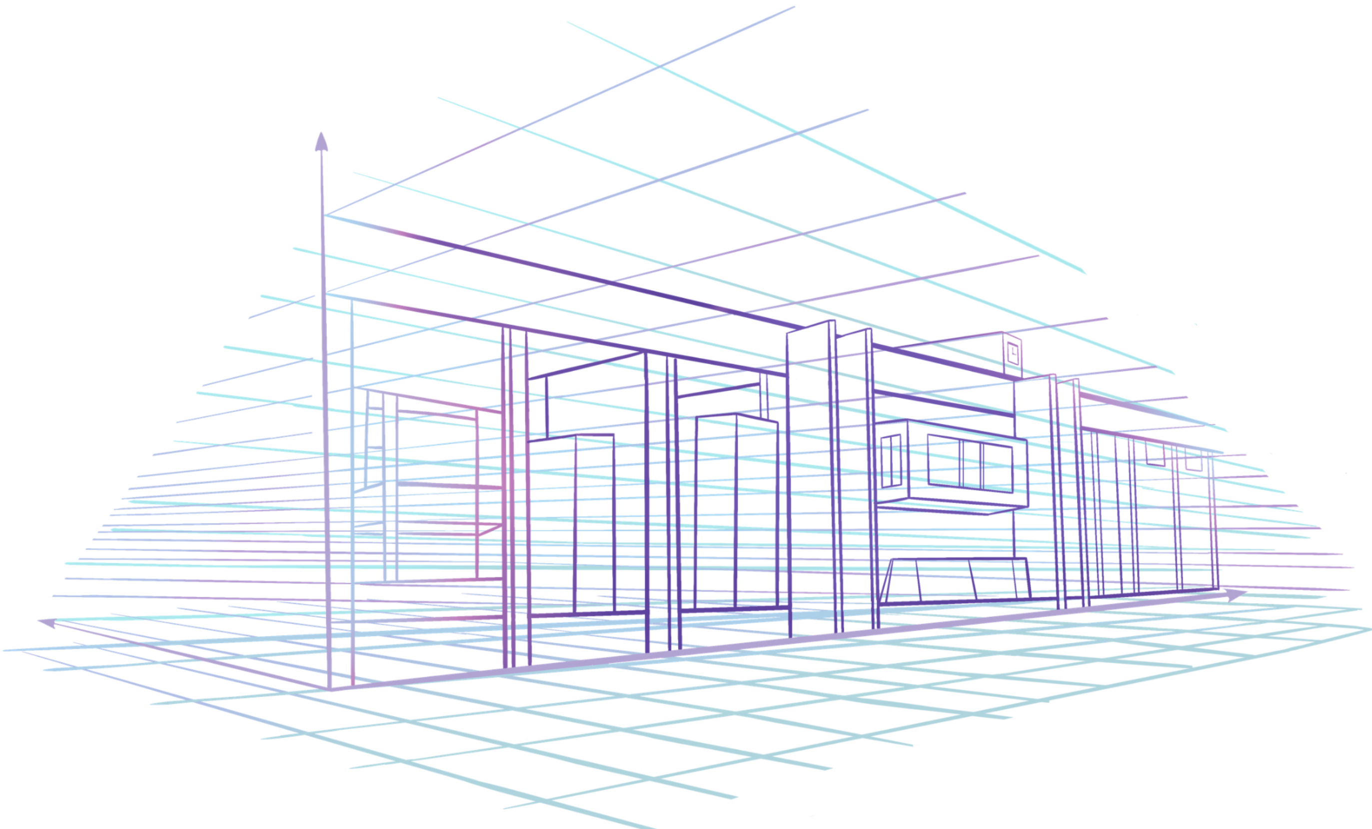
Linear Algebra

線形代数学

G30 Program, Nagoya University



Lecturer: Henrik Bachmann
Written by: Henrik Bachmann, Tue Tai Nguyen, Vic Austen
Illustrations: Hannah Patricia Halim

Version 22 (October 3, 2025)

Contents

About this course

These notes are based on the Linear Algebra I & II lectures of the G30 Program at Nagoya University given in the fall and spring semesters of 2019 - 2025. The concept of this course initially evolved from the Linear Algebra I & II given by Erik Darpö in the years before, which was based on the book [B]. The course is anticipated for two semesters (Linear Algebra I - fall semester, Linear Algebra II - spring semester) with 16 lectures of 90 minutes each. This includes a midterm and final exam each semester. In the year 2020, this course was (due to the pandemic) given online, and therefore, recordings of the lectures exist. The content and notation in the lecture videos are similar, but not exactly the same, as in these notes. However, the videos can be used for students who missed a lecture or who want to recall the content on their own. A possible schedule, together with links to the corresponding sections & videos, is given as follows:

Linear Algebra I

Week	**Content**	Section	Lecture video
01	Introduction & Linear systems	Introduction Chapter 1	Video 1
02	Matrices and vectors	Chapter 2	Video 2
03	Sets and functions	Chapter 3	Video 3
04	Linear maps	Chapter 4	Video 4
05	Linear maps in geometry	Chapter 5	Video 5
06	Matrix multiplication	Chapter 6	Video 6
07	*Midterm Exam*	Chapter 14	
08	The inverse of a linear map	Chapter 7	Video 7
09	Subspaces, Kernel & Image	Chapter 8	Video 8
10	Subspaces, Kernel & Image II	Chapter 8	Video 9
11	Linear independence & Bases I	Chapter 9 Chapter 10	Video 10
12	Bases II & Dimension	Chapter 10	part of Video 10
13	Coordinates & Orthogonal bases	Chapter 11 Chapter 12	Video 11
14	Orthogonal bases & The Gram-Schmidt algorithm	Chapter 12	Video 12
15	Orthogonal projection, Least square approximation	Chapter 13	Video 13
16	*Final Exam*	Exam LA1	

Linear Algebra II

Week	**Content**	Section	Lecture video
01	Recall Linear Algebra I & Overview	Introduction	Video 1.1
02	Vector spaces	Chapter 14	Video 1.2
03	Linear maps	Chapter 15	Video 2
04	The matrix of a linear map	Chapter 16	Video 3
05	Determinants & Mathematical induction	Chapter 17	Video 4
06	Properties of the determinant I	Chapter 17	Video 5
07	*Midterm exam*		
08	Properties of the determinant II	Chapter 17	Video 6
09	Eigenvalues and eigenvectors I	Chapter 18	Video 7
10	Eigenvalues and eigenvectors II	Chapter 18	Video 8
11	Eigenvalues and eigenvectors III (Spectral Theorem)	Chapter 18	Video 9
12	Applications	Chapter 19	Video 10
13	Continuous dynamical systems	Chapter 20	Video 11
14	Linear differential equations I	Chapter 20	Video 12
15	Linear differential equations II	Chapter 20	Video 13
16	*Final Exam*		Video 14 (Review)

A more visually pleasant overview of this course and its timeline is given by the Linear Algebra road map on the next page.

LINEAR ALGEBRA I
ようこそ!
START!
$2x_1 + x_2 = 1$
$-x_1 + 3x_2 = -2$
$\begin{pmatrix} 2 & 1 \\ -1 & 3 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$
LINEAR ALGEBRA
Road Map
$\mathbb{R}^n$
$A \cup B$
LINEAR MAPS
Week 4-6
$F: \mathbb{R}^n \to \mathbb{R}^m$
LINEAR SYSTEM MATRICES, VECTORS
SETS & MAPS
-Week- 3
Week 1-2
$A \cap B$
① $F(u+v)=F(u)+F(v)$
② $F(\lambda u)=\lambda F(u)$
$F(v)=Av.$
Week 9-10
$[F]=A \in \mathbb{R}^{m \times n}$
SUBSPACES
$U=\left\{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 \,\middle|\, x_1+x_2=0\right\}$
KERNEL
IMAGE
MIDTERM EXAM
W8. INVERSE
of linear maps
$F: \mathbb{R}^n \to \mathbb{R}^m$
$\cup$ Ker F $\cup$ im F
ORTHOGONALITY
Week 14-16
$u \cdot v = 0$
Week 13.
Winter Vacation
W11-12
Bases, linear independence
$B=\left\{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix},\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix},\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right\}$
$\sum_{i=1}^{n} \lambda_i v_i = 0 \Rightarrow \lambda_i = 0 \ \forall i$
FINAL EXAM
COORDINATES
LINEAR ALGEBRA II
VECTOR SPACES
WEEK 1-2
GOLDEN week
$F(\mathbb{R},\mathbb{R})$
$\mathcal{P}$
$\mathbb{C}$
$C^\infty(\mathbb{R},\mathbb{R})$
$\mathbb{R}^n$
$D: f \mapsto f'$
LINEAR MAPS
W2-3
$F: V \to W$
$\mathrm{im}(F) \subset W$
$\mathrm{Ker}(F) \subset V$
Spring break
MIDTERM
WEEK 4-6
determinants
$\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$
$F_n = F_{n-1} + F_{n-2}$
Eigenvalues
Eigenvectors
$F(v) = \lambda v$
Week 8-10
Dynamical Systems
$x_{t+1} = Ax_t$
θ
$f''(t)+f(t)=2t$
FINISH
Week 11 Applications
LINEAR DIFFERENTIAL EQUATIONS
FINALS
19.9%
4.2%
10.1%

Linear Algebra I

Introduction

In the realm of numbers and vectors we wade,
Linear Algebra, where scientific problems are laid.
Automotive, physics, chemistry, and bio too,
All fields that this essential course will accrue.

Linear systems we'll study, matrices we'll discern,
Through sets and functions, much knowledge we'll earn.
Linear maps transform, in spaces they unwind,
Subspaces and kernels, in invertible maps they're confined.

Independence and bases, dimensions we'll chart,
Each concept a masterpiece, a mathematical art.
As coordinates shift, new perspectives arise,
With Gram-Schmidt's process, orthonormal bases materialize.

Projections orthogonal, and least squares approximation,
Tools for data fitting, across each student's vocation.
"Linear Algebra II," a deeper journey we embark,
Where vector spaces and linear maps leave their mark.

Determinants and eigenvalues, the story further unfurls,
With linear differential equations, like precious pearls.
This course, a bridge from theory to practical narration,
Equipping students with robust mathematical foundation.

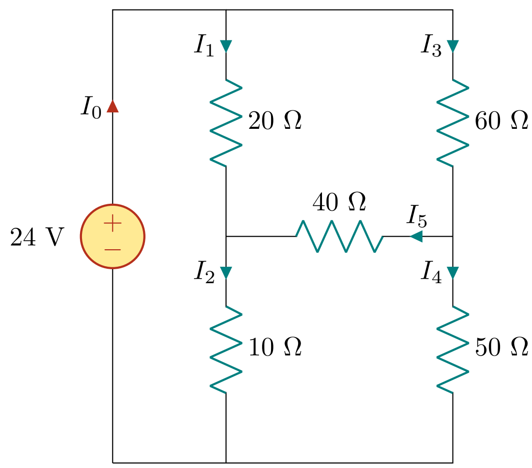
– ChatGPT

Linear Algebra, a fundamental branch of mathematics, offers essential tools and language to articulate and solve a broad array of problems across diverse scientific disciplines. For those studying automotive engineering, physics, chemistry, and bio-agriculture, the principles embedded within this field become invaluable in comprehending and modeling complex systems pertinent to their areas of study.

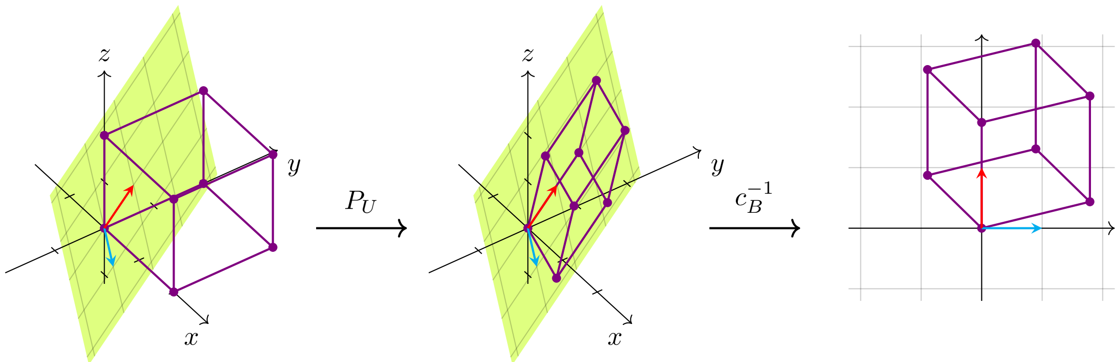
In this course, we will study vector spaces and linear transformations between these spaces, often referred to as "flat spaces" in a geometric sense. In these spaces, the basic operations of addition and scalar multiplication behave in a "linear" way, meaning they satisfy properties like distributivity, commutativity, and associativity. Unlike curved spaces, which are studied in differential geometry, linear spaces are characterized by the absence of curvature, making them easier to analyze and understand.

The study of linear spaces is crucial for various applications in science and engineering, as they often serve as good approximations for more complex structures. Whether it is solving systems of equations, analyzing data sets, or transforming shapes in computer graphics, the principles of linear algebra are foundational.

For example, in electrical circuits, solving linear systems of equations is often essential for applying Kirchhoff's laws, which govern the conservation of charge and energy in the circuit. Kirchhoff's current law states that the sum of currents entering a junction must equal the sum of currents leaving it, while Kirchhoff's voltage law states that the sum of the voltages around any closed loop in a circuit must be zero. These laws can be translated into a system of linear equations where the unknowns are the currents or voltages in the circuit componentes. In the diagram on the right, you see an example of an electrical circuit with unknown currents $I_0, \ldots, I_5$. In Example 7, we show how to calculate these by solving a linear system.



After studying linear systems and matrices and vectors, we will talk about linear maps. Often these maps have some kind of geometric interpretation, and their applications are endless. For example, if you want to program a 3D-game engine, you need to project 3-dimensional objects onto a 2-dimensional space (your monitor). This will be done by a linear map. Later in this course we will illustrate this by explaining how a 3-dimensional cube can be displayed in the plane after choosing a certain viewing angle and we will give a explanation of the following illustration:



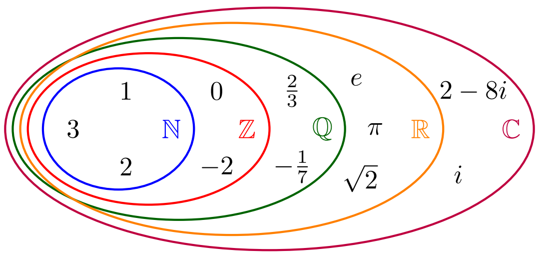
1

Linear systems

In linear algebra, linear systems play a crucial role in understanding and solving various real-world problems. A linear system is a collection of linear equations involving the same variables. These equations represent relationships between variables and can be used to model various scenarios from diverse fields such as physics, economics, and engineering.

We will denote the set of **real numbers** by $\mathbb{R}$. $\mathbb{R}$ contains all numbers usually considered in high school, such as $1, -1, 0, 2, 3, \frac{3}{8}, \pi, \sqrt{2}, e, \ldots$. There are rigorous definitions of the real numbers, which would be part of a pure mathematics lecture[1]. But in this course, we just assume that they exist and that everyone is familiar with them.

♡ *Remark.* Even though we will focus on real numbers most of the time in this course, we will also mention the usual notations for certain subsets of them.



The **natural numbers** $\mathbb{N}$ consists of the number $1, 2, 3, \ldots$. Allowing 0 and negative numbers leads to the integers $\mathbb{Z} = \{0, -1, 1, 2, -2, \ldots\}$. The set of fractions $\frac{a}{b}$ with $a \in \mathbb{Z}, b \in \mathbb{N}$ are called **rational numbers**, denoted by $\mathbb{Q}$. But not all numbers appearing "naturally" can be written as fractions. For example, the diagonal of a square of side-length one is, by Pythagoras, the positive solution of $x^2 = 1 + 1 = 2$, i.e. $x = \sqrt{2}$. But one can show that $\sqrt{2}$ is not rational.
The "completion" of rational numbers, which can be think of as filling up the missing gaps, leads then to the real numbers $\mathbb{R}$ mentioned above. The story does not end here, since there is an even bigger class of numbers which are often of interest. For example, one might be interested in solutions of the equation $x^2 = -1$, which does not exist if one just allows real numbers. This leads to the notion of **complex numbers** $\mathbb{C}$, which are numbers of the form $a + bi$ with $a, b \in \mathbb{R}$ and i being a new symbol satisfying $i^2 = -1$.

A linear equation can be written in the form:

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

where $a_1, a_2, \ldots, a_n \in \mathbb{R}$ and $b \in \mathbb{R}$ are constants, and $x_1, x_2, \ldots, x_n$ are variables.

Let us first consider a real-life situation that can be described by a linear system.

Example 1 (Cakes) A chocolate-obsessed patisserie only sells two types of cake: chocolate tart (henceforth, 'tarts') and chocolate cake (henceforth, 'cake'). To create one cake, 2 bars of chocolate, 3 tablespoons of sugar, and four eggs are required. For one tart, 3 bars of chocolate,

[1] See for example `https://en.wikipedia.org/wiki/Construction_of_the_real_numbers` for an overview of the "construction" of real numbers.

5 tablespoons of sugar, and 7 eggs are required.
On one particular day, the patisserie used 77 bars of chocolate, 124 tablespoons of sugar, and 137 eggs was used up. We want to find out how many cakes and tarts they made.
If we denote the number of cakes by x_1, and the number of tarts by x_2, then we can write three equations relating the number of cakes and tarts with the number of bars of chocolate, tablespoons of sugar, and eggs used:

$$\begin{aligned} 2x_1 + 3x_2 &= 77 \\ 3x_1 + 5x_2 &= 124 \\ 4x_1 + 7x_2 &= 171 \end{aligned}$$

To solve such simultaneous equations, one can use either substitution or elimination method. In this example, let us use the elimination method (which entails 'eliminating' variables).
First, we multiply the first equation by 2 to obtain:

$$4x_1 + 6x_2 = 154$$

By subtracting this from the third equation, we eliminate the variable x:

$$4x_1 + 7x_2 - (4x_1 + 6x_2) = 171 - 154 \Longrightarrow x_2 = 17$$

We can then find the value of x_1 by plugging $x_2 = 17$ into the first equation:

$$2x_1 + 3 \cdot 17 = 77 \Longrightarrow x_1 = \frac{1}{2}(77 - 51) = 13$$

As such, by solving the system of linear equations, we know that on that day, the patisserie made $x_1 = 13$ cakes and $x_2 = 17$ tarts.

No one reading these notes will probably own a patisserie or will be interested in solving problems as in the above example. But we will see that linear systems, and in particular the study of their solutions, arise in various serious real life applications. First let us fix the notation as follows:

Definition 1.1 (i) For real numbers $a_1, a_2, \dots, a_n, b \in \mathbb{R}$ an equation of the form

$$a_1x_1 + \dots + a_nx_n = b$$

is called a **linear equation**.
(ii) A finite collection of linear equations is called a **linear system**.
(iii) A **solution of a linear system** is a simultaneous solution for all of its equations.

Goal: Given a linear system we want to find all of its solutions.

While the method used in Example 1 is fine for smaller systems, larger, more complicated systems would be a nightmare to solve using such crude methods. As such, we must develop a more systematic method of solving such linear systems. One way is to add multiples of one equation to another one, or multiply an equation with a non-zero number. By doing this correctly, a new linear system with clearer solutions may be obtained from the original linear system.

Example 2 (Unique solution) Consider and solve the following linear system

$$\begin{array}{c} 2 \\ \end{array}\begin{cases} x_1 + 3x_2 = 1 \\ -2x_1 + x_2 = 2 \end{cases} \Rightarrow \begin{array}{c} \\ \frac{1}{7} \end{array}\begin{cases} x_1 + 3x_2 = 1 \\ 7x_2 = 4 \end{cases} \Rightarrow \begin{array}{c} \\ -3 \end{array}\begin{cases} x_1 + 3x_2 = 1 \\ x_2 = \frac{4}{7} \end{cases} \Rightarrow \begin{cases} x_1 = -\frac{5}{7} \\ x_2 = \frac{4}{7} \end{cases}$$

Is $x_1 = -\frac{5}{7}$, $x_2 = \frac{4}{7}$ a solution to the original linear system? The answer is yes because the operations work also in reverse.

$$\begin{cases} x_1 = -\frac{5}{7} \\ x_2 = \frac{4}{7} \end{cases} \Rightarrow \begin{cases} x_1 + 3x_2 = 1 \\ x_2 = \frac{4}{7} \end{cases} \Rightarrow \begin{cases} x_1 + 3x_2 = 1 \\ 7x_2 = 4 \end{cases} \Rightarrow \begin{cases} x_1 + 3x_2 = 1 \\ -2x_1 + x_2 = 2 \end{cases}$$

Therefore,

$$\begin{cases} x_1 + 3x_2 = 1 \\ -2x_1 + x_2 = 2 \end{cases} \iff \begin{cases} x_1 = -\frac{5}{7} \\ x_2 = \frac{4}{7} \end{cases}$$

This linear system has exactly one solution.

Remark. For anyone who is unfamiliar with the notations of implication "$\Rightarrow$" and equivalence "$\Leftrightarrow$", consider two statements denoted by p and q. In mathematics, statements are sentences or equations or inequalities, which are true or false, no ambivalence. Given the statements p and q, we can form new statements $p \Rightarrow q$ and $p \Leftrightarrow q$.

When p is true, $p \Rightarrow q$ is true if q is true and false if q is false. Otherwise, when p is false, $p \Rightarrow q$ is true regardless of q. The statement $p \Rightarrow q$ can be read as "the statement p implies the statement q" or "if p, then q". The meaning is that the truth of p leads to the truth of q, which makes $p \Rightarrow q$ true. For example, if $x = 2$, then $x^2 = x \cdot x = 2 \cdot 2 = 4$. For that, we can write $x = 2 \Rightarrow x^2 = 4$ and this statement is true. For another example, the statement $x < 0 \Rightarrow x^3 > 0$ is false because if $x = -1 < 0$, then $x^3 = (-1)^3 = -1 < 0$. We can sometimes see true statements $p \Rightarrow q$ where p is always false such as $x^2 < 0 \Rightarrow x = 100$. In these cases, we say that they are vacuously true.

The statement $p \Leftrightarrow q$ is a combination of $p \Rightarrow q$ and $q \Rightarrow p$. This statement is true when p, q are both true or both false. Therefore, the truth of $p \Leftrightarrow q$ means that p is equivalent to q and vice versa. It can be read as "The statement p is equivalent to the statement q" or "p if and only if q". For example, $x = 1 \Leftrightarrow x + 1 = 2$ is true because if $x = 1$ then $x + 1 = 1 + 1 = 2$, and if $x + 1 = 2$ then $x = (x + 1) - 1 = 2 - 1 = 1$. For another example, the statement $x = 2 \Leftrightarrow x^2 = 4$ is false because when $x = -2$, we have $x^2 = (-2)^2 = 4$ and hence, $x^2 = 4$ does not imply $x = 2$.

In Examples 1 and 2, the linear system has a unique solution. However, not all linear systems have unique solutions; some have no solutions (for example, the system $x_1 = 2$, $x_1 = 3$ has no solution), and some have infinitely many solutions. An example of such a linear system is given below.

Example 3 (Infinitely many solutions) Consider the following linear system:

$$\begin{matrix} -2 & -3 \end{matrix}\ \begin{cases} x_1 - 9x_2 - 3x_3 + x_4 = 4 \\ 3x_1 - 2x_2 + x_3 - 2x_4 = 2 \\ 2x_1 + 7x_2 + 4x_3 - 3x_4 = -2 \\ 25x_2 + 9x_3 - 2x_4 = -4 \end{cases} \Leftrightarrow \begin{matrix} -1 & -1 \end{matrix}\ \begin{cases} x_1 - 9x_2 - 3x_3 + x_4 = 4 \\ 25x_2 + 10x_3 - 5x_4 = -10 \\ 25x_2 + 10x_3 - 5x_4 = -10 \\ 25x_2 + 9x_3 - 2x_4 = -4 \end{cases}$$

$$\Leftrightarrow \begin{matrix} -3 & 10 \end{matrix}\ \begin{cases} x_1 - 9x_2 - 3x_3 + x_4 = 4 \\ 25x_2 + 10x_3 - 5x_4 = -10 \\ 0 = 0 \\ x_3 - 3x_4 = 6 \end{cases} \Leftrightarrow \begin{matrix} \frac{1}{25} \\ -1 \end{matrix}\ \begin{cases} x_1 - 9x_2 - 8x_4 = -14 \\ 25x_2 + 25x_4 = 50 \\ -x_3 + 3x_4 = 6 \end{cases}$$

row-reduced echelon form

$$\Leftrightarrow \quad 9 \quad \begin{cases} x_1 - 9x_2 - 8x_4 = -14 \\ x_2 + x_4 = 2 \\ x_3 - 3x_4 = -6 \end{cases} \Leftrightarrow \boxed{\begin{cases} x_1 + x_4 = 4 \\ x_2 + x_4 = 2 \\ x_3 - 3x_4 = -6 \end{cases}} \quad (*)$$

$$\Leftrightarrow \begin{cases} x_1 = 4 - x_4 \\ x_2 = 2 - x_4 \\ x_3 = -6 + 3x_4 \end{cases}$$

The variable x_4 is arbitrary.

We set $x_4 = t$ for some $t \in \mathbb{R}$.

All solutions are given by $\begin{cases} x_1 = 4 - t \\ x_2 = 2 - t \\ x_3 = -6 + 3t \\ x_4 = t \end{cases}$ for $t \in \mathbb{R}$. This means that this linear system has infinitely many solutions.

In the linear system $(*)$:

- Each equation contains a variable that occurs in no other equation: (x_1, x_2, x_3), called **pivot variables**.
- The other variables (x_4) are called **free variables**.

A linear system of this shape said to be on **row-reduced echelon form**. In general, this means that the following three conditions are satisfied:

(i) The first (that is, the leftmost) variable in each equation has coefficient 1.

(ii) If x_i is the first variable in one of the equations, then it does not occur in any other equation in the system.

(iii) If x_i is the first variable in one equation, then the equations below it do not contain any of the variables $x_1, x_2, \ldots, x_{i-1}$.

As we saw in the above example we only need three different operations to bring any linear system to row-reduced echelon form:

Definition 1.2 The following operations on a linear system are called **elementary row operations**.
(R1) Add a multiple of an equation to another.
(R2) Multiply an equation with a non-zero number.
(R3) Change the order of the equations.

Since all elementary row operations work in reverse (i.e., all elementary row operations can be undone),

Proposition 1.3 *Applying an elementary row operation to a linear system does not change the set of all solutions of said linear system.*

To bring an arbitrary linear system onto row-reduced echelon form we can use the following algorithm.

Algorithm 1.4 (Gaussian elimination / Row reduction) Given a linear system

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases},$$

The procedure for bringing this linear system to its row-reduced echelon form is as follows:

I. Downwards:

1) Make the first equation contain the first variable by using (R3).
2) Make the coefficient of this variable equal to 1 by using (R2).
3) Eliminate this variable from all other equations by using (R1).
4) Iterate with the first occurring variable in the remaining equations.

II. Upwards

1) Let x_i be the first variable in the last equation. Eliminate x_i from all other equations by using (R1).
2) Go to previous equations and iterate.

One example of the usage of Gaussian elimination to solve a linear system is given below:

Example 4 Consider the linear system below. First, the 'downwards' part:

$$\overset{\text{(R3)}}{\begin{cases} -x_3 + x_4 = 0 \\ x_1 + x_2 + x_3 - x_4 = 0 \\ x_1 + x_2 - x_3 = 1 \end{cases}} \Leftrightarrow \overset{\text{(R1)}}{-1\begin{cases} x_1 + x_2 + x_3 - x_4 = 0 \\ -x_3 + x_4 = 0 \\ x_1 + x_2 - x_3 = 1 \end{cases}}$$

$$\Leftrightarrow \overset{\text{(R2)}}{-1\begin{cases} x_1 + x_2 + x_3 - x_4 = 0 \\ -x_3 + x_4 = 0 \\ -2x_3 + x_4 = 1 \end{cases}} \Leftrightarrow \overset{\text{(R1)}}{2\begin{cases} x_1 + x_2 + x_3 - x_4 = 0 \\ x_3 - x_4 = 0 \\ -2x_3 + x_4 = 1 \end{cases}}$$

From here on upwards

$$\Leftrightarrow \overset{\text{(R2)}}{-1\begin{cases} x_1 + x_2 + x_3 - x_4 = 0 \\ x_3 - x_4 = 0 \\ -x_4 = 1 \end{cases}} \Leftrightarrow \overset{\text{(R1)}}{1 \quad 1\begin{cases} x_1 + x_2 + x_3 - x_4 = 0 \\ x_3 - x_4 = 0 \\ x_4 = -1 \end{cases}}$$

$$\Leftrightarrow \overset{\text{(R1)}}{-1\begin{cases} x_1 + x_2 + x_3 = -1 \\ x_3 = -1 \\ x_4 = -1 \end{cases}} \Leftrightarrow \begin{cases} x_1 + x_2 = 0 \\ x_3 = -1 \\ x_4 = -1 \end{cases}$$

pivot variables
free variables

The free variables can be chosen arbitrarily. We set $x_2 = t$ with an arbitrary $t \in \mathbb{R}$.

All solutions are given by $\begin{cases} x_1 = -t \\ x_2 = t \\ x_3 = -1 \\ x_4 = -1 \end{cases}$ for $t \in \mathbb{R}$.

So far, all linear systems considered had always one (or infinitely many) solutions. One example of using Gaussian elimination to prove that a linear system has no solution is given below.

Example 5 (No solution) Consider the linear system below.

$$\begin{cases} x_1 + 2x_2 + 3x_3 = 1 \\ x_1 + 3x_2 + 4x_3 = 3 \\ x_1 + 4x_2 + 5x_3 = 4 \end{cases} \Leftrightarrow \begin{cases} x_1 + 2x_2 + 3x_3 = 1 \\ \quad\;\; x_2 + x_3 = 2 \\ \quad 2x_2 + 2x_3 = 3 \end{cases} \Leftrightarrow \begin{cases} x_1 + 2x_2 + 3x_3 = 1 \\ \quad\;\; x_2 + x_3 = 2 \\ \qquad\qquad 0 = -1 \end{cases}$$

This shows that the original linear system has no solution because it is equivalent to a linear system containing a contradictory equation (of course, $0 \neq -1$!).

Another type of problem that can be solved using this method involves a linear system that is parameterized (i.e., some part of the linear system is determined by a parameter). Usually, one will be asked to consider how the solution to the linear system look with different values of the parameter. One typical problem of this kind is given below:

Example 6 (Parameterized) Consider the following linear system with a parameter $a \in \mathbb{R}$:

$$\begin{cases} (a-1)x_1 + 3x_2 = 2 \\ \qquad\;\; x_1 - x_2 = 1 \end{cases}.$$

We want to determine for which real numbers a the linear system has solutions and find all the solutions in these cases. To find the solutions of this linear system, we try to bring it on row-reduced echelon form.

$$\begin{cases} (a-1)x_1 + 3x_2 = 2 \\ \qquad\;\; x_1 - x_2 = 1 \end{cases} \Leftrightarrow \begin{cases} \qquad\;\; x_1 - x_2 = 1 \\ (a-1)x_1 + 3x_2 = 2 \end{cases} \Leftrightarrow \begin{cases} x_1 - \qquad\;\; x_2 = 1 \\ \qquad (a+2)x_2 = 3-a \end{cases}$$

Now we would like to divide by $2 + a$, but this is not possible if $a = -2$ (division by zero). Therefore, we assume that $a \neq -2$ and consider the $a = -2$ case separately.

- Case $a \neq -2$:

$$\begin{cases} x_1 - \qquad\;\; x_2 = 1 \\ \qquad (a+2)x_2 = 3-a \end{cases} \Leftrightarrow \begin{cases} x_1 - x_2 = 1 \\ \qquad x_2 = \frac{3-a}{2+a} \end{cases} \Leftrightarrow \begin{cases} x_1 \qquad = \frac{5}{2+a} \\ \qquad x_2 = \frac{3-a}{2+a} \end{cases}$$

  For the case $a \neq -2$ there is exactly one solution given by $\begin{cases} x_1 = \frac{5}{2+a} \\ x_2 = \frac{3-a}{2+a} \end{cases}$.

- Case $a = -2$:

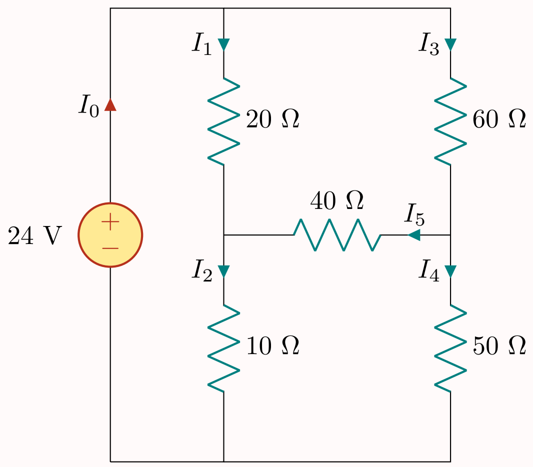
$$\begin{cases} x_1 - \qquad\;\; x_2 = 1 \\ \qquad (a+2)x_2 = 3-a \end{cases} \Leftrightarrow \begin{cases} x_1 - x_2 = 1 \\ \qquad\;\; 0 = 5 \end{cases}$$

  There are no solutions in the case $a = -2$.

Example 7 (Circuit analysis) Given the following circuit, we want to determine all currents through resistors.

By Kirchhoff's laws, we can derive the following linear system:

$$\begin{cases} I_1 - I_2 + I_5 = 0 \\ I_3 - I_4 - I_5 = 0 \\ 20I_1 + 10I_2 = 24 \\ -20I_1 + 60I_3 + 40I_5 = 0 \\ -10I_2 + 50I_4 - 40I_5 = 0 \end{cases}.$$



$$20 \quad -20 \qquad \left\{\begin{array}{rrrrrl} I_1 & - I_2 & & & + I_5 & = 0 \\ & & I_3 & - I_4 & - I_5 & = 0 \\ 20I_1 & + 10I_2 & & & & = 24 \\ -20I_1 & & + 60I_3 & & + 40I_5 & = 0 \\ & - 10I_2 & & + 50I_4 & - 40I_5 & = 0 \end{array}\right.$$

$$\Leftrightarrow \quad \tfrac{1}{30} \left\{\begin{array}{rrrrl} I_1 - I_2 & & & + I_5 & = 0 \\ & I_3 & - I_4 & - I_5 & = 0 \\ 30I_2 & & & - 20I_5 & = 24 \\ - 20I_2 & + 60I_3 & & + 60I_5 & = 0 \\ - 10I_2 & & + 50I_4 & - 40I_5 & = 0 \end{array}\right. \Leftrightarrow \left\{\begin{array}{rrrrl} I_1 - I_2 & & & + I_5 & = 0 \\ & I_3 & - I_4 & - I_5 & = 0 \\ I_2 & & & - \frac{2}{3}I_5 & = \frac{4}{5} \\ - 20I_2 & + 60I_3 & & + 60I_5 & = 0 \\ - 10I_2 & & + 50I_4 & - 40I_5 & = 0 \end{array}\right.$$

$$\Leftrightarrow \quad 10 \quad 20 \quad \left\{\begin{array}{rrrrl} I_1 - I_2 & & & + I_5 & = 0 \\ I_2 & & & - \frac{2}{3}I_5 & = \frac{4}{5} \\ & I_3 & - I_4 & - I_5 & = 0 \\ - 20I_2 & + 60I_3 & & + 60I_5 & = 0 \\ - 10I_2 & & + 50I_4 & - 40I_5 & = 0 \end{array}\right.$$

$$\Leftrightarrow \quad -60 \left\{\begin{array}{rrrrl} I_1 - I_2 & & & + I_5 & = 0 \\ I_2 & & & - \frac{2}{3}I_5 & = \frac{4}{5} \\ & I_3 & - I_4 & - I_5 & = 0 \\ & 60I_3 & & + \frac{140}{3}I_5 & = 16 \\ & & 50I_4 & - \frac{140}{3}I_5 & = 8 \end{array}\right. \Leftrightarrow \quad \tfrac{1}{60} \left\{\begin{array}{rrrrl} I_1 - I_2 & & & + I_5 & = 0 \\ I_2 & & & - \frac{2}{3}I_5 & = \frac{4}{5} \\ & I_3 & - I_4 & - I_5 & = 0 \\ & & 60I_4 & + \frac{320}{3}I_5 & = 16 \\ & & 50I_4 & - \frac{140}{3}I_5 & = 8 \end{array}\right.$$

$$\Leftrightarrow \quad -50 \left\{\begin{array}{rrrrl} I_1 - I_2 & & & + I_5 & = 0 \\ I_2 & & & - \frac{2}{3}I_5 & = \frac{4}{5} \\ & I_3 & - I_4 & - I_5 & = 0 \\ & & I_4 & + \frac{16}{9}I_5 & = \frac{4}{15} \\ & & 50I_4 & - \frac{140}{3}I_5 & = 8 \end{array}\right. \Leftrightarrow \quad -\tfrac{9}{1220} \left\{\begin{array}{rrrrl} I_1 - I_2 & & & + I_5 & = 0 \\ I_2 & & & - \frac{2}{3}I_5 & = \frac{4}{5} \\ & I_3 & - I_4 & - I_5 & = 0 \\ & & I_4 & + \frac{16}{9}I_5 & = \frac{4}{15} \\ & & & - \frac{1220}{9}I_5 & = -\frac{16}{3} \end{array}\right.$$

$$
\Leftrightarrow \quad
\begin{array}{r}
(-1) \\ (\tfrac{2}{3}) \\ (1) \\ (-\tfrac{16}{9}) \\ \\
\end{array}
\left\{
\begin{array}{ccccccccccc}
I_1 & - & I_2 & & & & & + & I_5 & = & 0 \\
 & & I_2 & & & & & - & \frac{2}{3}I_5 & = & \frac{4}{5} \\
 & & & & I_3 & - & I_4 & - & I_5 & = & 0 \\
 & & & & & & I_4 & + & \frac{16}{9}I_5 & = & \frac{4}{15} \\
 & & & & & & & & I_5 & = & \frac{12}{305}
\end{array}
\right.
\Leftrightarrow \quad
\begin{array}{r}
(1) \\ \\ (1) \\ \\ \\
\end{array}
\left\{
\begin{array}{ccccccccc}
I_1 & - & I_2 & & & & & = & -\frac{12}{305} \\
 & & I_2 & & & & & = & \frac{252}{305} \\
 & & & & I_3 & - & I_4 & = & \frac{12}{305} \\
 & & & & & & I_4 & = & \frac{12}{61} \\
 & & & & & & I_5 & = & \frac{12}{305}
\end{array}
\right.
$$

$$
\Leftrightarrow \left\{
\begin{array}{ccccccc}
I_1 & & & & & = & \frac{48}{61} \\
 & I_2 & & & & = & \frac{252}{305} \\
 & & I_3 & & & = & \frac{72}{305} \\
 & & & I_4 & & = & \frac{12}{61} \\
 & & & & I_5 & = & \frac{12}{305}
\end{array}
\right.
$$

In addition, we can determine the current coming directly from the source.

$$I_0 = I_1 + I_3 = I_2 + I_4 = \frac{312}{305} \text{ A.}$$

Exercises

Exercise 1. Which of the following linear systems are on row-reduced echelon form? For those that are not, find an equivalent system (i.e. one which has the same solutions) that is on row-reduced echelon form. For each system, find all solutions.

(i) $\left\{ \begin{array}{ccccccc} x_1 & + & 4x_2 & + & 7x_3 & = & 1 \\ 2x_1 & + & 5x_2 & + & 8x_3 & = & 2 \\ 3x_1 & + & 6x_2 & + & 10x_3 & = & 1 \end{array} \right.$

(ii) $x_1 + 2x_2 + 3x_3 + 4x_4 = 2022$

(iii) $\left\{ \begin{array}{ccccccccc} x_1 & + & x_2 & + & x_3 & + & 2x_4 & = & 0 \\ & & x_2 & & & + & x_4 & = & 0 \end{array} \right.$

(iv) $\left\{ \begin{array}{ccccc} x_1 & + & 2x_2 & = & 3 \\ 4x_1 & + & 8x_2 & = & 16 \end{array} \right.$

(v) $\left\{ \begin{array}{ccc} x_1 & = & 6 \\ x_2 & = & 9 \\ x_3 & = & 1 \end{array} \right.$

Exercise 2. Decide for which real numbers $a \in \mathbb{R}$ the following linear system has solutions. Give all the solutions in these cases.

$$\left\{ \begin{array}{ccccccc} (a-1)^2 x_1 & + & x_2 & + & ax_3 & = & 0 \\ x_1 & + & x_2 & & & = & 0 \\ 2x_1 & + & 2x_2 & + & x_3 & = & a \end{array} \right. .$$

Exercise 3. Let $a, b \in \mathbb{R}$ be two arbitrary real numbers. Consider the following linear system:

$$\left\{ \begin{array}{ccccc} x_1 & + & x_2 & = & 2 \\ ax_1 & + & 2x_2 & = & b \end{array} \right.$$

Find all the solutions to this linear system depending on a and b.
(Hint: You need to consider different special cases of a and b separately)

Exercise 4. Decide for which real numbers $a \in \mathbb{R}$ the following linear system has solutions. Give all the solutions in these cases.

$$\begin{cases} 2x_1 + 12x_2 + 7x_3 = 12a + 7 \\ 2x_1 + 4x_2 + 2x_3 = 12a \\ x_1 + 10x_2 + 6x_3 = 7a + 8 \end{cases}.$$

Exercise 5. A ramen store in Sakae offers three types of ramen: Miso ramen (price for one portion: 700¥), Taiwan ramen (800¥), and Tonkotsu ramen (850¥). For one portion of Miso ramen one needs 3 tablespoons (tbsp) of salt, one clove of garlic and no chili. One portion of Taiwan ramen needs 2 tbsp. of salt, 2 cloves of garlic and 4 tbsp. of chili. For one portion of Tonkotsu ramen 2 tbsp. of salt, 3 cloves of garlic and one tbsp. of chili is needed.[a] In one day the store uses 142 tbsp. of salt, 146 cloves of garlic, and 152 tbsp. of chili.
How much money (in ¥) did the store earn on this day? Describe this problem by using a linear system and then solve it.

[a] These amounts are made up and should probably not be used to make tasty ramen.

Exercise 6. (8 Points) A Japanese restaurant in 八事(やごと)(Yagoto, a neighbourhood in Nagoya) is holding an Ebi Festival, and thus is only selling three types of dishes: Ebi Sushi (¥370), Ebi Tempura Don (¥590), and Ebi Fry Bentō (¥830).
One serving of Ebi Sushi requires 3 ounces of shrimp, 1 cup of rice, and 3 tablespoon of shouyu. 5 ounces of shrimp, 4 cups of rice, and $\frac{5}{2}$ tablespoons of shouyu are needed for one portion of Ebi Tempura Don. For one serving of Ebi Fry Bento, 8 ounces of shrimp, 3 cups of rice, and $\frac{1}{2}$ tablespoons of shouyu are needed. In one certain day, the store expended 1000 ounces of shrimp, 500 cups of rice, and 500 tablespoons of shouyu.
The market prices are: ¥50 per ounce of shrimp, ¥30 per cup of rice, and ¥5 per tablespoon of shouyu. Given all these information, how much profit did the restaurant make on this certain day? Describe this problem by using a linear system, bring the linear system on row-reduced echelon form and solve it.

2

Matrices & Vectors

In the previous chapter, we studied linear systems and learned how to solve them using elementary operations. However, as the number of variables and equations increases, these methods become increasingly cumbersome. To overcome this difficulty, we introduce vectors and matrices in this chapter, which allow us to write linear systems in a more concise and elegant manner. Vectors and matrices are fundamental tools in linear algebra, and we will explore their properties and operations in detail.

Definition 2.1

(i) A $m \times n$-**matrix** is given by an array (m rows, n columns) of numbers $a_{ij} \in \mathbb{R}$

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} = (a_{ij})_{\substack{1 \le i \le m \\ 1 \le j \le n}} .$$

Notation: We often just write $A = (a_{ij})$ if the size of A, i.e. m and n, are known from context.
By $\mathbb{R}^{m \times n}$ we denote the set all of all $m \times n$-matrices.

(ii) A (column-) **vector** of size n is a $n \times 1$-matrix

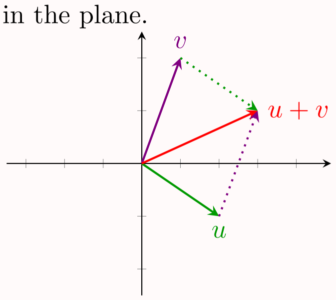
$$v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

and the set of all vectors of size n is denoted by $\mathbb{R}^n = \mathbb{R}^{n \times 1}$.

Example 8 For $n = 2$ we can visualize vectors in the plane.

$$v = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

$$u = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$$



We can also add vectors e.g. $u + v = \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$.

In general the sum of matrices is defined by just adding each entry.

Definition 2.2 For matrices $A = (a_{ij}), B = (b_{ij}) \in \mathbb{R}^{m\times n}$ and a real number $\lambda \in \mathbb{R}$ we define

$$\begin{aligned} A + B &= (a_{ij} + b_{ij}) \in \mathbb{R}^{m\times n} && \text{(Sum of two matrices)}\,, \\ \lambda A &= (\lambda a_{ij}) \in \mathbb{R}^{m\times n} && \text{(Scalar multiplication)}\,. \end{aligned}$$

In the case $\lambda = -1$ we write $(-1)A = -A$ and $A - B$ means $A + (-1)B$.

The matrices A and B need to be of the same size, otherwise the sum $A + B$ is not defined. A special case of the addition of matrices is given by the addition of vectors. For $u, v \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$ we have

$$u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}, v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}, \quad u + v = \begin{pmatrix} u_1 + v_1 \\ \vdots \\ u_n + v_n \end{pmatrix}, \quad \lambda v = \begin{pmatrix} \lambda v_1 \\ \vdots \\ \lambda v_n \end{pmatrix}.$$

Definition 2.3 The product of a matrix $A = (a_{ij}) \in \mathbb{R}^{m\times n}$ and a vector $v \in \mathbb{R}^n$ is defined by

$$Av = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} a_{11}v_1 + a_{12}v_2 + \cdots + a_{1n}v_n \\ a_{21}v_1 + a_{22}v_2 + \cdots + a_{2n}v_n \\ \vdots \\ a_{m1}v_1 + a_{m2}v_2 + \cdots + a_{mn}v_n \end{pmatrix} \in \mathbb{R}^m.$$

We have: ($m\times n$-matrix) $\cdot$ (vector of size n) $=$ (vector of size m).

Example 9 Here is the product of a 3×2 matrix and a vector in $\mathbb{R}^2$:

$$\begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 3 \end{pmatrix} = \begin{pmatrix} 1\cdot(-1) + 4\cdot 3 \\ 2\cdot(-1) + 5\cdot 3 \\ 3\cdot(-1) + 6\cdot 3 \end{pmatrix} = \begin{pmatrix} 11 \\ 13 \\ 15 \end{pmatrix},$$

where we get a vector in $\mathbb{R}^3$.

This product of a matrix and a vector satisfies the following rules.

Proposition 2.4 *We have for $A \in \mathbb{R}^{m\times n}$, $x, y \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$*

(i) $A(x + y) = Ax + Ay$,

(ii) $A(\lambda x) = \lambda(Ax)$.

Proof. This is Exercise 7. □

Example 10 Let $A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 4 \end{pmatrix}$, $b = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$. Find all $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3$ with $Ax = b$.

$$Ax = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 + 2x_2 + 3x_3 \\ x_1 + x_2 + 4x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} = b \quad \Leftrightarrow \quad \begin{cases} x_1 + 2x_2 + 3x_3 = 1 \\ x_1 + x_2 + 4x_3 = 2 \end{cases}.$$

This is a linear system. We also call $Ax = b$ a linear system because it gives us a linear system.

Solving:

$$\begin{matrix} (-1) \\ \end{matrix} \begin{cases} x_1 + 2x_2 + 3x_3 = 1 \\ x_1 + x_2 + 4x_3 = 2 \end{cases} \Leftrightarrow \begin{matrix} \\ (2) \end{matrix} \begin{cases} x_1 + 2x_2 + 3x_3 = 1 \\ \quad - x_2 + x_3 = 1 \end{cases}$$

$$\Leftrightarrow \quad \begin{matrix} \\ -1 \end{matrix}\begin{cases} x_1 & & + 5x_3 & = & 3 \\ & - x_2 & + \ \ x_3 & = & 1 \end{cases} \quad \Leftrightarrow \quad \overset{\text{pivot variables}}{\begin{cases} x_1 & & + 5x_3 & = & 3 \\ & x_2 & - \ \ x_3 & = & -1 \end{cases}}$$

free variable

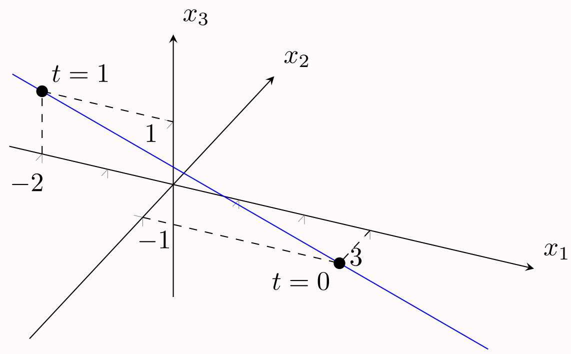
Solution: $\begin{cases} x_1 = 3 - 5t \\ x_2 = -1 + t \\ x_3 = t \end{cases}$ for $t \in \mathbb{R}$. Using the vector notation, this can be written as:

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 3 - 5t \\ -1 + t \\ t \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \\ 0 \end{pmatrix} + t \begin{pmatrix} -5 \\ 1 \\ 1 \end{pmatrix}.$$

Plotting the above solution for all possible values of t gives a line in $\mathbb{R}^3$.

$t = 0 : x = \begin{pmatrix} 3 \\ -1 \\ 0 \end{pmatrix}$

$t = 1 : x = \begin{pmatrix} 3 \\ -1 \\ 0 \end{pmatrix} + \begin{pmatrix} -5 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$



We can also use the matrix notation when solving a linear system to avoid having to write the symbols x_i of variables all the time as follows.

Definition 2.5 For a matrix $A = (a_{ij}) \in \mathbb{R}^{m \times n}$ and a vector $b \in \mathbb{R}^m$ the matrix

$$(A \mid b) = \left(\begin{array}{ccc|c} a_{11} & \dots & a_{1n} & b_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \dots & a_{mn} & b_m \end{array}\right) \in \mathbb{R}^{m \times (n+1)}$$

is called the **augmented matrix** of the linear system $Ax = b$.

The augmented matrix $(A \mid b)$ is just the matrix A where we append the vector b as a column. The line $\mid$ is a useful notation to distinguish between the left- and right-hand side of the corresponding linear system but it has no mathematical meaning. We will view $(A \mid b)$ as a usual matrix with m rows and $n + 1$ columns.

Definition 2.6 The following operations on a matrix are called **elementary row operations**.
(R1) Add a multiple of one row to another row.
(R2) Multiply a row with a non-zero number.
(R3) Interchange two rows.

Applying a row operation to a linear system (Definition 1.2) corresponds to the same row operation (Definition 2.6) on the corresponding augmented matrix of this linear system.

Definition 2.7 Two matrices A and B are called **row equivalent**, if B can be obtained from A by elementary row operations. In this case we write

$$A \sim B.$$

Notice that if $A \sim B$, then also $B \sim A$, i.e. A can be obtained from B by elementary row operations. In Example 10, we can use matrix notation for solving the linear system:

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

Each row operation creates two equivalent linear systems, which results in the following proposition.

Proposition 2.8 *Let $A, B \in \mathbb{R}^{m\times n}$ and $b, c \in \mathbb{R}^m$. If $(A \mid b) \sim (B \mid c)$ then the linear systems $Ax = b$ and $Bx = c$ have the same solutions.*

The final result after applying row operations which helps us directly obtain the solution of a linear system is defined as follows.

Definition 2.9 A matrix $A = (a_{ij}) \in \mathbb{R}^{m\times n}$ is on **row-reduced echelon form** if

(i) The first non-zero element on each row (if any) is equal to 1.
(ii) If there is a leading 1 in a row, then all rows above contain a leading 1 further to the left.
(iii) If a_{ij} is the first non-zero element in row i, then there are no other non-zero elements in the j-th column.

The first non-zero element in a row of a matrix in row-reduced echelon form is called **pivot element**.

In Example 10, we obtain the row-reduced echelon form as follows:

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

Is the row-reduced echelon form unique for every matrix? The following theorem will help us answer this question.

Theorem 2.10 *Every matrix A is row equivalent to a unique matrix B on row-reduced echelon form and we write*

$$B = \mathrm{rref}(A)\,.$$

Proof. We prove this theorem by induction on the number of columns of A (see Chapter 17 or `https://en.wikipedia.org/wiki/Mathematical_induction` if you are not familiar with this concept). If A has only one column, then there are only two different row-reduced echelon forms:

$$\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = e_1 \quad \text{or} \quad \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{0}.$$

Indeed, A cannot be row equivalent to both of these forms because there is no sequence of row operations leading one form to another (the zero vector is only row equivalent to itself). Hence, we already proved the base step (1 column).

Now, let $n > 1$ and assume that every matrix with $n - 1$ columns is equivalent to a unique matrix on row-reduced echelon form (RREF). Furthermore, assume A has m rows and n columns. Suppose that A is row equivalent to B and C, which are both on RREF. Let A_1, B_1, and C_1 be matrices formed by the first $n-1$ columns of A, B, and C respectively. Since B_1 and C_1 are both row equivalent to A_1 and on RREF, they are equal by the induction hypothesis. Suppose for a contradiction that $B \neq C$. Then, there exists an index j such that $b_{jn} \neq c_{jn}$.

From Proposition 2.8, the linear systems $Ax = \mathbf{0}$, $Bx = \mathbf{0}$, and $Cx = \mathbf{0}$ all have the same solutions. Let $v \in \mathbb{R}^n$ be any solution to $Ax = \mathbf{0}$. Then, we have $Bv = \mathbf{0}$ and $Cv = \mathbf{0}$, and therefore $(B - C)v = \mathbf{0}$.

Since the first $n-1$ columns of $B-C$ are zeroes (due to $B_1 = C_1$), we get the j^{th} row of the system: $(b_{jn} - c_{jn})v_n = 0$, so $v_n = 0$. Hence, every solution to $Ax = \mathbf{0}$ has zero at the last entry.

The matrix B_1 on RREF has some nonzero rows and then some zero rows. Say there are k zero rows. We can then write B_1 in the form

$$\begin{pmatrix} D \\ \mathbf{0}_{k\times(n-1)} \end{pmatrix},$$

where $D \in \mathbb{R}^{(m-k)\times(n-1)}$ with no zero rows and $\mathbf{0}_{k\times(n-1)}$ is the $k \times (n-1)$ zero matrix. Then, B and C have the form

$$B = \begin{pmatrix} D & b \\ \mathbf{0}_{k\times(n-1)} & t \end{pmatrix}, \qquad C = \begin{pmatrix} D & c \\ \mathbf{0}_{k\times(n-1)} & u \end{pmatrix},$$

where $b, c \in \mathbb{R}^{m-k}$ and $t, u \in \mathbb{R}^k$. Suppose that $t = \mathbf{0}$. The first $m-k$ rows of B all have left-most elements equal to 1. For $1 \le i \le m-k$, let such element in row i of B occur in column c_i. Also, let $b = \begin{pmatrix} b_1 \\ \vdots \\ b_{m-k} \end{pmatrix}$. Then, the linear system $Bx = \mathbf{0}$ has a solution with the c_i^{th} element equal to b_i, the last element equal to -1, and zeroes elsewhere. This contradicts to that every solution to $Ax = \mathbf{0}$ has zero at the last entry. Thus, $t \neq \mathbf{0}$.

Since B is on RREF, $t = e_1$ and $b = \mathbf{0}$. The same argument applies to C, so $u = e_1$ and $c = \mathbf{0}$. Hence, $B = C$, which is a contradiction to the assumption that $B \neq C$. Therefore, $B = C$, which completes the induction step. $\square$

For a general matrix $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ the row-reduced echelon form $(B \mid c)$ of the augmented matrix $(A \mid b)$ has the following shape

$$\operatorname{rref}(A \mid b) = (B \mid c) = \; m \updownarrow \underbrace{\left(\begin{array}{cccccccccccccccccc|c} 0 & \cdots & 0 & 1 & * & \cdots & * & 0 & * & \cdots & * & 0 & * & \cdots & * & 0 & & & * \\ & & & & & & & 1 & * & \cdots & * & 0 & * & \cdots & * & \vdots & & & \vdots \\ & & & & & & & & & & & 1 & * & \cdots & * & \vdots & & & \vdots \\ & & & & & & & & & & & & & & \ddots & \vdots & & & \vdots \\ & & & & & & & & & & & & & & & 0 & & & \vdots \\ & & & & & & & & & & & & & & & 1 & * & \cdots & * \\ & & & & & & & & & & & & & & & & & & \lambda \\ & & & & & & & & & & & & & & & & & & \end{array}\right)}_{\longleftarrow\; n \;\longrightarrow} \begin{array}{l} \left.\begin{array}{c} \\ \\ \\ \\ \\ \\ \end{array}\right\} c' \\ \\ \\ \end{array}$$

$\lambda \in \{0, 1\}$

We can read off solutions for linear system $Ax = b$ after finding $\operatorname{rref}(A \mid b)$ as follows:

$$(A \mid b) \sim (B \mid c) = \operatorname{rref}(A \mid b), \text{ for } A, B \in \mathbb{R}^{m\times n}, \text{ and } b, c \in \mathbb{R}^m$$

1) If the last column contains a pivot element $(\lambda = 1)$: **No solutions** (since $0 \neq 1$).
   Else $(\lambda = 0)$:

2) If every column of B contains a pivot element then we have the **unique solution** $x = c'$.

3) Some columns of B do not contain pivot elements: **Infinitely many solutions**.

From the above discussion, we see that the number of pivot elements and their location are important, which leads us to introduce the following definition.

> **Definition 2.11** Let $A \in \mathbb{R}^{m\times n}$ be a matrix. The **rank** $\operatorname{rk}(A)$ of A is the number of pivot elements in $\operatorname{rref}(A)$.

Following the discussion of three situations arising when we analyze the row-reduced echelon form of a matrix, we summarize them by using the notion of rank in the following proposition.

Proposition 2.12 *Let $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. The solution of $Ax = b$ depend on $\mathrm{rk}(A \mid b)$ and $\mathrm{rk}(A)$ as follows:*

(i) If $\mathrm{rk}(A \mid b) > \mathrm{rk}(A)$ then $Ax = b$ has ***no solutions****.*

(ii) If $\mathrm{rk}(A \mid b) = \mathrm{rk}(A) = n$ then $Ax = b$ has a ***unique solution****.*

(iii) If $\mathrm{rk}(A \mid b) = \mathrm{rk}(A) < n$ then $Ax = b$ has ***infinitely many solutions****.*

Proof. (i) If $\mathrm{rk}(A \mid b) > \mathrm{rk}(A)$, then the last column of $\mathrm{rref}(A \mid b)$ contains a pivot element. Hence, the linear system $Ax = b$ has no solution.

(ii) If $\mathrm{rk}(A \mid b) = \mathrm{rk}(A) = n$, then the last column of $\mathrm{rref}(A \mid b)$ does not contain a pivot element and every column of $\mathrm{rref}(A)$ contains a pivot element. Thus, the linear system $Ax = b$ has a unique solution given by $x = c'$.

(iii) If $\mathrm{rk}(A \mid b) = \mathrm{rk}(A) < n$, then the last column of $\mathrm{rref}(A \mid b)$ does not contain a pivot element but some columns of $\mathrm{rref}(A)$ do not contain pivot elements. Therefore, the linear system $Ax = b$ has infinitely many solutions.

□

Exercises

Exercise 7. Show that for all $A \in \mathbb{R}^{m\times n}$, $x, y \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$ we have

(i) $A(x + y) = Ax + Ay$,

(ii) $A(\lambda x) = \lambda(Ax)$.

(Without using Proposition 2.4).

Exercise 8. We define the following matrices and vectors:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad C = \begin{pmatrix} -1 & 0 \\ 2 & 4 \\ 0 & -3 \end{pmatrix}, \quad D = \begin{pmatrix} 0 & 8 & 0 \\ 1 & 2 & -1 \end{pmatrix}, \quad E = \begin{pmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{pmatrix},$$

$$t = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad u = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}, \quad v = \begin{pmatrix} 3 \\ -4 \end{pmatrix}, \quad w = \begin{pmatrix} -1 \\ 2 \\ 2 \end{pmatrix}.$$

(i) Decide which of the following expressions are defined. Evaluate them if possible.

$$At, \quad Au, \quad wA, \quad 2A, \quad A + B, \quad A + C, \quad A + D, \quad \frac{3}{4}Bt, \quad Bu, \quad B + B, \quad Dw,$$
$$Cv, \quad t + u, \quad tu, \quad -v, \quad u + w, \quad t - u, \quad \frac{1}{2}w, \quad C + w, \quad Et, \quad Ev \quad E\,(Ev)\,.$$

(ii) Draw the following vectors in $\mathbb{R}^2$

$$t, \quad ,v, \quad -2t, \quad t - \frac{1}{2}v, \quad v + t, \quad t + v, \quad Et, \quad Ev, \quad E\,(Ev)\,, \quad Bt, \quad Bv\,.$$

Can you guess what happens in general to a vector in $\mathbb{R}^2$ when you multiply it with B or E? Try to give a geometric interpretation. (without a proof)

Exercise 9. Let $a, b, c, d \in \mathbb{R}$ and $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$.

(i) Show that $\mathrm{rk}(A) = 2$ if and only if $ad - bc \neq 0$.

(ii) We define the following subset of $\mathbb{R}^2$

$$L = \{x \in \mathbb{R}^2 \mid x = Av \text{ for some } v \in \mathbb{R}^2\}.$$

How does L look like if $\operatorname{rk}(A) = 1$? How does it look like if $\operatorname{rk}(A) = 2$?

Exercise 10. Give examples of matrices $A, B, C \in \mathbb{R}^{3\times 3}$, which are all not on row-reduced echelon form, such that $\operatorname{rk}(A) = 1$, $\operatorname{rk}(B) = 2$, $\operatorname{rk}(C) = 3$.

Exercise 11. Let $A \in \mathbb{R}^{3\times 3}$ be a matrix.

(i) Show that if $\operatorname{rk}(A) = 3$ then there exists just one vector $x \in \mathbb{R}^3$ with $Ax = 0$.

(ii) Show that if $\operatorname{rk}(A) \leq 2$ then there exist infinitely many vectors $x \in \mathbb{R}^3$ with $Ax = 0$.

Exercise 12. Let $p(x) = a_0 + a_1x + a_2x^2 + a_3x^3$ be a polynomial of degree 3 with real coefficients $a_0, a_1, a_2, a_3 \in \mathbb{R}$. For this polynomial p we define the vector v_p by

$$v_p = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} \in \mathbb{R}^4 .$$

Find a matrix $D \in \mathbb{R}^{4\times 4}$, such that $v_{p'} = Dv_p$, where p' denotes the derivative of the polynomial p with respect to x. What is the rank of D?

3

Sets & Functions

A **set** is a collection of distinct objects, grouped together as a single entity. It is precisely, but not necessarily explicitly, defined. The objects that belong to a set are called its elements. If a set has finitely many elements we call it a finite set and otherwise infinite set. We have already seen examples of infinite sets: $\mathbb{R}$, $\mathbb{R}^n$, $\mathbb{R}^{m\times n}$.

Example 11
1) $\{2, 4, \pi\}$ is a finite set.
2) $\mathbb{N} = \{1, 2, 3, 4, \ldots\}$ is the set of natural numbers.
3) $\mathbb{Q}$ is the set of rational numbers.
4) $\emptyset = \{\}$ is the empty set, which has no element.

Given a set A, we write "$a \in A$" if a is an element of A and "$a \notin A$" if a is not an element of A.
A set A is a subset of another set B when every element of A belongs to B. That is, if $a \in A$, then $a \in B$. In this case, we write $A \subset B$. The empty set $\emptyset$ is subset of any other set.

Example 12
1) $2 \in \mathbb{N}$, $\frac{1}{2} \notin \mathbb{N}$, $\pi \in \mathbb{R}$, $\pi \notin \mathbb{Q}$.
2) $\emptyset \subset \mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$.
3) $\{1, 2, 3\} \subset \mathbb{N}$.

From a set A, we can define a subset of A that contains all elements of A satisfying a condition and we write it in the format $\{a \in A \mid \text{condition}\}$.

