Imperial College London

LECTURE NOTES

IMPERIAL COLLEGE LONDON

DEPARTMENT OF COMPUTING

Mathematics I: Foundation:

Part II: Linear Algebra

Abbas Edalat

Version: May 17, 2021

Chapter 1

Linear Algebra

This chapter is largely based on the lecture notes and books by Drumm and Weil (2001); Strang (2003); Hogben (2013); Liesen and Mehrmann (2015) as well as Pavel Grinfeld's Linear Algebra series[1]. Another excellent source is Gilbert Strang's Linear Algebra course at MIT[2].

Linear algebra is the study of vectors. The vectors we know from school are called "geometric vectors". In general, however, vectors are special objects that can be added together and multiplied by scalars to produce another object of the same kind. Any object that satisfies these two properties can be considered a vector. Here are some examples of such vector objects:

1. Geometric vectors. This example of a vector may be familiar from High School. Geometric vectors are directed segments, which can be drawn, see Fig. 1.1. Two vectors $\vec{x}, \vec{y}$ can be added, such that $\vec{x} + \vec{y} = \vec{z}$ is another geometric vector. Furthermore, $\lambda\vec{x}, \lambda \in \mathbb{R}$ is also a geometric vector. In fact, it is the original vector scaled by λ. Therefore, geometric vectors are instances of the vector concepts introduced above.

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

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

$$\boldsymbol{a} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \in \mathbb{R}^3 \tag{1.1}$$

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

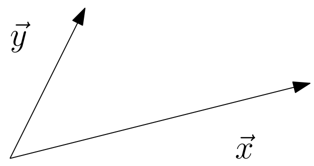


Figure 1.1: Example of two geometric vectors in two dimensions.

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

4. Audio signals are vectors, too. Audio signals are represented as a series of numbers. We can add audio signals together, and their sum is a new audio signals. If we scale an audio signal, we also obtain an audio signal. Therefore, audio signals are a type of vector, too.

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

Practical Applications of Linear Algebra

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

- Ranking of web pages (web search)
- Linear programming (optimziation)
- Error correcting codes (e.g., in DVDs)
- Machine learning, e.g., least square method (linear regression), principle component analysis
- Projections, rotations, scaling (computer graphics)
- Quantum computation
- En/Decryption algorithms (cryptography)
- State estimation and optimal control (e.g., in robotics and dynamical systems)
- Numerics: determine whether a computation is numerically stable (e.g., in large-scale data-analytics systems, optimization, machine learning)
- Find out whether a function is convex (e.g., in Optimization)

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

1.1 Linear Equation Systems

Systems of linear equations play a central part of linear algebra. Many problems can be formulated as systems of linear equations, and linear algebra gives us the tools for solving them.

Example

A company produces products $N_1, \dots, N_n$ for which resources $R_1, \dots, R_m$ are required. To produce a unit of product N_j, a_{ij} units of resource R_i are needed, where $i = 1, \dots, m$ and $j = 1, \dots, n$.

The objective is to find an optimal production plan, i.e., a plan how many units x_j of product N_j should be produced if a total of b_i units of resource R_i are available and (ideally) no resources are left over.

If we produce $x_1, \dots, x_n$ units of the corresponding products, we need a total of

$$a_{i1}x_1 + \cdots + a_{in}x_n \tag{1.2}$$

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

$$\begin{array}{c} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{array} = \begin{array}{c} b_1 \\ \vdots \\ b_m \end{array}, \tag{1.3}$$

where $a_{ij} \in \mathbb{R}$ and $b_i \in \mathbb{R}$.

Equation (1.3) is the general form of a **linear equation system**, and $x_1, \dots, x_n$ are the **unknowns** of this linear equation system. Every n-tuple $(x_1, \dots, x_n) \in \mathbb{R}^n$ that satisfies (1.3) is a **solution** of the linear equation system.

Example

The linear equation system

$$\begin{array}{rcrcrcrl} x_1 & + & x_2 & + & x_3 & = & 3 & (1) \\ x_1 & - & x_2 & + & 2x_3 & = & 2 & (2) \\ 2x_1 & & & + & 3x_3 & = & 1 & (3) \end{array} \tag{1.4}$$

has **no solution:** Adding the first two equations yields $(1)+(2) = 2x_1 + 3x_3 = 5$, which contradicts the third equation (3).

Let us have a look at the linear equation system

$$\begin{array}{rcrcrcrl} x_1 & + & x_2 & + & x_3 & = & 3 & (1) \\ x_1 & - & x_2 & + & 2x_3 & = & 2 & (2) \\ & & x_2 & + & x_3 & = & 2 & (3) \end{array}. \tag{1.5}$$

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

As a third example, we consider

$$\begin{array}{ccccccc} x_1 & + & x_2 & + & x_3 & = & 3 & (1) \\ x_1 & - & x_2 & + & 2x_3 & = & 2 & (2) \\ 2x_1 & & & + & 3x_3 & = & 5 & (3) \end{array}. \tag{1.6}$$

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

$$\left(\frac{5}{2} - \frac{3}{2}a, \frac{1}{2} + \frac{1}{2}a, a\right), \quad a \in \mathbb{R} \tag{1.7}$$

is a solution to the linear equation system, i.e., we obtain a solution set that contains **infinitely many** solutions.

In general, for a real-valued linear equation system we obtain either no, exactly one or infinitely many solutions.
For a systematic approach to solving linear equation systems, we will introduce a useful compact notation. We will write the linear equation system from (1.3) in the following form:

$$x_1 \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \Leftrightarrow \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}. \tag{1.8}$$

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

1.2 Matrices

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

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad a_{ij} \in \mathbb{R}. \tag{1.9}$$

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

$\mathbb{R}^{m \times n}$ is the set of all real-valued (m, n)-matrices. $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ can be equivalently represented as $\boldsymbol{A} \in \mathbb{R}^{mn}$.

1.2.1 Matrix Multiplication

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

$$c_{ij} = \sum_{l=1}^{n} a_{il} b_{lj}, \qquad i = 1,\ldots,m, \quad j = 1,\ldots,k. \tag{1.10}$$

This means, to compute element c_{ij} we multiply the elements of the ith row of $\boldsymbol{A}$ with the jth column of $\boldsymbol{B}$[4] and sum them up.[5]

Remark 1
Matrices can only be multiplied if their "neighbouring" dimensions match. For instance, an $n\times k$-matrix $\boldsymbol{A}$ can be multiplied with a $k\times m$-matrix $\boldsymbol{B}$, but only from the left side:

$$\underbrace{\boldsymbol{A}}_{n\times k} \underbrace{\boldsymbol{B}}_{k\times m} = \underbrace{\boldsymbol{C}}_{n\times m} \tag{1.11}$$

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

Remark 2
Note that matrix multiplication is **not** *defined as an element-wise operation on matrix elements, i.e., $c_{ij} \neq a_{ij} b_{ij}$ (even if the size of $\boldsymbol{A}, \boldsymbol{B}$ was chosen appropriately).*[6]

Example

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

$$\boldsymbol{AB} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 2 & 5 \end{bmatrix} \in \mathbb{R}^{2\times 2}, \tag{1.12}$$

$$\boldsymbol{BA} = \begin{bmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 4 & 2 \\ -2 & 0 & 2 \\ 3 & 2 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 3}. \tag{1.13}$$

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

[4] They are both of length k, such that we can compute $a_{il} b_{lj}$ for $l = 1,\ldots,n$.
[5] Later, we will call this the **scalar product** or **dot product** of the corresponding row and column.
[6] This kind of element-wise multiplication appears often in computer science where we multiply (multi-dimensional) arrays with each other.

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

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix} \in \mathbb{R}^{n\times n}. \tag{1.14}$$

With this, $A \cdot I_n = A = I_n A$ for all $A \in \mathbb{R}^{n\times n}$. We say, the identity matrix is the neutral element *with respect to matrix multiplication.*[7]

Properties

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

1.2.2 Inverse and Transpose

Definition 3 (Inverse)
For a square matrix[8] *$A \in \mathbb{R}^{n\times n}$ a matrix $B \in \mathbb{R}^{n\times n}$ with $AB = I_n = BA$ is called* **inverse** *and denoted by A^{-1}.*

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

Example
The matrix

$$A = \begin{bmatrix} 5 & 3 \\ 3 & 2 \end{bmatrix}$$

is invertible with

$$A^{-1} = \begin{bmatrix} 2 & -3 \\ -3 & 5 \end{bmatrix}$$

Exercise 1
Show that the inverse of an invertible matrix is unique.

[7] If $A \in \mathbb{R}^{m\times n}$ then I_n is only a right neutral element, such that $AI_n = A$. The corresponding left-neutral element would be I_m since $I_m A = A$.
[8] The number of columns equals the number of rows.

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

For a square matrix $\boldsymbol{A}^\top$ is the matrix we obtain when we "mirror" $\boldsymbol{A}$ on its main diagonal.[9] In general, $\boldsymbol{A}^\top$ can be obtained by writing the columns of $\boldsymbol{A}$ as the rows of $\boldsymbol{A}^\top$.

Example
For

$$\boldsymbol{A} = \begin{bmatrix} 5 & 4 & -1 \\ 3 & 2 & 0 \end{bmatrix}$$

we have

$$\boldsymbol{A}^\top = \begin{bmatrix} 5 & 3 \\ 4 & 2 \\ -1 & 0 \end{bmatrix}$$

Remark 3

(i) $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$

(ii) $(\boldsymbol{A}^\top)^\top = \boldsymbol{A}$

(iii) $(\boldsymbol{A}+\boldsymbol{B})^\top = \boldsymbol{A}^\top + \boldsymbol{B}^\top$

(iv) $(\boldsymbol{AB})^\top = \boldsymbol{B}^\top\boldsymbol{A}^\top$

(v) If $\boldsymbol{A}$ is invertible, $(\boldsymbol{A}^{-1})^\top = (\boldsymbol{A}^\top)^{-1}$

(vi) Note: $(\boldsymbol{A}+\boldsymbol{B})^{-1}$ may not exist even if $\boldsymbol{A}^{-1}$ and $\boldsymbol{B}^{-1}$ do both exist (why?). On the other hand, $(\boldsymbol{A}+\boldsymbol{B})^{-1}$ may exist even if neither $\boldsymbol{A}^{-1}$ nor $\boldsymbol{B}^{-1}$ exist (why?).

Exercise 2
Prove the statements in Remark 3(i)-(v).

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

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}. \tag{1.15}$$

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

1.2.3 Multiplication by a Scalar

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

- Distributivity:
  $(\lambda+\beta)\boldsymbol{C} = \lambda\boldsymbol{C} + \beta\boldsymbol{C}, \quad \boldsymbol{C} \in \mathbb{R}^{m\times n}$
  $\lambda(\boldsymbol{B}+\boldsymbol{C}) = \lambda\boldsymbol{B} + \lambda\boldsymbol{C}, \quad \boldsymbol{B}, \boldsymbol{C} \in \mathbb{R}^{m\times n}$

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

- $(\lambda\boldsymbol{C})^\top = \boldsymbol{C}^\top\lambda^\top = \boldsymbol{C}^\top\lambda = \lambda\boldsymbol{C}^\top$ since $\lambda = \lambda^\top$ for all $\lambda \in \mathbb{R}$.

1.2.4 Compact Representations of Linear Equation Systems

If we consider a linear equation system

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

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

$$\begin{bmatrix} 2 & 3 & 5 \\ 4 & -2 & -7 \\ 9 & 5 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 8 \\ 2 \end{bmatrix}. \tag{1.16}$$

Note that x_1 scales the first column, x_2 the second one, and x_3 the third one.
Generally, linear equation systems can be compactly represented in their matrix form as $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, see (1.3), and the product $\boldsymbol{A}\boldsymbol{x}$ is a (linear) combination of the columns of $\boldsymbol{A}$.[10]

1.3 Solve Linear Equations by Gaussian Elimination!

In (1.3), we have introduced the general form of an equation system, i.e.,

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ \vdots \qquad\qquad & \quad \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned}, \tag{1.17}$$

where $a_{ij} \in \mathbb{R}$ and $b_i \in \mathbb{R}$ are known constants and x_j are unknowns, $i = 1, \ldots, m$, $j = 1, \ldots, n$. Thus far, we have introduced matrices as a compact way of formulating

[10] We will discuss linear combinations in Section 1.4.

linear equation systems, i.e., such that we can write $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, see (1.8). If $\boldsymbol{b} = \boldsymbol{0}$, then we say that the linear equation system is **homogeneous**. If $\boldsymbol{b} \neq \boldsymbol{0}$, we say the linear equation system is **inhomogeneous**. Moreover, we defined basic matrix operations, such as addition and multiplication of matrices. In the following, we will introduce a constructive and systematic way of solving linear equation systems. Before doing this, we introduce the **augmented matrix** $\left[\boldsymbol{A}\,|\,\boldsymbol{b}\right]$ of the linear equation system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$. This augmented matrix will turn out to be useful when solving linear equation systems.

1.3.1 Example: Solving a Simple Linear Equation System

Now we are turning towards solving linear equation systems. Before doing this in a systematic way using Gaussian elimination, let us have a look at an example. Consider the following linear equation system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$:

$$\begin{bmatrix} 1 & 0 & 8 & -4 \\ 0 & 1 & 2 & 12 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 42 \\ 8 \end{bmatrix}. \tag{1.18}$$

This equation system is in a particularly easy form, where the first two columns consist of a 1 and a 0.[11] Remember that we want to find scalars $x_1, \dots, x_4$, such that $\sum_{i=1}^{4} x_i \boldsymbol{c}_i = \boldsymbol{b}$, where we define $\boldsymbol{c}_i$ to be the ith column of the matrix and $\boldsymbol{b}$ the right-hand-side of (1.18). A solution to the problem in (1.18) can be found immediately by taking 42 times the first column and 8 times the second column, i.e.,

$$\boldsymbol{b} = \begin{bmatrix} 42 \\ 8 \end{bmatrix} = 42 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 8 \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{1.19}$$

Therefore, one solution vector is $[42, 8, 0, 0]^\top$. This solution is called a **particular solution** or **special solution**.

However, this is not the only solution of this linear equation system. Note that if $\boldsymbol{x} = \boldsymbol{a}$ is a solution of $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ and if $\boldsymbol{A}\boldsymbol{y} = \boldsymbol{0}$, then $\boldsymbol{x} = \boldsymbol{a} + \lambda \boldsymbol{y}$ is also a solution of $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$.

Now observe that:

$$\begin{bmatrix} 1 & 0 & 8 & -4 \\ 0 & 1 & 2 & 12 \end{bmatrix} \begin{bmatrix} 8 \\ 2 \\ -1 \\ 0 \end{bmatrix} = \boldsymbol{0}. \tag{1.20}$$

We also observe that

$$\begin{bmatrix} 1 & 0 & 8 & -4 \\ 0 & 1 & 2 & 12 \end{bmatrix} \begin{bmatrix} -4 \\ 12 \\ 0 \\ -1 \end{bmatrix} = \boldsymbol{0}. \tag{1.21}$$

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

We can check that the general solution to

$$\begin{bmatrix} 1 & 0 & 8 & -4 \\ 0 & 1 & 2 & 12 \end{bmatrix} \boldsymbol{x} = \boldsymbol{0} \tag{1.22}$$

is given by

$$\boldsymbol{x} = \lambda_1 \begin{bmatrix} 8 \\ 2 \\ -1 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} -4 \\ 12 \\ 0 \\ -1 \end{bmatrix}. \tag{1.23}$$

Putting everything together, we obtain all solutions of the linear equation system in (1.18), which is called the **general solution**, as

$$\begin{bmatrix} 42 \\ 8 \\ 0 \\ 0 \end{bmatrix} + \lambda_1 \begin{bmatrix} 8 \\ 2 \\ -1 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} -4 \\ 12 \\ 0 \\ -1 \end{bmatrix}, \qquad \lambda_1, \lambda_2 \in \mathbb{R}. \tag{1.24}$$

Note that we are free to choose λ_1 and λ_2 as any real number, which is why they are called *free variables*.

Remark 4

- *The general approach we followed consisted of the following three steps:*
  1. *Find a particular solution to* $\boldsymbol{Ax} = \boldsymbol{b}$
  2. *Find all solutions to* $\boldsymbol{Ax} = \boldsymbol{0}$
  3. *Combine the solutions from (1.20) and (1.21) to the general solution.*
- *Neither the general nor the particular solution is unique.*

The linear equation system in the example above was easy to solve because the matrix in (1.18) has this particularly convenient form, which allowed us to find the particular and the general solution by inspection. However, general equation systems are not of this simple form. Fortunately, there exists a constructive way of transforming any linear equation system into this particularly simple form: **Gaussian elimination**.

The rest of this section will introduce Gaussian elimination, which will allow us to solve all kinds of linear equation systems by first bringing them into a simple form and then applying the three steps to the simple form that we just discussed in the context of the example in (1.18), see Remark 4.

1.3.2 Elementary Transformations

Key to solving linear equation systems are **elementary transformations or operations** that keep the solution set the same[12], but transform the equation system into a simpler form. Consider any matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$. We have three types of elementary row operations on $\boldsymbol{A}$.

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

- Exchange of two rows of $\boldsymbol{A}$.
- Multiply of a row with a constant $\lambda \in \mathbb{R}\backslash\{0\}$
- Add a row to another row.

Let's see how to use these elementary row operations to solve a system of linear equations.

Example

We want to find the solutions of the following system of equations:

$$\begin{array}{rcrcrcrcrcr}
-2x_1 & + & 4x_2 & - & 2x_3 & - & x_4 & + & 4x_5 & = & -3 \\
4x_1 & - & 8x_2 & + & 3x_3 & - & 3x_4 & + & x_5 & = & 2 \\
x_1 & - & 2x_2 & + & x_3 & - & x_4 & + & x_5 & = & 0 \\
x_1 & - & 2x_2 & & & - & 3x_4 & + & 4x_5 & = & a
\end{array}, \qquad a \in \mathbb{R} \tag{1.25}$$

We start by converting this system of equations into the compact matrix notation $\boldsymbol{Ax} = \boldsymbol{b}$. We no longer mention the variables $\boldsymbol{x}$ explicitly and build the **augmented matrix**

$$\left[\begin{array}{rrrrr|r}
-2 & 4 & -2 & -1 & 4 & -3 \\
4 & -8 & 3 & -3 & 1 & 2 \\
1 & -2 & 1 & -1 & 1 & 0 \\
1 & -2 & 0 & -3 & 4 & a
\end{array}\right]
\begin{array}{l}
\text{Swap with } R_3 \\
\\
\text{Swap with } R_1 \\
\\
\end{array}$$

where we used the vertical line to separate the left-hand-side from the right-hand-side in (1.25). Swapping rows 1 and 3 leads to

$$\left[\begin{array}{rrrrr|r}
1 & -2 & 1 & -1 & 1 & 0 \\
4 & -8 & 3 & -3 & 1 & 2 \\
-2 & 4 & -2 & -1 & 4 & -3 \\
1 & -2 & 0 & -3 & 4 & a
\end{array}\right]
\begin{array}{l}
\\
-4R_1 \\
+2R_1 \\
-R_1
\end{array}$$

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

$$\left[\begin{array}{rrrrr|r}
1 & -2 & 1 & -1 & 1 & 0 \\
0 & 0 & -1 & 1 & -3 & 2 \\
0 & 0 & 0 & -3 & 6 & -3 \\
0 & 0 & -1 & -2 & 3 & a
\end{array}\right]
\begin{array}{l}
\\
\\
\\
-R_2
\end{array}
\rightsquigarrow
\left[\begin{array}{rrrrr|r}
1 & -2 & 1 & -1 & 1 & 0 \\
0 & 0 & -1 & 1 & -3 & 2 \\
0 & 0 & 0 & -3 & 6 & -3 \\
0 & 0 & 0 & -3 & 6 & a-2
\end{array}\right]
\begin{array}{l}
\\
\\
\\
-R_3
\end{array}$$

$$\rightsquigarrow
\left[\begin{array}{rrrrr|r}
1 & -2 & 1 & -1 & 1 & 0 \\
0 & 0 & -1 & 1 & -3 & 2 \\
0 & 0 & 0 & -3 & 6 & -3 \\
0 & 0 & 0 & 0 & 0 & a+1
\end{array}\right]
\begin{array}{l}
\\
\cdot(-1) \\
\cdot(-\frac{1}{3}) \\
\\
\end{array}$$

$$\rightsquigarrow
\left[\begin{array}{rrrrr|r}
1 & -2 & 1 & -1 & 1 & 0 \\
0 & 0 & 1 & -1 & 3 & -2 \\
0 & 0 & 0 & 1 & -2 & 1 \\
0 & 0 & 0 & 0 & 0 & a+1
\end{array}\right]$$

This (augmented) matrix is in a convenient form, which we will define below as the **row-echelon form (REF)**. Reverting this compact notation back into the explicit notation with the variables we seek, we obtain

$$\begin{array}{ccccccccccc} x_1 & - & 2x_2 & + & x_3 & - & x_4 & + & x_5 & = & 0 \\ & & & & x_3 & - & x_4 & + & 3x_5 & = & -2 \\ & & & & & & x_4 & - & 2x_5 & = & 1 \\ & & & & & & & & 0 & = & a+1 \end{array} \quad . \tag{1.26}$$

Only for $a = -1$, this equation system can be solved. So assume $a = -1$. Notice that the leading variables in the three equations now are x_1, x_3 and x_4, which are called the **basic** or **bound** variables. The other variables, namely x_2 and x_5 are called **free** variables. The basic or bound variables x_1, x_3 and x_4 can be obtained in terms of the free variables x_2 and x_5 by starting at the last equation, obtaining x_4 in terms of x_5 and then, using the second equation, obtaining x_3 in terms of x_4 and x_5, hence, in terms of x_5 and then finally, using the first equation, obtaining x_1 in terms of x_2, x_3, x_4 and x_5 and hence in terms of x_5 and x_2 as follows:

$$\begin{array}{lcl} x_4 & = & 1 + 2x_5 \\ x_3 & = & -2 + x_4 - 3x_5 = -2 + 1 + 2x_5 - 3x_5 = -1 - x_5 \\ x_1 & = & 2x_2 - x_3 + x_4 - x_5 = 2x_2 + 1 + x_5 + 1 + 2x_5 - x_5 = 2 + 2x_2 + 2x_5 \end{array}$$

A **particular solution** can now be immediately obtained by freely choosing the free variables $x_2 = x_5 = 0$ to obtain:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} \tag{1.27}$$

For the **general solution**, which captures the set of all possible solutions, we need to find the general solution of the $\boldsymbol{Ax} = \boldsymbol{0}$ i.e., the system

$$\begin{array}{ccccccccccc} x_1 & - & 2x_2 & + & x_3 & - & x_4 & + & x_5 & = & 0 \\ & & & & x_3 & - & x_4 & + & 3x_5 & = & 0 \\ & & & & & & x_4 & - & 2x_5 & = & 0 \end{array}$$

This can again be easily obtained by in turn putting all but one free variable equal to zero. Putting $x_5 = 0$ we obtain

$$\boldsymbol{x} = x_2 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{1.28}$$

And putting $x_2 = 0$, we get:

$$\boldsymbol{x} = x_5 \begin{bmatrix} 2 \\ 0 \\ -1 \\ 2 \\ 1 \end{bmatrix} \tag{1.29}$$

Thus, finally, the **general solution** of equation system (1.25) is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 2 \\ 0 \\ -1 \\ 2 \\ 1 \end{bmatrix}, \quad x_2, x_5 \in \mathbb{R} \tag{1.30}$$

Exercise 3

Let's check why Gaussian elimination using elementary row operations actually works for solving linear equation systems.

(i) *Check that if P is an elementary row operation and $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, then the application of P on $\boldsymbol{A}$ produces the matrix $\boldsymbol{A}' = \boldsymbol{U}\boldsymbol{A} \in \mathbb{R}^{m\times n}$ where $\boldsymbol{U} \in \mathbb{R}^{m\times m}$ is the result of applying the row operation P on the identity matrix $\boldsymbol{I}_m$.*

(ii) *Show that any row operation is "invertible", i.e., for any elementary row operation P there is an elementary row operation P^* such that P^* applied to $\boldsymbol{A}'$ gives back $\boldsymbol{A}$. Verify that this is equivalent to say that the matrix $\boldsymbol{U}$ above is invertible so that $\boldsymbol{U}^{-1}\boldsymbol{A}' = \boldsymbol{U}^{-1}\boldsymbol{U}\boldsymbol{A} = \boldsymbol{A}$, i.e., the elementary row operation P^* is equivalent to pre-multiplication with $\boldsymbol{U}^{-1}$.*

(iii) *From (i) and (ii), deduce that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ if and only if $(\boldsymbol{U}\boldsymbol{A})\boldsymbol{x} = \boldsymbol{U}\boldsymbol{b}$ where $\boldsymbol{U}$ is the matrix representing an elementary row operation P as in (i).*

(iv) *Prove that if we apply a sequence of elementary row operations $P_1, P_2 \ldots P_k$ with matrix representations $\boldsymbol{U}_1, \boldsymbol{U}_2 \ldots \boldsymbol{U}_k$, then by iteration of (iii) we have $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ if and only if $\boldsymbol{U}_k\boldsymbol{U}_{k-1}\ldots\boldsymbol{U}_1\boldsymbol{A}\boldsymbol{x} = \boldsymbol{U}_k\boldsymbol{U}_{k-1}\ldots\boldsymbol{U}_1\boldsymbol{b}$.*

(v) *Deduce the correctness of the Gaussian elimination method.*

Remark 5 (Pivots and Staircase Structure)

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

Definition 5 (Row-Echelon Form)

A matrix is in **row-echelon form** *(REF) if*

- *All nonzero rows (rows with at least one nonzero element) are above any rows of all zeros (all zero rows, if any, belong at the bottom of the matrix), and*
- *The leading coefficient (the first nonzero number from the left, also called the pivot) of a nonzero row is always strictly to the right of the leading coefficient of the row above it*[13].

