

An endomorphism $\Phi : V \to V$, $V \subset \mathbb{R}^n$ (and equivalently the corresponding transformation matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$) possesses at most n different eigenvalues.
The following statements are equivalent:

- λ is eigenvalue of $\boldsymbol{A} \in \mathbb{R}^{n\times n}$

- There exists a $\boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{x} \neq \mathbf{0}$ with $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$ or, equivalently, $(\boldsymbol{A} - \lambda\boldsymbol{I}_n)\boldsymbol{x} = \mathbf{0}$

- $(\boldsymbol{A} - \lambda\boldsymbol{I}_n)\boldsymbol{x} = \mathbf{0}$ can be solved non-trivially, i.e., $\boldsymbol{x} \neq \mathbf{0}$.

- $\mathrm{rk}(\boldsymbol{A} - \lambda\boldsymbol{I}_n) < n$

- $\det(\boldsymbol{A} - \lambda\boldsymbol{I}_n) = 0$

- $\boldsymbol{A}$ and $\boldsymbol{A}^\top$ possess the same eigenvalues, but not the same eigenvectors.

2.11.1 Geometric Interpretation

Geometrically, an eigenvector corresponding to a real, nonzero eigenvalue points in a direction that is stretched, and the eigenvalue is the factor by which it is stretched. If the eigenvalue is negative, the direction is reversed. In particular, the eigenvector does not change its direction under Φ.

2.11.2 Characteristic Polynomial

In the following, we will discuss how to determine the eigenspaces of an endomorphism Φ.[20] For this, we need to introduce the characteristic polynomial first.

Definition 19 (Characteristic Polynomial)
For $\lambda \in \mathbb{R}$ and an endomorphism Φ on $\mathbb{R}^n$

$$p = \det(\boldsymbol{A} - \lambda \boldsymbol{I}) = a_0 + a_1\lambda + a_2\lambda^2 + \cdots + a_{n-1}\lambda^{n-1} + (-1)^n\lambda^n, \quad a_0, \ldots, a_{n-1} \in \mathbb{R}, \tag{2.147}$$

is the **characteristic polynomial** *of $\boldsymbol{A}$. In particular,*

$$a_0 = \det(\boldsymbol{A}), \tag{2.148}$$

$$a_{n-1} = (-1)^{n-1} tr(\boldsymbol{A}), \tag{2.149}$$

where $tr(\boldsymbol{A}) = \sum_{i=1}^n a_{ii}$ is the **trace** *of $\boldsymbol{A}$ and defined as the sum of the diagonal elements of $\boldsymbol{A}$.*

Theorem 8
$\lambda \in \mathbb{R}$ is eigenvalue of $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ if and only if λ is a root of the characteristic polynomial p of $\boldsymbol{A}$.

Remark 28

1. *If λ is an eigenvalue of $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ then the corresponding eigenspace E_λ is the solution space of the homogeneous linear equation system $(\boldsymbol{A} - \lambda \boldsymbol{I}_n)\boldsymbol{x} = \boldsymbol{0}$.*
2. *Similar matrices possess the same characteristic polynomial.*

2.11.3 Example: Eigenspace Computation

- $\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$

  1. Characteristic polynomial: $p = |\boldsymbol{A} - \lambda \boldsymbol{I}_2| = \begin{vmatrix} 1-\lambda & 0 \\ 1 & 1-\lambda \end{vmatrix} = (1-\lambda)^2$. Therefore $\lambda = 1$ is the only root of p and, therefore, the only eigenvalue of $\boldsymbol{A}$
  2. To compute the eigenspace for the eigenvalue $\lambda = 1$, we need to compute the null space of $\boldsymbol{A} - \boldsymbol{I}$:

$$\boldsymbol{A} - 1 \cdot \boldsymbol{I} = \boldsymbol{0} \Leftrightarrow \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \boldsymbol{0} \tag{2.150}$$

$$\Rightarrow E_1 = [\begin{bmatrix} 0 \\ 1 \end{bmatrix}] \tag{2.151}$$

[20] It turns out that it is sufficient to work directly with the corresponding transformation mappings $\boldsymbol{A}_\Phi \in \mathbb{R}^{n\times n}$.