Example 13
1) $\{m \in \mathbb{N} \mid m \text{ is even}\}$ is the set of all even numbers.
2) Let H be the set of all humans. $\mathrm{NU} = \{h \in \mathrm{H} \mid h \text{ is a student at Nagoya University}\} \subset \mathrm{H}$.
3) $\{x \in \mathbb{R}^n \mid Ax = b\}$ is the set of all solutions of $Ax = b$.

We can create new sets from given sets by operations on sets. Given two sets A and B, we define the following operations (that produce new sets):

(i) **Union:** $A \cup B$ is the set of all objects that are elements of A or B or both.

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}$$

(ii) **Intersection:** $A \cap B$ is the set of all objects that are elements of both A and B.

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}$$

(iii) **Difference:** $A \setminus B$ is the set of all objects that are elements of A but not B.

$$A \setminus B = \{x \in A \mid x \notin B\}$$

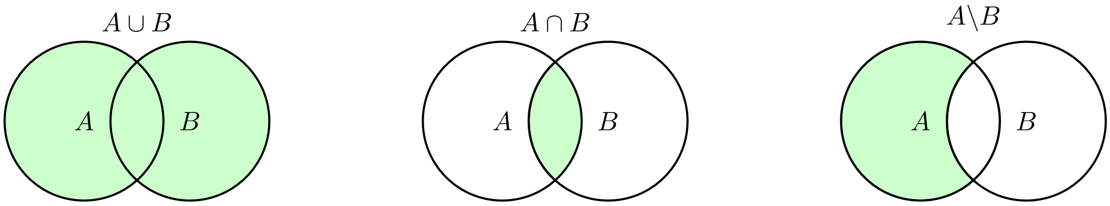


Figure 3.1: Visualization of the union, intersection and difference of two sets A and B.

Example 14 For the sets $A = \{-1, 2, 3\}$ and $B = \mathbb{N} = \{1, 2, 3, \dots\}$, we have
1) $A \cup B = \{-1, 1, 2, 3, \dots\}$,
2) $A \cap B = \{2, 3\}$,
3) $A \setminus B = \{-1\}$.

In order to introduce later the notion of **linear maps** in Chapter 4, which will be an important concept in linear algebra, now we want to examine the notion of **function** in general.

Definition 3.1 Let X and Y be two sets.
(i) A **function** $f : X \to Y$ is a rule, assigning to each element $x \in X$ an element $f(x) \in Y$. This is also denoted by

$$\begin{aligned} f : X &\longrightarrow Y \\ x &\longmapsto f(x) \,. \end{aligned}$$

(ii) For $f : X \to Y$, the set X is called the **domain of** f and Y is called the **codomain of** f.

A function is also sometimes called a **map**. These two names (for the exact same mathematical object) are used interchangeably in literature.

Definition 3.2 For a function $f : X \to Y$, the **image of** f is defined by

$$\mathrm{im}(f) = \{y \in Y \mid \exists x \in X : y = f(x)\} \,.$$

Another notation is $\mathrm{im}(f) = f(X)$. The image is a subset of the codomain, i.e., $\mathrm{im}(f) \subset Y$.

A fundamental operation for functions that allows us to create a new function from one function or several functions is as follows.

Definition 3.3 **Composition of functions:** For two functions $f : X \to Y$ and $g : Y \to Z$, the composition $g \circ f$ of f and g is defined by

$$\begin{aligned} g \circ f : X &\longrightarrow Z \\ x &\longmapsto (g \circ f)(x) = g(f(x)) \,. \end{aligned}$$

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[rr, bend right, "g \circ f"'] & Y \arrow[r, "g"] & Z \end{tikzcd}$$

The following notions are characteristics of certain special types of functions.

Definition 3.4 A function $f : X \to Y$ is called
(i) **injective** if $f(x_1) \neq f(x_2)$ whenever $x_1 \neq x_2$. $(x_1, x_2 \in X)$
(ii) **surjective** if $\mathrm{im}(f) = Y$.
(iii) **bijective** if it is both injective and surjective.

Given a set X, we define the **identity function** on X as follows:

$$\begin{aligned} \mathrm{id}_X : X &\longrightarrow X, \\ x &\longmapsto x\,. \end{aligned}$$

If $f : X \to Y$ is bijective or, equivalently, for *every* $y \in Y$ there exists a *unique* $x \in X$ with $f(x) = y$, then we can define $g : Y \to X$ by $g(y) = x$ for $f(x) = y$. In that case, we have

$$\begin{aligned} (g \circ f)(x) &= g(f(x)) = g(y) = x \quad \forall x \in X \quad \Rightarrow \quad g \circ f = \mathrm{id}_X, \\ (f \circ g)(y) &= f(g(y)) = f(x) = y \quad \forall y \in Y \quad \Rightarrow \quad f \circ g = \mathrm{id}_Y\,. \end{aligned}$$

For any function f, when there exists a function g which satisfies the above two conditions, we say that f is **invertible** and g is the **inverse** of f. Usually, we denote the inverse as $g = f^{-1}$. Hence, a bijective function is also invertible and vice versa.

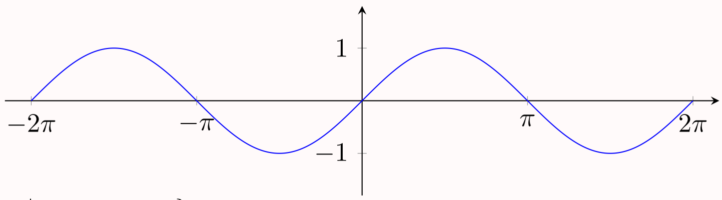
Example 15
1) Let H be the set of all humans. $\mathrm{NU} = \{h \in \mathrm{H} \mid h \text{ is a student at Nagoya University}\} \subset \mathrm{H}$.

$$\begin{aligned} f_1 : \mathrm{NU} &\longrightarrow \mathbb{N}, \\ s &\longmapsto \text{ Student ID of student } s. \end{aligned}$$

f_1 is injective because there are no two students with the same ID.
f_1 is not surjective because not every natural number is the student ID of a student.

2) Now consider the sine function with domain and codomain $\mathbb{R}$:

$$\begin{aligned} f_2 : \mathbb{R} &\longrightarrow \mathbb{R}, \\ x &\longmapsto \sin(x)\,. \end{aligned}$$

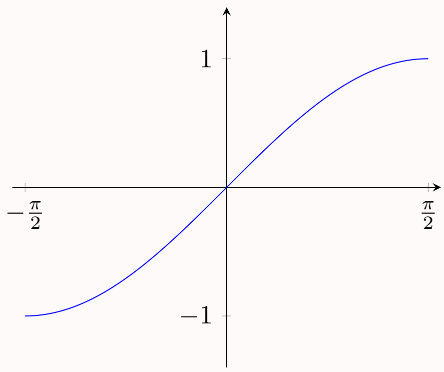


$\mathrm{im}(f_2) = [-1, 1] = \{x \in \mathbb{R} \mid -1 \leq x \leq 1\}$.
f_2 is not surjective because $2 \in \mathbb{R}$ but $2 \notin \mathrm{im}(f_2)$.
f_2 is not injective because $f_2(0) = f_2(2\pi) = 0$ but $0 \neq 2\pi$.

3) If we consider the sine function where we restrict the domain and codomain the situation changes:

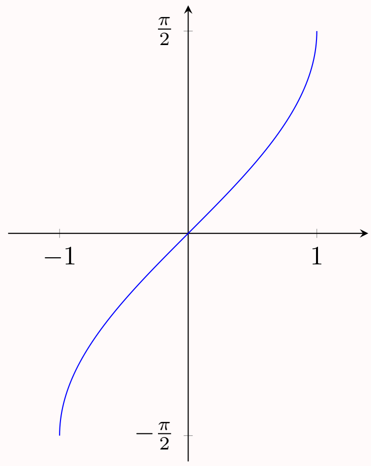
$$\begin{aligned} f_3 : \left[-\frac{\pi}{2}, \frac{\pi}{2}q\right] &\longrightarrow [-1, 1], \\ x &\longmapsto \sin(x). \end{aligned}$$

$\mathrm{im}(f_3) = [-1, 1] \;\Rightarrow\; f_3$ is surjective.
f_3 is also injective because, for each $y \in [-1, 1]$, there exists *exactly one* $x \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ with $f_3(x) = y$.

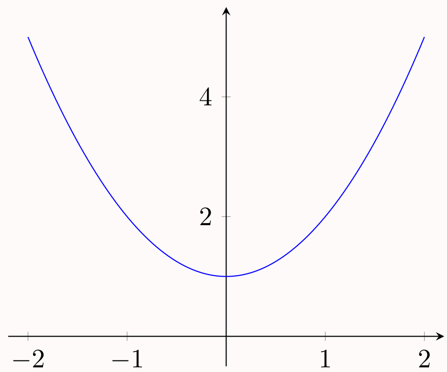


Hence, f_3 is bijective with the inverse

$$f_3^{-1} = \arcsin : [-1, 1] \longrightarrow \left[-\frac{\pi}{2}, \frac{\pi}{2}\right].$$



4) Consider the following function:



$$\begin{aligned} f_4 : \mathbb{R} &\longrightarrow \mathbb{R}. \\ x &\longmapsto x^2 + 1, \end{aligned}$$

$\mathrm{im}(f_4) = [1, \infty) = \{x \in \mathbb{R} \mid x \geq 1\} \neq \mathbb{R}$.
$f_4(-x) = x^2 = f_4(x) \quad \forall x \in \mathbb{R}$.
Hence, f_4 is neither injective nor surjective.

5) Consider the following function:

$$\begin{aligned} f_5 : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2, \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &\longmapsto \begin{pmatrix} x_1 + 2x_2 \\ 2x_1 + 4x_2 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = Ax, \end{aligned}$$

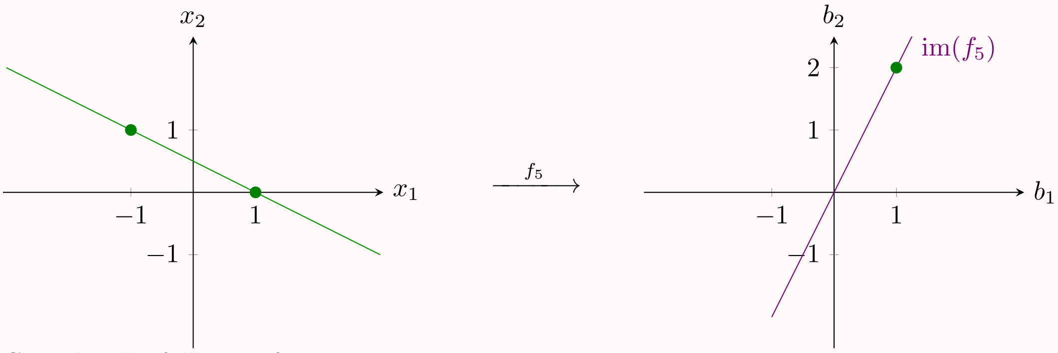
where $A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$ and $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$. To calculate the image of f_5 we need to find all $b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \in \mathbb{R}^2$ such that there exists a $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$ with $f_5(x) = Ax = b$, which means that for each $b \in \mathbb{R}^2$ we need to solve the linear system $Ax = b$.

$$(A \mid b) = \left(\begin{array}{cc|c} 1 & 2 & b_1 \\ 2 & 4 & b_2 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 2 & b_1 \\ 0 & 0 & b_2 - 2b_1 \end{array}\right).$$

Hence, $Ax = b$ has solutions when $b_2 = 2b_1$. Thus, $\mathrm{im}(f_5) = \left\{ \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \in \mathbb{R}^2 \,\middle|\, b_2 = 2b_1 \right\} \neq \mathbb{R}^2$.
Therefore, f_5 is not surjective.
If $b_2 = 2b_1$, we have the solution $\begin{cases} x_1 = b_1 - 2t \\ x_2 = t \end{cases}$, for $t \in \mathbb{R}$.
For $b_1 = 1$:

$$\begin{cases} x_1 = 1 - 2t \\ x_2 = t \end{cases} \quad \text{and } f_5 \begin{pmatrix} 1 - 2t \\ t \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad \forall t \in \mathbb{R}.$$

Hence, f_5 is not injective since $f_5 \begin{pmatrix} -1 \\ 1 \end{pmatrix} = f_5 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ but $\begin{pmatrix} -1 \\ 1 \end{pmatrix} \neq \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.



6) Consider the following function:

$$\begin{aligned} f_6 : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2, \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &\longmapsto \begin{pmatrix} x_1 + 2x_2 \\ 3x_1 + 4x_2 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = Ax, \end{aligned}$$

where $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ and $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$. In order to find

$$\mathrm{im}(f_6) = \{y \in \mathbb{R}^2 \mid \exists x \in \mathbb{R}^2 : f_6(x) = y\} = \{y \in \mathbb{R}^2 \mid Ax = y \text{ has a solution } x \in \mathbb{R}^2\},$$

we need to understand the solutions of $Ax = y$.

$$\begin{aligned} (A \mid y) = & \left(\begin{array}{cc|c} 1 & 2 & y_1 \\ 3 & 4 & y_2 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 2 & y_1 \\ 0 & -2 & y_2 - 3y_1 \end{array}\right) \\ \sim & \left(\begin{array}{cc|c} 1 & 2 & y_1 \\ 0 & 1 & \frac{3}{2}y_1 - \frac{1}{2}y_2 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & -2y_1 + y_2 \\ 0 & 1 & \frac{3}{2}y_1 - \frac{1}{2}y_2 \end{array}\right) \end{aligned}$$

This shows that the linear system $Ax = y$ has a unique solution

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -2y_1 + y_2 \\ \frac{3}{2}y_1 - \frac{1}{2}y_2 \end{pmatrix}$$

for every $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \in \mathbb{R}^2$. Hence, $\mathrm{im}(f_6) = \mathbb{R}^2$ and f_6 is bijective. In addition, the inverse $f_6^{-1} : \mathbb{R}^2 \to \mathbb{R}^2$ of f_6 is defined as follows:

$$f_6^{-1} = \begin{pmatrix} -2x_1 + x_2 \\ \frac{3}{2}x_1 - \frac{1}{2}x_2 \end{pmatrix} = \begin{pmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = Bx,$$

where $B = \begin{pmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix}$.

Later we will see that $B = A^{-1}$ is the inverse of A (Chapter 7).

7) Consider the following function:

$$\begin{aligned} f_7 : \mathbb{R}^2 &\longrightarrow \mathbb{R}^3 \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &\longmapsto \begin{pmatrix} x_1 \\ x_1 + x_2 \\ x_2 - x_1 \end{pmatrix} \end{aligned}$$

We want to calculate $\mathrm{im}(f_7) = \{y \in \mathbb{R}^3 \mid \exists x \in \mathbb{R}^2 : f_7(x) = y\}$. For this, we first rewrite

$$f_7(x) = f_7 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_1 + x_2 \\ x_2 - x_1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = Ax,$$

where $A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ -1 & 1 \end{pmatrix}$. Hence, $\mathrm{im}(f_7) = \{y \in \mathbb{R}^2 \mid Ax = y \text{ has a solution } x \in \mathbb{R}^2\}$. Therefore, we want to understand the solutions of $Ax = y$.

$$(A \mid y) = \left(\begin{array}{cc|c} 1 & 0 & y_1 \\ 1 & 1 & y_2 \\ -1 & 1 & y_3 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & y_1 \\ 0 & 1 & y_2 - y_1 \\ 0 & 1 & y_3 + y_1 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & y_1 \\ 0 & 1 & y_2 - y_1 \\ 0 & 0 & 2y_1 - y_2 + y_3 \end{array}\right)$$

This shows that $Ax = y$ has a solution if and only if $2y_1 - y_2 + y_3 = 0$.
Thus, $\mathrm{im}(f_7) = \left\{ \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \in \mathbb{R}^3 \;\middle|\; 2y_1 - y_2 + y_3 = 0 \right\}$ and f_7 is not surjective since $\mathrm{im}(f_7) \neq \mathbb{R}^3$.
For $y \in \mathrm{im}(f_7)$, the system $Ax = y$ has a unique solution because there are no free variables. Therefore, f_7 is injective (but not surjective).

Exercises

Exercise 13. Let X be a finite set. Show that a function $f : X \to X$ is injective if and only if it is surjective.

Exercise 14. Let $\mathbb{N}_0 = \{0, 1, 2, 3, \dots\}$ and $\mathbb{Z} = \{0, 1, -1, 2, -2, \dots\}$ denote the set of natural numbers (together with zero) and the integers. Decide if the following function is injective and/or surjective:

$$\begin{aligned} g : \mathbb{N}_0 &\longrightarrow \mathbb{Z} \\ n &\longmapsto \begin{cases} \frac{n}{2} & \text{if } n \text{ is even} \\ -\frac{n+1}{2} & \text{if } n \text{ is odd} \end{cases} . \end{aligned}$$

Exercise 15. Which of the following functions are injective and/or surjective?

$$\begin{aligned} f_1 : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto e^x , \end{aligned} \qquad \begin{aligned} f_2 : \mathbb{R}^2 &\longrightarrow \mathbb{R}^3 \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &\longmapsto \begin{pmatrix} x_1 + 2x_2 \\ 2x_1 + 4x_2 \\ x_1 - x_2 \end{pmatrix} , \end{aligned}$$

$$\begin{aligned} f_3 : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto 3x + 2 , \end{aligned} \qquad \begin{aligned} f_4 : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &\longmapsto \begin{pmatrix} x_1 + 2x_2 \\ x_1 x_2 \end{pmatrix} . \end{aligned}$$

4

Linear maps

In this chapter, we discuss linear maps, which serve as a fundamental concept in linear algebra. Up until now, we have explored the properties of vectors in $\mathbb{R}^n$ and their various operations. Linear maps, also known as linear transformations, are special functions that map one vector space to another while preserving the underlying structure and operations (vector addition and scalar multiplication) of the original space. As we have not yet formally introduced the notion of vector spaces (Chapter 15), it is essential to understand that these abstract mathematical structures provide a broader and more general framework for linear maps.

Let's consider a simple example of a linear map using matrix multiplication. Suppose we have an arbitrary $m \times n$ matrix A and a vector $x \in \mathbb{R}^n$. The linear map F represented by the matrix A can be defined as the function $F(x) = Ax$, which maps a vector x in $\mathbb{R}^n$ to a vector in $\mathbb{R}^m$. This linear map satisfies by Proposition 2.4 the property $F(x+y) = A(x+y) = Ax+Ay = F(x)+F(y)$ for any $x, y \in \mathbb{R}^n$ and $F(\lambda x) = A(\lambda x) = \lambda Ax = \lambda F(x)$ for $x \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$. This property will be the definition of a linear map and we will see, that all linear maps indeed come from a matrix in the above way.

As we explore the properties and applications of linear maps, you will see that such transformations play a significant role in various areas of mathematics and real-world applications, including computer graphics, data analysis, and engineering.

Definition 4.1 A function $F : \mathbb{R}^n \to \mathbb{R}^m$ is a **linear map** if for all $u, v \in \mathbb{R}^n$, $\lambda \in \mathbb{R}$ we have

(i) $F(u + v) = F(u) + F(v)$,
(ii) $F(\lambda u) = \lambda F(u)$.

Example 16

1) For any matrix $A \in \mathbb{R}^{m\times n}$, the function

$$\begin{aligned} F : \mathbb{R}^n &\longrightarrow \mathbb{R}^m \\ x &\longmapsto Ax \end{aligned}$$

is a linear map. This follows from Proposition 2.4 as follows: for any $u, v \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$

(i) $F(u + v) = A(u + v) = Au + Av = F(u) + F(v)$,
(ii) $F(\lambda u) = A(\lambda u) = \lambda(Au) = \lambda F(u)$.

Special case: When $n = m$ and $A = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} \overset{\text{def}}{=} I_n$, called identity matrix, we have $F(x) = x, \quad \forall x \in \mathbb{R}^n$. In this case, $F = \mathrm{id}_{\mathbb{R}^n}$.

2) The function

$$f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$$
$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \longmapsto \begin{pmatrix} x_1 x_2 \\ x_1 \end{pmatrix}$$

is not a linear map. For $\lambda = 2$ and $u = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$,

$$f(\lambda u) = f\left(2\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right) = f\begin{pmatrix} 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \cdot 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \end{pmatrix},$$
$$\lambda f(u) = 2f\begin{pmatrix} 1 \\ 1 \end{pmatrix} = 2\begin{pmatrix} 1 \cdot 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}.$$

Therefore, $f(\lambda u) \neq \lambda f(u)$ for the case $\lambda = 2$.

In fact we will see now that any linear map is given by a function like in Example 16 1).

Theorem 4.2 *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map. Then there exists a unique matrix $[F] \in \mathbb{R}^{m\times n}$, such that for all $x \in \mathbb{R}^n$ we have*

$$F(x) = [F]x\,.$$

Here the left-hand side is the evaluation of the function F at x and the right-hand side is the multiplication of the matrix $[F]$ with the vector x.

Proof. For $1 \leq j \leq n$, we consider vectors $e_j \in \mathbb{R}^n$ such that

$$e_j = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{matrix} \\ \\ \\ \leftarrow \text{the j}^{\text{th}} \text{ entry} \\ \\ \\ \\ \end{matrix}$$

Every $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ can be uniquely written as $x = x_1 e_1 + x_2 e_2 + \ldots + x_n e_n$.

Since F is linear, we have

$$\begin{aligned} F(x) &= F(x_1e_1 + x_2e_2 + \ldots + x_ne_n) \\ &\overset{\text{(i)}}{=} F(x_1e_1) + F(x_2e_2 + \ldots + x_ne_n) = \ldots = F(x_1e_1) + F(x_2e_2) + \ldots + F(x_ne_n) \\ &\overset{\text{(ii)}}{=} x_1F(e_1) + x_2F(e_2) + \ldots + x_nF(e_n) \end{aligned}$$

Now set $[F] = \begin{pmatrix} | & | & & | \\ F(e_1) & F(e_2) & \cdots & F(e_n) \\ | & | & & | \end{pmatrix} \in \mathbb{R}^{m\times n}$. With this, we have

$$[F]x = [F]\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1F(e_1) + x_2F(e_2) + \ldots + x_nF(e_n),$$

And therefore, $F(x) = [F]x$. $\square$

Definition 4.3 The matrix $[F]$ in Theorem 4.2 is called **the matrix of F**.

Remark. If F is a linear map and we know the values $F(e_j)$ $(1 \leq j \leq n)$, then we know the value of $F(x)$ for any x.

Example 17

1) If $F : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear map with $F\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}1\\2\end{pmatrix}$ and $F\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}3\\4\end{pmatrix}$, then

$$F\begin{pmatrix}-1\\3\end{pmatrix} = F\left(-1\begin{pmatrix}1\\0\end{pmatrix} + 3\begin{pmatrix}0\\1\end{pmatrix}\right) = -F\begin{pmatrix}1\\0\end{pmatrix} + 3F\begin{pmatrix}0\\1\end{pmatrix} = -\begin{pmatrix}1\\2\end{pmatrix} + 3\begin{pmatrix}3\\4\end{pmatrix} = \begin{pmatrix}8\\10\end{pmatrix}.$$

In general,

$$F\begin{pmatrix}x_1\\x_2\end{pmatrix} = F\left(x_1\begin{pmatrix}1\\0\end{pmatrix} + x_2\begin{pmatrix}0\\1\end{pmatrix}\right) = x_1F\begin{pmatrix}1\\0\end{pmatrix} + x_2F\begin{pmatrix}0\\1\end{pmatrix} = x_1\begin{pmatrix}1\\2\end{pmatrix} + x_2\begin{pmatrix}3\\4\end{pmatrix} = \begin{pmatrix}1 & 3\\2 & 4\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}.$$

The matrix of F is $[F] = \begin{pmatrix}1 & 3\\2 & 4\end{pmatrix}$.

2) This works in more general cases. Assume that $F : \mathbb{R}^2 \to \mathbb{R}^3$ is a linear map with

$$F\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}1\\1\\-1\end{pmatrix}, \qquad F\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}0\\1\\2\end{pmatrix}.$$

What is $F\begin{pmatrix}x_1\\x_2\end{pmatrix}$ for any x_1, x_2? We have

$$\begin{pmatrix}x_1\\x_2\end{pmatrix} = x_2\begin{pmatrix}1\\1\end{pmatrix} + (x_1 - x_2)\begin{pmatrix}1\\0\end{pmatrix}.$$

Hence,

$$\begin{aligned} F\begin{pmatrix}x_1\\x_2\end{pmatrix} &= F\left(x_2\begin{pmatrix}1\\1\end{pmatrix} + (x_1 - x_2)\begin{pmatrix}1\\0\end{pmatrix}\right) = x_2F\begin{pmatrix}1\\1\end{pmatrix} + (x_1 - x_2)F\begin{pmatrix}1\\0\end{pmatrix} \\ &= x_2\begin{pmatrix}0\\1\\2\end{pmatrix} + (x_1 - x_2)\begin{pmatrix}1\\1\\-1\end{pmatrix} = \begin{pmatrix}x_1 - x_2\\x_1\\-x_1 + 3x_2\end{pmatrix} = \begin{pmatrix}1 & -1\\1 & 0\\-1 & 3\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}.\end{aligned}$$

The matrix of F is $[F] = \begin{pmatrix}1 & -1\\1 & 0\\-1 & 3\end{pmatrix}$.

Remark. From the above examples, we see that, in order to know the value of a linear map F at any x, it suffices to know the value $F(v_1)$, …, $F(v_n)$ where v_1, …, v_n are vectors such that we can write for any x as

$$x = \alpha_1v_1 + \ldots + \alpha_nv_n,$$

for some $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$.

How to check if a given function $F : \mathbb{R}^n \to \mathbb{R}^m$ is linear or not?

- To show that F is linear, one can either show that:
  - There exists a matrix $A \in \mathbb{R}^{m\times n}$ with $F(x) = Ax$ (and hence, $A = [F]$), **or**

  - **For all** $u, v \in \mathbb{R}^n$, $\lambda \in \mathbb{R}$ that $F(u+v) = F(u) + F(v)$ and $F(\lambda u) = \lambda F(u)$.

- To show that F is not linear, it suffices to give **one** example of $u, v \in \mathbb{R}^n$ with $F(u+v) \neq F(u) + F(v)$, or **one** example of $u \in \mathbb{R}^n$, $\lambda \in \mathbb{R}$ with $F(\lambda u) \neq \lambda F(u)$.

Example 18 The function

$$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto e^x \end{aligned}$$

is not linear because the **one explicit example** with $u = v = 0$ gives:

$$\begin{aligned} f(u+v) &= f(0+0) = f(0) = e^0 = 1, \\ f(u) + f(v) &= f(0) + f(0) = e^0 + e^0 = 2. \end{aligned}$$

Therefore, $f(u+v) \neq f(u) + f(v)$ for the case $u = v = 0$.

What you should not do, when proving that this function is not linear, is to write "It is $e^{u+v} \neq e^u + e^v$ and therefore f is not linear". Even though $e^{u+v} \neq e^u + e^v$ is true for almost all $u, v \in \mathbb{R}$, there are cases where it is not true. For example, for $u = 1$ and $v = 0.4586\ldots$,

$$e^{u+v} = 4.3002\ldots = e^u + e^v.$$

Exercises

Exercise 16. Which of the following functions are linear maps?

$$\begin{aligned} f_1 : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto e^x\,, \end{aligned}$$

$$\begin{aligned} f_2 : \mathbb{R}^2 &\longrightarrow \mathbb{R}^3 \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &\longmapsto \begin{pmatrix} x_1 + 2x_2 \\ 2x_1 + 4x_2 \\ x_1 - x_2 \end{pmatrix}, \end{aligned}$$

$$\begin{aligned} f_3 : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto 3x + 2\,, \end{aligned}$$

$$\begin{aligned} f_4 : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &\longmapsto \begin{pmatrix} x_1 + 2x_2 \\ x_1 x_2 \end{pmatrix}. \end{aligned}$$

Exercise 17. We define the following four functions:

$$\begin{aligned} f_1 : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &\longmapsto \begin{pmatrix} 2x_1 + x_2 \\ x_1 x_2 \end{pmatrix}, \end{aligned}$$

$$\begin{aligned} f_2 : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto \frac{2x}{x^2+4}, \end{aligned}$$

$$\begin{aligned} f_3 : \mathbb{R} &\longrightarrow \mathbb{R}^2 \\ x &\longmapsto \begin{pmatrix} 3\cos(x) \\ 2\sin(x) \end{pmatrix}, \end{aligned}$$

$$\begin{aligned} f_4 : \mathbb{R}^2 &\longrightarrow \mathbb{R}^3 \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &\longmapsto \begin{pmatrix} 2x_1 - x_2 \\ x_1 - 3x_2 \\ x_1 - x_2 \end{pmatrix}. \end{aligned}$$

(i) Calculate the image of each function, i.e. describe $\mathrm{im}(f_j)$ for $j = 1, 2, 3, 4$ as explicit as possible. If you can not find a mathematical description try to describe the elements of the image in words.
(ii) Decide for each function if it is injective and/or surjective and/or bijective.
(iii) Decide which of the above functions are linear maps.

Justify your answers in (ii) and (iii).

Exercise 18. Show that there exist a unique linear map $G : \mathbb{R}^2 \to \mathbb{R}^3$ with the property

$$G\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \qquad G\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}.$$

What is the value of $G(x)$ for an arbitrary $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$? Determine the matrix of G.

Exercise 19. Let $F : \mathbb{R}^3 \to \mathbb{R}^2$ be a linear map. Show that F can not be injective.

Exercise 20. Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map. Show that the following two statements are equivalent:

(i) F is injective.

(ii) The only solution to $F(x) = 0$ is $x = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$.

To show that both statements are equivalent, you need to show that (i) implies (ii) and (ii) implies (i).

5

Linear maps in geometry

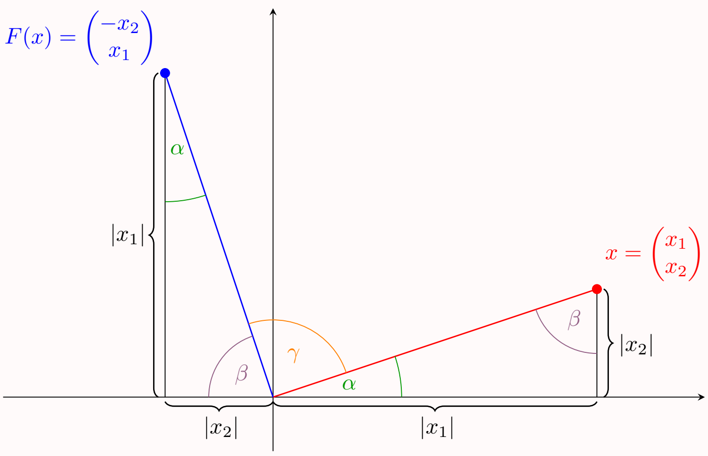
In the previous chapter, we introduced the concept of linear maps, which form the foundation of linear algebra and are essential tools for studying geometry. In this chapter, we will delve deeper into the topic of linear maps and discuss certain classes of them which have a geometric interpretation.

Example 19 Consider the function

$$F : \mathbb{R}^2 \longrightarrow \mathbb{R}^2.$$
$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \longmapsto \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix}$$

This function is indeed a linear map because

$$F\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$



We have

$$\begin{cases} \alpha + \beta + \gamma = 180^\circ \\ \alpha + \beta = 90^\circ \end{cases} \quad \Rightarrow \quad \gamma = 90^\circ.$$

Hence, F rotates x by 90° counterclockwise.

In Example 19, we say that x and $F(x)$ are "orthogonal" (meaning perpendicular) to each other. How to check if $x, y \in \mathbb{R}^2$ are orthogonal in general? What about $x, y \in \mathbb{R}^n$?

Definition 5.1 Let $u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$, $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^n$.

(i) The **dot product** of u and v is defined by

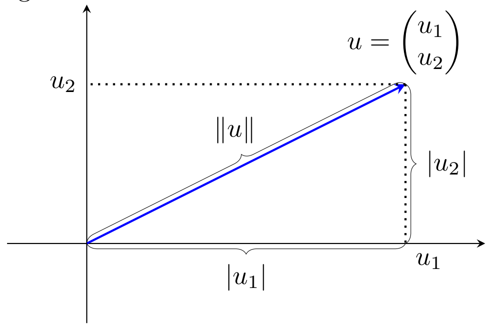
$$u \bullet v = u_1 v_1 + \cdots + u_n v_n \, .$$

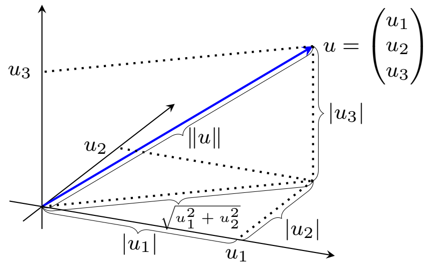
(ii) u and v are **orthogonal** (to each other) if $u \bullet v = 0$.

(iii) The **norm** of u is defined by

$$\|u\| = \sqrt{u \bullet u} = \sqrt{u_1^2 + \cdots + u_n^2} \, .$$

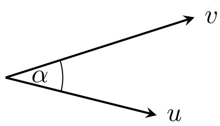
Remark. (i) The dot product (also called scalar product or inner product) allows us to speak about length and angle in $\mathbb{R}^n$. For $n = 2$ and $n = 3$, the norm of a vector is equal to its length by Pythagorean theorem:





$$\|u\| = \sqrt{u_1^2 + u_2^2} \qquad\qquad \|u\| = \sqrt{u_1^2 + u_2^2 + u_3^2}$$

Also, one can show that, for any non-zero vectors u and v,



$$u \bullet v = \|u\| \|v\| \cos(\alpha)$$
$$\Leftrightarrow \cos(\alpha) = \frac{u \bullet v}{\|u\| \|v\|} .$$

For $n > 3$, this gives the definition of an angle between u and v.

(ii) Write $u^T = \begin{pmatrix} u_1 & u_2 & \cdots & u_n \end{pmatrix} \in \mathbb{R}^{1 \times n}$ (T stands for "transpose", for which we will give a proper definition later) , then

$$u \bullet v = u^T v = \begin{pmatrix} u_1 & u_2 & \cdots & u_n \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = u_1 v_1 + \cdots + u_n v_n ,$$

where $u^T v$ is the product of the matrix u^T and the vector v.

Proposition 5.2 *The dot product satisfies the following properties for all $u, v, w \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$:*

(i) $u \bullet v = v \bullet u$,

(ii) $u \bullet (v + w) = u \bullet v + u \bullet w$,

(iii) $u \bullet (\lambda v) = \lambda(u \bullet v)$.

Proof. (i) $u \bullet v = u_1v_1 + \cdots + u_nv_n = v_1u_1 + \cdots + v_nu_n = v \bullet u$.

(ii)
$$\begin{aligned} u \bullet (v+w) &= u_1(v_1+w_1) + \cdots + u_n(v_n+w_n) \\ &= (u_1v_1 + u_1w_1) + \cdots + (u_nv_n + u_nw_n) \\ &= (u_1v_1 + \cdots + u_nv_n) + (u_1w_1 + \cdots + u_nw_n) \\ &= u \bullet v + u \bullet w. \end{aligned}$$

(iii) $u \bullet (\lambda v) = u_1(\lambda v_1) + \cdots + u_n(\lambda v_n) = \lambda(u_1v_1 + \cdots + u_nv_n) = \lambda(u \bullet v)$. □

In the following, we will give examples of linear maps which have geometric interpretations.

1. Scaling

Let $\lambda > 0$ and define the linear map

$$\begin{aligned} h_\lambda : \mathbb{R}^n &\longrightarrow \mathbb{R}^n. \\ x &\longmapsto \lambda x \end{aligned}$$

This map is indeed linear because it has the matrix

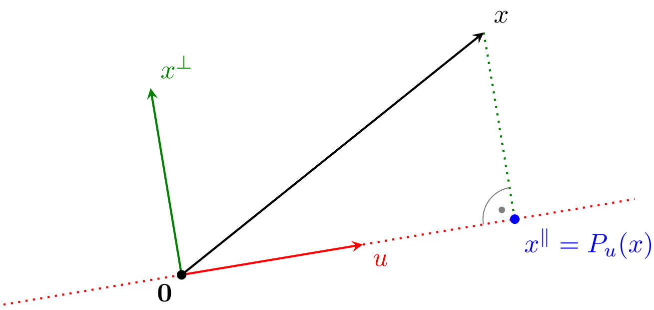
$$[h_\lambda] = \begin{pmatrix} \lambda & & 0 \\ & \ddots & \\ 0 & & \lambda \end{pmatrix} = \lambda \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} = \lambda I_n,$$

where I_n is the $n \times n$ identity matrix, the matrix of the identity map $\mathrm{id}_{\mathbb{R}^n}$. When we apply the map h_λ to a vector, we scale the norm (length) of this vector by a factor λ.

2. Orthogonal projection

Let $u \in \mathbb{R}^n$ with $u \neq \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{0}$. We want to define a map $P_u : \mathbb{R}^n \to \mathbb{R}^n$ that sends a vector $x \in \mathbb{R}^n$ to another vector $x^{\parallel}$ such that $x = x^{\perp} + x^{\parallel}$, where

$$\begin{aligned} x^{\parallel} &= \lambda u \qquad \text{for some unknown } \lambda \in \mathbb{R}, \\ x^{\perp} \bullet u &= 0. \end{aligned}$$



To find λ, we do the following calculation

$$u \bullet x = u \bullet (x^{\perp} + x^{\parallel}) = u \bullet x^{\perp} + u \bullet x^{\parallel} = 0 + u \bullet (\lambda u) = \lambda(u \bullet u).$$

Hence, $\lambda = \dfrac{u \bullet x}{u \bullet u}$. Observe that $\mathrm{im}(P_u) = \{x \in \mathbb{R}^n \mid \exists \lambda \in \mathbb{R} : x = \lambda u\} = \{\lambda u \in \mathbb{R}^n \mid \lambda \in \mathbb{R}\}$.

Definition 5.3 Let $u \in \mathbb{R}^n$ with $u \neq \mathbf{0}$. We define the **orthogonal projection** P_u onto the line spanned by u as

$$\begin{aligned} P_u : \mathbb{R}^n &\longrightarrow \mathbb{R}^n \\ x &\longmapsto \frac{u \bullet x}{u \bullet u} u \,. \end{aligned}$$

It is in our interest that the orthogonal projection is indeed a linear map.

Proposition 5.4 *$P_u : \mathbb{R}^n \to \mathbb{R}^n$ is a linear map.*

Proof. For $x, y \in \mathbb{R}^n$, $\lambda \in \mathbb{R}$, using Proposition 5.2, we have

$$\begin{aligned} P_u(x+y) &= \frac{u \bullet (x+y)}{u \bullet u} u = \frac{u \bullet x + u \bullet y}{u \bullet u} u \\ &= \frac{u \bullet x}{u \bullet u} u + \frac{u \bullet y}{u \bullet u} u = P_u(x) + P_u(y), \end{aligned}$$

$$P_u(\lambda x) = \frac{u \bullet (\lambda x)}{u \bullet u} u = \frac{\lambda (u \bullet x)}{u \bullet u} u = \lambda P_u(x).$$

As a result, P_u is a linear map. $\square$

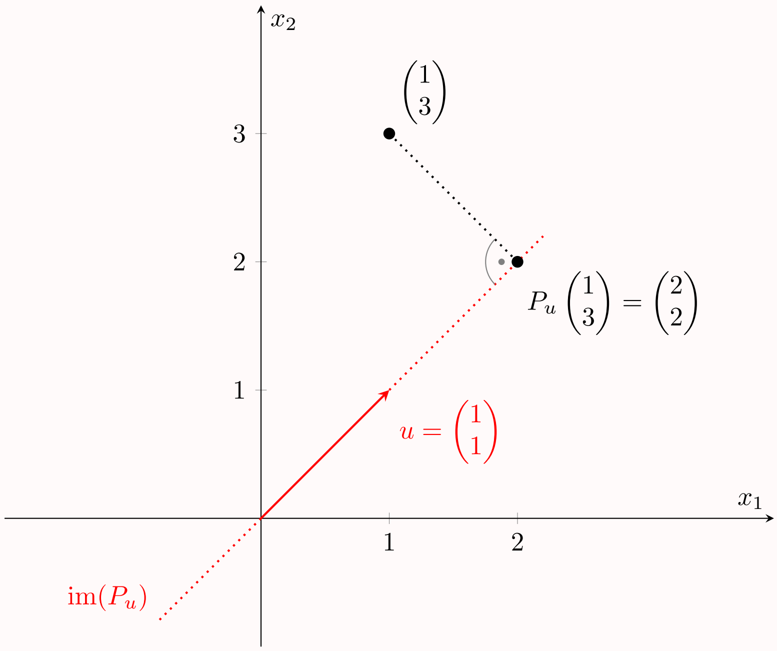
Example 20 For $n = 2$, consider $u = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$. We have

$$\begin{aligned} u \bullet u &= 1 \cdot 1 + 1 \cdot 1 = 2, \\ u \bullet x &= 1 \cdot x_1 + 1 \cdot x_2 = x_1 + x_2. \end{aligned}$$

Then,

$$\begin{aligned} P_u(x) &= \frac{u \bullet x}{u \bullet u} u = \frac{x_1 + x_2}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} \frac{x_1 + x_2}{2} \\ \frac{x_1 + x_2}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = [P_u] x, \end{aligned}$$

where $[P_u] = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$.



For example, we have

$$P_u\begin{pmatrix}1\\3\end{pmatrix} = \begin{pmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}\begin{pmatrix}1\\3\end{pmatrix} = \begin{pmatrix}2\\2\end{pmatrix}.$$

In the case $n = 1$, the dot product is just the multiplication of real numbers, and we have

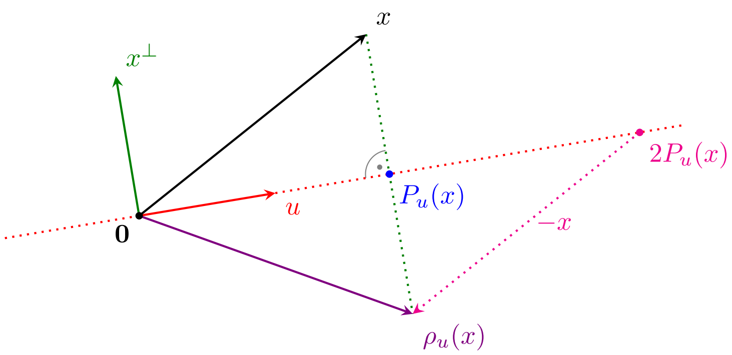
$$\begin{aligned} P_u : \mathbb{R} &\longrightarrow \mathbb{R}. \\ x &\longmapsto \frac{ux}{uu}u = x \end{aligned}$$

Hence, $P_u = \mathrm{id}_{\mathbb{R}}$ in the case $n = 1$. For $n > 1$, $P_u : \mathbb{R}^n \to \mathbb{R}^n$ is not injective and not surjective (check it yourself), and

$$\mathrm{im}(P_u) = \{\lambda u \mid \lambda \in \mathbb{R}\}.$$

3. Reflections

Now we want to define a map $\rho_u : \mathbb{R}^n \to \mathbb{R}^n$, which reflects $x \in \mathbb{R}^n$ along the line spanned by u.



To do that, we proceed as follows:

$$\begin{aligned}\rho_u(x) = x - 2x^\perp &= x - 2(x - P_u(x)) \\ &= 2P_u(x) - x \\ &= 2\frac{u \bullet x}{u \bullet u}u - x\end{aligned}$$

Definition 5.5 Let $u \in \mathbb{R}^n$ with $u \neq \mathbf{0}$. We define the **reflection** ρ_u along the line spanned by u as

$$\begin{aligned}\rho_u : \mathbb{R}^n &\longrightarrow \mathbb{R}^n \\ x &\longmapsto 2\frac{u \bullet x}{u \bullet u}u - x\,.\end{aligned}$$

Again, this map is also a linear map.

Proposition 5.6 *$\rho_u : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is a linear map.*

Proof. This is Exercise 21 □

Example 21 For $u = \begin{pmatrix}1\\1\end{pmatrix}$, the ρ_u is the reflection along the diagonal. What is the matrix of ρ_u? We have

$$[\rho_u] = \begin{pmatrix} | & | \\ \rho_u(e_1) & \rho_u(e_2) \\ | & | \end{pmatrix},$$

where

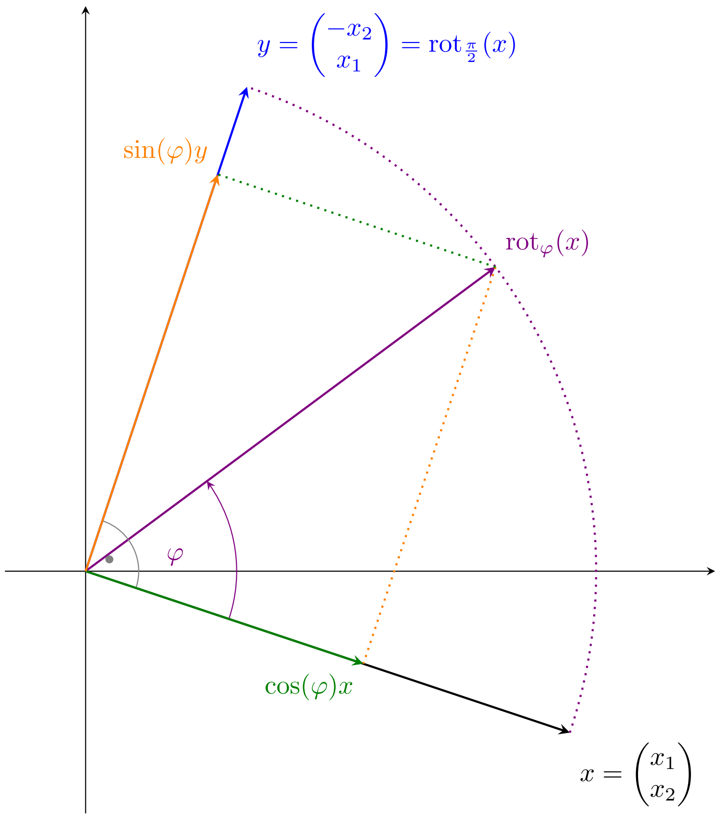
$$\begin{aligned}\rho_u(e_1) &= 2\frac{u \bullet e_1}{u \bullet u}u - e_1 = \left(2 \cdot \frac{1\cdot 1 + 1 \cdot 0}{1 \cdot 1 + 1 \cdot 1}\right)\begin{pmatrix}1\\1\end{pmatrix} - \begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}0\\1\end{pmatrix}, \\ \rho_u(e_2) &= 2\frac{u \bullet e_2}{u \bullet u}u - e_2 = \left(2 \cdot \frac{1\cdot 0 + 1 \cdot 1}{1 \cdot 1 + 1 \cdot 1}\right)\begin{pmatrix}1\\1\end{pmatrix} - \begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}1\\0\end{pmatrix}.\end{aligned}$$

Hence, $[\rho_u] = \begin{pmatrix}0 & 1\\1 & 0\end{pmatrix}$.

For any $n \geq 1$ and any $u \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, the map ρ_u is bijective with $\rho_u^{-1} = \rho_u$ because $\rho_u \circ \rho_u = \mathrm{id}_{\mathbb{R}^n}$ (this is supported with our common sense).

4. Rotations in $\mathbb{R}^2$

We want to define a map $\mathrm{rot}_\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ that describes a counterclockwise rotation with angle $\varphi \in \mathbb{R}$.



To do that, we proceed as follows:

$$\begin{aligned}
\mathrm{rot}_\varphi(x) &= \cos(\varphi)x + \sin(\varphi)y \\
&= \cos(\varphi)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \sin(\varphi)\begin{pmatrix} -x_2 \\ x_1 \end{pmatrix} \\
&= \begin{pmatrix} \cos(\varphi)x_1 - \sin(\varphi)x_2 \\ \cos(\varphi)x_2 + \sin(\varphi)x_1 \end{pmatrix} \\
&= \begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\
&= [\mathrm{rot}_\varphi]x,
\end{aligned}$$

where $[\mathrm{rot}_\varphi] = \begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix}$.

Definition 5.7 For $\varphi \in \mathbb{R}$ the counterclockwise **rotation by an angle** φ (in $\mathbb{R}^2$) is given by

$$\begin{aligned}
\mathrm{rot}_\varphi : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\
x &\longmapsto \begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix} x\,.
\end{aligned}$$

We have $\mathrm{rot}_{\varphi_1} \circ \mathrm{rot}_{\varphi_2} = \mathrm{rot}_{\varphi_1+\varphi_2}$, i.e. rot_φ is invertible with inverse $\mathrm{rot}_{-\varphi}$ because

$$\mathrm{rot}_\varphi \circ \mathrm{rot}_{-\varphi} = \mathrm{rot}_0 = \mathrm{id}_{\mathbb{R}^2}\,.$$

Exercises

Exercise 21. Let $u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} \in \mathbb{R}^n$ be with $u \neq \mathbf{0}$.

(i) Show that the reflection $\rho_u : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is a linear map.
(ii) Show that the matrix of the projection $P_u : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is given by

$$[P_u] = \frac{1}{u \bullet u} u u^T \in \mathbb{R}^{n \times n},$$

where $u^T = (u_1\, u_2\, \ldots\, u_n) \in \mathbb{R}^{1 \times n}$. Use this to give an expression for $[\rho_u]$.

Exercise 22. Show that for all $u \in \mathbb{R}^n$ with $u \neq \mathbf{0}$ the projection P_u and the reflection ρ_u satisfy for all $x \in \mathbb{R}^n$ the following two properties:

(i) $P_u(P_u(x)) = P_u(x)$.
(ii) $\rho_u(\rho_u(x)) = x$.

6

Composition of linear maps & Matrix multiplication

Linear maps are functions so we can compose them

$$
\begin{tikzcd}
\mathbb{R}^n \arrow[r, "F"] \arrow[rr, bend right, "G\circ F = GF"'] & \mathbb{R}^m \arrow[r, "G"] & \mathbb{R}^l.
\end{tikzcd}
$$

If F, G are linear maps, then does GF inherit the linearity property from them? The answer is yes by the following theorem.

Theorem 6.1 *If $F : \mathbb{R}^n \to \mathbb{R}^m$ and $G : \mathbb{R}^m \to \mathbb{R}^l$ are linear, then $GF : \mathbb{R}^n \to \mathbb{R}^l$ is linear.*

Proof. Since F and G are linear, we have for $x, y \in \mathbb{R}^n$,

$$GF(x+y) = G\big(F(x+y)\big) = G\big(F(x)+F(y)\big) = G\big(F(x)\big) + G\big(F(y)\big) = GF(x) + GF(y).$$

Also, we have for $\lambda \in \mathbb{R}, x \in \mathbb{R}^n$,

$$GF(\lambda x) = G\big(F(\lambda x)\big) = G\big(\lambda F(x)\big) = \lambda G\big(F(x)\big) = \lambda GF(x).$$

Therefore, GF is also a linear map. $\square$

A natural question follows: What is the matrix of GF?

Example 22 We consider the following linear maps

$$
\begin{aligned}
F : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2, \\
x &\longmapsto \begin{pmatrix} 1 & 2 \\ 3 & -1 \end{pmatrix} x,
\end{aligned}
\qquad
\begin{aligned}
G : \mathbb{R}^2 &\longrightarrow \mathbb{R}^3, \\
x &\longmapsto \begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} x.
\end{aligned}
$$

We have the matrices of the map F and G as follows:

$$[F] = \begin{pmatrix} 1 & 2 \\ 3 & -1 \end{pmatrix}, \qquad [G] = \begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

We want to calculate the matrix of $GF : \mathbb{R}^2 \to \mathbb{R}^3$. We have

$$[GF] = \begin{pmatrix} | & | \\ GF(e_1) & GF(e_2) \\ | & | \end{pmatrix},$$

where

$$GF(e_1) = G(F(e_1)) = G\left(F\begin{pmatrix}1\\0\end{pmatrix}\right) = G\begin{pmatrix}1\\3\end{pmatrix} = [G]\begin{pmatrix}1\\3\end{pmatrix} = \begin{pmatrix}-2\\3\\1\end{pmatrix},$$

$$GF(e_2) = G(F(e_2)) = G\left(F\begin{pmatrix}0\\1\end{pmatrix}\right) = G\begin{pmatrix}2\\-1\end{pmatrix} = [G]\begin{pmatrix}2\\-1\end{pmatrix} = \begin{pmatrix}3\\-1\\2\end{pmatrix}.$$

Hence,

$$[GF] = \begin{pmatrix} | & | \\ GF(e_1) & GF(e_2) \\ | & | \end{pmatrix} = \begin{pmatrix} | & | \\ [G]\begin{pmatrix}1\\3\end{pmatrix} & [G]\begin{pmatrix}2\\-1\end{pmatrix} \\ | & | \end{pmatrix} = \begin{pmatrix} -2 & 3 \\ 3 & -1 \\ 1 & 2 \end{pmatrix}.$$

We see that in order to find $[GF]$, we multiply $[G]$ with each of the columns of $[F]$ to get each of the corresponding columns of $[GF]$, which motivates us to define the matrix multiplication as follows.

Definition 6.2 Let $A \in \mathbb{R}^{l\times m}$ and $B \in \mathbb{R}^{m\times n}$, where B has the columns $v_1, \dots, v_n \in \mathbb{R}^m$, i.e.

$$B = \begin{pmatrix} | & & | \\ v_1 & \dots & v_n \\ | & & | \end{pmatrix}.$$

Then the **product of A and B** is the $l \times n$-matrix with columns $Av_1, \dots, Av_n \in \mathbb{R}^l$, i.e.

$$A \cdot B = \begin{pmatrix} | & & | \\ Av_1 & \dots & Av_n \\ | & & | \end{pmatrix} \in \mathbb{R}^{l\times n}.$$

Example 23 1) $A = \begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 2 \\ 3 & -1 \end{pmatrix}$ then

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

Compare this with Example 22, where $[G] = A$, $[F] = B$ and $[GF] = A \cdot B$.

2) Consider the following matrix multiplication:

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

3) Consider the following matrix multiplication:

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

In the next lecture, we will learn that $\begin{pmatrix} 3 & 2 \\ 1 & 1 \end{pmatrix}$ is the inverse of $\begin{pmatrix} 1 & -2 \\ -1 & 3 \end{pmatrix}$.

We already know that each matrix corresponds to a linear map. The following theorem confirms that the matrix multiplication is indeed equivalent to the composition of the corresponding linear maps.

Theorem 6.3 *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ and $G : \mathbb{R}^m \to \mathbb{R}^l$ be linear maps. Then the matrix of GF is given by the product of the matrices of G and F, i.e.*

$$[GF] = [G] \cdot [F] .$$

Proof. We have $[F] = \begin{pmatrix} | & & | \\ F(e_1) & \cdots & F(e_n) \\ | & & | \end{pmatrix}$ and

$$\begin{aligned}
[G] \cdot [F] &= \begin{pmatrix} | & & | \\ [G]F(e_1) & \cdots & [G]F(e_n) \\ | & & | \end{pmatrix} \\
&= \begin{pmatrix} | & & | \\ G\big(F(e_1)\big) & \cdots & G\big(F(e_n)\big) \\ | & & | \end{pmatrix} \\
&= \begin{pmatrix} | & & | \\ GF(e_1) & \cdots & GF(e_n) \\ | & & | \end{pmatrix} \\
&= [GF].
\end{aligned}$$

Hence, we have proved the theorem. □

Example 24 1) Given the linear map

$$\begin{aligned}
F : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2, \\
\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &\longmapsto \begin{pmatrix} 2x_1 - x_2 \\ x_1 + 3x_2 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} ,
\end{aligned}$$

what is the matrix of $F \circ F$?

- By hand:

$$\begin{aligned}
F \circ F(x) = F\big(F(x)\big) = F \begin{pmatrix} 2x_1 - x_2 \\ x_1 + 3x_2 \end{pmatrix} &= \begin{pmatrix} 2(2x_1 - x_2) - (x_1 + 3x_2) \\ (2x_1 - x_2) + 3(x_1 + 3x_2) \end{pmatrix} \\
&= \begin{pmatrix} 3x_1 - 5x_2 \\ 5x_1 + 8x_2 \end{pmatrix} = \begin{pmatrix} 3 & -5 \\ 5 & 8 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\
&\Rightarrow [FF] = \begin{pmatrix} 3 & -5 \\ 5 & 8 \end{pmatrix} .
\end{aligned}$$

- Using Theorem 6.3:

$$[FF] = [F] \cdot [F] = \begin{pmatrix} 2 & -1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 3 & -5 \\ 5 & 8 \end{pmatrix} .$$

2) In the last lecture, we define the rotation by angle φ as the linear map

$$\begin{aligned}
\mathrm{rot}_\varphi : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2, \\
x &\longmapsto \begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix} x.
\end{aligned}$$

The composition $\mathrm{rot}_{\varphi_1} \circ \mathrm{rot}_{\varphi_2}$ has the meaning as the rotation by φ_2 and then by φ_1. We also mention in the last lecture that

$$\mathrm{rot}_{\varphi_1} \circ \mathrm{rot}_{\varphi_2} = \mathrm{rot}_{\varphi_1+\varphi_2}\,. \tag{$*$}$$

Using Theorem 6.3, we have

$$\begin{aligned}
[\mathrm{rot}_{\varphi_1} \circ \mathrm{rot}_{\varphi_2}] &= [\mathrm{rot}_{\varphi_1}] \cdot [\mathrm{rot}_{\varphi_2}] \\
&= \begin{pmatrix} \cos(\varphi_1) & -\sin(\varphi_1) \\ \sin(\varphi_1) & \cos(\varphi_1) \end{pmatrix} \begin{pmatrix} \cos(\varphi_2) & -\sin(\varphi_2) \\ \sin(\varphi_2) & \cos(\varphi_2) \end{pmatrix} \\
&= \begin{pmatrix} \cos(\varphi_1)\cos(\varphi_2) - \sin(\varphi_1)\sin(\varphi_2) & -\cos(\varphi_1)\sin(\varphi_2) - \sin(\varphi_1)\cos(\varphi_2) \\ \sin(\varphi_1)\cos(\varphi_2) + \cos(\varphi_1)\sin(\varphi_2) & -\sin(\varphi_1)\sin(\varphi_2) + \cos(\varphi_1)\cos(\varphi_2) \end{pmatrix}.
\end{aligned}$$

In addition, $[\mathrm{rot}_{\varphi_1+\varphi_2}] = \begin{pmatrix} \cos(\varphi_1+\varphi_2) & -\sin(\varphi_1+\varphi_2) \\ \sin(\varphi_1+\varphi_2) & \cos(\varphi_1+\varphi_2) \end{pmatrix}$. By using $(*)$, we obtain the angle sum identities:

$$\begin{aligned}
\cos(\varphi_1+\varphi_2) &= \cos(\varphi_1)\cos(\varphi_2) - \sin(\varphi_1)\sin(\varphi_2), \\
\sin(\varphi_1+\varphi_2) &= \sin(\varphi_1)\cos(\varphi_2) + \cos(\varphi_1)\sin(\varphi_2).
\end{aligned}$$

The matrix multiplication has the following properties.

Proposition 6.4 *For all $A \in \mathbb{R}^{l\times m}, B, D \in \mathbb{R}^{m\times n}, C \in \mathbb{R}^{n\times p}$ and $\lambda \in \mathbb{R}$ we have*

(i) $A \cdot I_m = I_l \cdot A = A$, *where I_m denotes the $m\times m$-identity matrix.*
(ii) $(A\cdot B)\cdot C = A\cdot(B\cdot C)$.
(iii) $A\cdot(B+D) = A\cdot B + A\cdot D$.
(iv) $(B+D)\cdot C = B\cdot C + D\cdot C$.
(v) $\lambda(A\cdot B) = (\lambda A)\cdot B = A\cdot(\lambda B)$.

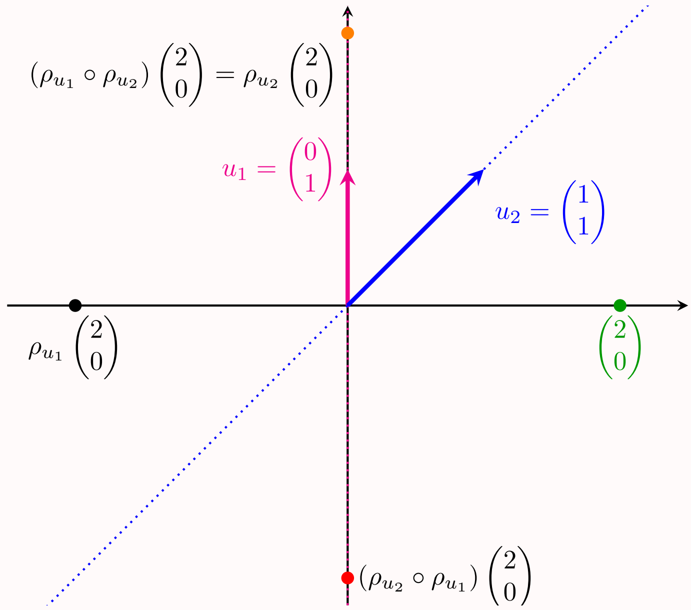
Proof. Check by yourself. Similar to Proposition 2.4. □

Remark. If $A, B \in \mathbb{R}^{n\times n}$ then in general we have $A\cdot B \neq B\cdot A$. For example,

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \text{ but } \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Example 25 In the last lecture, for $u \neq \mathbf{0}$, ρ_u denotes the reflection along the line spanned by u. Consider $u_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and $u_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, and their corresponding reflections with their matrices:

$$[\rho_{u_1}] = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad [\rho_{u_2}] = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$



We have

$$\rho_{u_1}\left(\rho_{u_2}\begin{pmatrix} 2 \\ 0 \end{pmatrix}\right) = \rho_{u_1}\begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$
$$\rho_{u_2}\left(\rho_{u_1}\begin{pmatrix} 2 \\ 0 \end{pmatrix}\right) = \rho_{u_2}\begin{pmatrix} -2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ -2 \end{pmatrix}$$

Hence, $\rho_{u_1}\left(\rho_{u_2}\begin{pmatrix} 2 \\ 0 \end{pmatrix}\right) \neq \rho_{u_2}\left(\rho_{u_1}\begin{pmatrix} 2 \\ 0 \end{pmatrix}\right)$. Generally, $\rho_{u_1}(\rho_{u_2}(x)) \neq \rho_{u_2}(\rho_{u_1}(x))$ for any vector $x \in \mathbb{R}^2$ because

$$[\rho_{u_1} \circ \rho_{u_2}] = [\rho_{u_1}] \cdot [\rho_{u_2}] = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$
$$[\rho_{u_2} \circ \rho_{u_1}] = [\rho_{u_2}] \cdot [\rho_{u_1}] = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

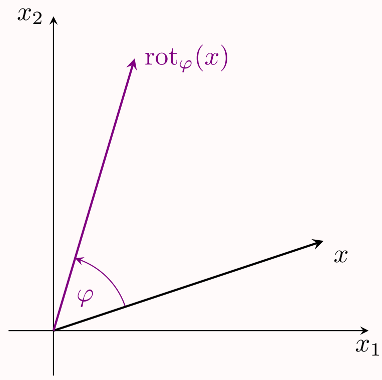
$$\Rightarrow [\rho_{u_1} \circ \rho_{u_2}] \neq [\rho_{u_2} \circ \rho_{u_1}].$$

Therefore, reflecting first along u_2 and then u_1 is different to first reflecting along u_1 and then u_2.

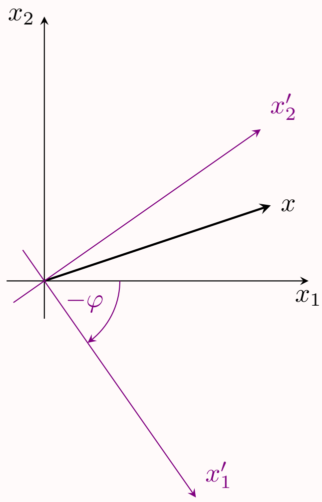
Remark. Notice that sometimes (really rare) we have $A \cdot B = B \cdot A$. For example,

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

Example 26 (Rotations in $\mathbb{R}^3$) There are two interpretations of rotation maps. Take the two-dimensional rotation rot_φ with $\varphi \in \mathbb{R}$ for example. For any vector $x \in \mathbb{R}^2$, the vector $\mathrm{rot}_\varphi(x)$ can be interpreted in two ways. In Chapter 5, we let the coordinate system be unchanged so $\mathrm{rot}_\varphi(x)$ is a new vector in the same coordinate system. In this case, the map rot_φ is called a **active transformation**.