Remark 6 (Basic or Bound and Free Variables)
The variables corresponding to the pivots in the row-echelon form are called **basic or bound variables***, the other variables are* **free variables***. For example, in* (1.26)*, x_1, x_3, x_4 are basic or bound variables, whereas x_2, x_5 are free variables.*

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

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 4 \\ 6 \\ 8 \end{bmatrix} \tag{1.31}$$

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

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

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

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

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

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -\frac{8}{3} \\ \frac{10}{3} \\ 0 \end{bmatrix}. \tag{1.32}$$

To find the general solution, we combine the particular solution with the solution of the homogeneous equation system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ using x_3 as the free variable. From the row

[13] In some literature, it is required that the leading coefficient is 1.

echelon form, we see that $x_2 + 2x_3 = 0 \Rightarrow x_2 = -2x_3$. With this, we now look at the first set of equations and obtain $x_1 + 2x_2 + 3x_3 = 0 \Rightarrow x_1 = -2x_2 - x_3 = 4x_3 - 3x_3 = x_3$. Thus, we arrive at the general solution

$$\begin{bmatrix} -\frac{8}{3} \\ \frac{10}{3} \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, x_3 \in \mathbb{R}. \tag{1.33}$$

Remark 7 (Reduced Row Echelon Form)
An equation system is in **reduced row echelon form**[14] *if*

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

Example (Reduced Row Echelon Form)
Solve for $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ where $\boldsymbol{A}$ is the following matrix which is in reduced row echelon form (the pivots are coloured in red):

$$\boldsymbol{A} = \begin{bmatrix} 1 & 3 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 & 9 \\ 0 & 0 & 0 & 1 & -4 \end{bmatrix} \tag{1.34}$$

The key idea for finding the solutions of $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ is to look at the *non-pivot columns*, which give us the free variables x_2 and x_5. Working backwards we find the basic or bound variables x_1, x_3 and x_4 in terms of the free variables x_2 and x_5:

$$\begin{aligned} x_4 &= 4x_5 \\ x_3 &= -9x_5 \\ x_1 &= -3x_2 - 3x_5 \end{aligned}$$

Putting first $x_5 = 0$ we obtain

$$\boldsymbol{x} = x_2 \begin{bmatrix} 3 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{1.35}$$

Next putting $x_2 = 0$, we get:

$$\boldsymbol{x} = x_5 \begin{bmatrix} 3 \\ 0 \\ 9 \\ -4 \\ -1 \end{bmatrix} \tag{1.36}$$

[14] Also called: **row-reduced echelon form** or **row canonical form**

Thus, all solutions of $\boldsymbol{Ax} = \boldsymbol{0}, \boldsymbol{x} \in \mathbb{R}^5$ are given by

$$\lambda_1 \begin{bmatrix} 3 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 3 \\ 0 \\ 9 \\ -4 \\ -1 \end{bmatrix}, \quad \lambda_1, \lambda_2 \in \mathbb{R}. \tag{1.37}$$

a

The general form of a matrix in reduced REF is given below where $*$ denotes any real number:

$$\boldsymbol{A} = \begin{bmatrix} 0 & \cdots & 0 & 1 & * & \cdots & * & 0 & * & \cdots & * & 0 & * & \cdots & * \\ \vdots & & \vdots & 0 & 0 & \cdots & 0 & 1 & * & \cdots & * & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & 0 & * & \ddots & * & 0 & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & * & \ddots & * & 0 & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & * & \cdots & * & 1 & * & \cdots & * \end{bmatrix} \tag{1.38}$$

1.3.3 Applications of Gaussian Elimination in Linear Algebra

Gaussian elimination can also be used to find the rank of a matrix (Chapter 1.6.2), to calculate the determinant of a matrix (discussed in a later section), the null space, and the inverse of an invertible square matrix. Because of its relevance to central concepts in Linear Algebra, Gaussian elimination is the most important algorithm we will cover.

Gaussian elimination is an important prerequisite for CO343 Operations Research. The initial part of the course discusses in depth systems of linear equations with more columns than rows, and the related notions of basic solutions and non-basic solutions.

Calculating the Inverse

To compute the inverse $\boldsymbol{A}^{-1}$ of $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, we need to satisfy $\boldsymbol{AA}^{-1} = \boldsymbol{I}_n$. We can write this down as a set of simultaneous linear equations $\boldsymbol{AX} = \boldsymbol{I}_n$, where we solve for $\boldsymbol{X} = [\boldsymbol{x}_1|\cdots|\boldsymbol{x}_n]$. We use the augmented matrix notation for a compact representation of this set of linear equation systems and obtain

$$\left[\boldsymbol{A}|\boldsymbol{I}_n\right] \quad \rightsquigarrow \cdots \rightsquigarrow \quad \left[\boldsymbol{I}_n|\boldsymbol{A}^{-1}\right]. \tag{1.39}$$

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

Example

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

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

We bring this system now into reduced row echelon form

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

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

$$A^{-1} = \begin{bmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix}. \tag{1.40}$$

Example
To determine the inverse of

$$A = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \tag{1.41}$$

we write down the augmented matrix

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

and transform it into reduced row echelon form

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

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

$$A^{-1} = \begin{bmatrix} -1 & 2 & -2 & 2 \\ 1 & -1 & 2 & -2 \\ 1 & -1 & 1 & -1 \\ -1 & 0 & -1 & 2 \end{bmatrix}. \tag{1.42}$$

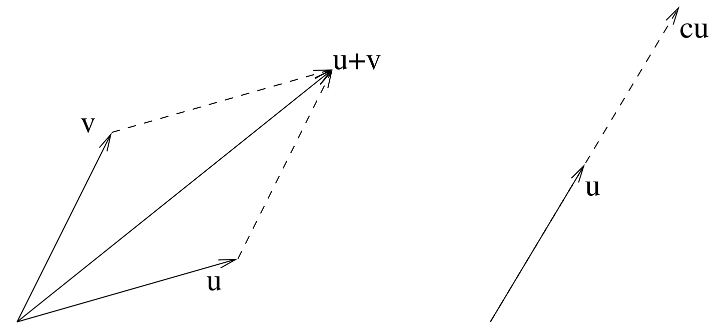


Figure 1.2: Addition and scalar multiplication of geometric vectors

Exercise 4
Use Exercise (3) to prove the correctness of the above algorithm to find the inverse of $\boldsymbol{A}$.

Remark 8
Inverse of a matrix can also be obtained by using co-factors. You may have encountered a way of computing the inverse of a matrix using cross-products in $\mathbb{R}^3$.

1.4 Finite dimensional vector spaces: $\mathbb{R}^n$

When we discussed vectors in the introduction, we pointed out that they can be added and multiplied by a scalar, in our case a real number. We call these two properties the linear structure in $\mathbb{R}^n$. We will focus on $\mathbb{R}^n$ and study its various properties as almost all algorithms in linear algebra are given in $\mathbb{R}^n$. We observe that the geometric interpretation of a vector gives us a way of interpreting the addition of two vectors by the "parallelogram law" and scalar multiplication as "re-scaling" (and possibly by also reorienting the vector in the opposite direction, when the scalar is a negative real number) as in Figure 1.2. The geometric interpretation of vectors and their linear structures is coordinate free.
The other standard way of representing a vector $\boldsymbol{v} \in \mathbb{R}^n$ is by regarding it as an $n \times 1$ matrix of real numbers and providing its *coordinates* with respect to the Cartesian coordinate system, i.e., $\boldsymbol{v} = v_1\boldsymbol{e}_1 + \ldots + v_n\boldsymbol{e}_n$, where $v_i \in \mathbb{R}$ for $i = 1, \ldots, n$ and

$$\boldsymbol{e}_i = [\overbrace{0, \ldots, 0, 1, 0, \ldots, 0}^{i}]^\top = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{1.43}$$

with 1 is in the $i$the place for $\boldsymbol{e}_i$, $i = 1, \ldots, n$. Note that with our notation, all vectors are written as "column" vectors.

Remark 9
We have the following two possible "matrix multiplications" for vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$: $\boldsymbol{a}\boldsymbol{b}^\top \in \mathbb{R}^{n\times n}$ (outer product), $\boldsymbol{a}^\top \boldsymbol{b} \in \mathbb{R}$ (inner/scalar product).

In $\mathbb{R}^n$, for $n \in \mathbb{N}$, we can alternatively represent a vectors by an n-tuples of real numbers in a column. In this representation, a vector $\boldsymbol{v} \in \mathbb{R}^n$ is simply a matrix in $\mathbb{R}^{n\times 1}$ and the two basic operations on vectors are simply the corresponding operations on the matrices, i.e., follows:

- Addition: $\boldsymbol{x}+\boldsymbol{y} = [x_1, \dots, x_n]^\top + [y_1, \dots, y_n]^\top = [x_1+y_1, \dots, x_n+y_n]^\top$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$
- Multiplication by scalars: $\lambda \boldsymbol{x} = \lambda[x_1, \dots, x_n]^\top = [\lambda x_1, \dots, \lambda x_n]^\top$ for all $\lambda \in \mathbb{R}, \boldsymbol{x} \in \mathbb{R}^n$

Remark 10
Note that we can regard the set $\mathbb{R}^{m\times n}, m, n \in \mathbb{N}$ of all $m \times n$ matrices equivalently as the set of vectors in $\mathbb{R}^{mn}$. Given $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, the **vectorisation** *of $\boldsymbol{A}$ is the vector $[\boldsymbol{a}_1^\top, \boldsymbol{a}_2^\top, \dots \boldsymbol{a}_n^\top]^\top \in \mathbb{R}^{mn}$ where $\boldsymbol{a}_i$ is the ith column of $\boldsymbol{A}$.*

Remark 11 (Notation)
The three vector spaces $\mathbb{R}^n, \mathbb{R}^{n\times 1}, \mathbb{R}^{1\times n}$ are only different with respect to the way of writing. In the following, we will not make a distinction between $\mathbb{R}^n$ and $\mathbb{R}^{n\times 1}$, which allows us to write n-tuples as **column vectors**

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}. \tag{1.44}$$

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

1.4.1 Vector Subspaces

Definition 6 (Vector Subspace)
Let $U \subset \mathbb{R}^n$, with $U \neq \emptyset$. Then, U is called a **vector subspace** *of $\mathbb{R}^n$ (or* **linear subspace***) if U is closed under addition and scalar multiplication, i.e., $\boldsymbol{a}, \boldsymbol{b} \in U \Rightarrow \boldsymbol{a}+\boldsymbol{b} \in U$ and $\boldsymbol{a} \in U \Rightarrow \lambda\boldsymbol{a} \in U$ for all $\lambda \in \mathbb{R}$.*

Examples

- There are two trivial subspaces of $\mathbb{R}^n$: $\{\boldsymbol{0}\}$ and $\mathbb{R}^n$ itself. Find out exactly the other subspaces of $\mathbb{R}^3$!
- The solution set of a homogeneous linear equation system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ with n unknowns $\boldsymbol{x} = [x_1, \dots, x_n]^\top$ is a subspace of $\mathbb{R}^n$ (why?)

- The solution set of an inhomogeneous equation system $\boldsymbol{Ax} = \boldsymbol{b}, \boldsymbol{b} \neq \boldsymbol{0}$ is not a subspace of $\mathbb{R}^n$, but it is of the form $\boldsymbol{a} + U = \{\boldsymbol{a} + \boldsymbol{u} : \boldsymbol{u} \in U\}$, where $\boldsymbol{a} \in \mathbb{R}^n$ is a particular solution of $\boldsymbol{Ax} = \boldsymbol{b}$ and U is the subspace of the solution set of the homogeneous equation $\boldsymbol{Ax} = \boldsymbol{0}$. The set $\boldsymbol{a} + U$ is called an **affine space**.
- The intersection of arbitrarily many subspaces is a subspace itself (why?).
- If $V \subset \mathbb{R}^n$ is any set, then the intersection of all subspaces $U \supset V$ is a subspace of $\mathbb{R}^n$ called the **linear hull** of V.

Remark 12
Every subspace $U \subset \mathbb{R}^n$ is the solution space of a homogeneous linear equation system $\boldsymbol{Ax} = \boldsymbol{0}$ (check this in $\mathbb{R}^3$!).

1.5 Linear Independence

In Section 1.4, we learned about linear combinations of vectors, see (1.45). We formalise this here.

Definition 7 (Linear Combination)
Consider a finite number of vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in \mathbb{R}^n$. Then, every vector $\boldsymbol{v} \in \mathbb{R}^n$ of the form

$$\boldsymbol{v} = \lambda_1 \boldsymbol{x}_1 + \cdots + \lambda_k \boldsymbol{x}_k = \sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i \in \mathbb{R}^n \tag{1.45}$$

with $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ is a **linear combination** *of the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$. It is a* **non-trivial linear combination** *if there exists m with $1 \leq m \leq k$ such that $\lambda_m \neq 0$.*

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

Definition 8 (Linear dependence and linear independence)
Consider $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in \mathbb{R}^n$. If there is a non-trivial linear combination, such that $\boldsymbol{0} = \sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i$, the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ are said to be **linearly dependent***. If only the trivial solution exists, i.e., $\lambda_1 = \ldots = \lambda_k = 0$ the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ are said to be* **linearly independent***.*

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

Example From Wikipedia (2015)
A geographic example may help to clarify the concept of linear independence. A person describing the location of a certain place might say, "It is 3 miles North and 4 miles East of here." This is sufficient information to describe the location, because

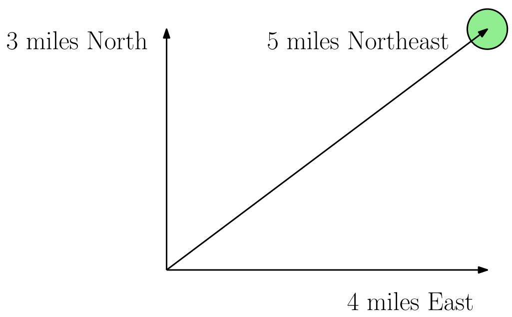


Figure 1.3: Linear dependence of three vectors in $\mathbb{R}^2$. (plane).

the geographic coordinate system may be considered as in $\mathbb{R}^2$ (ignoring altitude and the curvature of the Earth's surface). The person might add, "The place is 5 miles Northeast of here." Although this last statement is true, it is not necessary to find this place (see Fig. 1.3 for an illustration).
In this example, the "3 miles North" vector and the "4 miles East" vector are linearly independent. That is to say, the north vector cannot be described in terms of the east vector, and vice versa. The third "5 miles Northeast" vector is a linear combination of the other two vectors, and it makes the set of vectors linearly dependent, that is, one of the three vectors is unnecessary.

Example

(i) In $\mathbb{R}^3$ the three vectors $\boldsymbol{a}_1 = [1,-1,2]^\top$, $\boldsymbol{a}_2 = [2,1,-1]^\top$ and $\boldsymbol{a}_3 = [0,0,1]^\top$ are linearly independent since $x_1[1,-1,0]^\top + x_2[2,1,0d])^\top + x_3[0,0,1]^\top = 0$ implies $x_3 = 0$ and $x_1 + 2x_2 = 0$ and $-x_1 + x_2 = 0$ which result in $x_2 = x_3 = 0$.

(ii) In $\mathbb{R}^3$ the three vectors $\boldsymbol{a}_1 = [1,-1,2]^\top$, $\boldsymbol{a}_2 = [2,1,-1]^\top$ and $\boldsymbol{a}_3 = [4,-1,3]^\top$ are linearly dependent since $2\boldsymbol{a}_1 + \boldsymbol{a}_2 - \boldsymbol{a}_3 = 0$.

Exercise 5
Check that an invertible matrix preserves linear dependence or linear independence of a set of vectors: If $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ *is invertible and we have* k *vectors* $\boldsymbol{b}_1,\dots,\boldsymbol{b}_k \in \mathbb{R}^n$*, then* $\boldsymbol{b}_1,\dots,\boldsymbol{b}_k \in \mathbb{R}^n$ *are linearly independent iff* $\boldsymbol{A}\boldsymbol{b}_1,\dots,\boldsymbol{A}\boldsymbol{b}_k \in \mathbb{R}^n$ *are linearly independent.*

Exercise 6
Use Exercises 3 and 5 to show that elementary row operations preserve the linear dependence and the linear independence of the columns of any matrix in $\mathbb{R}^{m\times n}$.

Remark 13
The following properties are useful in finding out whether vectors are linearly independent.

(i) *Any k vectors are either linearly dependent or linearly independent by definition.*

(ii) *If at least one of the vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k$ is $\boldsymbol{0}$ then they are linearly dependent. The same holds if two vectors are identical.*

(iii) *The vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k \neq \boldsymbol{0}$, $k \geq 2$, are linearly dependent if and only if (at least) one of them is a linear combination of the others.*

(iv) *Note that $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k \in \mathbb{R}^n$ are linearly independent iff $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ implies $\boldsymbol{x} = 0$ where $\boldsymbol{A} = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_k] \in \mathbb{R}^{n \times k}$ (check this!). Therefore, a practical way of checking whether vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k \in \mathbb{R}^n$ are linearly independent is to use Gaussian elimination to see if $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ implies $\boldsymbol{x} = \boldsymbol{0}$. Gaussian elimination yields a matrix in (reduced) row echelon form which can determine whether or not the only solution is $\boldsymbol{x} = \boldsymbol{0}$.*

In fact, often you can find the answer in the process of reducing $\boldsymbol{A}$ to row echelon form: $\boldsymbol{x} = \boldsymbol{0}$ is the only solution iff all columns are pivot columns. Thus as soon as you obtain a non-pivot column then you know the there are non-trivial solutions and the columns $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k \in \mathbb{R}^n$ are linearly dependent:

- *The pivot columns indicate the vectors, which are linearly independent of the previous[15] vectors (note that there is an ordering of vectors when the matrix is built).*
- *The non-pivot columns can be expressed as linear combinations of the pivot columns on their left. For instance, in*

$$\begin{bmatrix} 1 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1.46}$$

the first and third column are pivot columns. The second column is a non-pivot column because it is 3 times the first column.

If all columns are pivot columns, the column vectors are linearly independent. If there is at least one non-pivot column, the columns are linearly dependent.

Exercise 7
Use Exercise 6 to show the correctness of the method in Remark 13(iv).

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

$$\boldsymbol{a}_1 = \begin{bmatrix} 1 \\ 2 \\ -3 \\ 4 \end{bmatrix}, \quad \boldsymbol{a}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 2 \end{bmatrix}, \quad \boldsymbol{a}_3 = \begin{bmatrix} -1 \\ -2 \\ 1 \\ 1 \end{bmatrix}. \tag{1.47}$$

[15] the vectors on the left

To check whether they are linearly dependent, we take the matrix $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3]$, whose columns are precisely the vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ and reduce $\boldsymbol{A}$ to REF to see if $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ has any non-trivial solution $\boldsymbol{x} \neq \boldsymbol{0}$.

$$
\begin{aligned}
&\begin{bmatrix} 1 & 1 & -1 \\ 2 & 1 & -2 \\ -3 & 0 & 1 \\ 4 & 2 & 1 \end{bmatrix} \begin{array}{l} \\ -2R_1 \\ +3R_1 \\ -4R_1 \end{array} &&\rightsquigarrow \begin{bmatrix} 1 & 1 & -1 \\ 0 & -1 & 0 \\ 0 & 3 & -2 \\ 0 & -2 & 5 \end{bmatrix} \begin{array}{l} \\ \\ +3R_2 \\ -2R_2 \end{array} \\
\rightsquigarrow &\begin{bmatrix} 1 & 1 & -1 \\ 0 & -1 & 0 \\ 0 & 0 & -2 \\ 0 & 0 & 5 \end{bmatrix} \begin{array}{l} \\ \cdot(-1) \\ \cdot(-\frac{1}{2}) \\ +\frac{5}{2}R_3 \end{array} &&\rightsquigarrow \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}
\end{aligned}
$$

Here, every column of the matrix is a pivot column[16], i.e., every column is linearly independent of the columns on its left. Therefore, there is no non-trivial solution, and we require $\lambda_1 = 0, \lambda_2 = 0, \lambda_3 = 0$. Hence, the vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ are linearly independent.

Theorem 1

Consider k linearly independent vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k \in \mathbb{R}^n$ and m linear combinations

$$\boldsymbol{x}_1 = \sum_{i=1}^{k} a_{i1}\boldsymbol{b}_i \quad \ldots \quad \boldsymbol{x}_j = \sum_{i=1}^{k} a_{ij}\boldsymbol{b}_i \quad \ldots \quad \boldsymbol{x}_m = \sum_{i=1}^{k} a_{im}\boldsymbol{b}_i. \tag{1.48}$$

Then $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ are linearly independent if and only if the vectors

$$\boldsymbol{a}_1 = \begin{bmatrix} a_{11} \\ \vdots \\ a_{k1} \end{bmatrix} \quad \ldots \quad \boldsymbol{a}_j = \begin{bmatrix} a_{1j} \\ \vdots \\ a_{kj} \end{bmatrix} \quad \ldots \quad \boldsymbol{a}_m = \begin{bmatrix} a_{1m} \\ \vdots \\ a_{km} \end{bmatrix} \in \mathbb{R}^k \tag{1.49}$$

are linearly independent, i.e., the columns of the matrix $\boldsymbol{A} \in \mathbb{R}^{k \times m}$ with $(\boldsymbol{A})_{ij} = a_{ij}$ where $1 \leq i \leq k$ and $1 \leq i \leq m$ are linearly independent.

Proof I. By a direct method. We first take any linear combination $\boldsymbol{y}$ of vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ and express it as a linear combination of vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k \in \mathbb{R}^n$:

$$\boldsymbol{y} = \sum_{j=1}^{m} c_j \boldsymbol{x}_j = \sum_{j=1}^{m} c_j \left(\sum_{i=1}^{k} a_{ij}\boldsymbol{b}_i \right) = \sum_{j=1}^{m} \sum_{i=1}^{k} c_j a_{ij} \boldsymbol{b}_i = \sum_{i=1}^{k} \left(\sum_{j=1}^{m} c_j a_{ij} \right) \boldsymbol{b}_i. \tag{1.50}$$

Assume now that the vectors $\boldsymbol{a}_j = [a_{1j}, a_{2j}, \ldots, a_{kj}]^\top$ for $j = 1, \ldots, m$ are linearly independent. To show that the vectors $\boldsymbol{x}_j$ for $j = 1, \ldots, m$ are linearly independent,

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

we need to show that whenever $\boldsymbol{y} = \boldsymbol{0}$ in (1.50), then $c_j = 0$ for $1 \le j \le m$. Let $\boldsymbol{y} = \boldsymbol{0}$ in (1.50). Since the vectors $\boldsymbol{b}_i$ for $1 \le i \le k$ are linearly independent, from the relation (1.50), we obtain for all $i = 1,\dots,k$. $\sum_{j=1}^m c_j a_{ij} = \boldsymbol{0}$, i.e., $\sum_{j=1}^m c_j \boldsymbol{a}_j = \boldsymbol{0}$. Then, since $\boldsymbol{a}_j$ for $j = 1,\dots,m$ are linearly independent we get $c_j = 0$ for $1 \le j \le m$. We conclude that $\boldsymbol{x}_1,\dots,\boldsymbol{x}_m$ are linearly independent. Next suppose $\boldsymbol{x}_1,\dots,\boldsymbol{x}_m$ are linearly independent. Assume $\sum_{j=1}^m c_j \boldsymbol{a}_j = \boldsymbol{0}$, i.e., $\sum_{j=1}^m c_j a_{ij} = 0$ for each $i = 1,\dots,k$. Thus, by (1.50), it follows that $\sum_{j=1}^m c_j \boldsymbol{x}_j = 0$ and hence, since $\boldsymbol{x}_1,\dots,\boldsymbol{x}_m$ are linearly independent, $c_j = 0$ for $j = 1,\dots,m$. We conclude that the vectors $\boldsymbol{a}_j$ for $j = 1,\dots,m$ are linearly independent.

II. A simpler proof. Note that if in Theorem 1 we put $\boldsymbol{X} = [\boldsymbol{x}_1,\dots,\boldsymbol{x}_m] \in \mathbb{R}^{n\times m}$ and $\boldsymbol{B} = [\boldsymbol{b}_1,\dots,\boldsymbol{b}_k] \in \mathbb{R}^{n\times k}$, then we have the compact representation $\boldsymbol{X} = \boldsymbol{B}\boldsymbol{A}$. So suppose the columns of $\boldsymbol{A}$ are linearly independent and $\boldsymbol{X}\boldsymbol{y} = \boldsymbol{0}$ for some $\boldsymbol{y} \in \mathbb{R}^n$. It is sufficient to show that $\boldsymbol{y} = \boldsymbol{0}$. But $\boldsymbol{B}(\boldsymbol{A}\boldsymbol{y}) = \boldsymbol{B}\boldsymbol{A}\boldsymbol{y} = \boldsymbol{X}\boldsymbol{y} = \boldsymbol{0}$ and since the columns of $\boldsymbol{B}$ are linearly independent we get $\boldsymbol{A}\boldsymbol{y} = \boldsymbol{0}$ which in turn (since the columns of $\boldsymbol{A}$ are linearly independent) implies $\boldsymbol{y} = \boldsymbol{0}$ are required. Next suppose, the columns of $\boldsymbol{X}$, i.e., $\boldsymbol{x}_1,\dots,\boldsymbol{x}_m$ are linearly independent and $\boldsymbol{A}\boldsymbol{y} = 0$. Then $\boldsymbol{X}\boldsymbol{y} = \boldsymbol{B}\boldsymbol{A}\boldsymbol{y} = \boldsymbol{0}$ and hence $\boldsymbol{y} = \boldsymbol{0}$, i.e., the columns of $\boldsymbol{A}$ are linearly independent.