- $\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$
  1. Characteristic polynomial: $p = \det(\boldsymbol{A} - \lambda \boldsymbol{I}) = \lambda^2 + 1$. For $\lambda \in \mathbb{R}$ there exist no eigenvalue of $\boldsymbol{A}$. However, for $\lambda \in \mathbb{C}$ we find $\lambda_1 = i$, $\lambda_2 = -i$.
  2. The corresponding eigenspaces (for $\lambda_i \in \mathbb{C}$) are

$$E_i = [\begin{bmatrix} 1 \\ i \end{bmatrix}], \quad E_{-i} = [\begin{bmatrix} 1 \\ -i \end{bmatrix}]. \tag{2.152}$$

- $\boldsymbol{A} = \begin{bmatrix} 0 & -1 & 1 & 1 \\ -1 & 1 & -2 & 3 \\ 2 & -1 & 0 & 0 \\ 1 & -1 & 1 & 0 \end{bmatrix}$
  1. Characteristic polynomial:

$$p = \begin{vmatrix} -\lambda & -1 & 1 & 1 \\ -1 & 1-\lambda & -2 & 3 \\ 2 & -1 & -\lambda & 0 \\ 1 & -1 & 1 & -\lambda \end{vmatrix} = \begin{vmatrix} -\lambda & -1 & 1 & 1 \\ 0 & -\lambda & -1 & 3-\lambda \\ 0 & 1 & -2-\lambda & 2\lambda \\ 1 & -1 & 1 & -\lambda \end{vmatrix} \tag{2.153}$$

$$= \begin{vmatrix} -\lambda & -1-\lambda & 0 & 1 \\ 0 & -\lambda & -1-\lambda & 3-\lambda \\ 0 & 1 & -1-\lambda & 2\lambda \\ 1 & 0 & 0 & -\lambda \end{vmatrix} \tag{2.154}$$

$$= (-\lambda)^2 \begin{vmatrix} -\lambda & -1-\lambda \\ 1 & -1-\lambda \end{vmatrix} - \begin{vmatrix} -1-\lambda & 0 & 1 \\ -\lambda & -1-\lambda & 3-\lambda \\ 1 & -1-\lambda & 2\lambda \end{vmatrix} \tag{2.155}$$

$$= (1+\lambda)^2(\lambda^2 - 3\lambda + 2) = (1+\lambda)^2(1-\lambda)(2-\lambda) \tag{2.156}$$

  Therefore, the eigenvalues of $\boldsymbol{A}$ are $\lambda_1 = -1, \lambda_2 = 1, \lambda_3 = 2$.
  2. The corresponding eigenspaces are the solutions of $(\boldsymbol{A} - \lambda_i \boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}, i = 1, 2, 3$, and given by

$$E_{\lambda_1} = [\begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}], \quad E_{\lambda_2} = [\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}], \quad E_{\lambda_3} = [\begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}]. \tag{2.157}$$

2.11.4 Applications

- Eigenvalues were used by Claude Shannon to determine the theoretical limit to how much information can be transmitted through a communication medium like your telephone line or through the air. This is done by calculating the eigenvectors and eigenvalues of the communication channel (expressed a matrix), and then waterfilling on the eigenvalues. The eigenvalues are then, in essence, the gains of the fundamental modes of the channel, which themselves are captured by the eigenvectors.