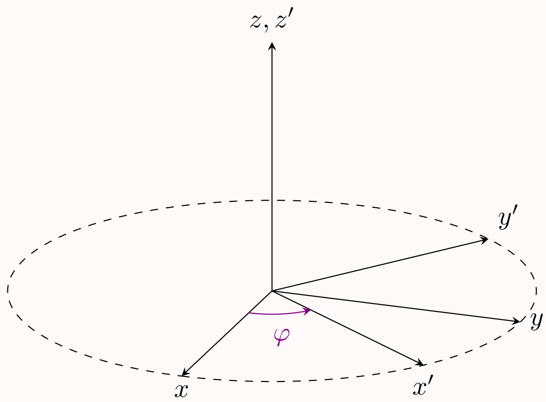


On the other hand, the vector x can be kept unchanged and the coordinate system is rotated instead. As a result, the components of $\mathrm{rot}_\varphi(x)$ are the coordinates of x in this new coordinate system. In this case, rot_φ is called a **passive transformation**.



Hence, for any $\varphi > 0$, the map rot_φ corresponds to a counterclockwise rotation by the angle φ when applied to the vector x. On the other hand, it also corresponds to a clockwise rotation by the angle φ when applied to the coordinate system. The same duality of roles often occurs with many transformations in physics.

For the following discussion of rotations in three dimension, we will interpret those rotations as passive transformations. The simplest three-dimensional rotations are rotations around any coordinate axes. For example, consider the following rotation.



Consider a vector $v = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$. This rotation is similar to a rotation in $\mathbb{R}^2$ except that there is one more component which is left unchanged. Hence, the new coordinates are as follows.

$$\begin{aligned} x' &= \cos(\varphi)x + \sin(\varphi)y \\ y' &= -\sin(\varphi)x + \cos(\varphi)y \\ z' &= z \end{aligned}$$

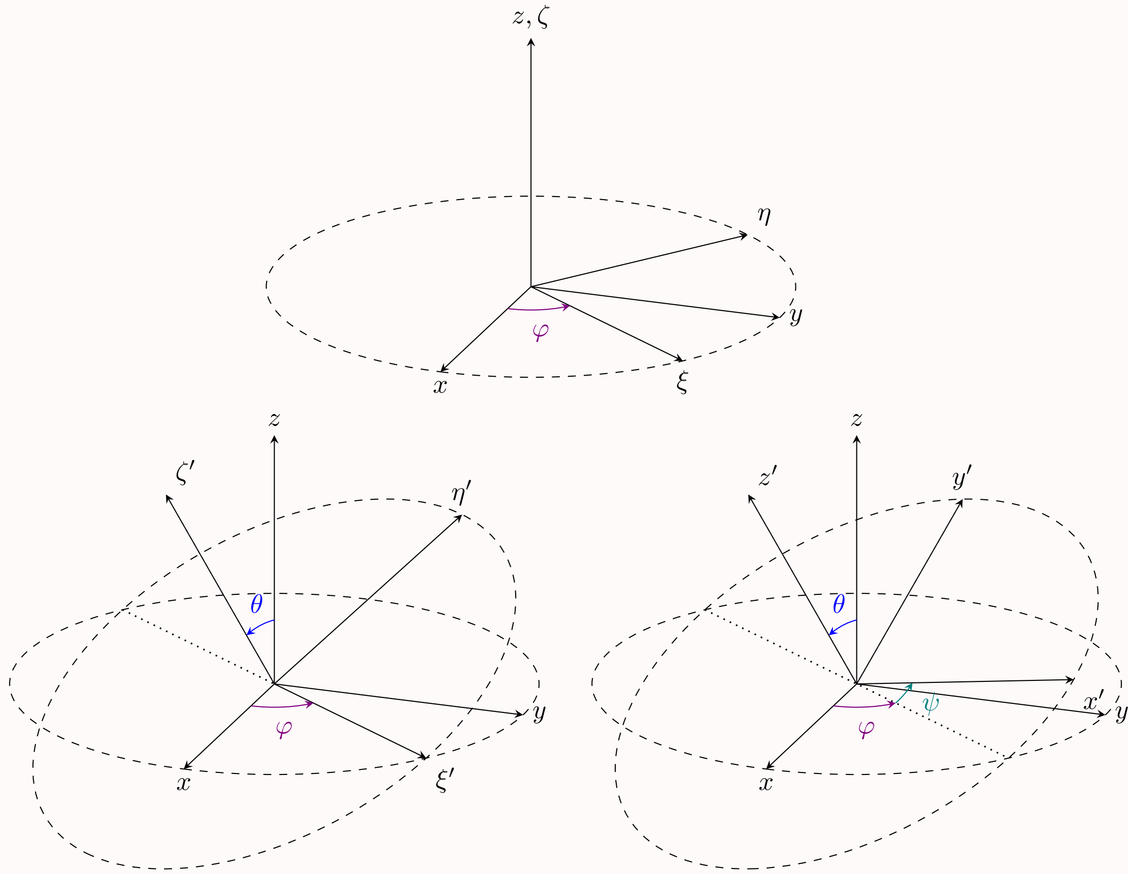
The new coordinates can be written as $v' = R_z(-\varphi)v$, where the matrix $R_z(\theta)$ is defined for any $\theta \in \mathbb{R}$ as

$$R_z(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Similarly, we can define the two matrices $R_x(\theta)$ and $R_y(\theta)$ from the rotations around x- and y-axes as follows:

$$R_x(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{pmatrix}, \qquad R_y(\theta) = \begin{pmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{pmatrix}.$$

Recall that every rotations rot_φ in $\mathbb{R}^2$ is specified by the angle φ. In $\mathbb{R}^3$, we need three parameters to describe any rotations. There are many sets of parameters that can be used but the most common and useful ones are **Euler angles** or **Eulerian angles**. The idea behind this is that every rotations can be decomposed into three successive rotations each of which is about one of the axes, with the condition that no two successive rotations can be about the same axis. Hence, there are totally 12 possible conventions in defining the Euler angles (in a right-handed coordinate system). We want to introduce 3 convenctions which are widely used in physics and engineering.
The first convention here is used widely in celestial mechanics, applied mechanics, and frequently in molecular and solid-state physics.



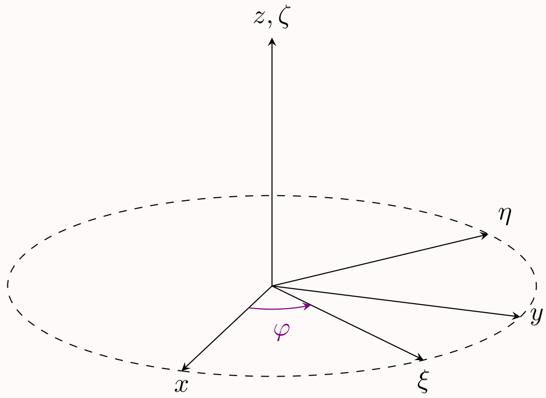
Consider any $v = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$. We get angles $u = \begin{pmatrix} \xi \\ \eta \\ \zeta \end{pmatrix}$, $u' = \begin{pmatrix} \xi' \\ \eta' \\ \zeta' \end{pmatrix}$, $v' = \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} \in \mathbb{R}^3$ after each rotations as follows: $u = R_z(-\varphi)v$, $u' = R_x(-\theta)u$, and $v' = R_z(-\psi)u'$. Hence, we get the rotated vector v' from v as follows:

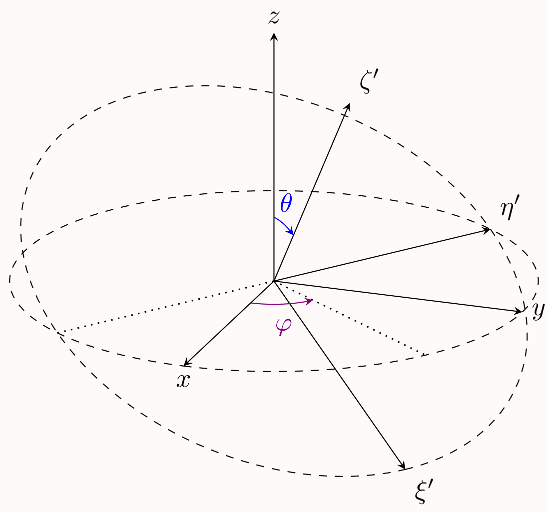
$$v' = R_z(-\psi)u' = R_z(-\psi)R_x(-\theta)u = R_z(-\psi)R_x(-\theta)R_z(-\varphi)v.$$

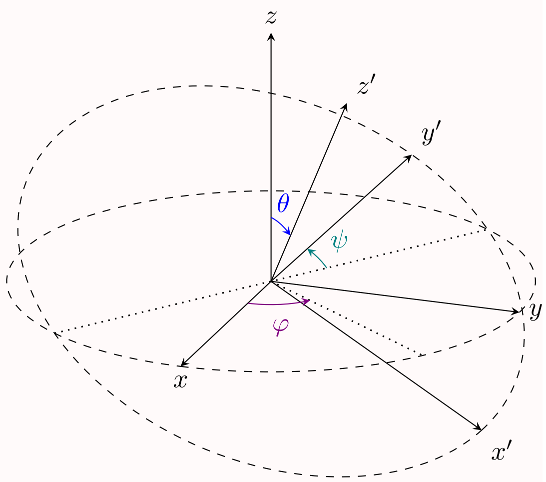
As a result, we get the matrix $R(\varphi, \theta, \psi)$ describing any general three-dimensional rotations as follows:

$$\begin{aligned}
&R(\varphi,\theta,\psi)\\
=&R_z(-\psi)R_x(-\theta)R_z(-\varphi)\\
=&\begin{pmatrix} \cos(\psi) & \sin(\psi) & 0 \\ -\sin(\psi) & \cos(\psi) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & \sin(\theta) \\ 0 & -\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} \cos(\varphi) & \sin(\varphi) & 0 \\ -\sin(\varphi) & \cos(\varphi) & 0 \\ 0 & 0 & 1 \end{pmatrix}\\
=&\begin{pmatrix} \cos(\psi)\cos(\varphi) - \cos(\theta)\sin(\varphi)\sin(\psi) & \cos(\psi)\sin(\varphi) + \cos(\theta)\cos(\varphi)\sin(\psi) & \sin(\psi)\sin(\theta) \\ -\sin(\psi)\cos(\varphi) - \cos(\theta)\sin(\varphi)\cos(\psi) & -\sin(\psi)\sin(\varphi) + \cos(\theta)\cos(\varphi)\cos(\psi) & \cos(\psi)\sin(\theta) \\ \sin(\theta)\sin(\varphi) & -\sin(\theta)\cos(\varphi) & \cos(\theta) \end{pmatrix}.
\end{aligned}$$

The second convention is used widely in quantum mechanics, nuclear physics, and particle physics.



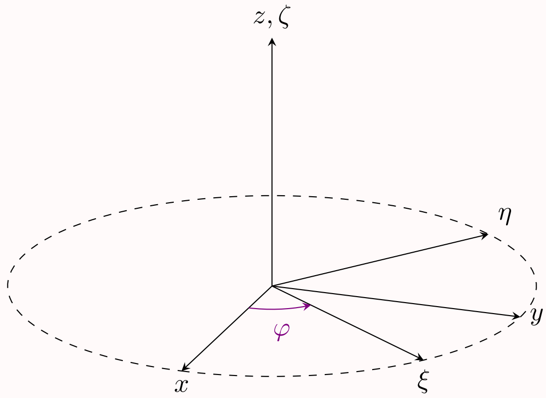


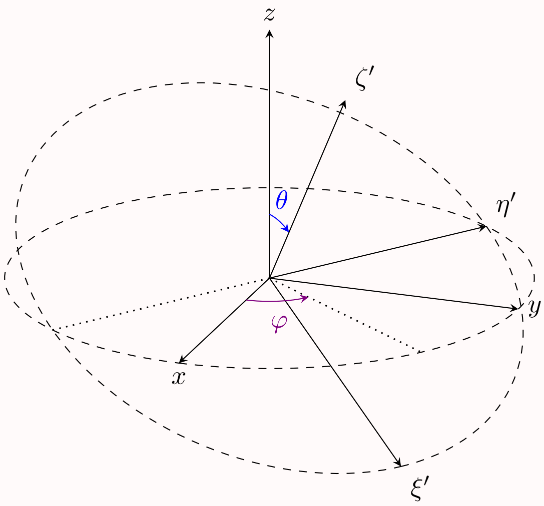


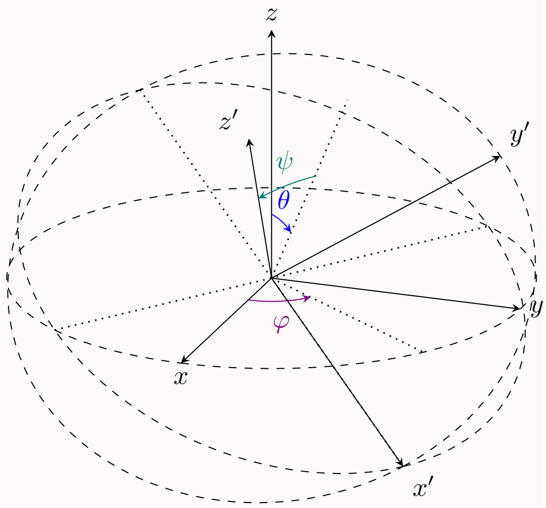
Similarly, we get

$$
\begin{aligned}
&R(\varphi,\theta,\psi)\\
=&R_z(-\psi)R_y(-\theta)R_z(-\varphi)\\
=&\begin{pmatrix}\cos(\psi) & \sin(\psi) & 0\\ -\sin(\psi) & \cos(\psi) & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}\cos(\theta) & 0 & -\sin(\theta)\\ 0 & 1 & 0\\ \sin(\theta) & 0 & \cos(\theta)\end{pmatrix}\begin{pmatrix}\cos(\varphi) & \sin(\varphi) & 0\\ -\sin(\varphi) & \cos(\varphi) & 0\\ 0 & 0 & 1\end{pmatrix}\\
=&\begin{pmatrix}-\sin(\psi)\sin(\varphi)+\cos(\theta)\cos(\varphi)\cos(\psi) & \sin(\psi)\cos(\varphi)+\cos(\theta)\sin(\varphi)\cos(\psi) & -\cos(\psi)\sin(\theta)\\ -\cos(\psi)\sin(\varphi)-\cos(\theta)\cos(\varphi)\sin(\psi) & \cos(\psi)\cos(\varphi)-\cos(\theta)\sin(\varphi)\sin(\psi) & \sin(\psi)\sin(\theta)\\ \sin(\theta)\cos(\varphi) & \sin(\theta)\sin(\varphi) & \cos(\theta)\end{pmatrix}.
\end{aligned}
$$

The last convention, also called **Tait-Bryan angles**, is widely used in engineering applications relating to the orientation of moving vehicles such as aircraft and satellites.







In a similar fashion, we get

$$
\begin{aligned}
&R(\varphi,\theta,\psi)\\
=&R_x(-\psi)R_y(-\theta)R_z(-\varphi)\\
=&\begin{pmatrix}1 & 0 & 0\\ 0 & \cos(\psi) & \sin(\psi)\\ 0 & -\sin(\psi) & \cos(\psi)\end{pmatrix}\begin{pmatrix}\cos(\theta) & 0 & -\sin(\theta)\\ 0 & 1 & 0\\ \sin(\theta) & 0 & \cos(\theta)\end{pmatrix}\begin{pmatrix}\cos(\varphi) & \sin(\varphi) & 0\\ -\sin(\varphi) & \cos(\varphi) & 0\\ 0 & 0 & 1\end{pmatrix}\\
=&\begin{pmatrix}\cos(\theta)\cos(\varphi) & \cos(\theta)\sin(\varphi) & -\sin(\theta)\\ \sin(\psi)\sin(\theta)\cos(\varphi)-\cos(\psi)\sin(\varphi) & \sin(\psi)\sin(\theta)\sin(\varphi)+\cos(\psi)\cos(\varphi) & \cos(\theta)\sin(\psi)\\ \cos(\psi)\sin(\theta)\cos(\varphi)+\sin(\psi)\sin(\varphi) & \cos(\psi)\sin(\theta)\sin(\varphi)-\sin(\psi)\cos(\varphi) & \cos(\theta)\cos(\psi)\end{pmatrix}
\end{aligned}
$$

In this case, the three parameters have their names: the angle φ of the rotation about the vertical axis (z- or ζ-axis) is the **heading** or **yaw** angle; the angle θ of the rotation around a perpendicular axis (y- or η-axis) fixed in the vehicle and normal to the figure axis (x- or ξ-axis) is the **pitch** or *attitude* angle; the angle ψ of the rotation about the figure axis of the vehicle is the **roll** or **bank** angle.
Furthermore, there is another useful set of parameters called the **Cayley-Klein parameters**. This set comprises 4 parameters that are better than the Euler angles to use in numerical computation due to the large number of trigonometric functions involved when using the Euler angles. Besides, the four-parameter sets are also useful in branches of physics, wherever rotations or rotational symmetry are involved. However, we will not introduce it here because complex numbers are involved. For more details, see [G].

Exercises

Exercise 23. Let $u = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, $d = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $x = \begin{pmatrix} 5 \\ 0 \end{pmatrix}$.

(i) Calculate the matrices $[P_u]$ and $[\rho_u]$ in this special case.

(ii) Calculate the following vectors and draw them in one picture together with u, d and x

$$P_u(x), \qquad \rho_u(x), \qquad (P_u \circ P_d)(x), \qquad \mathrm{rot}_{\frac{\pi}{2}}(x),$$
$$(P_u \circ \mathrm{rot}_{\frac{\pi}{2}})(x), \quad (\mathrm{rot}_{\frac{\pi}{2}} \circ P_u)(x), \quad (P_d \circ \mathrm{rot}_{\frac{\pi}{2}} \circ P_u)(x)\,.$$

7

The inverse of a linear map

In Chapter 3, we learned that a function $f : X \to Y$ is invertible if there exists a function $g : Y \to X$ such that for every $x \in X$ and $y \in Y$, $g(f(x)) = x$ and $f(g(y)) = y$. Then, $g = f^{-1}$ and it is called the inverse of f.

We also saw that invertibility is equivalent to bijectivity. A function f is bijective if for every $y \in Y$, there uniquely exists $x \in X$ such that $y = f(x)$. Equivalently, a function is bijective when it is injective and surjective.

In this chapter, we want to answer the following questions:

(a) When is a linear map invertible?

(b) Is the inverse also linear?

(c) How to calculate the inverse?

Example 27 Consider the linear map

$$\begin{aligned} F : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2, \\ x &\longmapsto \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} x \,. \end{aligned}$$

Is F invertible?

Take $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \in \mathbb{R}^2$ and check if $F(x) = y$ has a unique solution.

$$-2 \left(\begin{array}{cc|c} 1 & 3 & y_1 \\ 2 & 4 & y_2 \end{array}\right) \sim -\tfrac{1}{2} \left(\begin{array}{cc|c} 1 & 3 & y_1 \\ 0 & -2 & y_2 - 2y_1 \end{array}\right) \sim -3 \left(\begin{array}{cc|c} 1 & 3 & y_1 \\ 0 & 1 & y_1 - \frac{1}{2}y_2 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & -2y_1 + \frac{3}{2}y_2 \\ 0 & 1 & y_1 - \frac{1}{2}y_2 \end{array}\right)$$

We get the unique solution

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -2y_1 + \frac{3}{2}y_2 \\ y_1 - \frac{1}{2}y_2 \end{pmatrix} = \begin{pmatrix} -2 & \frac{3}{2} \\ 1 & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

Hence, F is invertible and its inverse is

$$\begin{aligned} F^{-1} : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2, \\ y &\longmapsto \begin{pmatrix} -2 & \frac{3}{2} \\ 1 & -\frac{1}{2} \end{pmatrix} y. \end{aligned}$$

The **rank of a linear map** $F : \mathbb{R}^n \to \mathbb{R}^m$ is defined by the rank of its matrix, i.e. $\operatorname{rk}(F) := \operatorname{rk}([F])$. The following theorem answers the question about when a linear map is invertible.

Theorem 7.1 *A linear map $F : \mathbb{R}^n \to \mathbb{R}^m$ is invertible if and only if $m = n = \operatorname{rk}(F)$.*

Remark. This theorem is in the format "Statement 1 if and only if Statement 2", which means that both the statements are either true or false. In addition, it also means both "Statement 1 implies Statement 2" and "Statement 2 implies Statement 1". To prove this theorem, we need to prove both the implications.

Proof. F is invertible $\iff$ $[F]x = y$ has a unique solution for all $y \in \mathbb{R}^m$.

$$\begin{array}{cc} & \scriptstyle\longleftarrow n+1 \longrightarrow \\ \scriptstyle m\updownarrow & \Big([F] \mid y\Big) \end{array} \sim \ldots \sim \big(B \mid z\big) = \operatorname{rref}\Big([F] \mid y\Big),$$

where $B = \operatorname{rref}\big([F]\big)$ and $z \in \mathbb{R}^m$ can be an arbitrary vector depending on y.

In one direction, suppose F is invertible. We want to show that $n = m = \operatorname{rk}(F) = \operatorname{rk}\big([F]\big)$.
If $\operatorname{rk}(F) < m$, then

$$\big(B \mid z\big) = \begin{array}{c} \left(\begin{array}{c|c} * & \\ \hline 0 & z \end{array}\right) \end{array} \begin{array}{c} \uparrow \\ m \\ \downarrow \end{array}$$

Hence, no solution for some z (and thus, some y). Therefore, $\operatorname{rk}(F) = m$.
If $\operatorname{rk}(F) < n$, then

$$\big(B \mid z\big) = \begin{array}{c} \longleftarrow\quad n \quad\longrightarrow \\ \left(\begin{array}{ccccc|c} 1 & & * & & & \\ & \ddots & * & & & \\ & & 1 \;\; * & & & \\ & & 0 & 1 & & z \\ & & \vdots & & \ddots & \\ & & 0 & & & \end{array}\right) \end{array}$$

In this case, there are columns without pivot elements so there is no unique solution (infinitely many solutions). Therefore, $\operatorname{rk}(F) = n$.

Conversely, if $m = n = \operatorname{rk}(F)$, then $B = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} = I_n$. Hence, $[F]x = y$ has a unique solution for all $y \in \mathbb{R}^m$; therefore, F is invertible. $\square$

Next, the following proposition confirms that the inverse of a linear map is also linear.

Proposition 7.2 *If a linear map $F : \mathbb{R}^n \to \mathbb{R}^n$ is invertible, then its inverse F^{-1} is also linear.*

Proof. Let $u, v \in \mathbb{R}^n$. Set $x = F^{-1}(u)$ and $y = F^{-1}(v)$, i.e. $F(x) = u$ and $F(y) = v$.
Then

$$\begin{aligned} F^{-1}(u) + F^{-1}(v) = x + y = \mathrm{id}(x+y) &= F^{-1}F(x+y) \\ &= F^{-1}\big(F(x) + F(y)\big) && \text{(because F is linear)} \\ &= F^{-1}(u+v). \end{aligned}$$

In addition, for $\lambda \in \mathbb{R}$,

$$\lambda F^{-1}(u) = \lambda x = F^{-1}F(\lambda x) = F^{-1}\big(\lambda F(x)\big) = F^{-1}(\lambda u).$$

Hence, F^{-1} is linear. □

Since each linear map corresponds to a matrix, we can define the inverse of a matrix corresponding to a invertible linear map as follows.

Definition 7.3 If $A \in \mathbb{R}^{n\times n}$ is the matrix of an invertible linear map $F : \mathbb{R}^n \to \mathbb{R}^n$ (i.e. $A = [F]$), then we define the **inverse of** A by $A^{-1} := \big[F^{-1}\big]$.

Naturally, we can ask when a matrix is invertible, which is answered by the following theorem.

Theorem 7.4 *The inverse of $A \in \mathbb{R}^{n\times n}$ exists (A is invertible) if and only if $\mathrm{rref}(A) = I_n$.*

Proof. This theorem follows from the proof of Theorem 7.1. □

The inverses of matrices have the following properties.

Proposition 7.5 *If $A, B \in \mathbb{R}^{n\times n}$ are invertible we have*
(i) $AA^{-1} = A^{-1}A = I_n$,
(ii) $(BA)^{-1} = A^{-1}B^{-1}$.

Proof. Suppose A and B are matrices of linear maps $F : \mathbb{R}^n \to \mathbb{R}^n$ and $G : \mathbb{R}^n \to \mathbb{R}^n$, respectively.

(i) Using Theorem 6.3, we have

$$\begin{aligned} AA^{-1} &= [F] \cdot \big[F^{-1}\big] = \big[F \circ F^{-1}\big] = [\mathrm{id}_{\mathbb{R}^n}] = I_n, \\ A^{-1}A &= \big[F^{-1}\big] \cdot [F] = \big[F^{-1} \circ F\big] = [\mathrm{id}_{\mathbb{R}^n}] = I_n. \end{aligned}$$

Hence, $AA^{-1} = A^{-1}A = I_n$.

(ii) We have $BA = [G] \cdot [F] = [G \circ F]$ and $(G \circ F)^{-1} = F^{-1} \circ G^{-1}$. Therefore,

$$(BA)^{-1} = \big[(G \circ F)^{-1}\big] = \big[F^{-1} \circ G^{-1}\big] = \big[F^{-1}\big] \cdot \big[G^{-1}\big] = A^{-1}B^{-1}.$$ □

In Example 27, we determined the inverse of $A = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$ by solving the linear system $Ax = y$. In general, if we want to determine the inverse of $A \in \mathbb{R}^{n\times n}$, we can use the following algorithm (a variant of Gaussian elimination).

Algorithm 7.6 (Gauss-Jordan elimination) Given a matrix $A \in \mathbb{R}^{n\times n}$, use the Gaussian elimination to bring the augmented matrix $(A \mid I_n)$ to the row-reduced echelon form $(B \mid C)$. There are 2 scenarios:

- If $B \neq I_n$, then A is not invertible.
- If $B = I_n$, then A is invertible and $C = A^{-1}$. That is, we get

$$(A \mid I_n) \sim \ldots \sim \left(I_n \mid A^{-1}\right) = \operatorname{rref}(A \mid I_n).$$

Example 28 Determine the inverse of $A = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$.

$$(A \mid I_2) = \begin{matrix} -2 \\ \hookrightarrow \end{matrix} \left(\begin{array}{cc|cc} 1 & 3 & 1 & 0 \\ 2 & 4 & 0 & 1 \end{array}\right) \sim \begin{matrix} \\ -\frac{1}{2} \end{matrix} \left(\begin{array}{cc|cc} 1 & 3 & 1 & 0 \\ 0 & -2 & -2 & 1 \end{array}\right)$$

$$\sim \begin{matrix} \hookleftarrow \\ -3 \end{matrix} \left(\begin{array}{cc|cc} 1 & 3 & 1 & 0 \\ 0 & 1 & 1 & -\frac{1}{2} \end{array}\right) \sim \left(\begin{array}{cc|cc} 1 & 0 & -2 & \frac{3}{2} \\ 0 & 1 & 1 & -\frac{1}{2} \end{array}\right) = \left(I_2 \mid A^{-1}\right).$$

We can check that

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

(Definition 7.7 and Theorem 7.8 are just a remark and they are not so important for the rest of this course. They will appear again in detail in Linear Algebra II)

Definition 7.7 For $\lambda \in \mathbb{R}$ with $\lambda \neq 0$ and $1 \leq i, j \leq n$ we define the **elementary matrices** $R_i^{\lambda,j}, R_i^{\lambda}, R_{i,j} \in \mathbb{R}^{n\times n}$ by

$$R_i^{\lambda,j} = \begin{pmatrix} 1 &&&&& \\ & \ddots &&&& \\ && 1 &&& \\ &&& \ddots && \\ && \lambda && 1 & \\ &&&&& \ddots \\ &&&&&& 1 \end{pmatrix}, \quad R_i^{\lambda} = \begin{pmatrix} 1 &&&&& \\ & \ddots &&&& \\ && 1 &&& \\ &&& \lambda && \\ &&&& 1 & \\ &&&&& \ddots & \\ &&&&&& 1 \end{pmatrix}, \quad R_{i,j} = \begin{pmatrix} 1 &&&&& \\ & \ddots &&&& \\ && 0 && 1 & \\ &&& \ddots && \\ && 1 && 0 & \\ &&&&& \ddots & \\ &&&&&& 1 \end{pmatrix}.$$

Here the λ in $R_i^{\lambda,j}$ is in the i-th row and j-th column, in R_i^{λ} it is in the i-th row, and in $R_{i,j}$ the 0 are on the diagonal in the i-th row and j-th column.

Multiplying with an elementary matrix from the left corresponds to the elementary row operations (Definition 2.6)

(R1) Multiyplying with $R_i^{\lambda,j}$: Add λ-times row j to row i.

(R2) Multiyplying with R_i^{λ}: Multiply row j by λ. $(\lambda \neq 0)$

(R3) Multiyplying with $R_{i,j}$: Change row i and j.

Taking a look at Example 28 again, we have

$$A = \begin{matrix} -2 \\ \hookrightarrow \end{matrix} \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \sim \begin{matrix} \\ -\frac{1}{2} \end{matrix} \begin{pmatrix} 1 & 3 \\ 0 & -2 \end{pmatrix} \sim \begin{matrix} \hookleftarrow \\ -3 \end{matrix} \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_2,$$

where each step corresponds to the multiplication with an elementary matrix:

$$\begin{aligned}
R_2^{-2,1}\begin{pmatrix}1 & 3\\ 2 & 4\end{pmatrix} &= \begin{pmatrix}1 & 0\\ -2 & 1\end{pmatrix}\begin{pmatrix}1 & 3\\ 2 & 4\end{pmatrix} = \begin{pmatrix}1 & 3\\ 0 & -2\end{pmatrix},\\
R_2^{-1/2}\begin{pmatrix}1 & 3\\ 0 & -2\end{pmatrix} &= \begin{pmatrix}1 & 0\\ 0 & -\frac{1}{2}\end{pmatrix}\begin{pmatrix}1 & 3\\ 0 & -2\end{pmatrix} = \begin{pmatrix}1 & 3\\ 0 & 1\end{pmatrix},\\
R_1^{-3,2}\begin{pmatrix}1 & 3\\ 0 & 1\end{pmatrix} &= \begin{pmatrix}1 & -3\\ 0 & 1\end{pmatrix}\begin{pmatrix}1 & 3\\ 0 & 1\end{pmatrix} = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}.
\end{aligned}$$

As a result, we get

$$\begin{aligned}
& R_1^{-3,2}R_2^{-1/2}R_2^{-2,1}A = I_2\\
\Leftrightarrow \quad & A = \left(R_1^{-3,2}R_2^{-1/2}R_2^{-2,1}\right)^{-1} = \left(R_2^{-2,1}\right)^{-1}\left(R_2^{-1/2}\right)^{-1}\left(R_1^{-3,2}\right)^{-1} = R_2^{2,1}R_2^{1/2}R_1^{3,2},
\end{aligned}$$

where we use the fact that all row operations are reversible, which means that all corresponding elementary matrices are invertible. The last result shows that the invertible matrix A can be written as a product of elementary matrices $R_2^{2,1}$, $R_2^{1/2}$, $R_1^{3,2}$. This result holds in general.

Theorem 7.8 *Every invertible matrix is a product of elementary matrices.*

Exercises

Exercise 24. Decide if the following two linear maps are invertible. Determine their inverses if they exist.

$$F : \mathbb{R}^3 \longrightarrow \mathbb{R}^3, \quad \begin{pmatrix}x_1\\ x_2\\ x_3\end{pmatrix} \longmapsto \begin{pmatrix}2x_2 + 2x_3\\ x_1 - 4x_2 + 6x_3\\ x_2 + x_3\end{pmatrix},$$

$$G : \mathbb{R}^3 \longrightarrow \mathbb{R}^3, \quad \begin{pmatrix}x_1\\ x_2\\ x_3\end{pmatrix} \longmapsto \begin{pmatrix}10x_1 + x_2 - 26x_3\\ x_1 - 2x_3\\ -x_1 + x_3\end{pmatrix}.$$

8

Subspaces, Kernel & Image

In the previous sections, we considered subsets of $\mathbb{R}^n$ which arose from the study of linear maps. For example, given a linear map $F : \mathbb{R}^n \to \mathbb{R}^m$, we could find its image $\mathrm{im}(F) \subset \mathbb{R}^m$. In the case $m = 3$, we saw that the image could be everything ($\mathbb{R}^3$), a plane, a line, or just contains only one point $\mathbf{0} \in \mathbb{R}^3$.

These sets are examples of subspaces, in which if you take any two vectors from a subspace, their sum and any scalar multiple of them also remain within the same subspace. The definition of subspaces is given as follows.

Definition 8.1 A subset $U \subset \mathbb{R}^n$ is a **subspace** of $\mathbb{R}^n$ if
(i) $\mathbf{0} \in U$,
(ii) for all $u, v \in U$ we have $u + v \in U$,
(iii) for all $u \in U$ and $\lambda \in \mathbb{R}$ we have $\lambda u \in U$.

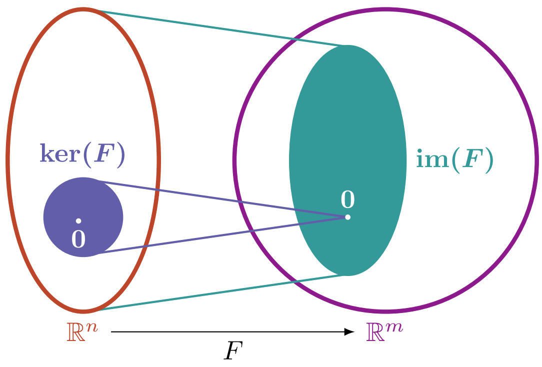
Example 29 1) $U = \{\mathbf{0}\}$ and $U = \mathbb{R}^n$ are always subspaces of $\mathbb{R}^n$ for all $n \geq 1$.
2) General subspaces of $\mathbb{R}^n$ for $n = 1,\ 2,\ 3$ are given as follows.
$n = 1$: $\{0\}$, $\mathbb{R}$.
$n = 2$: $\{\mathbf{0}\}$, $\mathbb{R}^2$,
$\{\lambda v \in \mathbb{R}^2 \mid \lambda \in \mathbb{R}\}$ for any $v \in \mathbb{R}^2$, $v \neq \mathbf{0}$.
$n = 3$: $\{\mathbf{0}\}$, $\mathbb{R}^3$,
$\{\lambda v \in \mathbb{R}^3 \mid \lambda \in \mathbb{R}\}$ for any $v \in \mathbb{R}^3$, $v \neq \mathbf{0}$,
$\{\lambda_1 v + \lambda_2 u \in \mathbb{R}^3 \mid \lambda_1, \lambda_2 \in \mathbb{R}\}$ for any $u, v \in \mathbb{R}^3$, $u, v \neq \mathbf{0}$.
3) $U = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3 \,\middle|\, x_1 + x_2 - 3x_3 = 4 \right\} \subset \mathbb{R}^3$ is not a subspace because $\mathbf{0} \notin U$.
4) $U = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3 \,\middle|\, -1 \leq x_1 \leq 1 \right\} \subset \mathbb{R}^3$ is also not a subspace because we have $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \in U$ but $2\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \notin U$.

A lot of subspaces come from linear maps (actually all of them). We will see that the image of a linear map is a subspace. Another subspace coming from a linear map is its **kernel**.

Definition 8.2 For a linear map $F : \mathbb{R}^n \to \mathbb{R}^m$ the **kernel of** F is defined by

$$\ker(F) = \{x \in \mathbb{R}^n \mid F(x) = \mathbf{0}\}.$$

In other words, the kernel of a linear map F is the set of all solutions to the linear system $[F]x = \mathbf{0}$. The following figure illustrates the kernel and the image of a linear map.



Proposition 8.3 *For any linear map $F : \mathbb{R}^n \to \mathbb{R}^m$ we have the following:*
(i) The kernel $\ker(F)$ *is a subspace of* $\mathbb{R}^n$*.*
(ii) The image $\operatorname{im}(F)$ *is a subspace of* $\mathbb{R}^m$*.*

Proof. To show that a subset U is a subspace, we need to check the 3 conditions from Definition 8.1.

(i) $\ker(F)$ is a subspace of $\mathbb{R}^n$ because it satisfies the following conditions.

(a) $\mathbf{0} \in \ker(F)$ because $F(\mathbf{0}) = [F]\mathbf{0} = \mathbf{0}$.

(b) For any $u, v \in \ker(F)$, we have $F(u + v) = F(u) + F(v) = \mathbf{0} + \mathbf{0} = \mathbf{0}$.
Therefore, $u + v \in \ker(F)$.

(c) For any $u \in \ker(F)$ and $\lambda \in \mathbb{R}$, we have $F(\lambda u) = \lambda F(u) = \lambda \cdot \mathbf{0} = \mathbf{0}$. Thus, $\lambda u \in \ker(F)$.

(ii) $\operatorname{im}(F)$ is a subspace of $\mathbb{R}^m$ because it satisfies the following conditions.

(a) We have $\mathbf{0} \in \operatorname{im}(F)$ since $F(\mathbf{0}) = \mathbf{0}$.

(b) Let $u, v \in \operatorname{im}(F)$. In other words, $u = F(x)$ and $v = F(y)$ for some $x, y \in \mathbb{R}^n$.
Then we have $u + v = F(x) + F(y) = F(x + y)$. Because $x + y \in \mathbb{R}^n$, we have $u + v \in \operatorname{im}(F)$.

(c) Let $u \in \operatorname{im}(F)$, $u = F(x)$ for some $x \in \mathbb{R}^n$, and $\lambda \in \mathbb{R}$.
We have $\lambda u = \lambda F(x) = F(\lambda x)$, which implies that $\lambda u \in \operatorname{im}(F)$. □

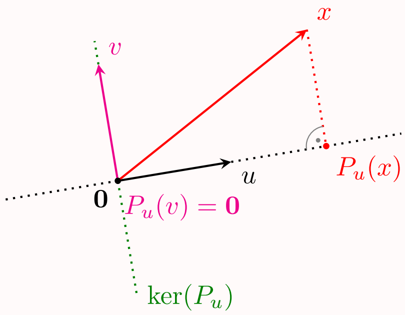
Remark. Actually, every subspace can be written as the kernel and the image of some linear maps. However, we cannot prove this yet.

Example 30 1) Let $u \in \mathbb{R}^n$ and $u \neq \mathbf{0}$. Consider the orthogonal projection $P_u : \mathbb{R}^n \to \mathbb{R}^n$.

The kernel of P_u is given by all vector $v \in \mathbb{R}^n$ such that $u \bullet v = 0$ because $u \neq \mathbf{0}$ and

$$P_u(v) = \frac{u \bullet v}{u \bullet u} u = \mathbf{0} \;\Leftrightarrow\; \frac{u \bullet v}{u \bullet u} = 0 \;\Leftrightarrow\; u \bullet v = 0.$$

Hence, $\ker(P_u) = \{v \in \mathbb{R}^n \mid u \bullet v = 0\}$.
For $n = 2$, $\ker(P_u)$ is a line.
For $n = 3$, $\ker(P_u)$ is a plane.



2) Consider the linear map

$$\begin{aligned} F : \mathbb{R}^2 &\longrightarrow \mathbb{R}^3 \\ x &\longmapsto \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 0 & 1 \end{pmatrix} x \end{aligned}$$

<u>Kernel:</u> We have $x \in \ker(F) \Leftrightarrow F(x) = \mathbf{0}$. Therefore, we need to find solutions to the linear system $[F]x = \mathbf{0}$.

$$\big([F] \mid \mathbf{0}\big) = \left(\begin{array}{cc|c} 1 & 1 & 0 \\ 2 & 1 & 0 \\ 0 & 1 & 0 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 1 & 0 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right)$$

Hence, $x = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \mathbf{0}$ and therefore, $\ker(F) = \{\mathbf{0}\}$.

<u>Image:</u> To calculate the image of F, we need to check for which $y \in \mathbb{R}^3$ there exist at least one $x \in \mathbb{R}^2$ with $F(x) = y$.

$$\big([F] \mid y\big) = \left(\begin{array}{cc|c} 1 & 1 & y_1 \\ 2 & 1 & y_2 \\ 0 & 1 & y_3 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 1 & y_1 \\ 0 & -1 & y_2 - y_1 \\ 0 & 1 & y_3 \end{array}\right)$$

$$\sim \left(\begin{array}{cc|c} 1 & 1 & y_1 \\ 0 & -1 & -2y_1 + y_2 \\ 0 & 0 & -2y_1 + y_2 + y_3 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & -y_1 + y_2 \\ 0 & 1 & 2y_1 - y_2 \\ 0 & 0 & -2y_1 + y_2 + y_3 \end{array}\right)$$

Thus, $\operatorname{im}(F) = \left\{ \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \in \mathbb{R}^3 \;\middle|\; y_3 = 2y_1 - y_2 \right\}$. Notice that this calculation can also be used to calculate the kernel by setting $y_1 = y_2 = y_3 = 0$. Besides, for any $y \in \operatorname{im}(F)$, we can set $y_1 = \lambda_1$, $y_2 = \lambda_2$, and write

$$y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ 2\lambda_1 - \lambda_2 \end{pmatrix} = \lambda_1 \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}.$$

Therefore, $\operatorname{im}(F) = \left\{ \lambda_1 \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \in \mathbb{R}^3 \;\middle|\; \lambda_1, \lambda_2 \in \mathbb{R} \right\}$.

3) Consider the linear map

$$\begin{aligned} G : \mathbb{R}^4 &\longrightarrow \mathbb{R}^2 \\ x &\longmapsto \begin{pmatrix} 1 & 2 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} x \end{aligned}$$

Kernel: We solve the linear system $[G]x = \mathbf{0}$.

$$\big([G] \mid \mathbf{0}\big) = \left(\begin{array}{cccc|c} 1 & 2 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 \end{array}\right) \sim \left(\begin{array}{cccc|c} 1 & 2 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 \end{array}\right) \sim \left(\begin{array}{cccc|c} 1 & 0 & 2 & 1 & 0 \\ 0 & 1 & -1 & 0 & 0 \end{array}\right)$$

Hence, the solution is $\begin{cases} x_1 = -2t_1 - t_2 \\ x_2 = t_1 \\ x_3 = t_1 \\ x_4 = t_2 \end{cases}$, for $t_1, t_2 \in \mathbb{R}$.
Another way of writing this solution is

$$x = t_1 \begin{pmatrix} -2 \\ 1 \\ 1 \\ 0 \end{pmatrix} + t_2 \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \quad \text{for } t_1, t_2 \in \mathbb{R}.$$

Therefore, $\ker(G) = \left\{ t_1 \begin{pmatrix} -2 \\ 1 \\ 1 \\ 0 \end{pmatrix} + t_2 \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \in \mathbb{R}^4 \;\middle|\; t_1, t_2 \in \mathbb{R} \right\}$.

Image: We check for which $y \in \mathbb{R}^2$ that the linear system $[G]x = y$ has solutions.

$$\big([G] \mid y\big) = \left(\begin{array}{cccc|c} 1 & 2 & 0 & 1 & y_1 \\ 1 & 1 & 1 & 1 & y_2 \end{array}\right) \sim \left(\begin{array}{cccc|c} 1 & 2 & 0 & 1 & y_1 \\ 0 & -1 & 1 & 0 & y_2 - y_1 \end{array}\right)$$
$$\sim \left(\begin{array}{cccc|c} 1 & 0 & 2 & 1 & 3y_1 - 2y_2 \\ 0 & 1 & -1 & 0 & y_1 - y_2 \end{array}\right).$$

Since $\operatorname{rk}\big([G] \mid y\big) = \operatorname{rk}\big([G]\big)$ for any $y \in \mathbb{R}^2$, the linear system $[G]x = y$ has solutions for any $y \in \mathbb{R}^2$. Hence, $\operatorname{im}(G) = \mathbb{R}^2$.

In Example 30, we see that the sets containing all sums of multiples of some vectors appear frequently when we try to determine the kernel and the image of a linear map.

Definition 8.4 (i) A **linear combination** of vectors $v_1, \dots, v_n \in \mathbb{R}^m$ is a vector of the form

$$u = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n \in \mathbb{R}^m$$

for some numbers $\lambda_1, \dots, \lambda_n \in \mathbb{R}$.

(ii) The **span** of $v_1, \dots, v_n \in \mathbb{R}^m$ is the set of all linear combinations

$$\operatorname{span}\{v_1, \dots, v_n\} = \{\lambda_1 v_1 + \cdots + \lambda_n v_n \in \mathbb{R}^m \mid \lambda_1, \dots, \lambda_n \in \mathbb{R}\}.$$

Example 31 1) $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ is a linear combination of $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix}$ since $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = 2\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix}$.

2) Every vector $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ is a linear combination of $e_1, \dots, e_n$ because

$$x = x_1 e_1 + \ldots + x_n e_n.$$

Therefore, $\mathbb{R}^n = \operatorname{span}\{e_1, \dots, e_n\}$.

3) From Example 30, we have

$$\operatorname{im}(F) = \operatorname{span}\left\{\begin{pmatrix}1\\0\\2\end{pmatrix}, \begin{pmatrix}0\\1\\-1\end{pmatrix}\right\},$$
$$\ker(G) = \operatorname{span}\left\{\begin{pmatrix}-2\\1\\1\\0\end{pmatrix}, \begin{pmatrix}-1\\0\\0\\1\end{pmatrix}\right\}.$$

Remark. Given a linear map

$$F : \mathbb{R}^n \longrightarrow \mathbb{R}^m,$$
$$x \longmapsto Ax,$$

with $A = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix}$, we can always write $\operatorname{im}(F) = \operatorname{span}\{v_1, \ldots, v_n\}$.

Proposition 8.5 *For $v_1, \ldots, v_n \in \mathbb{R}^m$ we have the following.*
(i) $\operatorname{span}\{v_1, \ldots, v_n\}$ *is a subspace of* $\mathbb{R}^m$.
(ii) If $U \subset \mathbb{R}^m$ is a subspace and $v_1, \ldots, v_n \in U$ then $\operatorname{span}\{v_1, \ldots, v_n\} \subset U$.

Proof. The proof is left as Exercise 25. □

Recall that a linear map $F : \mathbb{R}^n \to \mathbb{R}^m$ is surjective if and only if $\operatorname{im}(F) = \mathbb{R}^m$. A similar statement exists for injective functions as follows.

Proposition 8.6 *A linear map $F : \mathbb{R}^n \to \mathbb{R}^m$ is injective if and only if* $\ker(F) = \{\mathbf{0}\}$.

Proof. The proof is left as Exercise 20. □

We also know that F is injective if each column of $\operatorname{rref}([F])$ contains a pivot element, i.e., $\operatorname{rk}(F) = n$. Similarly, we know that F is surjective if each row of $\operatorname{rref}([F])$ contains a pivot element, i.e., $\operatorname{rk}(F) = m$. Summarizing everything, we get the following theorem.

Theorem 8.7 *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map.*
(i) We have the following equivalent statements for F being injective:

$$F \text{ is injective} \iff \ker(F) = \{\mathbf{0}\} \iff \operatorname{rk}([F]) = n\,.$$

(ii) We have the following equivalent statements for F being surjective:

$$F \text{ is surjective} \iff \operatorname{im}(F) = \mathbb{R}^m \iff \operatorname{rk}([F]) = m\,.$$

(iii) If $\underline{m = n}$ then the following statements are equivalent:

$$F \text{ is bijective} \iff F \text{ is injective} \iff F \text{ is surjective}\,.$$

Exercises

Exercise 25. Show the following without using Proposition 8.5:

(i) For $v_1, \ldots, v_n \in \mathbb{R}^m$ the set $\operatorname{span}\{v_1, \ldots, v_n\}$ is a subspace of $\mathbb{R}^m$.
(ii) If $U \subset \mathbb{R}^m$ is a subspace and $v_1, \ldots, v_n \in U$ then $\operatorname{span}\{v_1, \ldots, v_n\} \subset U$.

Exercise 26. Which of the following subsets are subspaces? Justify your answers.

$$\begin{aligned}
U_1 &= \left\{x \in \mathbb{R}^3 \mid x_1 + x_2 + x_3 = 0\right\} , \\
U_2 &= \left\{x \in \mathbb{R}^3 \mid x_1 \cdot x_2 \cdot x_3 = 0\right\} , \\
U_3 &= \{x \in \mathbb{R}^n \mid Ax = Bx\} , \quad \text{where } A, B \in \mathbb{R}^{m \times n}, \\
U_4 &= \left\{x \in \mathbb{R}^2 \mid x_1 \leq x_2\right\} .
\end{aligned}$$

Exercise 27.

(i) Which of the following subsets are subspaces? Justify your answers.

$$\begin{aligned}
U_1 &= \left\{x \in \mathbb{R}^3 \mid x_1 - x_2 = x_3\right\} , \\
U_2 &= \left\{x \in \mathbb{R}^2 \mid x_1^2 - x_2^2 = 0\right\} , \\
U_3 &= \{x \in \mathbb{R}^n \mid Ax = -2x\} , \quad \text{where } A \in \mathbb{R}^{n \times n} \text{ is a fixed matrix}, \\
U_4 &= \{x \in \mathbb{R}^n \mid x \bullet v = 0\} , \quad \text{for a fixed } v \in \mathbb{R}^n.
\end{aligned}$$

(ii) For each subset U in (i) which is a subspace, find numbers $a, b \geq 1$ and a linear map $F : \mathbb{R}^a \to \mathbb{R}^b$ such that $\ker(F) = U$.

Exercise 28.

(i) Which of the following subsets are subspaces? Justify your answers.

$$\begin{aligned}
U_1 &= \left\{x \in \mathbb{R}^3 \mid \|x\| \leq 1\right\} , \\
U_2 &= \left\{x \in \mathbb{R}^3 \mid 2x_1 - x_2 = x_3\right\} , \\
U_3 &= \left\{x \in \mathbb{R}^3 \mid x_1^2 - x_1 = 0\right\} , \\
U_4 &= \left\{x \in \mathbb{R}^4 \mid P_u(x) = \mathbf{0}\right\} , \quad \text{where } u = \begin{pmatrix} 1 \\ 0 \\ 2 \\ -1 \end{pmatrix} .
\end{aligned}$$

(ii) For the subspaces U in (i): Find vectors $v_1, \ldots, v_l$ such that $U = \operatorname{span}\{v_1, \ldots, v_l\}$.

(Challenge for (ii): Try to choose the $v_1, \ldots, v_l$ such that they are pairwise orthogonal and all of them have norm 1. Such a basis is called orthonormal basis and we will study them in Chapter 12)

Exercise 29. Consider the following subspace

$$W = \ker(P_u) = \left\{x \in \mathbb{R}^3 \mid P_u(x) = \mathbf{0}\right\} , \quad \text{where } u = \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} .$$

(i) Determine vectors $v_1, \ldots, v_m \in \mathbb{R}^3$ with $W = \operatorname{span}\{v_1, \ldots, v_m\}$.
(ii) Find a linear map $H : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ with $\operatorname{im}(H) = W$.
(iii) Calculate $\ker(H)$ and $\ker(P_u \circ H)$.

Exercise 30.
(i) Let $U, V \subset \mathbb{R}^m$ be subspaces. Decide whether the union $U \cup V$ is also a subspace or not.
(ii) Let $U, V \subset \mathbb{R}^m$ be subspaces. Decide whether the intersection $U \cap V$ is also a subspace or not.

Exercise 31.
(i) Decide if the following two linear maps are invertible. Determine their inverses if they exist.

$$F : \mathbb{R}^3 \longrightarrow \mathbb{R}^3\,, \qquad \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \longmapsto \begin{pmatrix} 2x_1 - x_2 \\ -4x_1 + x_2 - x_3 \\ 6x_1 - 2x_2 + x_3 \end{pmatrix}\,, \qquad G : \mathbb{R}^3 \longrightarrow \mathbb{R}^3\,, \qquad \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \longmapsto \begin{pmatrix} -x_1 + x_3 \\ -4x_1 + x_2 - x_3 \\ 6x_1 - 2x_2 + x_3 \end{pmatrix}\,.$$

(ii) Determine $\ker(F)$ and $\ker(G)$.

Exercise 32. Find an example of a subset $U \subset \mathbb{R}^2$ which is <u>not</u> a subspace, but which
(i) includes $\mathbf{0}$ and which is closed under addition.
(ii) includes $\mathbf{0}$ and which is closed under scalar multiplication.
In other words: Find examples of subsets, which just satisfy 2 of the 3 conditions for subspaces.

9

Linear independence

In Example 30, we considered the linear map

$$\begin{aligned} G : \mathbb{R}^4 &\longrightarrow \mathbb{R}^2 \\ x &\longmapsto \begin{pmatrix} 1 & 2 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} x \end{aligned}$$

and determined its image

$$\operatorname{im}(G) = \mathbb{R}^2 = \operatorname{span}\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}.$$

But we also learned that

$$\begin{aligned} \operatorname{im}(G) &= \text{ span of columns of } [G] \\ &= \operatorname{span}\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}. \end{aligned}$$

This shows that

$$\operatorname{span}\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} = \operatorname{span}\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}.$$

From the left-hand side, we just need 2 vectors to span $\operatorname{im}(G)$. Meanwhile, on the right-hand side, there are too many vectors, so we want to remove some extra vectors. In general, given a subspace as a span of some vectors, we may wonder about the minimum number of vectors and which vectors we need to keep to describe this subspace. In order to answer that question, we first need the following lemma.

Lemma 9.1 *Let $v_1, \dots, v_l \in \mathbb{R}^m$. If $v_l \in \operatorname{span}\{v_1, \dots, v_{l-1}\}$ then*

$$\operatorname{span}\{v_1, \dots, v_l\} = \operatorname{span}\{v_1, \dots, v_{l-1}\}.$$

Proof. Set $V = \operatorname{span}\{v_1, \dots, v_l\}$ and $W = \operatorname{span}\{v_1, \dots, v_{l-1}\}$. Clearly, we have $W \subset V$, so we want to show $V \subset W$. If $v \in \operatorname{span}\{v_1, \dots, v_l\} = V$, then there exist $\lambda_1, \dots, \lambda_l \in \mathbb{R}$ with

$$v = \lambda_1 v_1 + \ldots + \lambda_l v_l. \tag{$*$}$$

Since $v_l \in \operatorname{span}\{v_1, \dots, v_{l-1}\}$, there also exist $\alpha_1, \dots, \alpha_{l-1} \in \mathbb{R}$ with

$$v_l = \alpha_1 v_1 + \ldots + \alpha_{l-1} v_{l-1}. \tag{$**$}$$

Combining $(*)$ and $(**)$, we have

$$\begin{aligned} v &= \lambda_1 v_1 + \ldots + \lambda_{l-1} v_{l-1} + \lambda_l(\alpha_1 v_1 + \ldots + \alpha_{l-1} v_{l-1}) \\ &= (\lambda_1 + \lambda_l \alpha_1) v_1 + \ldots + (\lambda_{l-1} + \lambda_l \alpha_{l-1}) v_{l-1} \end{aligned}$$

And therefore $v \in \operatorname{span}\{v_1, \ldots, v_{l-1}\} = W$, i.e. $V \subset W$. As a result, $V = W$. $\square$

Example 32 For the linear map G in Example 30, we get

$$\operatorname{im}(G) = \operatorname{span}\left\{\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}2\\1\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}\right\}.$$

We can show directly that $\operatorname{im}(G) = \mathbb{R}^2$ by using Lemma 9.1 without solving the linear system $[G]x = y$. Since $\begin{pmatrix}2\\1\end{pmatrix} = 2\begin{pmatrix}1\\1\end{pmatrix} - \begin{pmatrix}0\\1\end{pmatrix} \in \operatorname{span}\left\{\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\right\}$ and also $\begin{pmatrix}1\\1\end{pmatrix} \in \operatorname{span}\left\{\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\right\}$, by Lemma 9.1, we get

$$\operatorname{im}(G) = \operatorname{span}\left\{\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\right\}.$$

Since $\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}1\\1\end{pmatrix} - \begin{pmatrix}0\\1\end{pmatrix}$, we get (again by Lemma 9.1)

$$\operatorname{im}(G) = \operatorname{span}\left\{\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\right\}.$$

Since $\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}1\\0\end{pmatrix} + \begin{pmatrix}0\\1\end{pmatrix}$, we get (again by Lemma 9.1)

$$\operatorname{im}(G) = \operatorname{span}\left\{\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\right\} = \mathbb{R}^2.$$

As a result,

$$\operatorname{span}\left\{\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}2\\1\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}\right\} = \operatorname{span}\left\{\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\right\}.$$

General question: When is it possible to remove elements from $\operatorname{span}\{v_1, \ldots, v_l\}$ without changing it?

Definition 9.2 (i) Vectors $v_1, \ldots, v_l \in \mathbb{R}^m$ are called **linearly independent** if the equation

$$\lambda_1 v_1 + \cdots + \lambda_l v_l = 0 \tag{9.0.1}$$

with $\lambda_1, \ldots, \lambda_l \in \mathbb{R}$ just has the unique solution $\lambda_1 = \cdots = \lambda_l = 0$.

(ii) If there exist another solution of (9.0.1), i.e. where at least for one $j = 1, \ldots, l$ we have $\lambda_j \neq 0$, then the vectors $v_1, \ldots, v_l$ are called **linearly dependent**.

Example 33 Are the vectors $v_1 = \begin{pmatrix}1\\1\\2\end{pmatrix}$, $v_2 = \begin{pmatrix}1\\-1\\1\end{pmatrix}$, and $v_3 = \begin{pmatrix}-1\\5\\1\end{pmatrix}$ linearly independent?

The equation $\lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3 = \mathbf{0}$ is equivalent to

$$\begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix}\begin{pmatrix}\lambda_1\\\lambda_2\\\lambda_3\end{pmatrix} = \begin{pmatrix}1 & 1 & -1\\1 & -1 & 5\\2 & 1 & 1\end{pmatrix}\begin{pmatrix}\lambda_1\\\lambda_2\\\lambda_3\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix}.$$

To solve this linear system, we need to find the row-reduced echelon form of the augmented matrix:

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

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

Hence, the solution is $\begin{cases} \lambda_1 = -2t \\ \lambda_2 = 3t \\ \lambda_3 = t \end{cases}$, for $t \in \mathbb{R}$. Therefore, v_1, v_2, v_3 are linearly dependent since there are non-zero solutions.
For $t = 1$, we get $\lambda_1 = -2$, $\lambda_2 = 3$, $\lambda_3 = 1$. Thus, we get

$$-2v_1 + 3v_2 + v_3 = -2\begin{pmatrix}1\\1\\2\end{pmatrix} + 3\begin{pmatrix}1\\-1\\1\end{pmatrix} + \begin{pmatrix}-1\\5\\1\end{pmatrix} = \mathbf{0}$$

$$\Rightarrow v_3 = 2v_1 - 3v_2.$$

Therefore, $v_3 \in \operatorname{span}\{v_1, v_2\}$ and by Lemma 9.1 $\operatorname{span}\{v_1, v_2, v_3\} = \operatorname{span}\{v_1, v_2\}$. On the other hand, v_1 and v_2 are linearly independent since

$$\left(\begin{array}{cc|c} | & | & 0 \\ v_1 & v_2 & 0 \\ | & | & 0 \end{array}\right) = \left(\begin{array}{cc|c} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 2 & 1 & 0 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right),$$

which shows that $\mathbf{0}$ is the only solution and hence, $\lambda_1 v_1 + \lambda_2 v_2 = \mathbf{0} \iff \lambda_1 = \lambda_2 = 0$. As a result, we cannot remove v_1 or v_2 without changing the span.

We can summarize in the following theorem what we have just done in Example 33.

Theorem 9.3 *Let $v_1, \ldots, v_l \in \mathbb{R}^m$. The following statements are equivalent:*

(i) $v_1, \ldots, v_l$ are linearly dependent.

(ii) There exists a $j = 1, \ldots, l$, such that v_j is a linear combination of the other vectors.

(iii) There exists a $j = 1, \ldots, l$ with

$$\operatorname{span}\{v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_l\} = \operatorname{span}\{v_1, \ldots, v_l\}.$$

Proof. The statement (ii) $\Rightarrow$ (iii) is Lemma 9.1. Now we prove the statement (iii) $\Rightarrow$ (ii) as follows:

$$v_j \in \operatorname{span}\{v_1, \ldots, v_l\} \overset{\text{(iii)}}{=} \operatorname{span}\{v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_l\} \quad \Rightarrow \quad v_j \in \operatorname{span}\{v_1, \ldots, \cancel{v_j}, \ldots, v_l\}.$$

Next, for proving the statement (ii) $\Rightarrow$ (i), suppose $v_j = \lambda_1 v_1 + \ldots + \lambda_{j-1} v_{j-1} + \lambda_{j+1} v_{j+1} + \ldots + \lambda_l v_l$. Then,

$$0 = \lambda_1 v_1 + \ldots + \lambda_{j-1} v_{j-1} - 1 \cdot v_j + \lambda_{j+1} v_{j+1} + \ldots + \lambda_l v_l.$$

Hence, $v_1, \ldots, v_l$ are linearly dependent.

Finally, for proving the statement (i) $\Rightarrow$ (ii), suppose $\lambda_1 v_1 + \ldots + \lambda_l v_l = 0$ with $\lambda_j \neq 0$. Then,

$$v_j = \left(\frac{\lambda_1}{\lambda_j}\right) v_1 + \ldots + \left(\frac{\lambda_{j-1}}{\lambda_j}\right) v_{j-1} + \left(\frac{\lambda_{j+1}}{\lambda_j}\right) v_{j+1} + \ldots + \left(\frac{\lambda_l}{\lambda_j}\right) v_l.$$

Thus, $v_j \in \operatorname{span}\{v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_l\}$. □

The following lemma shows that in a subspace, we cannot have more linearly independent vectors than the vectors spanning the subspace.

Lemma 9.4 *Let $V \subset \mathbb{R}^n$ be a subspace, $v_1, \dots, v_l \in V$ linearly independent and $V = \operatorname{span}\{w_1, \dots, w_m\}$ for some $w_1, \dots, w_m \in \mathbb{R}^n$. Then we have $l \leq m$.*

Proof. The proof is left as Exercise 33. □

The following lemma shows how to make a set of linearly independent vectors bigger.

Lemma 9.5 *If $v_1, \dots, v_l \in \mathbb{R}^n$ are linearly independent and $w \in \mathbb{R}^n$ with $w \notin \operatorname{span}\{v_1, \dots, v_l\}$ then $v_1, \dots, v_l, w$ are linearly independent.*

Proof. Assume that $\lambda_1 v_1 + \ldots + \lambda_l v_l + \mu w = 0$. If $\mu \neq 0$, then

$$w = \left(-\frac{\lambda_1}{\mu}\right) v_1 + \ldots + \left(-\frac{\lambda_l}{\mu}\right) v_l \in \operatorname{span}\{v_1, \dots, v_l\},$$

which contradicts to the assumption. Hence, $\mu = 0$, which gives $\lambda_1 v_1 + \ldots + \lambda_l v_l = 0$.

Then, $\lambda_1 = \ldots = \lambda_l = 0$ because $v_1, \dots, v_l$ are linearly independent. Thus, $v_1, \dots, v_l$, and w are linearly independent. □

Exercises

Exercise 33. Let $V \subset \mathbb{R}^n$ be a subspace, $v_1, \dots, v_l \in V$ linearly independent and $V = \operatorname{span}\{w_1, \dots, w_m\}$ for some $w_1, \dots, w_m \in \mathbb{R}^n$. Show that we have $l \leq m$. (Without using Lemma 9.4)
In other words: Show that a subspace spanned by m vectors can not contain more than m linearly independent vectors.

10

Bases & dimensions

We saw that if $v_1, \ldots, v_l \in \mathbb{R}^n$ are linearly dependent, then there exist a $1 \leq j \leq l$ such that

$$\operatorname{span}\{v_1, \ldots, v_l\} = \operatorname{span}\{v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_l\}.$$

Therefore, we will be just interested in the case when $v_1, \ldots, v_l$ are linearly independent.

Definition 10.1 Let $V \subset \mathbb{R}^n$ be a subspace. Vectors $v_1, \ldots, v_l \in V$ form a **basis of V** if
(i) $V = \operatorname{span}\{v_1, \ldots, v_l\}$,
(ii) $v_1, \ldots, v_l$ are linearly independent.
In this case we also say that $\{v_1, \ldots, v_l\}$ is a basis of V.

Later we will also be interested in the order of the v_j and write a basis as a tuple $(v_1, \ldots, v_l)$.

Example 34 1) $\{e_1, e_2\}$, $\left\{e_1, \begin{pmatrix}1\\1\end{pmatrix}\right\}$, $\left\{e_2, \begin{pmatrix}1\\1\end{pmatrix}\right\}$ are three different bases of $\mathbb{R}^2$.

2) Consider $U = \operatorname{span}\left\{\begin{pmatrix}1\\1\\2\end{pmatrix}, \begin{pmatrix}1\\-1\\1\end{pmatrix}, \begin{pmatrix}-1\\5\\1\end{pmatrix}\right\} = \operatorname{span}\{v_1, v_2, v_3\}$. We want to find a basis for the subspace U. We have seen in Example 33:
(a) v_1, v_2, v_3 are linearly dependent because $-2v_1 + 3v_2 + v_3 = \mathbf{0}$.
(b) $v_3 \in \operatorname{span}\{v_1, v_2\}$ and hence, $U = \operatorname{span}\{v_1, v_2\}$ by Lemma 9.1.
(c) v_1 and v_2 are linearly independent.
Therefore, $\{v_1, v_2\}$ is a basis of U.