Example

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

$$\begin{array}{rcrcrcrcr} \boldsymbol{x}_1 &=& \boldsymbol{b}_1 &-& 2\boldsymbol{b}_2 &+& \boldsymbol{b}_3 &-& \boldsymbol{b}_4 \\ \boldsymbol{x}_2 &=& -4\boldsymbol{b}_1 &-& 2\boldsymbol{b}_2 && &+& 4\boldsymbol{b}_4 \\ \boldsymbol{x}_3 &=& 2\boldsymbol{b}_1 &+& 3\boldsymbol{b}_2 &-& \boldsymbol{b}_3 &-& 3\boldsymbol{b}_4 \\ \boldsymbol{x}_4 &=& 17\boldsymbol{b}_1 &-& 10\boldsymbol{b}_2 &+& 11\boldsymbol{b}_3 &+& \boldsymbol{b}_4 \end{array} \tag{1.51}$$

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

$$\left\{ \begin{bmatrix} 1 \\ -2 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} -4 \\ -2 \\ 0 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ -1 \\ -3 \end{bmatrix}, \begin{bmatrix} 17 \\ -10 \\ 11 \\ 1 \end{bmatrix} \right\} \tag{1.52}$$

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

$$\boldsymbol{A} = \begin{bmatrix} 1 & -4 & 2 & 17 \\ -2 & -2 & 3 & -10 \\ 1 & 0 & -1 & 11 \\ -1 & 4 & -3 & 1 \end{bmatrix} \tag{1.53}$$

is given as

$$\begin{bmatrix} 1 & 0 & 0 & -7 \\ 0 & 1 & 0 & -15 \\ 0 & 0 & 1 & -18 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{1.54}$$

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

1.6 Generating Sets, Basis and Dimension

In $\mathbb{R}^n$, we are particularly interested in sets A of vectors that possess the property that any vector $\boldsymbol{v} \in \mathbb{R}^n$ can be obtained by a linear combination of vectors in A.

1.6.1 Generating Sets

Definition 9 (Generating Set/Span)
Consider $A = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_k\} \subset \mathbb{R}^n$ and let $U \subset \mathbb{R}^n$ be a subspace. If every vector $\boldsymbol{v} \in U$ can be expressed as a linear combination of $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k$, then A is called a **generating set** *for U and we say A* **spans** *U, writing $U = span(A)$ or $U = span\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_k\}$.*

1.6.2 Basis

We have just generating sets. Now, we will be now more specific and characterize the smallest generating set that spans a subspace.

Definition 10 (Dimension and Basis)
Consider a subspace $U \subset \mathbb{R}^k$.

(i) A generating set $A \subset U$ is called **minimal** *if there exists no smaller set $\tilde{A} \subset A \subset U$, which spans U.*

(ii) The number of vectors in a minimal generating set of U is called the **dimension** *of U*

(iii) A **basis** *of a subspace $U \subset \mathbb{R}^n$ is a minimal generating set of U.*

Example
The n vectors in (1.43) give a basis for $\mathbb{R}^n$, called the **standard basis**.

Example

- In $\mathbb{R}^3$, three different set of bases are given by

$$\left\{\begin{bmatrix}1\\1\\0\end{bmatrix},\begin{bmatrix}0\\1\\1\end{bmatrix},\begin{bmatrix}1\\0\\1\end{bmatrix}\right\},\quad \left\{\begin{bmatrix}1\\0\\0\end{bmatrix},\begin{bmatrix}1\\1\\0\end{bmatrix},\begin{bmatrix}1\\1\\1\end{bmatrix}\right\},\quad \left\{\begin{bmatrix}0.53\\0.86\\-0.43\end{bmatrix},\begin{bmatrix}1.83\\0.31\\0.34\end{bmatrix},\begin{bmatrix}-2.25\\-1.30\\3.57\end{bmatrix}\right\} \tag{1.55}$$

- The vectors in the set

$$A = \left\{\begin{bmatrix}1\\2\\3\\4\end{bmatrix},\begin{bmatrix}2\\-1\\0\\2\end{bmatrix},\begin{bmatrix}1\\1\\0\\-4\end{bmatrix}\right\} \tag{1.56}$$

are linearly independent, but A is not a generating set for $\mathbb{R}^4$ (and thus not a basis): For instance, the vector $[1,0,0,0]^\top$ cannot be obtained by a linear combination of elements in A.

Exercise 8
Show that the dimension of a subspace is well-defined, i.e., Definition 10(ii) does not depend on the minimal generating set of a subspace. (Use Theorem 1 and Remark **??** *to give a proof by contradiction.)*

Remark 14
Let $U \subset \mathbb{R}^n$ be a subspace and $B \subset U$. Then, the following statements are equivalent:

(i) B is a basis, i.e., a minimal generating set of U.

(ii) B is a maximal linearly independent subset of U.

(iii) Every vector $\boldsymbol{x} \in U$ is given as a unique linear combination of vectors from B, i.e., with

$$\boldsymbol{x} = \sum_{i=1}^{k} \lambda_i \boldsymbol{b}_i = \sum_{i=1}^{k} \beta_i \boldsymbol{b}_i \tag{1.57}$$

and $\lambda_i, \beta_i \in \mathbb{R}$, $\boldsymbol{b}_i \in B$ it follows that $\lambda_i = \beta_i,\ i = 1,\dots,k$.

Exercise 9
Prove the equivalence of the three statements in Remark 14.

Remark 15

- *Every subspace has a basis.*

- *The examples above show that there can be many bases of a subspace in particular $\mathbb{R}^n$ itself, i.e., there is no unique basis. However, all bases possess the same number of elements, the* **basis vectors**.

- *The* **dimension** *is the number of basis vectors, and we write it as* $\dim(V)$.
- *If* $U \subset V$ *is a subspace of* V *then* $\dim(U) \leq \dim(V)$ *and* $\dim(U) = \dim(V)$ *if and only if* $U = V$.

Remark 16
Suppose a subspace is spanned by the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \subset \mathbb{R}^n$. *Then a basis of* U *consisting of these spanning vectors can be found by executing the following steps:*

1. *Write the spanning vectors as columns of a matrix* $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m]$.
2. *Apply Gaussian elimination algorithm to* $\boldsymbol{X}$.
3. *The spanning vectors associated with the pivot columns form a basis of* U.

Example (Determining a Basis)
For a vector subspace $U \subset \mathbb{R}^5$, spanned by the vectors

$$\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \\ -1 \\ -1 \end{bmatrix}, \quad \boldsymbol{x}_2 = \begin{bmatrix} 2 \\ -1 \\ 1 \\ 2 \\ -2 \end{bmatrix}, \quad \boldsymbol{x}_3 = \begin{bmatrix} 3 \\ -4 \\ 3 \\ 5 \\ -3 \end{bmatrix}, \quad \boldsymbol{x}_4 = \begin{bmatrix} -1 \\ 8 \\ -5 \\ -6 \\ 1 \end{bmatrix} \in \mathbb{R}^5, \tag{1.58}$$

we are interested in finding out which vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_4$ are a basis for U. For this, we need to check whether $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_4$ are linearly independent. Therefore, we need to solve

$$\sum_{i=1}^{4} \lambda_i \boldsymbol{x}_i = \boldsymbol{0}, \tag{1.59}$$

which leads to a homogeneous equation system with the corresponding matrix

$$\left[\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4\right] = \begin{bmatrix} 1 & 2 & 3 & -1 \\ 2 & -1 & -4 & 8 \\ -1 & 1 & 3 & -5 \\ -1 & 2 & 5 & -6 \\ -1 & -2 & -3 & 1 \end{bmatrix}. \tag{1.60}$$

With the basic transformation of linear equation systems, we obtain

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

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

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

From this reduced-row echelon form we see that $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_4$ belong to the pivot columns, and, therefore, linearly independent (because the linear equation system $\lambda_1 \boldsymbol{x}_1 + \lambda_2 \boldsymbol{x}_2 + \lambda_4 \boldsymbol{x}_4 = \boldsymbol{0}$ can only be solved with $\lambda_1 = \lambda_2 = \lambda_4 = 0$). Therefore, $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_4\}$ is a basis of U.

Exercise 10
Show that the steps in Remark 16 correctly lead to a basis.

Example (Finding a Simple Basis)
Let us now consider a slightly different problem: Instead of finding out which vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_4$ of the span of U form a basis, we are interested in finding a "simple" basis for U. Here, "simple" means that we are interested in basis vectors with many coordinates equal to 0.
To solve this problem we replace the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_4$ with suitable linear combinations. In practice, we write $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_4$ as *row vectors* in a matrix and perform Gaussian elimination to obtain the reduced row echelon form:

$$\left[\begin{array}{ccccc} 1 & 2 & -1 & -1 & -1 \\ 2 & -1 & 1 & 2 & -2 \\ 3 & -4 & 3 & 5 & -3 \\ -1 & 8 & -5 & -6 & 1 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccccc} 1 & 0 & \frac{1}{5} & 0 & -1 \\ 0 & 1 & -\frac{3}{5} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] .$$

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

$$U = \operatorname{span}\left\{ \underbrace{\begin{bmatrix} 1 \\ 0 \\ \frac{1}{5} \\ 0 \\ -1 \end{bmatrix}}_{\boldsymbol{b}_1}, \underbrace{\begin{bmatrix} 0 \\ 1 \\ -\frac{3}{5} \\ 0 \\ 0 \end{bmatrix}}_{\boldsymbol{b}_2}, \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}}_{\boldsymbol{b}_3} \right\} \tag{1.61}$$

and $B = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3\}$ is a (simple) basis of U (check linear independence).

1.6.3 Matrix rank

Definition 11

The number of linearly independent columns of a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is called **rank** *of $\boldsymbol{A}$ and is denoted by $\mathrm{rk}(\boldsymbol{A})$.*

Remark 17

(i) $\mathrm{rk}(\boldsymbol{A}) = \mathrm{rk}(\boldsymbol{A}^\top)$*, i.e., the column rank equals the row rank.*

(ii) The columns of $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ span a subspace $U \subset \mathbb{R}^m$ with $\dim(U) = \mathrm{rk}(\boldsymbol{A})$. A basis of U can be found by applying Gaussian elimination to $\boldsymbol{A}$ to identify the pivot columns.

(iii) The rows of $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ span a subspace $W \subset \mathbb{R}^n$ with $\dim(W) = \mathrm{rk}(\boldsymbol{A})$. A basis of W can be found by applying the Gaussian elimination algorithm to $\boldsymbol{A}^\top$.

(iv) For all $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, we have: $\boldsymbol{A}$ is invertible if and only if $\mathrm{rk}(\boldsymbol{A}) = n$.

(v) For all $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and all $\boldsymbol{b} \in \mathbb{R}^m$: The linear equation system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ can be solved if and only if $\mathrm{rk}(\boldsymbol{A}) = \mathrm{rk}([\boldsymbol{A}|\boldsymbol{b}])$, where $[\boldsymbol{A}|\boldsymbol{b}]$ denotes the augmented system. In fact, $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ has a solution $x \in \mathbb{R}^m$ iff $x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_n\boldsymbol{a}_n = \boldsymbol{b}$, i.e., iff $\boldsymbol{b}$ is a linear combination of the column vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ of $\boldsymbol{A}$ iff $\mathrm{rk}(\boldsymbol{A}) = \mathrm{rk}([\boldsymbol{A}|\boldsymbol{b}])$.

(vi) A matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ has **full rank** *if its rank equals the largest possible for a matrix of the same dimensions, which is the lesser of the number of rows and columns, i.e., $\mathrm{rk}(\boldsymbol{A}) = \min(m, n)$. A matrix is said to be* **rank deficient** *if it does not have full rank.*

(vii) For $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ the subspace of solutions for $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ has dimension $n - \mathrm{rk}(\boldsymbol{A})$ [17], whereas the subspace of solutions for $\boldsymbol{A}^\top\boldsymbol{x} = \boldsymbol{0}$ has dimension $m - \mathrm{rk}(\boldsymbol{A})$.

Example

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

[17] Later, we will be calling this subspace **kernel** or **nullspace** of $\boldsymbol{A}$.

- $\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix}$ We use Gaussian elimination to determine the rank:

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

$$\rightsquigarrow \left[\begin{array}{ccc} 1 & 2 & 1 \\ 0 & -1 & 3 \\ 0 & 0 & 0 \end{array}\right]$$

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

Exercise 11
Prove the two statements Remark 17(i) and Remark 17(vii) for any matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ by the following steps:

(i) *Show that an elementary row operation does not change the row space of a matrix $\boldsymbol{A}$.*

(ii) *Show that an elementary row operation does not change the column rank of $\boldsymbol{A}$. (Use Exercises 3 and 5).*

(iii) *By reducing $\boldsymbol{A}$ to reduced row echelon form, show using (i) and (ii), that $\mathrm{rk}(\boldsymbol{A}) = \mathrm{rk}(\boldsymbol{A}^\top)$.*

(iv) *Show that an ERO does not change the null space of a matrix. Then, from the row echelon form obtained in (iii) above, deduce Remark 17(vii).*

Exercise 12

(i) *Show that if the vectors $\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_k \in \mathbb{R}^n$ are linearly dependent, then so are the vectors $\boldsymbol{A}\boldsymbol{b}_1, \boldsymbol{A}\boldsymbol{b}_2, \cdots, \boldsymbol{A}\boldsymbol{b}_k \in \mathbb{R}^m$ for any matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$.*

(ii) *From (i) deduce that $\mathrm{rk}(\boldsymbol{AB}) \leq \mathrm{rk}(\boldsymbol{B})$ for any matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{B} \in \mathbb{R}^{n \times k}$.*

(iii) *Using the property $\mathrm{rk}(\boldsymbol{C}) = \mathrm{rk}(\boldsymbol{C}^\top)$, deduce that $\mathrm{rk}(\boldsymbol{AB}) \leq \min\{\mathrm{rk}(\boldsymbol{A}), \mathrm{rk}(\boldsymbol{B})\}$.*

1.7 Intersection of Subspaces

In the following, we determine a basis of the intersection $U_1 \cap U_2$ of two subspaces $U_1, U_2 \subset \mathbb{R}^n$. This means, we are interested in finding all $\boldsymbol{x} \in \mathbb{R}^n$, such that $\boldsymbol{x} \in U_1$ and $\boldsymbol{x} \in U_2$.

Consider $U_1 = \text{span}\{\boldsymbol{b}_1, \dots, \boldsymbol{b}_k\} \subset V$ and $U_2 = \text{span}\{\boldsymbol{c}_1, \dots, \boldsymbol{c}_l\} \subset V$. We know that $\boldsymbol{x} \in U_1$ can be represented as a linear combination $\sum_{i=1}^{k} \lambda_i \boldsymbol{b}_i$ of the basis vectors (or spanning vectors) $\boldsymbol{b}_1, \dots, \boldsymbol{b}_k$. Equivalently $\boldsymbol{x} = \sum_{j=1}^{l} \gamma_j \boldsymbol{c}_j$. Therefore, the approach is to find $\lambda_1, \dots, \lambda_k$ and/or $\gamma_1, \dots, \gamma_l$, such that

$$\sum_{i=1}^{k} \lambda_i \boldsymbol{b}_i = \boldsymbol{x} = \sum_{j=1}^{l} \gamma_j \boldsymbol{c}_j \tag{1.62}$$

$$\Leftrightarrow \sum_{i=1}^{k} \lambda_i \boldsymbol{b}_i - \sum_{j=1}^{l} \gamma_j \boldsymbol{c}_j = \boldsymbol{0}. \tag{1.63}$$

This condition has the form of a homogeneous linear equation system, which we know how to solve: We write the basis vectors into a matrix

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{b}_1, & \cdots & ,\boldsymbol{b}_k, -\boldsymbol{c}_1, & \cdots & ,-\boldsymbol{c}_l, \end{bmatrix} \tag{1.64}$$

and solve the linear equation system

$$\boldsymbol{A} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_k \\ \gamma_1 \\ \vdots \\ \gamma_l \end{bmatrix} = \boldsymbol{0} \tag{1.65}$$

to find either $\lambda_1, \dots, \lambda_k$ or $\gamma_1, \dots, \gamma_l$, which we can then use to determine $U_1 \cap U_2$.

Example

We consider

$$U_1 = \text{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \right\} \subset \mathbb{R}^4, \quad U_2 = \text{span}\left\{ \begin{bmatrix} -1 \\ 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\} \subset \mathbb{R}^4. \tag{1.66}$$

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

$$\sum_{i=1}^{3} \lambda_i \boldsymbol{b}_i = \boldsymbol{x} = \sum_{j=1}^{2} \gamma_j \boldsymbol{c}_j, \tag{1.67}$$

where $\boldsymbol{b}_i$ and $\boldsymbol{c}_j$ are the basis vectors of U_1 and U_2, respectively. The matrix $\boldsymbol{A} = [\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3, -\boldsymbol{c}_1, -\boldsymbol{c}_2]$ from (1.64) is given as

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & -1 & -1 \\ 0 & 1 & 1 & -2 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}. \tag{1.68}$$

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

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -2 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{1.69}$$

We keep in mind that we are interested in finding $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$ and/or $\gamma_1, \gamma_2 \in \mathbb{R}$ with

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -2 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \gamma_1 \\ \gamma_2 \end{bmatrix} = \mathbf{0}. \tag{1.70}$$

From here, we can immediately see that $\gamma_2 = 0$ and $\gamma_1 \in \mathbb{R}$ is a free variable since it corresponds to a non-pivot column, and our solution is

$$U_1 \cap U_2 = \gamma_1 \boldsymbol{c}_1 = \gamma_1 \begin{bmatrix} -1 \\ 1 \\ 2 \\ 0 \end{bmatrix}, \quad \gamma_1 \in \mathbb{R}. \tag{1.71}$$

1.8 Linear Mappings

Earlier on in the course, we said that vectors are objects that can be added together and multiplied by a scalar, and the resulting object is still a vector. We have also seen that a subspace of $\mathbb{R}^n$, being closed under addition and scalar multiplication), shares the linear structure of $\mathbb{R}^n$ and has a basis and a finite dimension. In this course we only deal with subspaces of $\mathbb{R}^n$ for some positive integer n and these subspaces have dimensions bounded above by n. Thus, in this course, instead of presenting an abstract definition of a vector space (which would include the case of infinite dimensional vector spaces), it makes more sense to have a concrete meaning of a vector space that covers all the finite dimensional cases we deal with as follows:

Definition 12 (Vector Spaces)
A (real and finite dimensional) **vector space** *is any subspace of* $\mathbb{R}^n$ *for some positive integer* n.

In the following, we will study mappings on vector spaces that preserve their linear structure (addition and scalar multiplication). Consider two real vector spaces V and W, i.e., $V \subset \mathbb{R}^n$ and $W \subset \mathbb{R}^m$ are subspaces, for two positive integers m and n.

Definition 13 (Linear Mapping)
For vector spaces V, W, a mapping $f : V \to W$ is called **linear** *(or* **vector space homomorphism***) if it preserves the linear structure, i.e.,*

$$\forall \boldsymbol{x}, \boldsymbol{y} \in V \forall \lambda, \gamma \in \mathbb{R} : f(\lambda \boldsymbol{x} + \gamma \boldsymbol{y}) = \lambda f(\boldsymbol{x}) + \gamma f(\boldsymbol{y}). \tag{1.72}$$

Important special cases:

- **Isomorphism:** $f : V \to W$ linear and bijective, V and W are called **isomorphic**
- **Endomorphism:** $f : V \to V$ linear
- **Automorphism:** $f : V \to V$ linear and bijective
- **Identity map:** $I_V : V \to V$, $\boldsymbol{x} \mapsto \boldsymbol{x}$, which is clearly a linear map.

Theorem 2
Vector spaces V and W are isomorphic if and only if $\dim(V) = \dim(W)$.

Exercise 13
Prove Theorem 2 by defining two bases of V and W and using them to define as isomorphism between V and W.

1.8.1 Image and Kernel (Null Space)

Definition 14 (Image and Kernel)

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

$$\ker(f) := f^{-1}(\{\boldsymbol{0}_W\}) = \{\boldsymbol{v} \in V : f(\boldsymbol{v}) = \boldsymbol{0}_W\} \tag{1.73}$$

and the **image/range**

$$\mathrm{Im}(f) := f(V) = \{\boldsymbol{w} \in W | \exists \boldsymbol{v} \in V : f(\boldsymbol{v}) = \boldsymbol{w}\}. \tag{1.74}$$

More informally, the kernel is the set of vectors in $\boldsymbol{v} \in V$ that f maps onto the zero element $\boldsymbol{0}_W$ in W. The image is the set of vectors $\boldsymbol{w} \in W$ that can be "reached" by f from any vector in V. An illustration is given in Figure 1.4.

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

(i) Note that $f(\{\boldsymbol{0}_V\}) = \boldsymbol{0}_W$ and, therefore, $\boldsymbol{0}_v \in \ker(f)$. In particular, the null space is never empty.

(ii) $\mathrm{Im}(f) \subset W$ is a subspace of W, and $\ker(f) \subset V$ is a subspace of V.

(iii) f is injective (one-to-one) if and only if $\ker(f) = \{\boldsymbol{0}\}$

Exercise 14
Prove Remark18(ii).

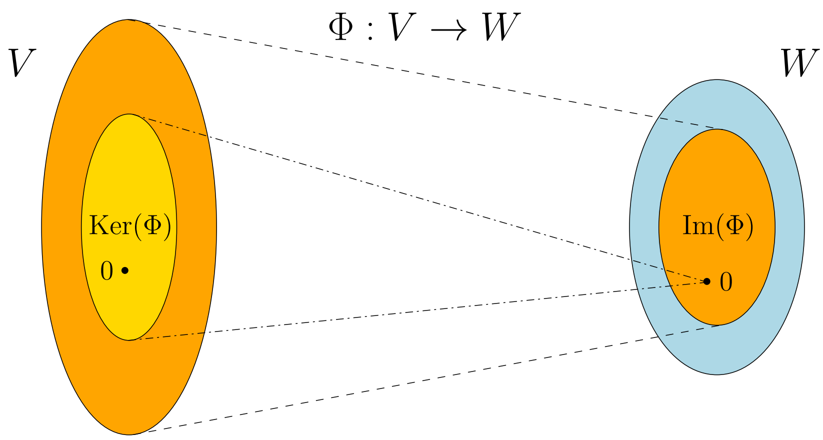


Figure 1.4: Kernel and Image of a linear mapping $f : V \to W$.

1.8.2 Linear maps and Matrices

What is the relationship between linear maps and matrices?

Definition 15

(i) If we have a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, then $\boldsymbol{A}$ defines a linear map $\phi_{\boldsymbol{A}} : \mathbb{R}^n \to \mathbb{R}^m$ which is given in the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ by $\phi_{\boldsymbol{A}}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$.

*(ii) Conversely, suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map and take any basis of $\mathbb{R}^n$, say $\boldsymbol{b}_j$ where $1 \le j \le n$, and any basis of $\mathbb{R}^m$, say $\boldsymbol{c}_i$ where $1 \le i \le m$, and let $a_{ij} \in \mathbb{R}$ for $1 \le j \le n$ such that $f(\boldsymbol{b}_j) = \sum_{i=1}^m a_{ij}\boldsymbol{c}_i$. Then the matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ with $\boldsymbol{A}_{ij} = a_{ij}$ for $1 \le j \le n$ and $1 \le i \le m$ is called the **matrix representation** of f **with respect to** the basis $\boldsymbol{b}_j$ $(1 \le j \le n)$ and the basis $\boldsymbol{c}_i$ $(1 \le i \le m)$.*

Exercise 15

Check that $\phi_{\boldsymbol{A}}$ in Definition 15(i) is a linear map and that $\boldsymbol{A}$ in Definition 15(ii) correctly represents f with respect to the two bases.

For $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, consider the mapping $f_{\boldsymbol{A}} : \mathbb{R}^n \to \mathbb{R}^m$ defined in the standard coordinate system of $\mathbb{R}^n$ and $\mathbb{R}^m$ by $f_{\boldsymbol{A}}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$.

Definition 16 (Null Space and Column Space of a matrix)

*The **column** or image space of $\boldsymbol{A}$ is defined to be $\mathrm{Im}\boldsymbol{A} := \mathrm{Im} f_{\boldsymbol{A}}$. The **kernel or null space** of $\boldsymbol{A}$ is defined to be $\ker \boldsymbol{A} := \ker f_{\boldsymbol{A}}$*

Remark 19

- *For $\boldsymbol{A} = [\boldsymbol{a}_1, ..., \boldsymbol{a}_n]$ we obtain*

$$\mathrm{Im}(f_{\boldsymbol{A}}) = \{\boldsymbol{A}\boldsymbol{x} : \boldsymbol{x} \in \mathbb{R}^n\} = \{x_1\boldsymbol{a}_1 + ... + x_n\boldsymbol{a}_n : x_1, \dots, x_n \in \mathbb{R}\} \subset \mathbb{R}^m, \tag{1.75}$$

  i.e., $\mathrm{Im}\boldsymbol{A} = \mathrm{Im} f_{\boldsymbol{A}}$ is the span of the columns of $\boldsymbol{A}$. Therefore, the column (image) space of $\boldsymbol{A}$ is a subspace of $\mathbb{R}^m$, where m is the "height" of the matrix.

