

2

761

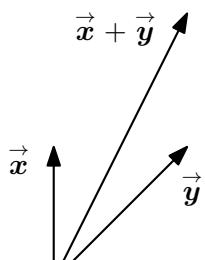
Linear Algebra

762 When formalizing intuitive concepts, a common approach is to construct
 763 a set of objects (symbols) and a set of rules to manipulate these objects.
 764 This is known as an *algebra*.

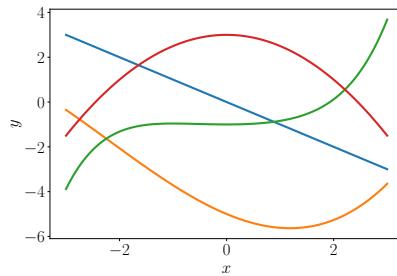
765 Linear algebra is the study of vectors. The vectors many of us know
 766 from school are called “geometric vectors”, which are usually denoted by
 767 having a small arrow above the letter, e.g., \vec{x} and \vec{y} . In this book, we
 768 discuss more general concepts of vectors and use a bold letter to represent
 769 them, e.g., x and y .

770 In general, vectors are special objects that can be added together and
 771 multiplied by scalars to produce another object of the same kind. Any
 772 object that satisfies these two properties can be considered a vector. Here
 773 are some examples of such vector objects:

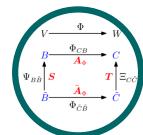
1. Geometric vectors. This example of a vector may be familiar from school.
 775 Geometric vectors are directed segments, which can be drawn, see
 776 Figure 2.1(a). Two geometric vectors \vec{x} , \vec{y} can be added, such that
 777 $\vec{x} + \vec{y} = \vec{z}$ is another geometric vector. Furthermore, multiplication
 778 by a scalar λ \vec{x} , $\lambda \in \mathbb{R}$ is also a geometric vector. In fact, it is the
 779 original vector scaled by λ . Therefore, geometric vectors are instances
 780 of the vector concepts introduced above.
2. Polynomials are also vectors, see Figure 2.1(b): Two polynomials can
 781 be added together, which results in another polynomial; and they can
 782 be multiplied by a scalar $\lambda \in \mathbb{R}$, and the result is a polynomial as
 783 well. Therefore, polynomials are (rather unusual) instances of vectors.



(a) Geometric vectors.



(b) Polynomials.



algebra

Figure 2.1
 Different types of
 vectors. Vectors can
 be surprising
 objects, including
 (a) geometric
 vectors and (b)
 polynomials.

785 Note that polynomials are very different from geometric vectors. While
 786 geometric vectors are concrete “drawings”, polynomials are abstract
 787 concepts. However, they are both vectors in the sense described above.

- 788 3. Audio signals are vectors. Audio signals are represented as a series of
 789 numbers. We can add audio signals together, and their sum is a new
 790 audio signal. If we scale an audio signal, we also obtain an audio signal.
 791 Therefore, audio signals are a type of vector, too.
- 792 4. Elements of \mathbb{R}^n are vectors. In other words, we can consider each el-
 793 ement of \mathbb{R}^n (the tuple of n real numbers) to be a vector. \mathbb{R}^n is more
 794 abstract than polynomials, and it is the concept we focus on in this
 book. For example,

$$\mathbf{a} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \in \mathbb{R}^3 \quad (2.1)$$

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

795 Linear algebra focuses on the similarities between these vector concepts.
 796 We can add them together and multiply them by scalars. We will largely
 797 focus on vectors in \mathbb{R}^n since most algorithms in linear algebra are for-
 798 mulated in \mathbb{R}^n . Recall that in machine learning, we often consider data
 799 to be represented as vectors in \mathbb{R}^n . In this book, we will focus on finite-
 800 dimensional vector spaces, in which case there is a 1:1 correspondence
 801 between any kind of (finite-dimensional) vector and \mathbb{R}^n . By studying \mathbb{R}^n ,
 802 we implicitly study all other vectors such as geometric vectors and poly-
 803 nomials. Although \mathbb{R}^n is rather abstract, it is most useful.

804 One major idea in mathematics is the idea of “closure”. This is the ques-
 805 tion: What is the set of all things that can result from my proposed oper-
 806 ations? In the case of vectors: What is the set of vectors that can result by
 807 starting with a small set of vectors, and adding them to each other and
 808 scaling them? This results in a vector space (Section 2.4). The concept of
 809 a vector space and its properties underlie much of machine learning.

810 A closely related concept is a *matrix*, which can be thought of as a
 811 collection of vectors. As can be expected, when talking about properties
 812 of a collection of vectors, we can use matrices as a representation. The
 813 concepts introduced in this chapter are shown in Figure 2.2

814 Pavel Grinfeld’s
 815 series on linear
 816 algebra:
 817 <http://tinyurl.com/nahclwm>

818 Gilbert Strang’s
 819 course on linear
 820 algebra:
 821 <http://tinyurl.com/29p5q8j>

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. Another excellent
 source is Gilbert Strang’s Linear Algebra course at MIT.

818 Linear algebra plays an important role in machine learning and gen-
 819 eral mathematics. In Chapter 5, we will discuss vector calculus, where
 820 a principled knowledge of matrix operations is essential. In Chapter 10,

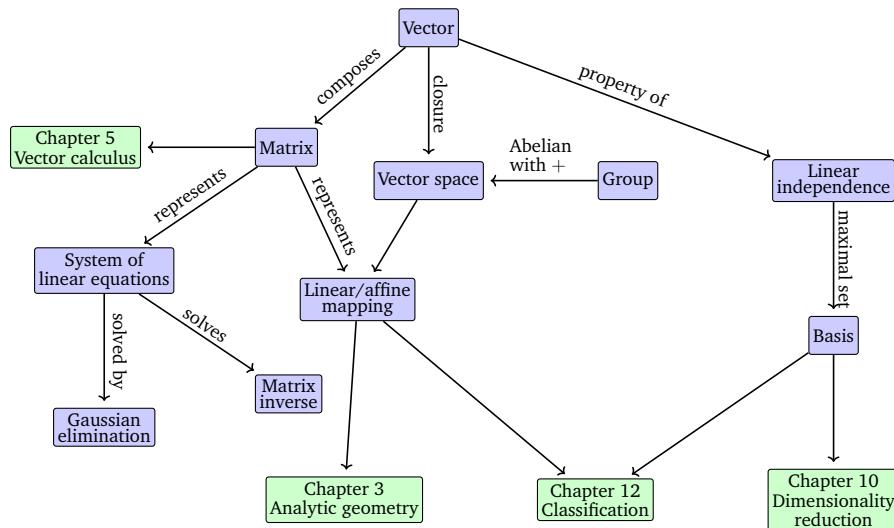


Figure 2.2 A mind map of the concepts introduced in this chapter, along with when they are used in other parts of the book.