3) For any $n \geq 1$, $\{e_1, \ldots, e_n\}$ is a basis of $\mathbb{R}^n$. This basis is called the **standard basis**.

We will see that a basis of a subspace is a convenient tool for us to study the subspace because we can represent every vector of the subspace in terms of vectors in the basis. By working with bases, we can focus on understanding a smaller set of vectors rather than dealing with the entire space, enabling us to analyze, manipulate, and comprehend subspaces with greater ease. Fortunately, the following theorem shows that bases always exist.

Theorem 10.2 *For any subspace $V \subset \mathbb{R}^n$ we have the following:*
(i) V has a basis.
(ii) All bases of V have the same number of elements.
(iii) If $v_1, \ldots, v_l \in V$ are linearly independent then there exist $u_{l+1}, \ldots, u_t \in V$, such that $\{v_1, \ldots, v_l, u_{l+1}, \ldots, u_t\}$ is a basis of V.
(iv) If $V = \operatorname{span}\{w_1, \ldots, w_m\}$ then there exists a subset $\{u_1, \ldots, u_t\} \subset \{w_1, \ldots, w_m\}$ such that $\{u_1, \ldots, u_t\}$ is a basis of V.

Proof. (ii) Let $\{v_1, \dots, v_l\}$ and $\{w_1, \dots, w_m\}$ be bases of V. Then, we have by Lemma 9.4,

$$v_1, \dots, v_l \in V \text{ are linearly independent and } V = \operatorname{span}\{w_1, \dots, w_m\} \quad \Rightarrow \quad l \le m,$$

$$w_1, \dots, w_m \in V \text{ are linearly independent and } V = \operatorname{span}\{v_1, \dots, v_l\} \quad \Rightarrow \quad m \le l.$$

Therefore, we have $l = m$.

(iv) 1. If $w_1, \dots, w_m$ are linearly independent, then $\{u_1, \dots, u_t\} = \{w_1, \dots, w_m\}$ is a basis of V.

2. Otherwise, if $\{w_1, \dots, w_m\}$ are linearly dependent, then By Theorem 9.3 there exist a $j = 1, \dots, m$ with $\operatorname{span}\{w_1, \dots, w_{j-1}, w_{j+1}, \dots, w_m\} = \operatorname{span}\{w_1, \dots, w_m\} = V$.

Now repeat 1. and 2. with $\operatorname{span}\{w_1, \dots, w_{j-1}, w_{j+1}, \dots, w_m\}$, i.e. remove vectors like w_j until when the remaining vectors are linearly independent. Eventually, we will get a basis $\{u_1, \dots, u_t\}$ of V.

(iii) Assume that $v_1, \dots, v_l \in V$ are linearly independent.

1. If $\operatorname{span}\{v_1, \dots, v_l\} = V$, then $\{v_1, \dots, v_l\}$ is a basis of V.

2. Otherwise, if $\operatorname{span}\{v_1, \dots, v_l\} \ne V$, then there exists $u \in V$ with $u \notin \operatorname{span}\{v_1, \dots, v_l\}$. In that case, we set $u_{l+1} = u$. By Lemma 9.5, $v_1, \dots, v_l, u_{l+1}$ are linearly independent.

Repeat 1. and 2. for $\{v_1, \dots, v_l, u_{l+1}\}$ until when $V = \operatorname{span}\{v_1, \dots, v_l, u_{l+1}, \dots, u_t\}$.

(i) If $V = \{\mathbf{0}\}$ then $\emptyset = \{\}$ is a basis because $\operatorname{span}(\emptyset) = \{\mathbf{0}\}$ by convention. Otherwise, we can construct a basis using (iii). □

Definition 10.3 Let $V \subset \mathbb{R}^n$ be a subspace. The **dimension of V**, denoted by $\dim(V)$, is the number of elements in a basis of V.

Example 35 1) $\dim(\mathbb{R}^n) = n$ because $\{e_1, \dots, e_n\}$ is a basis of $\mathbb{R}^n$.

2) The dimension of $U = \operatorname{span}\left\{\begin{pmatrix}1\\1\\2\end{pmatrix}, \begin{pmatrix}1\\-1\\1\end{pmatrix}, \begin{pmatrix}-1\\5\\1\end{pmatrix}\right\} = \operatorname{span}\{v_1, v_2, v_3\}$ is $\dim(U) = 2$ because $\{v_1, v_2\}$ is a basis.

Corollary 10.4 *Let $V \subset \mathbb{R}^n$ be a subspace with $\dim(V) = m$ and $v_1, \dots, v_m \in V$. Then the following statements are equivalent:*
(i) $v_1, \dots, v_m$ are linearly independent.
(ii) $V = \operatorname{span}\{v_1, \dots, v_m\}$.
(iii) $\{v_1, \dots, v_m\}$ is a basis of V.

Proof. **(i) ⇒ (ii):** If $V \ne \operatorname{span}\{v_1, \dots, v_m\}$, then by Theorem 10.2 (iii), there exists a basis with more than m elements, which contradicts to the assumption that $\dim(V) = m$. Hence, $V = \operatorname{span}\{v_1, \dots, v_m\}$.

(ii) ⇒ (i): If $v_1, \dots, v_m$ are linearly dependent, then by Theorem 9.3 and Theorem 10.2 (iv), there exists a basis with less than m elements, which contradicts to the assumption that $\dim(V) = m$. Thus, $v_1, \dots, v_m$ are linearly independent.

(i) + **(ii)** $\iff$ **(iii)** by definition 10.1. $\square$

Example 36 Determine bases for $\ker(F)$ and $\text{im}(F)$ of the following linear map:

$$\begin{aligned} F : \mathbb{R}^4 &\longrightarrow \mathbb{R}^3, \\ x &\longmapsto \begin{pmatrix} 1 & -2 & 1 & 2 \\ 2 & -4 & 1 & 3 \\ 0 & 0 & 1 & 1 \end{pmatrix} x. \end{aligned}$$

Kernel: We need to find $\text{rref}([F] \mid \mathbf{0})$ to solve the linear system $[F]x = \mathbf{0}$. Because the column corresponding to $\mathbf{0}$ does not change after row operations, we just need to find $\text{rref}([F])$.

$$\begin{pmatrix} 1 & -2 & 1 & 2 \\ 2 & -4 & 1 & 3 \\ 0 & 0 & 1 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & -2 & 1 & 2 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & -2 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \text{rref}([F]).$$

Hence,

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \in \ker(F) \iff [F]x = \mathbf{0} \iff \begin{cases} x_1 = 2t_1 - t_2 \\ x_2 = t_1 \\ x_3 = -t_2 \\ x_4 = t_2 \end{cases}, \quad \text{for } t_1, t_2 \in \mathbb{R}$$

$$\iff x = t_1 \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t_2 \begin{pmatrix} -1 \\ 0 \\ -1 \\ 1 \end{pmatrix} = t_1 v_1 + t_2 v_2, \quad \text{for } t_1, t_2 \in \mathbb{R},$$

where $v_1 = \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix}$ and $v_2 = \begin{pmatrix} -1 \\ 0 \\ -1 \\ 1 \end{pmatrix}$. Therefore, $\ker(F) = \text{span}\{v_1, v_2\}$.

Next, we need to check whether v_1 and v_2 are linearly independent or not.

$$\mathbf{0} = t_1 v_1 + t_2 v_2 = \begin{pmatrix} 2t_1 - t_2 \\ t_1 \\ -t_2 \\ t_2 \end{pmatrix} \implies t_1 = t_2 = 0.$$

Thus, v_1 and v_2 are linearly independent and $\{v_1, v_2\}$ is a basis of $\ker(F)$.

Image: We have $[F] = \text{span}\left\{ \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ -4 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} \right\} = \text{span}\{u_1, u_2, u_3, u_4\}$.

$$\lambda_1 u_1 + \lambda_2 u_2 + \lambda_3 u_3 + \lambda_4 u_4 = \mathbf{0} \iff [F] \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{pmatrix} = \mathbf{0} \iff \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{pmatrix} \in \ker(F).$$

Based on the result we got above,

$$\lambda_1 u_1 + \lambda_2 u_2 + \lambda_3 u_3 + \lambda_4 u_4 = \mathbf{0} \implies \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{pmatrix} = t_1 \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t_2 \begin{pmatrix} -1 \\ 0 \\ -1 \\ 1 \end{pmatrix}, \quad \text{for } t_1, t_2 \in \mathbb{R}.$$

When $t_1 = 1$ and $t_2 = 0$, we have

$$2u_1 + u_2 = \mathbf{0} \quad \Longleftrightarrow \quad u_2 = -2u_1 \quad \Longrightarrow \quad u_2 \in \operatorname{span}\{u_1, u_3\}.$$

When $t_1 = 0$ and $t_2 = 1$,

$$-u_1 - u_3 + u_4 = \mathbf{0} \quad \Longleftrightarrow \quad u_4 = -u_1 + u_3 \quad \Longrightarrow \quad u_3 \in \operatorname{span}\{u_1, u_3\}.$$

Hence, $\operatorname{im}(F) = \operatorname{span}\{u_1, u_3\}$. In addition, u_1 and u_3 are linear independent because when $\lambda_1 u_1 + \lambda_3 u_3 = \mathbf{0}$:

$$\left(\begin{array}{cc|c} | & | & 0 \\ u_1 & u_3 & 0 \\ | & | & 0 \end{array}\right) = \left(\begin{array}{cc|c} 1 & 1 & 0 \\ 2 & 1 & 0 \\ 0 & 1 & 0 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right) \quad \Longrightarrow \quad \lambda_1 = \lambda_3 = 0.$$

Thus, $\{u_1, u_3\}$ is a basis of $\operatorname{im}(F)$.

From Example 36, we can summarize the general calculation of bases for $\ker(F)$ and $\operatorname{im}(F)$ as follows:

Consider a linear map

$$\begin{aligned} F : \mathbb{R}^n &\longrightarrow \mathbb{R}^m, \\ x &\longmapsto Ax. \end{aligned}$$

- Let $\operatorname{rref}(A)$ have pivot elements in columns $c_1, \ldots, c_r$. Then, the columns $c_1, \ldots, c_r$ in the original matrix A form a basis of $\operatorname{im}(F)$. Hence, the dimension of $\operatorname{im}(F)$ is equal to the number of pivot elements in $\operatorname{rref}(A)$ or equal to $\operatorname{rk}(F)$.

$$\dim(\operatorname{im}(F)) = \text{ the number of pivot elements in } \operatorname{rref}(A).$$

- The vectors obtained in the "standard parametrization" (i.e. for each free variable x_i there is a parameter t_j) of the solutions to $F(x) = \mathbf{0}$ form a basis of $\ker(F)$. Thus, the dimension of $\ker(F)$ is equal to the number of free variables.

$$\dim(\ker(F)) = \text{ the number of free variables.}$$

As a consequence, we get the following theorem:

Theorem 10.5 *For a linear map $F : \mathbb{R}^n \to \mathbb{R}^m$ we have*

$$n = \dim(\ker(F)) + \dim(\operatorname{im}(F)).$$

Proof. The statement follows from the following facts:

$$\begin{aligned} n &= \text{ the number of columns of } [F], \\ \dim(\ker(F)) &= \text{ the number of columns without pivot elements.} \\ \dim(\operatorname{im}(F)) &= \text{ the number of columns with pivot elements.} \end{aligned}$$

□

Exercises

Exercise 34. Determine a basis of the following subspace

$$U = \operatorname{span}\left\{ \begin{pmatrix}1\\2\\3\end{pmatrix}, \begin{pmatrix}3\\3\\8\end{pmatrix}, \begin{pmatrix}1\\3\\2\end{pmatrix} \right\}.$$

Exercise 35. Determine bases for the kernel and the image of the following linear map

$$\begin{aligned} F : \mathbb{R}^5 &\longrightarrow \mathbb{R}^3 \\ x &\longmapsto \begin{pmatrix} 2 & -6 & -1 & 5 & 3 \\ -1 & 3 & 2 & -4 & -3 \\ 1 & -3 & -1 & 3 & 4 \end{pmatrix} x\,. \end{aligned}$$

Exercise 36. Let $U, V \subset \mathbb{R}^n$ be two subspaces. We define their sum by

$$U + V := \{x \in \mathbb{R}^n \mid \text{there exist } u \in U, v \in V \text{ with } x = u + v\}\,.$$

(i) Show that $U + V$ is a subspace of $\mathbb{R}^n$.
(ii) Show that we have

$$\dim(U + V) = \dim(U) + \dim(V) - \dim(U \cap V)\,.$$

Exercise 37. For $t \in \mathbb{R}$ we define

$$v_1 = \begin{pmatrix}1\\1\\2\end{pmatrix}, \quad v_2 = \begin{pmatrix}2\\2\\t\end{pmatrix}, \quad v_3 = \begin{pmatrix}t\\4\\(t-2)^2\end{pmatrix}$$

and set $V = \operatorname{span}\{v_1, v_2, v_3\}$. For each $t \in \mathbb{R}$ determine a basis of V and calculate its dimension.

Exercise 38. The following exercise is intended to show the basic idea of 3D computer graphics, by showing how to get a 2-dimensional picture (to be shown on a 2-dimensional monitor) from an 3-dimensional object.

(i) We define the corners of a cube with side length 18 in $\mathbb{R}^3$ by the following set of 8 points:

$$W = \left\{ \begin{pmatrix}w_1\\w_2\\w_3\end{pmatrix} \in \mathbb{R}^3 \mid w_1, w_2, w_3 \in \{0, 18\} \right\}.$$

Make a drawing of a cube with side length 18 in $\mathbb{R}^3$, i.e. draw the 8 points in the set W and connect two points if they differ just by one entry.

(This just means that you draw a cube like you would usually draw it. "Differ by one entry" just means that these points are on the same edge of the cube.)

(ii) Show that $D = (d_1, d_2, d_3)$ is a basis of $\mathbb{R}^3$, where

$$d_1 = \begin{pmatrix}2\\1\\0\end{pmatrix}, \quad d_2 = \begin{pmatrix}-1\\1\\3\end{pmatrix}, \quad d_3 = \begin{pmatrix}3\\-6\\3\end{pmatrix}.$$

(iii) Write each $x \in W$ as a linear combination in the basis D, i.e. for each x find $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$

with

$$x = \lambda_1 d_1 + \lambda_2 d_2 + \lambda_3 d_3 \,.$$

(iv) For each $x \in W$ draw the points (λ_1, λ_2) in $\mathbb{R}^2$. Connect two points if the corresponding elements in W just differ by one entry.

Explanation: What you should get in iv) is a drawing of the 3-dimensional cube in 2 dimensions. The basis D somehow describes from which direction you look at the cube. If you replaced the D by the standard basis (e_1, e_2, e_3), you would get a picture of the cube from the top (i.e., just a square). The λ_3, which you did not use for the drawing, describes the distance in the viewing direction.

11

Coordinates

From now on we will consider ordered bases, which means that we will write $(b_1, \ldots, b_n)$ (a tuple) for a basis instead of $\{b_1, \ldots, b_n\}$ (a set). The difference is, that we care about the order now. For example, the two sets $\{b_1, b_2\} = \{b_2, b_1\}$ are the same, but $(b_1, b_2) \neq (b_2, b_1)$.

Definition 11.1 Let $B = (b_1, \ldots, b_m)$ be a basis of a subspace $V \subset \mathbb{R}^n$. We define the **coordinate map** by

$$\begin{aligned} c_B : \mathbb{R}^m &\longrightarrow \mathbb{R}^n \\ \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{pmatrix} &\longmapsto \lambda_1 b_1 + \cdots + \lambda_m b_m \, . \end{aligned}$$

Clearly, the map c_B is a linear map. In addition, it is also a bijection, as stated in the following theorem.

Theorem 11.2 *Let $B = (b_1, \ldots, b_m)$ be a basis of a subspace $V \subset \mathbb{R}^n$.*
(i) The coordinate map $c_B : \mathbb{R}^m \longrightarrow V$ is bijective.
(ii) For all $x \in V$ there exist unique $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ such that

$$x = \lambda_1 b_1 + \cdots + \lambda_m b_m \, .$$

Proof. (i) The map $c_B : \mathbb{R}^m \longrightarrow V$ is surjective since $\mathrm{im}(c_B) = \mathrm{span}\{b_1, \ldots, b_m\} = V$.

Additionally,

$$\lambda = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{pmatrix} \in \ker(c_B) \quad \Longleftrightarrow \quad c_B(\lambda) = \lambda_1 b_1 + \ldots + \lambda_m b_m = \mathbf{0} \quad \Longrightarrow \quad \lambda = \mathbf{0},$$

where the implication comes from the assumption that $b_1, \ldots, b_m$ are linearly independent. Hence, $\ker(c_B) = \{\mathbf{0}\}$ and, by Theorem 8.7, c_B is injective.

As a result, the map c_B is bijective.

(ii) This is just a reformulation of (i). $\square$

As a consequence of Theorem 11.2, we can define **coordinates** and **coordinate vector** of any vectors in a subspace V as follows.

Definition 11.3 Let $B = (b_1, \dots, b_m)$ be a basis of a subspace $V \subset \mathbb{R}^n$ and $x \in V$ with

$$x = \lambda_1 b_1 + \cdots + \lambda_m b_m \,.$$

(i) The numbers $\lambda_1, \dots, \lambda_m \in \mathbb{R}$ are the **coordinates of x (in the basis B)**.
(ii) The **coordinate vector** of x (with respect to B) is given by

$$[x]_B = c_B^{-1}(x) = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{pmatrix} .$$

Example 37 1) $B = (e_1, \dots, e_n)$ is a basis of $\mathbb{R}^n$. For all $x \in \mathbb{R}^n$, we have $[x]_B = x$.

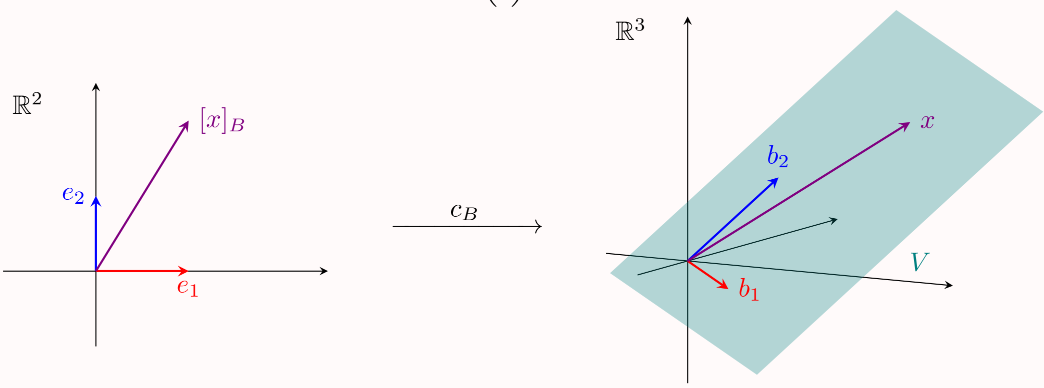
2) Consider $b_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$ and $b_2 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$. Clearly, b_1 and b_2 are linearly independent; therefore,

$B = (b_1, b_2)$ is a basis of $V = \operatorname{span}\{b_1, b_2\}$. Is it true that $x = \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix} \in V$? What is $[x]_B$?

In order to answer those questions, we need to solve $\lambda_1 b_1 + \lambda_2 b_2 = x$.

$$\left(\begin{array}{cc|c} | & | & | \\ b_1 & b_2 & x \\ | & | & | \end{array}\right) = 1 \left(\begin{array}{cc|c} 1 & 1 & 3 \\ 0 & 2 & 4 \\ -1 & 3 & 5 \end{array}\right) \sim \tfrac{1}{2}\ (-2) \left(\begin{array}{cc|c} 1 & 1 & 3 \\ 0 & 2 & 4 \\ 0 & 4 & 8 \end{array}\right) \sim (-1) \left(\begin{array}{cc|c} 1 & 1 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{array}\right)$$

Hence, $x = 1 \cdot b_1 + 2 \cdot b_2 \quad \Longrightarrow \quad [x]_B = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$.



Coordinate vectors have the following properties.

Proposition 11.4 *Let $B = (b_1, \dots, b_m)$ be a basis of a subspace $V \subset \mathbb{R}^n$. Then we have for all $x, y \in V$ and $\mu \in \mathbb{R}$*
(i) $[x + y]_B = [x]_B + [y]_B$,
(ii) $[\mu x]_B = \mu [x]_B$,
(iii) $[0]_B = 0$.

Proof. Firstly, we have $[x]_B = c_B^{-1}(x)$. In addition, c_B^{-1} is linear since c_B is linear (Proposition 7.2). Hence, all the properties follow from the linearity of c_B^{-1}. $\square$

Since c_B is a linear map, we can consider its matrix as follows.

Definition 11.5 Let $B = (b_1, \dots, b_n)$ be a basis of $\mathbb{R}^n$. The **change-of-basis matrix** associated with B is

$$S_B = [c_B] = \begin{pmatrix} | & & | \\ b_1 & \dots & b_n \\ | & & | \end{pmatrix}.$$

Remark. (a) S_B is invertible for any basis B of $\mathbb{R}^n$ since $S_B \in \mathbb{R}^{n\times n}$ and the linear system $S_B\lambda = \mathbf{0}$ or $\lambda_1 b_1 + \ldots + \lambda_n b_n = \mathbf{0}$ has a unique solution $x = \mathbf{0}$.

(b) For any $x \in \mathbb{R}^n$,

$$S_B[x]_B = c_B\big([x]_B\big) = c_B\big(c_B^{-1}(x)\big) = x.$$

The following definition may be a little complicated and confusing at first reading, but it will be useful later for the study of linear maps.

Definition 11.6 Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map, B_1 be a basis of $\mathbb{R}^n$ and B_2 be a basis of $\mathbb{R}^m$. The **matrix of** F **with respect to** B_1 **and** B_2 is the matrix

$$[F]_{B_1}^{B_2} := [c_{B_2}^{-1} \circ F \circ c_{B_1}].$$

In the case $n = m$ and $B_1 = B_2$ we just write $[F]_{B_1} := [F]_{B_1}^{B_1}$.

$$\begin{tikzcd} \mathbb{R}^n \arrow[r, "F"] & \mathbb{R}^m \arrow[d, "c_{B_2}^{-1}"] \\ \mathbb{R}^n \arrow[u, "c_{B_1}"] \arrow[r, "c_{B_2}^{-1}\circ F\circ c_{B_1}"'] & \mathbb{R}^m \end{tikzcd}$$

With this definition, we get the following proposition, which gives more insight into the definition of a matrix with respect to some bases.

Proposition 11.7 *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map, B_1 be a basis of $\mathbb{R}^n$ and B_2 be a basis of $\mathbb{R}^m$.*

(i) We have

$$[F]_{B_1}^{B_2} = S_{B_2}^{-1}[F]S_{B_1}.$$

(ii) If $B_1 = (b_1, \dots, b_n)$ then

$$[F]_{B_1}^{B_2} = \begin{pmatrix} | & & | \\ [F(b_1)]_{B_2} & \dots & [F(b_n)]_{B_2} \\ | & & | \end{pmatrix}.$$

Proof. (i) $[F]_{B_1}^{B_2} = [c_{B_2}^{-1} \circ F \circ c_{B_1}] = [c_{B_2}^{-1}][F][c_{B_1}] = S_{B_2}^{-1}[F]S_{B_1}$.

(ii) The i^{th} column of $[F]_{B_1}^{B_2}$ is

$$[F]_{B_1}^{B_2}e_i = c_{B_2}^{-1} \circ F \circ c_{B_1}(e_i) = c_{B_2}^{-1}\Big(F\big(c_{B_1}(e_i)\big)\Big) = c_{B_2}^{-1}\big(F(b_i)\big) = [F(b_i)]_{B_2}.$$ □

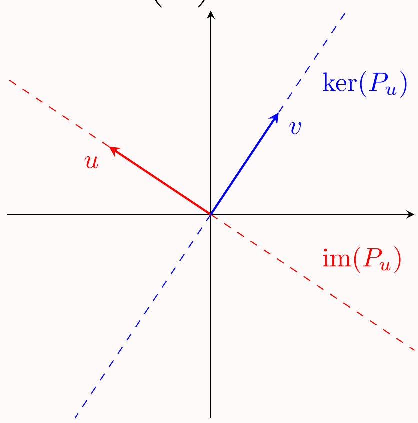
In many cases, we can write down $[F]_B$ with respect to some bases B much easier than $[F]$. After that, we can use Proposition 11.7 to obtain $[F]$.

Example 38

1) Consider $B_1 = (e_1, \dots, e_n)$ in $\mathbb{R}^n$ and $B_2 = (e_1, \dots, e_m)$ in $\mathbb{R}^m$.
   Then, $[c_{B_1}] = I_n$ and $[c_{B_2}] = I_m$. For any linear map $F : \mathbb{R}^n \to \mathbb{R}^m$, it is always true that
$$[F]_{B_1}^{B_2} = [F].$$

2) Consider the orthogonal projection $P_u : \mathbb{R}^2 \to \mathbb{R}^2$ with $u = \begin{pmatrix} -3 \\ 2 \end{pmatrix}$.

   We have
$$\begin{aligned} \ker(P_u) &= \{x \in \mathbb{R}^2 \mid u \bullet x = 0\} \\ &= \operatorname{span}\left\{\begin{pmatrix} 2 \\ 3 \end{pmatrix}\right\} = \operatorname{span}\{v\}, \end{aligned}$$
$$\operatorname{im}(P_u) = \operatorname{span}\{u\}.$$



   We also have
$$\lambda_1 u + \lambda_2 v = \mathbf{0} \quad \Rightarrow \quad \begin{cases} u \bullet (\lambda_1 u + \lambda_2 v) = 0 \\ v \bullet (\lambda_1 u + \lambda_2 v) = 0 \end{cases} \quad \Rightarrow \quad \begin{cases} \lambda_1 (u \bullet u) = 0 \\ \lambda_2 (v \bullet v) = 0 \end{cases} \quad \Rightarrow \quad \lambda_1 = \lambda_2 = 0.$$

   Hence, u and v are linearly independent, so $B = (u, v)$ is a basis of $\mathbb{R}^2$. Notice: this technique will be used again in the next chapter.

   We have $P_u(u) = u$ and $P_u(v) = \mathbf{0}$, so
$$[P_u]_B = [P_u]_B^B = \begin{pmatrix} | & | \\ [P_u(u)]_B & [P_u(v)]_B \\ | & | \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

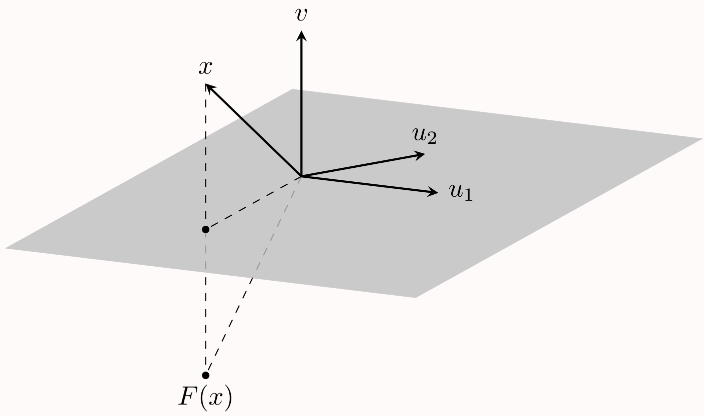
   Change-of-basis matrix is
$$S_B = \begin{pmatrix} | & | \\ u & v \\ | & | \end{pmatrix} = \begin{pmatrix} -3 & 2 \\ 2 & 3 \end{pmatrix}$$
   and its inverse is
$$S_B^{-1} = \frac{1}{13}\begin{pmatrix} -3 & 2 \\ 2 & 3 \end{pmatrix}.$$

   By Proposition 11.7 we have $[P_u]_B = [P_u]_B^B = S_B^{-1}[P_u]S_B$. Hence,
$$[P_u] = S_B [P_u]_B S_B^{-1} = \begin{pmatrix} -3 & 2 \\ 2 & 3 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\frac{1}{13}\begin{pmatrix} -3 & 2 \\ 2 & 3 \end{pmatrix} = \frac{1}{13}\begin{pmatrix} 9 & -6 \\ -6 & 4 \end{pmatrix}.$$

3) Let $F : \mathbb{R}^3 \to \mathbb{R}^3$ be the reflection through the plane
$$U = \operatorname{span}\left\{\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}\right\} = \operatorname{span}\{u_1, u_2\}.$$

   We want to determine the matrix $[F]$. Ideas for how to solve this problem are as follows.

(a) Find a "good" basis B where we can write down $[F]_B$ directly.
(b) For this, try to find $v \in \mathbb{R}^3$ which is orthogonal to u_1, u_2, and set $B = (u_1, u_2, v)$. Notice that u_1 and u_2 are linearly independent (check it!).



To find $v = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \in \mathbb{R}^3$ such that $v \bullet u_1 = v \bullet u_2 = 0$, we need to solve $\begin{cases} v_1 + v_2 = 0 \\ v_2 + v_3 = 0 \end{cases}$.

$$\begin{array}{r} \\ -1 \end{array} \left(\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 \end{array}\right)$$

Hence, all solutions are given by $t \cdot \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$ for $t \in \mathbb{R}$. We choose $t = 1$, i.e. set $v = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$.
We also have

$$\begin{aligned} \lambda_1 u_1 + \lambda_2 u_2 + \lambda_3 v = \mathbf{0} \quad &\Rightarrow \quad \begin{cases} \lambda_1 u_1 + \lambda_2 u_2 + \lambda_3 v = \mathbf{0} \\ v \bullet (\lambda_1 u_1 + \lambda_2 u_2 + \lambda_3 v) = 0 \end{cases} \quad \Rightarrow \begin{cases} \lambda_1 u_1 + \lambda_2 u_2 + \lambda_3 v = \mathbf{0} \\ \lambda_3 (v \bullet v) = 0 \end{cases} \\ &\Rightarrow \quad \begin{cases} \lambda_1 u_1 + \lambda_2 u_2 = \mathbf{0} \\ \lambda_3 = 0 \end{cases} \quad \Rightarrow \quad \lambda_1 = \lambda_2 = \lambda_3 = 0. \end{aligned}$$

Thus, u_1, u_2, v are linearly independent, so $B = (u_1, u_2, v)$ is a basis of $\mathbb{R}^3$.
We have

$$F(u_1) = u_1 \quad \Rightarrow \quad [F(u_1)]_B = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$F(u_2) = u_2 \quad \Rightarrow \quad [F(u_2)]_B = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

$$F(v) = -v \quad \Rightarrow \quad [F(v)]_B = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}$$

$$\Rightarrow \quad [F]_B = \begin{pmatrix} | & | & | \\ [F(u_1)]_B & [F(u_2)]_B & [F(v)]_B \\ | & | & | \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

In addition, the change-of-basis matrix is

$$S_B = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & -1 \\ 0 & 1 & 1 \end{pmatrix}$$

and its inverse is

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

Therefore,

$$[F] = S_B[F]_B S_B^{-1} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & -1 \\ 0 & 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}\frac{1}{3}\begin{pmatrix} 2 & 1 & 1 \\ -1 & 1 & 2 \\ 1 & -1 & 1 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 1 & 2 & -2 \\ 2 & 1 & 2 \\ -2 & 2 & 1 \end{pmatrix}.$$

Exercises

Exercise 39. Let $U = \operatorname{span}\{u_1, u_2, u_3, u_4\} \in \mathbb{R}^4$, where

$$u_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 1 \end{pmatrix}, \quad u_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad u_3 = \begin{pmatrix} 1 \\ 3 \\ 2 \\ 2 \end{pmatrix}, \quad u_4 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}.$$

(i) Determine a basis $B = (b_1, \ldots, b_m)$ of U.

(ii) Calculate the coordinate vectors $[u_j]_B \in \mathbb{R}^m$ for $j = 1, 2, 3, 4$.

(iii) For which values of $a \in \mathbb{R}$ does the vector $x = \begin{pmatrix} 3 \\ 3+a \\ 3+a \\ 2+2a \end{pmatrix}$ belong to U? For such a determine the coordinate vector $[x]_B$.

12

Orthonormal bases & Gram-Schmidt algorithm

In this chapter, we will discuss a special type of basis for a subspace. Before introducing any concepts in this chapter, let us recall some notations that were introduced in Definition 5.1 of Chapter 5:

(i) If $u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}, v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^n$, then the **dot product** of u and v is defined by

$$u \bullet v = u_1 v_1 + \cdots + u_n v_n \,.$$

(ii) $u, v \in \mathbb{R}^n$ are called **orthogonal** if $u \bullet v = 0$.

(iii) The **norm** of $u \in \mathbb{R}^n$ is defined by $\|u\| = \sqrt{u \bullet u}$.

In addition, the norm and the dot product have some useful properties as stated in the following proposition.

Proposition 12.1 *Let $x, y \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$.*
(i) $\|\lambda x\| = |\lambda| \cdot \|x\|$.
(ii) $|x \bullet y| \leq \|x\| \cdot \|y\|$ *(Cauchy-Schwartz inequality).*
The equality " $=$ " occurs when x, y are linearly dependent.
(iii) $\|x + y\| \leq \|x\| + \|y\|$ *(Triangle inequality).*
The equality " $=$ " also occurs for this inequality when x, y are linearly dependent.

Proof. Using the properties of the dot product from Proposition 5.2, we have

(i) $\|\lambda x\| = \sqrt{(\lambda x) \bullet (\lambda x)} = \sqrt{\lambda^2 (x \bullet x)} = |\lambda| \sqrt{x \bullet x} = |\lambda| \cdot \|x\|$.

(ii) For any $\mu \in \mathbb{R}$, we have

$$\begin{aligned} (x + \mu y) \bullet (x + \mu y) \geq 0 \quad &\Leftrightarrow \quad x \bullet x + (\mu y) \bullet x + x \bullet (\mu y) + (\mu y) \bullet (\mu y) \geq 0 \\ &\Leftrightarrow \quad \|x\|^2 + 2(x \bullet (\mu y)) + \|\mu y\|^2 \geq 0 \\ &\Leftrightarrow \quad \|x\|^2 + 2\mu(x \bullet y) + \mu^2 \|y\|^2 \geq 0. \end{aligned}$$

If $y = \mathbf{0}$, then the statement is trivial because $|x \bullet \mathbf{0}| = \|x\| \cdot \|\mathbf{0}\| = 0$. Therefore, we assume $y \neq \mathbf{0}$. Choosing $\mu = -(x \bullet y)/\|y\|^2$, we have

$$\|x\|^2 - 2\frac{x \bullet y}{\|y\|^2}(x \bullet y) + \frac{(x \bullet y)^2}{\|y\|^4}\|y\|^2 \geq 0 \quad \Leftrightarrow \quad \|x\|^2\|y\|^2 \geq (x \bullet y)^2 \quad \Leftrightarrow \quad \|x\| \cdot \|y\| \geq |x \bullet y|.$$

(iii) We have

$$\|x + y\|^2 = (x + y) \bullet (x + y) = \|x\|^2 + 2(x \bullet y) + \|y\|^2.$$

Using the Cauchy-Schwartz inequality, we get

$$\begin{aligned} & \|x + y\|^2 \leq \|x\|^2 + 2\|x\| \cdot \|y\| + \|y\|^2 = (\|x\| + \|y\|)^2 \\ \Leftrightarrow \quad & \|x + y\| \leq \|x\| + \|y\|, \end{aligned}$$

where the equivalence comes from the fact that $\|x + y\|$ and $\|x\| + \|y\|$ are both non-negative. □

Definition 12.2 (i) A vector $u \in \mathbb{R}^n$ is called a **unit vector** if $\|u\| = 1$. (i.e. $u \bullet u = 1$)

(ii) Every vector $u \in \mathbb{R}^n$ with $u \neq \mathbf{0}$ can be normalized by

$$\widehat{u} = \frac{1}{\|u\|} u\,.$$

The vector $\hat{u}$ is a unit vector and shows in the same direction as u.

(iii) Vectors $u_1, \ldots, u_l \in \mathbb{R}^n$ are called **orthonormal** if for $1 \leq i, j \leq l$,

$$u_i \bullet u_j = \begin{cases} 1 & \text{, if } i = j \\ 0 & \text{, if } i \neq j \end{cases}.$$

Equipped with these definitions, we can define the main object of this chapter as follows.

Definition 12.3 A basis $B = (b_1, \ldots, b_m)$ of a subspace U is called an **orthonormal basis (ONB)** of U if $b_1, \ldots, b_m$ are orthonormal.

Example 39 For the subspace $U = \mathbb{R}^n$, the standard basis $B = (e_1, \ldots, e_n)$ is an orthonormal basis.

The following proposition demonstrates some useful properties of orthonormal bases.

Proposition 12.4 *(i) If $v_1, \ldots, v_m \in \mathbb{R}^n$ are orthonormal (ON), then they are linearly independent.*

(ii) Let $B = (v_1, \ldots, v_m)$ be an ONB of $V \subset \mathbb{R}^n$ and $u \in V$. Then

$$[u]_B = \begin{pmatrix} u \bullet v_1 \\ \vdots \\ u \bullet v_m \end{pmatrix} \in \mathbb{R}^m\,,$$

i.e. $u = \sum_{i=1}^m (u \bullet v_i) v_i$.

(iii) If $B = (v_1, \ldots, v_m)$ is an ONB of $V \subset \mathbb{R}^n$ and $u, w \in V$, then

$$u \bullet w = [u]_B \bullet [w]_B\,.$$

Proof. (i) Assume that $v_1, \ldots, v_m$ are ON and $\lambda_1 v_1 + \ldots + \lambda_m v_m = \mathbf{0}$. For any $1 \leq j \leq m$, we have

$$v_j \bullet (\lambda_1 v_1 + \ldots + \lambda_m v_m) = 0 = \lambda_1 (v_j \bullet v_1) + \ldots + \lambda_m (v_j \bullet v_m) = \lambda_j.$$

Hence, $\lambda_1 = \ldots = \lambda_m = 0$, which shows that $v_1, \ldots, v_m$ are linearly independent.

(ii) Since $u \in V$, we can write $u = \lambda_1 v_1 + \ldots + \lambda_m v_m$. Doing the same calculations as in (i), we get

$$\text{For } 1 \le j \le m, \quad u \bullet v_j = \lambda_j \quad \Rightarrow \quad u = (u \bullet v_1)v_1 + \ldots + (u \bullet v_m)v_m \quad \Rightarrow \quad [u]_B = \begin{pmatrix} u \bullet v_1 \\ \vdots \\ u \bullet v_m \end{pmatrix}.$$

(iii) Let $[u]_B = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}$, $[w]_B = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$. We have

$$u \bullet w = (x_1 v_1 + \ldots + x_m v_m) \bullet (y_1 v_1 + \ldots + y_m v_m) = x_1 y_1 + \ldots + x_m y_m = [u]_B \bullet [w]_B. \qquad \square$$

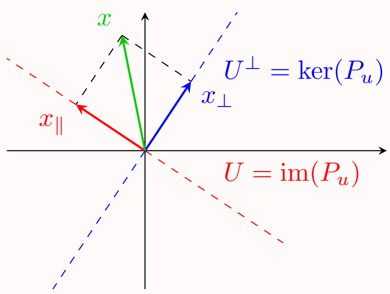
Now we want to introduce another important concept in the following definition.

Definition 12.5 For a subspace $U \subset \mathbb{R}^n$ we define the **orthogonal complement of U in $\mathbb{R}^n$** by

$$U^\perp = \{x \in \mathbb{R}^n \mid x \bullet u = 0 \text{ for all } u \in U\}\ .$$

Example 40 Consider the orthogonal projection $P_u : \mathbb{R}^2 \to \mathbb{R}^2$ with $u = \frac{1}{\sqrt{13}}\begin{pmatrix} -3 \\ 2 \end{pmatrix}$.
Recall that $P_u(x) = x_\parallel = \frac{u \bullet x}{u \bullet u} u$ for any $x \in \mathbb{R}^2$.
In this case, $\|u\| = 1$ so we have

$$P_u(x) = x_\parallel = (u \bullet x)u,$$



for any $x \in \mathbb{R}^2$. Clearly, $\{u\}$ is an ONB of the subspace $U = \mathrm{im}(P_u)$. We also know that $\ker(P_u) = \{x \in \mathbb{R}^2 \mid x \bullet u = 0\}$. Fix an element $v \in \ker(P_u)$. For any $w \in U$, we have $w = \lambda u$ for some $\lambda \in \mathbb{R}$ and

$$v \bullet w = v \bullet (\lambda u) = \lambda(v \bullet u) = 0,$$

which implies that $v \in U^\perp$. Thus, $\ker(P_u) \subset U^\perp$. Furthermore, it is clear that $U^\perp \subset \ker(P_u)$. As a result, $\ker(P_u) = U^\perp$, which shows that $U^\perp$ is a subspace.

Motivated by this example, we get the following lemma.

Lemma 12.6 *Let $U \subset \mathbb{R}^n$ be a subspace.*
(i) $U^\perp \subset \mathbb{R}^n$ is a subspace.
(ii) We have $U \cap U^\perp = \{\mathbf{0}\}$.
(iii) If $(u_1, \ldots, u_r)$ is a basis of U, $x \in \mathbb{R}^n$, then

$$x \in U^\perp \quad \Longleftrightarrow \quad x \bullet u_1 = \cdots = x \bullet u_r = 0\,.$$

(iv) Let $(f_1, \ldots, f_r)$ be an ONB of U and $x \in \mathbb{R}^n$. Then

$$x = x_\parallel + x_\perp\,,$$

where

$$x_{\|} = \sum_{i=1}^{r} (x \bullet f_i) f_i \in U$$
$$x_{\perp} = x - x_{\|} \in U^{\perp} .$$

Proof. (i) Clearly, $\mathbf{0} \in U^{\perp}$ since $\mathbf{0} \bullet u = 0$ for any $u \in U$. Given any $x, y \in U^{\perp}$ and $\lambda \in \mathbb{R}$, we have for all $u \in U$,

$$(x+y) \bullet u = x \bullet u + y \bullet u = 0 + 0 = 0 \quad \Rightarrow \quad x + y \in U^{\perp},$$

$$(\lambda x) \bullet u = \lambda(x \bullet u) = \lambda \cdot 0 = 0 \quad \Rightarrow \quad \lambda x \in U^{\perp}.$$

Therefore, $U^{\perp}$ is a subspace of $\mathbb{R}^n$.

(ii) If $x \in U \cap U^{\perp}$, then $x \bullet x = 0 = x_1^2 + \ldots + x_n^2 \quad \Rightarrow \quad x_1 = \ldots = x_n = 0 \quad \Rightarrow \quad x = \mathbf{0}$.

(iii) " $\Longrightarrow$ " is clear.

" $\Longleftarrow$ " : For all $w \in U$, we have $w = \lambda_1 u_1 + \ldots + \lambda_m u_m$ for some $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$. Then, we have

$$x \bullet w = x \bullet (\lambda_1 u_1 + \ldots + \lambda_m u_m) = \lambda_1 (x \bullet u_1) + \ldots + \lambda_m (x \bullet u_m) = 0 + \ldots + 0 = 0.$$

Hence, $x \in U^{\perp}$.

(iv) $x_{\|} = \sum_{i=1}^{r} (x \bullet f_i) f_i \in U$ since $(f_1, \ldots, f_r)$ is a basis of U. We want to show that $x_{\perp} \in U^{\perp}$.

For all $1 \leq j \leq r$:

$$f_j \bullet x_{\perp} = f_j \bullet (x - x_{\|}) = f_j \bullet x - f_j \bullet \sum_{i=1}^{r} (x \bullet f_i) f_i = f_j \bullet x - x \bullet f_j = 0.$$

Hence, $x_{\perp} \in U^{\perp}$ by using (iii). □

We see that orthonormal bases are extremely useful for many calculations, so we may be concerned about how to get them. Fortunately, the following algorithm allows us to obtain an orthonormal basis from an arbitrary basis of any subspace.

Algorithm 12.7 (Gram-Schmidt algorithm (GSA)) Let $B = (b_1, \ldots, b_m)$ be an arbitrary basis of a subspace $U \subset \mathbb{R}^n$. The GSA constructs an orthonormal basis $F = (f_1, \ldots, f_m)$ of U out of the basis B in the following m steps:

Step 1: Set $f_1 = \widehat{b_1} = \frac{1}{\|b_1\|} b_1$.

Step l $(2 \leq l \leq m)$: We have constructed orthonormal vectors $f_1, \ldots, f_{l-1}$ in the steps before. Now set

$$w_l = b_l - (b_l \bullet f_1) f_1 - \cdots - (b_l \bullet f_{l-1}) f_{l-1} = b_l - \sum_{i=1}^{l-1} (b_l \bullet f_i) f_i$$

and define $f_l = \widehat{w_l} = \frac{1}{\|w_l\|} w_l$.

Example 41 Consider $b_1 = \begin{pmatrix}1\\1\\1\end{pmatrix}$, $b_2 = \begin{pmatrix}1\\0\\2\end{pmatrix}$, $b_3 = \begin{pmatrix}-1\\2\\1\end{pmatrix}$. $B = (b_1, b_2, b_3)$ is a basis of $\mathbb{R}^3$. We will construct an ONB $F = (f_1, f_2, f_3)$ by GSA and demonstrate why it works.

Step 1: Set $f_1 = \widehat{b_1} = \frac{1}{\|b_1\|} b_1 = \frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix}$ and $U_1 = \operatorname{span}\{f_1\} = \{b_1\}$.

Step 2: We want to find a vector $f_2 \in \operatorname{span}\{f_1, b_2\}$ such that f_2 is orthogonal to f_1. From Lemma 12.6 (iv), we have

$$b_2 = b_{2\|} + b_{2\perp} = (b_2 \bullet f_1) f_1 + b_{2\perp},$$

where $b_{2\|} \in U_1$ and $b_{2\perp} \in U_1^{\perp}$.

Set $w_2 = b_{2\perp} = b_2 - (b_2 \bullet f_1) f_1 = \begin{pmatrix}1\\0\\2\end{pmatrix} - \frac{3}{\sqrt{3}} f_1 = \begin{pmatrix}1\\0\\2\end{pmatrix} - \begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}0\\-1\\1\end{pmatrix}$.

Then, set $f_2 = \widehat{w_2} = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\-1\\1\end{pmatrix}$ and $U_2 = \operatorname{span}\{f_1, f_2\} = \operatorname{span}\{b_1, b_2\}$.

Hence, (f_1, f_2) is an ONB of U_2.

Step 3: Now, we want to find a vector $f_3 \in \operatorname{span}\{f_1, f_2, b_3\}$ such that $f_3 \in U_2^{\perp}$. Again, from Lemma 12.6 (iv), we have

$$b_3 = b_{3\|} + b_{3\perp} = (b_3 \bullet f_1) f_1 + (b_3 \bullet f_2) f_2 + b_{3\perp},$$

where $b_3 = b_{3\|} \in U_2$ and $b_{3\perp} \in U_2^{\perp}$.

Set $w_3 = b_{3\perp} = b_3 - (b_3 \bullet f_1) f_1 - (b_3 \bullet f_2) f_2 = \begin{pmatrix}-1\\2\\1\end{pmatrix} - \frac{2}{3}\begin{pmatrix}1\\1\\1\end{pmatrix} + \frac{1}{2}\begin{pmatrix}0\\-1\\1\end{pmatrix} = \frac{5}{6}\begin{pmatrix}-2\\1\\1\end{pmatrix}$.

Then, set $f_3 = \widehat{w_3} = \frac{1}{\sqrt{6}}\begin{pmatrix}-2\\1\\1\end{pmatrix}$ and

$$U_3 = \operatorname{span}\{f_1, f_2, f_3\} = \operatorname{span}\{b_1, b_2, f_3\} = \operatorname{span}\{b_1, b_2, b_3\}.$$

Hence, $B = (f_1, f_2, f_3)$ is an ONB of $U_3 = \mathbb{R}^3$ from that basis.

As a consequence of the Gram-Schmidt algorithm, we get the following theorem.

Theorem 12.8 *Every subspace of $\mathbb{R}^n$ has an ONB.*

Proof. According to Theorem 10.2 (i), every subspace has a basis. Using GSA, we get an ONB. □

Corollary 12.9 *Let $U \subset \mathbb{R}^n$ be a subspace. For all $x \in \mathbb{R}^n$ there exist unique $x_{\|} \in U$ and $x_{\perp} \in U^{\perp}$ with*

$$x = x_{\|} + x_{\perp}.$$

Proof. **Existence:** By Theorem 12.8, there exists an ONB $(f_1, \dots, f_m)$ of U. And by Lemma 12.6 (iv), we get $x_{\|}$ and $x_{\perp}$.

Uniqueness: Let $x = x_\| + x_\perp = y_\| + y_\perp$ for $x_\|, y_\| \in U$ and $x_\perp, y_\perp \in U^\perp$. Then, we have

$$U \ni x_\| - y_\| = y_\perp - x_\perp \in U^\perp.$$

Hence, $x_\| - y_\| \in U \cap U^\perp$ and $y_\perp - x_\perp \in U \cap U^\perp$. However, since $U \cap U^\perp = \{\mathbf{0}\}$ by Lemma 12.6 (ii), we have

$$x_\| - y_\| = y_\perp - x_\perp = \mathbf{0} \quad \Longrightarrow \quad x_\| = y_\| \text{ and } y_\perp = x_\perp. \qquad \square$$

Exercises

Exercise 40. Let $U \subset \mathbb{R}^n$ be a subspace with orthonormal basis $(f_1, \dots, f_r)$. We define the orthogonal projection onto U by

$$\begin{aligned} P_U : \mathbb{R}^n &\longrightarrow \mathbb{R}^n \\ x &\longmapsto \sum_{i=1}^{r} (x \bullet f_i) f_i\,. \end{aligned}$$

Show the following properties of P_U:

(i) If $U = \operatorname{span}\{u\}$ with $u \in \mathbb{R}^n$ and $u \neq \mathbf{0}$ then P_U is the projection P_u we defined in Chapter 5.

(ii) P_U is a linear map.

(iii) $P_U \circ P_U = P_U$.

(iv) $\operatorname{im} P_U = U$ and $\ker(P_U) = U^\perp$, where $U^\perp$ is the orthogonal complement of U defined by

$$U^\perp = \{x \in \mathbb{R}^n \mid x \bullet u = 0 \text{ for all } u \in U\}\ .$$

Exercise 41. We define the following vectors

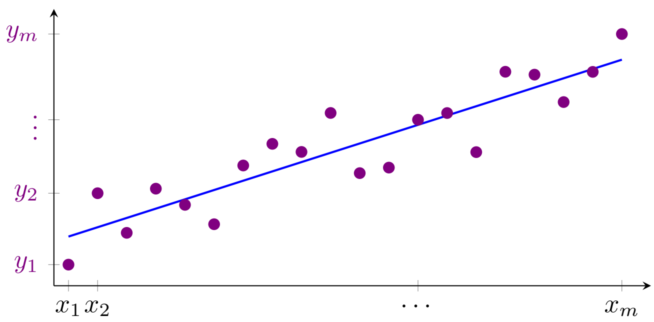
$$b_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \qquad b_2 = \begin{pmatrix} 1 \\ 2 \\ 0 \\ 1 \end{pmatrix}, \qquad b_3 = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 3 \end{pmatrix}.$$

These form a basis $B = (b_1, b_2, b_3)$ of the subspace $U = \operatorname{span}\{b_1, b_2, b_3\} \subset \mathbb{R}^4$ (You do not need to show this). Use the Gram-Schmidt algorithm to construct an orthonormal basis $F = (f_1, f_2, f_3)$ of U from B.

13

Orthogonal Projection & Least squares

Assume you measure some data $(x_1, y_1), \dots, (x_m, y_m)$, and you want to find a *line* which interpolates these points in the best possible way. If all points would lie on a line $\ell(x) = ax + b$, then they would satisfy



$$\begin{cases} ax_1 + b = y_1 \\ ax_2 + b = y_2 \\ \quad\vdots \\ ax_m + b = y_m \end{cases} \iff \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \\ x_m & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \iff A \begin{pmatrix} a \\ b \end{pmatrix} = y \quad (*).$$

However, if they are not on one line (like in the picture), then the linear system $(*)$ has no solutions because $y \notin \operatorname{im}(A)$. Nevertheless, in the picture, we see that there might be a "best possible" line.

This chapter aims to explain how to obtain this "best-fit line." In general, for some matrix $A \in \mathbb{R}^{m\times n}$ and some vector $y \in \mathbb{R}^m$, we want to find a vector $x \in \mathbb{R}^n$ such that Ax is the closest point to y. The main idea is to project y onto the image of A and then obtain a linear system we can solve for x. Later, we will see that x can be obtained by solving the normal equation

$$A^T Ax = A^T y.$$

From Corollary 12.9, for a subspace $U \subset \mathbb{R}^n$ and a vector $x \in \mathbb{R}^n$, there uniquely exist $x_\perp \in U^\perp$ and $x_\parallel \in U$ with $x = x_\perp + x_\parallel$. Hence, we have the following map.

Definition 13.1 Let $U \subset \mathbb{R}^n$ be a subspace. The map

$$\begin{aligned} P_U : \mathbb{R}^n &\longrightarrow \mathbb{R}^n \\ x &\longmapsto x_\parallel \end{aligned}$$

is the **orthogonal projection** onto U.

Remark. This generalizes the P_u for $u \in \mathbb{R}^n,\ u \neq 0$ we defined before by setting $U = \operatorname{span}\{u\}$.

Proposition 13.2 *Let $U \subset \mathbb{R}^n$ be a subspace.*

(i) P_U *is a linear map.*

(ii) $P_U^2 = P_U$.

(iii) $\ker(P_U) = U^\perp$ *and* $\operatorname{im} P_U = U$.

(iv) If $(f_1, \ldots, f_m)$ *is an ONB of* U*, then*

$$P_U(x) = (x \bullet f_1)f_1 + \cdots + (x \bullet f_m)f_m\,.$$

Proof. (iv) is exactly Lemma 12.6 (iv). Using that, we can prove the other statements.

(i) P_U is linear because for any $x,\ y \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$, we have

$$P_U(x+y) = \sum_{i=1}^m \big((x+y) \bullet f_i\big) f_i = \sum_{i=1}^m (x \bullet f_i) f_i + \sum_{i=1}^m (y \bullet f_i) f_i = P_U(x) + P_U(y)$$

$$P_U(\lambda x) = \sum_{i=1}^m \big((\lambda x) \bullet f_i\big) f_i = \lambda \sum_{i=1}^m (x \bullet f_i) f_i = \lambda P_U(x)$$

(ii) For any $x \in \mathbb{R}^n$ and any j such that $1 \leq j \leq m$, we have

$$P_U(x) \bullet f_j = f_j \bullet P_U(x) = f_j \bullet \sum_{i=1}^m (x \bullet f_i) f_i = \sum_{i=1}^m (x \bullet f_i)(f_j \bullet f_i) = x \bullet f_j.$$

Hence, for any $x \in \mathbb{R}^n$, we have

$$P_U \circ P_U(x) = P_U\big(P_U(x)\big) = \sum_{i=1}^m \big(P_U(x) \bullet f_i\big) f_i = \sum_{i=1}^m (x \bullet f_i) f_i = P_U(x).$$

Thus, $P_U^2 = P_U \circ P_U = P_U$.

(iii) For the kernel, we have

$$\begin{aligned}
x \in \ker(P_U) \quad &\Leftrightarrow \quad P_U(x) = \mathbf{0} \\
&\Leftrightarrow \quad \sum_{i=1}^m (x \bullet f_i) f_i = \mathbf{0} \\
&\Leftrightarrow \quad x \bullet f_1 = \ldots = x \bullet f_m = 0 && (f_1, \ldots, f_m \text{ are linearly independent}) \\
&\Leftrightarrow \quad x \in U^\perp && \text{(by Lemma 12.6 (iii))}
\end{aligned}$$

Hence, $\ker(P_U) = U^\perp$.

For the image, if $u \in \operatorname{im}(P_U)$, then clearly $u \in U$ because $u \in \operatorname{span}\{f_1, \ldots, f_m\} = U$. Otherwise, if $u \in U$, then we have by Proposition 12.4 (ii),

$$u = \sum_{i=1}^m (u \bullet f_i) f_i = P_U(x),$$

which implies that $u \in \operatorname{im}(P_U)$. Therefore, $u \in \operatorname{im}(P_U)$ if and only if $u \in U$, i.e. $\operatorname{im}(P_U) = U$. ☐

Proposition 13.3 *Let $U \subset \mathbb{R}^n$ be a subspace and $x \in \mathbb{R}^n$. Then for all $u \in U$ we have*

$$\|x - P_U(x)\| \leq \|x - u\| .$$

We just have equality in the case when $u = P_U(x)$. In other words, if x is outside of U, then $P_U(x)$ is the closest point to x which is in U.

Proof. For any $u \in U$, doing the same calculation as the one in the proof of Proposition 12.1 (ii), we get

$$\begin{aligned}\|x-u\|^2 &= \|(x - P_U(x)) + (P_U(x) - u)\|^2 \\ &= \|x - P_U(x)\|^2 + 2\big((x - P_U(x)) \bullet (P_U(x) - u)\big) + \|P_U(x) - u\|^2\end{aligned}$$

From Lemma 12.6 (iv), we have $(x - P_U(x)) = x_\perp \in U^\perp$. Because $(P_U(x) - u) \in U$, we have

$$(x - P_U(x)) \bullet (P_U(x) - u) = 0.$$

Therefore,

$$\|x-u\|^2 = \|x - P_U(x)\|^2 + \|P_U(x) - u\|^2 \geq \|x - P_U(x)\|^2 \quad \Leftrightarrow \quad \|x - u\| \geq \|x - P_U(x)\|.$$

The equality occurs when $\|P_U(x) - u\| = 0 \quad \Leftrightarrow \quad u = P_U(x)$. □

Definition 13.4 The **transpose** of a matrix $A = (a_{ij}) \in \mathbb{R}^{m \times n}$ is the matrix $A^T = (a_{ji}) \in \mathbb{R}^{n \times m}$.

Example 42 Given the matrix

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \in \mathbb{R}^{2\times 3},$$

its transpose is

$$A^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix} \in \mathbb{R}^{3\times 2}.$$

Proposition 13.5 *(i) For $A, B \in \mathbb{R}^{m\times n}$ and $\lambda \in \mathbb{R}$ we have*

$$(A+B)^T = A^T + B^T, \qquad (\lambda A)^T = \lambda A^T .$$

(ii) For $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n \times l}$ we have

$$(AB)^T = B^T A^T \in \mathbb{R}^{l \times m} .$$

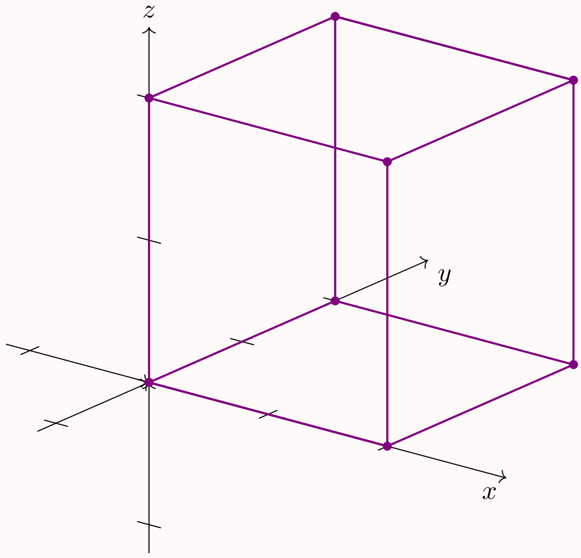
(iii) For $x, y \in \mathbb{R}^n$ we have $x \bullet y = x^T y$.

Proof. This can be checked by direct calculations. □

Example 43 (Basics behind 3D-Graphics) In this example, we want to project a cube in $\mathbb{R}^3$ to $\mathbb{R}^2$. This has the natural application of visualizing 3D-Graphics (in our case a cube) on a monitor. Our cube will have side-length 2 and its vertices are given by the following 8 points:

$$C = \left\{ \begin{pmatrix}0\\0\\0\end{pmatrix}, \begin{pmatrix}2\\0\\0\end{pmatrix}, \begin{pmatrix}0\\2\\0\end{pmatrix}, \begin{pmatrix}0\\0\\2\end{pmatrix}, \begin{pmatrix}2\\2\\0\end{pmatrix}, \begin{pmatrix}2\\0\\2\end{pmatrix}, \begin{pmatrix}0\\2\\2\end{pmatrix}, \begin{pmatrix}2\\2\\2\end{pmatrix} \right\}.$$

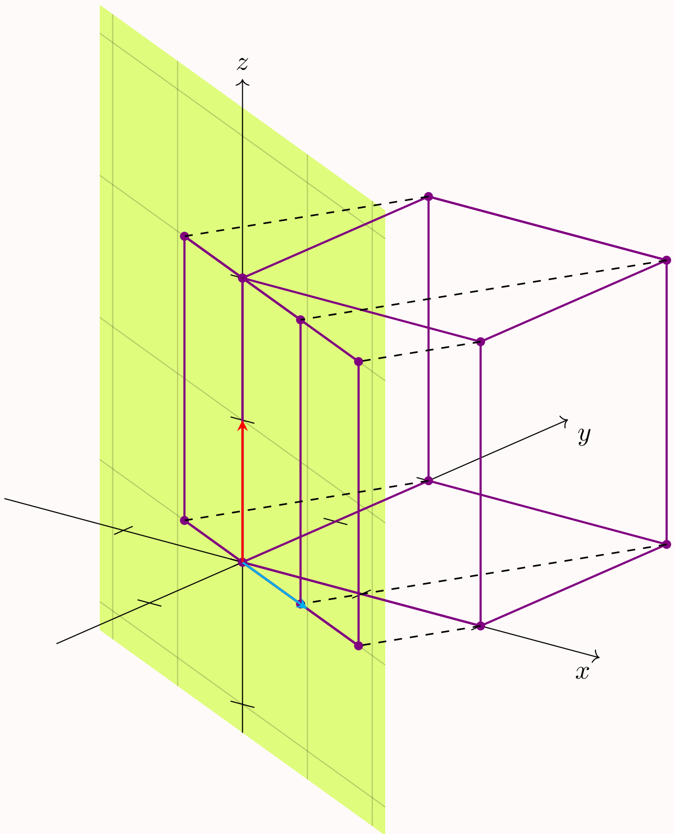
To make it look like a cube we connect two points if they share a facet, i.e. if their coordinates just differ by one entry. We obtain the following picture:

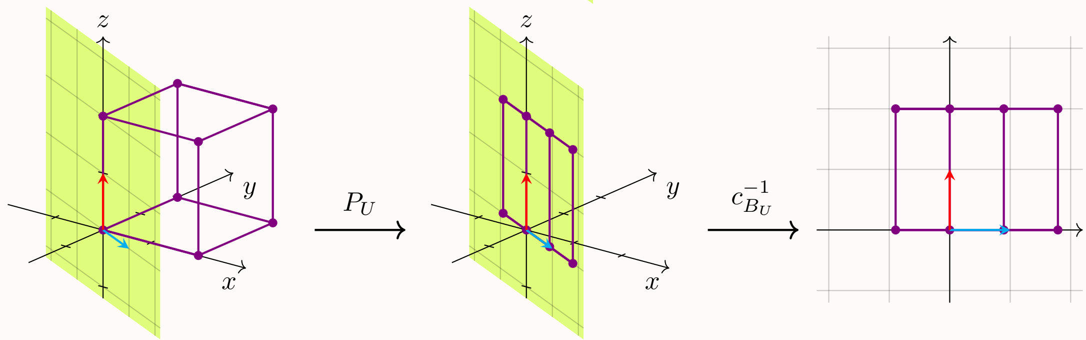


Now we want to project this cube onto a plane, which determines the viewing angle onto the scene. Let $U = \operatorname{span}\{f_1, f_2\} \subset \mathbb{R}^3$ with

$$f_1 = \frac{1}{\sqrt{5}} \begin{pmatrix}2\\-1\\0\end{pmatrix} \approx \begin{pmatrix}0.89\\-0.45\\0\end{pmatrix}, \qquad f_2 = \begin{pmatrix}0\\0\\1\end{pmatrix}.$$

We want to project the points in C onto the plane U and then calculate the coordinates with respect to the orthonormal basis $B_U = (f_1, f_2)$. For each $x \in C$ we now want to calculate $[P_U(x)]_{B_U}$. By Proposition 13.2 we have $P_U(x) = (x \bullet f_1)f_1 + (x \bullet f_2)f_2$.