- *The kernel/null space $\ker \boldsymbol{A} = \ker(f_{\boldsymbol{A}})$ is a subspace of $\mathbb{R}^n$, where n is the "width" of the matrix, and is the set of all solutions $\boldsymbol{x}$ to the linear homogeneous equation system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$.*

Example (Image and Kernel of a Linear Mapping)
The mapping

$$f : \mathbb{R}^4 \to \mathbb{R}^2, \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \mapsto \begin{bmatrix} 1 & 2 & -1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_1 + 2x_2 - x_3 \\ x_1 + x_4 \end{bmatrix} \tag{1.76}$$

$$= x_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{1.77}$$

is linear. Since the third and fourth columns form a basis, we have $\mathrm{Im}(f) = \mathbb{R}^2$.
To compute the kernel (null space) of f, we need to solve $\boldsymbol{Ax} = \boldsymbol{0}$, i.e., we need to solve a homogeneous equation system. To do this, we use Gaussian elimination to transform $\boldsymbol{A}$ into reduced row echelon form:

$$\left[\begin{array}{cccc} 1 & 2 & -1 & 0 \\ 1 & 0 & 0 & 1 \end{array}\right] \rightsquigarrow \left[\begin{array}{cccc} 1 & 0 & 0 & 1 \\ 1 & 2 & -1 & 0 \end{array}\right] \begin{array}{l} \\ -R_1 | \cdot (\frac{1}{2}) \end{array} \rightsquigarrow \left[\begin{array}{cccc} 1 & 0 & 0 & 1 \\ 0 & 1 & -\frac{1}{2} & -\frac{1}{2} \end{array}\right]$$

This matrix is now in reduced row echelon form, and we can solve it by expressing basic variables $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in terms of free variables $\boldsymbol{x}_3$ and $\boldsymbol{x}_4$: we have $x_1 = -x_4$ and $x_2 = (x_3 + x_4)/2$. Putting $x_3 = 0$ we get $\boldsymbol{x} = x_4[-1, 1/2, 0, 1]^\top$ and putting $x_4 = 0$ we get $\boldsymbol{x} = x_3[0, 1, 2, 0]^\top$. Thus, we find the general solution, i.e., the kernel (null space) of $\boldsymbol{A}$, as

$$\ker(f) = \mathrm{span}\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \\ 0 \end{bmatrix} \right\}. \tag{1.78}$$

Theorem 3 (Rank-Nullity Theorem)
For vector spaces V, W and a linear mapping $f : V \to W$, we have:

$$\dim(\ker(f)) + \dim(\mathrm{Im}(f)) = \dim(V) \tag{1.79}$$

Remark 20
Consider vector spaces V, W, X. Then:

- *For linear mappings $f : V \to W$ and $g : W \to X$ the mapping $g \circ f : V \to X$ is also linear.*
- *If $f : V \to W$ is an isomorphism then $f^{-1} : W \to V$ is an isomorphism as well.*
- *If $f : V \to W$, $g : V \to W$ are linear then $f + g$ and $\lambda f, \lambda \in \mathbb{R}$ are linear, too.*

Exercise 16
Prove the three statements in Remark 20.

Exercise 17
Prove Theorem 3 using Exercise 11(iv).

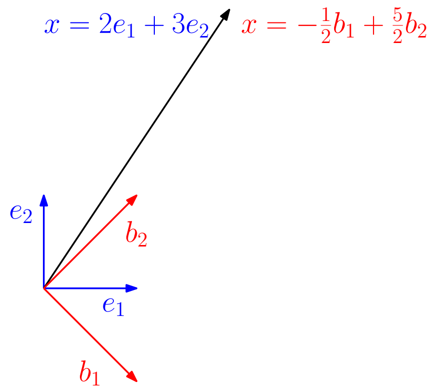


Figure 1.5: The coordinates of the vector $\boldsymbol{x}$ are the coefficients of the linear combination of the basis vectors to represent $\boldsymbol{x}$.

1.8.3 Matrix Representation of Linear Mappings

We have already seen in Definition 15 how a linear map $f : \mathbb{R}^n \to \mathbb{R}^m$ can be represented by a matrix $\boldsymbol{A}$ in the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$. In this section, we seek to generalise this to any linear map between two vector spaces. Any n-dimensional vector space is isomorphic to $\mathbb{R}^n$ (Theorem 2). If we define a basis $\{\boldsymbol{b}_1, \dots, \boldsymbol{b}_n\}$ of V we can construct an isomorphism concretely. In the following, the order of the basis vectors will be important. Therefore, we write

$$B = (\boldsymbol{b}_1, \dots, \boldsymbol{b}_n) \tag{1.80}$$

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

Definition 17 (Coordinates)
Consider a vector space V and an ordered basis $B = (\boldsymbol{b}_1, \dots, \boldsymbol{b}_n)$. For $\boldsymbol{x} \in V$ we obtain a unique representation (linear combination)

$$\boldsymbol{x} = \alpha_1 \boldsymbol{b}_1 + \dots + \alpha_n \boldsymbol{b}_n \tag{1.81}$$

of $\boldsymbol{x}$ with respect to B. Then $\alpha_1, \dots, \alpha_n$ are the **coordinates** *of $\boldsymbol{x}$ with respect to B and the vector*

$$\boldsymbol{x}_B = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \in \mathbb{R}^n \tag{1.82}$$

is the **coordinate vector/coordinate representation** *of $\boldsymbol{x}$ with respect to B. If the basis B is fixed, we write V_B for the collection of the coordinates of V in basis B, i.e., as a representation of V in the basis B.*

Remark 21
Intuitively, the basis vectors can be understood as units (including rather odd units, such as "apples", "bananas", "kilograms" or "seconds"). However, let us have a look at a geometric vector $\boldsymbol{x} \in \mathbb{R}^2$ with the coordinates

$$\boldsymbol{x}_E = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \tag{1.83}$$

with respect to the standard ordered basis $E = (\boldsymbol{e}_1, \boldsymbol{e}_2)$ *in* $\mathbb{R}^2$. *This means, we have* $\boldsymbol{x} = 2\boldsymbol{e}_1 + 3\boldsymbol{e}_2$. *However, we do not have to choose the standard basis to represent this vector. For the ordered basis* $B = (\boldsymbol{b}_1, \boldsymbol{b}_2)$ *with* $\boldsymbol{b}_1 = [1, -1]^\top, \boldsymbol{b}_2 = [1, 1]^\top$, *we have* $\boldsymbol{x} = 2\boldsymbol{e}_1 + 3\boldsymbol{e}_2 = -\frac{1}{2}\boldsymbol{b}_1 + \frac{5}{2}\boldsymbol{b}_2$ *and obtain the coordinates*

$$\boldsymbol{x}_B = \begin{bmatrix} -1/2 \\ 5/2 \end{bmatrix}$$

to represent the same vector $\boldsymbol{x}$ *with respect to the ordered basis* B *(see Fig. 1.5).*

In the following, we will be looking at mappings that transform coordinate vectors with respect to one basis into coordinate vectors with respect to a different basis.

Remark 22
For an n*-dimensional vector space* V *and a basis* B *of* V, *the mapping* $f : \mathbb{R}^n \to V$, $f(\boldsymbol{e}_i) = \boldsymbol{b}_i$, $i = 1, \dots, n$, *is linear (and because of Theorem 2 an isomorphism), where* $(\boldsymbol{e}_1, \dots, \boldsymbol{e}_n)$ *is the standard basis of* $\mathbb{R}^n$.

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

Definition 18 (Transformation matrix)
Consider vector spaces V, W *with corresponding (ordered) bases* $B = (\boldsymbol{b}_1, \dots, \boldsymbol{b}_n)$ *and* $D = (\boldsymbol{d}_1, \dots, \boldsymbol{d}_m)$. *Moreover, we consider a linear map* $\Phi : V \to W$. *For* $j \in \{1, \dots, n\}$

$$\Phi(\boldsymbol{b}_j) = \alpha_{1j}\boldsymbol{d}_1 + \cdots + \alpha_{mj}\boldsymbol{d}_m = \sum_{i=1}^{m} \alpha_{ij}\boldsymbol{d}_i \tag{1.84}$$

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

$$\boldsymbol{\Phi}_{DB} := ((\alpha_{ij})) \tag{1.85}$$

the **transformation matrix** *of* Φ *with respect to the ordered bases* B *of* V *and* D *of* W, *which has* V_B *as input space and* W_D *as output space, i.e.,* $\boldsymbol{\Phi}_{DB} : V_B \to W_D$ *that takes the coordinates* $\boldsymbol{x}_B \in V_B$ *of a vector* $\boldsymbol{x} \in V$ *to the vector* $\boldsymbol{y} = \Phi(\boldsymbol{x}) \in W$ *whose coordinates* $\boldsymbol{y}_D$ *with respect to the ordered basis* D *is given by* $\boldsymbol{\Phi}_{DB}(\boldsymbol{x}_B) \in W_D$.

Remark 23

(i) *The coordinates of* $\Phi(\boldsymbol{b}_j)$ *with respect to the ordered basis* D *of* W *are the* j*-th column of* $\boldsymbol{\Phi}_{DB}$.

(ii) $\mathrm{rk}(\boldsymbol{\Phi}_{DB}) = \dim(\mathrm{Im}(\Phi))$

(iii) *Consider vector spaces* V, W *with ordered bases* B, D *and a linear mapping* $\Phi : V \to W$ *with transformation matrix* $\boldsymbol{\Phi}_{DB}$. *Since* $\boldsymbol{x}_B$ *is the coordinate vector of* $\boldsymbol{x} \in V$ *with respect to the ordered basis* B *and* $\boldsymbol{y}_D$ *the coordinate vector of* $\boldsymbol{y} = \Phi_{DB}(\boldsymbol{x}) \in W$ *with respect to* D, *then*

$$\boldsymbol{y}_D = \boldsymbol{\Phi}_{DB}\boldsymbol{x}_B. \tag{1.86}$$

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

Note that in Φ_{DB} we are taking coordinates with respect to B (first from right in the subscript DB) to coordinates with respect to D (second from right in the subscript DB). Clearly, Definition 15(ii) is a particular case of Definition 18.

Example Transformation Matrix
Consider a homomorphism $\Phi : V \to W$ and ordered bases $B = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_3)$ of V and $D = (\boldsymbol{d}_1, \ldots, \boldsymbol{d}_4)$ of W. With

$$\begin{aligned} \Phi(\boldsymbol{b}_1) &= \boldsymbol{d}_1 - \boldsymbol{d}_2 + 3\boldsymbol{d}_3 - \boldsymbol{d}_4 \\ \Phi(\boldsymbol{b}_2) &= 2\boldsymbol{d}_1 + \boldsymbol{d}_2 + 7\boldsymbol{d}_3 + 2\boldsymbol{d}_4 \\ \Phi(\boldsymbol{b}_3) &= 3\boldsymbol{d}_2 + \boldsymbol{d}_3 + 4\boldsymbol{d}_4 \end{aligned} \tag{1.87}$$

the transformation matrix $((\alpha_{ik}))$ with respect to B and D satisfies $\Phi(\boldsymbol{b}_k) = \sum_{i=1}^{4} \alpha_{ik}\boldsymbol{d}_i$ for $k = 1, \ldots, 3$ and is given as

$$\boldsymbol{\Phi}_{DB} = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3] = \begin{bmatrix} 1 & 2 & 0 \\ -1 & 1 & 3 \\ 3 & 7 & 1 \\ -1 & 2 & 4 \end{bmatrix}, \tag{1.88}$$

where the $\boldsymbol{\alpha}_j,\ j = 1, 2, 3,$ are the coordinate vectors of $\Phi(\boldsymbol{b}_j)$ with respect to D.

Exercise 18
Check the correctness of Remark 23(iii).

1.8.4 Transformation matrix for basis Change

Suppose V is a vector space and

$$B = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n), \quad \tilde{B} = (\tilde{\boldsymbol{b}}_1, \ldots, \tilde{\boldsymbol{b}}_n) \tag{1.89}$$

are two ordered bases of V with B considered as the old basis and $\tilde{B}$ as the new basis. Since B is a basis, any basis element in $\tilde{B}$ can be expressed as a linear combination of vectors in B, i.e., for $j = 1, \ldots, n$ we have:

$$\tilde{\boldsymbol{b}}_j = s_{1j}\boldsymbol{b}_1 + \cdots + s_{nj}\boldsymbol{b}_n, \quad j = 1, \ldots, n. \tag{1.90}$$

for some matrix $\boldsymbol{S} = ((s_{ij}) \in \mathbb{R}^{n \times n}$. Comparing this with the definition of the transformation matrix of a linear map in Definition 18, we see that $\boldsymbol{S}$ is the transformation matrix of the identity map represented in the basis B and $\tilde{B}$. Thus,

$$\boldsymbol{S} = \boldsymbol{I}_{B\tilde{B}},$$

where the letter $\boldsymbol{I}$ in $\boldsymbol{I}_{B\tilde{B}}$ stands for "Identity map" and the subscript "$B\tilde{B}$" means the transformation is taking the coordinates with respect to the new basis $\tilde{B}$ (which from the right hand side in $B\tilde{B}$ comes first) to the coordinates with respect to old basis B (which comes second from the right hand side). Since it is clear from the context that B and $\tilde{B}$ are bases of V, we do not need to mention V explicitly in $I_{B\tilde{B}}$ and thus can avoid cluttering the notation.

1.8.5 Composition of linear maps

Consider vector spaces V, W, X respectively with dimensions n, k and m with, respectively, ordered bases $B = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$, $C = (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_k)$ and $D = (\boldsymbol{d}_1, \ldots, \boldsymbol{d}_m)$. From Remark 20 we already know that for linear mappings $\Phi : V \to W$ and $\Psi : W \to X$ the mapping $\Psi \circ \Phi : V \to X$ is also linear. We thus have transformation matrices Φ_{CB}, Ψ_{DC} and $(\Psi \circ \Phi)_{DB}$. What is the relationship between the matrix $(\Psi \circ \Phi)_{DB}$ and the two matrices Ψ_{DC} and Φ_{CB}? As one may expect, the first matrix is the product of the latter two matrices:

Theorem 4

$$(\Psi \circ \Phi)_{DB} = \Psi_{DC}\Phi_{CB}.$$

Proof By definition of the transformation matrix for Φ and Ψ, for each $j = 1, \ldots, n$, we have:

$$\Phi(\boldsymbol{b}_j) = \sum_{\ell=1}^{k} (\Phi_{CB})_{\ell j} \boldsymbol{c}_\ell$$

which implies by applying Ψ to both sides, using the linearity of Ψ and the definition of the transformation matrix for Ψ that

$$\begin{aligned}(\Psi \circ \Phi)(\boldsymbol{b}_j) = \Psi(\Phi(\boldsymbol{b}_j)) &= \textstyle\sum_{\ell=1}^{k} (\Phi_{CB})_{\ell j} \Psi(\boldsymbol{c}_\ell) = \sum_{\ell=1}^{k} (\Phi_{CB})_{\ell j} \sum_{i=1}^{m} (\Psi_{DC})_{i\ell} \boldsymbol{d}_i \\ &= \textstyle\sum_{i=1}^{m} \sum_{\ell=1}^{k} (\Psi_{DC})_{i\ell} (\Phi_{CB})_{\ell j} \boldsymbol{d}_i = \sum_{i=1}^{m} (\Psi_{DC}\Phi_{CB})_{ij} \boldsymbol{d}_i\end{aligned}$$

Thus, by the definition of the transformation matrix for $\Psi \circ \Phi$, we obtain:

$$((\Psi \circ \Phi)_{DB})_{ij} = (\Psi_{DC}\Phi_{CB})_{ij}$$

and it follows that $(\Psi \circ \Phi)_{DB} = \Psi_{DC}\Phi_{CB}$ as required.

Exercise 19
Find I_{BB} and then show that $I_{\tilde{B}B} = (I_{B\tilde{B}})^{-1}$.

1.8.6 Transformation matrix after bases change

Now we can see how the transformation matrix of a linear map $\Phi : V \to W$ changes if we change the bases both in V and W. Consider

$$B = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n), \quad \tilde{B} = (\tilde{\boldsymbol{b}}_1, \ldots, \tilde{\boldsymbol{b}}_n) \tag{1.91}$$

ordered bases of V, where, as above, B is the old basis and $\tilde{B}$ is the new basis with $I_{B\tilde{B}}$ the transformation matrix from basis $\tilde{B}$ to basis B. Similarly,

$$D = (\boldsymbol{d}_1, \ldots, \boldsymbol{d}_m), \quad \tilde{D} = (\tilde{\boldsymbol{d}}_1, \ldots, \tilde{\boldsymbol{d}}_m) \tag{1.92}$$

ordered bases of W, where D is the old basis and $\tilde{D}$ is the new basis with $I_{\tilde{D}D}$ the transformation matrix from basis D to basis $\tilde{D}$. Moreover, $\boldsymbol{\Phi}_{DB}$ is the transformation matrix of the linear mapping $\Phi : V \to W$ with respect to the old bases B and D, and

$$\begin{array}{ll} \text{Vector spaces} & V \xrightarrow{\Phi} W \\ \\ \text{Old coordinates} & \begin{tikzcd} V_B \arrow[r, "\boldsymbol{\Phi}_{DB}"] & W_D \arrow[d, "\boldsymbol{I}_{\tilde{D}D}"] \\ V_{\tilde{B}} \arrow[u, "\boldsymbol{I}_{B\tilde{B}}"] \arrow[r, "\boldsymbol{\Phi}_{\tilde{D}\tilde{B}}"'] & W_{\tilde{D}} \end{tikzcd} \\ \text{New coordinates} & \end{array}$$

Figure 1.6: For a homomorphism $\Phi : V \to W$ and ordered old bases B, D respectively of of V, W and new ordered bases $\tilde{B}, \tilde{D}$ respectively of V, W, there are two different ways of going from $V_{\tilde{B}}$ to $W_{\tilde{D}}$ to obtain the image of a vector in V as the arrows in the digaram shows: Thus we can express the matrix $\boldsymbol{\Phi}_{\tilde{D}\tilde{B}}$, representing the map Φ with respect to the new bases $\tilde{B}, \tilde{D}$ equivalently by a product of matrices $\boldsymbol{\Phi}_{\tilde{D}\tilde{B}} = \boldsymbol{I}_{\tilde{D}D}\boldsymbol{\Phi}_{DB}\boldsymbol{I}_{B\tilde{B}}$.

$\boldsymbol{\Phi}_{\tilde{D}\tilde{B}}$ is the corresponding transformation mapping with respect to the new basis $\tilde{B}$ and $\tilde{D}$. What is the relationship between $\boldsymbol{\Phi}_{DB}$ and $\boldsymbol{\Phi}_{\tilde{D}\tilde{B}}$?
Assume $\boldsymbol{x} \in V$ with $\boldsymbol{y} := \Phi(\boldsymbol{x}) \in W$. Note that $\boldsymbol{x} \in V$ and $\Phi(\boldsymbol{x}) \in W$ are geometric vectors and are independent of any basis we use in V or in W. If we choose the coordinates $\boldsymbol{x}_{\tilde{B}}$ of $\boldsymbol{x}$ with respect to the new basis $\tilde{B}$ then there are two ways to obtain the coordinates $\boldsymbol{y}_{\tilde{D}}$ of $\boldsymbol{y} = \Phi(\boldsymbol{x})$ with respect to $\tilde{D}$; see Figure 1.6. These two ways are as follows: (i) Going directly from coordinates with respect to $\tilde{B}$ to coordinates with respect to $\tilde{D}$, i.e, with $\boldsymbol{y}_{\tilde{D}} = \Phi_{\tilde{D}\tilde{B}}$, and (ii) Going via three maps: changing first the coordinates with respect the basis $\tilde{B}$ to coordinates with respect to B, then applying Φ_{DB} to get the coordinates of $\boldsymbol{y}$ with respect to $\tilde{D}$ and finally changing the coordinates from D to $\tilde{D}$, i.e., $\boldsymbol{y}_{\tilde{D}} = \boldsymbol{I}_{\tilde{D}D}\Phi_{DB}\boldsymbol{I}_{B\tilde{B}}$. Since $\boldsymbol{y}_{\tilde{D}}$ is uniquely defined by Remark 14(iii), the two results are the same and we have actually proved:

Theorem 5

$$\Phi_{\tilde{D}\tilde{B}} = \boldsymbol{I}_{\tilde{D}D}\Phi_{DB}\boldsymbol{I}_{B\tilde{B}}.$$

Recall that $\boldsymbol{I}_{B\tilde{B}} = (\boldsymbol{I}_{\tilde{B}B})^{-1}$ and $\boldsymbol{I}_{D\tilde{D}} = (\boldsymbol{I}_{\tilde{D}D})^{-1}$. This relation and Theorem 5 motivates the following:

Definition 19 (Equivalence)
Two matrices $\boldsymbol{A}, \tilde{\boldsymbol{A}} \in \mathbb{R}^{m\times n}$ *are* **equivalent** *if there exist invertible matrices* $\boldsymbol{S} \in \mathbb{R}^{n\times n}$ *and* $\boldsymbol{T} \in \mathbb{R}^{m\times m}$*, such that* $\tilde{\boldsymbol{A}} = \boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{S}$.

Note that equivalence of matrices is an equivalence relation (why?).

Corollary 1
$\boldsymbol{\Phi}_{\tilde{D}\tilde{B}}$ *and* $\boldsymbol{\Phi}_{DB}$ *are equivalent.*

Definition 20 (Similarity)
Two matrices $\boldsymbol{A}, \tilde{\boldsymbol{A}} \in \mathbb{R}^{n\times n}$ *are* **similar** *if there exists a invertible matrix* $\boldsymbol{S} \in \mathbb{R}^{n\times n}$ *with* $\tilde{\boldsymbol{A}} = \boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}$

Remark 24
Similar matrices are always equivalent. However, equivalent matrices are not necessarily similar.

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

$$\boldsymbol{\Phi}_{CB} = \begin{bmatrix} 1 & 2 & 0 \\ -1 & 1 & 3 \\ 3 & 7 & 1 \\ -1 & 2 & 4 \end{bmatrix} \tag{1.93}$$

with respect to the standard ordered bases

$$B = \left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right), \quad C = \left(\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right). \tag{1.94}$$

We seek the transformation matrix $\Phi_{\tilde{C}\tilde{B}}$ of Φ with respect to the new ordered bases

$$\tilde{B} = \left(\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right), \quad \tilde{C} = \left(\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right). \tag{1.95}$$

Then, it is easy to compute

$$\boldsymbol{I}_{B\tilde{B}} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \qquad \boldsymbol{I}_{C\tilde{C}} = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{1.96}$$

where the ith column of $\boldsymbol{S}$ is the coordinate representation of $\tilde{\boldsymbol{b}}_i$ in terms of the basis vectors of B.[18] Similarly, the jth column of $\boldsymbol{T}$ is the coordinate representation of $\tilde{\boldsymbol{c}}_j$ in terms of the basis vectors of $\boldsymbol{C}$. Since $\boldsymbol{\Phi}_{\tilde{C}\tilde{B}} = \boldsymbol{I}_{\tilde{C}C}\boldsymbol{\Phi}_{CB}\boldsymbol{I}_{B\tilde{B}}$ and $\boldsymbol{I}_{\tilde{C}C} = \boldsymbol{I}^{-1}_{C\tilde{C}}$, we compute $\boldsymbol{I}_{\tilde{C}C}$ using ERO:

$$\boldsymbol{I}_{\tilde{C}C} = (\boldsymbol{I}_{C\tilde{C}})^{-1} = \frac{1}{2}\begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

[18] Since B is the standard basis, this representation is straightforward to find. For a general basis B we would need to solve a linear equation system to find the λ_i such that $\sum_{i=1}^{3} \lambda_i \boldsymbol{b}_i = \tilde{\boldsymbol{b}}_j$, $j = 1, \ldots, 3$.

Therefore, we obtain

$$\boldsymbol{\Phi}_{\tilde{C}\tilde{B}} = \boldsymbol{I}_{\tilde{C}C}\boldsymbol{\Phi}_{CB}\boldsymbol{I}_{B\tilde{B}} = \begin{bmatrix} 1/2 & 1/2 & -1/2 & -1/2 \\ 1/2 & -1/2 & 1/2 & -1/2 \\ -1/2 & 1/2 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 2 & 1 \\ 0 & 4 & 2 \\ 10 & 8 & 4 \\ 1 & 6 & 3 \end{bmatrix} = \begin{bmatrix} -4 & -4 & -2 \\ 6 & 0 & 0 \\ 4 & 8 & 4 \\ 1 & 6 & 3 \end{bmatrix}. \tag{1.97}$$

Therefore the following two matrices are equivalent

$$\begin{bmatrix} 1 & 2 & 0 \\ -1 & 1 & 3 \\ 3 & 7 & 1 \\ -1 & 2 & 4 \end{bmatrix} \in \mathbb{R}^{4\times 3} \qquad \begin{bmatrix} -4 & -4 & -2 \\ 6 & 0 & 0 \\ 4 & 8 & 4 \\ 1 & 6 & 3 \end{bmatrix} \in \mathbb{R}^{4\times 3}$$

Soon, we will be able to exploit the concept of a basis change to find a basis with respect to which the transformation matrix of an endomorphism has a particularly simple (diagonal) form.

1.9 Determinants

The determinant of a square matrix is a key concept in linear algebra. For instance, it indicates whether a square matrix can be inverted or we can use it to check for linear independence of the columns of such a matrix. A geometric intuition is that the absolute value of the determinant of real vectors is equal to the volume of the parallelepiped spanned by those vectors. Determinants will play a very important role for determining eigenvalues and eigenvectors (Section 1.10).
Determinants are only defined for square matrices $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, and we write $\det(\boldsymbol{A})$ or $|\boldsymbol{A}|$.

Remark 25

- *For $n = 1$, $\det(\boldsymbol{A}) = \det(a_{11}) = a_{11}$ with $|\det \boldsymbol{A}| = \det(a_{11})| = |a_{11}|$ the length of the 1-dim vector $a_{11} \in \mathbb{R}$.*

- *For $n = 2$,*

$$\det(\boldsymbol{A}) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21} \tag{1.98}$$

  with $|\det \boldsymbol{A}| = |a_{11}a_{22} - a_{12}a_{21}|$, i.e., the area of the parallelogram induced by the columns $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ of $\boldsymbol{A}$.

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

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23} \tag{1.99}$$

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

with $|\det \boldsymbol{A}|$ *being the volume of the parallelepiped induced by the three columns of* $\boldsymbol{A}$.

- *In general, there is a recursive formula to define the determinant of a square matrix* $\boldsymbol{A} = ((a_{ij})) \in \mathbb{R}^{n\times n}$ *to the computation of the determinants of* n *square matrices of dimension* $n-1$*, i.e.,* n *matrices in* $\mathbb{R}^{(n-1)\times(n-1)}$*. This recursive formula is called the* **co-factor** *method.*

  We define $\boldsymbol{A}_{i,j}$ *to be the matrix that remains if we delete the* i*th row and the* j*th column from* $\boldsymbol{A}$*. Then, for all* $i = 1,\dots,n$*, the determinant* $\det(A)$ *is computed recursively by "Expansion along the first row", i.e.,*

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

  The scalar $C_{1k} := (-1)^{1+k} \det(A_{1,k})$ *is also called the* **co-factor** *for* $1k$ *entry of* $\boldsymbol{A}$*. The absolute value* $|\det \boldsymbol{A}|$ *is the* n*-dimensional volume of the parallelepiped induced by the* n *columns of* $\boldsymbol{A}$.