- 821 we will use projections (to be introduced in Section 3.7) for dimensionality reduction with Principal Component Analysis (PCA). In Chapter 9, we
 822 will discuss linear regression where linear algebra plays a central role for
 823 solving least-squares problems.
 824

2.1 Systems of Linear Equations

- 825 Systems of linear equations play a central part of linear algebra. Many
 826 problems can be formulated as systems of linear equations, and linear
 827 algebra gives us the tools for solving them.
 828

Example 2.1

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 of 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

$$a_{11}x_1 + \dots + a_{in}x_n \quad (2.2)$$

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

$$\begin{aligned} a_{11}x_1 + \dots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + \dots + a_{mn}x_n &= b_m \end{aligned} \quad (2.3)$$

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

system of linear equations
 x_1, \dots, x_n are the *unknowns* of this system of linear equations. Every n -tuple $(x_1, \dots, x_n) \in \mathbb{R}^n$ that satisfies (2.3) is a *solution* of the linear equation system.
solution

Example 2.2

The system of linear equations

$$\begin{aligned} x_1 + x_2 + x_3 &= 3 & (1) \\ x_1 - x_2 + 2x_3 &= 2 & (2) \\ 2x_1 + 3x_3 &= 1 & (3) \end{aligned} \quad (2.4)$$

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

Let us have a look at the system of linear equations

$$\begin{aligned} x_1 + x_2 + x_3 &= 3 & (1) \\ x_1 - x_2 + 2x_3 &= 2 & (2) \\ x_2 + x_3 &= 2 & (3) \end{aligned} \quad (2.5)$$

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

As a third example, we consider

$$\begin{aligned} x_1 + x_2 + x_3 &= 3 & (1) \\ x_1 - x_2 + 2x_3 &= 2 & (2) \\ 2x_1 + 3x_3 &= 5 & (3) \end{aligned} \quad (2.6)$$

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

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

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

In general, for a real-valued system of linear equations we obtain either no, exactly one or infinitely many solutions.

Remark (Geometric Interpretation of Systems of Linear Equations). In a system of linear equations with two variables x_1, x_2 , each linear equation determines a line on the $x_1 x_2$ -plane. Since a solution to a system of lin-

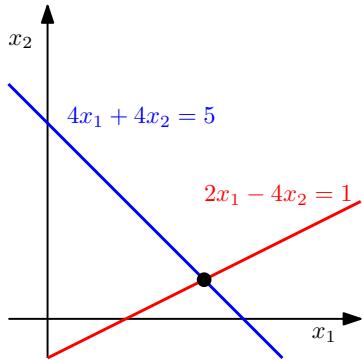


Figure 2.3 The solution space of a system of two linear equations with two variables can be geometrically interpreted as the intersection of two lines. Every linear equation represents a line.

ear equations must satisfy all equations simultaneously, the solution set
is the intersection of these line. This intersection can be a line (if the lin-
ear equations describe the same line), a point, or empty (when the lines
are parallel). An illustration is given in Figure 2.3. Similarly, for three
variables, each linear equation determines a plane in three-dimensional
space. When we intersect these planes, i.e., satisfy all linear equations at
the same time, we can end up with solution set that is a plane, a line, a
point or empty (when the planes are parallel). ◇

For a systematic approach to solving systems of linear equations, we will introduce a useful compact notation. We will write the system from (2.3) in the following form:

$$x_1 \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \quad (2.8)$$

$$\iff \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}. \quad (2.9)$$

In the following, we will have a close look at these *matrices* and define computation rules.

848

2.2 Matrices

Matrices play a central role in linear algebra. They can be used to compactly represent systems of linear equations, but they also represent linear functions (linear mappings) as we will see later in Section 2.7. Before we discuss some of these interesting topics, let us first define what a matrix is and what kind of operations we can do with matrices.

Definition 2.1 (Matrix). With $m, n \in \mathbb{N}$ a real-valued (m, n) *matrix* \mathbf{A} is an $m \cdot n$ -tuple of elements a_{ij} , $i = 1, \dots, m$, $j = 1, \dots, n$, which is ordered

matrix

according to a rectangular scheme consisting of m rows and n columns:

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

rows 854 $(1, n)$ -matrices are called *rows*, $(m, 1)$ -matrices are called *columns*. These
columns 855 special matrices are also called *row/column vectors*.

row/column vectors 856 $\mathbb{R}^{m \times n}$ is the set of all real-valued (m, n) -matrices. $\mathbf{A} \in \mathbb{R}^{m \times n}$ can be
857 equivalently represented as $\mathbf{a} \in \mathbb{R}^{mn}$ by stacking all n columns of the
858 matrix into a long vector.

859

2.2.1 Matrix Addition and Multiplication

The sum of two matrices $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{B} \in \mathbb{R}^{m \times n}$ is defined as the element-wise sum, i.e.,

$$\mathbf{A} + \mathbf{B} := \begin{bmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{bmatrix} \in \mathbb{R}^{m \times n}. \quad (2.11)$$

Note the size of the
matrices.

$\mathbf{C} =$
`np.einsum('il,
lj', A, B)`

There are n columns
in \mathbf{A} and n rows in
 \mathbf{B} , such that we can
compute $a_{il}b_{lj}$ for
 $l = 1, \dots, n$.

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

$$c_{ij} = \sum_{l=1}^n a_{il}b_{lj}, \quad i = 1, \dots, m, \quad j = 1, \dots, k. \quad (2.12)$$

This means, to compute element c_{ij} we multiply the elements of the i th row of \mathbf{A} with the j th column of \mathbf{B} and sum them up. Later in Section 3.2, we will call this the *dot product* of the corresponding row and column.

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

$$\underbrace{\mathbf{A}}_{n \times k} \underbrace{\mathbf{B}}_{k \times m} = \underbrace{\mathbf{C}}_{n \times m} \quad (2.13)$$

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

865 *Remark.* Matrix multiplication is *not* defined as an element-wise operation
866 on matrix elements, i.e., $c_{ij} \neq a_{ij}b_{ij}$ (even if the size of \mathbf{A} , \mathbf{B} was chosen
867 appropriately). This kind of element-wise multiplication often appears
868 in programming languages when we multiply (multi-dimensional) arrays
869 with each other. \diamond

Example 2.3

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

$$\mathbf{AB} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 2 & 5 \end{bmatrix} \in \mathbb{R}^{2 \times 2}, \quad (2.14)$$

$$\mathbf{BA} = \begin{bmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 4 & 2 \\ -2 & 0 & 2 \\ 3 & 2 & 1 \end{bmatrix} \in \mathbb{R}^{3 \times 3}. \quad (2.15)$$

From this example, we can already see that matrix multiplication is not commutative, i.e., $\mathbf{AB} \neq \mathbf{BA}$, see also Figure 2.4 for an illustration.

Definition 2.2 (Identity Matrix). In $\mathbb{R}^{n \times n}$, we define the *identity matrix* as

$$\mathbf{I}_n = \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{n \times n} \quad (2.16)$$

as the $n \times n$ -matrix containing 1 on the diagonal and 0 everywhere else. With this, $\mathbf{A} \cdot \mathbf{I}_n = \mathbf{A} = \mathbf{I}_n \cdot \mathbf{A}$ for all $\mathbf{A} \in \mathbb{R}^{n \times n}$.