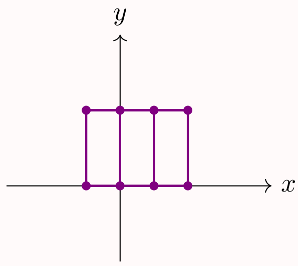
The coordinates are $[P_U(x)]_{B_U} = \begin{pmatrix} x \bullet f_1 \\ x \bullet f_2 \end{pmatrix} = \begin{pmatrix} f_1^T x \\ f_2^T x \end{pmatrix}$. Or in other words, we have

$$[P_U(x)]_{B_U} = (c_{B_U}^{-1} \circ P_U)(x) = \begin{pmatrix} —f_1^T— \\ —f_2^T— \end{pmatrix} x = \begin{pmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} & 0 \\ 0 & 0 & 1 \end{pmatrix} x.$$

Multiplying each element in C with this matrix gives the following set

$$P = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{4}{\sqrt{5}} \\ 0 \end{pmatrix}, \begin{pmatrix} -\frac{2}{\sqrt{5}} \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \begin{pmatrix} \frac{2}{\sqrt{5}} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{4}{\sqrt{5}} \\ 2 \end{pmatrix}, \begin{pmatrix} -\frac{2}{\sqrt{5}} \\ 2 \end{pmatrix}, \begin{pmatrix} \frac{2}{\sqrt{5}} \\ 2 \end{pmatrix} \right\}.$$

Drawing these points in $\mathbb{R}^2$ gives a picture of the cube viewed from the side:



Now consider another plane $V = \operatorname{span}\{v_1, v_2\}$ spanned by

$$v_1 = \frac{1}{\sqrt{5}}\begin{pmatrix}2\\-1\\0\end{pmatrix}, \qquad v_2 = \frac{1}{\sqrt{21}}\begin{pmatrix}1\\2\\4\end{pmatrix}.$$

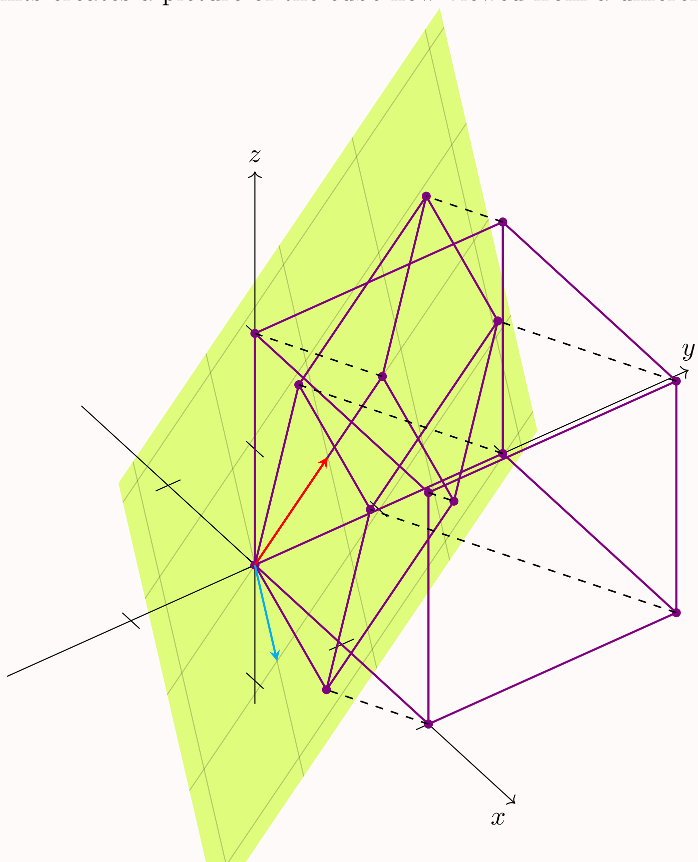
Notice that this again gives an orthonormal basis $B_V = (v_1, v_2)$ of V. As before we get

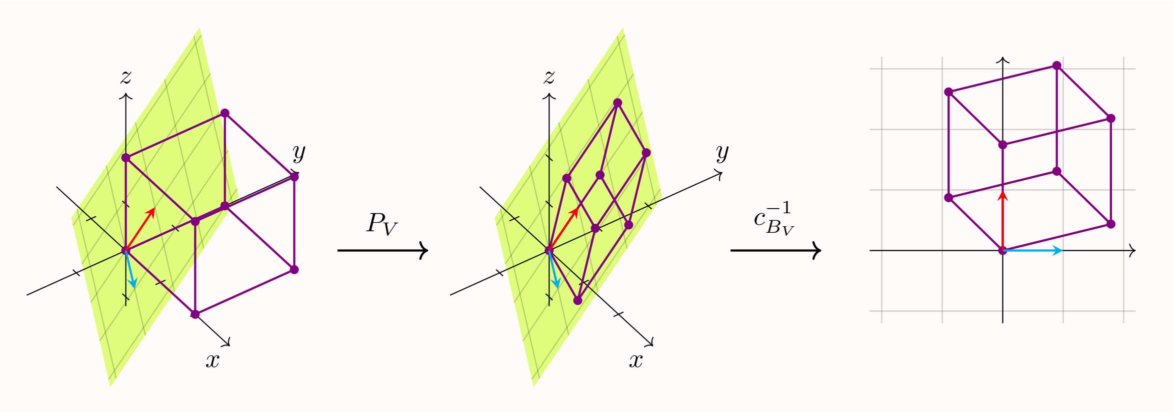
$$[P_V(x)]_{B_V} = (c_{B_V}^{-1} \circ P_V)(x) = \begin{pmatrix}—v_1^T—\\—v_2^T—\end{pmatrix} x = \begin{pmatrix}\frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} & 0\\ \frac{1}{\sqrt{21}} & \frac{2}{\sqrt{21}} & \frac{4}{\sqrt{21}}\end{pmatrix} x.$$

and applying this to each element in C gives

$$Q = \left\{ \begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\frac{4}{\sqrt{5}}\\ \frac{2}{\sqrt{21}}\end{pmatrix}, \begin{pmatrix}-\frac{2}{\sqrt{5}}\\ \frac{4}{\sqrt{21}}\end{pmatrix}, \begin{pmatrix}0\\ \frac{8}{\sqrt{21}}\end{pmatrix}, \begin{pmatrix}\frac{2}{\sqrt{5}}\\ \frac{6}{\sqrt{21}}\end{pmatrix}, \begin{pmatrix}\frac{4}{\sqrt{5}}\\ \frac{10}{\sqrt{21}}\end{pmatrix}, \begin{pmatrix}-\frac{2}{\sqrt{5}}\\ \frac{12}{\sqrt{21}}\end{pmatrix}, \begin{pmatrix}\frac{2}{\sqrt{5}}\\ \frac{14}{\sqrt{21}}\end{pmatrix} \right\}.$$

Plotting these points creates a picture of the cube now viewed from a different angle:





For any $A \in \mathbb{R}^{m\times n}$, we can define a linear map $F : x \mapsto Ax$. Using that, we define the image and kernel of the matrix A by $\operatorname{im}(A) = \operatorname{im}(F)$ and $\ker(A) = \ker(F)$.

Proposition 13.6 *For all $A \in \mathbb{R}^{m\times n}$ we have $\operatorname{im}(A)^\perp = \ker(A^T)$.*

Proof. Let $x \in \mathbb{R}^n$. Then we have

$$\begin{aligned}
x \in \operatorname{im}(A)^\perp \quad &\Leftrightarrow \quad y \bullet x = 0, \quad \forall y \in \operatorname{im}(A) \\
&\Leftrightarrow \quad (Av) \bullet x = 0, \quad \forall v \in \mathbb{R}^n \\
&\Leftrightarrow \quad (Av)^T x = 0, \quad \forall v \in \mathbb{R}^n && \text{(by Proposition 13.5 (iii))} \\
&\Leftrightarrow \quad v^T A^T x = 0, \quad \forall v \in \mathbb{R}^n && \text{(by Proposition 13.5 (ii))} \\
&\Leftrightarrow \quad v \bullet \left(A^T x\right) = 0, \quad \forall v \in \mathbb{R}^n && \text{(by Proposition 13.5 (iii))} \\
&\Leftrightarrow \quad A^T x = \mathbf{0} \\
&\Leftrightarrow \quad x \in \ker\left(A^T\right).
\end{aligned}$$

□

Corollary 13.7 *Let $A \in \mathbb{R}^{m\times n}$.*

(i) We have $\ker(A^T A) = \ker(A)$.

(ii) We have the following equivalence

$$\ker(A) = \{\mathbf{0}\} \quad \Longleftrightarrow \quad A^T A \in \mathbb{R}^{n\times n} \text{ is invertible}.$$

Proof. (i) We have

$$\begin{aligned}
x \in \ker\left(A^T A\right) \quad \Leftrightarrow \quad A^T A x = \mathbf{0} \quad &\Leftrightarrow \quad Ax \in \ker\left(A^T\right) = \operatorname{im}(A)^\perp && \text{(Proposition 13.6)} \\
&\Leftrightarrow \quad Ax \in \operatorname{im}(A) \cap \operatorname{im}(A)^\perp = \{\mathbf{0}\} && \text{(Lemma 12.6 (ii))} \\
&\Leftrightarrow \quad Ax = \mathbf{0} \\
&\Leftrightarrow \quad x \in \ker(A).
\end{aligned}$$

(ii) Since $A^T A$ is a $n \times n$ matrix, we have

$$\begin{aligned}
\ker(A) = \{\mathbf{0}\} \quad &\Longleftrightarrow \quad \ker(A^T A) = \{\mathbf{0}\} \\
&\Longleftrightarrow \quad A^T A \text{ is invertible} && \text{(Theorem 8.7)}
\end{aligned}$$

□

We can now use our results to answer the question at the beginning of this chapter. Here, we will present the **least squares method**.

Problem: Given a linear map $F : \mathbb{R}^n \to \mathbb{R}^m$ and $b \in \mathbb{R}^m$, we want to find $x \in \mathbb{R}^n$ that minimizes the quantity

$$\delta = \|F(x) - b\|.$$

The "least squares" stems from the fact that if $b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$ and $F(x) = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$, then

$$\delta = \|F(x) - b\| = \sqrt{(y_1 - b_1)^2 + \ldots + (y_m - b_m)^2}.$$

In other words, the problem is finding $x \in \mathbb{R}^n$ that minimizes the sum of squares of the difference $y_i - b_i$.

Remark. Minimal δ is 0 if and only if $b \in \operatorname{im}(F)$, i.e. $F(x) = b$ has a solution.

By Proposition 13.3, the minimal δ is given in the case $F(x) = P_{\operatorname{im}(F)}(b)$. Writing $[F] = A$, we want to find $x \in \mathbb{R}^n$ such that $Ax = P_{\operatorname{im}(F)}(b)$. We have

$$\begin{aligned} Ax = P_{\operatorname{im}(F)} \quad &\Leftrightarrow \quad (Ax - b) \in (\operatorname{im}(A))^\perp = \ker(A^T) && \text{(by Proposition 13.6)} \\ &\Leftrightarrow \quad A^T(Ax - b) = \mathbf{0} \\ &\Leftrightarrow \quad \underbrace{\boxed{A^T Ax = A^T b}}_{\textbf{normal equation}} \end{aligned}$$

Therefore, if $\ker(A) = \{\mathbf{0}\}$ (i.e., the columns of A are linearly independent), then by Corollary 13.7, $A^T A$ is invertible and we get the unique solution to our problem by

$$x = \left(A^T A\right)^{-1} A^T b.$$

Example 44 Find the best possible quadratic polynomial $f(t) = a_0 + a_1 t + a_2 t^2$ to fit the data points $(0, 2),\ (1, 1),\ (2, 2),\ (3, 3)$.
We first translate this problem into linear algebra: We want to minimize

$$(f(0) - 2)^2 + (f(1) - 1)^2 + (f(2) - 2)^2 + (f(3) - 3)^2,$$

so we define the linear map $F : \mathbb{R}^3 \to \mathbb{R}^4$ for $x = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} \in \mathbb{R}^3$ by

$$F(x) = \begin{pmatrix} f(0) \\ f(1) \\ f(2) \\ f(3) \end{pmatrix} = \begin{pmatrix} a_0 + a_1 \cdot 0 + a_2 \cdot 0^2 \\ a_0 + a_1 \cdot 1 + a_2 \cdot 1^2 \\ a_0 + a_1 \cdot 2 + a_2 \cdot 2^2 \\ a_0 + a_1 \cdot 3 + a_2 \cdot 3^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = Ax,$$

where $A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix}$. Then, we want to find $x \in \mathbb{R}^3$ such that $\|Ax - b\|$ with $b = \begin{pmatrix} 2 \\ 1 \\ 2 \\ 3 \end{pmatrix}$ is minimal. Therefore, we need to solve the normal equation $A^T Ax = A^T b$.

We have

$$A^T = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 4 & 9 \end{pmatrix}, \qquad A^T A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 4 & 9 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix} = \begin{pmatrix} 4 & 6 & 14 \\ 6 & 14 & 36 \\ 14 & 36 & 98 \end{pmatrix},$$

$$A^T b = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 4 & 9 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 8 \\ 14 \\ 36 \end{pmatrix}.$$

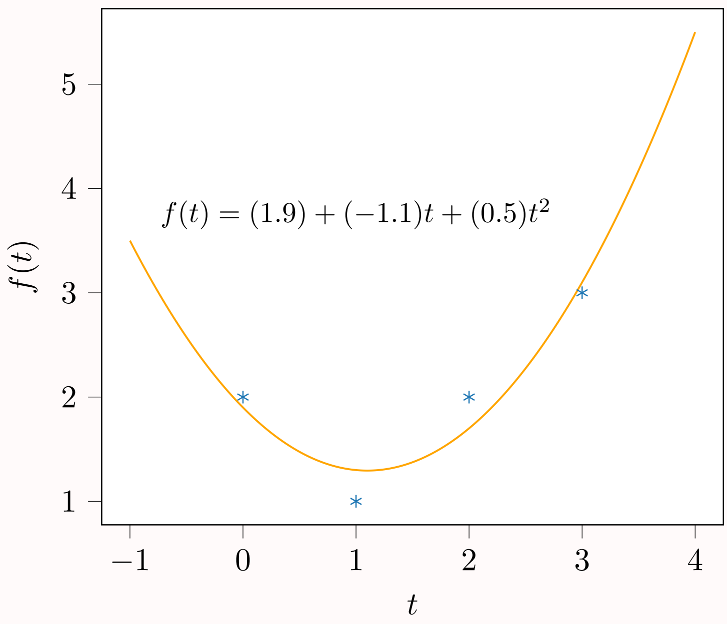
Then, we want solve the linear system

$$\begin{pmatrix} 4 & 6 & 14 \\ 6 & 14 & 36 \\ 14 & 36 & 98 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 8 \\ 14 \\ 36 \end{pmatrix}.$$

This linear system has a unique solution $\begin{cases} a_0 = \frac{19}{10} \\ a_1 = -\frac{11}{10} \\ a_2 = \frac{1}{2} \end{cases}$.

Hence, the best fit polynomial is

$$f(t) = \frac{19}{10} - \frac{11}{10}t + \frac{1}{2}t^2.$$



Remark. This method works for arbitrary polynomials (i.e. in particular for lines). In addition, the normal equation always have a unique solution if the columns of A are linearly indepdendent. In applications, this is usually the case since the number of data points (the number m of rows of A) is usually greater than the degree of the polynomial $(n-1)$, where n is the number of columns of A.

Exercises

Exercise 42. Three points (x_1, y_1), (x_2, y_2) and (x_3, y_3) lie on one (non-vertical) line, if there exist $a, b \in \mathbb{R}$ with $ax_j + b = y_j$ for $j = 1, 2, 3$. In other words the linear system

$$\underbrace{\begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \end{pmatrix}}_{=A} \begin{pmatrix} a \\ b \end{pmatrix} = \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}}_{=y}$$

has a solution, i.e. $y \in \mathrm{im}(A)$.

(i) Show that the points $(0, 1)$, $(1, 3)$ and $(2, 2)$ do <u>not</u> lie on one line.

For (ii) - (iv) we assume that $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix}$ and $y = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}$.

(ii) Calculate an orthonormal basis $F = (f_1, f_2)$ of $\mathrm{im}(A)$ by using the GSA for the columns of A.
(Hint: The result becomes nicer if you set b_1 = second column of A and b_2 = first column of A.)

(iii) Calculate $z = (y \bullet f_1)f_1 + (y \bullet f_2)f_2$ and show that $z \in \mathrm{im}(A)$.

(iv) Solve the linear system $A\begin{pmatrix} a \\ b \end{pmatrix} = z$ and draw the graph of $f(x) = ax + b$ together with the three points in i). Can you interpret the connection between the graph and the points?

Exercise 43. Assume we have the following data points

i	1	2	3	4
x_i	0	1	2	3
y_i	2	1	3	4

(i) Find the line of best fit for the above data, i.e. find $a, b \in \mathbb{R}$ such that the function $l(x) = ax + b$ minimizes the sum of squares $\sum_{i=1}^{4}(l(x_i) - y_i)^2$.

(ii) Interpolate the data by a quadratic polynomial. For this find $c, d, e \in \mathbb{R}$ such that the function $p(x) = cx^2 + dx + e$ minimizes $\sum_{i=1}^{4}(p(x_i) - y_i)^2$.

(iii) Draw the data points and the graphs of l and p into one diagram.

For both (i) and (ii) solve the exercise by finding the solutions to the normal equation.

Midterm & Final exams

In the following you can find the midterms and final exams given in the fall semesters 2019 - 2022 at Nagoya University. Solutions can be found on the corresponding pages of the lectures at `https://www.henrikbachmann.com/teaching.html`.

13.1 Linear Algebra I - Midterm 2019

Exercise 1. (10 Points) Consider the following linear system

$$\left\{\begin{array}{ccccccccc} 2x_1 & + & 3x_2 & + & 4x_3 & + & 5x_4 & = & 6 \\ x_1 & + & 2x_2 & + & 3x_3 & + & 4x_4 & = & 5 \\ 3x_1 & + & 4x_2 & + & 5x_3 & + & 6x_4 & = & 7 \end{array}\right. .$$

(i) Find a matrix $A \in \mathbb{R}^{3\times 4}$ and and a vector $b \in \mathbb{R}^3$, such that the solutions of the above linear system are given by the vectors $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \in \mathbb{R}^4$ satisfying $Ax = b$.

(ii) Determine the row-reduced echelon forms of the matrices $(A \mid b)$ and A.

(iii) Find all the solutions to the linear system.

(iv) Calculate the rank of $(A \mid b)$ and A.

(v) Find a vector $c \in \mathbb{R}^3$, such that $Ax = c$ has no solutions. Calculate the rank of $(A \mid c)$.

Exercise 2. (10 Points) Let $u = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \in \mathbb{R}^2$ and define the following four functions:

$$\begin{aligned} f_1 : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ x &\longmapsto (u \bullet x)u + x \,, \end{aligned} \qquad \begin{aligned} f_2 : \mathbb{R} &\longrightarrow \mathbb{R}^2 \\ x &\longmapsto \begin{pmatrix} 2\cos(x) \\ \sin(x) \end{pmatrix} , \end{aligned}$$

$$\begin{aligned} f_3 : \mathbb{R}^3 &\longrightarrow \mathbb{R}^2 \\ x &\longmapsto \frac{x \bullet x}{u \bullet u} u \,, \end{aligned} \qquad \begin{aligned} f_4 : \mathbb{R}^3 &\longrightarrow \mathbb{R}^4 \\ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} &\longmapsto \begin{pmatrix} 0 \\ x_1 + 2x_2 + 3x_3 \\ x_1 - x_3 \\ x_2 \end{pmatrix} . \end{aligned}$$

(i) Which of the above functions f_1, f_2, f_3, f_4 are linear maps? For each one that is linear, determine its matrix.

(ii) Draw a picture of the image of f_2. Is f_2 injective and/or surjective?

Exercise 3. (6 Points) Let $G : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear map with

$$G\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix} , \quad G\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \end{pmatrix} .$$

(i) Determine the matrix of G.

(ii) Determine the matrix of $G \circ G$.

Exercise 4. (6 Points) We define the following linear map

$$\begin{aligned} H : \mathbb{R}^3 &\longrightarrow \mathbb{R}^3 \\ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} &\longmapsto \begin{pmatrix} x_1 \\ x_2 + x_3 \\ x_1 + x_2 + x_3 \end{pmatrix} . \end{aligned}$$

(i) Calculate the image of H.

(ii) Decide if H is injective and/or surjective.

(iii) Find all vectors $v \in \mathbb{R}^3$, which are orthogonal to all vectors in the image of H.

13.2 Linear Algebra I - Midterm 2020

Exercise 1. (10 Points) Consider the following linear system

$$\left\{\begin{array}{rcrcrcrcr} -2x_1 & + & 4x_2 & + & x_3 & + & x_4 & = & 6 \\ -3x_1 & + & 6x_2 & + & x_3 & & & = & 7 \\ x_1 & - & 2x_2 & & & + & x_4 & = & -1 \end{array}\right. .$$

(i) Find a matrix $A \in \mathbb{R}^{3\times 4}$ and and a vector $b \in \mathbb{R}^3$, such that the solutions of the above linear system are given by the vectors $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \in \mathbb{R}^4$ satisfying $Ax = b$.
(ii) Determine the row-reduced echelon forms of the matrices $(A \mid b)$ and A.
(iii) Find all the solutions to the linear system.
(iv) Calculate the rank of $(A \mid b)$ and A.
(v) Find all $y \in \mathbb{R}^4$ with $Ay = 2b$ by using your result for iii).

Exercise 2. (8 Points) Let $u = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \in \mathbb{R}^2$ and define the following four functions:

$$f_1 : \mathbb{R}^2 \longrightarrow \mathbb{R}^3 \qquad x \longmapsto \begin{pmatrix} u \bullet x \\ 0 \\ x \bullet u \end{pmatrix}, \qquad f_2 : \mathbb{R}^2 \longrightarrow \mathbb{R} \qquad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \longmapsto 2^{x_1+x_2} - 1, \qquad f_3 : \mathbb{R}^3 \longrightarrow \mathbb{R}^2 \qquad \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \longmapsto \begin{pmatrix} x_1 - 3x_2 \\ 2x_1 + x_2x_3 \end{pmatrix}.$$

(i) Which of the above functions f_1, f_2, f_3 are linear maps? For each one that is linear, determine its matrix.
(ii) Is f_2 injective and/or surjective?

Exercise 3. (8 Points)
(i) Let $G : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear map with

$$G\begin{pmatrix} -1 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad G\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}.$$

Determine the matrix of G.
(ii) Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be a function with

$$F\begin{pmatrix} -1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad F\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad F\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix}.$$

Show that F is <u>not</u> a linear map.

Exercise 4. (8 Points) We define the following linear map

$$H : \mathbb{R}^3 \longrightarrow \mathbb{R}^3 \qquad \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \longmapsto \begin{pmatrix} x_1 + x_2 \\ x_1 - x_3 \\ x_2 + x_3 \end{pmatrix}.$$

(i) Calculate the image of H.
(ii) Decide if H is injective and/or surjective.
(iii) Find a non-zero vector $v \in \mathbb{R}^3$, such that v is orthogonal to $H(v)$. (Just <u>one explicit</u> vector is enough)

13.3 Linear Algebra I - Midterm 2021

Exercise 1. (10 Points) Consider the following linear system

$$\left\{\begin{array}{rcrcrcrcr} 3x_1 & - & 6x_2 & + & x_3 & + & 5x_4 & = & 5 \\ 2x_1 & - & 4x_2 & + & x_3 & + & 3x_4 & = & 4 \\ -x_1 & + & 2x_2 & - & 2x_3 & & & = & -5 \end{array}\right. .$$

(i) Find a matrix $A \in \mathbb{R}^{3\times 4}$ and and a vector $b \in \mathbb{R}^3$, such that the solutions of the above linear system are given by the vectors $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \in \mathbb{R}^4$ satisfying $Ax = b$.

(ii) Determine the row-reduced echelon forms of the matrices $(A \mid b)$ and A and calculate their ranks.

(iii) Find all the solutions to the linear system.

(iv) Determine all $x \in \mathbb{R}^4$ which satisfy $Ax = b$ and which are orthogonal to the vector $u = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 1 \end{pmatrix}$.

Exercise 2. (8 Points) Let $u = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \in \mathbb{R}^2$ and define the following three functions:

$$f_1 : \mathbb{R}^3 \longrightarrow \mathbb{R}^2, \quad \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \longmapsto \begin{pmatrix} 2x_1 + 3x_2 \\ x_1 + (u \bullet u)x_3 \end{pmatrix}, \qquad f_2 : \mathbb{R}^2 \longrightarrow \mathbb{R}, \quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \longmapsto \sin(x_1) + \cos(x_2), \qquad f_3 : \mathbb{R}^2 \longrightarrow \mathbb{R}^3, \quad x \longmapsto \begin{pmatrix} x \bullet x \\ 0 \\ u \bullet u \end{pmatrix}.$$

(i) Which of the above functions f_1, f_2, f_3 are linear maps? For each one that is linear, determine its matrix.

(ii) Is f_2 injective and/or surjective?

Exercise 3. (8 Points)

(i) Let $G : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear map with

$$G\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad G\begin{pmatrix} -2 \\ -1 \end{pmatrix} = \begin{pmatrix} -2 \\ 2 \end{pmatrix}.$$

Determine the matrix of G.

(ii) Let $F : \mathbb{R}^3 \to \mathbb{R}^3$ be a linear map with

$$F\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ 3 \end{pmatrix}, \quad F\begin{pmatrix} 1 \\ -1 \\ 5 \end{pmatrix} = \begin{pmatrix} 6 \\ 4 \\ 6 \end{pmatrix}.$$

Show that F is not injective.

Exercise 4. (8 Points) We define the following linear map

$$H : \mathbb{R}^3 \longrightarrow \mathbb{R}^3, \quad \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \longmapsto \begin{pmatrix} x_1 + x_2 - x_3 \\ x_1 + 2x_2 \\ x_2 + x_3 \end{pmatrix}.$$

(i) Calculate the image of H.

(ii) Decide if H is injective and/or surjective.

(iii) Find all vectors $x \in \mathbb{R}^3$ with $H(x) = 2x$.

13.4 Linear Algebra I - Midterm 2022

Exercise 1. (10 Points) Consider the following linear system

$$\left\{\begin{array}{rcrcrcrcr} x_1 & + & 3x_2 & & & + & x_4 & = & 1 \\ & & x_2 & + & 2x_3 & - & 2x_4 & = & 2 \\ 2x_1 & - & 2x_2 & + & x_3 & + & x_4 & = & 3 \end{array}\right. .$$

(i) Find a matrix $A \in \mathbb{R}^{3\times 4}$ and and a vector $b \in \mathbb{R}^3$, such that the solutions of the above linear system are given by the vectors $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \in \mathbb{R}^4$ satisfying $Ax = b$.
(ii) Determine the row-reduced echelon forms of the matrices $(A \mid b)$ and A and calculate their ranks.
(iii) Find all the solutions to the linear system.
(iv) Determine all $x \in \mathbb{R}^4$ which satisfy $Ax = b$ and which have norm $\|x\| = \sqrt{14}$.

Exercise 2. (8 Points) Let $u = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \in \mathbb{R}^2$ and define the following three functions:

$$f_1 : \mathbb{R}^3 \longrightarrow \mathbb{R}^2, \quad \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \longmapsto \begin{pmatrix} (u \bullet u) - 2 \\ x_1 + (u \bullet u)x_3 \end{pmatrix},$$

$$f_2 : \mathbb{R}^2 \longrightarrow \mathbb{R}, \quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \longmapsto e^{x_1} - e^{x_2},$$

$$f_3 : \mathbb{R}^2 \longrightarrow \mathbb{R}^2, \quad x \longmapsto (x \bullet u)u\,.$$

(i) Which of the above functions f_1, f_2, f_3 are linear maps? For each one that is linear, determine its matrix.
(ii) Is f_2 injective and/or surjective?

Exercise 3. (8 Points) Let $G : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear map with

$$G\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ -2 \end{pmatrix}, \quad G\begin{pmatrix} 2 \\ 2 \end{pmatrix} = \begin{pmatrix} -4 \\ 4 \end{pmatrix}.$$

(i) Determine the matrix of G.
(ii) Find all vectors $x \in \mathbb{R}^2$ such that x is orthogonal to $G(x)$.

Exercise 4. (8 Points) We define the following linear map

$$H : \mathbb{R}^3 \longrightarrow \mathbb{R}^4, \quad \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \longmapsto \begin{pmatrix} x_1 + x_2 \\ x_2 + x_3 \\ x_1 + 2x_2 + x_3 \\ 2x_1 - 2x_3 \end{pmatrix}.$$

(i) Calculate the image of H.
(ii) Decide if H is injective and/or surjective.
(iii) Find a linear map $F : \mathbb{R}^2 \to \mathbb{R}^4$ with $\mathrm{im}(F) = \mathrm{im}(H)$.

13.5 Linear Algebra I - Finals 2019

Exercise 1. (12 Points) Let $A = \begin{pmatrix} 0 & 1 & -2 & 3 \\ 1 & -2 & 3 & -4 \\ -2 & 3 & -4 & 5 \end{pmatrix}$ and $B = \begin{pmatrix} 2 & 1 & 1 \\ 3 & 2 & 1 \\ 2 & 1 & 2 \end{pmatrix}$.

(i) Compute the products AB and BA, or explain why they are not defined.
(ii) Determine whether or not the matrices A and B are invertible and, if they are, compute their inverses.
(iii) Calculate $\operatorname{Im}(B)$ and $\ker(B)$.

Exercise 2. (14 Points) We define the subspace $U = \operatorname{span}\{u_1, u_2, u_3\} \subset \mathbb{R}^3$, where

$$u_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad u_2 = \begin{pmatrix} 1 \\ 0 \\ -3 \end{pmatrix}, \quad u_3 = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}.$$

(i) Determine a basis $B = (b_1, \dots, b_m)$ of U and calculate its dimension.
(ii) Calculate the coordinate vectors $[u_1]_B$, $[u_2]_B$ and $[u_3]_B$, where B is the basis you determined in i).
(iii) Determine a basis for $U^\perp$.
(iv) Find a linear map $G : \mathbb{R}^2 \to \mathbb{R}^3$ with $\ker(G) = \{0\}$ and $\operatorname{Im}(G) = U$.

Exercise 3. (10 Points) Which of the following subsets of $\mathbb{R}^2$ are subspaces? Justify your answers.

(i) $U_1 = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 \mid x_1 + x_2 = x_1 x_2 \right\}$.
(ii) $U_2 = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 \mid 2x_1 = x_1 + x_2 \right\}$.
(iii) $U_3 = \operatorname{span}\left\{ \begin{pmatrix} 2 \\ 2 \end{pmatrix} \right\} \bigcup \operatorname{span}\left\{ \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\}$.

(Friendly reminder: $\cup$ is the union of two sets)

Exercise 4. (14 Points) We define the following linear map

$$T : \mathbb{R}^2 \longrightarrow \mathbb{R}^3$$
$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \longmapsto \begin{pmatrix} 1 & 2 \\ 0 & 3 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

(i) Calculate an orthonormal basis $F = (f_1, \dots, f_r)$ for $\operatorname{Im}(T)$.
(ii) Check for which $t \in \mathbb{R}$ the vector $v = \begin{pmatrix} 1 \\ t \\ 1 \end{pmatrix}$ is an element in $\operatorname{Im}(T)$. Determine the coordinate vector $[v]_F$ in this case.
(iii) Find a $w \in \mathbb{R}^3$ with $[w]_F = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$.
(iv) Find a $x \in \mathbb{R}^2$ such that $\|T(x) - b\|$ is minimal, where $b = \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}$.

(In (ii) and (iii) the F is the basis of $\operatorname{Im}(T)$ you calculated in (i)).

13.6 Linear Algebra I - Finals 2020

Exercise 1. (12 Points) Let $A = \begin{pmatrix} 1 & -1 & 0 \\ -2 & 3 & -4 \\ -2 & 3 & -3 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & -1 \\ -2 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$.

(i) Compute the products AB and BA, or explain why they are not defined.
(ii) Determine whether or not the matrices A and B are invertible and, if they are, compute their inverses.
(iii) Find all $x \in \mathbb{R}^3$ with $A^T AA^T AA^T Ax = 0$. Justify you answer.

Exercise 2. (12 Points) We define the subspace $U = \operatorname{span}\{u_1, u_2, u_3, u_4\} \subset \mathbb{R}^3$, where

$$u_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad u_2 = \begin{pmatrix} -1 \\ 0 \\ 3 \end{pmatrix}, \quad u_3 = \begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix}, \quad u_4 = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}.$$

(i) Determine a basis $B = (b_1, \dots, b_m)$ of U and calculate its dimension.
(ii) Calculate the coordinate vectors $[u_1]_B$, $[u_2]_B$,$[u_3]_B$ and $[u_4]_B$, where B is the basis from i).
(iii) Find a linear map $G : \mathbb{R}^3 \to \mathbb{R}^3$ with $\operatorname{Im}(G) = U$. What is the dimension of $\ker(G)$?

Exercise 3. (12 Points) Set $u = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$. Which of the following subsets of $\mathbb{R}^2$ are subspaces? Justify your answers.

(i) $U_1 = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 \mid 2x_1 - 3x_2 = x_1 \right\}$.
(ii) $U_2 = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 \mid x_1 \text{ is an integer, i.e. } x_1 \in \{\dots, -2, -1, 0, 1, 2, \dots\} \right\}$.
(iii) $U_3 = \left\{ x \in \mathbb{R}^2 \mid x \notin \operatorname{span}\{u\} \right\}$.
(iv) $U_4 = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 \mid \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \bullet u = x_1 \right\}$.

Exercise 4. (14 Points) We define the following linear map

$$\begin{aligned} H : \mathbb{R}^2 &\longrightarrow \mathbb{R}^3 \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &\longmapsto \begin{pmatrix} 1 & 1 \\ -2 & 4 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \end{aligned}$$

(i) Show that $\dim(\operatorname{Im}(H)) = 2$.
(ii) Calculate an orthonormal basis (f_1, f_2) for $\operatorname{Im}(H)$.
(iii) Find a vector $v \in \mathbb{R}^3$, such that $B = (f_1, f_2, v)$ is an orthonormal basis for $\mathbb{R}^3$.
(iv) Calculate $[H(x)]_B$ for any $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$.
(v) Find a $x \in \mathbb{R}^2$ such that $\|H(x) - b\|$ is minimal, where $b = \begin{pmatrix} -5 \\ 1 \\ -1 \end{pmatrix}$.

13.7 Linear Algebra I - Finals 2021

Exercise 1. (14 Points) Let $A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & -2 \\ 1 & 2 & 5 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 1 & -2 \\ -1 & 2 & 5 \end{pmatrix}$.

(i) Determine whether or not the matrices A and B are invertible and, if they are, compute their inverses.
(ii) Calculate the matrix BA and decide if BA^n is invertible for any integer $n \geq 1$.
(iii) Determine if $\mathrm{Im}(A) \cup \mathrm{Im}(B)$ and $\mathrm{Im}(A) \cap \mathrm{Im}(B)$ are subspaces and, if they are, determine a basis.

Exercise 2. (12 Points) We define the subspace $U = \mathrm{span}\{u_1, u_2, u_3, u_4\} \subset \mathbb{R}^3$, where

$$u_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad u_2 = \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}, \quad u_3 = \begin{pmatrix} 3 \\ -4 \\ -1 \end{pmatrix}, \quad u_4 = \begin{pmatrix} 4 \\ -2 \\ 2 \end{pmatrix}.$$

(i) Determine a basis $B = (b_1, \ldots, b_m)$ of U and calculate its dimension.
(ii) Calculate the coordinate vectors $[u_1]_B$, $[u_2]_B$,$[u_3]_B$ and $[u_4]_B$, where B is the basis from i).
(iii) Calculate an orthonormal basis $F = (f_1, \ldots, f_m)$ for U and determine $[u_1]_F$, $[u_2]_F$,$[u_3]_F$ and $[u_4]_F$
(iv) Determine a basis for $U^\perp$.

Exercise 3. (12 Points) Set $u = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$ and let $C \in \mathbb{R}^{2\times 2}$ be an arbitrary matrix. Which of the following subsets of $\mathbb{R}^2$ are subspaces? Justify your answers.

(i) $U_1 = \{x \in \mathbb{R}^2 \mid x \bullet x = x \bullet u\}$.
(ii) $U_2 = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 \mid 2x_1 - x_2 = 3x_2 \right\}$.
(iii) $U_3 = \{x \in \mathbb{R}^2 \mid Cx = x\}$.
(iv) $U_4 = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 \mid x_1 \cdot x_2 \geq x_1 \right\}$.

Exercise 4. (12 Points) Assume we have the following data points

i	1	2	3
x_i	1	2	3
y_i	0	-1	-3

(i) Find the line of best fit for the above data, i.e. find $m, n \in \mathbb{R}$ such that the function $l(x) = mx + n$ minimizes the sum of squares $\sum_{i=1}^{3} (l(x_i) - y_i)^2$.
(ii) We define the following linear map

$$H : \mathbb{R}^2 \longrightarrow \mathbb{R}^3$$
$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \longmapsto \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

and set $V = \mathrm{Im}(H)$. Determine a basis of V and for $b = \begin{pmatrix} 0 \\ -1 \\ -3 \end{pmatrix}$ calculate the orthogonal projection $P_V(b)$. Use your result to show that b is not an element in V.

13.8 Linear Algebra I - Finals 2022

Exercise 1. (12 Points) Let $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$, $B = \begin{pmatrix} 2 & -4 & 1 \\ -1 & 2 & -1 \\ 1 & -1 & 2 \end{pmatrix}$ and $C = A^T BA$.

(i) Determine whether or not the matrices A, B, and C are invertible and, if they are, compute their inverses.
(ii) Determine $\operatorname{im}(C)$, $\ker(C)$ and $\operatorname{im}(C) \cap \ker(C)$.
(iii) Give a basis for $\ker(C^n)$ for all $n \geq 1$.

Exercise 2. (14 Points) We define the subspace $U = \operatorname{span}\{u_1, u_2, u_3, u_4\} \subset \mathbb{R}^3$, where

$$u_1 = \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}, \quad u_2 = \begin{pmatrix} -6 \\ -6 \\ -3 \end{pmatrix}, \quad u_3 = \begin{pmatrix} -1 \\ 3 \\ 1 \end{pmatrix}, \quad u_4 = \begin{pmatrix} 6 \\ -2 \\ 0 \end{pmatrix}.$$

(i) Determine a basis $B = (b_1, \ldots, b_m)$ of U and calculate its dimension.
(ii) Calculate the coordinate vectors $[u_1]_B$, $[u_2]_B$,$[u_3]_B$ and $[u_4]_B$, where B is the basis from i).
(iii) Find a linear map $F : \mathbb{R}^3 \to \mathbb{R}^3$ with $\ker(F) = U^\perp$ and determine a basis of $\operatorname{im}(F)$.

Exercise 3. (12 Points) Let $D \in \mathbb{R}^{2\times 2}$ be an arbitrary matrix. Which of the following sets are subspaces? Justify your answers.

(i) $U_1 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3 \mid 2x_1 - x_2 = -x_3 + x_2 \right\}$.
(ii) $U_2 = \left\{ x \in \mathbb{R}^{2023} \mid x \bullet x \geq -2023 \right\}$.
(iii) $U_3 = \left\{ x \in \mathbb{R}^2 \mid \text{ There exists a } y \in \mathbb{R}^2 \text{ with } Dy = 3x \right\}$.
(iv) $U_4 = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 \mid x_1 + x_2 = 0 \right\} \cup \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 \mid x_1 \cdot x_2 \leq 0 \right\}$.

Exercise 4. (12 Points) We consider the vector $b = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$, the linear map

$$G : \mathbb{R}^2 \longrightarrow \mathbb{R}^4$$
$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \longmapsto \begin{pmatrix} 1 & 1 \\ -1 & 1 \\ 1 & 3 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

and define the subspace $U = \operatorname{im}(G)$.

(i) Show that $\dim(U) = 2$ and find an orthonormal basis $F = (f_1, f_2)$ of U.
(ii) Determine the orthogonal projection $y = P_U(b)$ of b onto U and calculate $[y]_F$.
(iii) Find a $x \in \mathbb{R}^2$ such that $\|G(x) - b\|$ is minimal.

Linear Algebra II

Introduction

In Linear Algebra I, we focused on concepts primarily within the familiar framework of $\mathbb{R}^n$. This allowed us to visualize and understand these ideas concretely, facilitating intuition and practical computations. Often, the proofs presented relied only on the defining properties of addition and scalar multiplication, rather than on specific characteristics of $\mathbb{R}^n$. Consequently, nearly all results from Linear Algebra I remain valid in the more general context of arbitrary vector spaces. This realization provides a powerful perspective: the fundamental principles we learned are not limited to specific numerical vectors but extend broadly across mathematical structures known as **vector spaces**.
Specifically, we introduced foundational concepts such as:

- Linear systems and their solutions ($Ax = b$)
- Matrices ($\in \mathbb{R}^{m\times n}$), vectors ($\in \mathbb{R}^n$), and basic operations, specifically:
  (i) Addition: $x, y \mapsto (x + y) \in \mathbb{R}^n$
  (ii) Scalar multiplication: $x \mapsto \lambda x \in \mathbb{R}^n$
- Linear maps, functions $F : \mathbb{R}^m \to \mathbb{R}^n$ satisfying:
  (i) $F(x + y) = F(x) + F(y)$
  (ii) $F(\lambda x) = \lambda F(x)$
- Subspaces $U \subset \mathbb{R}^n$ defined by the properties:
  (i) $\mathbf{0} \in U$
  (ii) $x, y \in U \Rightarrow x + y \in U$ (closure under addition)
  (iii) $x \in U \Rightarrow \lambda x \in U$ (closure under scalar multiplication)
- Image and kernel of linear maps
- Linear independence and bases

However, these concepts encountered are not restricted to $\mathbb{R}^n$; they appear naturally in various mathematical contexts, forming vector spaces. Vector spaces generalize the essential properties of addition and scalar multiplication that we encountered. Examples of vector spaces beyond $\mathbb{R}^n$ include:

- The space of all real-valued continuous functions on $\mathbb{R}^n$ or an interval, for example $C([0, 1])$, with function addition and scalar multiplication defined pointwise.
- The space of polynomials with real coefficients $\mathbb{R}[X]$, where addition and scalar multiplication are defined by combining coefficients.
- The space of all $m\times n$ real matrices $\mathbb{R}^{m\times n}$, with standard matrix addition and scalar multiplication.
- The space of infinite sequences of real numbers, with addition and scalar multiplication performed term-by-term.

In Linear Algebra II, we will explore these concepts in their full generality, providing a deeper understanding of vector spaces and laying the foundation for applications across various mathematical disciplines.

Example 45 Consider $\mathcal{F}(\mathbb{R}, \mathbb{R})$ to be the set of all functions $\mathbb{R} \to \mathbb{R}$.
For $f, g \in \mathcal{F}(\mathbb{R}, \mathbb{R})$ we can also define $f + g \in \mathcal{F}(\mathbb{R}, \mathbb{R})$ and $\lambda f \in \mathcal{F}(\mathbb{R}, \mathbb{R})$ by

$$(f + g)(x) = f(x) + g(x) \quad \text{and} \quad (\lambda f)(x) = \lambda f(x).$$

Next, consider the subset $\mathcal{U} = \{f \in \mathcal{F}(\mathbb{R}, \mathbb{R}) \mid f'' = f\}$. Observe that:
(i) The function $n(x) = 0$ (the zero function) is an element of $\mathcal{U}$.
(ii) If $f, g \in \mathcal{U}$, i.e. $f'' = f$ and $g'' = g$, then

$$(f + g)'' = f'' + g'' = f + g \Longrightarrow f + g \in \mathcal{U}.$$

(iii) For $\lambda \in \mathbb{R}$, $f \in \mathcal{U} \Longrightarrow \lambda f \in \mathcal{U}$.
As such, $\mathcal{U}$ can be seen as a **subspace** of $\mathcal{F}(\mathbb{R}, \mathbb{R})$.
By solving the differential equation $f'' = f$ one can show that $f_1(x) = e^x$, $f_2(x) = e^{-x}$, and their linear combinations $\lambda_1 f_1 + \lambda_2 f_2$ ($\lambda_1, \lambda_2 \in \mathbb{R}$) are the only elements in $\mathcal{U}$. As such, we can write $\mathcal{U} = \operatorname{span}\{f_1, f_2\}$.
Moreover, f_1 and f_2 are **linearly independent**, since

$$\lambda_1 f_1 + \lambda_2 f_2 = 0 \quad \Longleftrightarrow \quad \lambda_1 = \lambda_2 = 0.$$

Because the "$\Leftarrow$" statement is obvious, we just prove the "$\Rightarrow$" statement. We can prove it in two ways. One way is substituting some x's into the equation (because the equation is true for every $x \in \mathbb{R}$):

$$\begin{aligned} \lambda_1 f_1 + \lambda_2 f_2 = 0 \quad &\Longrightarrow \quad \begin{cases} \lambda_1 f_1(0) + \lambda_2 f_2(0) = 0 \\ \lambda_1 f_1(1) + \lambda_2 f_2(1) = 0 \end{cases} \quad \Longrightarrow \quad \begin{cases} \lambda_1 + \lambda_2 = 0 \\ \lambda_1 e + \lambda_2 e^{-1} = 0 \end{cases} \\ &\Longrightarrow \quad \begin{cases} \lambda_1 - \lambda_1 e^2 = 0 \\ \lambda_2 = -\lambda_1 e^2 \end{cases} \\ &\Longrightarrow \quad \begin{cases} \lambda_1 = 0 \\ \lambda_2 = 0 \end{cases}. \end{aligned}$$

Another way is using the fact that $f_1' = f_1$ and $f_2' = -f_2$:

$$\begin{aligned} \lambda_1 f_1 + \lambda_2 f_2 = 0 \quad &\Longrightarrow \quad \begin{cases} \lambda_1 f_1 + \lambda_2 f_2 = 0 \\ \lambda_1 f_1' + \lambda_2 f_2' = 0 \end{cases} \quad \Longrightarrow \quad \begin{cases} \lambda_1 f_1 + \lambda_2 f_2 = 0 \\ \lambda_1 f_1 - \lambda_2 f_2 = 0 \end{cases} \\ &\Longrightarrow \quad \begin{cases} \lambda_1 f_1 = 0 \\ \lambda_2 f_2 = 0 \end{cases} \\ &\Longrightarrow \quad \begin{cases} \lambda_1 = 0 \\ \lambda_2 = 0 \end{cases} \quad (f_1,\ f_2 \text{ are not zero functions}). \end{aligned}$$

Therefore, a **basis** of $\mathcal{U}$ is (f_1, f_2), and $\dim \mathcal{U} = 2$.

This example shows just one object in which the concepts in Linear Algebra I can be applied. As such, the Linear Algebra II course aims to generalize these concepts that were introduced in Linear Algebra I for $\mathbb{R}^n$ to more general spaces (vector spaces), discuss the importance of eigenvalues and eigenvectors, and give explicit examples of applications of the concepts taught in this course.

14

Vector Spaces

Roughly speaking, **vector spaces** are defined as spaces where addition and scalar multiplication are defined in such a way that they satisfy the same computational rules as $\mathbb{R}^n$.

Definition 14.1 A (real) **vector space** is a tuple $(V, +, \cdot)$, where V is a set together with two functions

$$+ : V \times V \longrightarrow V \qquad\qquad \cdot : \mathbb{R} \times V \longrightarrow V$$
$$(u, v) \longmapsto u + v \qquad\qquad (\lambda, v) \longmapsto \lambda \cdot v$$

such that the following properties are satisfied:

- Properties of the addition:
  (A.1) $\forall u, v, w \in V\colon\ (u + v) + w = u + (v + w)$. (Associativity)
  (A.2) $\forall u, v \in V\colon\ u + v = v + u$. (Commutativity)
  (A.3) $\exists n \in V,\ \forall u \in V\colon\ n + u = u$. (Identity/neutral element of addition)
  (A.4) $\forall u \in V,\ \exists v \in V\colon\ u + v = n$. (Inverse elements of addition)
- Compatibility of addition and scalar multiplication:
  (C.1) $\forall u, v \in V,\ \lambda \in \mathbb{R}\colon\ \lambda \cdot (u + v) = \lambda \cdot u + \lambda \cdot v$. (Distributivity I)
  (C.2) $\forall u \in V,\ \lambda, \mu \in \mathbb{R}\colon\ (\lambda + \mu) \cdot u = \lambda \cdot u + \mu \cdot u$. (Distributivity II)
  (C.3) $\forall u \in V,\ \lambda, \mu \in \mathbb{R}\colon\ \lambda \cdot (\mu \cdot u) = (\lambda\mu) \cdot u$.
  (C.4) $\forall u \in V\colon\ 1 \cdot u = u$.

We write the vector space as $(V, +, \cdot)$ if we want to emphasize the addition and scalar multiplication operator used. Otherwise, the vector space is commonly written as only V.

Following immediately from Definition 14.1, one observes a few properties of vector spaces:

Proposition 14.2 *Let V be a vector space, with $u \in V$. Then,*

(i) $u + n = u$.
(ii) If $n, \tilde{n} \in V$ both satisfy (A.3) in Definition 14.1, then $n = \tilde{n}$. (The Identity element is unique)
(iii) If for a fixed $u \in V$ the elements $v, \tilde{v} \in V$ both satisfy (A.4), i.e. $u + v = u + \tilde{v} = n$, then $v = \tilde{v}$. (The inverse of an element u is unique)
(iv) $u + (-1) \cdot u = 0$.

Proof. (i) By (A.2), $u + n = n + u = u$.
(ii) By (A.3), $n = \tilde{n} + n$, and by i), $\tilde{n} + n = \tilde{n}$, so $n = \tilde{n}$.
(iii) Consider the expression $\tilde{v} + (u + v)$. By assumption, $\tilde{v} + (u + v) = \tilde{v} + n = \tilde{v}$.
By (A.1), $\tilde{v} + (u + v) = (\tilde{v} + u) + v$. Then, by assumption, $(\tilde{v} + u) + v = n + v = v$. Thus, $v = \tilde{v}$.
(iv) By (C.2), we have that $0 \cdot u = (0 + 0) \cdot u = 0 \cdot u + 0 \cdot u$. Then, by (A.4), $\exists(-(0 \cdot u))$, the inverse of $0 \cdot u$.
Then, adding $(-(0 \cdot u))$ we obtain $n = n + 0 \cdot u \Longrightarrow 0 \cdot u = n$.
By (C.2) we have $u + (-1) \cdot u = (1 - 1) \cdot u = 0 \cdot u = n$. □

Remark. 1) The operations $+$ and $\cdot$ must be defined when specifying a vector space.
2) We often just write λu instead of $\lambda \cdot u$ and it will be clear from the context if we mean scalar multiplication or multiplication of real numbers. The same is true for $+$.
3) By Definition 14.1, $u + v + w = (u + v) + w = u + (v + w)$ is well-defined.
4) By Proposition 14.2 (ii) the neutral element n is unique, and we denote it as 0.
5) By Proposition 14.2 (iii) the inverse of an element $u \in V$ is unique, and we write $v =: -u$. By Proposition 14.2 (iv) we have $(-1)u = -u$.

The element n in (A.3), the identity (or neutral) element of the vector space V is usually (by abuse of notation) also denoted by 0. Going forward, be always aware if 0 means the real number 0, the zero vector, or the identity element of a vector space.

Example 46 A few examples of vector spaces, along with their neutral element and inverses.

1) $V = \mathbb{R}^n$ with the usual operations, namely
$$u + v = \begin{pmatrix} u_1 + v_1 \\ \vdots \\ u_n + v_n \end{pmatrix} \quad \text{and} \quad \lambda u = \begin{pmatrix} \lambda u_1 \\ \vdots \\ \lambda u_n \end{pmatrix}.$$
The neutral element is the zero vector $\mathbf{0} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$ and the inverse is given by $-u = \begin{pmatrix} -u_1 \\ \vdots \\ -u_n \end{pmatrix}$.

2) The set of all real functions, $\mathcal{F}(\mathbb{R}, \mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R}\}$ (see Example 45), with operations
$$(f + g)(x) = f(x) + g(x) \quad \text{and} \quad (\lambda f)(x) = \lambda \cdot f(x).$$
The neutral element n is defined by $n(x) = 0 \ \forall x \in \mathbb{R}$.
The inverse of f is $-f$, defined by $(-f)(x) = -f(x) \ \forall x \in \mathbb{R}$.
In addition, we also have the following space of functions as vector spaces
$$\begin{aligned} \mathcal{C}^0(\mathbb{R}, \mathbb{R}) &= \{f \in \mathcal{F}(\mathbb{R}, \mathbb{R}) \mid f \text{ is continuous}\}, \\ \mathcal{C}^n(\mathbb{R}, \mathbb{R}) &= \{f \in \mathcal{F}(\mathbb{R}, \mathbb{R}) \mid f^{(n)} \text{ exists and is continuous}\}, \\ \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}) &= \{f \in \mathcal{F}(\mathbb{R}, \mathbb{R}) \mid f^{(n)} \text{ exists for all } n \geq 0\}, \\ \mathcal{P} &= \{f \in \mathcal{F}(\mathbb{R}, \mathbb{R}) \mid f \text{ is a polynomial function}\}, \\ \mathcal{P}_n &= \left\{ f \in \mathcal{F}(\mathbb{R}, \mathbb{R}) \mid f(x) = \sum_{j=0}^{n} a_j x^j \text{ for some } a_0, \ldots, a_n \in \mathbb{R} \right\}. \end{aligned}$$
Furthermore, $\mathcal{P}_n \subset \mathcal{P} \subset \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}) \subset \mathcal{C}^n(\mathbb{R}, \mathbb{R}) \subset \mathcal{C}^0(\mathbb{R}, \mathbb{R})$.

3) The set of all infinite sequences $\mathcal{J} = \{(a_n)_{n\geq 1}\}$, with operations for $a = (a_n)_{n\geq 1}$, $b = (b_n)_{n\geq 1} \in \mathcal{J}$, and $\lambda \in \mathbb{R}$,
$$a + b = (a_n + b_n)_{n\geq 1} \quad \text{and} \quad \lambda a = (\lambda a_n)_{n\geq 1}.$$
The neutral element is the sequence with all zeroes (namely, the sequence $0 = (0, 0, \ldots)$).
The inverse of $(a_n)_{n\geq 1}$ is given by $(-a_n)_{n\geq 1} = (-a_1, -a_2, \ldots)$.

4) The set of $m \times n$ matrices, $\mathbb{R}^{m\times n}$, along with the usual addition and scalar multiplication. The neutral element is the zero matrix, while the inverse of an element $(a_{ij}) \in \mathbb{R}^{m\times n}$ is the matrix $(-a_{ij})$.

5) If $(V, +, \cdot)$ is a vector space and $f : V \to W$ is a bijective map, then you can define for $u, v \in W$ and $\lambda \in \mathbb{R}$:
$$\begin{aligned} u \oplus v &= f\left(f^{-1}(u) + f^{-1}(v)\right), \\ \lambda \odot u &= f\left(\lambda \cdot f^{-1}(u)\right) \end{aligned}$$
and obtain a vector space $(W, \oplus, \odot)$.

To prove that the combination of a set and two operators forms a vector space, one must refer to Definition 14.1 and confirm that the alleged vector space satisfies both the conditions on addition and the compatibility conditions between addition and scalar multiplication. This was not done for the five vector spaces mentioned above and is left as a simple exercise.
Next, we define subspaces in this framework of vector spaces.

Definition 14.3 Let V be a vector space. A subset $U \subset V$ is a **subspace** if
(i) $0 \in U$.
(ii) $\forall u, v \in U\colon\ u + v \in U$.
(iii) $\forall u \in U,\ \lambda \in \mathbb{R}\colon\ \lambda u \in U$.

As a reminder, 0 refers to the neutral element, not the real number 0 (as not all vector spaces contain the real number 0). Due to conditions (ii) and (iii), a subspace is said to be *closed under addition and scalar multiplication* (identical to Linear Algebra I subspaces).

Example 47 Some subsets of vector spaces, and whether or not they are subspaces.
1) $\mathcal{P} = \{f : \mathbb{R} \to \mathbb{R} \mid f \text{ is a polynomial function}\}$ is a subspace of $\mathcal{F}(\mathbb{R}, \mathbb{R})$.
2) Subspaces of $\mathbb{R}^n$.
3) $\mathcal{J}^0 = \{(a_n)_{n\geq 1} \mid \lim_{n\to\infty} a_n \text{ exists}\}$ is a subspace of $\mathcal{J} = \{(a_n)_{n\geq 1}\}$.
4) $\mathrm{GL}_n(\mathbb{R}) = \left\{A \in \mathbb{R}^{n\times n} \mid A^{-1} \text{ exists}\right\} \subset \mathbb{R}^{n\times n}$ is **NOT** a subspace of $\mathbb{R}^{n\times n}$ because the zero matrix is not in $\mathrm{GL}_n(\mathbb{R})$.

To prove if a subset is a subspace, one must check if it satisfies the conditions in Definition 14.3.

Proposition 14.4 *If $U \subset V$ is a subspace, then U is also a vector space with the operations inherited from V.*

Proof. The proof of this proposition follows directly from the definition of a subspace; since subspaces contain the neutral element and are closed under addition and scalar multiplication, conditions (A.1) to (A.4) and (C.1) to (C.4) will be satisfied. An explicit proof can be done as a small exercise. □

We now introduce some familiar concepts, namely spans and basis, in this framework of vector spaces.

Definition 14.5 Let V be a vector space and $v_1, \dots, v_n \in V$.
(i) The **span** of the elements $v_1, \dots, v_n$ is given by the set of all their **linear combinations**,

$$\operatorname{span}\{v_1, \dots, v_n\} = \left\{ \sum_{i=1}^{n} \lambda_i v_i \in V \ \middle|\ \lambda_1, \dots, \lambda_n \in \mathbb{R} \right\}.$$

(ii) The elements $v_1, \dots, v_n$ **span (or generate) the space** V if $\operatorname{span}\{v_1, \dots, v_n\} = V$.
(iii) V is **finitely generated** if there exist $v_1, \dots, v_n \in V$ with $\operatorname{span}\{v_1, \dots, v_n\} = V$. (i.e. one just needs finitely many elements to generate the space)
(iv) The elements $v_1, \dots, v_n$ are **linearly independent** if

$$\lambda_1 v_1 + \cdots + \lambda_n v_n = 0 \implies \lambda_1 = \ldots = \lambda_n = 0\,.$$

(v) $B = (v_1, \dots, v_n)$ is a **basis** of V if $v_1, \dots, v_n$ are linearly independent and $\operatorname{span}\{v_1, \dots, v_n\} = V$.

On the other hand, a vector space V that could not be spanned by a finite number of elements is not finitely generated. Such spaces do not have a finite basis.
By convention, the vector space $\{0\}$, spanned by the empty set, is a finitely-generated space.

Example 48 Some subspaces, and whether or not they are finitely generated.
1) $\mathbb{R}^n$ is finitely generated, since $\mathbb{R}^n = \operatorname{span}\{e_1, \dots, e_n\}$.
2) $\mathcal{P} = \left\{f : \mathbb{R} \to \mathbb{R} \mid f(x) = \sum_{i=1}^{m} a_i x^i \quad a_1, \dots, a_m \in \mathbb{R}\right\}$ is not finitely generated.

Suppose that $f_1, \ldots, f_m \in \mathcal{P}$ generates $\mathcal{P}$, with $d = \max\{\deg(f_i) \mid i = 1, \ldots, m\}$. Then, $g(x) = x^{d+1} \in \mathcal{P}$ would not be in $\operatorname{span}\{f_1, \ldots, f_m\}$.

3) $\mathcal{P}_n = \{f \in \mathcal{P} \mid \deg(f) \leq n\}$ is finitely generated with basis $B = (1, x, \ldots, x^n)$. We will show that $1, x, \ldots, x^n$ are linearly independent later by using determinants.

Most of the following propositions, lemmas, and theorems are generalizations of results in $\mathbb{R}^n$ (LA1) to those in the framework of abstract vector spaces. The proofs of these statements are mostly identical to the proofs in LA1.

Proposition 14.6 *Let V be a vector space and $v_1, \ldots, v_n \in V$. These statements are equivalent:*

(i) $v_1, \ldots, v_n$ *are linearly dependent.*

(ii) There exist a $1 \leq j \leq n$ *such that* $v_j \in \operatorname{span}\{v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_n\}$.

(iii) There exist a $1 \leq j \leq n$ *such that* $\operatorname{span}\{v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_n\} = \operatorname{span}\{v_1, \ldots, v_n\}$.

Proof. The proof of this proposition is exactly the same as the proof for $V = \mathbb{R}^n$ (LA1, Theorem 9.3). □

Lemma 14.7 *If $v_1, \ldots, v_l \in V$ are linearly independent and $V = \operatorname{span}\{w_1, \ldots, w_m\}$, then $l \leq m$.*

Proof. Suppose by contradiction that $l > m$. Because $v_1, \ldots, v_l \in V = \operatorname{span}\{w_1, \ldots, w_m\}$, we have

$$\begin{cases} v_1 = a_{11}w_1 + a_{12}w_2 + \ldots + a_{1m}w_m \\ v_2 = a_{21}w_1 + a_{22}w_2 + \ldots + a_{2m}w_m \\ \qquad\qquad\vdots \\ v_l = a_{l1}w_1 + a_{l2}w_2 + \ldots + a_{lm}w_m \end{cases},$$

for some $a_{ij} \in \mathbb{R}$ with $1 \leq i \leq l$ and $1 \leq j \leq m$. Now, consider the equation

$$\sum_{i=1}^{l} \lambda_i v_i = 0,$$

which is equivalent to

$$\sum_{i=1}^{l} \lambda_i \left(\sum_{j=1}^{m} a_{ij} w_j \right) = 0.$$

Rearranging the sum, we get

$$\sum_{j=1}^{m} \left(\sum_{i=1}^{l} a_{ij} \lambda_i \right) w_j = 0. \tag{$*$}$$

The solution $(\lambda_1, \ldots, \lambda_l)$ to the following linear system

$$\begin{cases} a_{11}\lambda_1 + a_{21}\lambda_2 + \ldots + a_{l1}\lambda_l = 0 \\ a_{12}\lambda_1 + a_{22}\lambda_2 + \ldots + a_{l2}\lambda_l = 0 \\ \qquad\qquad\vdots \\ a_{1m}\lambda_1 + a_{2m}\lambda_2 + \ldots + a_{lm}\lambda_l = 0 \end{cases}. \tag{$**$}$$

also satisfies $(*)$. Because there are more variables than equations ($l > m$), the linear system $(**)$ has some free variables, which implies that there are infinitely many solutions to $(**)$. Hence, there exists a non-zero solution to $(**)$ and therefore, there exist $\lambda_1, \ldots, \lambda_n$ that do not simultaneously vanish and satisfy $(*)$. As a result, $v_1, \ldots, v_l$ are linearly dependent, which contradicts the assumption. Consequently, $l \leq m$. □

Theorem 14.8 *Let V be a finitely generated vector space. Then we have the following*

(i) V has a (finite) basis.

(ii) All bases of V have the same number of elements.

(iii) If $v_1, \dots, v_l \in V$ are linearly independent then there exist $v_{l+1}, \dots, v_n \in V$ such that $(v_1, \dots, v_n)$ is a basis of V.

(iv) If $V = \operatorname{span}\{w_1, \dots, w_m\}$, then there exist a subset $\{u_1, \dots, u_l\} \subset \{w_1, \dots, w_m\}$, such that $(u_1, \dots, u_l)$ is a basis of V.

Proof. The proof of this proposition is exactly the same as the proof for $V = \mathbb{R}^n$ (LA1, Theorem 10.2). □

Definition 14.9 Let V be a finitely generated vector space with basis $(v_1, \dots, v_n)$. Then $\dim(V) = n$ is the **dimension of V**.

Corollary 14.10 *Let V be a vector space with $\dim(V) = n$ and $v_1, \dots, v_n \in V$. Then the following statements are equivalent.*

(i) $v_1, \dots, v_n$ are linearly independent.

(ii) $V = \operatorname{span}\{v_1, \dots, v_n\}$.

(iii) $(v_1, \dots, v_n)$ is a basis of V.