- *Let's see the co-factor method for* $n = 3$ *expanding along the first row.*

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

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

  i.e., we go through the elements of the first row of $\boldsymbol{A}$ *one by one, and each time we multiply that element by the determinant of the* $n-1$ *dimensional matrix obtained from* $\boldsymbol{A}$ *when we remove the first row and the column of the element under consideration. Then the* n *numbers that are resulted are added with alternating signs. The coefficient of* a_{1i} *in the sum is the* **co-factor** C_{1i} *corresponding to the* $1i$ *entry, i.e.,* $|\boldsymbol{A}| = a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13}$ *with*

$$C_{11} = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} \qquad C_{12} = -\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} \qquad C_{13} = \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

- *Thus for* $\boldsymbol{A} \in \mathbb{R}^{4\times 4}$ *we have*

$$|\boldsymbol{A}| = a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13} + a_{14}C_{14}$$

  where C_{1j} *is* $(-1)^{1+j}$ *times the determinant of the 3-dimensional matrix with the first row and column* j *removed.*

  We will see that in practice there are easier ways of computing the determinant for higher dimensions.

- *It follows from the recursive formula that for an upper or a lower triangular matrix $\boldsymbol{A}$, the determinant is the product of the diagonal elements:* $\det(\boldsymbol{A}) = \prod_{i=1}^{n} a_{ii}$

Theorem 6
The determinant $\det(A)$ *can be computed by "Expansion along ith row" for any* $1 \leq i \leq n$*, i.e.,*

$$\det(\boldsymbol{A}) = \sum_{k=1}^{n} (-1)^{i+k} a_{ik} \det(\boldsymbol{A}_{i,k}).$$

Proof This can be proved by induction on n but the proof is rather long.[19]

Remark 26 (Properties of Determinants)

(i) The determinant is linear in each column and in each row, i.e., it is a multi-linear function of the columns and the rows: E.g., we spell out multi-linearity with respect to the columns of the matrix $\boldsymbol{A} = ((a_{ij})) \in \mathbb{R}^{n\times n}$*. Let* $\boldsymbol{a}_j$ *denote the* j *the column, then*

$$\det([\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots \boldsymbol{a}_{j-1}, \lambda \boldsymbol{a}_j + \lambda' \boldsymbol{a}'_j, \boldsymbol{a}_{j+1} \cdots, \boldsymbol{a}_n])$$

$$= \lambda \det([\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots \boldsymbol{a}_{j-1}, \boldsymbol{a}_j, \boldsymbol{a}_{j+1} \cdots, \boldsymbol{a}_n]) + \lambda' \det([\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_{j-1}, \boldsymbol{a}'_j, \boldsymbol{a}_{j+1} \cdots, \boldsymbol{a}_n]).$$

(ii) If two rows or two columns of $\boldsymbol{A}$ *are swapped then the determinant changes sign.*

(iii) If two rows or two columns of $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ *are the same then* $\det(\boldsymbol{A}) = 0$.

(iv) $\det(\boldsymbol{A}) = \det(\boldsymbol{A}^\top)$*. This is equivalent (using Theorem 6) to saying that* $\det(A)$ *can be computed by "Expansion along* j*th column" for* $1 \leq j \leq n$*:*

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

(v) Adding a multiple of a column/row to another one does not change $\det(\boldsymbol{A})$.

(vi) $\det(\boldsymbol{A}) = 0 \Leftrightarrow \boldsymbol{A}$ *is singular (not invertible). Alternatively:* $\boldsymbol{A}$ *is invertible* $\Leftrightarrow \det(\boldsymbol{A}) \neq 0$.

(vii) $\det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det(\boldsymbol{B})$

(viii) If $\boldsymbol{A}$ *is invertible then* $\det(\boldsymbol{A}^{-1}) = 1/\det(\boldsymbol{A})$.

(ix) Similar matrices possess the same determinant, a property which follows from (iv) above. Therefore, for a linear mapping $\Phi : V \to V$ *all transformation matrices* $\boldsymbol{\Phi}_{BB}$ *of* Φ *with respect to any basis* B *have the same determinant.*

(x) If $\boldsymbol{A}$ *is invertible, then* $A^{-1} = \frac{1}{\det(A)} \boldsymbol{C}^\top$ *where* $\boldsymbol{C} = (C_{ij})$ *is the matrix of co-factors of* $\boldsymbol{A}$.

[19] For a proof see http://linear.ups.edu/html/section-DM.html

Because of the properties stated in Remark 26(i), (ii) and (v), we can use Gaussian elimination to compute $\det(\boldsymbol{A})$. However, we need to pay attention to swapping the sign when swapping rows.

Example

$$\begin{vmatrix} 2 & 0 & 1 & 2 & 0 \\ 2 & -1 & 0 & 1 & 1 \\ 0 & 1 & 2 & 1 & 2 \\ -2 & 0 & 2 & -1 & 2 \\ 2 & 0 & 0 & 1 & 1 \end{vmatrix} = \begin{vmatrix} 2 & 0 & 1 & 2 & 0 \\ 0 & -1 & -1 & -1 & 1 \\ 0 & 1 & 2 & 1 & 2 \\ 0 & 0 & 3 & 1 & 2 \\ 0 & 0 & -1 & -1 & 1 \end{vmatrix} = \begin{vmatrix} 2 & 0 & 1 & 2 & 0 \\ 0 & -1 & -1 & -1 & 1 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 3 & 1 & 2 \\ 0 & 0 & -1 & -1 & 1 \end{vmatrix} \tag{1.100}$$

$$= \begin{vmatrix} 2 & 0 & 1 & 2 & 0 \\ 0 & -1 & -1 & -1 & 1 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & -7 \\ 0 & 0 & 0 & -1 & 4 \end{vmatrix} = \begin{vmatrix} 2 & 0 & 1 & 2 & 0 \\ 0 & -1 & -1 & -1 & 1 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & -7 \\ 0 & 0 & 0 & 0 & -3 \end{vmatrix} = 6 \tag{1.101}$$

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

Remark 27 (Checkerboard Pattern)
The signs +/− resulting from $(-1)^{k+j}$ *follow a checkerboard pattern, e.g., for a* 3×3-*matrix it looks as follows:*

$$\begin{bmatrix} + & - & + \\ - & + & - \\ + & - & + \end{bmatrix}. \tag{1.102}$$

Example (Expansion along Rows)
In the following, we will go through an example of how we can simplify a determinant by expanding along a row. In particular, we expand along the first row in the following example (the Sarrus rule can be recovered from this):

$$\begin{vmatrix} 1 & -2 & -3 \\ 1 & 0 & 1 \\ 2 & 4 & 4 \end{vmatrix} = 1(-1)^{1+1}\begin{vmatrix} 0 & 1 \\ 4 & 4 \end{vmatrix} + (-2)(-1)^{1+2}\begin{vmatrix} 1 & 1 \\ 2 & 4 \end{vmatrix} + (-3)(-1)^{1+3}\begin{vmatrix} 1 & 0 \\ 2 & 4 \end{vmatrix} \tag{1.103}$$

$$= -4 + 2(4-2) - 3 \cdot 4 = -4 + 4 - 12 = -12 \tag{1.104}$$

It is usually advisable to expand along rows/columns with many 0 entries.

Example
Let us re-compute the example in (1.100)

$$\begin{vmatrix} 2 & 0 & 1 & 2 & 0 \\ 2 & -1 & 0 & 1 & 1 \\ 0 & 1 & 2 & 1 & 2 \\ -2 & 0 & 2 & -1 & 2 \\ 2 & 0 & 0 & 1 & 1 \end{vmatrix} = \begin{vmatrix} 2 & 0 & 1 & 2 & 0 \\ 0 & -1 & -1 & -1 & 1 \\ 0 & 1 & 2 & 1 & 2 \\ 0 & 0 & 3 & 1 & 2 \\ 0 & 0 & -1 & -1 & 1 \end{vmatrix} \overset{\text{1st col.}}{=} (-1)^{1+1} 2 \cdot \begin{vmatrix} -1 & -1 & -1 & 1 \\ 1 & 2 & 1 & 2 \\ 0 & 3 & 1 & 2 \\ 0 & -1 & -1 & 1 \end{vmatrix} \tag{1.105}$$

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

$$2 \begin{vmatrix} -1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 3 & 1 & 0 \\ 0 & -1 & -1 & 3 \end{vmatrix} \overset{\text{1st row}}{=} -2 \begin{vmatrix} 2 & 1 & 0 \\ 3 & 1 & 0 \\ -1 & -1 & 3 \end{vmatrix} \overset{\text{3rd col.}}{=} (-2) \cdot 3(-1)^{3+3} \cdot \begin{vmatrix} 2 & 1 \\ 3 & 1 \end{vmatrix} = 6 \tag{1.106}$$

In the following set of exercises, you will prove all the basic properties of determinants listed in Remark 26.

Exercise 20
Prove the statements regarding rows in Remark 26(i)-(iii) using induction on n.

Exercise 21
For $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, use Exercise 20 to show Remark 26(v) that adding a multiple of a row to another one does not change $\det(\boldsymbol{A})$.

Exercise 22
Let $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ and let $\boldsymbol{U}_{swap}$, $\boldsymbol{U}_{\times\lambda}$ and $\boldsymbol{U}_{+R}$ be respectively the $n \times n$ matrices obtained by applying the ERO for "swapping two rows", "multiplying a row by $\lambda \neq 0$" and "adding a row to another row" to the identity matrix $\boldsymbol{I}_n$ (cf. Exercise 3).

- *Show that $\det(\boldsymbol{U}_{swap}) = -1$, $\det(\boldsymbol{U}_{\times\lambda}) = \lambda$ and $\det(\boldsymbol{U}_{+R}) = 1$ with $\det(\boldsymbol{U}^{-1}) = 1/\det(\boldsymbol{U})$ for all three types of $\boldsymbol{U}$.*
- *Prove that $\det(\boldsymbol{UA}) = \det(\boldsymbol{U})\det(\boldsymbol{A})$ for $\boldsymbol{U}$ of any of the three types above.*

Exercise 23
Prove that $\det(\boldsymbol{A}) = 0$ iff $\boldsymbol{A}$ is singular by reducing $\boldsymbol{A}$ to REF and using Exercise 22.

Exercise 24
Let $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times n}$.

- *Show that $\det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det \boldsymbol{B} = 0$ if $\det(\boldsymbol{A}) = 0$ using Exercises 12 and 23.*
- *Assume now that $\det(\boldsymbol{A}) \neq 0$. Use Exercise 23 to show that $\boldsymbol{A}^{-1}$ exists. Then, using Exercise 22, show that $\det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det(\boldsymbol{B})$ by representing $\boldsymbol{A}$ as a product of matrices $\boldsymbol{U}$ induced by the ERO's on the identity matrix.*

- *Deduce Remark 26(viii): If $\boldsymbol{A}$ is invertible then* $\det(\boldsymbol{A}^{-1}) = 1/\det(\boldsymbol{A})$ *by using the previous part.*

Exercise 25
Show that $\det(U) = \det(U^\top)$ *for any matrix induced by an elementary row operation as in Exercise 24. Then, use the technique in Exercise 24 to show Remark 26(iv):* $\det(A) = \det(A^\top)$

1.10 Eigenvalues

Until now we have restricted ourselves, for simplicity and convenience, to finite dimensional real vector spaces $\mathbb{R}^n$ where the scalars have been real numbers. It was natural to make this choice since in applications in most areas of computer science (Quantum Computation being a notable exception) it is real vector spaces that are used. However, when we deal with eigenvalues and eigenvectors of matrices it is natural to allow our vector space be $\mathbb{C}^n$, since the eigenvalues and eigenvectors of a *real* square matrix of type $\mathbb{R}^{n\times n}$ can indeed be complex numbers and complex vectors respectively. Thus, although we will still only study matrices of type $\mathbb{R}^{n\times n}$, we recognise that they may have complex eigenvalues and complex eigenvectors.
The good news is that all that we have covered so far (in terms of linear dependence/independence, subspaces, kernel/image spaces and matrix representation etc.) holds for the complex vector space $\mathbb{C}^n$ and complex matrices of type $\mathbb{C}^{m\times n}$: the only difference is that the scalars are in $\mathbb{C}$ and not necessarily in $\mathbb{R}$. For example, $U \subset \mathbb{C}^n$ is a subspace if $U \neq \emptyset$ and is closed under addition of vectors and multiplication of vectors by *complex numbers*.
Interestingly, the set of real vectors in Equations (1.43) provides also a standard basis for $\mathbb{C}^n$, i.e., any vector in $\mathbb{C}^n$ can be obtained as a linear combination of these basis vectors where the scalars are now complex number. In addition the identity matrix I_n is still the identity matrix on $\mathbb{C}^n$.

Definition 21 (Eigenvalue, Eigenvector)
For a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ *, the scalar* $\lambda \in \mathbb{C}$ *is called an* **eigenvalue** *of* $\boldsymbol{A}$ *if there exists* $\boldsymbol{x} \in \mathbb{C}^n \setminus \{\boldsymbol{0}\}$*, with*

$$\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}. \tag{1.107}$$

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

Definition 22 (Eigenspace and Spectrum)

- *The set of all eigenvectors of* $\boldsymbol{A}$ *associated with an eigenvalue* λ *spans a subspace of* $\mathbb{C}^n$*, which is called the* **eigenspace** *of* $\boldsymbol{A}$ *with respect to* λ *and is denoted by* E_λ*.*

- *The set of all eigenvalues of* $\boldsymbol{A}$ *is called the* **spectrum** *of* $\boldsymbol{A}$*.*

Remark 28

- *Note that* $E_\lambda = \ker(\boldsymbol{A} - \lambda I_n)$ *since*

$$\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x} \Leftrightarrow \boldsymbol{A}\boldsymbol{x} - \lambda\boldsymbol{x} = \boldsymbol{0} \Leftrightarrow (\boldsymbol{A} - \lambda I_n)\boldsymbol{x} = \boldsymbol{0} \Leftrightarrow \boldsymbol{x} \in \ker(\boldsymbol{A} - \lambda I_n). \tag{1.108}$$

- *Similar matrices possess the same eigenvalues*
- *Eigenvectors are not unique: If $\boldsymbol{x}$ is an eigenvector of $\boldsymbol{A}$ with eigenvalue λ, then $\alpha\boldsymbol{x}$, $\alpha \in \mathbb{C}\backslash\{0\}$, is an eigenvector with the same eigenvalue. Therefore, there exists an infinite number of eigenvectors for every eigenvalue λ.*

Remark 29
If $\Phi : V \to V$ is a linear map of the vector space V, then the eigenvalues and eigenvectors of any matrix representation of Φ in some basis of V are defined to be the eigenvalues and eigenvectors of Φ. It is straightforward to check that the set of eigenvalues and eigenvectors of Φ do not depend on the matrix representation with respect to any particular basis.

Theorem 7
Consider a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ with pairwise different eigenvalues $\lambda_1, \ldots, \lambda_k \in \mathbb{C}$ and corresponding eigenvectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$. Then the eigenvectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in \mathbb{C}^n$ are linearly independent.

Exercise 26
Prove Theorem 7 by induction on k.

More informally, Theorem 7 states that eigenvectors belonging to different eigenvalues are linearly independent.[20]
The following statements are equivalent:

- $\lambda \in \mathbb{C}$ is eigenvalue of $\boldsymbol{A} \in \mathbb{R}^{n\times n}$
- There exists an $\boldsymbol{x} \in \mathbb{C}^n \setminus \{\boldsymbol{0}\}$ with $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$ or, equivalently, $(\boldsymbol{A} - \lambda\boldsymbol{I}_n)\boldsymbol{x} = \boldsymbol{0}$
- $(\boldsymbol{A} - \lambda\boldsymbol{I}_n)\boldsymbol{x} = \boldsymbol{0}$ can be solved non-trivially, i.e., $\boldsymbol{x} \neq \boldsymbol{0}$.
- $\mathrm{rk}(\boldsymbol{A} - \lambda\boldsymbol{I}_n) < n$
- $\det(\boldsymbol{A} - \lambda\boldsymbol{I}_n) = 0$

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

Remark 30 (Geometric Interpretation)
Geometrically, an eigenvector corresponding to a real, nonzero eigenvalue points in a direction that is stretched, and the eigenvalue is the factor by which it is stretched. If the eigenvalue is negative, the direction is reversed. In particular, the eigenvector does not change its direction under $\boldsymbol{A}$; in other words the subspace generated by the eigenvector is invariant[21] An eigenvector corresponding to a complex, nonzero eigenvalue represents a complex direction (complex 1 dim subspace) which is invariant by $\boldsymbol{A}$. In terms of the real vector space, a complex eigenvector represents a rotation as well as a stretching by the matrix.

[20] Here is a short video that goes through a proof (by contradiction):
`http://tinyurl.com/hhxkslk`

[21] A short video on the geometric interpretation of eigenvalues is available at
`http://tinyurl.com/p6nhkvf`.

1.10.1 Characteristic Polynomial

In the following, we will discuss how to determine the eigenspaces of an endomorphism Φ.[22] For this, we need to introduce the characteristic polynomial first.

Definition 23 (Characteristic Polynomial)
For $\lambda \in \mathbb{R}$ and an endomorphism Φ on $\mathbb{R}^n$ with transformation matrix $\boldsymbol{A}$

$$p(\lambda) = \det(\boldsymbol{A} - \lambda \boldsymbol{I}) = a_0 + a_1\lambda + a_2\lambda^2 + \cdots + a_{n-1}\lambda^{n-1} + (-1)^n\lambda^n, \quad a_0, \ldots, a_{n-1} \in \mathbb{R}, \tag{1.109}$$

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

$$a_0 = \det(\boldsymbol{A}), \tag{1.110}$$

$$a_{n-1} = (-1)^{n-1} tr(\boldsymbol{A}), \tag{1.111}$$

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

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

Remark 31
The fundamental theorem of algebra says that any (non-constant) polynomial of degree n (with real or complex coefficients) has precisely n real or complex roots counting their multiplicity. Therefore, the characteristic polynomial of a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ has precisely n roots (real or complex) counting multiplicity.

Remark 32

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

2. *Similar matrices possess the same characteristic polynomial.*

1.10.2 Example: Eigenspace Computation

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

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

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

2. To compute the eigenspace for the eigenvalue $\lambda = 1$, we need to compute the null space of $\boldsymbol{A} - \boldsymbol{I}$:

$$(\boldsymbol{A} - 1 \cdot \boldsymbol{I})\boldsymbol{x} = \boldsymbol{0} \Leftrightarrow \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \boldsymbol{x} = \boldsymbol{0} \tag{1.112}$$

$$\Rightarrow E_1 = \operatorname{span}\left\{ \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\} \tag{1.113}$$

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

1. Characteristic polynomial: $p(\lambda) = \det(\boldsymbol{A} - \lambda \boldsymbol{I}) = \lambda^2 + 1$. We find $\lambda_1 = i$, $\lambda_2 = -i$.
2. The corresponding eigenspaces (for $\lambda_i \in \mathbb{C}$) are

$$E_i = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ i \end{bmatrix} \right\}, \quad E_{-i} = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ -i \end{bmatrix} \right\}. \tag{1.114}$$

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

1. Characteristic polynomial:

$$p(\lambda) = \begin{vmatrix} -\lambda & -1 & 1 & 1 \\ -1 & 1-\lambda & -2 & 3 \\ 2 & -1 & -\lambda & 0 \\ 1 & -1 & 1 & -\lambda \end{vmatrix} = \begin{vmatrix} -\lambda & -1 & 1 & 1 \\ 0 & -\lambda & -1 & 3-\lambda \\ 0 & 1 & -2-\lambda & 2\lambda \\ 1 & -1 & 1 & -\lambda \end{vmatrix} \tag{1.115}$$

$$= \begin{vmatrix} -\lambda & -1-\lambda & 0 & 1 \\ 0 & -\lambda & -1-\lambda & 3-\lambda \\ 0 & 1 & -1-\lambda & 2\lambda \\ 1 & 0 & 0 & -\lambda \end{vmatrix} \tag{1.116}$$

$$= (-\lambda) \begin{vmatrix} -\lambda & -1-\lambda & 3-\lambda \\ 1 & -1-\lambda & 2\lambda \\ 0 & 0 & -\lambda \end{vmatrix} - \begin{vmatrix} -1-\lambda & 0 & 1 \\ -\lambda & -1-\lambda & 3-\lambda \\ 1 & -1-\lambda & 2\lambda \end{vmatrix} \tag{1.117}$$

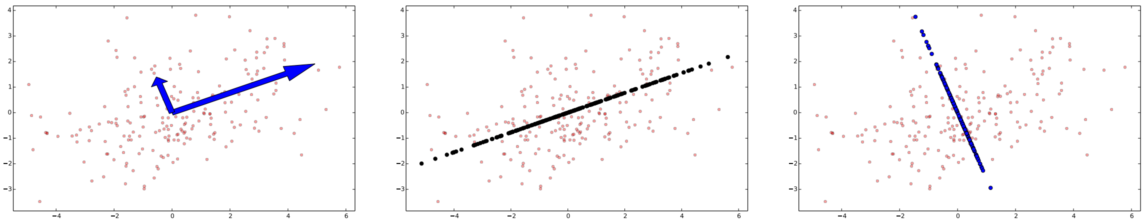
$$= (-\lambda)^2 \begin{vmatrix} -\lambda & -1-\lambda \\ 1 & -1-\lambda \end{vmatrix} - \begin{vmatrix} -1-\lambda & 0 & 1 \\ -\lambda & -1-\lambda & 3-\lambda \\ 1 & -1-\lambda & 2\lambda \end{vmatrix} \tag{1.118}$$

$$= (1+\lambda)^2(\lambda^2 - 3\lambda + 2) = (1+\lambda)^2(1-\lambda)(2-\lambda) \tag{1.119}$$

Therefore, the eigenvalues of $\boldsymbol{A}$ are $\lambda_1 = -1, \lambda_2 = 1, \lambda_3 = 2$.

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

$$E_{-1} = \operatorname{span}\left\{ \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \right\}, \quad E_1 = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\}, \quad E_2 = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} \right\}. \tag{1.120}$$



(a) Data set with eigenvectors (b) Data set with projection onto the first principal component (c) Data set with projection onto the second principal component

Figure 1.7: (a) Two-dimensional data set (red) with corresponding eigenvectors (blue), scaled by the magnitude of the corresponding eigenvectors. The longer the eigenvector the higher the variability (spread) of the data along this axis. (b) For optimal (linear) dimensionality reduction, we would project the data onto the subspace spanned by the eigenvector associated with the largest eigenvalue. (c) Projection of the data set onto the subspace spanned by the eigenvector associated with the smaller eigenvalue leads to a larger projection error.

1.10.3 Eigenvalues in Practical Applications

Eigenvalues and eigenvectors are essential concepts that are widely used.

- Eigenvectors and eigenvalues are fundamental to **principal component analysis** (PCA[23], Hotelling (1933)), which is commonly used for dimensionality reduction in face recognition, data visualization and other machine learning applications. Eigenvalues of a "data matrix" tell us, which dimensions of the high-dimensional data contain a significant signal/information.[24] These dimensions are important to keep, whereas dimensions associated with small eigenvalues can be discarded without much loss. The eigenvectors associated with the largest eigenvalues are called **principal components**. Figure 1.7 illustrates this for two-dimensional data.

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

- Google uses the eigenvector corresponding to the maximal eigenvalue of the Google matrix to determine the rank of a page for search. The idea that the

[23] also known as Kosambi-KarhunenLoève Transform

[24] To be more precise, PCA can be considered a method for (1) performing a basis change from the standard basis toward the eigenbasis in $\mathbb{R}^d$, $d < n$, (2) projecting the data in $\mathbb{R}^n$ onto the subspace spanned by the d eigenvectors corresponding to the d largest eigenvalues (which are called the principal components).

PageRank algorithm[25] brought up was that the importance of any web page can be judged by looking at the pages that link to it. For this, we write down all websites as a huge directed graph that shows which page links to which. PageRank computes the weight (importance) $x_i \geq 0$ of a website a_i by counting the number of pages pointing to a_i. PageRank also take the importance of the website into account that links to a website to a_i. Then, the navigation behavior of a user can be described by a transition matrix $\boldsymbol{A}$ of this graph that tells us with what (click) probability somebody will end up on a different website. The matrix $\boldsymbol{A}$ has the property that for any initial rank/importance vector $\boldsymbol{x}$ of a website the sequence $\boldsymbol{x}, \boldsymbol{A}\boldsymbol{x}, \boldsymbol{A}^2\boldsymbol{x}, \ldots$ converges to a vector $\boldsymbol{x}^*$. This vector is called the **PageRank** and satisfies $\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{x}^*$, i.e., it is an eigenvector (with corresponding eigenvalue 1) of $\boldsymbol{A}$.[26] Liesen and Mehrmann (2015) provide a more detailed description of PageRank with some examples.More details and different perspectives on PageRank can be found at `http://tinyurl.com/83tehpk`.

- Eigenvalues are frequently used to determine **numerical stability**, e.g., when inverting matrices. Since a computer can only represent numbers with a finite precision, we often look at **condition numbers** of matrices, i.e., the ratio $\frac{|\lambda_{\max}|}{|\lambda_{\min}|}$ of the biggest to the smallest eigenvalue. If this ratio exceeds a threshold (e.g., 10^8), matrix inversion may become numerically unstable and lead to inaccurate results.

- Consider a linear time-invariant system that evolves according to $\boldsymbol{x}_{t+1} = \boldsymbol{A}\boldsymbol{x}_t$, where $\boldsymbol{x}$ is called the "state" and $\boldsymbol{A}$ the "transition matrix". We are often interested in analysing the limit-behavior of such a system for $t \to \infty$. This requires us to compute $\lim_{t\to\infty} \boldsymbol{A}^t \boldsymbol{x}_0$ for some initial state $\boldsymbol{x}_0$. To analyse the limit behavior (convergence to a fixed point or a limit cycle or divergence), we perform a basis change into the eigenbasis, such that $\boldsymbol{A} = \boldsymbol{S}^{-1}\boldsymbol{D}\boldsymbol{S}$, where $\boldsymbol{D}$ is a diagonal matrix with the eigenvalues on its diagonal.[27] Then, $\lim_{t\to\infty} \boldsymbol{A}^t \boldsymbol{x}_0 = (\boldsymbol{S}^{-1}\boldsymbol{D}\boldsymbol{S})^t \boldsymbol{x}_0 = \boldsymbol{S}^{-1}\boldsymbol{D}^t\boldsymbol{S}\boldsymbol{x}_0$. Therefore, the eigenvalues of $\boldsymbol{A}$ can tell us something about convergence/divergence (e.g., if there is an eigenvalue $\lambda > 1$ of $\boldsymbol{A}$, the system diverges for $t \to \infty$).