Now that we have defined matrix multiplication, matrix addition and the identity matrix, let us have a look at some properties of matrices, where we will omit the “.” for matrix multiplication:

- Associativity:

$$\forall \mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{B} \in \mathbb{R}^{n \times p}, \mathbf{C} \in \mathbb{R}^{p \times q} : (\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}) \quad (2.17)$$

- Distributivity:

$$\forall \mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times n}, \mathbf{C}, \mathbf{D} \in \mathbb{R}^{n \times p} : (\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC} \quad (2.18a)$$

$$\mathbf{A}(\mathbf{C} + \mathbf{D}) = \mathbf{AC} + \mathbf{AD} \quad (2.18b)$$

- Neutral element:

$$\forall \mathbf{A} \in \mathbb{R}^{m \times n} : \mathbf{I}_m \mathbf{A} = \mathbf{A} \mathbf{I}_n = \mathbf{A} \quad (2.19)$$

Note that $\mathbf{I}_m \neq \mathbf{I}_n$ for $m \neq n$.

Figure 2.4 Even if both matrix multiplications \mathbf{AB} and \mathbf{BA} are defined, the dimensions of the results can be different.

identity matrix

878

2.2.2 Inverse and Transpose

A square matrix 879 possesses the same 880 number of columns 881 and rows.

inverse

regular

invertible

non-singular

singular

non-invertible

Definition 2.3 (Inverse). For a square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ a matrix $\mathbf{B} \in \mathbb{R}^{n \times n}$ with $\mathbf{AB} = \mathbf{I}_n = \mathbf{BA}$ the matrix \mathbf{B} is called *inverse* and denoted by \mathbf{A}^{-1} .

Unfortunately, not every matrix \mathbf{A} possesses an inverse \mathbf{A}^{-1} . If this inverse does exist, \mathbf{A} is called *regular/invertible/non-singular*, otherwise *singular/non-invertible*.

Remark (Existence of the Inverse of a 2×2 -Matrix). Consider a matrix

$$\mathbf{A} := \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \in \mathbb{R}^{2 \times 2}. \quad (2.20)$$

If we multiply \mathbf{A} with

$$\mathbf{B} := \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \quad (2.21)$$

we obtain

$$\mathbf{AB} = \begin{bmatrix} a_{11}a_{22} - a_{12}a_{21} & 0 \\ 0 & a_{11}a_{22} - a_{12}a_{21} \end{bmatrix} = (a_{11}a_{22} - a_{12}a_{21})\mathbf{I} \quad (2.22)$$

so that

$$\mathbf{A}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \quad (2.23)$$

if and only if $a_{11}a_{22} - a_{12}a_{21} \neq 0$. In Section 4.1, we will see that $a_{11}a_{22} - a_{12}a_{21}$ is the determinant of a 2×2 -matrix. Furthermore, we can generally use the determinant to check whether a matrix is invertible. ◇

Example 2.4 (Inverse Matrix)

The matrices

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 1 \\ 4 & 4 & 5 \\ 6 & 7 & 7 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} -7 & -7 & 6 \\ 2 & 1 & -1 \\ 4 & 5 & -4 \end{bmatrix} \quad (2.24)$$

are inverse to each other since $\mathbf{AB} = \mathbf{I} = \mathbf{BA}$.

Definition 2.4 (Transpose). For $\mathbf{A} \in \mathbb{R}^{m \times n}$ the matrix $\mathbf{B} \in \mathbb{R}^{n \times m}$ with $b_{ij} = a_{ji}$ is called the *transpose* of \mathbf{A} . We write $\mathbf{B} = \mathbf{A}^\top$.

For a square matrix \mathbf{A}^\top is the matrix we obtain when we “mirror” \mathbf{A} on its main diagonal. In general, \mathbf{A}^\top can be obtained by writing the columns of \mathbf{A} as the rows of \mathbf{A}^\top .

Some important properties of inverses and transposes are:

$$\mathbf{AA}^{-1} = \mathbf{I} = \mathbf{A}^{-1}\mathbf{A} \quad (2.25)$$

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \quad (2.26)$$

transpose 888
The main diagonal 889 (sometimes called 890 “principal diagonal” 891 “primary diagonal” 892 “leading diagonal”, 893 or “major diagonal”) 894 of a matrix \mathbf{A} is the 895 collection of entries 896 A_{ij} where $i = j$.

In the scalar case

$$\frac{1}{2+4} = \frac{1}{6} \neq \frac{1}{2} + \frac{1}{4}.$$

$$(\mathbf{A} + \mathbf{B})^{-1} \neq \mathbf{A}^{-1} + \mathbf{B}^{-1} \quad (2.27)$$

$$(\mathbf{A}^\top)^\top = \mathbf{A} \quad (2.28)$$

$$(\mathbf{A} + \mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top \quad (2.29)$$

$$(\mathbf{AB})^\top = \mathbf{B}^\top \mathbf{A}^\top \quad (2.30)$$

Moreover, if \mathbf{A} is invertible then so is \mathbf{A}^\top and $(\mathbf{A}^{-1})^\top = (\mathbf{A}^\top)^{-1} =: \mathbf{A}^{-\top}$

A matrix \mathbf{A} is *symmetric* if $\mathbf{A} = \mathbf{A}^\top$. Note that this can only hold for

(n, n) -matrices, which we also call *square matrices* because they possess

the same number of rows and columns.

symmetric
square matrices

Remark (Sum and Product of Symmetric Matrices). The sum of symmetric matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$ is always symmetric. However, although their product is always defined, it is generally not symmetric:

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}. \quad (2.31)$$

◇

897

2.2.3 Multiplication by a Scalar

Let us have a brief look at what happens to matrices when they are multiplied by a scalar $\lambda \in \mathbb{R}$. Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\lambda \in \mathbb{R}$. Then $\lambda \mathbf{A} = \mathbf{K}$, $K_{ij} = \lambda a_{ij}$. Practically, λ scales each element of \mathbf{A} . For $\lambda, \psi \in \mathbb{R}$ it holds:

- Distributivity:

$$(\lambda + \psi)\mathbf{C} = \lambda\mathbf{C} + \psi\mathbf{C}, \quad \mathbf{C} \in \mathbb{R}^{m \times n}$$

$$\lambda(\mathbf{B} + \mathbf{C}) = \lambda\mathbf{B} + \lambda\mathbf{C}, \quad \mathbf{B}, \mathbf{C} \in \mathbb{R}^{m \times n}$$

- Associativity:

$$(\lambda\psi)\mathbf{C} = \lambda(\psi\mathbf{C}), \quad \mathbf{C} \in \mathbb{R}^{m \times n}$$

$$\lambda(\mathbf{BC}) = (\lambda\mathbf{B})\mathbf{C} = \mathbf{B}(\lambda\mathbf{C}) = (\mathbf{BC})\lambda, \quad \mathbf{B} \in \mathbb{R}^{m \times n}, \mathbf{C} \in \mathbb{R}^{n \times k}.$$

Note that this allows us to move scalar values around.