Proof. The proof of this proposition is exactly the same as the proof for $V = \mathbb{R}^n$ (LA1, Corollary 10.4). □

Proposition 14.11 *Let V be finitely generated and $U \subset V$ a subspace. Then U is also finitely generated.*

Proof. Suppose $V = \operatorname{span}\{w_1, w_2, \dots, w_n\}$ with $w_1, w_2, \dots, w_n \in V$. We prove the statement by doing the following process:
Step 0**:** If $U = \{0\}$, then U is finitely generated and the process ends. Otherwise, if $U \neq \{0\}$, we can choose a non-zero vector $v_1 \in U$, which is linearly independent by itself.
Step l $(l \geq 1)$**:** Suppose that we have chosen l linearly independent vectors $v_1, \dots, v_l \in U$. If $U = \operatorname{span}\{v_1, \dots, v_l\}$, then U is finitely generated, and the process ends. Otherwise, if $U \neq \operatorname{span}\{v_1, \dots, v_l\}$, then we can choose a non-zero vector $v_{l+1} \in U$ such that $v_{l+1} \notin \operatorname{span}\{v_1, \dots, v_l\}$. As a result, we get $l+1$ linearly independent vectors $v_1, \dots, v_{l+1}$ (Lemma 9.5 in chapter 9, which can be proved for the case of abstract vector spaces similarly to the case of $\mathbb{R}^n$).

Due to Lemma 14.7, the number of linearly independent vectors in V cannot exceed the number of vectors spanning V. Therefore, we can only obtain at most n linearly independent vectors spanning U by the above process. Consequently, U is finitely generated. □

The contrapositive statement of Preposition 14.11 is that if U is not finitely generated, then V is also not finitely generated. An application is given below.

Example 49 Applications of Proposition 14.11

1) The space of $m \times n$ matrices, $\mathbb{R}^{m \times n}$ has a basis $(E_{ij})_{i,j}$, so $\dim \mathbb{R}^{m \times n} = mn$. Here, E_{ij} is a matrix of all zeros except for one 1 entry in the i^{th} row and the j^{th} column.
   As a consequence of Proposition 14.11, all subspaces of this space are finitely generated.
2) The space of all real functions $\mathcal{F}(\mathbb{R}, \mathbb{R})$ is not finitely generated since one of its subspaces, the set of all polynomial functions $\mathcal{P}$, is not finitely generated (see Example 48).

Proposition 14.12 *Let $B = (v_1, \dots, v_n)$ be a basis of V. Then for all $u \in V$ there exist unique $\lambda_1, \dots, \lambda_n \in \mathbb{R}$, such that*

$$u = \sum_{i=1}^{n} \lambda_i v_i \,.$$

Proof. Since $B = (v_1, \dots, v_n)$ is a basis of V, there always exists for any $u \in V$ a set of real numbers $\lambda_1, \dots, \lambda_n$ such that $u = \sum_{i=1}^n \lambda_i v_i$. We only have to prove the uniqueness of this set of numbers for each $u \in V$.
Suppose that for any $u \in V$, there exist two sets of real numbers $\lambda_1, \dots, \lambda_n$ and $\mu_1, \dots, \mu_n$ such that

$$u = \sum_{i=1}^n \lambda_i v_i = \sum_{i=1}^n \mu_i v_i.$$

Then we have

$$\sum_{i=1}^n (\lambda_i - \mu_i) v_i = 0.$$

Since $v_1, \dots, v_n$ are linearly independent, we have $\lambda_i = \mu_i$ for $1 \leq i \leq n$. This completes the proof. □

Due to this proposition, it is natural to have the following definition.

Definition 14.13 Let $B = (v_1, \dots, v_n)$ be a basis of V.

(i) The $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ in Proposition 14.12 are called the **coordinates** of $u \in V$ in the basis B.

(ii) The vector $[u]_B \in \mathbb{R}^n$ given by

$$[u]_B = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$$

is called the **coordinate vector** of u with respect to the basis B.

15

Linear maps

In the following we consider vector spaces $(V, +_V, \cdot_V)$ and $(W, +_W, \cdot_W)$. We denote 0_V and 0_W for the neutral elements in $(V, +_V, \cdot_V)$ and $(W, +_W, \cdot_W)$, respectively. Later we will usually just write "+", "$\cdot$", and "0".

Definition 15.1 Let $(V, +_V, \cdot_V)$, $(W, +_W, \cdot_W)$ be vector spaces. A **linear map** is a function $F : V \to W$ satisfying

(i) $F(u +_V v) = F(u) +_W F(v)$ for all $u, v \in V$.
(ii) $F(\lambda \cdot_V u) = \lambda \cdot_W F(u)$ for all $u \in V, \lambda \in \mathbb{R}$.

In LA1, we have studied the case where $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$. In this case, any linear map $F : V \to W$ can be expressed as the multiplication of vectors by a matrix $[F] = A \in \mathbb{R}^{m\times n}$. However, in general setup, it is not straightforward to have a matrix multiplied with a vector in V to get a vector in W since V and W can contain arbitrary objects such as polynomial functions and matrices. Nevertheless, for any finitely-generated vector spaces, we can get matrices of linear maps by establishing an equivalence relation, called isomorphism, between a vector space V and $\mathbb{R}^{\dim(V)}$. We will discuss this topic later in this chapter and the next chapter. Now, we revisit other notions from LA1.

Definition 15.2 Let $F : V \to W$ be a linear map.

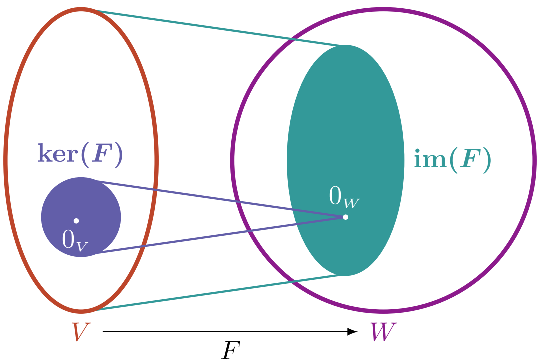
(i) The **kernel of** F is given by

$$\ker(F) = \{u \in V \mid F(u) = 0_W\} \subset V .$$

(ii) The **image of** F is given by

$$\operatorname{im}(F) = \{w \in W \mid \exists u \in V : w = F(u)\} \subset W .$$

With the same arguments as in the $\mathbb{R}^n$-case we see that $\ker(F)$ is a subspace of V and $\operatorname{im}(F)$ is a subspace of W. If $\operatorname{im}(F)$ is finitely generated, we define the **rank of** F by $\operatorname{rk}(F) = \dim(\operatorname{im}(F))$.



Example 50 1) For $n \geq 1$, consider the map

$$\begin{aligned} D : \mathcal{C}^n(\mathbb{R},\mathbb{R}) &\longrightarrow \mathcal{F}(\mathbb{R},\mathbb{R}), \\ f &\longmapsto f'. \end{aligned}$$

D is a linear map since for any $f, g \in \mathcal{C}^n(\mathbb{R},\mathbb{R})$ and $\lambda \in \mathbb{R}$,
(i) $D(f+g) = (f+g)' = f' + g' = D(f) + D(g)$,
(ii) $D(\lambda f) = (\lambda f)' = \lambda f' = \lambda D(f)$.
The kernel and the image of D are as follows

$$\begin{aligned} \ker(D) &= \text{ The set of constant functions } = \mathcal{P}_0, \\ \operatorname{im}(D) &= \mathcal{C}^{n-1}(\mathbb{R},\mathbb{R}). \end{aligned}$$

2) Consider the map

$$\begin{aligned} \mathrm{ev}_a : \mathcal{F}(\mathbb{R},\mathbb{R}) &\longrightarrow \mathbb{R}. \\ f &\longmapsto f(a). \end{aligned}$$

ev_a is linear for any $a \in \mathbb{R}$ because
(i) $\mathrm{ev}_a(f+g) = (f+g)(a) = f(a) + g(a) = \mathrm{ev}_a(f) + \mathrm{ev}_a(g)$,
(ii) $\mathrm{ev}_a(\lambda f) = (\lambda f)(a) = \lambda f(a) = \lambda \cdot \mathrm{ev}_a(f)$.
The kernel and the image of ev_a are as follows:

$$\begin{aligned} \ker(\mathrm{ev}_a) &= \{f \in \mathcal{F}(\mathbb{R},\mathbb{R}) \mid f(a) = 0\}, \\ \operatorname{im}(\mathrm{ev}_a) &= \mathbb{R}. \end{aligned}$$

3) Consider the map

$$\begin{aligned} F : \mathbb{R}^{2\times 2} &\longrightarrow \mathbb{R}^2, \\ \begin{pmatrix} a & b \\ c & d \end{pmatrix} &\longmapsto \begin{pmatrix} a \\ d \end{pmatrix}. \end{aligned}$$

It is easy to see that F is linear. Then, we have

$$\begin{aligned} \ker(F) &= \operatorname{span}\left\{\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\right\}, & \dim(\ker(F)) &= 2, \\ \operatorname{im}(F) &= \mathbb{R}^2, & \dim(\operatorname{im}(F)) &= 2. \end{aligned}$$

Here, we see that $\dim(\mathbb{R}^{2\times 2}) = 4 = 2 + 2 = \dim(\ker(F)) + \dim(\operatorname{im}(F))$.

Theorem 15.3 (kernel-image theorem) *Let V be finitely generated and let $F : V \to W$ be a linear map to an arbitrary vector space W. Then*

$$\dim V = \dim(\ker(F)) + \dim(\operatorname{im}(F)).$$

Proof. Because $\ker(F) \subset V$ is a subspace and V is finitely generated, $\ker(F)$ is also finitely generated by Proposition 14.11. Therefore, there exists a basis $B_1 = (v_1, \ldots, v_l)$ of $\ker(F)$. By Theorem 14.8 (iii), there exist $u_1, \ldots, u_m \in V$ such that $B = (v_1, \ldots, v_l, u_1, \ldots, u_m)$ is a basis of V. Now, we want to find a basis of $\operatorname{im}(F)$. We have

$$\begin{aligned} \mu_1 F(u_1) + \ldots + \mu_m F(u_m) = 0 \quad &\Rightarrow \quad F(\mu_1 u_1 + \ldots + \mu_m u_m) = 0 \\ &\Rightarrow \quad \mu_1 u_1 + \ldots + \mu_m u_m \in \ker(F) = \operatorname{span}\{v_1, \ldots, v_l\} \\ &\Rightarrow \quad \mu_1 u_1 + \ldots + \mu_m u_m = \lambda_1 v_1 + \ldots + \lambda_l v_l \\ &\Rightarrow \quad -\lambda_1 v_1 - \ldots - \lambda_l v_l + \mu_1 u_1 + \ldots + \mu_m u_m = 0. \end{aligned}$$

Because $v_1, \ldots, v_l, u_1, \ldots, u_m$ are linearly independent, we have $\lambda_1 = \ldots = \lambda_l = \mu_1 = \ldots = \mu_m = 0$.

Hence,

$$\mu_1 F(u_1) + \ldots + \mu_m F(u_m) = 0 \quad \Rightarrow \quad \mu_1 = \ldots = \mu_m = 0$$

and therefore, $F(u_1), \ldots, F(u_m)$ are linearly independent. For any $u \in \operatorname{im}(F)$, there exists $v \in V = \operatorname{span}\{v_1, \ldots, v_l, u_1, \ldots, u_m\}$ such that

$$u = F(v) = F(\lambda_1 v_1 + \ldots + \lambda_l v_l + \mu_1 u_1 + \ldots + \mu_m u_m) = F(\lambda_1 v_1 + \ldots + \lambda_l v_l) + F(\mu_1 u_1 + \ldots + \mu_m u_m).$$

Because $\lambda_1 v_1 + \ldots + \lambda_l v_l \in \ker(F)$, we have

$$u = 0 + F(\mu_1 u_1 + \ldots + \mu_m u_m) = \mu_1 F(u_1) + \ldots + \mu_m F(u_m).$$

Thus, $\operatorname{im}(F) = \operatorname{span}\{F(u_1), \ldots, F(u_m)\}$ and $B_2 = (F(u_1), \ldots, F(u_m))$ is a basis of $\operatorname{im}(F)$. As a result,

$$\dim(\ker(F)) + \dim(\operatorname{im}(F)) = l + m = \dim(V).$$

□

Example 51 Consider the linear map

$$\begin{aligned} D : \mathcal{P}_2 &\longrightarrow \mathcal{P}_2 \\ f &\longmapsto f' \end{aligned}$$

For any $f \in \mathcal{P}_2$, it can be written for any $x \in \mathbb{R}$ as $f(x) = ax^2 + bx + c$ for some $a, b, c \in \mathbb{R}$. Then, we have

$$(D(f))(x) = f'(x) = 2ax + b.$$

Therefore, $\ker(D) = \mathcal{P}_0$, $\operatorname{im}(D) = \mathcal{P}_1$. Notice: $\mathcal{P}_2$ "behaves" like $\mathbb{R}^3$ because we can associate, one by one, $f(x) = ax^2 + bx + c$ with a vector $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$. We can view D as a linear map

$$\begin{aligned} F : \mathbb{R}^3 &\longrightarrow \mathbb{R}^3, \\ \begin{pmatrix} a \\ b \\ c \end{pmatrix} &\longmapsto \begin{pmatrix} 0 \\ 2a \\ b \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix}. \end{aligned}$$

We make the above statement about $\mathcal{P}_2$ and $\mathbb{R}^3$ precise by the definition of isomorphism as follows:

Definition 15.4 (i) (Recall) A function $f : X \to Y$ is **invertible** if there exist a function $g : Y \to X$ such that $f \circ g = \mathrm{id}_Y$ and $g \circ f = \mathrm{id}_X$. f is invertible iff f is bijective, i.e. injective and surjective.

(ii) An invertible linear map $F : V \to W$ is called an **isomorphism**.

(iii) Two vector spaces V and W are called **isomorphic** (Notation: $V \cong W$) if there exists an isomorphism $F : V \to W$.

By this definition, $\mathcal{P}_2 \cong \mathbb{R}^3$ in Example 51.

Theorem 15.5 *(i) A linear map $F : V \to W$ is an isomorphism iff $\ker(F) = \{0_V\}$ (F is injective) and $\operatorname{im}(F) = W$ (F is surjective).*

(ii) Let $F : V \to W$ be an isomorphism and $(b_1, \ldots, b_n)$ a basis of V. Then $(F(b_1), \ldots, F(b_n))$ is a basis of W.

(iii) Let V, W be finitely generated and $V \cong W$ then $\dim(V) = \dim(W)$.

(iv) Let V, W be finitely generated and $\dim(V) = \dim(W)$. Then for a linear map $F : V \to W$ the following three statements are equivalent

(a) F is an isomorphism.

(b) $\ker(F) = \{0_V\}$.

(c) $\operatorname{im}(F) = W$.

Example 52 Examples of linear maps that are isomorphism and those that are not isomorphisms are given as follows.
1) The following linear maps are isomorphisms:

$$F: \quad \mathbb{R}^{2\times 2} \longrightarrow \mathbb{R}^4, \qquad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \longmapsto \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix},$$

$$G: \quad \mathbb{R}^3 \longrightarrow \mathcal{P}_2, \qquad \begin{pmatrix} a \\ b \\ c \end{pmatrix} \longmapsto ax^2 + bx + c.$$

2) For $n \geq 1$, the following linear map

$$D : \mathcal{C}^n(\mathbb{R}, \mathbb{R}) \longrightarrow \mathcal{C}^{n-1}(\mathbb{R}, \mathbb{R}), \quad f \longmapsto f'$$

is not an isomorphism. Indeed, since $\ker(D) = \{f : \mathbb{R} \to \mathbb{R} \mid f \text{ is a constant function}\} \neq \{0\}$, the linear map D is not injective.

Proposition 15.6 *Let V be finitely generated with basis $B = (b_1, \dots, b_n)$, i.e.* $\dim(V) = n$. *Then the* **coordinate map**

$$c_B : \mathbb{R}^n \longrightarrow V, \qquad \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} \longmapsto \sum_{i=1}^{n} \lambda_i b_i$$

is an isomorphism. The inverse is given by $c_B^{-1}(u) = [u]_B$ for $u \in V$.

As a result, for any finite-generated vector space V, we have $V \cong \mathbb{R}^n$ for some $n \geq 1$.

Example 53 Consider $V = \mathcal{P}_2$ with the basis $B = (x+1, x^2-1, x+3) = (b_1(x), b_2(x), b_3(x))$. We have the coordinate map

$$c_B : \mathbb{R}^3 \longrightarrow \mathcal{P}_2, \qquad \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} \longmapsto \lambda_1 b_1 + \lambda_2 b_2 + \lambda_3 b_3,$$

where for $u = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} \in \mathbb{R}^3$ and $x \in \mathbb{R}$,

$$(c_B(u))(x) = \lambda_1(x+1) + \lambda_2(x^2-1) + \lambda_3(x+3) = \lambda_2 x^2 + (\lambda_1 + \lambda_3)x + (\lambda_1 - \lambda_2 + 3\lambda_3).$$

For example, with $u = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$, we have $(c_B(u))(x) = -x^2 + 3x + 8$.

Hence, if $f(x) = -x^2 + 3x + 8$, then $[f]_B = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$.

Corollary 15.7 *Let V, W be finitely generated. Then the following two statements are equivalent*

(i) $V \cong W$.

(ii) $\dim(V) = \dim(W)$.

Proof. The statement that (i) implies (ii) is Theorem 15.5 (iii). Therefore, we only need to prove the converse.

Assume that two vector spaces V and W have the same dimension $\dim(V) = \dim(W) = n$. Let $B_V = (v_1, \ldots, v_n)$ and $B_W = (w_1, \ldots, w_n)$ be bases of V and W, respectively. By Proposition 15.6, $c_{B_V} : \mathbb{R}^n \to V$ and $c_{B_W} : \mathbb{R}^n \to W$ are isomorphisms. Similar to LA1, we can prove that compositions between linear maps and their inverses are also linear by using the definition of linear maps (see Theorem 6.1 and Proposition 7.2). As a result, $c_{B_W} \circ c_{B_V}^{-1} : V \to W$ is linear and invertible, and therefore, it is an isomorphism from V to W. Consequently, we have $V \cong W$. □

Example 54 Consider the subspace $U = \{f \in \mathcal{P}_3 \mid f(1) = f(2) = 0\} \subset \mathcal{P}_3$. A possible basis of U is $B = (b_1, b_2)$, where

$$b_1(x) = (x-1)(x-2), \qquad\qquad b_2(x) = (x-1)(x-2)(x+3)$$
$$= x^2 - 3x + 2 \qquad\qquad = x^3 - 7x + 6.$$

Hence, $\dim(U) = 2$ and by Corollary 15.7, $U \cong \mathbb{R}^2$.

Let $f(x) = (x-1)^2(x-2)$. What is $[f]_B$? First of all, we need to write f as a linear combination of b_1 and b_2. For all $x \in \mathbb{R}$,

$$\begin{aligned} f(x) &= \lambda_1 b_1(x) + \lambda_2 b_2(x) \\ \Leftrightarrow \quad (x-1)^2(x-2) &= \lambda_1(x-1)(x-2) + \lambda_2(x-1)(x-2)(x+3) \\ \Leftrightarrow \quad x-1 &= \lambda_1 + \lambda_2(x+3) \\ \Leftrightarrow \quad 0 &= (\lambda_2 - 1)x + (\lambda_1 + 3\lambda_2 + 1) \end{aligned}$$

By the linear independence of the functions $f_1(x) = x$ and $f_2(x) = 1$, we have

$$\begin{cases} \lambda_2 - 1 = 0 \\ (\lambda_2 - 1)x + \lambda_1 + 3\lambda_2 + 1 = 0 \end{cases} \Leftrightarrow \begin{cases} \lambda_1 = -4 \\ \lambda_2 = 1 \end{cases}.$$

Thus, $[f]_B = \begin{pmatrix} -4 \\ 1 \end{pmatrix}$.

16

The matrix of a linear map

In the following V and W are finitely generated vector spaces. The goal of chapter is to define the matrix of a linear map $F : V \to W$, where $V \cong \mathbb{R}^n$ and $W \cong \mathbb{R}^m$. There are a lot of choices for this matrix depending on bases of V and W.

Definition 16.1 Let $B_V = (v_1, \dots, v_n)$ be a basis of V, $B_W = (w_1, \dots, w_m)$ be a basis of W and let $F : V \to W$ be a linear map. The **matrix of F with respect to B_V and B_W** is defined by

$$[F]_{B_V}^{B_W} = \left[c_{B_W}^{-1} \circ F \circ c_{B_V}\right].$$

Here $c_{B_W}^{-1} \circ F \circ c_{B_V}$ is the linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ for which the corresponding matrix was defined before. We have the following diagram

$$\begin{tikzcd} V \arrow[r, "F"] & W \arrow[d, "c_{B_W}^{-1}"] \\ \mathbb{R}^n \arrow[u, "c_{B_V}"] \arrow[r, "c_{B_W}^{-1} \circ F \circ c_{B_V}"'] & \mathbb{R}^m \end{tikzcd}$$

Let $G = c_{B_W}^{-1} \circ F \circ c_{B_V}$. From Chapter 11, we have

$$[G] = \begin{pmatrix} | & & | \\ G(e_1) & \cdots & G(e_n) \\ | & & | \end{pmatrix}.$$

For $1 \le j \le n$, we have

$$G(e_j) = c_{B_W}^{-1}\left(F\left(c_{B_V}(e_j)\right)\right) = c_{B_W}^{-1}\left(F\left(v_j\right)\right) = [F(v_j)]_{B_W}.$$

Hence, we get

$$[F]_{B_V}^{B_W} = \begin{pmatrix} | & & | \\ [F(v_1)]_{B_W} & \dots & [F(v_n)]_{B_W} \\ | & & | \end{pmatrix}.$$

In other words: The j-th column of $[F]_{B_V}^{B_W}$ is given by the vector $\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{pmatrix}$, where $F(v_j) = \sum_{i=1}^m \lambda_i w_i$.

In the special case that $V = W$ and $F = \mathrm{id}_V$, we have the following definition.

Definition 16.2 Let $B_1 = (v_1, \dots, v_n)$ and $B_2 = (u_1, \dots, u_n)$ be bases of V. The **change-of-basis matrix from B_1 to B_2** is the matrix

$$S_{B_1}^{B_2} = [\mathrm{id}_V]_{B_1}^{B_2} = [c_{B_2}^{-1} \circ c_{B_1}] = \begin{pmatrix} | & & | \\ [v_1]_{B_2} & \dots & [v_n]_{B_2} \\ | & & | \end{pmatrix}.$$

For any $v \in V$, we have $S_{B_1}^{B_2}[v]_{B_1} = [v]_{B_2}$.

Example 55 Consider the linear map

$$\begin{aligned} D : \mathcal{P}_2 &\longrightarrow \mathcal{P}_2, \\ f &\longmapsto f', \end{aligned}$$

and two bases of $\mathcal{P}_2$, which are

$$B = \left(1, x, x^2\right) = (b_1(x), b_2(x), b_3(x)) \quad \text{and} \quad C = \left(2, 2x+4, x^2\right) = (c_1(x), c_2(x), c_3(x)).$$

We would like to calculate $[D]_B^B$, $[D]_B^C$, $[D]_C^C$, $[D]_C^B$. For every $f \in \mathcal{P}_2$, we have $f(x) = ax^2 + bx + c$ for some $a,\ b,\ c \in \mathbb{R}$ and

$$f(x) = ax^2 + bx + c \quad \xmapsto{\quad D \quad} \quad f'(x) = 2ax + b.$$

Then, we write the coordinate vectors of $f,\ f'$ in the bases $B,\ C$ and get the following diagram.

$$\begin{tikzcd}
[f]_B = \begin{pmatrix} c \\ b \\ a \end{pmatrix} \arrow[r, mapsto, "{[D]_B^B}"] \arrow[dr, mapsto, "{[D]_B^C}"] & \begin{pmatrix} b \\ 2a \\ 0 \end{pmatrix} = [D(f)]_B \\
[f]_C = \begin{pmatrix} \frac{1}{2}c - b \\ \frac{1}{2}b \\ a \end{pmatrix} \arrow[u, mapsto, "S_C^B"] \arrow[r, mapsto, "{[D]_C^C}"'] \arrow[ur, mapsto, "{[D]_C^B}"'] & \begin{pmatrix} \frac{1}{2}b - 2a \\ a \\ 0 \end{pmatrix} = [D(f)]_C
\end{tikzcd}$$

Now, we calculate the matrices:

$$[D]_B^B = \begin{pmatrix} | & | & | \\ [D(b_1)]_B & [D(b_2)]_B & [D(b_3)]_B \\ | & | & | \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix},$$

$$[D]_B^C = \begin{pmatrix} | & | & | \\ [D(b_1)]_C & [D(b_2)]_C & [D(b_3)]_C \\ | & | & | \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{2} & -2 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$

where

$$D(b_1) = 0, \qquad [D(b_1)]_B = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \qquad [D(b_1)]_C = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

$$D(b_2) = 1, \qquad [D(b_2)]_B = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad [D(b_2)]_C = \begin{pmatrix} \frac{1}{2} \\ 0 \\ 0 \end{pmatrix},$$

$$D(b_3) = 2x, \qquad [D(b_2)]_B = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}, \qquad [D(b_3)]_C = \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}.$$

Next, we calculate the change of basis matrix:

$$S_C^B = \begin{pmatrix} | & | & | \\ [c_1]_B & [c_2]_B & [c_3]_B \\ | & | & | \end{pmatrix} = \begin{pmatrix} 2 & 4 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Then, we have

$$[D]_C^C = [D]_B^C S_C^B = \begin{pmatrix} 0 & \frac{1}{2} & -2 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 2 & 4 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & -2 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$

$$[D]_C^B = [D]_B^B S_C^B = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 2 & 4 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 2 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Remark. 1) We have $\left(S_{B_1}^{B_2}\right)^{-1} = S_{B_2}^{B_1}$.

2) If we have bases $B_1,\ B_2,\ B_3,\ B_4$ of V and a linear map $F : V \to V$, then

$$[F]_{B_1}^{B_4} = S_{B_3}^{B_4} [F]_{B_2}^{B_3} S_{B_1}^{B_2}.$$

Often, $B_1 = B_4$ and $B_2 = B_3$. In that case, we use the notation $[F]_{B_1} := [F]_{B_1}^{B_1}$. Then,

$$[F]_{B_1} = \left(S_{B_1}^{B_2}\right)^{-1} [F]_{B_2} S_{B_1}^{B_2}.$$

3) Using the above formula, we have

$$\begin{aligned} [F]_{B_1}^2 &= \left(S_{B_1}^{B_2}\right)^{-1} [F]_{B_2} \cancel{S_{B_1}^{B_2} \left(S_{B_1}^{B_2}\right)^{-1}} [F]_{B_2} S_{B_1}^{B_2} \\ &= \left(S_{B_1}^{B_2}\right)^{-1} [F]_{B_2}^2 S_{B_1}^{B_2}. \end{aligned}$$

The motivation for this formula is that, in applications, we want to calculate $[F]_{B_1}^n$ for any basis B_1 of V and $n \geq 1$. For this, we try to find another basis B_2 such that $[F]_{B_2}^n$ is easy to calculate. For example, we can try to find B_2 such that $[F]_{B_2}$ is a diagonal matrix since it is easy to calculate n-th power of diagonal matrices.

17

Determinants

17.1 Mathematical Induction

Given a statement, denoted P_n, depending on a natural number n, our goal is proving that P_n is true for all n. To accomplish that, we do the following procedure.
1) Base step: Show that P_1 is true.
2) Induction step: Assume that P_m is true for all $m < n$. Then, show that P_n is true.
Henceforth, 1) and 2) show that P_n is true for all $n \geq 1$ because

$$P_1 \overset{2)}{\Longrightarrow} P_2 \overset{2)}{\Longrightarrow} P_3 \overset{2)}{\Longrightarrow} \dots$$

where P_1 is true due to 1).
Remark. If you want to show that P_n is just true for all $n \geq K$, then you prove P_K in 1) and assume P_m is true for $K \leq m < n$ in 2).

Example 56 (Sum of the first n odd numbers) We see the following pattern.

$$\begin{aligned} n = 1: &\quad 1 = 1^2 \\ n = 2: &\quad 1 + 3 = 2^2 \\ n = 3: &\quad 1 + 3 + 5 = 3^2 \\ &\vdots \end{aligned}$$

Hence, we propose the statement

$$P_n : \sum_{i=1}^{n} (2i - 1) = n^2.$$

We prove P_n for all n by induction:
1) Base step $(n = 1)$: $P_n : \sum_{i=1}^{1} (2i - 1) = 1 = 1^2$. Hence, P_1 is true.
2) Induction step: Fix n and assume P_m is true for all $m < n$. In particular, for $m = n - 1$, we assume that

$$\sum_{i=1}^{n-1} (2i - 1) = (n - 1)^2.$$

Then, we get

$$\begin{aligned}\sum_{i=1}^{n}(2i-1) &= 2n-1+\sum_{i=1}^{n-1}(2i-1)\\ &= 2n-1+(n-1)^2\\ &= 2n-1+(n^2-2n+1)\\ &= n^2.\end{aligned}$$

Hence, we conclude that P_n is true for all n.

17.2 Determinants

Definition 17.1 A **pattern** in an $n \times n$-matrix is a choice of entries, such that precisely one entry from each row and each column is chosen.
Notation: $P = \{(i_1, j_1), \ldots, (i_n, j_n)\}$.

Example 57 i) Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{R}^{2\times 2}$. Then there are 2 patterns:

$$P_1 = \{(1,1),(2,2)\}$$
$$P_2 = \{(1,2),(2,1)\}$$

ii) For the case that $A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \in \mathbb{R}^3$, the patterns are

$$\begin{aligned} P_1 &= \{(1,1),(2,2),(3,3)\} & P_4 &= \{(1,2),(2,3),(3,1)\},\\ P_2 &= \{(1,1),(2,3),(3,2)\} & P_5 &= \{(1,3),(2,1),(3,2)\},\\ P_3 &= \{(1,2),(2,1),(3,3)\} & P_6 &= \{(1,3),(2,2),(3,1)\}.\end{aligned}$$

Remark. In general, there are $n!$ patterns (where $n! = 1 \cdot 2 \cdot \ldots \cdot n$) for any matrix $A \in \mathbb{R}^{n\times n}$.

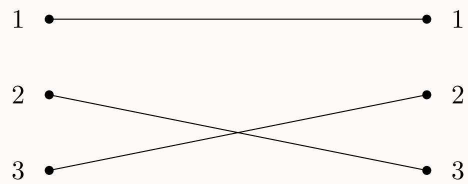
Definition 17.2 i) A bijective map $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ is called a **permutation** of $\{1, \ldots, n\}$.
ii) S_n denotes the set of all permutations of $\{1, \ldots, n\}$.

Patterns in an $n \times n$-matrix corresponds exactly to the permutations of $\{1, \ldots, n\}$. For each $\sigma \in S_n$ we have the pattern

$$P = \left\{(1,\sigma(1)),(2,\sigma(2)),\ldots,(n,\sigma(n))\right\},$$

where (i,j) denotes the choice of the i-th row and the j-th column.

Example 58 Consider the case $n = 3$ and a permutation $\sigma : \{1,2,3\} \to \{1,2,3\}$ defined by $\sigma(1) = 1,\ \sigma(2) = 3,\ \sigma(3) = 2$. The permutation $\sigma \in S_3$ can be described by the following diagram.



In addition, the patterns P corresponding to the permutation σ is $P = \{(1,1),\ (2,3),\ (3,2)\}$.

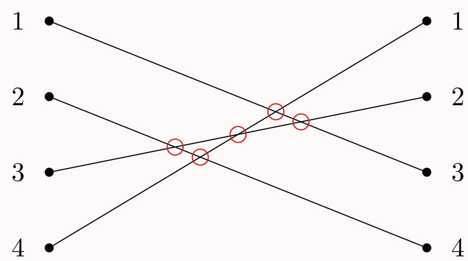
Definition 17.3 i) The **number of inversion** of a permutation $\sigma \in S_n$, denoted by $\operatorname{inv}(\sigma)$, is the number of pairs $(i, \sigma(i)), (j, \sigma(j))$ with $i < j$ and $\sigma(i) > \sigma(j)$.

ii) The sign of a permutation $\sigma \in S_n$ is defined by

$$\operatorname{sign}(\sigma) = (-1)^{\operatorname{inv}(\sigma)}\,.$$

Remark. Given a permutation $\sigma \in S_n$, we can calculate the inversion $\operatorname{inv}(\sigma)$ by counting the number of intersections between any two lines in the diagram describing σ.

Example 59 Consider a permutation σ described by the following diagram.



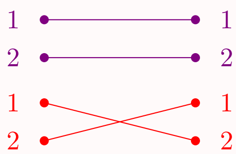
There are 5 intersections marked by red circles in the diagram. Then, the inversion of the permutation is $\operatorname{inv}(\sigma) = 5$.

Definition 17.4 The **determinant** of a $n \times n$-matrix $A = (a_{i,j}) \in \mathbb{R}^{n \times n}$ is defined by

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)}\,.$$

Example 60 i) In the case that $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, we have $S_2 = \{\sigma_1, \sigma_2\}$, where

$\sigma_1 :$ $\sigma_2 :$



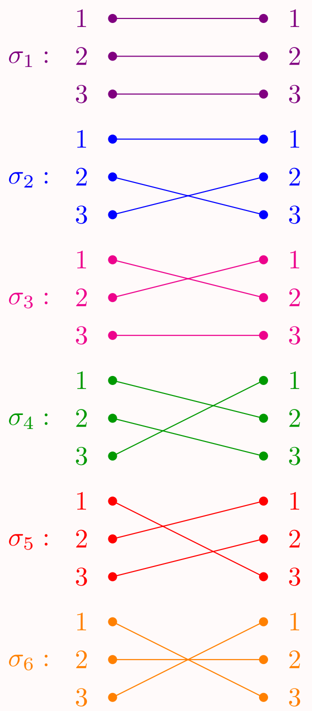
$$\operatorname{inv}(\sigma_1) = 0, \quad \operatorname{sign}(\sigma_1) = (-1)^{\operatorname{inv}(\sigma_1)} = (-1)^0 = 1,$$

$$\operatorname{inv}(\sigma_2) = 1, \quad \operatorname{sign}(\sigma_2) = (-1)^{\operatorname{inv}(\sigma_2)} = (-1)^1 = -1.$$

Then, we calculate the determinant of A as follows

$$\begin{aligned}\det(A) &= \operatorname{sign}(\sigma_1) \cdot a_{1,\sigma_1(1)} \cdot a_{2,\sigma_1(2)} + \operatorname{sign}(\sigma_2) \cdot a_{1,\sigma_2(1)} \cdot a_{2,\sigma_2(2)} \\ &= 1 \cdot a \cdot d + (-1) \cdot b \cdot c \\ &= ad - bc\end{aligned}$$

ii) In the case that $A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$, we have $S_3 = \{\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6\}$, where



$\operatorname{inv}(\sigma_1) = 0, \quad \operatorname{sign}(\sigma_1) = (-1)^{\operatorname{inv}(\sigma_1)} = (-1)^0 = 1,$

$\operatorname{inv}(\sigma_2) = 1, \quad \operatorname{sign}(\sigma_2) = (-1)^{\operatorname{inv}(\sigma_2)} = (-1)^1 = -1,$

$\operatorname{inv}(\sigma_3) = 1, \quad \operatorname{sign}(\sigma_3) = (-1)^{\operatorname{inv}(\sigma_3)} = (-1)^1 = -1,$

$\operatorname{inv}(\sigma_4) = 2, \quad \operatorname{sign}(\sigma_4) = (-1)^{\operatorname{inv}(\sigma_4)} = (-1)^2 = 1,$

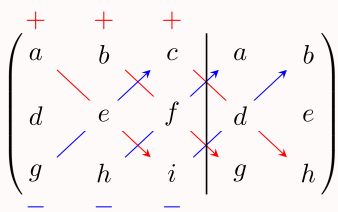
$\operatorname{inv}(\sigma_5) = 2, \quad \operatorname{sign}(\sigma_5) = (-1)^{\operatorname{inv}(\sigma_5)} = (-1)^2 = 1,$

$\operatorname{inv}(\sigma_6) = 3, \quad \operatorname{sign}(\sigma_6) = (-1)^{\operatorname{inv}(\sigma_6)} = (-1)^3 = -1.$

These permutations correspond to the patterns which we encountered in Example 57. As a result, we get the determinant of A

$$\begin{aligned}
\det(A) = \; & \operatorname{sign}(\sigma_1) \cdot a_{1,\sigma_1(1)} \cdot a_{2,\sigma_1(2)} \cdot a_{3,\sigma_1(3)} \\
& + \operatorname{sign}(\sigma_2) \cdot a_{1,\sigma_2(1)} \cdot a_{2,\sigma_2(2)} \cdot a_{3,\sigma_2(3)} \\
& + \operatorname{sign}(\sigma_3) \cdot a_{1,\sigma_3(1)} \cdot a_{2,\sigma_3(2)} \cdot a_{3,\sigma_3(3)} \\
& + \operatorname{sign}(\sigma_4) \cdot a_{1,\sigma_4(1)} \cdot a_{2,\sigma_4(2)} \cdot a_{3,\sigma_4(3)} \\
& + \operatorname{sign}(\sigma_5) \cdot a_{1,\sigma_5(1)} \cdot a_{2,\sigma_5(2)} \cdot a_{3,\sigma_5(3)} \\
& + \operatorname{sign}(\sigma_6) \cdot a_{1,\sigma_6(1)} \cdot a_{2,\sigma_6(2)} \cdot a_{3,\sigma_6(3)} \\
= \; & 1 \cdot a \cdot e \cdot i + (-1) \cdot a \cdot f \cdot h + (-1) \cdot b \cdot d \cdot i + 1 \cdot b \cdot f \cdot g + 1 \cdot c \cdot d \cdot h + (-1) \cdot c \cdot e \cdot g \\
= \; & aei + bfg + cdh - gec - hfa - idb.
\end{aligned}$$

The end result can be also obtained by the Sarrus rule, which is illustrated in the following diagram where the copies of the two first columns are added to the right of the matrix.



The diagram shows that the determinant can be calculated as follows:

$$\det(A) = +aei + bfg + cdh - gec - hfa - idb.$$

17.3 Properties of determinants

Lemma 17.5 *For all $\sigma \in S_n$ we have* $\mathrm{inv}(\sigma) = \mathrm{inv}(\sigma^{-1})$.

Proof. Consider any pairs $(i, \sigma(i))$, $(j, \sigma(j))$ with $i < j$ and $\sigma(i) > \sigma(j)$. Let $k = \sigma(i)$ and $l = \sigma(j)$. Then, $i = \sigma^{-1}(k)$ and $j = \sigma^{-1}(l)$. As a result, we get pairs $(k, \sigma^{-1}(k))$, $(l, \sigma^{-1}(l))$ with $l < k$ and $\sigma^{-1}(l) > \sigma^{-1}(k)$.
Also, consider any pairs $(i, \sigma^{-1}(i))$, $(j, \sigma^{-1}(j))$ with $i < j$ and $\sigma^{-1}(i) > \sigma^{-1}(j)$. Similarly, we can define $k = \sigma^{-1}(i)$, $l = \sigma^{-1}(j)$, and get pairs $(k, \sigma(k))$, $(l, \sigma(l))$ with $l < k$ and $\sigma(l) > \sigma(k)$. Consequently, we get $\mathrm{inv}(\sigma) = \mathrm{inv}(\sigma^{-1})$. □

Proposition 17.6 *For any $A \in \mathbb{R}^{n\times n}$ we have* $\det(A) = \det\big(A^T\big)$.

Proof. Consider $A = (a_{i,j})$ and $A^T = (b_{i,j})$, where $a_{i,j} = b_{j,i}$ for $i, j \in \mathbb{N}$ such that $1 \le i, j \le n$. We have

$$\det(A) = \sum_{\sigma\in S_n} \mathrm{sign}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)},$$
$$\det\big(A^T\big) = \sum_{\tau\in S_n} \mathrm{sign}(\tau) \prod_{j=1}^{n} b_{j,\tau(j)} = \sum_{\tau\in S_n} \mathrm{sign}(\tau) \prod_{j=1}^{n} a_{\tau(j),j}.$$

Now, since τ is a bijective map, we can consider $k = \tau(j)$ and $j = \tau^{-1}(k)$. Therefore, we have

$$\det\big(A^T\big) = \sum_{\tau\in S_n} \mathrm{sign}(\tau) \prod_{k=1}^{n} a_{k,\tau^{-1}(k)}.$$

Taking $\tau^{-1} = \sigma$ and $\tau = \sigma^{-1}$, and using the fact that $\mathrm{sign}(\sigma) = \mathrm{sign}(\sigma^{-1})$ (a direct result from Lemma 17.5), we have

$$\begin{aligned}\det\big(A^T\big) &= \sum_{\sigma\in S_n} \mathrm{sign}(\sigma^{-1}) \prod_{k=1}^{n} a_{k,\sigma(k)} \\ &= \sum_{\sigma\in S_n} \mathrm{sign}(\sigma) \prod_{k=1}^{n} a_{k,\sigma(k)} = \det(A).\end{aligned}$$ □

For $A = (a_{i,j}) \in \mathbb{R}^{n\times n}$ define for a vector $x \in \mathbb{R}^n$ and $1 \le l \le n$ the matrix $A(l; x)$ as the matrix where the l-th row of A gets replaced by x, i.e.

$$A(l;x) = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ \vdots & & & \vdots \\ a_{l-1,1} & a_{l-1,2} & \cdots & a_{l-1,n} \\ x_1 & x_2 & \cdots & x_n \\ a_{l+1,1} & a_{l+1,2} & \cdots & a_{l+1,n} \\ \vdots & & & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{pmatrix} \in \mathbb{R}^{n\times n}, \qquad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n .$$

Proposition 17.7 *For any $A = (a_{i,j}) \in \mathbb{R}^{n\times n}$ and $1 \le l \le n$ the map*

$$\begin{aligned} F_{A,l} : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ x &\longmapsto \det(A(l;x)) \end{aligned}$$

is a linear map, i.e. the determinant is linear in each row,

Example 61 For $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$ and $l = 2$, the map $F_{A,2}$ is given by

$$\begin{aligned} F_{A,2}: \quad & \mathbb{R}^n \longrightarrow \mathbb{R}, \\ & \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \longmapsto \det \begin{pmatrix} 1 & 2 & 3 \\ x_1 & x_2 & x_3 \\ 7 & 8 & 9 \end{pmatrix}. \end{aligned}$$

By using the formula for 3×3 determinants, we get for any $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3$,

$$\begin{aligned} F_{A,2} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \det \begin{pmatrix} 1 & 2 & 3 \\ x_1 & x_2 & x_3 \\ 7 & 8 & 9 \end{pmatrix} &= 9x_2 + 14x_3 + 24x_1 - 21x_2 - 8x_3 - 18x_1 \\ &= 6x_1 - 12x_2 + 6x_3 \\ &= \begin{pmatrix} 6 & -12 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}. \end{aligned}$$

In this case, it is easy to see that this map is linear. In addition, the matrix of this map is $[F_{A,2}] = \begin{pmatrix} 6 & -12 & 6 \end{pmatrix}$.

Proposition 17.8 *For $A \in \mathbb{R}^{n \times n}$ let $B \in \mathbb{R}^{n \times n}$ be a matrix obtained from the matrix A by swapping two rows. Then we have*

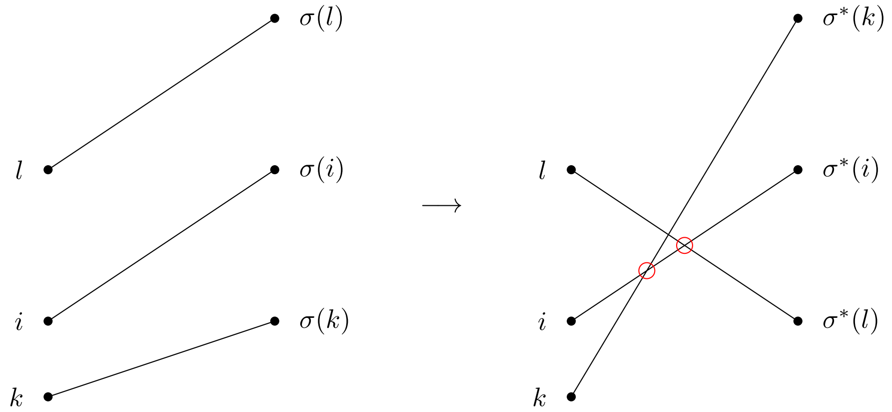
$$\det(A) = -\det(B).$$

Proof. Suppose $B = (b_{i,j})$ is obtained by swapping the l-th and k-th rows of $A = (a_{i,j})$ for $1 \leq l, k \leq n$ and $l \neq k$. Then, we have

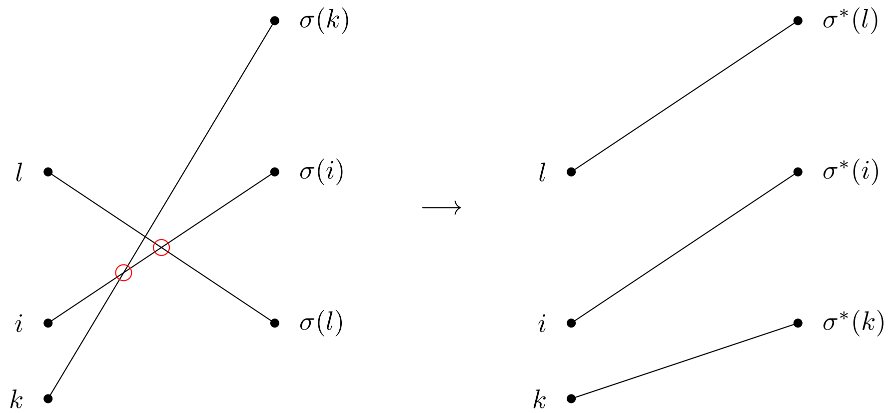
$$\begin{aligned} \det(B) = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) \prod_{i=1}^{n} b_{i,\sigma(i)} &= \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) b_{l,\sigma(l)} b_{k,\sigma(k)} \prod_{i=1, i \neq l, i \neq k}^{n} b_{i,\sigma(i)} \\ &= \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{k,\sigma^*(k)} a_{l,\sigma^*(l)} \prod_{i=1, i \neq l, i \neq k}^{n} a_{i,\sigma^*(i)} \\ &= \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) \prod_{i=1}^{n} a_{i,\sigma^*(i)}, \end{aligned} \qquad (*)$$

where we define σ^* exactly like σ except that $\sigma^*(l) = \sigma(k)$ and $\sigma^*(k) = \sigma(l)$. Now, we will prove that $\operatorname{sign}(\sigma^*) = -\operatorname{sign}(\sigma)$. Without loss of generality, assume that $l < k$. First, if $\sigma(l) < \sigma(k)$, then $\sigma^*(l) > \sigma^*(k)$, and vice versa. Hence, the net change in the number of inversion for this pair is 1 or -1. Next, consider any i such that $l < i < k$. If $\sigma(l) < \sigma(i)$ and $\sigma(i) < \sigma(k)$, then $\sigma^*(k) < \sigma^*(i)$ and

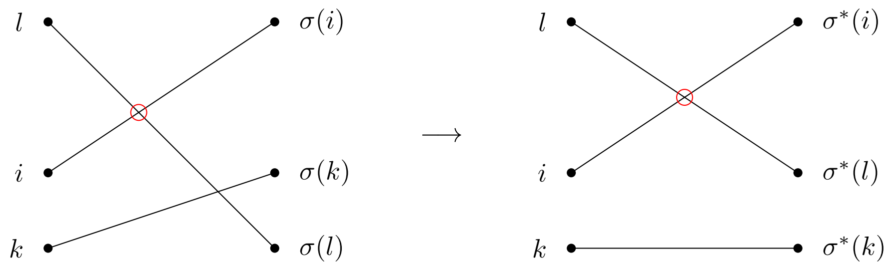
$\sigma(l)^* < \sigma^*(i)$.



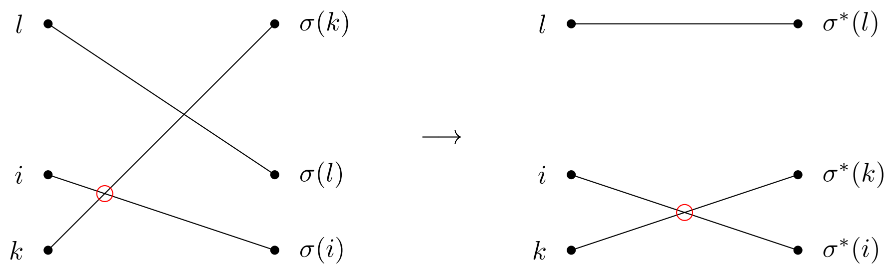
Otherwise, if $\sigma(l) > \sigma(i)$ and $\sigma(i) > \sigma(k)$, then $\sigma^*(k) > \sigma^*(i)$ and $\sigma(l)^* > \sigma^*(i)$.



If $\sigma(l) > \sigma(i)$ and $\sigma(i) < \sigma(k)$, then $\sigma^*(k) > \sigma^*(i)$ and $\sigma^*(i) < \sigma^*(l)$.



If $\sigma(k) < \sigma(i)$ and $\sigma(l) < \sigma(i)$, then $\sigma^*(l) < \sigma^*(i)$ and $\sigma^*(k) < \sigma^*(i)$.



Overall, the net change in the number of inversion for each i between l and k is 2, -2, or 0. For each i such that $i < l < k$ or $l < k < i$, there is no net change in the number of inversion. As a result, the difference between $\mathrm{inv}(\sigma)$ and $\mathrm{inv}(\sigma^*)$ is an odd number, which implies that $\mathrm{sign}(\sigma) = -\mathrm{sign}(\sigma^*)$. Since we take the sum in Equation ($*$) over all permutations in S_n, the transition from σ to σ^* also lets us take the sum over all permutations even though we define σ^* in term of σ. Therefore,

$$\begin{aligned}\det(B) &= \sum_{\sigma\in S_n} \operatorname{sign}(\sigma)\prod_{i=1}^{n} a_{i,\sigma^*(i)}\\ &= -\sum_{\sigma\in S_n} \operatorname{sign}(\sigma^*)\prod_{i=1}^{n} a_{i,\sigma^*(i)}\\ &= -\sum_{\sigma^*\in S_n} \operatorname{sign}(\sigma^*)\prod_{i=1}^{n} a_{i,\sigma^*(i)} = -\det(A)\end{aligned}$$ □

Corollary 17.9 *If a matrix $A \in \mathbb{R}^{n\times n}$ contains two equal rows or columns, then* $\det(A) = 0$.

Recall from Linear Algebra I that there are three types of **row operations** for a matrix $A \in \mathbb{R}^{n\times n}$. $(1 \le i, j \le n\,, i \ne j\,, \lambda \in \mathbb{R})$.
(R1) Add λ-times the j-th row to the i-th row.
(R2) For $\lambda \ne 0$ multiply the i-th row with λ.
(R3) Swap the j-th row with the i-th row.
Two matrices $A, B \in \mathbb{R}^{n\times n}$ are called **row equivalent**, if one can obtain B from A by using the row operations (R1), (R2) and (R3). Notation: $A \sim B$.

Proposition 17.10 *Let $A, B \in \mathbb{R}^{n\times n}$.*
i) If B is obtained from A by using (R1), then $\det(B) = \det(A)$.
ii) If B is obtained from A by using (R2), then $\det(B) = \lambda\det(A)$.
iii) If B is obtained from A by using (R3), then $\det(B) = -\det(A)$.

Proof. Each statement is proved as follows:
i) Consider the row operation where we add λ times j-th row to the i-th row for $\lambda \ne 0$ and $1 \le i, j \le n$, i.e.

$$A = \begin{pmatrix} r_1^T \\ \vdots \\ r_j^T \\ \vdots \\ r_i^T \\ \vdots \\ r_n^T \end{pmatrix} \overset{R1}{\sim} \begin{pmatrix} r_1^T \\ \vdots \\ r_j^T \\ \vdots \\ r_i^T + \lambda r_j^T \\ \vdots \\ r_n^T \end{pmatrix} = B$$

Since determinant is linear in the rows (Proposition 17.7), we get

$$\det(B) = \det\begin{pmatrix} r_1^T \\ \vdots \\ r_j^T \\ \vdots \\ r_i^T \\ \vdots \\ r_n^T \end{pmatrix} + \lambda\cdot\det\begin{pmatrix} r_1^T \\ \vdots \\ r_j^T \\ \vdots \\ r_j^T \\ \vdots \\ r_n^T \end{pmatrix} = \det(A) + 0 = \det(A).$$

In the second equality, the second term is zero by Corollary 17.9 since it has two equal rows.
ii) This statement follows from Proposition 17.7 since the determinant is linear in the rows.
iii) This statement is precisely Proposition 17.8.

$\square$

Example 62 Let $\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 4 \\ 1 & 4 & 6 \end{pmatrix}$. We perform row operations to find the rref of A:

$$A = \begin{array}{c} -1 \\ \\ \end{array} \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 4 \\ 1 & 4 & 6 \end{pmatrix} \overset{\text{R1}}{\sim} \begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & 1 \\ 0 & 2 & 3 \end{pmatrix} \overset{\text{R3}}{\sim} \begin{array}{c} \\ \\ -3 \end{array} \begin{pmatrix} 1 & 2 & 3 \\ 0 & 2 & 3 \\ 0 & 0 & 1 \end{pmatrix} \overset{\text{R1}}{\sim} \begin{array}{c} \\ \frac{1}{2} \\ \end{array} \begin{pmatrix} 1 & 2 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\overset{\text{R2}}{\sim} \begin{array}{c} \\ -2 \\ \end{array} \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \overset{\text{R1}}{\sim} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = I_3 = \text{rref}(A)$$

Hence, the rref of A is the identity matrix with the determinant equal to 1. By Proposition 17.10, we can calculate the determinant of A as follows:

$$1 = \det(I_3) = 1 \cdot \frac{1}{2} \cdot 1 \cdot -1 \cdot 1 \cdot \det(A) \quad \Rightarrow \quad \det(A) = -2$$

We see that if $A \sim B$, then $\det(A) = 0 \Leftrightarrow \det(B) = 0$ and $\det(A) \neq 0 \Leftrightarrow \det(B) \neq 0$. Therefore, in general, $\det(A) = c \cdot \det(B)$ for some $c \neq 0$. In particular, $\det(A) = 0 \Leftrightarrow \det(\text{rref}(A)) = 0$. From Linear Algebra I, we know that a matrix $A \in \mathbb{R}^{n\times n}$ is invertible if and only if $\text{rref}(A) = I_n$. Now, we can show another necessary and sufficient condition for the invertibility of a matrix.

Theorem 17.11 *A matrix $A \in \mathbb{R}^{n\times n}$ is invertible if and only if $\det(A) \neq 0$.*

Proof. Assume that A is invertible. As mentioned above, $A \sim \text{rref}(A) = I_n$ and $\det(I_n) = 1$. As a result, for some constant $c \neq 0$ depending on the row operations to get from A to $\text{rref}(A)$, we have $\det(A) = c \det(\text{rref}(A)) = c \neq 0$.
Conversely, assuming that A is not invertible, we will show that $\det(A) = 0$. If A is not invertible, then we have $\text{rk}(A) < n$, i.e. $\text{rref}(A) = \begin{pmatrix} & * & \\ 0 & \cdots & 0 \end{pmatrix}$. In addition, determinant is linear in rows (Proposition 17.7) so the determinant of any matrix containing a zero row is zero since any linear map transform a zero vector to another zero vector. Hence,

$$\det(\text{rref}(A)) = 0 \quad \Rightarrow \quad \det(A) = 0.$$

$\square$

Theorem 17.12 *i) For all $A, B \in \mathbb{R}^{n\times n}$ we have $\det(AB) = \det(A)\det(B)$.*
ii) If A is invertible, then $\det\left(A^{-1}\right) = \frac{1}{\det(A)}$.

Proof. The two statements are proven as follows:
i) This statement will be proven later when we develop enough tools.
ii) Since $AA^{-1} = I_n$ and $\det(I_n) = 1$, we get by using i),

$$1 = \det\left(AA^{-1}\right) = \det(A)\det\left(A^{-1}\right) \quad \Rightarrow \quad \det\left(A^{-1}\right) = \frac{1}{\det(A)}.$$

$\square$

Corollary 17.13 *Let V be a finitely generated vector space, $F : V \to V$ a linear map and B_1, B_2 two bases of V. Then*

$$\det\left([F]_{B_1}\right) = \det\left([F]_{B_2}\right),$$

where $[F]_B = [F]_B^B$ denotes the matrix of F with respect to the basis B (Definition 16.1).

Proof. Since $[F]_{B_1} = (S_{B_1}^{B_2})^{-1}[F]_{B_2}S_{B_1}^{B_2}$, we get by using Theorem 17.12,

$$\begin{aligned}\det([F]_{B_1}) &= \det\left((S_{B_1}^{B_2})^{-1}\right)\det([F]_{B_2})\det\left(S_{B_1}^{B_2}\right)\\ &= \frac{1}{\det\left(S_{B_1}^{B_2}\right)}\det([F]_{B_2})\det\left(S_{B_1}^{B_2}\right)\\ &= \det([F]_{B_2})\,.\end{aligned}$$

□

Thanks to this corollary, the following definition is well-defined.

Definition 17.14 Let V be a finitely generated vector space, $F : V \to V$ a linear map and B any basis of V. We define the **determinant of the linear map** F by

$$\det(F) = \det([F]_B)\ .$$

Now we describe row operations by matrices, which helps us prove Theorem 17.12 i).

Definition 17.15 For $\lambda \in \mathbb{R}$ with $\lambda \neq 0$ and $1 \leq i, j \leq n$ we define the **elementary matrices** $R_i^{\lambda,j}, R_i^\lambda, R_{i,j} \in \mathbb{R}^{n\times n}$ by

$$R_i^{\lambda,j} = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & \ddots & & \\ & & \lambda & & 1 & \\ & & & & & \ddots & \\ & & & & & & 1\end{pmatrix},\quad R_i^\lambda = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & \lambda & \\ & & & & 1 \\ & & & & & \ddots \\ & & & & & & 1\end{pmatrix},\quad R_{i,j} = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 0 & & 1 \\ & & & \ddots & \\ & & 1 & & 0 \\ & & & & & \ddots \\ & & & & & & 1\end{pmatrix}.$$

Here the λ in $R_i^{\lambda,j}$ is in the i-th row and j-th column, in R_i^λ it is in the i-th row, and in $R_{i,j}$ the 0 are on the diagonal in the i-th row and j-th column.

Example 63 For any $\lambda \in \mathbb{R}\setminus\{0\}$,

$$R_3^{\lambda,1} = \begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ \lambda & 0 & 1\end{pmatrix},\quad R_2^\lambda = \begin{pmatrix}1 & 0 & 0\\ 0 & \lambda & 0\\ 0 & 0 & 1\end{pmatrix},\quad R_{1,2} = \begin{pmatrix}0 & 1 & 0\\ 1 & 0 & 0\\ 0 & 0 & 1\end{pmatrix}.$$

The elementary matrices satisfy the following properties
1) $\det\left(R_i^{\lambda,j}\right) = 1,\ \det\left(R_i^\lambda\right) = \lambda,\ \det(R_{i,j}) = -1,$
2) $(R_i^{\lambda,j})^{-1} = R_i^{-\lambda,j},\ (R_i^\lambda)^{-1} = R_i^{1/\lambda},\ (R_{i,j})^{-1} = R_{i,j}.$
In particular, inverses of elementary matrices are again elementary matrices.

Proposition 17.16 *Let $A \in \mathbb{R}^{n\times n}$.*
i) $R_i^{\lambda,j}A$ is the matrix obtained from A by row operation (R1). *(Add λ-times the j-th row to the i-th row)*
ii) $R_i^\lambda A$ is the matrix obtained from A by row operation (R2). *(Multiply the i-th row with λ)*
iii) $R_{i,j}A$ is the matrix obtained from A by row operation (R3). *(Swap the j-th row with the i-th row)*

Proof. The statements are proven as follows:

i) We have

$$R_i^{\lambda,j} = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & \ddots & \\ & & \lambda & & 1 & \\ & & & & & \ddots \\ & & & & & & 1 \end{pmatrix} = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 & \\ & & & & & \ddots \\ & & & & & & 1 \end{pmatrix} + \lambda \begin{pmatrix} 0 & & & & \\ & \ddots & & & \\ & & 0 & & \\ & & & \ddots & \\ & & 1 & & 0 & \\ & & & & & \ddots \\ & & & & & & 0 \end{pmatrix} = I_n + \lambda E_{i,j}$$

where $E_{i,j} = (e_{kl})_{1\le k,l\le n}$ with $e_{kl} = \begin{cases} 1, & k=i,\ l=j \\ 0, & \text{else} \end{cases}$.

Therefore,

$$R_i^{\lambda,j} A = (I_n + \lambda E_{i,j})A = A + \lambda E_{i,j}A.$$

Let $A = (a_{kl})_{k,l}$ and $E_{i,j}A = (b_{kl})_{k,l}$. Then, we have for any $1 \le k,l \le n$,

$$b_{kl} = \sum_{m=1}^{n} e_{km}a_{ml} = \begin{cases} \sum_{m=1}^{n} e_{im}a_{ml}, & k=i \\ 0 \quad\quad\quad\quad\quad , & k \neq i \end{cases} = \begin{cases} e_{ij}a_{jl}, & k=i \\ 0 \quad\quad , & k\neq i \end{cases} = \begin{cases} a_{jl}, & k=i \\ 0 \ \ , & k \neq i \end{cases}$$

The final result is a matrix which just contains the j-th row of A in its i-th row. As a result, $A + \lambda E_{i,j}A$ is the matrix obtained by adding λ-times the j-th row to the i-th row of A (row operation R1)

ii) $R_i^{\lambda} = (c_{kl})_{k,l}$ is a diagonal matrix with element $c_{kk} = \begin{cases} 1, & k \neq i \\ \lambda, & k = i \end{cases}$. Let $R_i^{\lambda}A = (d_{kl})_{k,l}$. By direction computation, we have for any $1 \le k,l \le n$,

$$d_{kl} = \sum_{m=1}^{n} c_{km}a_{ml} = c_{kk}a_{kl} = \begin{cases} \lambda a_{il}, & k=i \\ a_{kl}\ , & k \neq i \end{cases}.$$

Hence, $R_i^{\lambda}A$ is the matrix obtained by multipling the i-th row of A by $\lambda \neq 0$.

iii) Let $I_n = (\delta_{kl})_{k,l}$ where I_n is the idenity matrix. Note that $\delta_{kl} = \begin{cases} 1, & k = l \\ 0, & k \neq l \end{cases}$ is also called the Kronecker delta. In addition, let $R_{i,j} = (f_{kl})_{k,l}$ and $R_{i,j}A = (h_{kl})_{k,l}$. Then, we have $f_{kl} = \delta_{kl}$ for $1 \le k,l \le n$ except that $f_{ii} = f_{jj} = 0$ and $f_{ij} = f_{ji} = 1$. For $k \neq i$ and $k \neq j$, we have

$$h_{kl} = \sum_{m=1}^{n} f_{km}a_{ml} = f_{kk}a_{kl} = a_{kl}.$$

For $k = i$, we have

$$h_{il} = \sum_{m=1}^{n} f_{im}a_{ml} = f_{ij}a_{jl} = a_{jl}.$$

For $k = j$, we have

$$h_{jl} = \sum_{m=1}^{n} f_{jm}a_{ml} = f_{ji}a_{il} = a_{il}.$$

Hence, $R_{i,j}A$ is the matrix obtained by exchanging i-th row and j-th row of A. □

Corollary 17.17 *Let* $A \in \mathbb{R}^{n\times n}$.

i) The matrix A is invertible if and only if it is a product of elementary matrices.

ii) If C is an elementary matrix then $\det(CA) = \det(C)\det(A)$.

Proof. The statements are proven as follows:

i) Assume that A is invertible. Then, we can peform a combination of row operations R1, R2, and R3 to obtain $\mathrm{rref}(A) = I_n$. By Proposition 17.16, these operations correspond to multiplication of elementary matrices, i.e.

$$A \sim \ldots \sim I_n = \mathrm{rref}(A) \quad \Rightarrow \quad A = C_1 \cdot \ldots \cdot C_m \cdot I_n = C_1 \cdot \ldots \cdot C_m,$$

where $C_1, \ldots, C_m$ are elementary matrices.
Conversely, assume that $A = C_1 \cdot \ldots \cdot C_m$ where $C_1, \ldots, C_m$ are elementary matrices. Since elementary matrices are always invertible, we have

$$A^{-1} = (C_1 \cdot \ldots \cdot C_m)^{-1} = C_m^{-1} \cdot \ldots \cdot C_1^{-1},$$

where $C_1^{-1}, \ldots, C_m^{-1}$ are also elementary matrices.

ii) This is just a reformulation of Proposition 17.16 together with Proposition 17.10 and the properties

$$\det\left(R_i^{\lambda,j}\right) = 1, \quad \det\left(R_i^{\lambda}\right) = \lambda, \quad \det(R_{i,j}) = -1$$

□

Now we can prove Theorem 17.12 i), which states that for any $A, B \in \mathbb{R}^{n \times n}$,

$$\det(AB) = \det(A)\det(B).$$

Proof of Theorem 17.12 i). We divide this statement into two cases.