1.11 Diagonalisation

Diagonal matrices are possess a very simple structure and they allow for a very fast computation of determinants and inverses, for instance. In this section, we will have a closer look at how to transform matrices into diagonal form. More specifically, we will look at endomorphisms of finite-dimensional vector spaces, which are similar to a diagonal matrix, i.e., endomorphisms whose transformation matrix attains diagonal structure for a suitable basis. Here, we finally have a practically important

[25] Developed at Stanford University by Larry Page and Sergey Brin in 1996.
[26] When normalising $\boldsymbol{x}^*$, such that $\|\boldsymbol{x}^*\| = 1$ we can interpret the entries as probabilities.
[27] We will discuss this in a later section.

application of the basis change we discussed in Section 1.8.4 and eigenvalues form Section 1.10.

Definition 24 (Diagonal Form)
A matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is **diagonalisable** *if it is similar*[28] *to a diagonal matrix*

$$\begin{bmatrix} c_1 & 0 & \cdots & \cdots & 0 \\ 0 & c_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ 0 & \cdots & 0 & c_{n-1} & 0 \\ 0 & \cdots & \cdots & 0 & c_n \end{bmatrix}. \tag{1.121}$$

A linear map $\Phi : V \to V$ on a vector space V is **diagonalisable** *if it has a matrix representation $\boldsymbol{A}$ in some basis that is diagonalisable.*

Theorem 9
For a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ the following statements are equivalent:

1. *$\boldsymbol{A}$ is diagonalisable.*
2. *There exists a basis of $\mathbb{C}^{n\times n}$ consisting of the eigenvectors of $\boldsymbol{A}$.*[29]
3. *The sum of the dimensions of the eigenspaces of $\boldsymbol{A}$ is n.*[30]

Suppose a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ has characteristic polynomial $p(\lambda)$ with

$$p(\lambda) = (-1)^n(\lambda - c_1)^{r_1} \cdots (\lambda - c_k)^{r_k} \tag{1.122}$$

with $r_i \in \mathbb{N}$ and pairwise different roots $c_i \in \mathbb{R}$. Then, we have:

Theorem 10
$\boldsymbol{A}$ is diagonalisable iff

$$\dim(\mathrm{Im}(\boldsymbol{A} - c_i I_n)) = n - r_i \tag{1.123}$$

$$\overset{\text{Rank-Nullity}}{\Longleftrightarrow} \quad \dim(\ker(\boldsymbol{A} - c_i I_n)) = r_i\,. \tag{1.124}$$

Definition 25
The **algebraic multiplicity** *of the eigenvalue c_i of $\boldsymbol{A}$ is the number r_i for which c_i is repeated as a root of the characteristic polynomial in Equation 1.122. The* **geometric multiplicity** *of c_i is the dimension of the eigenstate E_{c_i} of c_i, i.e., dimension of the kernel/null space of $\boldsymbol{A} - c_i I_n$*

[28]Remember: Two matrices $\boldsymbol{A}, \boldsymbol{D}$ are similar if and only if there exists an invertible matrix $\boldsymbol{S}$, such that $\boldsymbol{D} = \boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}$.
[29]Therefore, we need n eigenvectors.
[30]In particular, $\boldsymbol{A}$ is diagonalisable if it has n different eigenvalues.

The requirement in (1.123) says that for each eigenvalue c_i the geometric multiplicity of c_i, must be equal to its algebraic multiplicity r_i: The dimension of the eigenspace E_{c_i} must be the dimension of the kernel/null space of $\boldsymbol{A} - c_i I_n$.
Thus, $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is diagonalisable iff for each eigenvalue c_i there are r_i linearly independent eigenvectors $\boldsymbol{v}_{ij}$ with $j = 1,\dots,r_i$ for c_i. If E is the standard basis and we take all such eigenvectors for all c_i with $i = 1,\dots,k$ as the basis

$$B = (\boldsymbol{v}_{11},\dots,v_{1r_1},v_{21}\dots,v_{2r_2},\dots,\boldsymbol{v}_{k1},\dots,\boldsymbol{v}_{kr_k}),$$

then we have:

$$\boldsymbol{I}_{BE}\boldsymbol{A}\boldsymbol{I}_{EB} = \boldsymbol{I}_{EB}^{-1}\boldsymbol{A}\boldsymbol{I}_{EB} = \begin{bmatrix} c_1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & & & \vdots \\ \vdots & \ddots & c_1 & \ddots & & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & c_k & \ddots & \vdots \\ \vdots & & & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & c_k \end{bmatrix}, \tag{1.125}$$

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

Remark 33
The dimension of the eigenspace $\mathbb{E}_\lambda$ i.e., the geometric multiplicity of λ is clearly less than or equal to the algebraic multiplicity of λ.

Remark 34
So far, we computed diagonal matrices as $\boldsymbol{D} = \boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}$. However, we can equally write $\boldsymbol{A} = \boldsymbol{S}\boldsymbol{D}\boldsymbol{S}^{-1}$. Here, we can interpret the transformation matrix $\boldsymbol{A}$ as follows: $\boldsymbol{S}^{-1}$ performs a basis change from the standard basis into the eigenbasis. Then, $\boldsymbol{D}$ then scales the vector along the axes of the eigenbasis, and $\boldsymbol{S}$ transforms the scaled vectors back into the standard/canonical coordinates.

Example

- $\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$.
  1. Characteristic polynomial: $p(\lambda) = (1-\lambda)^2$
  2. The dimension of eigenspace is $2 - \text{rk}(\boldsymbol{A} - \boldsymbol{I}) = 1 \neq 2$. Because of Theorem 10 $\boldsymbol{A}$ is not diagonalisable.
- $\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$.

- Characteristic polynomial: $p(\lambda) = 1 + \lambda^2$.
  - For $\lambda \in \mathbb{R}$ there exist no roots of $p(\lambda)$, and the characteristic polynomial does not decompose into linear real factors.
  - If we consider a $\mathbb{C}^2$ however and $\lambda \in \mathbb{C}$, then $p(\lambda) = (i - \lambda)(-i - \lambda)$ and $\boldsymbol{A}$ has two eigenvalues, the characteristic polynomial decomposes into linear factors, and $\boldsymbol{A}$ is diagonalisable (verify the dimension of the eigenspaces).

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

  1. Characteristic polynomial: $p(\lambda) = (1 + \lambda)^2(1 - \lambda)(2 - \lambda)$. The eigenvalues are $c_1 = -1, c_2 = 1, c_3 = 2$ with algebraic multiplicities $r_1 = 2, r_2 = 1, r_3 = 1$, respectively.
  2. Dimension of eigenspaces: $\dim(E_{c_1}) = 1 \neq r_1$.

  Therefore, $\boldsymbol{A}$ cannot be diagonalised.

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

  1. Characteristic polynomial: $p(\lambda) = (2 - \lambda)^2(7 - \lambda)$. Therefore, $c_1 = 2, c_2 = 7, r_1 = 2, r_2 = 1$
  2. Dimension of eigenspaces: $\dim(\boldsymbol{A} - c_1\boldsymbol{I}_3) = 2 = r_1$, $\dim(\boldsymbol{A} - c_2\boldsymbol{I}_3) = 1 = r_2$

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

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

Remark 35
If $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is diagonalisable and $(\boldsymbol{b}_1, \dots, \boldsymbol{b}_n)$ is an ordered basis of eigenvectors of $\boldsymbol{A}$ with $\boldsymbol{A}\boldsymbol{b}_i = c_i \boldsymbol{b}_i$, $i = 1, \dots, n$, then it follows that for the invertible matrix $\boldsymbol{S} = [\boldsymbol{b}_1 | \dots | \boldsymbol{b}_n]$

$$\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S} = \begin{bmatrix} c_1 & 0 & \cdots & \cdots & 0 \\ 0 & c_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \cdots & 0 & c_{n-1} & 0 \\ 0 & \cdots & \cdots & 0 & c_n \end{bmatrix}. \tag{1.126}$$

The diagonal matrix in (1.126) *is the transformation matrix of $\boldsymbol{x} \mapsto \boldsymbol{A}\boldsymbol{x}$ with respect to the eigenbasis $(\boldsymbol{b}_1, \dots, \boldsymbol{b}_n)$.*

Example

Coming back to the above example, where we wanted to determine the diagonal form of

$$A = \begin{bmatrix} 3 & 2 & -1 \\ 2 & 6 & -2 \\ 0 & 0 & 2 \end{bmatrix}. \tag{1.127}$$

From above, we already know that A is diagonalisable. We now determine the eigenbasis of $\mathbb{R}^3$ that allows us to transform A into a similar matrix in diagonal form via $S^{-1}AS$:

1. The eigenspaces are

$$E_2 = \operatorname{span}\left\{ \underbrace{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}}_{=:b_1}, \underbrace{\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}}_{=:b_2} \right\}, \quad E_7 = \operatorname{span}\left\{ \underbrace{\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}}_{=:b_3} \right\} \tag{1.128}$$

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

$$S = (b_1|b_2|b_3) = \begin{bmatrix} 1 & -2 & 1 \\ 0 & 1 & 2 \\ 1 & 0 & 0 \end{bmatrix} \tag{1.129}$$

   such that

$$S^{-1}AS = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 7 \end{bmatrix}. \tag{1.130}$$

1.11.1 Applications

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

$$A^k = (SDS^{-1})^k = SD^kS^{-1}. \tag{1.131}$$

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

A different property of diagonal matrices is that they decouple variables. This is important in probability theory to interpret random variables, e.g., for Gaussian distributions.

With diagonal matrices, it is easier to analyse properties of differential equations, which play an important role in any kind of (linear) dynamical system.

1.11.2 Cayley-Hamilton Theorem

Theorem 11 (Cayley-Hamilton)
Let $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ with characteristic polynomial p. Then,

$$p(\boldsymbol{A}) = 0 \tag{1.132}$$

Remark 36

- *Note that the right hand side of* (1.132) *is the zero matrix.*
- *The importance of the Cayley-Hamilton theorem is not the existence of a (non-trivial) polynomial q, such that $q(\boldsymbol{A}) = 0$, but that the characteristic polynomial has this property.*

Exercise 27
Prove Cayley-Hamilton Theorem for any diagonalisable matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$.

Applications

- Find an expression for $\boldsymbol{A}^{-1}$ in terms of $\boldsymbol{I}, \boldsymbol{A}, \boldsymbol{A}^2, \dots, \boldsymbol{A}^{n-1}$.

Example
$\boldsymbol{A} = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix}$ has the characteristic polynomial $p(\lambda) = \lambda^2 - 2\lambda + 3$. Then Theorem 11 states that $\boldsymbol{A}^2 - 2\boldsymbol{A} + 3\boldsymbol{I} = \boldsymbol{0}$ and, therefore,

$$-\boldsymbol{A}^2 + 2\boldsymbol{A} = 3\boldsymbol{I} \tag{1.133}$$
$$\Leftrightarrow \boldsymbol{A}\tfrac{1}{3}(-\boldsymbol{A} + 2\boldsymbol{I}) = \boldsymbol{I} \tag{1.134}$$
$$\Leftrightarrow \boldsymbol{A}^{-1} = \tfrac{1}{3}(2\boldsymbol{I} - \boldsymbol{A}). \tag{1.135}$$

Generally, for an invertible matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{A}^{-1}$ can be written as an $(n-1)$-th order polynomial expression in $\boldsymbol{A}$.

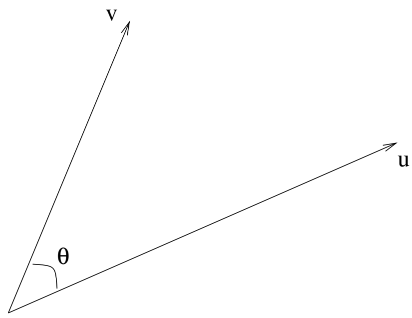
- Find an expression of $\boldsymbol{A}^m$, $m \geq n$, in terms of $\boldsymbol{I}, \boldsymbol{A}, \boldsymbol{A}^2, \dots, \boldsymbol{A}^{n-1}$

1.12 Scalar Product

The **scalar** or **dot** product of two geometric vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^2$ is defined using the cosine formula

$$\boldsymbol{u} \cdot \boldsymbol{v} = \|\boldsymbol{u}\|\|\boldsymbol{v}\| \cos\theta \tag{1.136}$$

where $\|\boldsymbol{u}\|$ is the length of the vector $\boldsymbol{u}$ and θ is the angle between the two vectors. Note that this definition is independent of any basis of $\mathbb{R}^2$ as it only depends on the length of the two vectors and the angle between them.



The two basic cases: (i) If $\theta = 0$ then $\boldsymbol{u} \cdot \boldsymbol{v} = ||\boldsymbol{u}||||\boldsymbol{v}||$, thus in particular $\boldsymbol{u} \cdot \boldsymbol{u} = ||\boldsymbol{u}||^2$ so that $||\boldsymbol{u}|| = \sqrt{\boldsymbol{u} \cdot \boldsymbol{u}}$ and (ii) if $\theta = \pi/2$ then $\boldsymbol{u} \cdot \boldsymbol{v} = 0$.
If $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$, then there exists a unique (two dimensional) plane going through them and one can define the scalar product of them with respect to the angle they make with each other on this plane and Equation 1.136 defines their scalar product.
Clearly the scalar product is commutative $\boldsymbol{a} \cdot \boldsymbol{b} = \boldsymbol{b} \cdot \boldsymbol{a}$. It is also linear in each of its two arguments. In fact, it is a simple exercise in geometry to check that with the geometric definition we have:

$$\boldsymbol{a} \cdot (\boldsymbol{b} + \boldsymbol{c}) = \boldsymbol{a} \cdot \boldsymbol{b} + \boldsymbol{a} \cdot \boldsymbol{c}.$$

If we take any basis consisting of two perpendicular unit vectors (e.g., then standard basis of $\mathbb{R}^2$) then:

$$\boldsymbol{a} \cdot \boldsymbol{b} = a_1 b_1 + a_2 b_2 \tag{1.137}$$

where $\boldsymbol{a} = (a_1, a_2)^\top$ and $\boldsymbol{b} = (b_1, b_2)^\top$ are the components of $\boldsymbol{a}, \boldsymbol{b}$ with respect to this basis.

Exercise 28
Prove Equation 1.137.

In fact, in general, if we have any basis of $\mathbb{R}^n$ consisting of mutually perpendicular basis vectors $\boldsymbol{e}_i$ with $1 \leq i \leq n$, and two vectors $\boldsymbol{a} = \sum_{i=1}^n a_i \boldsymbol{e}_i$ and $\boldsymbol{b} = \sum_{i=1}^n b_i \boldsymbol{e}_i$, then $\boldsymbol{e}_i \cdot \boldsymbol{e}_j = 1$ if $i = j$ and 0 otherwise. Then, we have:

Theorem 12

$$\boldsymbol{a} \cdot \boldsymbol{b} = \sum_{i=1}^n a_i b_i$$

Proof By using the distributivity of the scalar product in the first component we have:

$$\boldsymbol{a} \cdot \boldsymbol{b} = \left(\sum_{i=1}^n a_i \boldsymbol{e}_i \right) \cdot \boldsymbol{b} = \sum_{i=1}^n a_i (\boldsymbol{e}_i \cdot \boldsymbol{b}) = \sum_{i=1}^n a_i \left(\boldsymbol{e}_i \cdot \left(\sum_{j=1}^n b_j \boldsymbol{e}_j \right) \right) = \sum_{i=1}^n a_i b_i$$

since $\boldsymbol{e}_i \cdot \boldsymbol{e}_j = 1$ if $i = j$ and 0 otherwise.

Definition 26
A basis $(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ *of* $\mathbb{R}^n$ *is* **orthonormal** *if* $\boldsymbol{b}_i \cdot \boldsymbol{b}_j = 1$ *if* $i = j$ *and* 0 *otherwise, i.e., each basis vector has length* 1 *and the basis vectors are pairwise perpendicular.*

Corollary 2
If the coordinates of two vectors $\boldsymbol{a}$ *and* $\boldsymbol{b}$ *in an orthonormal basis is given by* $\boldsymbol{a} = [a_1, \ldots a_n]^\top$ *and* $\boldsymbol{b} = [b_1, \ldots b_n]^\top$*, then*

$$\boldsymbol{a} \cdot \boldsymbol{b} = \sum_{i=1}^{n} a_i b_i = \boldsymbol{a}^\top \boldsymbol{b} = \boldsymbol{b}^\top \boldsymbol{a}.$$

Exercise 29
Assuming an orthonormal basis, show that the angle θ *between two vectors* $\boldsymbol{a}$ *and* $\boldsymbol{b}$ *in* $\mathbb{R}^n$ *satisfies:*

$$\cos\theta = \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} a_i^2}\sqrt{\sum_{i=1}^{n} b_i^2}}$$

Remark 37
The **Euclidean length or Euclidean norm** $\|\boldsymbol{x}\|$ *of vectors possesses the following properties:*

1. $\|\boldsymbol{x}\| \geq 0$ *for all* $\boldsymbol{x} \in V$ *and* $\|\boldsymbol{x}\| = 0 \Leftrightarrow \boldsymbol{x} = \boldsymbol{0}$

2. $\|\lambda \boldsymbol{x}\| = |\lambda| \cdot \|\boldsymbol{x}\|$ *for all* $\boldsymbol{x} \in V$ *and* $\lambda \in \mathbb{R}$

3. **Triangular inequality:** $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$ *for all* $\boldsymbol{x}, \boldsymbol{y} \in V$

Definition 27 (Orthogonality)
Vectors $\boldsymbol{x}$ *and* $\boldsymbol{y}$ *are* **orthogonal** *if* $\boldsymbol{x} \cdot \boldsymbol{y} = 0$ *(i.e., if they are perpendicular), and we write* $\boldsymbol{x} \perp \boldsymbol{y}$

Theorem 13
The scalar or dot product $\boldsymbol{a} \cdot \boldsymbol{b}$ *of vectors in* $\mathbb{R}^n$ *has the following properties:*

1. **Cauchy-Schwarz inequality:** $|\boldsymbol{x} \cdot \boldsymbol{y}| \leq \|\boldsymbol{x}\| \|\boldsymbol{y}\|$

2. **Parallelogram law:** $\|\boldsymbol{x} + \boldsymbol{y}\|^2 + \|\boldsymbol{x} - \boldsymbol{y}\|^2 = 2\|\boldsymbol{x}\|^2 + 2\|\boldsymbol{y}\|^2$

3. $4\boldsymbol{x} \cdot \boldsymbol{y} = \|\boldsymbol{x} + \boldsymbol{y}\|^2 - \|\boldsymbol{x} - \boldsymbol{y}\|^2$

4. $\boldsymbol{x} \perp \boldsymbol{y} \Leftrightarrow \|\boldsymbol{x} + \boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2$

Exercise 30
Prove Theorem 13.

1.12.1 Applications of Inner Products

The scalar product allows us to compute angles between vectors as well as distances. A major purpose of the scalar product is to determine whether vectors are orthogonal to each other; in this case $\boldsymbol{x} \cdot \boldsymbol{y} = 0$. This will play an important role when we discuss projections in a later section . The scalar product also allows us to determine specific bases of vector (sub)spaces, where each vector is orthogonal to all others (orthogonal bases) using the Gram-Schmidt method[31]. These bases are important optimisation and numerical algorithms for solving linear equation systems. For instance, Krylov subspace methods[32], such as Conjugate Gradients or GMRES, minimise residual errors that are orthogonal to each other (Stoer and Burlirsch, 2002).
In machine learning, a generalisation of the scalar product (called the **inner product**) is important in the context of kernel methods (Schölkopf and Smola, 2002). Kernel methods exploit the fact that many linear algorithms can be expressed purely by inner product computations.[33] Then, the "kernel trick" allows us to compute these inner products implicitly in a (potentially infinite-dimensional) feature space, without even knowing this feature space explicitly. This allowed the "non-linearisation" of many algorithms used in machine learning, such as kernel-PCA (Schölkopf et al., 1998) for dimensionality reduction. Gaussian processes (Rasmussen and Williams, 2006) also fall into the category of kernel methods and are the current state-of-the-art in probabilistic regression (fitting curves to data points).

1.13 Orthogonal Projections

Projections are an important class of linear transformations (besides rotations and reflections). Projections play an important role in graphics (see e.g., Figure 1.8), coding theory, statistics and machine learning. We often deal with data that is very high-dimensional. However, often only a few dimensions are important. In this case, we can project the original very high-dimensional data into a lower-dimensional feature space[34] and work in this lower-dimensional space to learn more about the data set and extract patterns. For example, machine learning tools such as Principal Component Analysis (PCA) by Hotelling (1933) and Deep Neural Networks (e.g., deep auto-encoders, first applied by Deng et al. (2010)) heavily exploit this idea.
In the following, we will focus on linear orthogonal projections.

Definition 28 (Projection)
Let $W \subset \mathbb{R}^n$ a subspace. A linear onto mapping $\pi : \mathbb{R}^n \to W$ is called a **projection** *onto W if $\pi^2 = \pi \circ \pi = \pi$.*

As always, we can study projection maps by studying their matrix representation.

[31]This is not discussed in this course

[32]The basis for the Krylov subspace is derived from the Cayley-Hamilton theorem, which allows us to compute the inverse of a matrix in terms of a linear combination of its powers.

[33]Matrix-vector multiplication $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ falls into this category since b_i is dot product of the ith row of $\boldsymbol{A}$ with $\boldsymbol{x}$.

[34]"Feature" is just a commonly used word for "data representation".

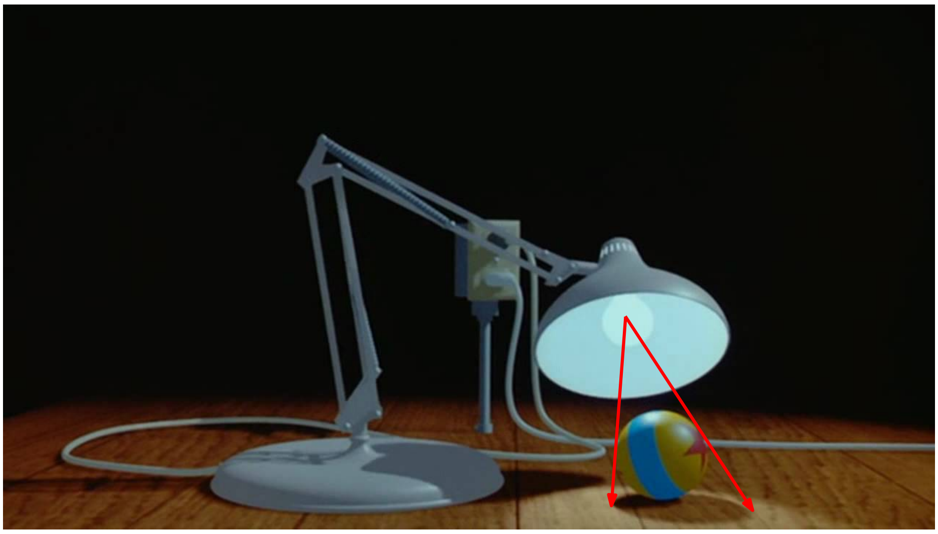


Figure 1.8: The shade is the projection of the ball onto a plane (table) with the center being the light source. Adapted from `http://tinyurl.com/ka4t28t`.

Remark 38 (Projection matrix)
Since linear mappings can be expressed by transformation matrices, the definition above applies equally to a special kind of transformation matrices, the **projection matrices** $\boldsymbol{P}_\pi$*, which exhibit the property that* $\boldsymbol{P}_\pi^2 = \boldsymbol{P}_\pi$*,i.e.,* $\boldsymbol{P}_\pi(\boldsymbol{P}_\pi - \boldsymbol{I}_n) = \boldsymbol{0}$*. A projection matrix is an* **orthogonal** *projection if* $\boldsymbol{P}_\pi = \boldsymbol{P}_\pi^\top$*, i.e,* $(\boldsymbol{P}_\pi \boldsymbol{x}) \cdot \boldsymbol{y} = \boldsymbol{x} \cdot (\boldsymbol{P}_\pi \boldsymbol{y})$*. We have* $\mathrm{Im}\boldsymbol{P}_\pi = W$*. Since* $\boldsymbol{P}_\pi^2 = \boldsymbol{P}_\pi$*, it follows that* $\boldsymbol{P}_\pi(\boldsymbol{v}) = \boldsymbol{v}$ *for any* $\boldsymbol{v} \in W$*. This means that* $\boldsymbol{P}_\pi$ *has eigenvalue* 1 *with algebraic multiplicity* $\dim W$*. By rank-nullity theorem* $n - \dim W$ *is the nullity of* $\boldsymbol{P}_\pi$ *and* $\boldsymbol{P}_\pi$ *has also eigenvalue* 0 *with algebraic multiplicity* $n - \dim W$*. Thus, all eigenvalues of* $\boldsymbol{P}_\pi$ *are either 1 or 0. The corresponding eigenspaces are the image space or the null space of the projection, respectively.*[35]

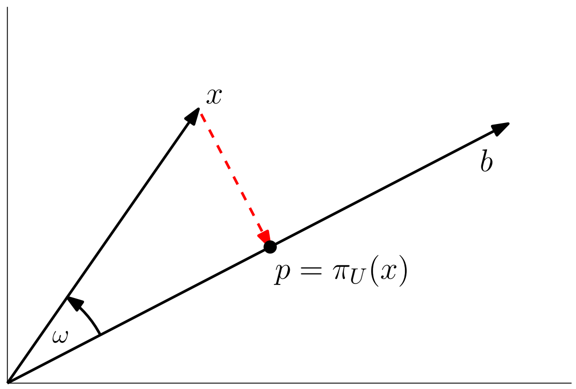
In the following, we will derive projections by using the scalar product. We will start with one-dimensional subspaces, the lines.