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

Example 2.5 (Distributivity)

If we define

$$\mathbf{C} := \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad (2.32)$$

then for any $\lambda, \psi \in \mathbb{R}$ we obtain

$$(\lambda + \psi)\mathbf{C} = \begin{bmatrix} (\lambda + \psi)1 & (\lambda + \psi)2 \\ (\lambda + \psi)3 & (\lambda + \psi)4 \end{bmatrix} = \begin{bmatrix} \lambda + \psi & 2\lambda + 2\psi \\ 3\lambda + 3\psi & 4\lambda + 4\psi \end{bmatrix} \quad (2.33a)$$

$$= \begin{bmatrix} \lambda & 2\lambda \\ 3\lambda & 4\lambda \end{bmatrix} + \begin{bmatrix} \psi & 2\psi \\ 3\psi & 4\psi \end{bmatrix} = \lambda\mathbf{C} + \psi\mathbf{C} \quad (2.33b)$$

910 **2.2.4 Compact Representations of Systems of Linear Equations**

If we consider the system of linear equations

$$\begin{aligned} 2x_1 + 3x_2 + 5x_3 &= 1 \\ 4x_1 - 2x_2 - 7x_3 &= 8 \\ 9x_1 + 5x_2 - 3x_3 &= 2 \end{aligned} \quad (2.34)$$

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

$$\begin{bmatrix} 2 & 3 & 5 \\ 4 & -2 & -7 \\ 9 & 5 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 8 \\ 2 \end{bmatrix}. \quad (2.35)$$

911 Note that x_1 scales the first column, x_2 the second one, and x_3 the third
912 one.

913 Generally, system of linear equations can be compactly represented in
914 their matrix form as $\mathbf{Ax} = \mathbf{b}$, see (2.3), and the product \mathbf{Ax} is a (linear)
915 combination of the columns of \mathbf{A} . We will discuss linear combinations in
916 more detail in Section 2.5.

917 **2.3 Solving Systems of Linear Equations**

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

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m, \end{aligned} \quad (2.36)$$

918 where $a_{ij} \in \mathbb{R}$ and $b_i \in \mathbb{R}$ are known constants and x_j are unknowns,
919 $i = 1, \dots, m$, $j = 1, \dots, n$. Thus far, we saw that matrices can be used as
920 a compact way of formulating systems of linear equations so that we can
921 write $\mathbf{Ax} = \mathbf{b}$, see (2.9). Moreover, we defined basic matrix operations,
922 such as addition and multiplication of matrices. In the following, we will
923 focus on solving systems of linear equations.

924 **2.3.1 Particular and General Solution**

Before discussing how to solve systems of linear equations systematically,
let us have a look at an example. Consider the system of equations

$$\begin{bmatrix} 1 & 0 & 8 & -4 \\ 0 & 1 & 2 & 12 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 42 \\ 8 \end{bmatrix}. \quad (2.37)$$

This system of equations is in a particularly easy form, where the first two columns consist of a 1 and a 0. Remember that we want to find scalars x_1, \dots, x_4 , such that $\sum_{i=1}^4 x_i c_i = b$, where we define c_i to be the i th column of the matrix and b the right-hand-side of (2.37). A solution to the problem in (2.37) can be found immediately by taking 42 times the first column and 8 times the second column so that

$$b = \begin{bmatrix} 42 \\ 8 \end{bmatrix} = 42 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 8 \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \quad (2.38)$$

Therefore, a 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 system of linear equations. To capture all the other solutions, we need to be creative of generating $\mathbf{0}$ in a non-trivial way using the columns of the matrix: Adding $\mathbf{0}$ to our special solution does not change the special solution. To do so, we express the third column using the first two columns (which are of this very simple form)

$$\begin{bmatrix} 8 \\ 2 \end{bmatrix} = 8 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (2.39)$$

so that $\mathbf{0} = 8c_1 + 2c_2 - 1c_3 + 0c_4$ and $(x_1, x_2, x_3, x_4) = (8, 2, -1, 0)$. In fact, any scaling of this solution by $\lambda_1 \in \mathbb{R}$ produces the $\mathbf{0}$ vector, i.e.,

$$\begin{bmatrix} 1 & 0 & 8 & -4 \\ 0 & 1 & 2 & 12 \end{bmatrix} \left(\lambda_1 \begin{bmatrix} 8 \\ 2 \\ -1 \\ 0 \end{bmatrix} \right) = \lambda_1(8c_1 + 2c_2 - c_3) = \mathbf{0}. \quad (2.40)$$

Following the same line of reasoning, we express the fourth column of the matrix in (2.37) using the first two columns and generate another set of non-trivial versions of $\mathbf{0}$ as

$$\begin{bmatrix} 1 & 0 & 8 & -4 \\ 0 & 1 & 2 & 12 \end{bmatrix} \left(\lambda_2 \begin{bmatrix} -4 \\ 12 \\ 0 \\ -1 \end{bmatrix} \right) = \lambda_2(-4c_1 + 12c_2 - c_4) = \mathbf{0} \quad (2.41)$$

for any $\lambda_2 \in \mathbb{R}$. Putting everything together, we obtain all solutions of the equation system in (2.37), which is called the *general solution*, as the set

$$\left\{ \mathbf{x} \in \mathbb{R}^4 : \mathbf{x} = \begin{bmatrix} 42 \\ 8 \\ 0 \\ 0 \end{bmatrix} + \lambda_1 \begin{bmatrix} 8 \\ 2 \\ -1 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} -4 \\ 12 \\ 0 \\ -1 \end{bmatrix}, \lambda_1, \lambda_2 \in \mathbb{R} \right\}. \quad (2.42)$$

⁹²⁵ Remark. The general approach we followed consisted of the following three steps:

- ⁹²⁷ 1. Find a particular solution to $A\mathbf{x} = b$
- ⁹²⁸ 2. Find all solutions to $A\mathbf{x} = \mathbf{0}$

general solution

929 3. Combine the solutions from 1. and 2. to the general solution.

930 Neither the general nor the particular solution is unique. \diamond

931 The system of linear equations in the example above was easy to solve
 932 because the matrix in (2.37) has this particularly convenient form, which
 933 allowed us to find the particular and the general solution by inspection.
 934 However, general equation systems are not of this simple form. Fortunately,
 935 there exists a constructive algorithmic way of transforming any
 936 system of linear equations into this particularly simple form: Gaussian
 937 elimination. Key to Gaussian elimination are elementary transformations
 938 of systems of linear equations, which transform the equation system into
 939 a simple form. Then, we can apply the three steps to the simple form that
 940 we just discussed in the context of the example in (2.37), see the remark
 941 above.

942 2.3.2 Elementary Transformations

elementary 943 Key to solving a system of linear equations are *elementary transformations*
 transformations 944 that keep the solution set the same, but that transform the equation system
 945 into a simpler form:

- 946 • Exchange of two equations (or: rows in the matrix representing the
 947 equation system)
- 948 • Multiplication of an equation (row) with a constant $\lambda \in \mathbb{R} \setminus \{0\}$
- 949 • Addition an equation (row) to another equation (row)

Example 2.6

For $a \in \mathbb{R}$, we seek all solutions of the following system of equations:

$$\begin{array}{ccccccccc} -2x_1 & + & 4x_2 & - & 2x_3 & - & x_4 & + & 4x_5 = -3 \\ 4x_1 & - & 8x_2 & + & 3x_3 & - & 3x_4 & + & x_5 = 2 \\ x_1 & - & 2x_2 & + & x_3 & - & x_4 & + & x_5 = 0 \\ x_1 & - & 2x_2 & & & - & 3x_4 & + & 4x_5 = a \end{array} \quad (2.43)$$

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

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

where we used the vertical line to separate the left-hand-side from the right-hand-side in (2.43). We use \rightsquigarrow to indicate a transformation of the left-hand-side into the right-hand-side using elementary transformations.