- Google uses the eigenvector corresponding to the maximal eigenvalue of the Google matrix to determine the rank of a page for search. The idea that the PageRank algorithm[21] brought up was that the importance of any web page can be judged by looking at the pages that link to it. For this, we write down all websites as a huge (weighted) directed graph that shows which page links to which. Then, the navigation behavior of a user can be described by a transition matrix $\boldsymbol{A}$ of this graph that tells us with what (click) probability somebody will end up on a different website. The matrix $\boldsymbol{A}$ has the property that for any initial rank/importance vector $\boldsymbol{x}$ of a website the sequence $\boldsymbol{x}, \boldsymbol{A}\boldsymbol{x}, \boldsymbol{A}^2\boldsymbol{x}, \ldots$ converges to a vector $\boldsymbol{x}^*$. This vector is called the **PageRank** and satisfies $\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{x}^*$, i.e., it is an eigenvector (with corresponding eigenvalue 1).[22]

- Eigenvectors are fundamental to principal components analysis (PCA, Hotelling (1936)), which is commonly used for dimensionality reduction in face recognition and other machine learning applications.

- Geometry and computer graphics

2.12 Diagonalization

Diagonal matrices are possess a very simple structure and they allow for very fast computation of determinants and inverses, for instance. In this section, we will have a closer look at endomorphisms of finite-dimensional vector spaces, which are similar to a diagonal matrix, i.e., endomorphisms whose transformation matrix attains diagonal structure for a suitable basis.
For this purpose, we will exploit learned concepts about basis change (Chapter 2.9.3) and eigenvalues (Chapter 2.11).

Definition 20 (Diagonal Form)
A matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is **diagonalizable** *if it is similar to a diagonal matrix*

$$\begin{bmatrix} c_1 & 0 & \cdots & \cdots & 0 \\ 0 & c_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ 0 & \cdots & 0 & c_{n-1} & 0 \\ 0 & \cdots & \cdots & 0 & c_n \end{bmatrix} \tag{2.158}$$

.

Theorem 9
For an endomorphism Φ of an n-dimensional $\mathbb{R}$-vector space V the following statements are equivalent:

1. *Φ is diagonalizable.*

2. *The transformation matrix $\boldsymbol{A}_\Phi$ is diagonalizable.*

[21]Developed at Stanford University by Larry Page and Sergey Brin in 1996
[22]When normalizing $\boldsymbol{x}^*$, such that $\|\boldsymbol{x}^*\| = 1$ we can interpret the entries as probabilities.

3. *There exists a basis in V consisting of eigenvectors of Φ.*

4. *The sum of the dimensions of the eigenspaces of Φ is n.*[23]

Theorem 10
For an n-dimensional $\mathbb{R}$-vector space V and a linear mapping $\Phi : V \to V$ the following holds: Φ is diagonalizable if and only if

1. *Its characteristic polynomial p is given in the form*

$$p = (-1)^n(\lambda - c_1)^{r_1} \cdots (\lambda - c_k)^{r_k} \tag{2.159}$$

   with $r_i \in \mathbb{N}$ and pairwise different roots $c_i \in \mathbb{R}$ and

2. *For $i = 1, \ldots, k$*

$$\dim(\mathrm{Im}(\Phi - c_i \mathrm{id}_V)) = n - r_i \tag{2.160}$$

In (2.159) we say that the characteristic polynomial decomposes into **linear factors**. The second requirement in (2.160) says that the dimension of the eigenspace E_{c_i} must correspond to the multiplicity r_i of the eigenvalues in the characteristic polynomial, $i = 1, \ldots, k$. The dimension of the eigenspace E_{c_i} is the dimension of the kernel/null space of $\Phi - c_i \mathrm{id}_V$.
Theorem 10 holds equivalently if we replace Φ with $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ and id_V with $\boldsymbol{I}_n$.
If Φ is diagonalizable it possesses a transformation matrix of the form

$$\boldsymbol{A}_\Phi = \begin{bmatrix} c_1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & & & \vdots \\ \vdots & \ddots & c_1 & \ddots & & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & c_k & \ddots & \vdots \\ \vdots & & & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & c_k \end{bmatrix} \tag{2.161}$$

where each eigenvalue c_i appears r_i times (its multiplicity in the characteristic polynomial and the dimension of the corresponding eigenspace) on the diagonal.