Case 1: Assume that A is invertible, i.e. $A =_1 \cdot \ldots \cdot C_m$ for elementary matrices $C_1, \ldots, C_m$ (Corollary 17.17 i)). We want to show that $\det(AB) = \det(A)\det(B)$ by induction on m.

1) Base step m=1: In this case, $A = C_1$. Hence, we have by Corollary 17.17 ii),

$$\det(AB) = \det(C_1 B) = \det(C_1)\det(B) = \det(A)\det(B).$$

2) Induction step: Assume that $\det(C_2 \cdot \ldots \cdot C_m \cdot B) = \det(C_2 \cdot \ldots \cdot C_m)\det(B)$, which is the induction hypothesis. Then, we have for $A = C_1 \cdot \ldots \cdot C_m$,

$$\begin{aligned}
\det(AB) &= \det(C_1 \cdot C_2 \cdot \ldots \cdot C_m \cdot B) \\
&= \det(C_1)\det(C_2 \cdot \ldots \cdot C_m \cdot B) \\
&= \det(C_1)\det(C_2 \cdot \ldots \cdot C_m)\det(B) \\
&= \det(C_1 \cdot C_2 \cdot \ldots \cdot C_m)\det(B) \\
&= \det(A)\det(B).
\end{aligned}$$

We apply Corollary 17.17 ii) in the second and the fourth equality while the induction hypothesis is applied in the third equality.

Case 2: Assume that A is not invertible, which implies that $\det(A) = 0$ (Theorem 17.11). Recall from Linear Algebra I that a matrix $A \in \mathbb{R}^{n \times n}$ is not invertible iff $\mathrm{im}(A) \neq \mathbb{R}^n$. In addition, AB is the matrix of a linear map $F \circ G$ with $F : x \mapsto Ax$ and $G : x \mapsto Bx$. The image of G is generally a subset of the domain of F (which is $\mathbb{R}^n$). Consequently, we have $\mathrm{im}(AB) \subset \mathrm{im}(A) \neq \mathbb{R}^n$, which implies that AB is not invertible. As a result, we have

$$\det(AB) = 0 = 0 \cdot \det(B) = \det(A)\det(B).$$

□

There is a way to calculate determinants of matrices with a lot of zero entries, which is called **Laplace expansion**. For a matrix $A \in \mathbb{R}^{n \times n}$ and $1 \leq i, j \leq n$ we denote by $A_{i,j} \in \mathbb{R}^{(n-1) \times (n-1)}$ the matrix which is obtained from A by removing the i-th row and the j-th column.

Theorem 17.18 (Laplace expansion) *For a matrix* $A = (a_{i,j}) \in \mathbb{R}^{n\times n}$ *and* $1 \leq i, j \leq n$ *we have*

$$\begin{aligned}\det(A) &= \sum_{l=1}^{n} (-1)^{i+l} a_{i,l} \det(A_{i,l}) \\ &= \sum_{l=1}^{n} (-1)^{j+l} a_{l,j} \det(A_{l,j})\,.\end{aligned}$$

Example 64 Consider the matrix $A = \begin{pmatrix} 0 & 1 & 4 \\ 8 & 0 & -1 \\ 2 & 1 & 3 \end{pmatrix}$. We calculate its determinant by expansion along the first row ($i = 1$):

$$\begin{aligned}\det(A) = \det\begin{pmatrix} \color{red}{0} & \color{purple}{1} & \color{green}{4} \\ 8 & 0 & -1 \\ 2 & 1 & 3 \end{pmatrix} &= \sum_{l=1}^{3} (-1)^{l+1} a_{1,l} \det(A_{i,l}) \\ &= +\color{red}{0} \cdot \det\begin{pmatrix} \cancel{0} & \cancel{1} & \cancel{4} \\ \cancel{8} & 0 & -1 \\ \cancel{2} & 1 & 3 \end{pmatrix} - \color{purple}{1} \cdot \det\begin{pmatrix} \cancel{0} & \cancel{1} & \cancel{4} \\ 8 & \cancel{0} & -1 \\ 2 & \cancel{1} & 3 \end{pmatrix} + \color{green}{4} \cdot \det\begin{pmatrix} \cancel{0} & \cancel{1} & \cancel{4} \\ 8 & 0 & \cancel{-1} \\ 2 & 1 & \cancel{3} \end{pmatrix} \\ &= 0 \cdot \det\begin{pmatrix} 0 & -1 \\ 1 & 3 \end{pmatrix} - 1 \cdot \det\begin{pmatrix} 8 & -1 \\ 2 & 3 \end{pmatrix} + 4 \cdot \det\begin{pmatrix} 8 & 0 \\ 2 & 1 \end{pmatrix} \\ &= -1 \cdot (8 \cdot 3 + 2) + 4 \cdot 8 \\ &= -26 + 32 = 6.\end{aligned}$$

18

Eigenvalues and eigenvectors

18.1 Eigenvalues and eigenvectors

In this section V always denotes a vector space. In the previous section, we introduced the determinant of a matrix A and more generally, of a linear map $F : V \to V$. The determinant tells us if a linear map is invertible or not. However, there are more important properties of a linear map or matrix. In this section, we will introduce one of them, which are the eigenvalues and eigenvectors of a linear map or matrix. The general goal of this chapter is to understand linear maps $F : V \to V$, where V is finitely generated, by finding subspaces $U \subset V$ such that $F(U) \subset U$. The most important case is when $\dim(U) = 1$, i.e. for any $u \in U$,

$$F(u) = \lambda u \quad \text{for some } \lambda \in \mathbb{R},$$

where λ and u are called an eigenvalue and an eigenvector, respectively. Calculating eigenvalues and eigenvectors has many applications, e.g. computer vision, machine learning, google page rank, quantum mechanics (e.g. solving the Schrodinger equation), spectrosopy (chemistry), etc.

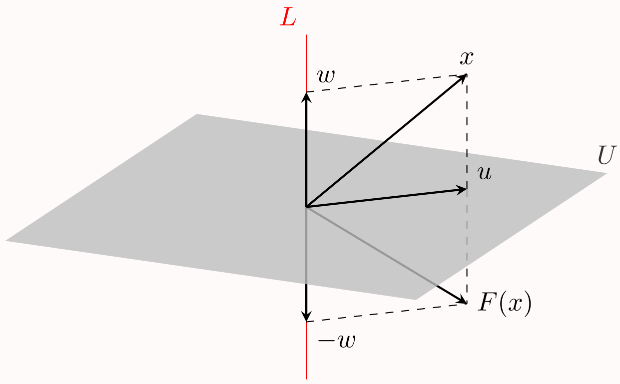
Example 65 Consider the linear map

$$\begin{aligned} F : \mathbb{R}^3 &\longrightarrow \mathbb{R}^3, \\ x &\longmapsto Ax, \end{aligned} \qquad A = \frac{1}{3}\begin{pmatrix} 1 & 2 & -2 \\ 2 & 1 & 2 \\ -2 & 2 & 1 \end{pmatrix}.$$

The subspace U given by $U = \operatorname{span}\left\{\begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}-1\\0\\1\end{pmatrix}\right\} = \operatorname{span}\{u_1, u_2\} \subset \mathbb{R}^3$ satisfies $F(u) = u$ for all $u \in U$. The subspace L given by $L = \operatorname{span}\left\{\begin{pmatrix}1\\-1\\1\end{pmatrix}\right\} = \operatorname{span}\{u_3\}$ satisfies $F(w) = -w$ for all $w \in L$. Moreoever, $L = U^\perp$ since $u_3 \bullet u_1 = u_3 \bullet u_2 = 0$. and any $x \in \mathbb{R}^3$ can be written uniquely as $x = u + w$ for $u \in U,\ w \in L$. Therefore, we have for any $x \in \mathbb{R}^3$,

$$F(x) = F(u + w) = F(u) + F(w) = u - w.$$

As the following figure illustrates, the map F is the reflection through the plane $U \subset \mathbb{R}^3$.



Definition 18.1 Let $F : V \to V$ be a linear map.

i) A $\lambda \in \mathbb{R}$ is called an **eigenvalue** of F, if there exist a vector $v \in V$ with $v \neq 0$, such that

$$F(v) = \lambda v\,. \tag{18.1.1}$$

ii) A vector $v \in V$ with $v \neq 0$, satisfying (18.1.1), is called an **eigenvector** of F with eigenvalue λ.

Notice that $v = 0$ always satisfies (18.1.1) for any $\lambda \in \mathbb{R}$, since F is a linear map. This is one of many reasons why $v = 0$ is not called an eigenvector of F.
In Example 65, $\lambda_1 = 1$ and $\lambda_2 = -1$ are eigenvalues of F since $F(u_1) = 1 \cdot u_1$ and $F(u_3) = -1 \cdot u_3$ ($u_1 \neq \mathbf{0}$ and $u_3 \neq \mathbf{0}$). In fact, any $u \in U$ with $u \neq \mathbf{0}$ is an eigenvector with eigenvalue 1 and any $w \in L$ with $w \neq \mathbf{0}$ is an eigenvector with eigenvalue -1.
In the following, we always assume that V is a finitely generated vector space. Now, we will discuss about the general way to calculate the eigenvalues and eigenvectors of a linear map F. Let λ be an eigenvalue of F and we have by definition

$$\begin{aligned}
\exists v \in V,\ v \neq \mathbf{0} \ \text{ and } \ F(v) = \lambda v \quad &\Leftrightarrow \quad \exists v \in V,\ v \neq \mathbf{0} \ \text{ and } \ (F - \lambda \cdot \mathrm{id}_V)(v) = \mathbf{0} \\
&\Leftrightarrow \quad \exists v \in V,\ v \neq \mathbf{0} \ \text{ and } \ v \in \ker(F - \lambda \cdot \mathrm{id}_V) \\
&\Leftrightarrow \quad \ker(F - \lambda \cdot \mathrm{id}_V) \neq \{\mathbf{0}\}.
\end{aligned}$$

By Theorem 15.5 and Theorem 17.11, this is equivalent to the condition that $(F - \lambda \cdot \mathrm{id}_V)$ is not invertible or $\det(F - \lambda \cdot \mathrm{id}_V) = 0$. If B is a basis of V and $n = \dim(V)$, then

$$\lambda \text{ is an eigenvalue of } F \quad \Leftrightarrow \quad \det(F - \lambda \cdot \mathrm{id}_V) = 0 \quad \Leftrightarrow \quad \det([F]_B - \lambda I_n) = 0.$$

From the above derivation, the eigenvalues of F are given as the solutions of $\det(F - \lambda \cdot \mathrm{id}_V) = 0$ while the eigenvectors for an eigenvalue λ are given as the non-zero elements in $\ker(F - \lambda \cdot \mathrm{id}_V)$.

Example 66 Take again the linear map

$$\begin{aligned} F : \mathbb{R}^3 &\longrightarrow \mathbb{R}^3, \\ x &\longmapsto Ax, \end{aligned} \qquad A = \frac{1}{3}\begin{pmatrix} 1 & 2 & -2 \\ 2 & 1 & 2 \\ -2 & 2 & 1 \end{pmatrix}.$$

For the eigenvalues, we want to find the solutions of $\det(A - \lambda I_3) = 0$. First of all, we need to calculate the determinant

$$\det(A - \lambda I_3) = \det\left(\frac{1}{3}\begin{pmatrix} 1 & 2 & -2 \\ 2 & 1 & 2 \\ -2 & 2 & 1 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right) = \det\begin{pmatrix} \frac{1}{3} - \lambda & \frac{2}{3} & -\frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} - \lambda & \frac{2}{3} \\ -\frac{2}{3} & \frac{2}{3} & \frac{1}{3} - \lambda \end{pmatrix}.$$

We can calculate this determinant by performing row operations:

$$\begin{array}{c} 3 \\ 3 \\ 3 \end{array}\begin{pmatrix} \frac{1}{3}-\lambda & \frac{2}{3} & -\frac{2}{3} \\ \frac{2}{3} & \frac{1}{3}-\lambda & \frac{2}{3} \\ -\frac{2}{3} & \frac{2}{3} & \frac{1}{3}-\lambda \end{pmatrix} \overset{\text{R2}}{\sim} \begin{array}{c} \\ \\ -1 \quad 1 \end{array}\begin{pmatrix} 1-3\lambda & 2 & -2 \\ 2 & 1-3\lambda & 2 \\ -2 & 2 & 1-3\lambda \end{pmatrix} \overset{\text{R1}}{\sim} \begin{pmatrix} 3-3\lambda & 0 & -3+3\lambda \\ 0 & 3-3\lambda & 3-3\lambda \\ -2 & 2 & 1-3\lambda \end{pmatrix}$$

$$\begin{aligned} \Rightarrow \quad \det(A-\lambda I_3) = \det\begin{pmatrix} \frac{1}{3}-\lambda & \frac{2}{3} & -\frac{2}{3} \\ \frac{2}{3} & \frac{1}{3}-\lambda & \frac{2}{3} \\ -\frac{2}{3} & \frac{2}{3} & \frac{1}{3}-\lambda \end{pmatrix} &= \left(\frac{1}{3}\right)^3 \det\begin{pmatrix} 1-3\lambda & 2 & -2 \\ 2 & 1-3\lambda & 2 \\ -2 & 2 & 1-3\lambda \end{pmatrix} \\ &= \left(\frac{1}{3}\right)^3 \cdot 1 \det\begin{pmatrix} 3-3\lambda & 0 & -3+3\lambda \\ 0 & 3-3\lambda & 3-3\lambda \\ -2 & 2 & 1-3\lambda \end{pmatrix} \\ &= \left(\frac{1}{3}\right)^3 (3-3\lambda)^2 \det\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ -2 & 2 & 1-3\lambda \end{pmatrix}, \end{aligned}$$

where we use the fact that the determinant is linear in rows in last equality. Next, we continue to use row operations:

$$\begin{array}{c} 2 \\ -2 \\ \end{array}\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ -2 & 2 & 1-3\lambda \end{pmatrix} \overset{\text{R1}}{\sim} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & -3-3\lambda \end{pmatrix}$$

$$\begin{aligned} \Rightarrow \quad \det(A-\lambda I_3) &= \left(\frac{1}{3}\right)^3 (3-3\lambda)^2 \det\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ -2 & 2 & 1-3\lambda \end{pmatrix} \\ &= \left(\frac{1}{3}\right)^3 (3-3\lambda)^2 \cdot 1 \det\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & -3-3\lambda \end{pmatrix} \\ &= \left(\frac{1}{3}\right)^3 (3-3\lambda)^2(-3-3\lambda) \det\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \left(\frac{1}{3}\right)^3 (3-3\lambda)^2(-3-3\lambda)\cdot 1 \\ &= -(1-\lambda)^2(1+\lambda). \end{aligned}$$

As a result, we get the eigenvalues of F:

$$\det(A-\lambda I_3) = 0 \quad \Leftrightarrow \quad -(1-\lambda)^2(1+\lambda) = 0 \quad \Leftrightarrow \quad \lambda = 1 \text{ or } \lambda = -1.$$

Note that the equivalence indicates that there are no more eigenvalues.

About the eigenvectors, we want to calculate $\ker(A-\lambda I_3)$ for each eigenvalue λ. For the eigenvalue $\lambda_1 = 1$, we have $\ker(A - \lambda_1 I_3) = \ker(A - I_3)$ and

$$A - I_3 = \begin{array}{c} -1 \quad 1 \\ \\ \\ \end{array} \begin{pmatrix} -\frac{2}{3} & \frac{2}{3} & -\frac{2}{3} \\ \frac{2}{3} & -\frac{2}{3} & \frac{2}{3} \\ -\frac{2}{3} & \frac{2}{3} & -\frac{2}{3} \end{pmatrix} \sim \begin{array}{c} -\frac{3}{2} \\ \\ \\ \end{array} \begin{pmatrix} -\frac{2}{3} & \frac{2}{3} & -\frac{2}{3} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & -1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Translating this to linear equations, any element $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \ker(A-\lambda_1 I_3)$ satisfies the equation $x_1 - x_2 + x_3 = 0$ and hence,

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_2 - x_3 \\ x_2 \\ x_3 \end{pmatrix} = x_2 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}.$$

Therefore, $\ker(A - \lambda_1 I_3) = \operatorname{span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right\} = U$ (Example 65).

For the eigenvalue $\lambda_2 = -1$, we have $\ker(A - \lambda_2 I_3) = \ker(A + I_3)$ and

$$A+I_3 = \begin{array}{c} \frac{3}{2} \\ \frac{3}{2} \\ \frac{3}{2} \end{array} \begin{pmatrix} \frac{4}{3} & \frac{2}{3} & -\frac{2}{3} \\ \frac{2}{3} & \frac{4}{3} & \frac{2}{3} \\ -\frac{2}{3} & \frac{2}{3} & \frac{4}{3} \end{pmatrix} \sim \begin{array}{c} \\ -2 \quad 1 \\ \\ \end{array} \begin{pmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix} \sim \begin{array}{c} -\frac{1}{3} \quad 1 \\ \\ \\ \end{array} \begin{pmatrix} 0 & -3 & -3 \\ 1 & 2 & 1 \\ 0 & 3 & 3 \end{pmatrix} \sim \begin{array}{c} -2 \\ \\ \\ \end{array} \begin{pmatrix} 0 & 1 & 1 \\ 1 & 2 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

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

Translating this to linear equations, any element $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \ker(A - \lambda_2 I_3)$ satisfies the equations $\begin{cases} x_1 - x_3 = 0 \\ x_2 + x_3 = 0 \end{cases}$ and hence, $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_3 \\ -x_3 \\ x_3 \end{pmatrix} = x_3 \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$.

As a result, $\ker(A - \lambda_2 I_3) = \operatorname{span}\left\{ \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \right\} = L$ (Example 65).

In general, if $F : V \to V$ is a linear map and B is any basis of V, then

$$\det(F - \lambda \cdot \mathrm{id}_V) = \det([F]_B - \lambda I_n) = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) \prod_{i=1}^{n} b_{i,\sigma(i)},$$

where $[F] = (a_{ij})_{i,j}$ and $b_{ij} = \begin{cases} a_{ij} & , \quad i \neq j \\ a_{ii} - \lambda, & i = j \end{cases}$. The term corresponding to $\sigma = \mathrm{id}_{\{1,\ldots,n\}}$ contains the term λ^n. Hence, $\det(F - \lambda \cdot \mathrm{id}_V)$ is a polynomial of degree n in λ. Therefore, eigenvalues are zeros of this polynomial while the eigenvectors are non-zero elements in $\ker(F - \lambda \cdot \mathrm{id}_V)$ for each eigenvalue λ.

Definition 18.2 Let $F : V \to V$ be a linear map let $\mathrm{id}_V : V \to V$ be the identity map on V.

i) The polynomial $f_F(\lambda) = \det(F - \lambda\,\mathrm{id}_V)$ is called the **characteristic polynomial** of F.

ii) Let $\lambda \in \mathbb{R}$ be an eigenvalue of F. Then the space

$$\begin{aligned} E_\lambda(F) &= \ker(F - \lambda\,\mathrm{id}_V) \\ &= \{v \in V \mid F(v) = \lambda v\} \end{aligned}$$

is called the **eigenspace** of F with respect to the eigenvalue λ.

The eigenspace $E_\lambda(F)$ contains therefore all eigenvectors of F with eigenvalue λ and the zero vector. The degree of $f_F(\lambda)$ is $n = \dim(V)$ and therefore F can have <u>at most</u> n different eigenvalues. However, F can also have no eigenvalues at all. For example,

$$\begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix} \quad \text{has no eigenvalues,}$$

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{has one eigenvalue,}$$

$$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} \quad \text{has two eigenvalues.}$$

Definition 18.3 i) Let $\dim V = n$. A linear map $F : V \to V$ is called **diagonalizable** if there exist a basis B of V, such that

$$[F]_B = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}$$

for some $d_1, \ldots, d_n \in \mathbb{R}$.

ii) A matrix $A \in \mathbb{R}^{n \times n}$ is called **diagonalizable** if there exists an invertible matrix $S \in \mathbb{R}^{n \times n}$ with

$$S^{-1}AS = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}$$

for some $d_1, \ldots, d_n \in \mathbb{R}$.

Lemma 18.4 *Let B be a basis of V and let $F : V \to V$ be a linear map. Then the following two statements are equivalent*

i) The linear map F is diagonalizable.

ii) The matrix $[F]_B$ is diagonalizable.

Proof. Assume F is diagonalizable, which means that there exist a basis B' of V such that

$$[F]_{B'} = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}.$$

Then, we we have

$$S_B^{B'}[F]_B S_{B'}^B = [F]_{B'} = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}.$$

Hence, by setting $S = S_{B'}^B$, we have

$$S^{-1}[F]_B S = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix},$$

which implies that $[F]_B$ is diagonalizable.
Conversely, assume that $[F]_B$ is diagonalizable, which implies that there exists an invertible matrix S such that

$$S^{-1}[F]_B S = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}.$$

We want to find a basis B' such that

$$[F]_{B'} = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}.$$

Let $B = (b_1, \dots, b_n)$. Then we define $B' = (b'_1, \dots, b'_n)$ by $b'_j = c_B(Se_j)$ for $1 \leq j \leq n$, where $c_B : \mathbb{R}^n \to V$ is the coordinate map. In addition, let $\varphi_S : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map defined by $\varphi_S(x) = Sx$. With this, we have $c_{B'} = c_B \circ \varphi_S$ and

$$[F]_{B'} = [c_B^{-1} \circ F \circ c_{B'}] = [\varphi_S^{-1} \circ c_B^{-1} \circ F \circ c_B \circ \varphi_S] = [\varphi_S^{-1}][c_B^{-1} \circ F \circ c_B][\varphi_S] = S^{-1}[F]_B S$$

$$\Rightarrow \quad [F]_{B'} = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}.$$

Therefore, F is diagonalizable. □

Lemma 18.5 *Let $F : V \to V$ be a linear map and $B = (b_1, \dots, b_n)$ be a basis of V, such that all b_i are eigenvectors of F, i.e. $F(b_i) = d_i b_i$ for some $d_i \in \mathbb{R}$ and $i = 1, \dots, n$. Then F is diagonalizable and*

$$[F]_B = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}$$

Conversely, if F is diagonalizable then there exists a basis of eigenvectors.

Proof. If $B = (b_1, \dots, b_n)$ be a basis with $F(b_i) = d_i b_i$ for $i = 1, \dots, n$, then we have

$$[F(b_i)]_B = [d_i b_i]_B = d_i [b_i]_B = d_i e_i.$$

As a result,

$$[F]_B = \begin{pmatrix} | & & | \\ [F(b_1)]_B & \cdots & [F(b_n)]_B \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ d_1 e_1 & \cdots & d_n e_n \\ | & & | \end{pmatrix} = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix},$$

which implies that F is diagonalizable. Conversely, if F is diagonalizable, then there exists a basis $B = (b_1, \ldots, b_n)$ of V such that

$$[F]_B = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}.$$

Then, we have for any $i = 1, \ldots, n$,

$$F(b_i) = c_B \circ c_B^{-1} \circ F \circ c_B(e_i) = c_B([F]_B e_i) = c_B(d_i e_i) = d_i b_i.$$

Therefore, $b_1, \ldots, b_n$ are eigenvectors of F. $\square$

Example 67 1) $A = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$ has eigenvectors $b_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $b_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ with eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 2$. Setting $S = \begin{pmatrix} | & | \\ b_1 & b_2 \\ | & | \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, we have

$$S^{-1}AS \begin{pmatrix} 1 \\ 0 \end{pmatrix} = S^{-1}Ab_1 = \lambda_1 S^{-1} b_1 = \lambda_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$
$$S^{-1}AS \begin{pmatrix} 0 \\ 1 \end{pmatrix} = S^{-1}Ab_2 = \lambda_2 S^{-1} b_2 = \lambda_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$\Rightarrow \quad S^{-1}AS = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.$$

2) $B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ has just one eigenvalue $\lambda = 1$ and all vectors in $\operatorname{span}\left\{\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right\} \setminus \left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix}\right\}$ are the only eigenvectors of this matrix. Consequently, there exists no basis of $\mathbb{R}^2$ consisting of eigenvectors of B and hence, B is not diagonalizable.

Also, we have an interesting example related to Fibonacci sequence as follows.

Example 68 (counting bunnies) Assume that we have a pair of rabbits at the beginning. Each pair mates at the age of one month and then produces a pair of rabbits every one month after that. Assume also that the rabbit do not die. In total, how many pairs of rabbit do we have after n months? The answer is given by the sequence

$$1, 1, 2, 3, 5, 8, 13, 21, \ldots$$

The initial pair cannot produce offspring before reaching the age of one month so there is still one pair after one month. Observe that the difference between two consecutive numbers of pairs is the number of pairs that have not reached the age of one month. Since only pairs being at the age of over one month can produce other pairs, we take the sum of the current number of pairs and the previous number of pairs to get the next number of pairs. Formally, let F_n be the number of pairs in the n-th month. Then we have

$$\begin{cases} F_0 = 0, \ F_1 = 1, \\ F_n = F_{n-1} + F_{n-2}, \quad n \geq 2. \end{cases}$$

Alternatively, letting $x_n = \begin{pmatrix} F_n \\ F_{n+1} \end{pmatrix}$, we have

$$\begin{cases} x_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \\ x_n = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} x_{n-1}, \quad n \geq 1. \end{cases}$$

Let $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$. We want to calculate x_n from x_0 for any $n \geq 0$. We can do that if we can calculate A^n for $n \geq 0$. We can find that the eigenvalues are $d_1 = \frac{1+\sqrt{5}}{2}$ and $d_2 = \frac{1-\sqrt{5}}{2}$, along with eigenspaces

$$E_{d_1}(A) = \operatorname{span}\left\{\begin{pmatrix} 2 \\ 1+\sqrt{5} \end{pmatrix}\right\} = \operatorname{span}\{v_1\}, \qquad E_{d_2}(A) = \operatorname{span}\left\{\begin{pmatrix} 2 \\ 1-\sqrt{5} \end{pmatrix}\right\} = \operatorname{span}\{v_2\}.$$

As a result, A is diagonalizable with

$$S = \begin{pmatrix} | & | \\ v_1 & v_2 \\ | & | \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 1+\sqrt{5} & 1-\sqrt{5} \end{pmatrix}$$

$$S^{-1}AS = \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix} = \begin{pmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{pmatrix}$$

As a result, we have for $n \geq 0$,

$$A^n = S \begin{pmatrix} d_1^n & 0 \\ 0 & d_2^n \end{pmatrix} S^{-1} = \frac{1}{\sqrt{5}} \begin{pmatrix} \left(\frac{1+\sqrt{5}}{2}\right)^{n-1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n-1} & \left(\frac{1+\sqrt{5}}{2}\right)^{n} - \left(\frac{1-\sqrt{5}}{2}\right)^{n} \\ \left(\frac{1+\sqrt{5}}{2}\right)^{n} - \left(\frac{1-\sqrt{5}}{2}\right)^{n} & \left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+1} \end{pmatrix}.$$

Ultimately, we get

$$x_n = A^n x_0 = \frac{1}{\sqrt{5}} \begin{pmatrix} \left(\frac{1+\sqrt{5}}{2}\right)^{n} - \left(\frac{1-\sqrt{5}}{2}\right)^{n} \\ \left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+1} \end{pmatrix} = \begin{pmatrix} F_n \\ F_{n+1} \end{pmatrix}.$$

Hence, we have for $n \geq 0$,

$$F_n = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right) = \frac{1}{2^n\sqrt{5}}\left((1+\sqrt{5})^n - (1-\sqrt{5})^n\right).$$

Theorem 18.6 *Let $v_1, \ldots, v_m \in V$ be eigenvectors of a linear map $F : V \to V$ with <u>different</u> eigenvalues $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$. Then $v_1, \ldots, v_m$ are linearly independent.*

Proof. We prove this statement by induction on m.
1) <u>Base step m=1:</u> v_1 is an eigenvector so $v_1 \neq \mathbf{0}$ and therefore, it is linearly independent.
2) <u>Induction step:</u> Assume that $v_1, \ldots, v_l$, for $1 \leq l < m$, are linearly independent. Now, assume that

we have for some $c_i \in \mathbb{R}$ and $1 \le i \le m$,

$$\sum_{i=1}^{m} c_i v_i = \mathbf{0}. \tag{$*$}$$

Then, we get the equation

$$\mathbf{0} = F(\mathbf{0}) = F\left(\sum_{i=1}^{m} c_i v_i\right) = \sum_{i=1}^{m} c_i F(v_i)$$

$$\Rightarrow \quad \sum_{i=1}^{m} c_i \lambda_i v_i = \mathbf{0}. \tag{$**$}$$

Now, we multiply $(*)$ by λ_m, then subtract $(**)$ from it, and get

$$\mathbf{0} = \lambda_m \sum_{i=1}^{m} c_i v_i - \sum_{i=1}^{m} c_i \lambda_i v_i = \sum_{i=1}^{m} c_i(\lambda_m - \lambda_i) v_i = \sum_{i=1}^{m-1} c_i(\lambda_m - \lambda_i) v_i.$$

Since $v_1, \ldots, v_{m-1}$ are linearly independent by the induction hypothesis, we have for any $1 \le i \le m-1$

$$c_i(\lambda_m - \lambda_i) = 0.$$

Also, since all eigenvalues considered are different, which implies that $\lambda_m \neq \lambda_i$ for any $1 \le i \le m-1$, we get $c_i = 0$ for any $1 \le i \le m-1$. By $(*)$, we get

$$c_m v_m = 0 \quad \Rightarrow \quad c_m = 0.$$

As a result, $v_1, \ldots, v_m$ are linearly independent. □

Corollary 18.7 *Let $F : V \to V$ be a linear map with eigenvalues $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ and $\dim V = n$.*

i) *If F has n distinct eigenvalues, i.e. $m = n$, then F is diagonalizable.*

ii) *If $B_1, \ldots, B_m$ are bases of $E_{\lambda_1}(F), \ldots, E_{\lambda_m}(F)$, then $B_1 \cup \cdots \cup B_m$ are linearly independent.*

iii) *The map F is diagonalizable if and only if*

$$\sum_{j=1}^{m} \dim E_{\lambda_j}(F) = n\,.$$

Proof. The proof is left as exercise. □

Recall: Let f be a polynomial and λ be any real number such that $f(\lambda) = 0$. Then we have

$$f(\lambda) = 0 \quad \Leftrightarrow \quad f(x) = (x - \lambda) g(x)$$

for some polynomial g. Also, if $f(x) = (x - \lambda)^k g(x)$ where $g(\lambda) \neq 0$, then k is called the **multiplicity of $\boldsymbol{\lambda}$** in f. For example, $f(x) = (x-2)^3(x-5)$ has two roots, $\lambda_1 = 5$ with multiplicity 1 and $\lambda_2 = 2$ with multiplicity 3.

Definition 18.8 Let $F : V \to V$ be a linear map and let $\lambda \in \mathbb{R}$ be an eigenvalue of F.

i) The **algebraic multiplicity** of λ, denoted by $\operatorname{algmu}_F(\lambda)$, is the multiplicity of λ in the characteristic polynomial f_F.

ii) The **geometric multiplicity** of λ is given by $\operatorname{geomu}_F(\lambda) = \dim E_\lambda(F)$.

Example 69 1) In Example 66, we considered the linear map

$$\begin{aligned} F : \mathbb{R}^3 &\longrightarrow \mathbb{R}^3, \\ x &\longmapsto Ax, \end{aligned} \qquad A = \frac{1}{3}\begin{pmatrix} 1 & 2 & -2 \\ 2 & 1 & 2 \\ -2 & 2 & 1 \end{pmatrix}.$$

The characteristic polynomial is $f_F(\lambda) = -(1-\lambda)^2(1+\lambda)$, which has two roots, $\lambda_1 = 1$ with $\operatorname{algmu}_F(\lambda_1) = 2$ and $\lambda_2 = -1$ with $\operatorname{algmu}_F(\lambda_2) = 1$. The eigenspaces of F are

$$E_{\lambda_1}(F) = \operatorname{span}\left\{\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}\right\} \qquad \text{with } \operatorname{geomu}_F(\lambda_1) = 2,$$

$$E_{\lambda_2}(F) = \operatorname{span}\left\{\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}\right\} \qquad \text{with } \operatorname{geomu}_F(\lambda_2) = 1.$$

2) Consider $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. We can easily obtain $f_A(\lambda) = \det(A - \lambda I_2) = (1-\lambda)^2$. The eigenvalue is $\lambda = 1$ with the algebraic multiplicity $\operatorname{algmu}_A(\lambda) = 2$. However, the eigenspace is $E_\lambda(A) = \operatorname{span}\left\{\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right\}$ with the geometric multiplicity $\operatorname{geomu}_A(\lambda) = 1$.

Theorem 18.9 *Let $F : V \to V$ be a linear map and $\lambda \in \mathbb{R}$ be an eigenvalue of F. Then*

$$\operatorname{geomu}_F(\lambda) \leq \operatorname{algmu}_F(\lambda).$$

Proof. Let $m = \operatorname{geomu}_F(\lambda)$ and $(v_1, \ldots, v_m)$ be a basis of $E_\lambda(F)$. We extend this basis to a basis of V, which is $B = (v_1, \ldots, v_m, w_{m+1}, \ldots, w_n)$. Then we get

$$[F]_B = \left(\begin{array}{ccc|c} \lambda & & 0 & \\ & \ddots & & * \\ 0 & & \lambda & \\ \hline & 0 & & N \end{array}\right),$$

where upper left quarter is a $m \times m$ diagonal matrix, the upper right one is some $m \times (n-m)$ matrix, the lower left one is a zero matrix, and the lower right one is a $(n-m) \times (n-m)$ matrix N. With this, we have the characteristic polynomial

$$f_F(x) = \det([F]_B - xI_n) = \left(\begin{array}{ccc|c} \lambda - x & & 0 & \\ & \ddots & & * \\ 0 & & \lambda - x & \\ \hline & 0 & & N - xI_{n-m} \end{array}\right) = (\lambda - x)^m \det(N - xI_{n-m}).$$

If $\det(N - \lambda I_{n-m}) \neq 0$, then $\operatorname{algmu}_F(\lambda) = m$. Otherwise, if $\det(N - \lambda I_{n-m}) = 0$, then $\operatorname{algmu}_F(\lambda) > m$. Overall, $\operatorname{algmu}_F(\lambda) \geq m = \operatorname{geomu}_F(\lambda)$. □

Corollary 18.10 *If F is diagonalizable then $\operatorname{geomu}_F(\lambda) = \operatorname{algmu}_F(\lambda)$ for all eigenvalues λ of F.*

Proof. Let $\lambda_1, \dots, \lambda_m$ be eigenvalues of F. We have

$$\sum_{j=1}^{m} \mathrm{geomu}_F(\lambda_j) \leq \sum_{j=1}^{m} \mathrm{algmu}_F(\lambda_j) \leq \dim(V),$$

where we use Theorem 18.9 in the first inequality. If F is diagonalizable, then we have by Corollary 18.7

$$\sum_{j=1}^{m} \mathrm{geomu}_F(\lambda_j) = \dim(V).$$

As a result, we get

$$\dim(V) = \sum_{j=1}^{m} \mathrm{geomu}_F(\lambda_j) \leq \sum_{j=1}^{m} \mathrm{algmu}_F(\lambda_j) \leq \dim(V),$$

which implies that $\sum_{j=1}^{m} \mathrm{algmu}_F(\lambda_j) = \dim(V)$ and $\mathrm{geomu}_F(\lambda_i) = \mathrm{algmu}_F(\lambda_i)$ for all $i = 1, \dots, m$. $\square$

Remark. The converse is not true since there are matrices with algebraic multiplicities of eigenvalues that do not add up to the order of the characteristic polynomial. For example, the matrix $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$ has only one eigenvalue $\lambda = 1$ with $\mathrm{geomu}_A(\lambda) = \mathrm{algmu}_A(\lambda) = 1$. However, the converse is true if we consider complex matrices because of the Fundamental Theorem of Algebra.

18.2 The spectral theorem

In this section we will just consider the vector space $V = \mathbb{R}^n$. The goal of this lecture is to explain the spectral theorem, which states that every symmetric matrix A is diagonalizable and that there exist an orthnormal basis of eigenvectors of A. Recall that the **norm** of a vector $x \in \mathbb{R}^n$ is defined by

$$\|x\| = \sqrt{x_1^2 + \cdots + x_n^2}\,, \qquad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n\,.$$

Definition 18.11 An **orthogonal map** is a linear map $F : \mathbb{R}^n \to \mathbb{R}^n$, such that

$$\|F(x)\| = \|x\|\,, \qquad \forall x \in \mathbb{R}^n\,,$$

i.e. the map F does not change the norm of a vector. We call a matrix $A \in \mathbb{R}^{n \times n}$ orthogonal if $\|Ax\| = \|x\|$ for all $x \in \mathbb{R}^n$.

Recall that the **dot product** $\bullet$ for two vectors $x, y \in \mathbb{R}^n$ is defined by

$$x \bullet y = x^T y = x_1 y_1 + \cdots + x_n y_n \in \mathbb{R}\,.$$

With this the norm of a vector can also be written as $\|x\| = \sqrt{x \bullet x}$.

Lemma 18.12 *For all $x, y \in \mathbb{R}^n$ we have*

$$x \bullet y = \frac{1}{2}\left(\|x+y\|^2 - \|x\|^2 - \|y\|^2\right)\,.$$

Proof. By direct computation, we have

$$\|x+y\|^2 = (x+y)\bullet(x+y) = \|x\|^2 + 2x\bullet y + \|y\|^2$$

$$\Rightarrow \quad x\bullet y = \frac{1}{2}\left(\|x+y\|^2 - \|x\|^2 - \|y\|^2\right) \qquad \square$$

Proposition 18.13 *A linear map $F:\mathbb{R}^n\to\mathbb{R}^n$ is orthogonal if and only if*

$$F(x)\bullet F(y) = x\bullet y$$

for all $x,y\in\mathbb{R}^n$.

Proof. Assume F is orthogonal. Then, we get by Lemma 18.12

$$\begin{aligned} F(x)\bullet F(y) &= \frac{1}{2}\left(\|F(x)+F(y)\|^2 - \|F(x)\|^2 - \|F(y)\|^2\right) \\ &= \frac{1}{2}\left(\|F(x+y)\|^2 - \|F(x)\|^2 - \|F(y)\|^2\right) \\ &= \frac{1}{2}\left(\|x+y\|^2 - \|x\|^2 - \|y\|^2\right) = x\bullet y. \\ &= x\bullet y. \end{aligned}$$

Conversely, assume that F satisfies $F(x)\bullet F(y) = x\bullet y$ for any $x,y\in\mathbb{R}^n$. By considering $x=y$, we have

$$\|F(x)\|^2 = F(x)\bullet F(x) = x\bullet x = \|x\|^2 \quad \Rightarrow \quad \|F(x)\| = \|x\|. \qquad \square$$

Example 70 1) The rotation map $\mathrm{rot}_\varphi:\mathbb{R}^2\to\mathbb{R}^2$ is orthogonal for any $\varphi\in\mathbb{R}$ since we have for any $x\in\mathbb{R}^2$,

$$\|\mathrm{rot}_\varphi(x)\|^2 = \mathrm{rot}_\varphi(x)\bullet\mathrm{rot}_\varphi(x) = \mathrm{rot}_\varphi(x)^T\,\mathrm{rot}_\varphi(x) = x^T[\mathrm{rot}_\varphi]^T[\mathrm{rot}_\varphi]x = x^Tx = \|x\|^2,$$

where

$$[\mathrm{rot}_\varphi]^T[\mathrm{rot}_\varphi] = \begin{pmatrix}\cos(\varphi) & \sin(\varphi)\\ -\sin(\varphi) & \cos(\varphi)\end{pmatrix}\begin{pmatrix}\cos(\varphi) & -\sin(\varphi)\\ \sin(\varphi) & \cos(\varphi)\end{pmatrix} = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} = I_2.$$

2) The projection map $P_U:\mathbb{R}^n\to\mathbb{R}^n$ is not orthogonal for any subspace $U\subset\mathbb{R}^n$ such that $U\neq\{\mathbf{0}\}$ and $U\neq\mathbb{R}^n$. Indeed, we can decompose any $x\notin U$ into a sum of $P_U(x)=x_\parallel\in U$ and $x_\perp\in U^\perp$. As a result, we have $x_\parallel\bullet x_\perp = 0$ and

$$\|x\|^2 = \|x_\parallel + x_\perp\|^2 = \|x_\parallel\|^2 + \|x_\perp\|^2 + 2x_\parallel\bullet x_\perp = \|P_U(x)\|^2 + \|x_\perp\|^2 \neq \|P_U(x)\|^2.$$

3) We can check that $A = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}$ is orthogonal: for any $x = \begin{pmatrix}x_1\\ x_2\end{pmatrix}\in\mathbb{R}^2$,

$$Ax = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}x_1+x_2\\ x_1-x_2\end{pmatrix}$$

$$\Rightarrow \quad \|Ax\| = \frac{1}{\sqrt{2}}\sqrt{(x_1+x_2)^2+(x_1-x_2)^2} = \sqrt{x_1^2+x_2^2} = \|x\|.$$

4) We can check that $B = \begin{pmatrix}1 & 2\\ 3 & 4\end{pmatrix}$ is not orthogonal: we have in general for any arbitrary vector

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2,$$

$$\|Bx\|^2 = Bx \bullet Bx = x^T B^T Bx = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 10 & 14 \\ 14 & 20 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 10x_1^2 + 20x_2^2 + 28x_1x_2 \neq \|x\|^2,$$

where
$$B^T B = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 10 & 14 \\ 14 & 20 \end{pmatrix}.$$

Recall: We say that x and y are **orthogonal** if $x \bullet y = 0$. A basis $B = (b_1, \ldots, b_n)$ of $\mathbb{R}^n$ is called an **orthonormal basis** if b_i and b_j for $i \neq j$ are orthogonal and $\|b_i\| = 1$ for all i, i.e.

$$b_i \bullet b_j = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}.$$

Theorem 18.14 *Let $F : \mathbb{R}^n \to \mathbb{R}^n$ a linear map and $A = [F]_B$ the matrix of F for $B = (e_1, \ldots, e_n)$. The following statements are equivalent.*

i) F is orthogonal.
ii) A is orthogonal.
iii) For all $x, y \in \mathbb{R}^n$ we have $F(x) \bullet F(y) = x \bullet y$.
iv) A is invertible and $A^{-1} = A^T$.
v) $(F(e_1), \ldots, F(e_n))$ (the columns of A) is an orthonormal basis of $\mathbb{R}^n$.
vi) If $(b_1, \ldots, b_n)$ is an orthonormal basis of $\mathbb{R}^n$ then $(F(b_1), \ldots, F(b_n))$ is also an orthonormal basis.

Proof. The equivalence i) $\Leftrightarrow$ ii) is clear by definition. Also, the equivalence i) $\Leftrightarrow$ iii) is Proposition 18.13. To prove the rest of this proposition, it is enough to just prove that iii) $\Rightarrow$ vi) $\Rightarrow$ v) $\Rightarrow$ iv) $\Rightarrow$ iii). Assume that the statement iii) is true, i.e. $F(x) \bullet F(y) = x \bullet y$ for any $x, y \in \mathbb{R}^n$. Then, let $(b_1, \ldots, b_n)$ be an ONB. Hence, we get by using iii),

$$F(b_i) \bullet F(b_j) = b_i \bullet b_j = \delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}.$$

Therefore, $(F(b_1), \ldots, F(b_n))$ is an ONB and we have proven iii) $\Rightarrow$ vi).
Now, assume that the statement vi) is true. Since the basis $(e_1, \ldots, e_n)$ is an ONB, the statement vi) implies that $(F(e_1), \ldots, F(e_n))$ is also an ONB of $\mathbb{R}^n$. Thus, we completed the proof of vi) $\Rightarrow$ v).
The proof of v) $\Rightarrow$ iv) $\Rightarrow$ iii) is left as exercise. □

Corollary 18.15 *i) $A \in \mathbb{R}^{n\times n}$ is orthogonal if and only if A^T is orthogonal.*
ii) If $A, B \in \mathbb{R}^{n\times n}$ are orthogonal then AB is orthogonal.
iii) If B_1 and B_2 are two orthonormal bases, then the change of basis matrix $S_{B_1}^{B_2}$ is orthogonal.

Proof. The statements are proven as follows.

i) By Theorem 18.14, the matrix A is orthogonal if and only if A is invertible and $A^{-1} = A^T$. Using identities $A = (A^{-1})^{-1} = (A^T)^T$, we get
$$A^{-1} = A^T \quad \Leftrightarrow \quad (A^{-1})^{-1} = (A^T)^{-1} \quad \Leftrightarrow \quad A = (A^T)^{-1} \quad \Leftrightarrow \quad (A^T)^T = (A^T)^{-1}.$$
Equivalently, A^T is orthogonal by Theorem 18.14.

ii) If A and B are orthogonal, then $A^{-1} = A^T$ and $B^{-1} = B^T$ by Theorem 18.14. Consequently, we get
$$(AB)^{-1} = B^{-1}A^{-1} = B^T A^T = (AB)^T.$$
Therefore, AB is also orthogonal.

iii) The proof is left as exercise. □

Definition 18.16 i) An **eigenbasis** of a linear map $F : \mathbb{R}^n \to \mathbb{R}^n$ is a basis consisting of eigenvectors of F.

ii) Let $U \subset \mathbb{R}^n$ be a subspace. A linear map $F : U \to U$ is called **symmetric** if we have for all $x, y \in U$

$$x \bullet F(y) = F(x) \bullet y\,.$$

Remark. 1) F has an eigenbasis if and only if F is diagonalizable.

2) Consider any linear map $F : \mathbb{R}^n \to \mathbb{R}^n$ with $A = [F]$. We have for any $x, y \in \mathbb{R}^n$

$$\begin{aligned} F(x) \bullet y &= (Ax)^T y = x^T A^T y, \\ x \bullet F(y) &= x^T A y. \end{aligned}$$

If F is symmetric, i.e. $F(x) \bullet y = x \bullet F(y)$ for any x and y, then we have $x^T A y = x^T A^T y$ for any x and y, which implies that $A = A^T$. Any matrix A such that $A = A^T$ is said to be a **symmetric matrix**. Conversely, if A is symmetric, then it is straightforward to show that F is symmetric.

3) Let $U \subset \mathbb{R}^n$ be any subspace of $\mathbb{R}^n$ and B' be an ONB of U. For any linear map $F : U \to U$, we have

$$\begin{aligned} F(x) \bullet y &= [F(x)]_{B'} \bullet [y]_{B'} = ([F]_{B'}[x]_{B'})^T [y]_{B'} = [x]_{B'}^T [F]_{B'}^T [y]_{B'}, \\ x \bullet F(y) &= [x]_{B'} \bullet [F(y)]_{B'} = [x]_{B'}^T [F]_{B'} [y]_{B'}. \end{aligned}$$

By the same argument as above, F is symmetric if and only if $[F]_{B'}$ is symmetric for any ONB B'.

Theorem 18.17 (Spectral theorem) *Let $U \subset \mathbb{R}^n$ be a subspace and $F : U \to U$ a linear map. Then F is symmetric if and only if there exists an orthonormal eigenbasis of F.*

Proof. We will prove this theorem later. □

Corollary 18.18 *A matrix $A \in \mathbb{R}^{n\times n}$ is symmetric if and only if there exists an orthogonal matrix $S \in \mathbb{R}^{n\times n}$, such that*

$$S^T A S = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix},$$

for some $d_1, \dots, d_n \in \mathbb{R}$.

Proof. Consider $F : \mathbb{R}^n \to \mathbb{R}^n$ defined by $F(x) = Ax$. We know that A is symmetric if and only if F is symmetric. Also, by Theorem 18.17, this is equivalent to that there exists an ONB B' such that $[F]_{B'} = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}$.

Now, assume that A is symmetric or equivalently, F is symmetric. Then, we set $S = S_{B'}^{B}$ where $B = (e_1, \dots, e_n)$ and B' is the ONB as described above. By Corollary 18.15, S is orthogonal. As a result,

$$\begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix} = [F]_{B'} = S_B^{B'} [F]_B S_{B'}^{B} = S^{-1} A S = S^T A S.$$

Conversely, assume that there exists an orthogonal matrix $S \in \mathbb{R}^{n\times n}$ such that $S^T AS = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}$.
Define $b_i = Se_i$ (the i-th column of S) for $1 \le i \le n$. By Theorem 18.14, $B' = (b_1, \dots, b_n)$ is an ONB of $\mathbb{R}^n$. As a result, we get $S = S^B_{B'}$ and

$$\begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix} = S^T AS = S^{B'}_B [F]_B S^B_{B'} = [F]_{B'}.$$

Consequently, $[F]_{B'}$ is symmetric and therefore, F is symmetric. Equivalently, A is symmetric. □

Example 71 In Example 66, we considered the reflection

$$F : \mathbb{R}^3 \longrightarrow \mathbb{R}^3, \quad x \longmapsto Ax, \qquad A = \frac{1}{3}\begin{pmatrix} 1 & 2 & -2 \\ 2 & 1 & 2 \\ -2 & 2 & 1 \end{pmatrix}.$$

The eigenvalues are $\lambda_1 = 1$ and $\lambda_2 = -1$. The eigenspaces are

$$E_{\lambda_1}(F) = \operatorname{span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right\} = \operatorname{span}\{v_1, v_2\}, \qquad E_{\lambda_2}(F) = \operatorname{span}\left\{ \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \right\} = \operatorname{span}\{v_3\}.$$

Since A is symmetric, the spectral theorem dictates that there exists an orthonormal eigenbasis of F. Observe that the eigenspaces $E_{\lambda_1}(F)$ and $E_{\lambda_2}(F)$ are orthogonal to each other since $v_1 \bullet v_3 = v_2 \bullet v_3 = 0$. In general, for any symmetric matrix, eigenvectors corresponding to different eigenvalues are orthogonal to each other. Indeed, suppose that v_1, v_2 are eigenvectors of B with eigenvalues λ_1, λ_2 such that $\lambda_1 \neq \lambda_2$. We have

$$\begin{aligned} & v_1 \bullet (Bv_2) = v_1 \bullet (\lambda_2 v_2) = \lambda_2 (v_1 \bullet v_2), \\ & v_1 \bullet (Bv_2) = (Bv_1) \bullet v_2, \\ & (Bv_1) \bullet v_2 = (\lambda_1 v_1) \bullet v_2 = \lambda_1 (v_1 \bullet v_2) \\ \Rightarrow \quad & (\lambda_1 - \lambda_2)(v_1 \bullet v_2) = 0 \quad \Rightarrow \quad v_1 \bullet v_2 = 0. \end{aligned}$$

Now, coming back to A, we can obtain an orthonormal eigenbasis by using the Gram-Schmidt process for $E_{\lambda_1}(F)$ and $E_{\lambda_2}(F)$. First of all, we have

$$b_1 = \frac{v_1}{\|v_1\|} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \qquad b_3 = \frac{v_3}{\|v_3\|} = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}.$$

Next, a vector in $E_{\lambda_1}(F)$ that is orthogonal to b_1 is

$$w_2 = v_2 - (v_2 \bullet b_1) b_1 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} - \left(-\frac{1}{\sqrt{2}}\right) \cdot \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix}.$$

Then, we get

$$b_2 = \frac{w_2}{\|w_2\|} = \frac{1}{\sqrt{6}}\begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix}.$$

As a result, we get an orthonormal eigenbasis (b_1, b_2, b_3) with $E_{\lambda_1}(F) = \text{span}\{b_1, b_2\}$ and $E_{\lambda_2}(F) = \text{span}\{b_3\}$, and get an orthogonal matrix

$$S = \begin{pmatrix} | & | & | \\ b_1 & b_2 & b_3 \\ | & | & | \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{pmatrix}.$$

With this, we can diagonalize the matrix A, i.e. you can check that

$$S^T A S = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

To prove the spectral theorem, we need the following lemma.

Lemma 18.19 *Every symmetric linear map $F : U \to U$ has an eigenvalue.*

Proof. In this proof, we will employ complex numbers and the fundamental theorem of algebra which states that every polynomial with complex coefficients has at least one complex root (this topic is covered in the course on Complex Analysis). By this theorem, the characteristic polynomial f_F always has at least one complex zero $\lambda \in \mathbb{C}$ and hence, there exists $x \in \mathbb{C}^n$ such that $Ax = \lambda x$. In the context of real vector space, λ is an eigenvalue of the linear map F if $\lambda \in \mathbb{R}$. Since F is real and symmetric, we get

$$Ax = \lambda x \quad \Rightarrow \quad A\overline{x} = \overline{\lambda}\overline{x} \quad \Rightarrow \quad \overline{x}^T A = \overline{\lambda}\overline{x}^T \quad \Rightarrow \quad \overline{\lambda}\overline{x}^T x = \overline{x}^T A x = \overline{x}^T(\lambda x) \quad \Rightarrow \quad (\lambda - \overline{\lambda})\overline{x}^T x = 0.$$

Since $x \neq \mathbf{0}$ and therefore, $\overline{x}^T x > 0$, the last equality implies that $\lambda = \overline{\lambda}$. As a result, λ is a real number. $\square$

Now we can prove the spectral theorem

Proof of Theorem 18.17. First, assume that there exist an orthonormal eigenbasis B' of F. Then, we have

$$[F]_{B'} = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_m \end{pmatrix}.$$

Since the diagonal matrix $[F]_{B'}$ is symmetric, it follows that F is symmetric.
Conversely, assume that $F : U \to U$ is symmetric. We want to prove that there exists an orthonormal eigenbasis of F by induction on the dimension $m = \dim(U)$.

1) <u>Base step m=1:</u> We have $U = \text{span}\{u\}$ for some $u \neq \mathbf{0}$ and we can set $b_1 = u/\|u\|$. Since the linear map F is a function from U to U, there exists $\lambda \in \mathbb{R}$ such that $F(u) = \lambda u$. As result, $B' = (u_1)$ is an orthonormal eigenbasis of F.
2) <u>Induction step:</u> Assume that there exist an orthonormal eigenbasis for any symmetric linear map $F : U' \to U'$ with $1 \leq \dim(U') < m$. Consider now any symmetric linear map $F : U \to U$ with $\dim(U) = m$. By Lemma 18.2, F has an eigenvalue λ. Let $v \in U$ be an eigenvector with eigenvalue λ and it satisfies $\|v\| = 1$. From the definition of symmetric matrices, we have for every $x \in U$,

$$F(x) \bullet v = x \bullet F(v) = x \bullet (\lambda v) = \lambda(x \bullet v).$$

Therefore, if $x \bullet v = 0$, then $F(x) \bullet v = 0$. Hence, $F(x) \in \text{span}\{v\}^\perp$ for every $x \in \text{span}\{v\}^\perp$. We can define a new linear map $F' : \text{span}\{v\}^\perp \to \text{span}\{v\}^\perp$ by $F'(x) = F(x)$ for any $x \in \text{span}\{v\}^\perp$. The linear map F' inherits the symmetry from F. In addition, $\dim(\text{span}\{v\}^\perp) = \dim(U) - 1 = m - 1 < m$. By the induction hypothesis, there exists an orthonormal eigenbasis $(b_1, \ldots, b_{m-1})$ of $\text{span}\{v\}^\perp$ consisting of eigenvectors of F'. As a result, $B' = (b_1, \ldots, b_{m-1}, v)$ is an orthonormal eigenbasis of U consisting of of eigenvectors of F. $\square$

Remark. In this chapter, we worked with subspaces of $\mathbb{R}^n$ only since we needed the dot product $\bullet$. But an analogue of the dot product, which is called scalar product or inner product, can also be defined for other types of vector spaces. General inner product $\langle\cdot,\cdot\rangle : V\times V\to\mathbb{R}$ for a vector space V satisfies the following conditions: for any $x,y,z\in V$ and $a,b\in\mathbb{R}$,

(i) Symmetric: $\langle x,y\rangle=\langle y,x\rangle$,
(ii) Linear: $\langle ax+by,z\rangle=a\langle x,z\rangle+b\langle y,z\rangle$,
(iii) Positive-definiteness: if $v\neq 0$, then $\langle x,x\rangle>0$.

From an inner product $\langle\cdot,\cdot\rangle$, we can always define a norm $\|\cdot\|$ by $\|x\|=\sqrt{\langle x,x\rangle}$ for any $x\in V$. This definition of norm makes sense due to the positive-definiteness of inner products. From this definition, we also get for $x,y\in V$,

$$\langle x,y\rangle=\frac{1}{2}\left(\|x+y\|^2-\|x\|^2-\|y\|^2\right).$$

Having a norm $\|\cdot\|$, we can try to define an inner product by this formula but it is not always possible since $\langle\cdot,\cdot\rangle$ defined in this way may not satisfy condition (ii). For example, consider $V=\mathbb{R}^n$ with a norm $\|\cdot\|_1$ defined by

$$\|x\|_1=|x_1|+|x_2|+\cdots+|x_n|.$$

Let $\langle\cdot,\cdot\rangle_1$ be defined by $\langle x,y\rangle_1=\frac{1}{2}\left(\|x+y\|_1^2-\|x\|_1^2-\|y\|_1^2\right)$. Now consider $x=-e_1$, $y=e_2$, and $z=e_1$, where $e_1=\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}$ and $e_2=\begin{pmatrix}0\\1\\0\\\vdots\\0\end{pmatrix}$. We have

$$\begin{aligned}
\langle x,z\rangle_1&=\frac{1}{2}\left(\|-e_1+e_1\|_1^2-\|-e_1\|_1^2-\|e_1\|_1^2\right)=\frac{1}{2}(0-1-1)=-1\\
\langle y,z\rangle_1&=\frac{1}{2}\left(\|e_2+e_1\|_1^2-\|e_2\|_1^2-\|e_1\|_1^2\right)=\frac{1}{2}(2^2-1-1)=1\\
\langle x+y,z\rangle_1&=\frac{1}{2}\left(\|-e_1+e_2+e_1\|_1^2-\|-e_1+e_2\|_1^2-\|e_1\|_1^2\right)=\frac{1}{2}(1-2^2-1)=-2\\
\Rightarrow\quad &\langle x,z\rangle_1+\langle y,z\rangle_1=0\neq-2=\langle x+y,z\rangle_1.
\end{aligned}$$

Therefore, $\langle\cdot,\cdot\rangle_1$ is not an inner product. A vector space with a norm is called a normed vector space. From our discussion, a vector space with an inner product is also a normed vector space but not every normed vector space has an inner product.

Example 72 Consider $V=\mathcal{C}^0([-\pi,\pi])$, space of continuous functions on $[-\pi,\pi]$, with an inner product $\langle\cdot,\cdot\rangle$ defined by

$$\langle f,g\rangle=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)g(x)\,\mathrm{d}x\,.$$

By direct computations, we can show for any $n,m\geq 0$ that

$$\begin{aligned}
\langle\sin(nx),\sin(mx)\rangle&=\begin{cases}1, & \text{if } n=m,\\ 0, & \text{otherwise},\end{cases}\\
\langle\cos(nx),\cos(mx)\rangle&=\begin{cases}2, & \text{if } n=m=0,\\ 1, & \text{if } n=m\neq 0,\\ 0, & \text{otherwise},\end{cases}\\
\langle\sin(nx),\cos(mx)\rangle&=0.
\end{aligned}$$

The vector space $T=\operatorname{span}\{1,\sin(x),\cos(x),\sin(2x),\cos(2x),\dots\}$ is called the space of “real trigonometric polynomials”. There is a theorem showing that any continuous function in $[-\pi,\pi]$ can be “approximated well” by trigonometric polynomials. This has applications in signal processing where any wave with a complicated shape can be decomposed into a sum of sine wave.

How can we do that? For simplicity, consider that we can write for some $f \in V$,

$$f(x) = \frac{a_0}{2} + \sum_{j=1}^{m} a_j \cos(jx) + \sum_{k=1}^{n} b_k \sin(kx).$$

We can interpret j as frequencies and a_j, b_k as "volume" of these frequencies. We learned from Linear Algebra 1 that if $B = (b_1, \ldots, b_n)$ is an ONB then for any $u \in \mathbb{R}^n$, we have $u = \sum_{j=1}^{n} \lambda_j b_j$ with $\lambda_j = u \bullet b_j$. Similarly, here we also have for $1 \leq j \leq m$ and $1 \leq k \leq n$,

$$a_j = \langle f, \cos(jx) \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(jx) \, \mathrm{d}x \,,$$
$$b_k = \langle f, \sin(kx) \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(kx) \, \mathrm{d}x \,,$$

although the difference is that $a_0/2 = \langle f, 1 \rangle / 2$. In general, we have $n, m \to \infty$ and the above sum turns into a **Fourier series**.

Remark. The spectral theorem holds in general for all finite generated vector spaces but there is also a version of this theorem for infinite dimensional vector spaces.

19

Applications

In this chapter, we give some "real life" application examples of the theory learned in this course. We start by presenting a way of numerically calculating the eigenvalues. Since we will use Python to give some explicit example, we start with a general example on how to represent matrices and vectors in Python (You can try out this code by yourself in the following Colab Notebook: `https://colab.research.google.com/drive/1wMY-LkJszvg1QJeKhrIjpxPjsjlKbPQk?usp=sharing`)

Example 73 Previously we considered the matrix

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

and calculated its eigenvalues & eigenvectors. We saw that the eigenvalues are $\lambda_1 = 1$ and $\lambda_2 = -1$. The eigenspaces are

$$E_1(A) = \text{span}\left\{\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}\right\}, \qquad E_{-1}(A) = \text{span}\left\{\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}\right\}.$$

Let us check this in python now. The following code defines the matrix A and the 3 vectors which span the eigenspaces. After this we multiply the vectors with the matrix A and check what we get.

```python
# In python you can import a lot of packages which offer functions.
# For example we want to import the package "numpy" which includes numerical operations such as matrix multiplication.
# The following command includes the package numpy and calls it "np" in the following.
import numpy as np

# The matrix A
A=1/3*np.array([[1,2,-2],[2,1,2],[-2,2,1]])

# The print command gives an output.
print("Hello. This is the matrix A:")
print(A)

# The basisvectors of the first eigenspace (eigenvalue 1)
v1 = np.array([1,1,0])
v2 = np.array([-1,0,1])

# The basisvector of the second eigenspace (eigenvalue -1)
v3 = np.array([1,-1,1])

# The command np.dot(A,B) multiplies a matrix A with a matrix B (for example B could be a vector)

# Multiply A with the vectors
print("A * v1 = ",np.dot(A,v1))
print("A * v2 = ",np.dot(A,v2))
print("A * v3 = ",np.dot(A,v3))
```

Output:

```
Hello. This is the matrix A:
[[ 0.33333333  0.66666667 -0.66666667]
 [ 0.66666667  0.33333333  0.66666667]
 [-0.66666667  0.66666667  0.33333333]]
A * v1 =  [1. 1. 0.]
A * v2 =  [-1.  0.  1.]
A * v3 =  [-1.  1. -1.]
```

19.1 Power iteration

In this section, we answer the question how one can calculate eigenvalues for big matrices numerically. Calculating the characteristic polynomial of a 100×100 matrix and then finding its zeros is clearly not practical. The rough idea instead is the following: If we have a matrix A and a random vector v, then $A^k v$ for large k will give an approximation for an eigenvector of A for its "largest" eigenvalue.

More precisely we consider the following situation:
(i) Let $A \in \mathbb{R}^{n\times n}$ be diagonalizable.
(ii) Eigenvalues of A: $\lambda_1, \ldots, \lambda_n$ with $\lambda_1 = \cdots = \lambda_d$ and $|\lambda_1| > |\lambda_j|$ for $j = d+1, \ldots, n$. (Here λ_1 is called the **dominant eigenvalue** of A).
(iii) Let $(b_1, \ldots, b_n)$ be a basis of $\mathbb{R}^n$ of eigenvectors b_j with eigenvalue λ_j.
(iv) Suppose that we have $v = \sum_{j=1}^n \alpha_j b_j$ for $\alpha_j \in \mathbb{R}$, such that $\sum_{j=1}^d \alpha_j b_j \neq 0$.

In this situation, we calculate, by using $A^k b_j = A^{k-1}\lambda_j b_j = \cdots = \lambda_j^k b_j$,

$$\begin{aligned} A^k v = A^k \sum_{j=1}^n \alpha_j b_j &= \sum_{j=1}^n \alpha_j A^k b_j = \sum_{j=1}^n \alpha_j \lambda_j^k b_j \\ &= \sum_{j=1}^d \alpha_j \lambda_1^k b_j + \sum_{j=d+1}^n \alpha_j \lambda_j^k b_j \\ &= \lambda_1^k \left(\sum_{j=1}^d \alpha_j b_j + \sum_{j=d+1}^n \alpha_j \left(\frac{\lambda_j}{\lambda_1}\right)^k b_j \right) . \end{aligned}$$

Since $|\frac{\lambda_j}{\lambda_1}| < 1$ for $j > d$ we have $\lim_{k\to\infty} \left(\frac{\lambda_j}{\lambda_1}\right)^k = 0$. The vector $\sum_{j=1}^d \alpha_j b_j$ is an eigenvector of A for eigenvalue λ_1 and therefore the above expression gives an approximation for an eigenvector for big k. But because of the factor λ_1^k it also makes sense to normalize the vector in each step. Motivated by this the power iteration algorithm works as follows:

Algorithm 19.1 (Power iteration) Let $A \in \mathbb{R}^{n\times n}$.
(i) Start with a random vector $v_0 \in \mathbb{R}^n$.
(ii) Define for $k \geq 0$

$$v_{k+1} = \frac{Av_k}{||Av_k||}$$

Then for large k the vector v_k will give an approximation for an eigenvector of A for its dominant eigenvalue.