The augmented matrix $[\mathbf{A} | \mathbf{b}]$ compactly represents the system of linear equations $\mathbf{A}\mathbf{x} = \mathbf{b}$.

Swapping rows 1 and 3 leads to

$$\left[\begin{array}{ccccc|c} 1 & -2 & 1 & -1 & 1 & 0 \\ 4 & -8 & 3 & -3 & 1 & 2 \\ -2 & 4 & -2 & -1 & 4 & -3 \\ 1 & -2 & 0 & -3 & 4 & a \end{array} \right] \begin{matrix} \\ -4R_1 \\ +2R_1 \\ -R_1 \end{matrix}$$

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

$$\begin{aligned} & \left[\begin{array}{ccccc|c} 1 & -2 & 1 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 & -3 & 2 \\ 0 & 0 & 0 & -3 & 6 & -3 \\ 0 & 0 & -1 & -2 & 3 & a \end{array} \right] \begin{matrix} \\ \\ \\ -R_2 - R_3 \end{matrix} \\ \rightsquigarrow & \left[\begin{array}{ccccc|c} 1 & -2 & 1 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 & -3 & 2 \\ 0 & 0 & 0 & -3 & 6 & -3 \\ 0 & 0 & 0 & 0 & 0 & a+1 \end{array} \right] \begin{matrix} \\ \\ \cdot(-1) \\ \cdot(-\frac{1}{3}) \end{matrix} \\ \rightsquigarrow & \left[\begin{array}{ccccc|c} 1 & -2 & 1 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 3 & -2 \\ 0 & 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 & a+1 \end{array} \right] \end{aligned}$$

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

$$\begin{aligned} x_1 - 2x_2 + x_3 - x_4 + x_5 &= 0 \\ x_3 - x_4 + 3x_5 &= -2 \\ x_4 - 2x_5 &= 1 \\ 0 &= a+1 \end{aligned} \quad . \quad (2.44)$$

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

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} . \quad (2.45)$$

The *general solution*, which captures the set of all possible solutions, is

$$\left\{ \mathbf{x} \in \mathbb{R}^5 : \mathbf{x} = \begin{bmatrix} 2 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} + \lambda_1 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 2 \\ 0 \\ -1 \\ 2 \\ 1 \end{bmatrix}, \quad \lambda_1, \lambda_2 \in \mathbb{R} \right\} . \quad (2.46)$$

row-echelon form
(REF)

particular solution

general solution

950 In the following, we will detail a constructive way to obtain a particular
 951 and general solution of a system of linear equations.

952 *Remark* (Pivots and Staircase Structure). The leading coefficient of a row
 pivot 953 (first non-zero number from the left) is called the *pivot* and is always
 954 strictly to the right of the pivot of the row above it. Therefore, any equa-
 955 tion system in row echelon form always has a “staircase” structure. ◇

row echelon form 956 **Definition 2.5** (Row Echelon Form). A matrix is in *row echelon form* (REF)
 957 if

- 958 • All rows that contain only zeros are at the bottom of the matrix; corre-
 959 spondingly, all rows that contain at least one non-zero element are on
 960 top of rows that contain only zeros.
- 961 • Looking at non-zero rows only, the first non-zero number from the left
 pivot 962 (also called the *pivot* or the *leading coefficient*) is always strictly to the
 963 leading coefficient 964 right of the pivot of the row above it.

In other books, it is
 sometimes required 964 that the pivot is 1. 965
 basic variables 966 free variables 967 *Remark* (Basic and Free Variables). The variables corresponding to the
 pivots in the row-echelon form are called *basic variables*, the other vari-
 ables are *free variables*. For example, in (2.44), x_1, x_3, x_4 are basic vari-
 ables, whereas x_2, x_5 are free variables. ◇

968 *Remark* (Obtaining a Particular Solution). The row echelon form makes
 969 our lives easier when we need to determine a particular solution. To do
 970 this, we express the right-hand side of the equation system using the pivot
 971 columns, such that $\mathbf{b} = \sum_{i=1}^P \lambda_i \mathbf{p}_i$, where \mathbf{p}_i , $i = 1, \dots, P$, are the pivot
 972 columns. The λ_i are determined easiest if we start with the most-right
 973 pivot column and work our way to the left.

In the above example, we would try to find $\lambda_1, \lambda_2, \lambda_3$ such that

$$\lambda_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \lambda_3 \begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \end{bmatrix}. \quad (2.47)$$

974 From here, we find relatively directly that $\lambda_3 = 1, \lambda_2 = -1, \lambda_1 = 2$. When
 975 we put everything together, we must not forget the non-pivot columns
 976 for which we set the coefficients implicitly to 0. Therefore, we get the
 977 particular solution $\mathbf{x} = [2, 0, -1, 1, 0]^\top$. ◇

reduced row 978 *Remark* (Reduced Row Echelon Form). An equation system is in *reduced
 979 row echelon form* (also: *row-reduced echelon form* or *row canonical form*) if

- 980 • It is in row echelon form.
- 981 • Every pivot is 1.
- 982 • The pivot is the only non-zero entry in its column.

◇

984 The reduced row echelon form will play an important role later in Sec-
 985 tion 2.3.3 because it allows us to determine the general solution of a sys-
 986 tem of linear equations in a straightforward way.

987 *Remark* (Gaussian Elimination). *Gaussian elimination* is an algorithm that
 988 performs elementary transformations to bring a system of linear equations
 989 into reduced row echelon form. ◇

Gaussian
elimination

Example 2.7 (Reduced Row Echelon Form)

Verify that the following matrix is in reduced row echelon form (the pivots are in **bold**):

$$\mathbf{A} = \begin{bmatrix} \mathbf{1} & 3 & 0 & 0 & 3 \\ 0 & 0 & \mathbf{1} & 0 & 9 \\ 0 & 0 & 0 & \mathbf{1} & -4 \end{bmatrix} \quad (2.48)$$

The key idea for finding the solutions of $\mathbf{A}\mathbf{x} = \mathbf{0}$ is to look at the *non-pivot columns*, which we will need to express as a (linear) combination of the pivot columns. The reduced row echelon form makes this relatively straightforward, and we express the non-pivot columns in terms of sums and multiples of the pivot columns that are on their left: The second column is 3 times the first column (we can ignore the pivot columns on the right of the second column). Therefore, to obtain $\mathbf{0}$, we need to subtract the second column from three times the first column. Now, we look at the fifth column, which is our second non-pivot column. The fifth column can be expressed as 3 times the first pivot column, 9 times the second pivot column, and -4 times the third pivot column. We need to keep track of the indices of the pivot columns and translate this into 3 times the first column, 0 times the second column (which is a non-pivot column), 9 times the third pivot column (which is our second pivot column), and -4 times the fourth column (which is the third pivot column). Then we need to subtract the fifth column to obtain $\mathbf{0}$. In the end, we are still solving a homogeneous equation system.