2.12.1 Examples

- $\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$.

  1. Characteristic polynomial: $p = (1 - \lambda)^2$

[23] In particular, an endomorphism Φ of an n-dimensional $\mathbb{R}$-vector space with n different eigenvalues is diagonalizable.

2. Dimension of eigenspace: $\text{rk}(\boldsymbol{A}-\boldsymbol{I}) = 1 \neq 2$. Because of Theorem 10 $\boldsymbol{A}$ is not diagonalizable.

- $\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$.
- Characteristic polynomial: $p = 1 + \lambda^2$.
  - For $\lambda \in \mathbb{R}$ there exist no roots of p and $\boldsymbol{A}$ is not diagonalizable.
  - For $\lambda \in \mathbb{C}$, $p = (i-\lambda)(-i-\lambda)$ and $\boldsymbol{A}$ has two eigenvalues and is therefore diagonalizable.
- $\boldsymbol{A} = \begin{bmatrix} 0 & -1 & 1 & 1 \\ -1 & 1 & -2 & 3 \\ 2 & -1 & 0 & 0 \\ 1 & -1 & 1 & 0 \end{bmatrix}$.
  1. Characteristic polynomial: $p = (1+\lambda)^2(1-\lambda)(2-\lambda)$. The eigenvalues are $c_1 = -1, c_2 = 1, c_3 = 2$ with multiplicities $r_1 = 2, r_2 = 1, r_3 = 1$, respectively.
  2. Dimension of eigenspaces: $\dim(E_{c_1}) = 1 \neq r_1$.

  Therefore, $\boldsymbol{A}$ cannot be diagonalized.
- $\boldsymbol{A} = \begin{bmatrix} 3 & 2 & -1 \\ 2 & 6 & -2 \\ 0 & 0 & 2 \end{bmatrix}$.
  1. Characteristic polynomial: $p = (2-\lambda)^2(7-\lambda)$. Therefore, $c_1 = 2, c_2 = 7, r_1 = 2, r_2 = 1$
  2. Dimension of eigenspaces: $\text{rk}(\boldsymbol{A}-c_1\boldsymbol{I}_3) = 1 = n-r_1$, $\text{rk}(\boldsymbol{A}-c_2\boldsymbol{I}_3) = 2 = n-r_2$

  Therefore, $\boldsymbol{A}$ is diagonalizable.

Let us now discuss a concrete way of constructing diagonal matrices.

Remark 29

If $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is diagonalizable and $(\boldsymbol{b}_1,\dots,\boldsymbol{b}_n)$ is a basis of eigenvectors of $\boldsymbol{A}$ with $\boldsymbol{A}\boldsymbol{b}_i = c_i\boldsymbol{b}_i$, $i = 1,\dots,n$ then it holds that for the regular matrix $\boldsymbol{S} = (\boldsymbol{b}_1|\dots|\boldsymbol{b}_n)$

$$\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S} = \begin{bmatrix} c_1 & 0 & \cdots & \cdots & 0 \\ 0 & c_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \cdots & 0 & c_{n-1} & 0 \\ 0 & \cdots & \cdots & 0 & c_n \end{bmatrix} \tag{2.162}$$

The diagonal matrix in (2.162) is the transformation matrix of $\boldsymbol{x} \mapsto \boldsymbol{A}\boldsymbol{x}$ with respect to the eigenbasis $(\boldsymbol{b}_1,\dots,\boldsymbol{b}_n)$.