Example 74 Let us consider the following matrix

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

This has eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 2$, i.e. 2 is the dominant eigenvalue. The eigenspace to the eigenvalue 2 is given by

$$E_2(B) = \operatorname{span}\left\{ \begin{pmatrix} 0 \\ 2 \\ 0 \\ 1 \end{pmatrix} \right\}$$

```python
# We will do the power iteration in numpy

# Define the matrix A in numpy as
A = np.array([[1,1,6,-2],[0,1,-3,2],[0,0,1,0],[0,0,-2,2]])

# Start vector
v = np.array([1,2,3,4])

# for i in range(1,n) does something n-1 times.
for i in range(1,100):
  # Multiply the vector v with the matrix B
  v = np.dot(A,v)
  # normalize the vector
  v = 1 / np.linalg.norm(v) * v

# The round functions round the vector
print("New vector:\n",v.round(1))
product=np.dot(A,v)
print("This vector multiplied with A:\n", product.round(1))
print("possible eigenvalue:",product[1]/v[1])
```

Output:

```
New vector:
 [ 0.   -0.9  0.   -0.4]
This vector multiplied with A:
 [ 0.   -1.8  0.   -0.9]
possible eigenvalue: 2.0
```

19.2 Population growth - Leslie matrices

The graph shows the population of Japan in different age groups in the year 2022. From this data, we define

$$v_{2022} = \begin{pmatrix} 11.7 & 14.6 & 17.4 & 21.4 & 19.7 & 10.0 & 5.2 \end{pmatrix}^T .$$

Assume that we also know for each age group the fertility rate f_j (how much new babies are created per person) and the survival rates s_j for each age group.
With this, we can define a matrix which describes the change of the population v_{2022} in the year 2017 to the next generation $2022 + 15 = 2037$:

$$v_{2037} = Lv_{2022} = \begin{pmatrix} f_1 & f_2 & f_3 & f_4 & f_5 & f_6 & f_7 \\ s_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & s_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & s_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & s_4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & s_5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & s_6 & 0 \end{pmatrix} \begin{pmatrix} 11.7 \\ 14.6 \\ 17.4 \\ 21.4 \\ 19.7 \\ 10.0 \\ 5.2 \end{pmatrix},$$

where a matrix with the form of the matrix L is called **Leslie matrix**. Eigenvectors of this matrix are stable population distributions. The dominant eigenvalue λ ($|\lambda| \geq |\lambda_j|$) of this matrix indicates the growth rate. There are 3 cases: if $\lambda < 1$, the population dies out; if $\lambda = 1$, the population stays the same; if $\lambda > 1$, the population gets bigger.

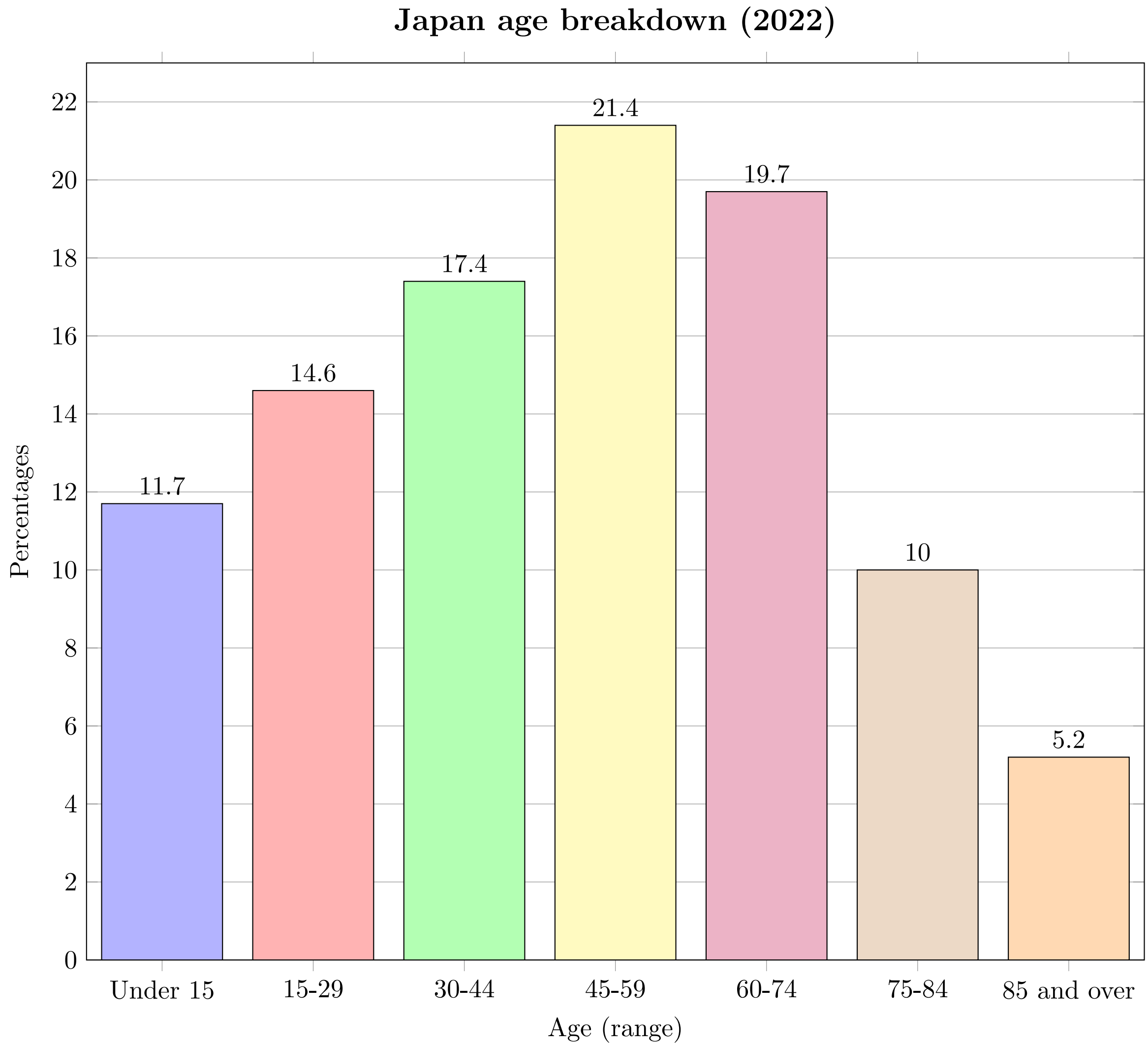


Figure 19.1: Source: `https://kids.britannica.com/students/assembly/view/208943`

Example 75 Consider the Leslie matrix

$$A = \begin{pmatrix} 0 & 3 & 2 \\ \frac{4}{5} & 0 & 0 \\ 0 & \frac{2}{5} & 0 \end{pmatrix}$$

This matrix describes the change of 3 age groups (a_1, a_2, a_3) after one generation. From the Leslie matrix A, we have after one generation:

a_1 does not breed;
a_2 creates in average 3 babies;
a_3 creates in average 2 babies;
$\frac{4}{5}$ or 80% of a_1 survives and becomes the new a_2;
$\frac{2}{5}$ or 40% of a_2 survives and becomes the new a_3;
a_3 dies out after one generation.

The characteristic polynomial of A is

$$f_A(\lambda) = \det \begin{pmatrix} -\lambda & 3 & 2 \\ \frac{4}{5} & -\lambda & 0 \\ 0 & \frac{2}{5} & -\lambda \end{pmatrix} = -\lambda^3 + \frac{12}{5}\lambda + \frac{16}{25}.$$

The eigenvalues of the matrix A are

$$\begin{aligned} \lambda_1 &= 1.668..., \\ \lambda_2 &= -1.39..., \\ \lambda_3 &= -0.275..., \end{aligned}$$

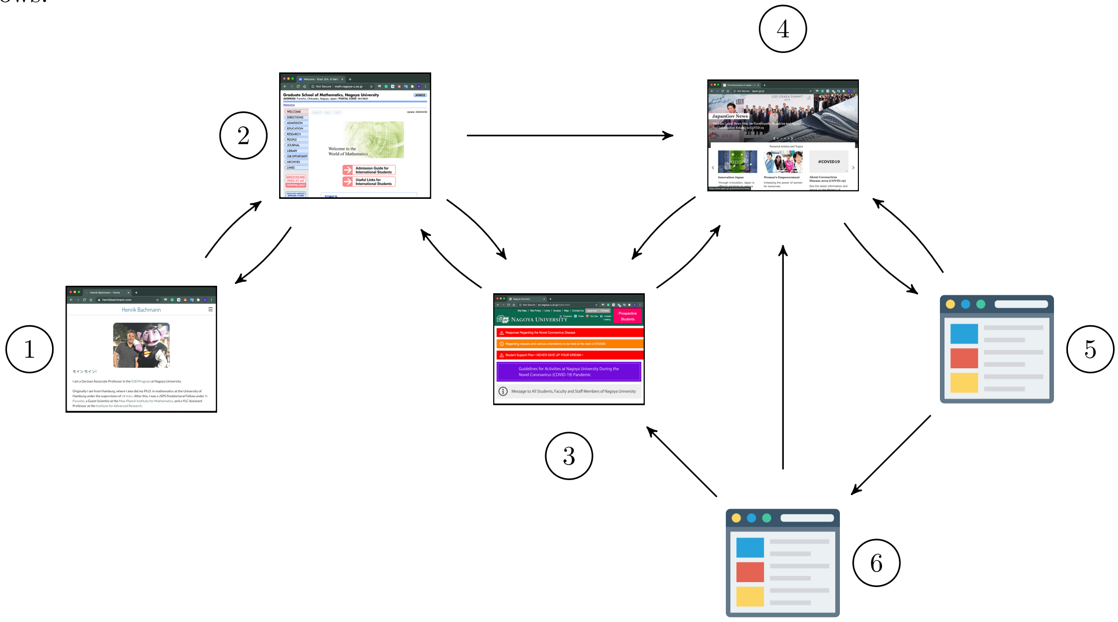
where λ_1 is the dominant eigenvalue. From this, we interpret that after each generation we have an approximate growth of the population by 66.8%. An eigenvector corresponding to λ_1 is

$$v \approx \begin{pmatrix} 89 \\ 43 \\ 10 \end{pmatrix}.$$

If the age distribution is $\begin{pmatrix} 89 \\ 43 \\ 10 \end{pmatrix}$, then the age distribution stays the same in the sense that $\dfrac{Av}{\|Av\|} = \dfrac{v}{\|v\|}$.

19.3 Google page rank

The Google PageRank algorithm assign a number indicating "importance" to each webpages. These numbers can be used to sort webpages in a search result. The Google PageRank algorithm was used by Google in the early days (about 20 years ago). Assume that we have a mini version of the web as follows.

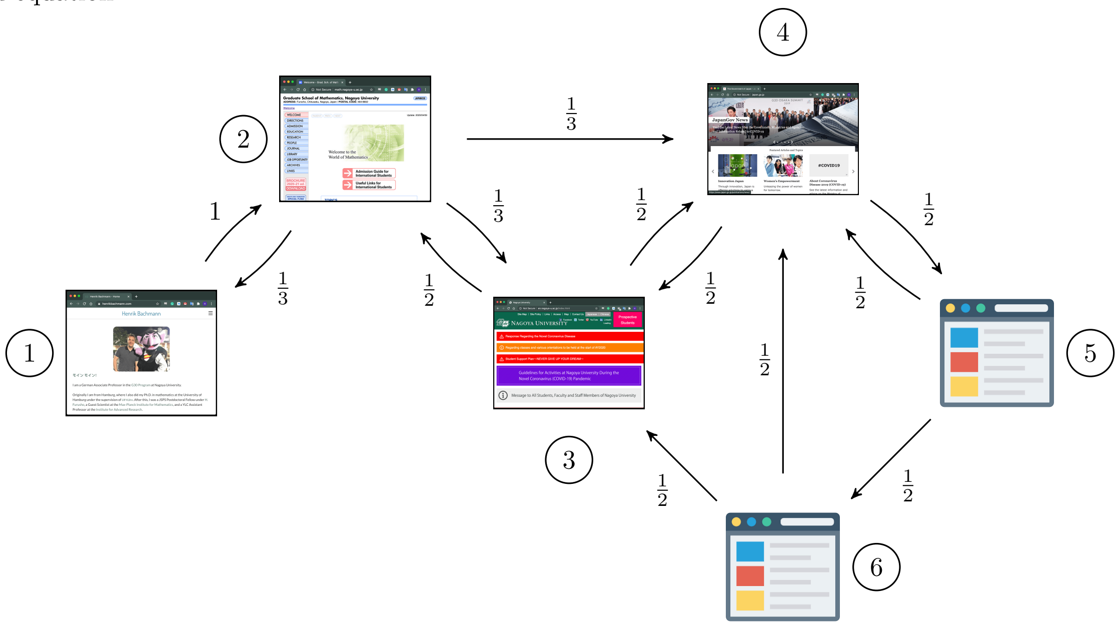


Each arrow indicates a link from one page to another. The question is how to rank these 6 pages by importance? The idea of the Google PageRank algorithm is follows:

- A page is important if another page links to it.
- Getting a link from an important should count more.

- The importance from one page is distributed uniformly to all pages it links to.

Let $x_1, x_2, \dots, x_6$ denote the importance of pages $1, 2, \dots, 6$, respectively. We get the following graph and equation



$$\begin{pmatrix} 0 & \frac{1}{3} & 0 & 0 & 0 & 0 \\ 1 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} \quad \Leftrightarrow \quad Ax = x,$$

where the i-th row of the matrix A represents links to page i and the j-th column represents links from j page. For example, since page 2 links to 3 pages (page 1, 3, 4) so we have $a_{12} = a_{32} = a_{42} = \frac{1}{3}$.
To find the importance of webpages, we want to find an eigenvector of A with eigenvalue 1. In our example, we have an eigenvector with eigenvalue 1 given by

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} = \begin{pmatrix} \frac{5}{6} \\ \frac{5}{2} \\ \frac{10}{3} \\ 4 \\ 2 \\ 1 \end{pmatrix} \approx \begin{pmatrix} 0.8 \\ 2.5 \\ 3.3 \\ 4 \\ 2 \\ 1 \end{pmatrix}.$$

In the interpretation of this vector x, page 4 is the most important one then page 3, page 2, page 4, page 5, and at last the most unimportant page 1.
One may question whether we can always have eigenvalue 1 for the matrix A in a more general situation. The answer is yes due to the following reason. Observe that the sum of the entries in each column of A is 1. Therefore, the sum of the entries in each row of A^T is 1. As a result, A^T has eigenvalue 1 because $A^T \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$. Hence, A also has eigenvalue 1 since $f_A(\lambda) = f_{A^T}(\lambda)$.

19.4 Decision making - Analytic hierarchy process

Similar to the Page rank, let's assume that we want order n object $x_1, \ldots, x_n$ by importance. However, this time the input we are given is their pairwise comparison, i.e. we know the values $\frac{x_i}{x_j}$ for $1 \leq i, j \leq n$. If we know $\frac{x_i}{x_j}$, how can we obtain $x_1, \ldots, x_n$? What would be a practical application of this? Let start with an (stupid) example. Assume that you need to rank your 4 most favorite "restaurants".

1	2	3	4

Looking at the 4 options at once, it might by difficult to rank them. Usually, it is easier to compare them pairwise. Let's assume that we have the following preferences:

	VS		:	$4:1$	$\Leftrightarrow$	$\frac{x_1}{x_2} = 4$
	VS		:	$2:1$	$\Leftrightarrow$	$\frac{x_1}{x_3} = 2$
	VS		:	$3:2$	$\Leftrightarrow$	$\frac{x_1}{x_4} = \frac{3}{2}$
	VS		:	$1:2$	$\Leftrightarrow$	$\frac{x_2}{x_3} = \frac{1}{2}$
	VS		:	$1:3$	$\Leftrightarrow$	$\frac{x_2}{x_4} = \frac{1}{3}$
	VS		:	$1:1$	$\Leftrightarrow$	$\frac{x_1}{x_2} = 1$

Now consider the matrix

$$A = \left(\frac{x_i}{x_j}\right)_{1 \leq i,j \leq 4} = \begin{pmatrix} 1 & 4 & 2 & \frac{3}{2} \\ \frac{1}{4} & 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & 2 & 1 & 1 \\ \frac{2}{3} & 3 & 1 & 1 \end{pmatrix}.$$

We claim that an eigenvector of A corresponding to the dominant eigenvalue gives a good approximation for a ranking. In our example, we have eigenvalues $\lambda_1 = 4$ and $\lambda_2 = 0$. For the dominant eigenvalue λ_1, we have an eigenvector v given by

$$v = \begin{pmatrix} 0.75 \\ 0.18 \\ 0.41 \\ 0.48 \end{pmatrix} \quad \begin{array}{l} \text{1st} \\ \text{4th} \\ \text{3rd} \\ \text{2nd} \end{array}.$$

Why does this process work? In general, the matrix $A = \left(\frac{x_i}{x_j}\right)_{1 \leq i,j \leq n}$ has the eigenvalues $\lambda_1 = n$ and

$\lambda_2 = \lambda_3 = ... = \lambda_n = 0$ with eigenvectors $\begin{pmatrix} x_1 \\ \cdot \\ x_n \end{pmatrix}, \begin{pmatrix} x_1 \\ -x_2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} x_1 \\ 0 \\ -x_3 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} x_1 \\ 0 \\ \vdots \\ 0 \\ -x_n \end{pmatrix}$. For example, we have for $n = 3$,

$$\begin{pmatrix} \frac{x_1}{x_1} & \frac{x_1}{x_2} & \frac{x_1}{x_3} \\ \frac{x_2}{x_1} & \frac{x_2}{x_2} & \frac{x_2}{x_3} \\ \frac{x_3}{x_1} & \frac{x_3}{x_2} & \frac{x_3}{x_3} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 + x_1 + x_1 \\ x_2 + x_2 + x_2 \\ x_3 + x_3 + x_3 \end{pmatrix} = 3 \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix},$$

$$\begin{pmatrix} \frac{x_1}{x_1} & \frac{x_1}{x_2} & \frac{x_1}{x_3} \\ \frac{x_2}{x_1} & \frac{x_2}{x_2} & \frac{x_2}{x_3} \\ \frac{x_3}{x_1} & \frac{x_3}{x_2} & \frac{x_3}{x_3} \end{pmatrix} \begin{pmatrix} x_1 \\ -x_2 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{x_1}{x_1} & \frac{x_1}{x_2} & \frac{x_1}{x_3} \\ \frac{x_2}{x_1} & \frac{x_2}{x_2} & \frac{x_2}{x_3} \\ \frac{x_3}{x_1} & \frac{x_3}{x_2} & \frac{x_3}{x_3} \end{pmatrix} \begin{pmatrix} x_1 \\ 0 \\ -x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Since the pairwise comparison gives a good approximation for $\frac{x_i}{x_j}$, the eigenvector to the dominant eigenvalue $\lambda_1 = n$ will give a good approximation for the importances $x_1, \dots, x_n$.
There are much more applications of eigenvalues and eigenvectors.

19.5 Principal component analysis (PCA)

20

Linear differential equations

20.1 Continuous dynamical systems

In the population behavior example (and the bunny counting), we considered vectors $x_0, x_1, \dots$ in $\mathbb{R}^n$ with $A \in \mathbb{R}^{n\times n}$ and

$$x_{t+1} = Ax_t, \qquad t = 0, 1, 2, \dots \tag{20.1.1}$$

This is called a **discrete (linear) dynamical system**. Another example for a discrete dynamical system would be to describe the money in a saving account.

Example 76 Assume that you have a saving account with 7% interest rate each year. If $x_t \in \mathbb{R}$ is your money in year t, then

$$x_{t+1} = 1.07x_t.$$

This is an example for (20.1.1) in the one-dimensional case with the 1×1 matrix $A = (1.07)$. However, usually interest rate is paid continuously, i.e. not just after one year. So instead of considering $x_0, x_1, \dots,$ one considers $x(t)$, i.e. the money is a function in a time variable $t \in \mathbb{R}$. The interest rate then describes the change of x in time. In this example, we have $x_t = (1.07)^t x_0$ for $t = 0, 1, 2, \dots$ If we let t vary continuously in $\mathbb{R}$, then

$$x'(t) = (1.07)^t x_0 \ln(1.07) = \ln(1.07)x(t).$$

As a result, we get the equation

$$x'(t) = (1.07)^t x_0 \ln(1.07) = \ln(1.07)x(t),$$

which is an example of **one dimensional continuous dynamical system**.

Now we consider the problem when t varies continuously in $\mathbb{R}$. Let $x : \mathbb{R} \to \mathbb{R}^n$ be a function written as

$$x(t) = \begin{pmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{pmatrix},$$

where the entries $x_1, \dots, x_n$ are differentiable functions in $\mathcal{C}^1(\mathbb{R}, \mathbb{R})$. By $x'(t) = \frac{d}{dt}x(t)$ we denote

$$x'(t) = \begin{pmatrix} x_1'(t) \\ \vdots \\ x_n'(t) \end{pmatrix}.$$

For a matrix $A \in \mathbb{R}^{n\times n}$ the equation

$$x'(t) = Ax(t)$$

is called a **continuous (linear) dynamical system**. The goal of this chapter is that for a given $A \in \mathbb{R}^{n\times n}$, we want to find $x(t)$ such that $x'(t) = Ax(t)$.
One dimensional ($n = 1$) continuous dynamical systems have the following solutions:

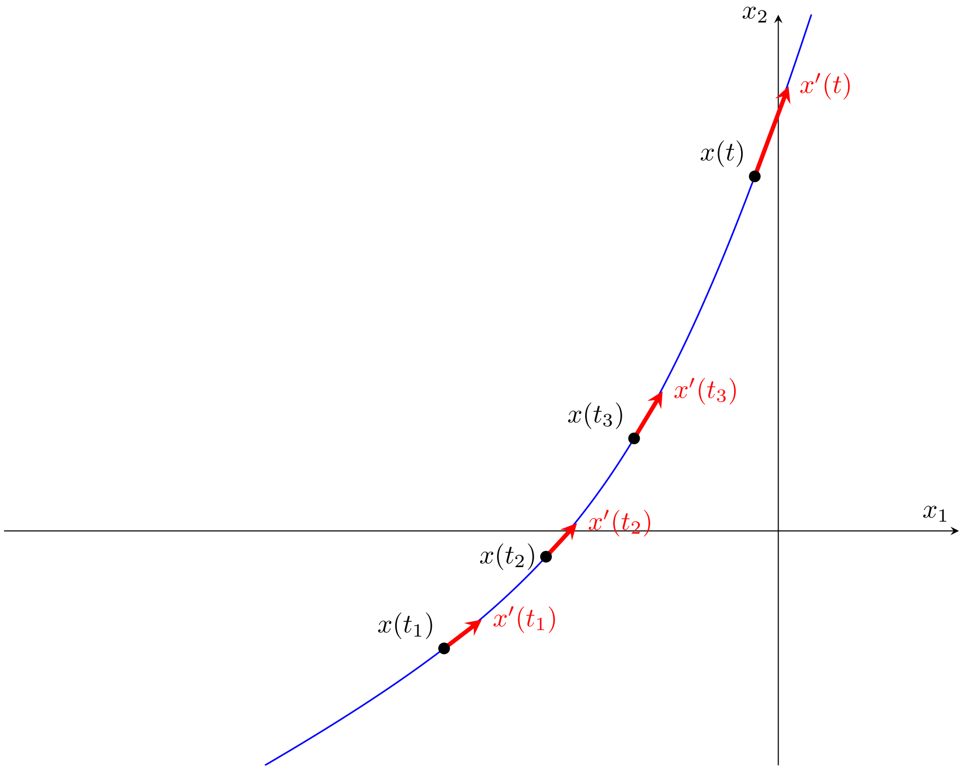
Proposition 20.1 *Let $a \in \mathbb{R}$. The only solutions to*

$$x'(t) = a\,x(t)$$

in $\mathcal{C}^{(1)}(\mathbb{R}, \mathbb{R})$ are given by $x(t) = c\,e^{at}$ for $c \in \mathbb{R}$.

Proof. The proof can be found in Calculus I. □

In two dimensional case, we want to find $x(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}$ such that $x'(t) = Ax(t)$ for a given $A \in \mathbb{R}^{2\times 2}$. Visually, $x(t)$ can be drawn as a curve in $\mathbb{R}^2$. The vector $x'(t)$ describes the tangential vector at each position $x(t)$.

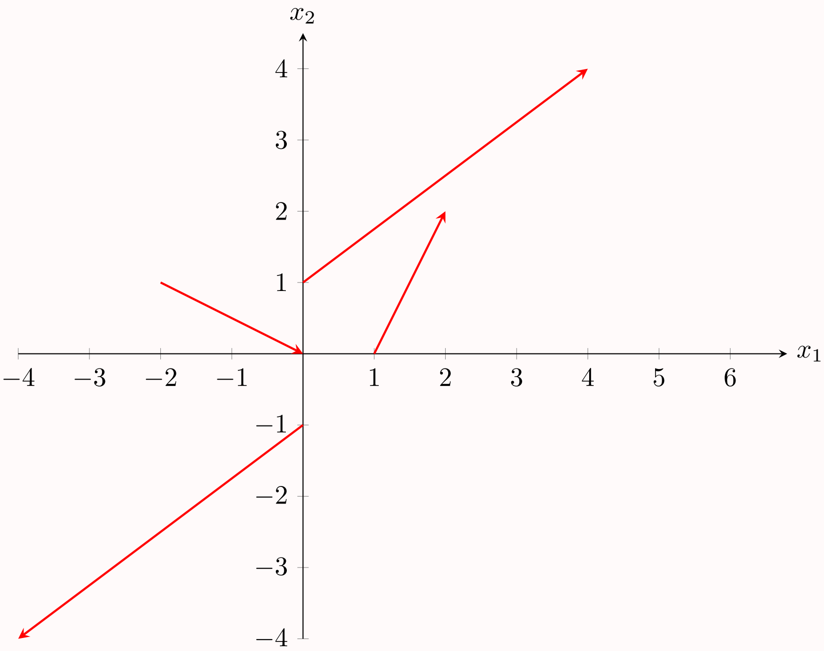


Example 77 1) Let $A = \begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix}$ and consider the dynamical system $x'(t) = Ax(t)$, which can be written as

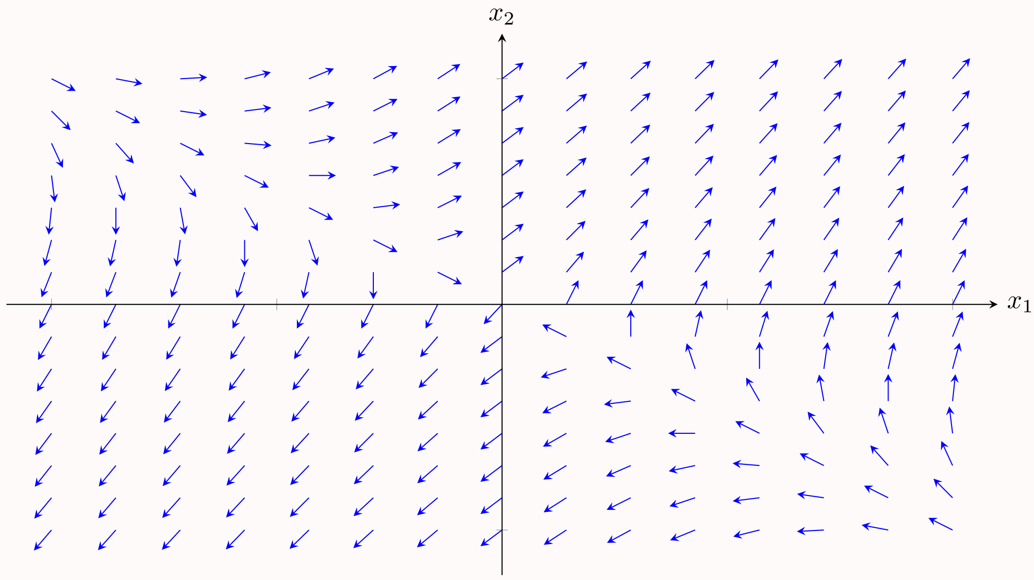
$$\begin{pmatrix} x_1'(t) \\ x_2'(t) \end{pmatrix} = \begin{pmatrix} x_1(t) + 4x_2(t) \\ 2x_1(t) + 3x_3(t) \end{pmatrix}.$$

For each point $x \in \mathbb{R}^2$, we can calculate the corresponding vector $y = Ax$, which is a tangent vector to a particular solution of $x'(t) = Ax(t)$. For example, we have

x	$\begin{pmatrix}1\\0\end{pmatrix}$	$\begin{pmatrix}0\\-1\end{pmatrix}$	$\begin{pmatrix}0\\1\end{pmatrix}$	$\begin{pmatrix}-2\\1\end{pmatrix}$
y	$\begin{pmatrix}1\\2\end{pmatrix}$	$\begin{pmatrix}-4\\-3\end{pmatrix}$	$\begin{pmatrix}4\\3\end{pmatrix}$	$\begin{pmatrix}2\\-1\end{pmatrix}$



If we normalize vector $y = Ax$ for every $x \in \mathbb{R}^2$, we will get a vector field.



We will solve this dynamical system below after considering a more easier example.

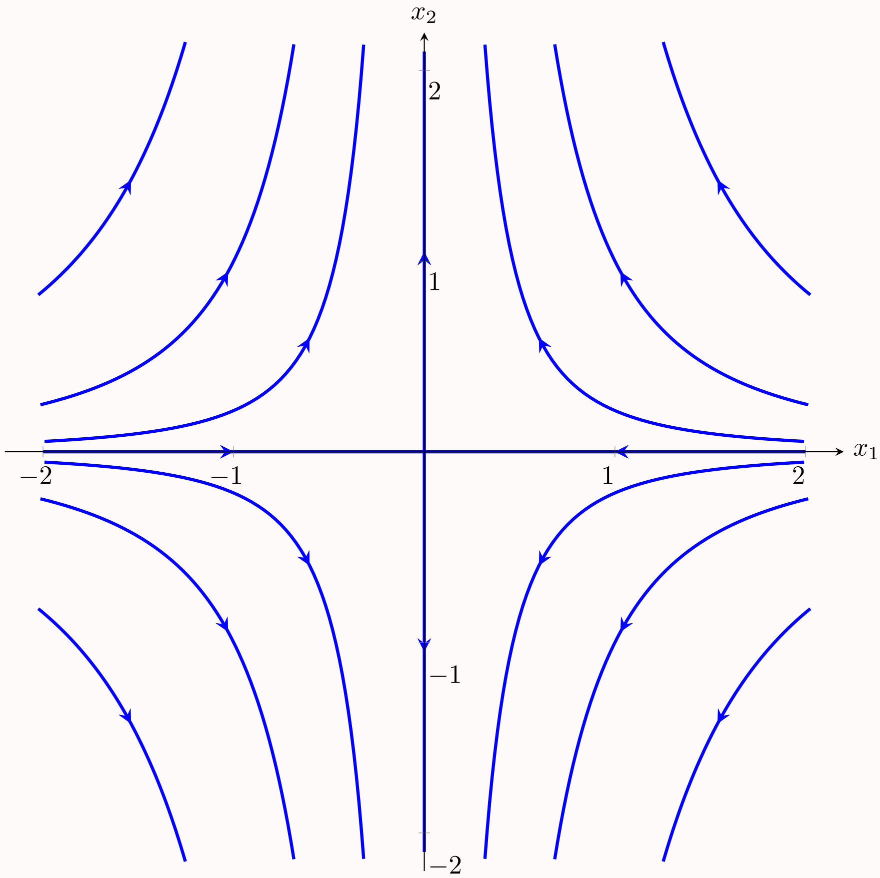
2) Let's try to solve a simple equation first. Consider $A = \begin{pmatrix}-1 & 0\\0 & 2\end{pmatrix}$ and the dynamical system

$$x'(t) = Ax(t) \quad \Leftrightarrow \quad \begin{cases} x_1'(t) &= -x_1(t) \\ x_2'(t) &= 2x_2(t) \end{cases}.$$

x_1 and x_2 are independent of each other and by Proposition 20.1 we get the solutions

$$\begin{cases} x_1(t) & = c_1 e^{-t} \\ x_2(t) & = c_2 e^{2t} \end{cases}, \qquad \text{for } c_1, c_2 \in \mathbb{R}.$$

We have $x(0) = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$, which means that the choices of different c_1, c_2 correspond to choosing the point $x(0)$ on our curve $x(t)$. The graph of different solutions is as bellow.



In general, in order to solve $x'(t) = Ax(t)$, we try to diagonalize A.

Suppose $S^{-1}AS = D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$ for an invertible matrix $S = \begin{pmatrix} | & & | \\ b_1 & \cdots & b_n \\ | & & | \end{pmatrix}$, which means that $(b_1, \dots, b_n)$ is an eigenbasis of A.

Now consider $u(t) = S^{-1}x(t)$ or $x(t) = Su(t)$. Then we have

$$u'(t) = S^{-1}x'(t) = S^{-1}Ax(t) = S^{-1}ASu(t) = Du(t) \quad \Leftrightarrow \quad u'(t) = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} u(t).$$

These equations can be solved as in Example 77 2):

$$\begin{cases} u_1'(t) = \lambda_1 u_1(t) \\ \vdots \\ u_n'(t) = \lambda_n u_n(t) \end{cases} \Leftrightarrow \begin{cases} u_1(t) = c_1 e^{\lambda_1 t} \\ \vdots \\ u_n(t) = c_n e^{\lambda_n t} \end{cases} \Leftrightarrow u(t) = \begin{pmatrix} c_1 e^{\lambda_1 t} \\ \vdots \\ c_n e^{\lambda_n t} \end{pmatrix}, \qquad \text{for } c_1, \ldots, c_n \in \mathbb{R}.$$

The solution to the original dynamical system is

$$x(t) = Su(t) = \begin{pmatrix} | & & | \\ b_1 & \cdots & b_n \\ | & & | \end{pmatrix} \begin{pmatrix} c_1 e^{\lambda_1 t} \\ \vdots \\ c_n e^{\lambda_n t} \end{pmatrix} = c_1 b_1 e^{\lambda_1 t} + \ldots + c_n b_n e^{\lambda_n t}.$$

Example 78 (Example 77 1) continued) Let $A = \begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix}$ and consider the dynamical system $x'(t) = Ax(t)$. The characteristic polynomial is

$$f_A(\lambda) = \det \begin{pmatrix} 1-\lambda & 4 \\ 2 & 3-\lambda \end{pmatrix} = (1-\lambda)(3-\lambda) - 8 = \lambda^2 - 4\lambda - 5 = (\lambda - 5)(\lambda + 1).$$

Hence, the eigenvalues are $\lambda_1 = -1$ and $\lambda_2 = 5$. The corresponding eigenspaces are

$$E_{-1}(A) = \ker(A + I_2) = \ker \begin{pmatrix} 2 & 4 \\ 2 & 4 \end{pmatrix} = \ker \begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix} == \operatorname{span}\left\{ \begin{pmatrix} -2 \\ 1 \end{pmatrix} \right\} = \operatorname{span}\{b_1\},$$
$$E_5(A) = \ker(A - 5I_2) = \ker \begin{pmatrix} -4 & 4 \\ 2 & -2 \end{pmatrix} = \ker \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix} == \operatorname{span}\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\} = \operatorname{span}\{b_2\}.$$

Writing $S = \begin{pmatrix} | & & | \\ b_1 & \cdots & b_n \\ | & & | \end{pmatrix} = \begin{pmatrix} -2 & 1 \\ 1 & 1 \end{pmatrix}$, we get

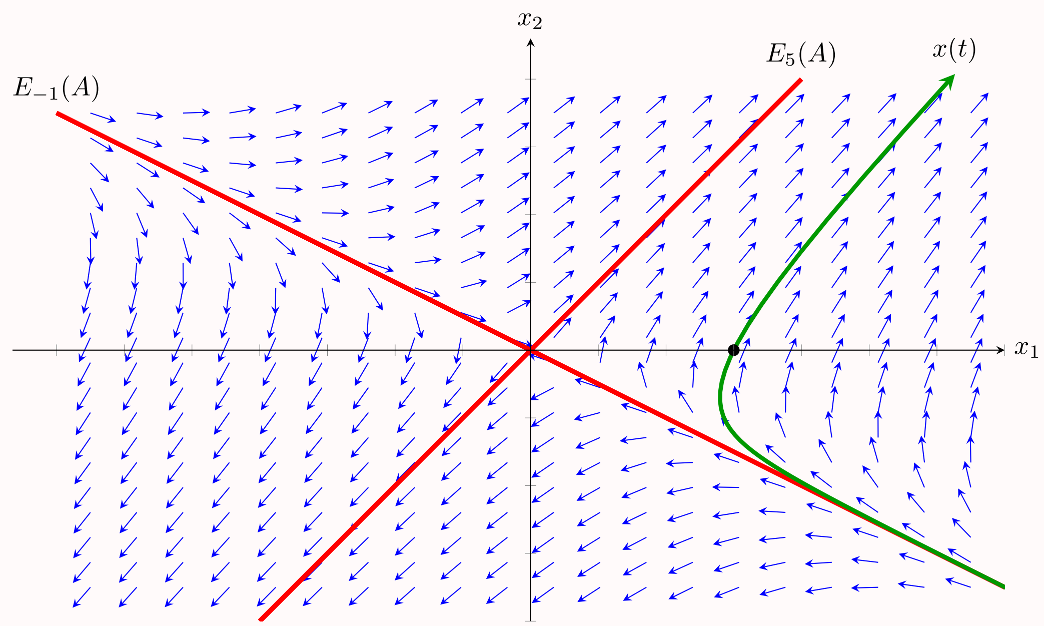
$$S^{-1}AS = D = \begin{pmatrix} -1 & 0 \\ 0 & 5 \end{pmatrix}.$$

The solution of $u'(t) = Du(t)$ is given by $u(t) = \begin{pmatrix} c_1 e^{-t} \\ c_2 e^{5t} \end{pmatrix}$ for some $c_1, c_2 \in \mathbb{R}$.
The solution to $x'(t) = Ax(t)$ is therefore

$$x(t) = Su(t) = \begin{pmatrix} -2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} c_1 e^{-t} \\ c_2 e^{5t} \end{pmatrix} = \begin{pmatrix} -2c_1 e^{-t} + c_2 e^{5t} \\ c_1 e^{-t} + c_2 e^{5t} \end{pmatrix} = c_1 e^{-t} \begin{pmatrix} -2 \\ 1 \end{pmatrix} + c_2 e^{5t} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

For example, for $c_1 = -1$ and $c_2 = 1$, we have $x(0) = \begin{pmatrix} 3 \\ 0 \end{pmatrix}$ and get the solution

$$x(t) = \begin{pmatrix} 2e^{-t} + e^{5t} \\ -e^{-t} + e^{5t} \end{pmatrix} = -e^{-t} \begin{pmatrix} -2 \\ 1 \end{pmatrix} + e^{5t} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$



20.2 Linear differential equations

In the case of continuous dynamical systems we considered several functions and their first derivatives. In this section, we will consider equations involving one function with higher order derivatives.
Recall that the space $\mathcal{C}^\infty(\mathbb{R},\mathbb{R})$, the space of **smooth functions**, denotes the space of all functions $f:\mathbb{R}\to\mathbb{R}$ for which derivatives of all orders exist. This means that for any $n\geq 0$ and $f\in\mathcal{C}^\infty(\mathbb{R},\mathbb{R})$, the n-th derivative $f^{(n)}\in\mathcal{C}^\infty(\mathbb{R},\mathbb{R})$ exists. The space $\mathcal{C}^\infty(\mathbb{R},\mathbb{R})$ is a vector space.

Definition 20.2 i) A **differential operator of order** n is a map $T:\mathcal{C}^\infty(\mathbb{R},\mathbb{R})\to\mathcal{C}^\infty(\mathbb{R},\mathbb{R})$ of the form

$$T(f)=a_0f+a_1f'+a_2f^{(2)}+\cdots+a_nf^{(n)}$$

for some $a_0,a_1,\ldots,a_n\in\mathbb{R}$ with $a_n\neq 0$.
(More precisely this is a "linear differential operator of order n with constant coefficients".)

ii) A **linear differential equation** is an equation of the form $T(f)=g$, where T is a differential operator and $g\in\mathcal{C}^\infty(\mathbb{R},\mathbb{R})$.

iii) A linear differential equation is called **homogeneous** if $g=0$, i.e. if $T(f)=0$.

Remark. 1) Differential operators are linear maps.
2) Solving a homogeneous differential equation $T(f)=0$ means to determine $\ker(T)$.

Example 79 1) Consider the operator

$$\begin{aligned} T=D:\mathcal{C}^\infty(\mathbb{R},\mathbb{R}) &\longrightarrow \mathcal{C}^\infty(\mathbb{R},\mathbb{R}),\\ f &\longmapsto f'\,. \end{aligned}$$

The differential equation $D(f)=g$ for any function $g\in\mathcal{C}^\infty(\mathbb{R},\mathbb{R})$ has solutions of the form $f(t)=\int g(t)\,\mathrm{d}t$, which is any antiderivative of f. In particular, if $g=0$, then the solutions are constant functions $f=c$ for any $c\in\mathbb{R}$. Therefore, $\ker(D)=\operatorname{span}\{1\}$.

2) Consider the differential operator $T=D-a\overset{\text{def}}{=}D-a\cdot\mathrm{id}$. We have

$$T(f)=0\quad\Leftrightarrow\quad f'-af=0\quad\Leftrightarrow\quad f(t)=ce^{at},\quad\text{for } c\in\mathbb{R}.$$

As a result, $\ker(D-a) = \operatorname{span}\{e^{at}\}$. Also, the differential equation $(D-a)f = g$ has solutions of the form $f(t) = e^{at}\int e^{-at}g(t)\,\mathrm{d}t$.

In general, the strategy to find all solutions of linear differential equations is as follows. Let $f_p \in \mathcal{C}^\infty(\mathbb{R},\mathbb{R})$ be any function satisfying $T(f_p) = g$. Then, all of solutions of $T(f) = g$ are of the form $f = f_p + f_h$, where $f_h \in \ker(T)$. In addition, f_p is called a **particular solution** and f_h is any solution to the homogeneous equation $T(f) = 0$. This strategy follows from the following lemma.

Lemma 20.3 *Let $F : V \to W$ be a linear map between two vector spaces V and W. Assume that $F(v) = w$ for a fixed $v \in V$ and $w \in W$. Then the following two statements are equivalent:*
i) $F(x) = w$.
ii) $x = v + u$ for some $u \in \ker(F)$.

Proof. The proof is left as Exercise **??**. □

Example 80 Consider the differential equation $f''(t) + f(t) = 2t$. In this case, $T = D^2 + 1$ and $g(t) = 2t$. A particular solution is $f_p(t) = 2t$. The homogeneous equation $f'' + f = 0$ has solutions of the form
$$f_h(t) = c_1\cos(t) + c_2\sin(t).$$
Therefore, $\ker(T) = \operatorname{span}\{\cos, \sin\}$. The general solution to the above differential equation $T(f) = g$ is
$$f(t) = f_p(t) + f_h(t) = 2t + c_1\cos(t) + c_2\sin(t).$$

The following theorem explains why we only get 2 linearly independent solutions to $T(f) = 0$ in the above example.

Theorem 20.4 *Let $T : \mathcal{C}^\infty(\mathbb{R},\mathbb{R}) \to \mathcal{C}^\infty(\mathbb{R},\mathbb{R})$ be a differential operator of order n. Then we have*
$$\dim(\ker(T)) = n\,.$$

The differential operator $T : f \mapsto a_0 f + a_1 f' + \ldots + a_n f^{(n)}$ can be written as $T = \sum_{i=0}^n a_i D^i$. Note that $D^0 = \mathrm{id}$.

Definition 20.5 Let $T(f) = a_0 f + a_1 f' + \cdots + a_n f^{(n)}$ be a differential operator of order n. The **characteristic polynomial** of T is defined by
$$p_T(x) = \sum_{i=0}^n a_i x^i = a_0 + a_1 x + \cdots + a_n x^n\,.$$

Remark. The characteristic polynomial of a differential operator T has nothing to do with the characteristic polynomials in the eigenvalue theory. Also, notice that $p_T(D) = T$.
In the following statements, T always denotes a differential operator.

Proposition 20.6 *(i) The function $e^{\lambda t}$ is an eigenvector of T with eigenvalue $p_T(\lambda)$.*
(ii) We have $e^{\lambda t} \in \ker(T)$ if and only if $p_T(\lambda) = 0$.

Proof. (i) We have $D(e^{\lambda t}) = (e^{\lambda t})' = \lambda e^{\lambda t}$. Consequently, we get $D^i(e^{\lambda t}) = \lambda^i e^{\lambda t}$ for any $i \in \mathbb{N}$. Then we have
$$T(e^{\lambda t}) = \sum_{i=0}^n a_i D^i(e^{\lambda t}) = \sum_{i=0}^n a_i \lambda^i (e^{\lambda t}) = p_T(\lambda)e^{\lambda t}.$$
(ii) We have
$$T(e^{\lambda t}) = 0 \quad\Leftrightarrow\quad p_T(\lambda)e^{\lambda t} = 0 \quad\Leftrightarrow\quad p_T(\lambda) = 0.$$
□

Corollary 20.7 *Let T be a differential operator of order n.*
(i) If $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ are distinct, then $e^{\lambda_1 t}, \dots, e^{\lambda_n t}$ are linearly independent.
(ii) If p_T has n distinct zeroes $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ then $(e^{\lambda_1 t}, \dots, e^{\lambda_n t})$ is a basis of $\ker(T)$.

Proof. (i) From Proposition 20.6, $e^{\lambda_1 t}, \dots, e^{\lambda_n t}$ are eigenvectors of D with distinct eigenvalues $\lambda_1, \dots, \lambda_n$. By Theorem 18.6, we conclude that $e^{\lambda_1 t}, \dots, e^{\lambda_n t}$ are linearly independent.
(ii) Since $p_T(\lambda_i) = 0$ for $1 \leq i \leq n$, we have $e^{\lambda_1 t}, \dots, e^{\lambda_n t} \in \ker(T)$ by Proposition 20.6. Also, $e^{\lambda_1 t}, \dots, e^{\lambda_n t}$ are linearly independent since $\lambda_1, \dots, \lambda_n$ are distinct. Since we have $\dim(\ker(T)) = n$ by Theorem 20.4, $(e^{\lambda_1 t}, \dots, e^{\lambda_n t})$ is a basis of $\ker(T)$. $\square$

Example 81 Consider the differential equation $f^{(3)} - 3f'' - f' + 3f = 0$. Here, the differential operator is $T = D^3 - 3D^2 - D + 3$ and the characteristic polynomial is

$$p_T(x) = x^3 - 3x^2 - x + 3 = (x+1)(x-1)(x-3).$$

The zeros of this polynomial are $-1, 1, 3$. Therefore, $\ker(T) = \operatorname{span}\{e^{-t}, e^t, e^{3t}\}$ and the general solution to the above differential equation is $f(t) = c_1 e^{-t} + c_2 e^t + c_3 e^{3t}$.

In fact, every real polynomial p factors uniquely as

$$p = c \cdot p_1^{m_1} \cdot \dots \cdot p_r^{m_r},$$

where $c \in \mathbb{R}$, with maximal $m_1, \dots, m_r \geq 1$, and the polynomials p_j for $1 \leq j \leq r$ are of the form

$$p_j(t) = \begin{cases} t - \lambda, & \lambda \in \mathbb{R} \\ (t-a)^2 + b^2, & a, b \in \mathbb{R}, b \neq 0 \end{cases}.$$

How does $\ker(T)$ look like if p_T has a factor $(t-\lambda)^m$ for $m \geq 2$ or $(t-a)^2 + b^2$? First of all, we need the following result.

Lemma 20.8 *For two differential operators T_1 and T_2 we have $T_1 \circ T_2 = T_2 \circ T_1$.*

Proof. Let $T_1 = \sum_{i=0}^n a_i D^i$ and $T_2 = \sum_{j=0}^m b_j D^j$. For any $f \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R})$, we have

$$T_1 \circ T_2(f) = \sum_{i=0}^n a_i D^i \left(\sum_{j=0}^m b_j D^j(f) \right) = \sum_{i=0}^n \sum_{j=0}^m a_i b_j D^i \left(D^j(f) \right).$$

Since $D^i \left(D^j(f) \right) = D^{i+j}(f) = D^j \left(D^i(f) \right)$, we have

$$T_1 \circ T_2(f) = \sum_{i=0}^n \sum_{j=0}^m a_i b_j D^j \left(D^i(f) \right) = \sum_{j=0}^m \sum_{i=0}^n a_i b_j D^j \left(D^i(f) \right) = \sum_{j=0}^m b_j D^j \left(\sum_{i=0}^n a_i D^i(f) \right) = T_2 \circ T_1(f).$$

$\square$

This means that if the characteristic polynomial $p_T(x)$ of a differential operator T has, for example, the factorization

$$p_T(x) = (x-2)^2(x-3),$$

we have $T = (D-2)^2 \circ (D-3) = (D-3) \circ (D-2)^2$, where $(D-2)^2 = (D-2) \circ (D-2)$.

Example 82 Consider the differential operator $T = D^2 - 2\lambda D + \lambda^2 = (D-\lambda)^2$. We claim that $(e^{\lambda t}, te^{\lambda t})$ is a basis of $\ker(T)$. Indeed, we have first of all, $e^{\lambda t} \in \ker(D - \lambda)$ and hence, $e^{\lambda t} \in \ker(T)$. In addition,

$$(D-\lambda)(te^{\lambda t}) = D(te^{\lambda t}) - \lambda te^{\lambda t} = e^{\lambda t} + \lambda te^{\lambda t} - \lambda te^{\lambda t} = e^{\lambda t}$$
$$\Rightarrow \quad T(te^{\lambda t}) = (D-\lambda)^2(te^{\lambda t}) = (D-\lambda)(e^{\lambda t}) = 0.$$

As a result, $te^{\lambda t} \in \ker(T)$. Notice that $e^{\lambda t} \in \ker(D-\lambda)$ but $te^{\lambda t} \notin \ker(D-\lambda)$. Therefore, $e^{\lambda t}$ and $te^{\lambda t}$ are linearly independent. Since $\dim(\ker(T)) = 2$ by Theorem 20.4, the claim is proven.

More generally, we have the following result.

Theorem 20.9 *Let T be a differential operator with characteristic polynomial*

$$p_T(x) = (x-\lambda_1)^{m_1} \cdots (x-\lambda_r)^{m_r}$$

where $\lambda_1, \ldots, \lambda_r \in \mathbb{R}$ with $\lambda_i \neq \lambda_j$ for $i \neq j$.
Then $B = B_1 \cup \cdots \cup B_r$ is a basis of $\ker(T)$, where we have for $1 \leq j \leq r$

$$B_j = (e^{\lambda_j t}, te^{\lambda_j t}, \ldots, t^{m_j-1}e^{\lambda_j t}).$$

Proof. We have for any $1 \leq j \leq r$ and $1 \leq n \leq m_j - 1$,

$$(D-\lambda_j)(t^n e^{\lambda_j t}) = nt^{n-1}e^{\lambda_j t} + \lambda_j t^n e^{\lambda_j t} - \lambda_j t^n e^{\lambda_j t} = nt^{n-1}e^{\lambda_j t}.$$

Since $(D-\lambda_j)(e^{\lambda_j t}) = 0$, we have $(D-\lambda_j)^{n+1}(t^n e^{\lambda_j t}) = 0$. As a result, we see that

$$e^{\lambda_j t}, te^{\lambda_j t}, \ldots, t^{m_j-1}e^{\lambda_j t} \in \ker(D-\lambda_j)^{m_j} \subset \ker(T).$$

The linear independence of these elements comes from the linear independence of $1, t, t^2, \ldots, t^{m_j-1}$. Since $\dim(\ker(D-\lambda_j)^{m_j}) = m_j$ by Theorem 20.4, $(e^{\lambda_j t}, te^{\lambda_j t}, \ldots, t^{m_j-1}e^{\lambda_j t})$ is a basis of $\ker(D-\lambda_j)^{m_j}$. Also, it is clear that $\ker(D-\lambda_i)^{m_i} \cap \ker(D-\lambda_j)^{m_j} = \{0\}$ for $i \neq j$. In total, we have $m_1 + \ldots + m_r = n$ linearly independent elements in $\ker(T)$. Since $\dim(\ker(T)) = n$ by Theorem 20.4, $B = B_1 \cup \ldots \cup B_r$ as defined above is a basis of $\ker(T)$. $\square$

Example 83 Consider the differential equation $T(f) = f^{(3)} - 3f' + 2f = 0$. Here, we have $T = D^3 - 3D + 2 = (D-1)^2(D+2)$. By Theorem 20.9, (e^t, te^t, e^{-2t}) is a basis of $\ker(T)$. All solutions of $T(f) = 0$ are given by

$$f(t) = c_1 e^t + c_2 te^t + c_3 e^{-2t},$$

for $c_1, c_2, c_3 \in \mathbb{R}$.

Now, what if p_T contains a factor of the form $(x-a)^2 + b^2$? We can do the same as before by using complex number. The result is that we get two linearly independent elements $e^{at}\cos(bt)$ and $e^{at}\sin(bt)$. More generally, we have the following result.

Theorem 20.10 *Let T be a differential operator. If $p_T(x)$ contains a factor $((x-a)^2+b^2)^m$, then*

$$\{e^{at}\cos(bt), e^{at}\sin(bt), te^{at}\cos(bt), te^{at}\sin(bt), \ldots, t^{m-1}e^{at}\cos(bt), t^{m-1}e^{at}\sin(bt)\}$$

are $2m$ linearly independent elements in $\ker(T)$.

Proof. We have

$$((x-a)^2+b^2)^m = (x-\lambda)^m(x-\overline{\lambda})^m,$$

where $\lambda = a + bi$ and $\overline{\lambda} = a - bi$.
With the same arguments as before, $e^{\lambda t}, te^{\lambda t}, \ldots, t^{m-1}e^{\lambda t}, e^{\overline{\lambda}t}, te^{\overline{\lambda}t}, \ldots, t^{m-1}e^{\overline{\lambda}t}$ are linearly independent elements in $\ker(T)$. In addition, using Euler identity $e^{ix} = \cos(x) + i\sin(x)$, we have

$$e^{\lambda t} = e^{(a+bi)t} = e^{at}e^{ibt} = e^{at}(\cos(bt) + i\sin(bt)),$$
$$e^{\overline{\lambda}t} = e^{(a-bi)t} = e^{at}e^{-ibt} = e^{at}(\cos(bt) - i\sin(bt)).$$

Then, we have for any $0 \le n \le m-1$,

$$t^n e^{at}\cos(bt) = \frac{1}{2}t^n(e^{\lambda t} + e^{\overline{\lambda}t}),$$
$$t^n e^{at}\sin(bt) = \frac{1}{2i}t^n(e^{\lambda t} - e^{\overline{\lambda}t}).$$

Hence, $e^{at}\cos(bt), e^{at}\sin(bt), te^{at}\cos(bt), te^{at}\sin(bt), \ldots, t^{m-1}e^{at}\cos(bt), t^{m-1}e^{at}\sin(bt)$ are also elements in $\ker(T)$. Due to the linear independence of $e^{\lambda t}, te^{\lambda t}, \ldots, t^{m-1}e^{\lambda t}, e^{\overline{\lambda}t}, te^{\overline{\lambda}t}, \ldots, t^{m-1}e^{\overline{\lambda}t}$, we have for $n \neq k$,

$$\operatorname{span}\{t^n e^{\lambda}, t^n e^{\overline{\lambda}}\} \cap \operatorname{span}\{t^k e^{\lambda}, t^k e^{\overline{\lambda}}\} = \{0\}.$$

Since $t^n e^{at}\cos(bt), t^n e^{at}\sin(bt) \in \operatorname{span}\{t^n e^{\lambda}, t^n e^{\overline{\lambda}}\}$, we only need to prove the linear independence between them in order to prove the linear independence of

$$e^{at}\cos(bt), e^{at}\sin(bt), te^{at}\cos(bt), te^{at}\sin(bt), \ldots, t^{m-1}e^{at}\cos(bt), t^{m-1}e^{at}\sin(bt).$$

We have

$$\begin{aligned} c_1 t^n e^{at}\cos(bt) + c_2 t^n e^{at}\sin(bt) = 0, \quad \forall t \in \mathbb{R} \quad &\Rightarrow \quad c_1 t^n\cos(bt) + c_2 t^n \sin(bt) = 0, \quad \forall t \in \mathbb{R} \\ &\Rightarrow \quad \begin{cases} c_1\left(\frac{\pi}{b}\right)^n \cos(\pi) + c_2\left(\frac{\pi}{b}\right)^n \sin(\pi) = 0 \\ c_1\left(\frac{\pi}{2b}\right)^n \cos\left(\frac{\pi}{2}\right) + c_2\left(\frac{\pi}{2b}\right)^n \sin\left(\frac{\pi}{2}\right) = 0 \end{cases} \\ &\Rightarrow \quad \begin{cases} c_1 = 0 \\ c_2 = 0 \end{cases}. \end{aligned}$$

As a result, $e^{at}\cos(bt), e^{at}\sin(bt), te^{at}\cos(bt), te^{at}\sin(bt), \ldots, t^{m-1}e^{at}\cos(bt), t^{m-1}e^{at}\sin(bt)$ are linearly independent. $\square$

Theorem 20.9 and 20.10 gives us n linearly independent elements in $\ker(T)$ for a differential operator T of order n. Now, we want to come back to prove Theorem 20.4. First, we need the following lemma.

Lemma 20.11 *Let $F : U \to V$ and $G : V \to W$ be surjective linear maps between vector spaces U, V, W, such that $\ker(F)$ and $\ker(G)$ are finitely generated. Then we have*

$$\dim(\ker(G \circ F)) = \dim(\ker(F)) + \dim(\ker(G)).$$

Proof. Let $(v_1, \ldots, v_m)$ be a basis of $\ker(G)$. Since F is surjective, we can find $u_1, \ldots, u_m \in U$ such that $F(u_j) = v_j$ for $j = 1, \ldots, m$. Let $(x_1, \ldots, x_l)$ be a basis of $\ker(F)$. We claim that $(x_1, \ldots, x_l, u_1, \ldots, u_m)$ is a basis of $\ker(G \circ F)$.
If $y \in \ker(G \circ F)$ then $F(y) \in \ker(G)$ and therefore, $F(y) = \sum_{i=1}^m \lambda_i v_i$. Set $w = \sum_{i=1}^m \lambda_i u_i$. Then we have

$$\begin{aligned} & F(w) = \sum_{i=1}^m \lambda_i F(u_i) = \sum_{i=1}^m \lambda_i v_i = F(y) \\ \Rightarrow \quad & F(y-w) = 0 \quad \Rightarrow \quad y - w \in \ker(F) \\ \Rightarrow \quad & y - w = \sum_{j=1}^l \mu_j x_j \quad \Rightarrow \quad y = \sum_{j=1}^l \mu_j x_j + \sum_{i=1}^m \lambda_i u_i. \end{aligned}$$

Therefore, $\ker(G \circ F) = \operatorname{span}\{x_1, \ldots, x_l, u_1, \ldots, u_m\}$. What remains is the linear independence of $x_1, \ldots, x_l, u_1, \ldots, u_m$. Indeed, we have

$$\sum_{j=1}^l \mu_j x_j + \sum_{i=1}^m \lambda_i u_i = 0 \quad \Rightarrow \quad \sum_{j=1}^l \mu_j F(x_j) + \sum_{i=1}^m \lambda_i F(u_i) = 0 \quad \Rightarrow \quad \sum_{i=1}^m \lambda_i v_i = 0.$$

Since $v_1, \ldots, v_m$ are linearly independent, we get $\lambda_1 = \cdots = \lambda_m = 0$. In turn, we have $\mu_1 = \cdots = \mu_l = 0$ by the linear independence of $x_1, \ldots, x_l$. $\square$

Now we are in the position to prove Theorem 20.4.

Proof of Theorem 20.4. Let $T = (D-\lambda_1)(D-\lambda_2)\dots(D-\lambda_n)$ be a differential operator of order n for $\lambda_1, \dots, \lambda_n \in \mathbb{R}$. The differential equation $(D-\lambda)(f) = g$ always has a solution

$$f(t) = e^{\lambda t} \int e^{-\lambda t} g(t)\, \mathrm{d}t\,.$$

Therefore, the differential operator $(D-\lambda) : \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}) \to \mathcal{C}^\infty(\mathbb{R}, \mathbb{R})$ is a surjective linear map with $\dim(\ker(D-\lambda)) = 1$ since $\ker(D-\lambda) = \operatorname{span}\{e^{\lambda t}\}$. The theorem can now be proven by induction on n together with Lemma 20.11.

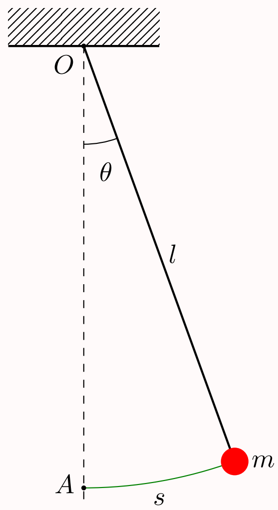
1) Base step ($n = 1$): $T = D - \lambda_1$ has $\dim(\ker(T)) = 1$.
2) Induction step: Assume that

$$\dim\left(\ker\left((D-\lambda_1)\dots(D-\lambda_m)\right)\right) = m, \quad \text{for } m < n.$$

Now consider $T = G \circ F$ where $G = (D-\lambda_1)\dots(D-\lambda_{n-1})$ and $F = D - \lambda_n$. By the base step and the induction hypothesis, we have $\dim(\ker(F)) = 1$ and $\dim(\ker(G)) = n-1$. By Lemma 20.11, we have

$$\dim(\ker(T)) = \dim(\ker(G)) + \dim(\ker(F)) = n - 1 + 1 = n.$$

This proves Theorem 20.4 for the case $T = (D-\lambda_1)\dots(D-\lambda_n)$. In general, if T contains a factor $((x-a)^2 + b^2)$, we use the same argument as before by using complex numbers and $(x-a)^2 + b^2 = (x-\lambda)(x-\overline{\lambda})$ with $\lambda = a + bi$ and $\overline{\lambda} = a - bi$. □



Example 84 Assume that we want to describe the movement of a pendulum of length l with mass m. Denote by $\theta(t)$ the angle at time t. The position of th mass on the circular path at time t is given by $s(t) = l \cdot \theta(t)$. The acceleration at time t is therefore

$$a(t) = \frac{\mathrm{d}^2}{\mathrm{d}t^2} s(t) = l\theta''(t).$$

By Newton's second law, we have the force

$$F(t) = ma(t) = ml\theta''(t). \tag{$*$}$$

On the other hand the force acting on the pendulum is given by gravity. The part acting in the direction of the pendulum is given by $-\sin(\theta)mg$. Therefore,

$$F(t) = -\sin(\theta(t))mg. \tag{$**$}$$

Combining $(*)$ and $(**)$, we get

$$ml\theta''(t) = -\sin(\theta)mg$$

$$\theta''(t) + \frac{g}{l}\sin(\theta(t)) = 0.$$

For small θ, one has $\theta(t) \approx \sin(\theta(t))$. Therefore, for small θ, we obtain the differential equation

$$\theta''(t) + \frac{g}{l}\theta(t) = 0.$$

This is a homogeneous differential equation with the differential operator

$$T = D^2 + \frac{g}{l} = (D-a)^2 + b^2,$$

with $a = 0$ and $b = \sqrt{\frac{g}{l}}$. The solution has the form

$$\theta(t) = c_1 \cos\left(\sqrt{\frac{g}{l}}t\right) + c_2 \sin\left(\sqrt{\frac{g}{l}}t\right)).$$

If the pendulum is at rest at $t = 0$ with the angle $\theta(0) = \theta_0$, then the solution is

$$\theta(t) = \theta_0 \cos\left(\sqrt{\frac{g}{l}}t\right).$$

Bibliography

[B] O. Bretscher: *Linear Algebra with Applications*, 4th edition, Pearson 2009.
[G] H. Goldstein, C. Poole, J. Safko: *Classical Mechanics*, 3th edition, Addison-Wesley 2001.