To summarize, all solutions of $\mathbf{A}\mathbf{x} = \mathbf{0}, \mathbf{x} \in \mathbb{R}^5$ are given by

$$\left\{ \mathbf{x} \in \mathbb{R}^5 : \mathbf{x} = \lambda_1 \begin{bmatrix} 3 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 3 \\ 0 \\ 9 \\ -4 \\ -1 \end{bmatrix}, \quad \lambda_1, \lambda_2 \in \mathbb{R} \right\}. \quad (2.49)$$

990 **2.3.3 The Minus-1 Trick**

991 In the following, we introduce a practical trick for reading out the solutions
 992 \mathbf{x} of a homogeneous system of linear equations $\mathbf{A}\mathbf{x} = \mathbf{0}$, where
 993 $\mathbf{A} \in \mathbb{R}^{k \times n}$, $\mathbf{x} \in \mathbb{R}^n$.

To start, we assume that \mathbf{A} is in reduced row echelon form without any rows that just contain zeros, i.e.,

$$\mathbf{A} = \begin{bmatrix} 0 & \cdots & 0 & \mathbf{1} & * & \cdots & * & 0 & * & \cdots & * & 0 & * & \cdots & * \\ \vdots & & \vdots & 0 & 0 & \cdots & 0 & \mathbf{1} & * & \cdots & * & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & 0 & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & 0 & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \mathbf{1} & * & \cdots & * \end{bmatrix}, \quad (2.50)$$

where $*$ can be an arbitrary real number, with the constraints that the first non-zero entry per row must be 1 and all other entries in the corresponding column must be 0. The columns j_1, \dots, j_k with the pivots (marked in **bold**) are the standard unit vectors $e_1, \dots, e_k \in \mathbb{R}^k$. We extend this matrix to an $n \times n$ -matrix $\tilde{\mathbf{A}}$ by adding $n - k$ rows of the form

$$[0 \quad \cdots \quad 0 \quad -1 \quad 0 \quad \cdots \quad 0] \quad (2.51)$$

994 so that the diagonal of the augmented matrix $\tilde{\mathbf{A}}$ contains either 1 or -1 .
 995 Then, the columns of $\tilde{\mathbf{A}}$, which contain the -1 as pivots are solutions of
 996 the homogeneous equation system $\mathbf{A}\mathbf{x} = \mathbf{0}$. To be more precise, these
 997 columns form a basis (Section 2.6.1) of the solution space of $\mathbf{A}\mathbf{x} = \mathbf{0}$,
 998 which we will later call the *kernel* or *null space* (see Section 2.7.3).

kernel
null space

Example 2.8 (Minus-1 Trick)

Let us revisit the matrix in (2.48), which is already in REF:

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 & 9 \\ 0 & 0 & 0 & 1 & -4 \end{bmatrix}. \quad (2.52)$$

We now augment this matrix to a 5×5 matrix by adding rows of the form (2.51) at the places where the pivots on the diagonal are missing and obtain

$$\tilde{\mathbf{A}} = \begin{bmatrix} 1 & 3 & 0 & 0 & 3 \\ \color{blue}{0} & \color{blue}{-1} & \color{blue}{0} & \color{blue}{0} & \color{blue}{0} \\ 0 & 0 & 1 & 0 & 9 \\ 0 & 0 & 0 & 1 & -4 \\ \color{blue}{0} & \color{blue}{0} & \color{blue}{0} & \color{blue}{0} & \color{blue}{-1} \end{bmatrix} \quad (2.53)$$

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

$$\left\{ \mathbf{x} \in \mathbb{R}^5 : \mathbf{x} = \lambda_1 \begin{bmatrix} 3 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 3 \\ 0 \\ 9 \\ -4 \\ -1 \end{bmatrix}, \quad \lambda_1, \lambda_2 \in \mathbb{R} \right\}, \quad (2.54)$$

which is identical to the solution in (2.49) that we obtained by “insight”.

999

Calculating the Inverse

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

$$[\mathbf{A} | \mathbf{I}_n] \rightsquigarrow \dots \rightsquigarrow [\mathbf{I}_n | \mathbf{A}^{-1}]. \quad (2.55)$$

1000 This means that if we bring the augmented equation system into reduced
 1001 row echelon form, we can read out the inverse on the right-hand side of
 1002 the equation system. Hence, determining the inverse of a matrix is equiv-
 1003 alent to solving systems of linear equations.

Example 2.9 (Calculating an Inverse Matrix by Gaussian Elimination)

To determine the inverse of

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad (2.56)$$

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 use Gaussian elimination to bring 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:

$$\mathbf{A}^{-1} = \begin{bmatrix} -1 & 2 & -2 & 2 \\ 1 & -1 & 2 & -2 \\ 1 & -1 & 1 & -1 \\ -1 & 0 & -1 & 2 \end{bmatrix}. \quad (2.57)$$

1004 2.3.4 Algorithms for Solving a System of Linear Equations

1005 In the following, we briefly discuss approaches to solving a system of linear
1006 equations of the form $\mathbf{A}\mathbf{x} = \mathbf{b}$.

In special cases, we may be able to determine the inverse \mathbf{A}^{-1} , such that the solution of $\mathbf{A}\mathbf{x} = \mathbf{b}$ is given as $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$. However, this is only possible if \mathbf{A} is a square matrix and invertible, which is often not the case. Otherwise, under mild assumptions (i.e., \mathbf{A} needs to have linearly independent columns) we can use the transformation

$$\mathbf{A}\mathbf{x} = \mathbf{b} \iff \mathbf{A}^\top \mathbf{A}\mathbf{x} = \mathbf{A}^\top \mathbf{b} \iff \mathbf{x} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b} \quad (2.58)$$

Moore-Penrose
1007 pseudo-inverse 1007 and use the *Moore-Penrose pseudo-inverse* $(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top$ to determine the
1008 solution (2.58) that solves $\mathbf{A}\mathbf{x} = \mathbf{b}$, which also corresponds to the minimum
1009 norm least-squares solution. A disadvantage of this approach is that
1010 it requires many computations for the matrix-matrix product and computing
1011 the inverse of $\mathbf{A}^\top \mathbf{A}$. Moreover, for reasons of numerical precision it
1012 is generally not recommended to compute the inverse or pseudo-inverse.
1013 In the following, we therefore briefly discuss alternative approaches to
1014 solving systems of linear equations.

1015 Gaussian elimination plays an important role when computing determinants (Section 4.1), checking whether a set of vectors is linearly independent (Section 2.5), computing the inverse of a matrix (Section 2.2.2),
1016 computing the rank of a matrix (Section 2.6.2) and a basis of a vector space (Section 2.6.1). We will discuss all these topics later on. Gaussian
1017 elimination is an intuitive and constructive way to solve a system of linear
1018 equations with thousands of variables. However, for systems with millions
1019 of variables, it is impractical as the required number of arithmetic operations
1020 scales cubically in the number of simultaneous equations.

1021 In practice, systems of many linear equations are solved indirectly, by either stationary iterative methods, such as the Richardson method, the
1022 Jacobi method, the Gauß-Seidel method, or the successive over-relaxation method,
1023 or Krylov subspace methods, such as conjugate gradients, generalized minimal residual, or biconjugate gradients.

Let \mathbf{x}_* be a solution of $\mathbf{A}\mathbf{x} = \mathbf{b}$. The key idea of these iterative methods

is to set up an iteration of the form

$$\mathbf{x}^{(k+1)} = \mathbf{A}\mathbf{x}^{(k)} \quad (2.59)$$

that reduces the residual error $\|\mathbf{x}^{(k+1)} - \mathbf{x}_*\|$ in every iteration and finally converges to \mathbf{x}_* . We will introduce norms $\|\cdot\|$, which allow us to compute similarities between vectors, in Section 3.1.