1.13.1 Projection onto 1-Dimensional Subspaces (Lines)

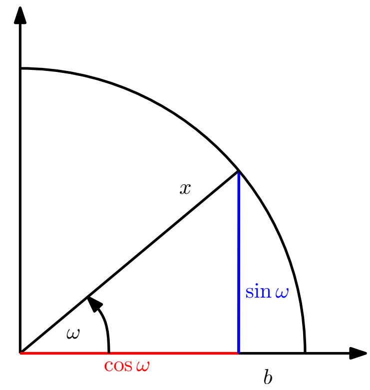
Assuming we are given a line (1-dimensional subspace) through the origin with basis vector $\boldsymbol{b} \in \mathbb{R}^m$.[36] This is a one-dimensional subspace $U \subset \mathbb{R}^m$ spanned by $\boldsymbol{b}$. When we project $\boldsymbol{x} \in \mathbb{R}^m$ onto U, we want to find the point $\pi_U(\boldsymbol{x}) = \boldsymbol{p} \in U$ that is closest to $\boldsymbol{x}$. Using geometric arguments, let us characterize some properties of the projection $\boldsymbol{p} = \pi_U(\boldsymbol{x})$ (Fig. 1.9(a) serves as an illustration):

- The projection point $\boldsymbol{p}$ is closest to $\boldsymbol{x}$, where "closest" implies that the distance $\|\boldsymbol{x}-\boldsymbol{p}\|$ is minimal. This means, the segment $\boldsymbol{p}-\boldsymbol{x}$ from $\boldsymbol{p}$ to $\boldsymbol{x}$ is orthogonal to U and, therefore, the basis $\boldsymbol{b}$ of U. The orthogonality condition yields $(\boldsymbol{p}-\boldsymbol{x}) \cdot \boldsymbol{b} = (\boldsymbol{p}-\boldsymbol{x})^\top \boldsymbol{b} = 0$.

[35] A good illustration is given here: `http://tinyurl.com/p5jn5ws`.
[36] In affine geometry, the space spanned by $\boldsymbol{b}$ is also called a **direction (space)**.



(a) General case



(b) Special case: Unit circle with $\|\boldsymbol{x}\| = 1$

Figure 1.9: Projection $\pi_U(x)$ of x onto a subspace U with basis b.

- The projection $\boldsymbol{p} = \pi_U(\boldsymbol{x})$ of $\boldsymbol{x}$ onto U must be a multiple of the basis/direction vector $\boldsymbol{b}$ that spans U. Thus, $\boldsymbol{p} = \lambda\boldsymbol{b}$, for some $\lambda \in \mathbb{R}$.

In the following three steps, we determine λ, the projection point $\pi_U(\boldsymbol{x}) = \boldsymbol{p} \in U$ and the projection matrix $\boldsymbol{P}_\pi$ that maps arbitrary $\boldsymbol{x} \in \mathbb{R}^m$ onto U.

1. Finding λ. We know that $\boldsymbol{p} = \lambda\boldsymbol{b}$. Therefore,

$$\boldsymbol{x} - \boldsymbol{p} = \boldsymbol{x} - \lambda\boldsymbol{b} \perp \boldsymbol{b} \tag{1.138}$$

$$\Leftrightarrow (\boldsymbol{x} - \lambda\boldsymbol{b}) \cdot \boldsymbol{b} = 0 \tag{1.139}$$

$$\Leftrightarrow \boldsymbol{x} \cdot \boldsymbol{b} - \lambda\boldsymbol{b} \cdot \boldsymbol{b} = 0 \tag{1.140}$$

$$\Leftrightarrow \lambda = \frac{\boldsymbol{x} \cdot \boldsymbol{b}}{\boldsymbol{b} \cdot \boldsymbol{b}} = \frac{\boldsymbol{x} \cdot \boldsymbol{b}}{\|\boldsymbol{b}\|^2} \tag{1.141}$$

Thus, we get

$$\lambda = \frac{\boldsymbol{b}^\top \boldsymbol{x}}{\boldsymbol{b}^\top \boldsymbol{b}} = \frac{\boldsymbol{b}^\top \boldsymbol{x}}{\|\boldsymbol{b}\|^2} \tag{1.142}$$

2. Finding the projection point $\boldsymbol{p} = \pi_U(\boldsymbol{x}) \in U$. Since $\boldsymbol{p} = \lambda\boldsymbol{b}$ we immediately obtain with (1.142) that

$$\boldsymbol{p} = \lambda\boldsymbol{b} = \frac{\boldsymbol{b}^\top \boldsymbol{x}}{\|\boldsymbol{b}\|^2}\boldsymbol{b}. \tag{1.143}$$

We can also compute the length of $\boldsymbol{p}$ (i.e., the distance from $\boldsymbol{0}$):

$$\|\boldsymbol{p}\| = \|\lambda\boldsymbol{b}\| = |\lambda|\,\|\boldsymbol{b}\| = \frac{|\boldsymbol{b}^\top \boldsymbol{x}|}{\|\boldsymbol{b}\|^2}\|\boldsymbol{b}\| = |\cos\omega|\,\|\boldsymbol{x}\|\,\|\boldsymbol{b}\|\frac{\|\boldsymbol{b}\|}{\|\boldsymbol{b}\|^2} = |\cos\omega|\,\|\boldsymbol{x}\|. \tag{1.144}$$

Here, ω is the angle between $\boldsymbol{x}$ and $\boldsymbol{b}$. This equation should be familiar from trigonometry: If $\|\boldsymbol{x}\| = 1$ it lies on the unit circle. Then the projection onto the horizontal axis $\boldsymbol{b}$ is exactly $|\cos\omega|$ in magnitude. An illustration is given in Figure 1.9(b)

3. Finding the projection matrix $\boldsymbol{P}_\pi$. We know that a projection is a linear mapping (see Definition 28). Therefore, there exists a projection matrix $\boldsymbol{P}_\pi$, such that $\boldsymbol{\pi}_U(\boldsymbol{x}) = \boldsymbol{p} = \boldsymbol{P}_\pi \boldsymbol{x}$. With

$$\boldsymbol{p} = \lambda \boldsymbol{b} = \boldsymbol{b}\frac{\boldsymbol{b}^\top \boldsymbol{x}}{\|\boldsymbol{b}\|^2} = \frac{\boldsymbol{b}\boldsymbol{b}^\top}{\|\boldsymbol{b}\|^2}\boldsymbol{x} \tag{1.145}$$

we immediately see that

$$\boldsymbol{P}_\pi = \frac{\boldsymbol{b}\boldsymbol{b}^\top}{\|\boldsymbol{b}\|^2}. \tag{1.146}$$

Note that $\boldsymbol{b}\boldsymbol{b}^\top$ is a matrix (with rank 1) and $\|\boldsymbol{b}\|^2$ is of course a scalar.

The projection matrix $\boldsymbol{P}_\pi$ projects any vector $\boldsymbol{x} \in \mathbb{R}^m$ onto the line through the origin with direction $\boldsymbol{b}$ (equivalently, the subspace U spanned by $\boldsymbol{b}$).

Example (Projection onto a Line)
Find the projection matrix $\boldsymbol{P}_\pi$ onto the line through the origin spanned by $\boldsymbol{b} = \begin{bmatrix}1 & 2 & 2\end{bmatrix}^\top$. $\boldsymbol{b}$ is a direction and a basis of the one-dimensional subspace (line through origin).
With (1.146), we obtain

$$\boldsymbol{P}_\pi = \frac{\boldsymbol{b}\boldsymbol{b}^\top}{\boldsymbol{b}^\top \boldsymbol{b}} = \frac{1}{9}\begin{bmatrix}1\\2\\2\end{bmatrix}\begin{bmatrix}1 & 2 & 2\end{bmatrix} = \frac{1}{9}\begin{bmatrix}1 & 2 & 2\\2 & 4 & 4\\2 & 4 & 4\end{bmatrix}. \tag{1.147}$$

Let us now choose a particular $\boldsymbol{x}$ and see whether it lies in the subspace spanned by $\boldsymbol{b}$. For $\boldsymbol{x} = \begin{bmatrix}1 & 1 & 1\end{bmatrix}^\top$, the projected point is

$$\boldsymbol{p} = \boldsymbol{P}_\pi \boldsymbol{x} = \frac{1}{9}\begin{bmatrix}1 & 2 & 2\\2 & 4 & 4\\2 & 4 & 4\end{bmatrix}\begin{bmatrix}1\\1\\1\end{bmatrix} = \frac{1}{9}\begin{bmatrix}5\\10\\10\end{bmatrix} \in \operatorname{span}\left\{\begin{bmatrix}1\\2\\2\end{bmatrix}\right\}. \tag{1.148}$$

Note that the application of $\boldsymbol{P}_\pi$ to $\boldsymbol{p}$ does not change anything, i.e., $\boldsymbol{P}_\pi \boldsymbol{p} = \boldsymbol{p}$.[37] This is expected because according to Definition 28 we know that a projection matrix $\boldsymbol{P}_\pi$ satisfies $\boldsymbol{P}_\pi^2 \boldsymbol{x} = \boldsymbol{P}_\pi \boldsymbol{x}$.

1.13.2 Projection onto General Subspaces

In the following, we generalise projections to the case where we project vectors $\boldsymbol{x} \in \mathbb{R}^n$ onto general subspaces $U \subset \mathbb{R}^n$. Assume that $(\boldsymbol{b}_1, \dots, \boldsymbol{b}_m)$ is an ordered basis of U.[38]

[37] $\boldsymbol{p}$ is therefore an eigenvector of $\boldsymbol{P}_\pi$, and the corresponding eigenvalue is 1.

[38] If U is given by a set of spanning vectors, make sure you determine a basis $\boldsymbol{b}_1, \dots, \boldsymbol{b}_m$ before proceeding.

Projections $\pi_U(\boldsymbol{x})$ onto U are elements of U. Therefore, they exhibit the property that they can be represented as linear combinations of the basis vectors of U, i.e., a projected point is mapped onto the subspace spanned by the columns of $\boldsymbol{B} \in \mathbb{R}^{n\times m}$, where $\boldsymbol{B} = [\boldsymbol{b}_1, ..., \boldsymbol{b}_m]$. Projections $\pi_U(\boldsymbol{x}) \in U$ of $\boldsymbol{x}$ onto the subspace U spanned by $(\boldsymbol{b}_1, \dots, \boldsymbol{b}_m)$ can be expressed as linear combinations of the basis vectors of U, such that $\boldsymbol{p} = \sum_{i=1}^m \lambda_i \boldsymbol{b}_i$.
As before, we follow a three-step procedure to find $\boldsymbol{p}$ and the projection matrix $\boldsymbol{P}_\pi$:

1. Find $\lambda_1, \dots, \lambda_m$, such that the linear combination $\boldsymbol{p} = \sum_{i=1}^m \lambda_i \boldsymbol{b}_i = \boldsymbol{B}\boldsymbol{\lambda}$ is closest to $\boldsymbol{x} \in \mathbb{R}^n$. As in the 1D case, “closest” means “minimum distance”, which implies that the line connecting $\boldsymbol{p} \in U$ and $\boldsymbol{x} \in \mathbb{R}^n$ must be orthogonal to all basis vectors of U. Therefore, we obtain the conditions:

$$\boldsymbol{b}_1 \cdot (\boldsymbol{x} - \boldsymbol{p}) = \boldsymbol{b}_1^\top (\boldsymbol{x} - \boldsymbol{p}) = 0 \tag{1.149}$$

$$\vdots \tag{1.150}$$

$$\boldsymbol{b}_m \cdot (\boldsymbol{x} - \boldsymbol{p}) = \boldsymbol{b}_m^\top (\boldsymbol{x} - \boldsymbol{p}) = 0 \tag{1.151}$$

which, with $\boldsymbol{p} = \boldsymbol{B}\boldsymbol{\lambda}$, can be written as

$$\boldsymbol{b}_1^\top (\boldsymbol{x} - \boldsymbol{B}\boldsymbol{\lambda}) = 0 \tag{1.152}$$

$$\vdots \tag{1.153}$$

$$\boldsymbol{b}_m^\top (\boldsymbol{x} - \boldsymbol{B}\boldsymbol{\lambda}) = 0 \tag{1.154}$$

such that we obtain a homogeneous linear equation system

$$\begin{bmatrix} \boldsymbol{b}_1^\top \\ \vdots \\ \boldsymbol{b}_m^\top \end{bmatrix} (\boldsymbol{x} - \boldsymbol{B}\boldsymbol{\lambda}) = \boldsymbol{0} \quad \Leftrightarrow \quad \boldsymbol{B}^\top (\boldsymbol{x} - \boldsymbol{B}\boldsymbol{\lambda}) = \boldsymbol{0}. \tag{1.155}$$

Factorising yields

$$\boldsymbol{B}^\top (\boldsymbol{x} - \boldsymbol{B}\boldsymbol{\lambda}) = \boldsymbol{0} \quad \Leftrightarrow \quad \boldsymbol{B}^\top \boldsymbol{B}\boldsymbol{\lambda} = \boldsymbol{B}^\top \boldsymbol{x}, \tag{1.156}$$

and the expression on the right-hand side is called **normal equation**.[39] Since the vectors $\boldsymbol{b}_1, \dots, \boldsymbol{b}_m$ are a basis and, therefore, linearly independent, $\boldsymbol{B}^\top \boldsymbol{B}$ is invertible and can be inverted. This allows us to solve for the optimal coefficients

$$\boldsymbol{\lambda} = (\boldsymbol{B}^\top \boldsymbol{B})^{-1} \boldsymbol{B}^\top \boldsymbol{x}. \tag{1.157}$$

The matrix $(\boldsymbol{B}^\top \boldsymbol{B})^{-1} \boldsymbol{B}^\top$ is often called the **pseudo-inverse** of $\boldsymbol{B}$, which can be computed for non-square matrices $\boldsymbol{B}$. It only requires that $\boldsymbol{B}^\top \boldsymbol{B}$ is positive definite, i.e., all its eigenvalues are strictly positive. (Since $\boldsymbol{B}^\top \boldsymbol{B}$ is a symmetric real matrix all its eigenvalues will be real, as you will see in Computational

[39] You may see the normal equation again when you take courses on machine learning, state estimation or robotics.

Techniques course next year.) In practical applications (e.g., linear regression), we often add a "jitter term" $\epsilon \boldsymbol{I}$ to $\boldsymbol{B}^\top \boldsymbol{B}$ to guarantee positive definiteness or increase numerical stability. This "ridge" can be rigorously derived using Bayesian inference.

2. Find the projection of $\pi_U(\boldsymbol{x}) = \boldsymbol{p} \in U$. We already established that $\boldsymbol{p} = \boldsymbol{B}\boldsymbol{\lambda}$. Therefore, with (1.157)

$$\boldsymbol{p} = \boldsymbol{B}(\boldsymbol{B}^\top \boldsymbol{B})^{-1}\boldsymbol{B}^\top \boldsymbol{x}. \tag{1.158}$$

3. Find the projection matrix $\boldsymbol{P}_\pi$. From (1.158) we can immediately see that the projection matrix that solves $\boldsymbol{P}_\pi \boldsymbol{x} = \boldsymbol{p}$ must be

$$\boldsymbol{P}_\pi = \boldsymbol{B}(\boldsymbol{B}^\top \boldsymbol{B})^{-1}\boldsymbol{B}^\top. \tag{1.159}$$

Remark 39
Comparing the solutions for projecting onto a one-dimensional subspace and the general case, we see that the general case includes the 1D case as a special case: If $\dim(U) = 1$ *then* $\boldsymbol{B}^\top \boldsymbol{B}$ *is just a scalar and we can rewrite the projection matrix in* (1.159) $\boldsymbol{P}_\pi = \boldsymbol{B}(\boldsymbol{B}^\top \boldsymbol{B})^{-1}\boldsymbol{B}^\top$ *as* $\boldsymbol{P}_\pi = \frac{\boldsymbol{B}\boldsymbol{B}^\top}{\boldsymbol{B}^\top \boldsymbol{B}}$*, which is exactly the projection matrix in (1.146).*

Example

For a subspace $U = \operatorname{span}\left\{\begin{bmatrix}1\\1\\1\end{bmatrix}, \begin{bmatrix}0\\1\\2\end{bmatrix}\right\} \subset \mathbb{R}^3$ and $\boldsymbol{x} = \begin{bmatrix}6\\0\\0\end{bmatrix} \in \mathbb{R}^3$ find $\boldsymbol{\lambda}$, the projection point $\boldsymbol{p}$ and the projection matrix $\boldsymbol{P}_\pi$.
First, we see that the generating set of U is a basis (linear independence) and write the basis vectors of U into a matrix $\boldsymbol{B} = \begin{bmatrix}1 & 0\\1 & 1\\1 & 2\end{bmatrix}$.
Second, we compute the matrix $\boldsymbol{B}^\top \boldsymbol{B}$ and the vector $\boldsymbol{B}^\top \boldsymbol{x}$ as

$$\boldsymbol{B}^\top \boldsymbol{B} = \begin{bmatrix}1 & 1 & 1\\0 & 1 & 2\end{bmatrix}\begin{bmatrix}1 & 0\\1 & 1\\1 & 2\end{bmatrix} = \begin{bmatrix}3 & 3\\3 & 5\end{bmatrix} \quad \boldsymbol{B}^\top \boldsymbol{x} = \begin{bmatrix}1 & 1 & 1\\0 & 1 & 2\end{bmatrix}\begin{bmatrix}6\\0\\0\end{bmatrix} = \begin{bmatrix}6\\0\end{bmatrix} \tag{1.160}$$

Third, we solve the normal equation $\boldsymbol{B}^\top \boldsymbol{B}\boldsymbol{\lambda} = \boldsymbol{B}^\top \boldsymbol{x}$ to find $\boldsymbol{\lambda}$:

$$\begin{bmatrix}3 & 3\\3 & 5\end{bmatrix}\begin{bmatrix}\lambda_1\\\lambda_2\end{bmatrix} = \begin{bmatrix}6\\0\end{bmatrix} \quad \Rightarrow \boldsymbol{\lambda} = \begin{bmatrix}5\\-3\end{bmatrix} \tag{1.161}$$

Fourth, the projection $\pi_U(\boldsymbol{x}) = \boldsymbol{p}$ of $\boldsymbol{x}$ onto U, i.e., into the column space of $\boldsymbol{B}$, can be directly computed via

$$\boldsymbol{p} = \boldsymbol{B}\boldsymbol{\lambda} = \begin{bmatrix}5\\2\\-1\end{bmatrix}. \tag{1.162}$$

The corresponding **projection/reconstruction error** is $\|x-p\| = \left\|\begin{bmatrix}1 & -2 & 1\end{bmatrix}^\top\right\| = \sqrt{6}$
Fifth, the general projection matrix (for and $x \in \mathbb{R}^3$) is given by

$$P_\pi = B(B^\top B)^{-1}B^\top = \frac{1}{6}\begin{bmatrix} 5 & 2 & 1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix} \tag{1.163}$$

To verify the results, we can (a) check whether the error vector $p - x$ is orthogonal to all basis vectors of U, (b) verify that $P_\pi = P_\pi^2$ (see Definition 28).

1.13.3 Applications of Projections

Projections allow us to answer the question of what to do when we have a linear system $Ax = b$ without a solution. Recall that this means that b does not lie in the span of A, i.e., the vector b does not lie in the subspace spanned by the columns of A. Given that the linear equation cannot be solved exactly, we will find an **approximate solution**. The idea is to find the vector in the subspace spanned by the columns of A that is closest to b, where "closest" is defined as the vector in the span of the column space of A with the greatest inner product with b. This problem arises often in practice, and the solution is called the **least squares solution** of an overdetermined system. This turns out to be equivalent to finding the orthogonal projection of b onto the subspace defined by the columns of A.
Projections are often used in computer graphics, e.g., to generate shadows, see Figure 1.8. In optimisation, orthogonal projections are often used to (iteratively) minimise residual errors. This also has applications in machine learning, e.g., in linear regression where we want to find a (linear) function that minimises the residual errors, i.e., the lengths of the orthogonal projections of the data onto the line (Bishop, 2006). PCA (Hotelling, 1933) also uses projections to reduce the dimensionality of high-dimensional data: First, PCA determines an orthogonal basis of the data space. It turns out that this basis is the eigenbasis of the data matrix. The importance of each individual dimension of the data is proportional to the corresponding eigenvalue. Finally, we can select the eigenvectors corresponding to the largest eigenvalues to reduce the dimensionality of the data, and this selection results in the minimal residual error of the data points projected onto the subspace spanned by these principal components. Figure 1.7 illustrates the projection onto the first principal component for a two-dimensional data set.

1.14 Affine Subspaces

In the following, we will have a closer look at geometric properties of vector spaces. For this purpose, we define an affine space.

Definition 29 (Affine Subspace)
Let V be a vector space, $x_0 \in V$ and $U \subset V$ a subspace. Then the subset

$$L = x_0 + U := \{x_0 + u : u \in U\} = \{v \in V | \exists u \in U : v = x_0 + u\} \subset V \tag{1.164}$$

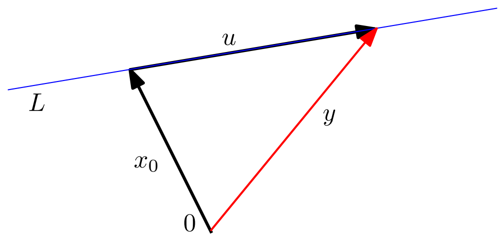


Figure 1.10: Points $\boldsymbol{y}$ on a line lie in an affine subspace L with support point $\boldsymbol{x}_0$ and direction $\boldsymbol{u}$.

is called **affine subspace** *of* V.[40] *The subspace* U *is called* **direction** *or* **direction space**, *and* $\boldsymbol{x}_0$ *is called* **support point**. *The* **dimension** *of* L *is defined to be the dimension of* U, *i.e.,* $\dim L := \dim U$.

Note that an affine subspace may not contain $\boldsymbol{0}$ if the support point $\boldsymbol{x}_0 \notin U$. Therefore, an affine subspace is not necessarily a subspace of V for $\boldsymbol{x}_0 \notin U$.
Examples of affine subspaces are points, lines and planes in $\mathbb{R}^2$, which do not necessarily go through the origin.

Remark 40

(i) *Consider two affine subspaces* $L = \boldsymbol{x}_0 + U$ *and* $\tilde{L} = \tilde{\boldsymbol{x}}_0 + \tilde{U}$ *of a vector space* V. $L \subset \tilde{L}$ *if and only if* $U \subset \tilde{U}$ *and* $\boldsymbol{x}_0 - \tilde{\boldsymbol{x}}_0 \in \tilde{U}$.

(ii) *Affine subspaces are often described by* **parameters**: *Consider a* k*-dimensional affine space* $L = \boldsymbol{x}_0 + U$ *of* V. *If* $(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k)$ *is an (ordered) basis of* U, *then every element* $\boldsymbol{x} \in L$ *can be (uniquely) described as*

$$\boldsymbol{x} = \boldsymbol{x}_0 + \lambda_1 \boldsymbol{b}_1 + \ldots + \lambda_k \boldsymbol{b}_k , \tag{1.165}$$

where $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$.

This representation is called **parametric equation** *of* L *with directional vectors* $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k$ *and* **parameters** $\lambda_1, \ldots, \lambda_k$.

Exercise 31
Prove Remark 40(i)

Example

- One-dimensional affine subspaces are called **lines** and can be written as $\boldsymbol{y} = \boldsymbol{x}_0 + \lambda \boldsymbol{x}_1$, where $\lambda \in \mathbb{R}$, where $U = \mathrm{span}\{\boldsymbol{x}_1\} \subset \mathbb{R}^n$ is a one-dimensional subspace of $\mathbb{R}^n$.
- Two-dimensional affine subspaces of $\mathbb{R}^n$ are called **planes**. The parametric equation for planes is $\boldsymbol{y} = \boldsymbol{x}_0 + \lambda_1 \boldsymbol{x}_1 + \lambda_2 \boldsymbol{x}_2$, where $\lambda_1, \lambda_2 \in \mathbb{R}$ and $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is a basis of $U \subset \mathbb{R}^n$.

[40] L is also called **linear manifold**.

- In an n-dimensional vector space V, the $(n-1)$-dimensional affine subspaces are called **hyperplanes**.
- For $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$ the solution of the linear equation system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ is either the empty set or an affine subspace of $\mathbb{R}^n$ of dimension $n - \text{rk}(\boldsymbol{A})$. In particular, the solution of the linear equation $\lambda_1\boldsymbol{x}_1 + \ldots + \lambda_n\boldsymbol{x}_n = \boldsymbol{b}$, where $(\lambda_1,\ldots,\lambda_n) \neq (0,\ldots,0)$, is a hyperplane in $\mathbb{R}^n$.