Coming back to the above example, where we wanted to determine the diagonal form of $A = \begin{bmatrix} 3 & 2 & -1 \\ 2 & 6 & -2 \\ 0 & 0 & 2 \end{bmatrix}$. We already know that A is diagonalizable. We now determine the eigenbasis of $\mathbb{R}^3$ that allows us to transform A into a similar matrix in diagonal form via $S^{-1}AS$:

1. The eigenspaces are

$$E_{c_1} = [\ \underbrace{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}}_{=:b_1}, \underbrace{\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}}_{=:b_2}], \quad E_{c_2} = [\ \underbrace{\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}}_{=:b_3}\ ] \tag{2.163}$$

2. We now collect the eigenvectors in a matrix and obtain

$$S = (b_1|b_2|b_3) = \begin{bmatrix} 1 & -2 & 1 \\ 0 & 1 & 2 \\ 1 & 0 & 0 \end{bmatrix} \tag{2.164}$$

such that

$$S^{-1}AS = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 7 \end{bmatrix}. \tag{2.165}$$

Remark 30

So far, we computed diagonal matrices as $D = S^{-1}AS$. However, we can equally write $A = SDS^{-1}$. Here, we can interpret the transformation matrix A as follows: S performs a basis change from the eigenbasis into the standard basis, D then scales the vector along the axes of the standard basis, and S^{-1} transforms the scaled vectors back into the eigenbasis coordinates. In Section **??***, we will discover that S represents a rotation.*

2.12.2 Applications

Diagonal matrices $D = S^{-1}AS$ exhibit the nice properties that they can be easily raised to a power.

$$A^k = (S^{-1}DS)^k = S^{-1}D^kS \tag{2.166}$$

Computing D^k is easy because we apply this operation individually to any diagonal element. As an example, this allows to compute inverses of D in $\mathcal{O}(n)$ instead of $\mathcal{O}(n^3)$.

2.12.3 Cayley-Hamilton Theorem*

Theorem 11 (Cayley-Hamilton)
Let V be an n-dimensional $\mathbb{R}$-vector space and $\Phi : V \to V$ an endomorphism with transformation matrix $\boldsymbol{A}_\Phi$ and characteristic polynomial p. Then,

$$p(\Phi) = 0 \tag{2.167}$$

(and equivalently, $p(\boldsymbol{A}_\Phi) = \boldsymbol{0}$).

Remark 31

- *Note that the right hand side of (2.167) is the zero mapping (or the $\boldsymbol{0}$-matrix when we use the transformation matrix $\boldsymbol{A}_\Phi$).*
- *The importance of the Cayley-Hamilton theorem is not the existence of a (non-trivial) polynomial q, such that $q(\Phi) = 0$, but that the characteristic polynomial has this property.*

Applications

- Find an expression for $\boldsymbol{A}^{-1}$ in terms of $\boldsymbol{I}, \boldsymbol{A}, \boldsymbol{A}^2, \ldots, \boldsymbol{A}^{n-1}$. Example: $\boldsymbol{A} = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix}$ has the characteristic polynomial $p(\lambda) = \lambda^2 - 2\lambda + 3$. Then Theorem 11 states that $\boldsymbol{A}^2 - 2\boldsymbol{A} + 3\boldsymbol{I} = \boldsymbol{0}$ and, therefore, $-\boldsymbol{A}^2 + 2\boldsymbol{A} = 3\boldsymbol{I} \Leftrightarrow \boldsymbol{A}^{-1} = \frac{1}{3}(2\boldsymbol{I} - \boldsymbol{A})$
- Find an expression of $\boldsymbol{A}^m$, $m \geq n$, in terms of $\boldsymbol{I}, \boldsymbol{A}, \boldsymbol{A}^2, \ldots, \boldsymbol{A}^{n-1}$

2.13 Scalar Products

Definition 21
Let $\beta : V \times V \to \mathbb{R}$ be a bilinear mapping (i.e., linear in both arguments).

- *β is called* **symmetric** *if $\beta(\boldsymbol{x}, \boldsymbol{y}) = \beta(\boldsymbol{y}, \boldsymbol{x})$ for all $\boldsymbol{x}, \boldsymbol{y} \in V$.*
- *β is called* **positive definite** *if for all $\boldsymbol{x} \neq \boldsymbol{0}$: $\beta(\boldsymbol{x}, \boldsymbol{x}) > 0$. $\beta(\boldsymbol{0}, \boldsymbol{0}) = 0$.*
- *A positive definite, symmetric bilinear mapping $\beta : V \times V \to \mathbb{R}$ is called* **scalar product/dot product/inner product** *on V. We typically write $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ instead of $\beta(\boldsymbol{x}, \boldsymbol{y})$.*
- *The pair $(V, \langle \cdot, \cdot \rangle)$ is called* **Euclidean vector space** *or (real)* **vector space with scalar product**.