2.4 Vector Spaces

Thus far, we have looked at systems of linear equations and how to solve them. We saw that systems of linear equations can be compactly represented using matrix-vector notations. In the following, we will have a closer look at vector spaces, i.e., a structured space in which vectors live.

In the beginning of this chapter, we informally characterized vectors as objects that can be added together and multiplied by a scalar, and they remain objects of the same type (see page 17). Now, we are ready to formalize this, and we will start by introducing the concept of a group, which is a set of elements and an operation defined on these elements that keeps some structure of the set intact.

2.4.1 Groups

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

Definition 2.6 (Group). Consider a set \mathcal{G} and an operation $\otimes : \mathcal{G} \times \mathcal{G} \rightarrow \mathcal{G}$ defined on \mathcal{G} .

Then $G := (\mathcal{G}, \otimes)$ is called a *group* if the following hold:

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

group
Closure
Associativity:
Neutral element:
Inverse element:

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

Example 2.10 (Groups)

Let us have a look at some examples of sets with associated operations and see whether they are groups.

- $(\mathbb{Z}, +)$ is a group.

$\mathbb{N}_0 := \mathbb{N} \cup \{0\}$

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

$$(x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1 + y_1, \dots, x_n + y_n). \quad (2.60)$$

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

- $(\mathbb{R}^{m \times n}, +)$, the set of $m \times n$ -matrices is Abelian (with componentwise addition as defined in (2.60)).
- Let us have a closer look at $(\mathbb{R}^{n \times n}, \cdot)$, i.e., the set of $n \times n$ -matrices with matrix multiplication as defined in (2.12).
 - Closure and associativity follow directly from the definition of matrix multiplication.
 - Neutral element: The identity matrix I_n is the neutral element with respect to matrix multiplication “.” in $(\mathbb{R}^{n \times n}, \cdot)$.
 - Inverse element: If the inverse exists then A^{-1} is the inverse element of $A \in \mathbb{R}^{n \times n}$.

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
 $I_m A = A$.

1057
1058

Remark. The inverse element is defined with respect to the operation \otimes and does not necessarily mean $\frac{1}{x}$. \diamond

1059
1060
general linear group

1061
1062

Definition 2.7 (General Linear Group). The set of regular (invertible) matrices $A \in \mathbb{R}^{n \times n}$ is a group with respect to matrix multiplication as defined in (2.12) and is called *general linear group* $GL(n, \mathbb{R})$. However, since matrix multiplication is not commutative, the group is not Abelian.

1063

2.4.2 Vector Spaces

1064
1065
1066
1067
1068

When we discussed groups, we looked at sets \mathcal{G} and inner operations on \mathcal{G} , i.e., mappings $\mathcal{G} \times \mathcal{G} \rightarrow \mathcal{G}$ that only operate on elements in \mathcal{G} . In the following, we will consider sets that in addition to an inner operation + also contain an outer operation \cdot , the multiplication of a vector $x \in \mathcal{V}$ by a scalar $\lambda \in \mathbb{R}$.

vector space

Definition 2.8 (Vector space). A real-valued *vector space* $V = (\mathcal{V}, +, \cdot)$ is a set \mathcal{V} with two operations

$$+ : \mathcal{V} \times \mathcal{V} \rightarrow \mathcal{V} \quad (2.61)$$

$$\cdot : \mathbb{R} \times \mathcal{V} \rightarrow \mathcal{V} \quad (2.62)$$

1069 where

- 1070 1. $(\mathcal{V}, +)$ is an Abelian group
 1071 2. Distributivity:
 1072 1. $\forall \lambda \in \mathbb{R}, \mathbf{x}, \mathbf{y} \in \mathcal{V} : \lambda \cdot (\mathbf{x} + \mathbf{y}) = \lambda \cdot \mathbf{x} + \lambda \cdot \mathbf{y}$
 1073 2. $\forall \lambda, \psi \in \mathbb{R}, \mathbf{x} \in \mathcal{V} : (\lambda + \psi) \cdot \mathbf{x} = \lambda \cdot \mathbf{x} + \psi \cdot \mathbf{x}$
 1074 3. Associativity (outer operation): $\forall \lambda, \psi \in \mathbb{R}, \mathbf{x} \in \mathcal{V} : \lambda \cdot (\psi \cdot \mathbf{x}) = (\lambda \psi) \cdot \mathbf{x}$
 1075 4. Neutral element with respect to the outer operation: $\forall \mathbf{x} \in \mathcal{V} : 1 \cdot \mathbf{x} = \mathbf{x}$

1076 The elements $\mathbf{x} \in V$ are called *vectors*. The neutral element of $(\mathcal{V}, +)$ is
 1077 the zero vector $\mathbf{0} = [0, \dots, 0]^\top$, and the inner operation $+$ is called *vector*
 1078 *addition*. The elements $\lambda \in \mathbb{R}$ are called *scalars* and the outer operation
 1079 \cdot is a *multiplication by scalars*. Note that a scalar product is something
 1080 different, and we will get to this in Section 3.2.

vectors

vector addition

scalars

multiplication by
scalars

1081 *Remark.* A “vector multiplication” \mathbf{ab} , $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, is not defined. Theoretically, we could define an element-wise multiplication, such that $\mathbf{c} = \mathbf{ab}$ with $c_j = a_j b_j$. This “array multiplication” is common to many programming languages but makes mathematically limited sense using the standard rules for matrix multiplication: By treating vectors as $n \times 1$ matrices (which we usually do), we can use the matrix multiplication as defined in (2.12). However, then the dimensions of the vectors do not match. Only the following multiplications for vectors are defined: $\mathbf{ab}^\top \in \mathbb{R}^{n \times n}$ (outer product), $\mathbf{a}^\top \mathbf{b} \in \mathbb{R}$ (inner/scalar/dot product). ◇

Example 2.11 (Vector Spaces)

Let us have a look at some important examples.

- $\mathcal{V} = \mathbb{R}^n, n \in \mathbb{N}$ is a vector space with operations defined as follows:
 - Addition: $\mathbf{x} + \mathbf{y} = (x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1 + y_1, \dots, x_n + y_n)$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$
 - Multiplication by scalars: $\lambda \mathbf{x} = \lambda(x_1, \dots, x_n) = (\lambda x_1, \dots, \lambda x_n)$ for all $\lambda \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^n$
- $\mathcal{V} = \mathbb{R}^{m \times n}, m, n \in \mathbb{N}$ is a vector space with
 - Addition: $\mathbf{A} + \mathbf{B} = \begin{bmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{bmatrix}$ is defined elementwise for all $\mathbf{A}, \mathbf{B} \in \mathcal{V}$
 - Multiplication by scalars: $\lambda \mathbf{A} = \begin{bmatrix} \lambda a_{11} & \cdots & \lambda a_{1n} \\ \vdots & & \vdots \\ \lambda a_{m1} & \cdots & \lambda a_{mn} \end{bmatrix}$ as defined in Section 2.2. Remember that $\mathbb{R}^{m \times n}$ is equivalent to \mathbb{R}^{mn} .
- $\mathcal{V} = \mathbb{C}$, with the standard definition of addition of complex numbers.