- In $\mathbb{R}^n$ every k-dimensional affine subspace is the solution of a linear inhomogeneous equation system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, where $\boldsymbol{A} \in \mathbb{R}^{m\times n}, \boldsymbol{b} \in \mathbb{R}^m$ and $\text{rk}(\boldsymbol{A}) = n - k$.[41]

1.14.1 Intersection of Affine Subspaces

In the following, we consider the problem of determining the intersection $L_1 \cap L_2$ of two affine subspaces $L_1 = \boldsymbol{x}_1 + U_1$ and $L_2 = \boldsymbol{x}_2 + U_2$ where $U_1, U_2 \subset V$ are subspaces with ordered bases $(\boldsymbol{b}_1,\ldots,\boldsymbol{b}_k)$ and $(c_1,\ldots,\boldsymbol{c}_m)$, respectively.
We follow the same approach that we used in the discussion of the intersection of vector subspace (Section 1.7):

$$\boldsymbol{x} \in L_1 \Leftrightarrow \exists \lambda_1,\ldots,\lambda_k \in \mathbb{R} : \boldsymbol{x} = \boldsymbol{x}_1 + \sum_{i=1}^{k} \lambda_i \boldsymbol{b}_i \tag{1.166}$$

$$\boldsymbol{x} \in L_2 \Leftrightarrow \exists \psi_1,\ldots,\psi_m \in \mathbb{R} : \boldsymbol{x} = \boldsymbol{x}_2 + \sum_{j=1}^{k} \psi_j \boldsymbol{c}_j \tag{1.167}$$

We exploited that since $\boldsymbol{b}_1,\ldots,\boldsymbol{b}_k$ are a basis of U_1, every $\boldsymbol{x} \in L_1$ can (uniquely) be represented as a linear combination of the basis vectors plus the support point $\boldsymbol{x}_1$. The same reasoning applies to $\boldsymbol{x} \in L_2$ with the basis $\boldsymbol{c}_1,\ldots,\boldsymbol{c}_m$ of U_2.

$$\forall \boldsymbol{x} \in L_1 \cap L_2 : \boldsymbol{x} \in L_1 \wedge \boldsymbol{x} \in L_2 \tag{1.168}$$

$$\Leftrightarrow \exists \lambda_1,\ldots,\lambda_k,\psi_1,\ldots,\psi_m \in \mathbb{R} : \boldsymbol{x}_1 + \sum_{i=1}^{k} \lambda_i \boldsymbol{b}_i = \boldsymbol{x} = \boldsymbol{x}_2 + \sum_{j=1}^{m} \psi_j \boldsymbol{c}_j \tag{1.169}$$

$$\Leftrightarrow \exists \lambda_1,\ldots,\lambda_k,\psi_1,\ldots,\psi_m \in \mathbb{R} : \sum_{i=1}^{k} \lambda_i \boldsymbol{b}_i - \sum_{j=1}^{m} \psi_j \boldsymbol{c}_j = \boldsymbol{x}_2 - \boldsymbol{x}_1 \tag{1.170}$$

[41] Recall that for homogeneous equation systems $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ the solution was a vector subspace (not affine).

$$\Leftrightarrow \exists \lambda_1, \dots, \lambda_k, \psi_1, \dots, \psi_m \in \mathbb{R} : \underbrace{\begin{bmatrix} \boldsymbol{b}_1 & \cdots & \boldsymbol{b}_k & -\boldsymbol{c}_1 & \cdots & -\boldsymbol{c}_m \end{bmatrix}}_{=:\boldsymbol{A}} \underbrace{\begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_k \\ \psi_1 \\ \vdots \\ \psi_m \end{bmatrix}}_{=:\boldsymbol{y}} = \underbrace{\boldsymbol{x}_2 - \boldsymbol{x}_1}_{=:\boldsymbol{z}} \quad (1.171)$$

Because of this, finding the intersection of two affine spaces reduces to solving an inhomogeneous linear equation system $\boldsymbol{A}\boldsymbol{y} = \boldsymbol{z}$ to determine either $\lambda_1, \dots, \lambda_k$ or $\psi_1, \dots, \psi_m$. When we have these coefficients, $L_1 \cap L_2$ can be described using (1.166) or (1.167).

Example

Consider two linear spaces $L_1, L_2 \subset \mathbb{R}^4$:

$$L_1 = \underbrace{\begin{bmatrix} 2 \\ 0 \\ 0 \\ 1 \end{bmatrix}}_{\boldsymbol{x}_1} + \underbrace{\text{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \right\}}_{U_1}, \qquad L_2 = \underbrace{\begin{bmatrix} 3 \\ 1 \\ 0 \\ 0 \end{bmatrix}}_{\boldsymbol{x}_2} + \underbrace{\text{span}\left\{ \begin{bmatrix} -1 \\ 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\}}_{U_2} \quad (1.172)$$

We want to determine the intersection $L_1 \cap L_2$. We know that we have: $\boldsymbol{x} \in L_1 \cap L_2 \Leftrightarrow \boldsymbol{x} \in L_1 \wedge \boldsymbol{x} \in L_2 \Leftrightarrow \exists \lambda_1, \lambda_2, \lambda_3, \gamma_1, \gamma_2 \in \mathbb{R}$, such that

$$\underbrace{\begin{bmatrix} 2 \\ 0 \\ 0 \\ 1 \end{bmatrix} + \lambda_1 \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} + \lambda_3 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}}_{\in L_1} = \boldsymbol{x} = \underbrace{\begin{bmatrix} 3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \gamma_1 \begin{bmatrix} -1 \\ 1 \\ 2 \\ 0 \end{bmatrix} + \gamma_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}}_{\in L_2}. \quad (1.173)$$

From this, we obtain the following inhomogeneous equation system with unknowns $\lambda_1, \lambda_2, \lambda_3, \gamma_1, \gamma_2$:

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

Using Gaussian elimination, we quickly obtain the reduced row echelon form

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

We exploit the reduced row echelon form to obtain the general solution

$$\begin{bmatrix}\lambda_1\\\lambda_2\\\lambda_3\\\gamma_1\\\gamma_2\end{bmatrix}=\begin{bmatrix}1\\1\\-1\\0\\1\end{bmatrix}+\alpha\begin{bmatrix}1\\-2\\0\\-1\\0\end{bmatrix},\quad \alpha\in\mathbb{R}. \tag{1.174}$$

Note that we have only determined the general solution for the coordinates of the linear combinations that describe the intersection $L_1 \cap L_2$, see (1.173). To obtain the intersection, we use $\lambda_1, \lambda_2, \lambda_3$ from (1.174) in the left-hand-side of (1.173) or, alternatively, γ_1, γ_2 in the left-hand-side of (1.173) to obtain the intersection.
Let us do both—we should obtain the same result.

- Let us start with $\lambda_1, \lambda_2, \lambda_3$ to obtain $L_1 \cap L_2$ via L_1. From (1.173), we know that $\boldsymbol{x} \in L_1 \cap L_2$ can be written as $\boldsymbol{x}_1 + \sum_{i=1}^{3} \lambda_i \boldsymbol{b}_i$ where the $\boldsymbol{b}_i$ are the basis vectors of U_1. With the λ_i in (1.174), we, therefore, obtain

$$L_1\cap L_2=\begin{bmatrix}2\\0\\0\\1\end{bmatrix}+(1+\alpha)\begin{bmatrix}1\\1\\0\\0\end{bmatrix}+(1-2\alpha)\begin{bmatrix}0\\1\\1\\0\end{bmatrix}-\begin{bmatrix}0\\0\\1\\1\end{bmatrix} \tag{1.175}$$

$$=\begin{bmatrix}3\\2\\0\\0\end{bmatrix}+\alpha\begin{bmatrix}1\\-1\\-2\\0\end{bmatrix}=\begin{bmatrix}3\\2\\0\\0\end{bmatrix}+\operatorname{span}\left\{\begin{bmatrix}-1\\1\\2\\0\end{bmatrix}\right\}. \tag{1.176}$$

- Now, we consider γ_1, γ_2 to obtain $L_1 \cap L_2$ via L_2. From (1.174), we obtain $\gamma_1 = \alpha \in \mathbb{R}$ is a free parameter, and $\gamma_2 = 1$. Plugging this into (1.173), we obtain that any $\boldsymbol{x} \in L_1 \cap L_2$ satisfies

$$L_1\cap L_2=\begin{bmatrix}3\\1\\0\\0\end{bmatrix}+\gamma_1\begin{bmatrix}-1\\1\\2\\0\end{bmatrix}+1\cdot\begin{bmatrix}0\\1\\0\\0\end{bmatrix}=\begin{bmatrix}3\\2\\0\\0\end{bmatrix}+\gamma_1\begin{bmatrix}-1\\1\\2\\0\end{bmatrix}=\begin{bmatrix}3\\2\\0\\0\end{bmatrix}+\operatorname{span}\left\{\begin{bmatrix}-1\\1\\2\\0\end{bmatrix}\right\}, \tag{1.177}$$

We see that the solutions in (1.176) and (1.177) are identical. Overall, $L_1 \cap L_2$ is a line (1D affine subspace) in $\mathbb{R}^4$.

Remark 41
The intersection of a finite set of hyperplanes in $\mathbb{R}^n$ is either empty or an affine subspace. Moreover, every k-dimensional affine subspace of $\mathbb{R}^n$ is the intersection of finitely many hyperplanes.

Definition 30 (Parallelism)
Two affine subspaces L_1 and L_2 are **parallel** *($L_1 \| L_2$) if the following holds for the corresponding direction spaces U_1, U_2: $U_1 \subset U_2$ or $U_2 \subset U_1$.*[42]

[42] Note that this definition of parallel allows for $L_1 \subset L_2$ or $L_2 \subset L_1$.

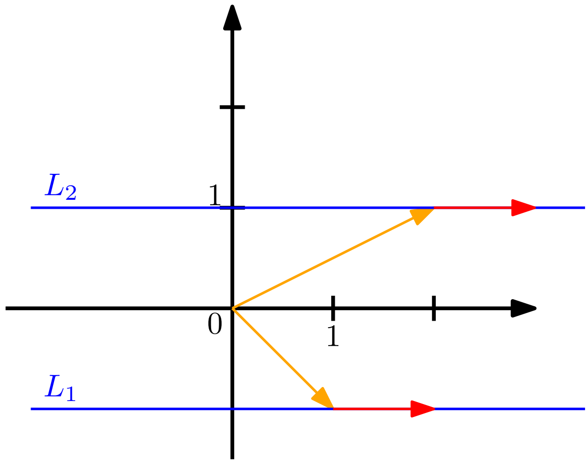


Figure 1.11: Parallel lines. The affine subspaces L_1 and L_2 are parallel with $L_1 \cap L_2 = \emptyset$.

Remark 42
Parallel affine subspaces, which do not contain each other (such that $L_1 \subset L_2$ or $L_2 \subset L_1$), have no points in common, i.e., $L_1 \cap L_2 = \emptyset$.

Example (Parallel affine subspaces)
The affine subspaces

$$L_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \operatorname{span}\left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right\}, \quad L_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \operatorname{span}\left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right\} \tag{1.178}$$

are parallel ($U_1 = U_2$ in this case) and $L_1 \cap L_2 = \emptyset$ because the lines are offset as illustrated in Figure 1.11.

We talk about **skew lines** if they are neither parallel nor have an intersection point. Imagine non-parallel lines in 3D that "miss" each other.

Example

1. Consider two lines $g = \boldsymbol{x}_1 + U_1, h = \boldsymbol{x}_2 + U_2$ in $\mathbb{R}^2$, where $U_1, U_2 \subset \mathbb{R}$ are one-dimensional subspaces.
   (a) If $g \cap h \neq \emptyset$:
      - For $\dim(U_1 \cap U_2) = 0$, we get a single point as the intersection.
      - For $\dim(U_1 \cap U_2) = 1$ it follows that $U_1 = U_2$ and $g = h$.

   (b) If $g \cap h = \emptyset$:
      - For $\dim(U_1 \cap U_2) = 1$ it follows again that $U_1 = U_2$, and $g \| h$, $g \neq h$ (because they do not intersect).
      - The case $\dim(U_1 \cap U_2) = 0$ cannot happen in $\mathbb{R}^2$.
2. Consider two lines $g = \boldsymbol{x}_1 + U_1, h = \boldsymbol{x}_2 + U_2$ in $\mathbb{R}^n, n \geq 3$

(a) For $g \cap h \neq \emptyset$ we obtain (as in $\mathbb{R}^2$) that either g and h intersect in a single point or $g = h$.

(b) If $g \cap h = \emptyset$:

- For $\dim(U_1 \cap U_2) = 0$, we obtain that g and h are skew lines (this cannot happen in $\mathbb{R}^2$). This means, there exists no plane that contains both g and h.
- If $\dim(U_1 \cap U_2) = 1$ it follows that $g \| h$.

3. Consider two hyper-planes $L_1 = \boldsymbol{x}_1 + U_1$ and $L_2 = \boldsymbol{x}_2 + U_2$ in $\mathbb{R}^n, n = 3$, where $U_1, U_2 \subset \mathbb{R}^2$ are two-dimensional subspaces.

(a) If $L_1 \cap L_2 \neq \emptyset$ the intersection is an affine subspace. The kind of affine subspace depends on the dimension $\dim(U_1 \cap U_2)$ of the intersection of the corresponding direction spaces.

- $\dim(U_1 \cap U_2) = 2$: Then $U_1 = U_2$ and, therefore, $L_1 = L_2$.
- $\dim(U_1 \cap U_2) = 1$: The intersection is a line.
- $\dim(U_1 \cap U_2) = 0$: Cannot happen in $\mathbb{R}^3$.

(b) If $L_1 \cap L_2 = \emptyset$, then $\dim(U_1 \cap U_2) = 2$ (no other option possible in $\mathbb{R}^3$) and $L_1 \| L_2$.

4. Consider two planes $L_1 = \boldsymbol{x}_1 + U_1$ and $L_2 = \boldsymbol{x}_2 + U_2$ in $\mathbb{R}^n, n = 4$, where $U_1, U_2 \subset \mathbb{R}^2$ are two-dimensional subspaces.

(a) For $L_1 \cap L_2 \neq \emptyset$ the additional case is possible that the planes intersect in a point.

(b) For $L_1 \cap L_2 = \emptyset$ the additional case is possible that $\dim(U_1 \cap U_2) = 1$. This means that the planes are not parallel, they have no point in common, but there is a line g such that $g \| L_1$ and $g \| L_2$.

5. For two planes $L_1 = \boldsymbol{x}_1 + U_1$ and $L_2 = \boldsymbol{x}_2 + U_2$ in $\mathbb{R}^n, n > 4$, where $U_1, U_2 \subset \mathbb{R}^2$ are two-dimensional subspaces, all kinds of intersections are possible.

1.15 Affine Mappings

Definition 31 (Affine mapping)
For two vector spaces V, W and a linear mapping $\Phi : V \to W$ and $\boldsymbol{a} \in W$ the mapping

$$\psi : V \to W \tag{1.179}$$

$$\boldsymbol{x} \mapsto \boldsymbol{a} + \Phi(\boldsymbol{x}) \tag{1.180}$$

is an **affine mapping** *from V to W. The vector $\boldsymbol{a}$ is called the* **translation vector** *of ψ.*

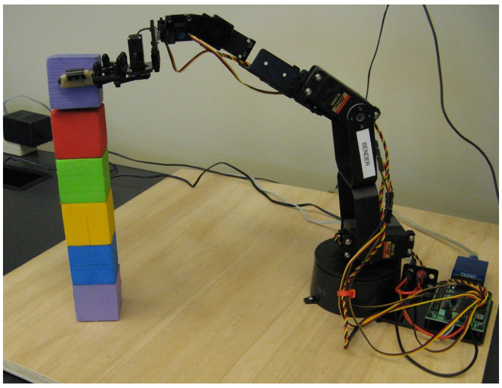


Figure 1.12: The robotic arm needs to rotate its joints in order to pick up objects or to place them correctly. Figure taken from (Deisenroth et al., 2015).

- Every affine mapping $\psi : V \to W$ is also the composition of a linear mapping $\Phi : V \to W$ and a translation $\tau : W \to W$ in W, such that $\psi = \tau \circ \Phi$. The mappings Φ and τ are uniquely determined.
- The composition $\psi' \circ \psi$ of affine mappings $\psi : V \to W$, $\psi' : W \to X$ is affine.
- Affine mappings keep the geometric structure invariant, and preserve the dimension and parallelism.

Definition 32
For a map $f : A \to B$ of sets A and B and a subset $C \subset A$, the forward image of C by f is defined as $f[C] = \{f(x)|x \in C\}$.

For example, for a linear map $\Phi : \mathbb{R}^n \to \mathbb{R}^m$, the forward image of $\mathbb{R}^n$ by Φ is simply the image space of Φ, i.e., $\Phi[\mathbb{R}^n] = \mathrm{Im}(\Phi)$.

Theorem 14
Let V, W be vector spaces and $\psi : V \to W$ an affine mapping. Then:

- *If L is an affine subspace of V then $\psi[L]$ is an affine subspace of W and* $\dim(\psi[L]) \leq \dim(W)$.
- *ψ preserves parallelism, i.e., for all affine subspaces $L_1, L_2 \subset V$ it follows that if $L_1 \| L_2$ in V then $\psi[L_1] \| \psi[L_2]$ in W.*

1.16 Rotations

An important category of linear mappings are rotations. A rotation in$\mathbb{R}^2$ rotates an object (counterclockwise) by an angle θ about the origin. Important application areas of rotations include computer graphics and robotics. For example, in robotics, it is often important to know how to rotate the joints of a robotic arm in order to pick up or place an object, see Figure 1.12.

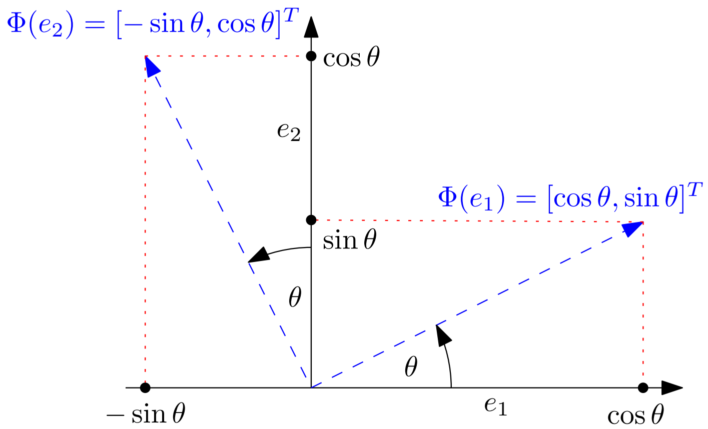


Figure 1.13: Rotation of the standard basis in $\mathbb{R}^2$ by an angle θ.

1.16.1 Rotations in the Plane

Consider the standard basis in $\mathbb{R}^2$ given by $(\boldsymbol{e}_1, \boldsymbol{e}_2)$ with

$$\boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

which defines the standard coordinate system in $\mathbb{R}^2$. Assume we want to rotate this coordinate system by an angle θ as illustrated in Figure 1.13. Note that the rotated vectors are still linearly independent and, therefore, are a basis of $\mathbb{R}^2$. This means that the rotation performs a basis change.
Since rotations Φ are linear mappings, we can express them by a **rotation matrix** $\boldsymbol{R}(\theta)$. Trigonometry allows us to determine the coordinates of the rotated axes with respect to the standard basis in $\mathbb{R}^2$. We obtain

$$\Phi(\boldsymbol{e}_1) = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}, \quad \Phi(\boldsymbol{e}_2) = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}. \tag{1.181}$$

Therefore, the rotation matrix that performs the basis change into the rotated coordinates $\boldsymbol{R}(\theta)$ is given as

$$\boldsymbol{R}(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{1.182}$$

Generally, in $\mathbb{R}^2$ rotations of any vector $\boldsymbol{x} \in \mathbb{R}^2$ by an angle θ are given by

$$\boldsymbol{x}' = \Phi(\boldsymbol{x}) = \boldsymbol{R}(\theta)\boldsymbol{x} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1\cos\theta - x_2\sin\theta \\ x_1\cos\theta + x_2\sin\theta \end{bmatrix}. \tag{1.183}$$

1.16.2 Rotations in Three Dimensions

In three dimensions, we have to define what "counterclockwise" means: We will go by the convention that a "counterclockwise" (planar) rotation about an axis refers to a rotation about an axis when we look at the axis "from the end toward the origin". In $\mathbb{R}^3$ there are three (planar) rotations about three standard basis vectors (see Figure 1.14):

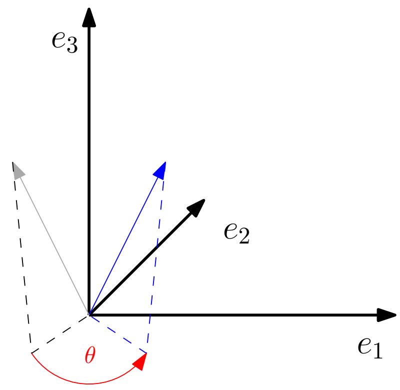


Figure 1.14: Rotation of a general vector (gray) in $\mathbb{R}^3$ by an angle θ about the e_3-axis. The rotated vector is shown in blue.

- Counterclockwise rotation about the e_1-axis

$$R_1(\theta) = \begin{bmatrix}\Phi(e_1) & \Phi(e_2) & \Phi(e_3)\end{bmatrix} = \begin{bmatrix}1 & 0 & 0\\ 0 & \cos\theta & -\sin\theta\\ 0 & \sin\theta & \cos\theta\end{bmatrix} \tag{1.184}$$

  Here, the e_1 coordinate is fixed, and the counterclockwise rotation is performed in the e_2e_3 plane.

- Counterclockwise rotation about the e_2-axis

$$R_2(\theta) = \begin{bmatrix}\cos\theta & 0 & \sin\theta\\ 0 & 1 & 0\\ -\sin\theta & 0 & \cos\theta\end{bmatrix} \tag{1.185}$$

  If we rotate the e_1e_3 plane about the e_2 axis, we need to look at the e_2 axis from its “tip” toward the origin.

- Counterclockwise rotation about the e_3-axis

$$R_3(\theta) = \begin{bmatrix}\cos\theta & -\sin\theta & 0\\ \sin\theta & \cos\theta & 0\\ 0 & 0 & 1\end{bmatrix} \tag{1.186}$$

  Figure 1.14 to illustrates this.

1.16.3 Rotations in n Dimensions

The generalization of rotations from 2D and 3D to n-dimensional Euclidean vector spaces can be intuitively described as keeping $n-2$ dimensions fix and restrict the rotation to a two-dimensional plane in the n-dimensional space.

Definition 33 (Givens Rotation)
Let V be an n-dimensional Euclidean vector space and $\Phi : V \to V$ an automorphism with transformation matrix

$$\boldsymbol{R}_{ij}(\theta) := \begin{bmatrix} \boldsymbol{I}_{i-1} & \boldsymbol{0} & \cdots & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \cos\theta & \boldsymbol{0} & -\sin\theta & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{j-i-1} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \sin\theta & \boldsymbol{0} & \cos\theta & \boldsymbol{0} \\ \boldsymbol{0} & \cdots & \cdots & \boldsymbol{0} & \boldsymbol{I}_{n-j} \end{bmatrix} \in \mathbb{R}^{n\times n}, \qquad \theta \in \mathbb{R}, \tag{1.187}$$

for $1 \leq i < j \leq n$. Then $\boldsymbol{R}_{ij}(\theta)$ is called a **Givens rotation** *(named after Wallace Givens). Essentially, $\boldsymbol{R}_{ij}(\theta)$ is the identity matrix $\boldsymbol{I}_n$ with*

$$r_{ii} = \cos\theta, \quad r_{ij} = -\sin\theta, \quad r_{ji} = \sin\theta, \quad r_{jj} = \cos\theta. \tag{1.188}$$

In two dimensions (i.e., $n = 2$), we obtain (1.182) *as a special case.*

1.16.4 Properties of Rotations

Rotations exhibit some useful properties:

- The composition of rotations can be expressed as a single rotation where the angles are summed up:

$$\boldsymbol{R}(\phi)\boldsymbol{R}(\theta) = \boldsymbol{R}(\phi + \theta) \tag{1.189}$$

- Rotations preserve lengths, i.e., $\|\boldsymbol{x}\| = \|\boldsymbol{R}(\phi)\boldsymbol{x}\|$. The original vector and the rotated vector have the same length. This also implies that $\det(\boldsymbol{R}(\phi)) = 1$

- Rotations preserve distances, i.e., $\|\boldsymbol{x} - \boldsymbol{y}\| = \|\boldsymbol{R}_\theta(\boldsymbol{x}) - \boldsymbol{R}_\theta(\boldsymbol{y})\|$. In other words, rotations leave the distance between any two points unchanged after the transformation

- Rotations in three (or more) dimensions are generally not commutative. Therefore, the order in which rotations are applied is important, even if they rotate about the same point. Only in two dimensions, rotations are commutative, such that $\boldsymbol{R}(\phi)\boldsymbol{R}(\theta) = \boldsymbol{R}(\theta)\boldsymbol{R}(\phi)$ for all $\phi, \theta \in [0, 2\pi)$.

- Rotations in the plane have eigenvalues $\exp \pm i\theta$, where θ is the angle of rotation.

Bibliography

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer-Verlag. pages 66

Deisenroth, M. P., Fox, D., and Rasmussen, C. E. (2015). Gaussian Processes for Data-Efficient Learning in Robotics and Control. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2):408–423. pages 73

Deng, L., Seltzer, M. L., Yu, D., Acero, A., Mohamed, A.-r., and Hinton, G. E. (2010). Binary Coding of Speech Spectrograms using a Deep Auto-Encoder. In *Interspeech*, pages 1692–1695. pages 60

Drumm, V. and Weil, W. (2001). Lineare Algebra und Analytische Geometrie. Lecture Notes, Universität Karlsruhe. pages 1

Hogben, L., editor (2013). *Handbook of Linear Algebra*. Discrete Mathematics and Its Applications. Chapman and Hall, 2nd edition. pages 1

Hotelling, H. (1933). Analysis of a Complex of Statistical Variables into Principal Components. *Journal of Educational Psychology*, 24:417–441. pages 51, 60, 66

Liesen, J. and Mehrmann, V. (2015). *Linear Algebra*. Springer Undergraduate Mathematics Series. Springer. pages 1, 52

Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. The MIT Press, Cambridge, MA, USA. pages 60

Schölkopf, B. and Smola, A. J. (2002). *Learning with Kernels—Support Vector Machines, Regularization, Optimization, and Beyond*. Adaptive Computation and Machine Learning. The MIT Press, Cambridge, MA, USA. pages 60

Schölkopf, B., Smola, A. J., and Müller, K.-R. (1998). Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Computation*, 10(5):1299–1319. pages 60

Stoer, J. and Burlirsch, R. (2002). *Introduction to Numerical Analysis*. Springer. pages 60

Strang, G. (2003). *Introduction to Linear Algebra*. Wellesley-Cambridge Press, 3rd edition. pages 1

Wikipedia (2015). Linear Dependence. https://en.wikipedia.org/wiki/Linear_independence. pages 20