2.13.1 Examples

- For $V = \mathbb{R}^n, \beta(\boldsymbol{x}, \boldsymbol{y}) = \langle \boldsymbol{x}, \boldsymbol{y} \rangle := \boldsymbol{x}^\top \boldsymbol{y} = \sum_{i=1}^n x_i y_i$ is called **standard scalar product**.

- $V = \mathbb{R}^2$. If we define $\beta(\boldsymbol{x},\boldsymbol{y}) := x_1y_1 - (x_1y_2 + x_2y_1) + 2x_2y_2$ then β is a scalar product but different from the standard scalar product $\langle\cdot,\cdot\rangle$.

In a Euclidean vector space, the scalar product allows us to introduce concepts, such as lengths, distances and orthogonality.

2.13.2 Lengths, Distances, Orthogonality

Definition 22 (Norm)
Consider a Euclidean vector space $(V,\langle\cdot,\cdot\rangle)$. Then $\|\boldsymbol{x}\| := \sqrt{\langle\boldsymbol{x},\boldsymbol{x}\rangle}$ is the **length** *or* **norm** *of $\boldsymbol{x} \in V$. The mapping*

$$\|\cdot\| : V \to \mathbb{R} \tag{2.168}$$
$$\boldsymbol{x} \mapsto \|\boldsymbol{x}\| \tag{2.169}$$

is called **norm**.

2.13.3 Example

In geometry, we are often interested in lengths of vectors. We can now use scalar product to compute them. For instance, if $\boldsymbol{x} = [1,2]^\top$ then its length is $\sqrt{1^2+2^2} = \sqrt{5}$

Remark 32
The norm $\|\cdot\|$ possesses the following properties:

1. $\|\boldsymbol{x}\| \geq 0$ *for all* $\boldsymbol{x} \in V$ *and* $\|\boldsymbol{x}\| = 0 \Leftrightarrow \boldsymbol{x} = \boldsymbol{0}$
2. $\|\lambda\boldsymbol{x}\| = |\lambda| \cdot \|\boldsymbol{x}\|$ *for all* $\boldsymbol{x} \in V$ *and* $\lambda \in \mathbb{R}$
3. **Minkowski inequality:** $\|\boldsymbol{x}+\boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$ *for all* $\boldsymbol{x},\boldsymbol{y} \in V$

Definition 23 (Distance and Metric)
Consider a Euclidean vector space $(V,\langle\cdot,\cdot\rangle)$. Then $d(\boldsymbol{x},\boldsymbol{y}) := \|\boldsymbol{x}-\boldsymbol{y}\|$ is called **distance** *of $\boldsymbol{x},\boldsymbol{y} \in V$. The mapping*

$$d : V \times V \to \mathbb{R} \tag{2.170}$$
$$(\boldsymbol{x},\boldsymbol{y}) \mapsto d(\boldsymbol{x},\boldsymbol{y}) \tag{2.171}$$

is called **metric**.

A metric d satisfies:

1. d is positive definite, i.e., $d(\boldsymbol{x},\boldsymbol{y}) \geq 0$ for all $\boldsymbol{x},\boldsymbol{y} \in V$ and $d(\boldsymbol{x},\boldsymbol{y}) = 0 \Leftrightarrow \boldsymbol{x} = \boldsymbol{y}$
2. d is symmetric, i.e., $d(\boldsymbol{x},\boldsymbol{y}) = d(\boldsymbol{y},\boldsymbol{x})$ for all $\boldsymbol{x},\boldsymbol{y} \in V$.
3. **Triangular inequality:** $d(\boldsymbol{x},\boldsymbol{z}) \leq d(\boldsymbol{x},\boldsymbol{y}) + d(\boldsymbol{y},\boldsymbol{z})$.

Definition 24 (Orthogonality)
$\boldsymbol{x}$ and $\boldsymbol{y}$ are **orthogonal** *if $\langle\boldsymbol{x},\boldsymbol{y}\rangle = 0$, and we write $\boldsymbol{x} \perp \boldsymbol{y}$*

Theorem 12
Let V be a Euclidean vector space $(V, \langle\cdot,\cdot\rangle)$ and $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$. Then:

1. **Cauchy-Schwarz inequality:** $|\langle \boldsymbol{x}, \boldsymbol{y}\rangle| \leq \|\boldsymbol{x}\| \|\boldsymbol{y}\|$
2. **Minkowski inequality:** $\|\boldsymbol{x}+\boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$
3. **Triangular inequality:** $d(\boldsymbol{x}, \boldsymbol{z}) \leq d(\boldsymbol{x}, \boldsymbol{y}) + d(\boldsymbol{y}, \boldsymbol{z})$
4. **Parallelogram law:** $\|\boldsymbol{x}+\boldsymbol{y}\| + \|\boldsymbol{x}-\boldsymbol{y}\| = 2\|\boldsymbol{x}\|^2 + 2\|\boldsymbol{y}\|^2$
5. $4\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \|\boldsymbol{x}+\boldsymbol{y}\|^2 - \|\boldsymbol{x}-\boldsymbol{y}\|^2$
6. $\boldsymbol{x} \perp \boldsymbol{y} \Leftrightarrow \|\boldsymbol{x}+\boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2$

The Cauchy-Schwarz inequality allows us to define angles ω in Euclidean vector spaces between two vectors $\boldsymbol{x}, \boldsymbol{y}$. Assume that $\boldsymbol{x} \neq \boldsymbol{0}, \boldsymbol{y} \neq \boldsymbol{0}$. Then

$$-1 \leq \frac{\langle \boldsymbol{x}, \boldsymbol{y}\rangle}{\|\boldsymbol{x}\| \|\boldsymbol{y}\|} \leq 1 \tag{2.172}$$

Therefore, there exists a unique $\omega \in [0, \pi]$ with

$$\cos\omega = \frac{\langle \boldsymbol{x}, \boldsymbol{y}\rangle}{\|\boldsymbol{x}\| \|\boldsymbol{y}\|} \tag{2.173}$$

The number ω is the **angle** between $\boldsymbol{x}$ and $\boldsymbol{y}$.

Bibliography

Box, G. E. P. and Muller, M. E. (1958). A Note on the Generation of Random Normal Deviates. *Annals of Mathematical Statistics*, 29(2):610–611. pages 2

Deng, L., Seltzer, M. L., Yu, D., Acero, A., Mohamed, A.-r., and Hinton, G. E. (2010). Binary Coding of Speech Spectrograms using a Deep Auto-Encoder. In *Interspeech*, pages 1692–1695. pages

Drumm, V. and Weil, W. (2001). Lineare Algebra und Analytische Geometrie. Lecture Notes, Universität Karlsruhe. pages 21

Hogben, L., editor (2013). *Handbook of Linear Algebra*. Discrete Mathematics and Its Applications. Chapman and Hall, 2nd edition. pages 21

Hotelling, H. (1936). Analysis of a Complex of Statistical Variables into Principal Components. *Journal of Educational Psychology*, 24:417–441. pages 66

Knop, R. (1969). Remark on Algorithm 334 [G5]: Normal Random Deviates. *Communications of the ACM*, 12(5):281. pages 2

Strang, G. (2003). *Introduction to Linear Algebra*. Wellesley-Cambridge Press, 3rd edition. pages 21

Wikipedia (2015). Linear Dependence. https://en.wikipedia.org/wiki/Linear_independence. pages 40