1090 1091 1092 1093 *Remark.* In the following, we will denote a vector space $(\mathcal{V}, +, \cdot)$ by V when $+$ and \cdot are the standard vector addition and scalar multiplication. Moreover, we will use the notation $\mathbf{x} \in V$ for vectors in \mathcal{V} to simplify notation. \diamond

column vectors 1094 *Remark.* The vector spaces $\mathbb{R}^n, \mathbb{R}^{n \times 1}, \mathbb{R}^{1 \times n}$ are only different in the way we write vectors. 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*

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}. \quad (2.63)$$

row vectors 1094 1095 1096 1097 This simplifies the notation regarding vector space operations. However, we do 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 \mathbf{x} to denote a column vector, and a row vector is denoted by \mathbf{x}^\top , the *transpose* of \mathbf{x} . \diamond

1098 2.4.3 Vector Subspaces

1099 1100 1101 1102 In the following, we will introduce vector subspaces. Intuitively, they are sets contained in the original vector space with the property that when we perform vector space operations on elements within this subspace, we will never leave it. In this sense, they are “closed”.

vector subspace 1103 1104 1105 1106 1107 **Definition 2.9** (Vector Subspace). Let $V = (\mathcal{V}, +, \cdot)$ be a vector space and $\mathcal{U} \subseteq \mathcal{V}, \mathcal{U} \neq \emptyset$. Then $U = (\mathcal{U}, +, \cdot)$ is called *vector subspace* of V (or *linear subspace*) if U is a vector space with the vector space operations $+$ and \cdot restricted to $\mathcal{U} \times \mathcal{U}$ and $\mathbb{R} \times \mathcal{U}$. We write $U \subseteq V$ to denote a subspace U of V .

1108 1109 1110 1111 1112 If $\mathcal{U} \subseteq \mathcal{V}$ and V is a vector space, then U naturally inherits many properties directly from V because they are true for all $\mathbf{x} \in \mathcal{V}$, and in particular for all $\mathbf{x} \in \mathcal{U} \subseteq \mathcal{V}$. This includes the Abelian group properties, the distributivity, the associativity and the neutral element. To determine whether $(\mathcal{U}, +, \cdot)$ is a subspace of V we still do need to show

- 1113 1. $\mathcal{U} \neq \emptyset$, in particular: $\mathbf{0} \in \mathcal{U}$
- 1114 2. Closure of U :

- 1115 1. With respect to the outer operation: $\forall \lambda \in \mathbb{R} \forall \mathbf{x} \in \mathcal{U} : \lambda \mathbf{x} \in \mathcal{U}$.
- 1116 2. With respect to the inner operation: $\forall \mathbf{x}, \mathbf{y} \in \mathcal{U} : \mathbf{x} + \mathbf{y} \in \mathcal{U}$.

Example 2.12 (Vector Subspaces)

Let us have a look at some subspaces.

- For every vector space V the trivial subspaces are V itself and $\{\mathbf{0}\}$.

- Only example D in Figure 2.5 is a subspace of \mathbb{R}^2 (with the usual inner/outer operations). In A and C, the closure property is violated; B does not contain $\mathbf{0}$.
- The solution set of a homogeneous linear equation system $A\mathbf{x} = \mathbf{0}$ with n unknowns $\mathbf{x} = [x_1, \dots, x_n]^\top$ is a subspace of \mathbb{R}^n .
- The solution of an inhomogeneous equation system $A\mathbf{x} = \mathbf{b}$, $\mathbf{b} \neq \mathbf{0}$ is not a subspace of \mathbb{R}^n .
- The intersection of arbitrarily many subspaces is a subspace itself.

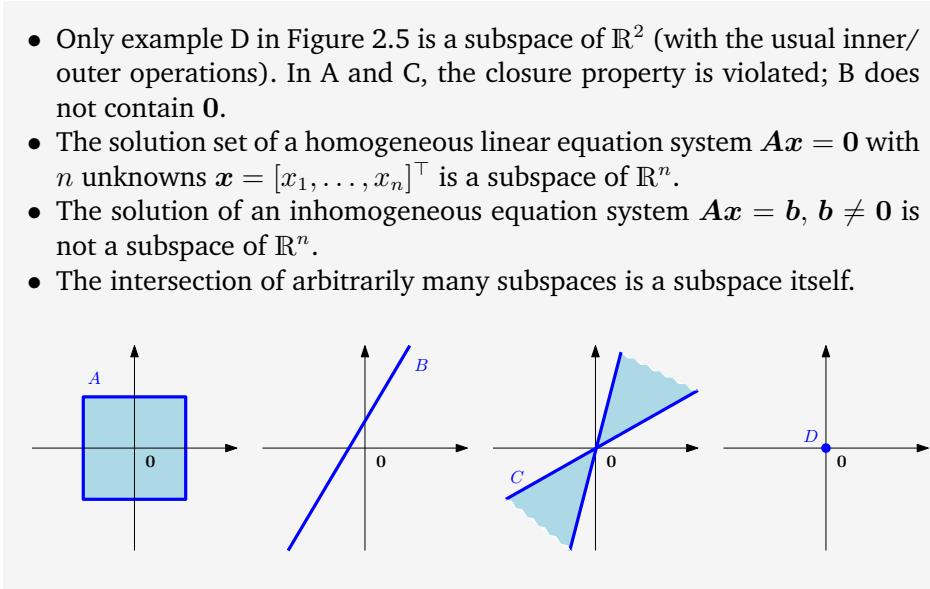


Figure 2.5 Not all subsets of \mathbb{R}^2 are subspaces. In A and C, the closure property is violated; B does not contain $\mathbf{0}$. Only D is a subspace.

1117 *Remark.* Every subspace $U \subseteq (\mathbb{R}^n, +, \cdot)$ is the solution space of a homogeneous linear equation system $A\mathbf{x} = \mathbf{0}$. 1118 ◇

1119 2.5 Linear Independence

1120 So far, we looked at vector spaces and some of their properties, e.g., closure. Now, we will look at what we can do with vectors (elements of 1121 the vector space). In particular, we can add vectors together and multiply them with scalars. The closure property guarantees that we end up 1122 with another vector in the same vector space. Let us formalize this: 1123

Definition 2.10 (Linear Combination). Consider a vector space V and a finite number of vectors $\mathbf{x}_1, \dots, \mathbf{x}_k \in V$. Then, every $\mathbf{v} \in V$ of the form

$$\mathbf{v} = \lambda_1 \mathbf{x}_1 + \dots + \lambda_k \mathbf{x}_k = \sum_{i=1}^k \lambda_i \mathbf{x}_i \in V \quad (2.64)$$

1125 with $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ is a *linear combination* of the vectors $\mathbf{x}_1, \dots, \mathbf{x}_k$.

linear combination

1126 The $\mathbf{0}$ -vector can always be written as the linear combination of k vectors 1127 $\mathbf{x}_1, \dots, \mathbf{x}_k$ because $\mathbf{0} = \sum_{i=1}^k 0 \mathbf{x}_i$ is always true. In the following, 1128 we are interested in non-trivial linear combinations of a set of vectors to 1129 represent $\mathbf{0}$, i.e., linear combinations of vectors $\mathbf{x}_1, \dots, \mathbf{x}_k$ where not all 1130 coefficients λ_i in (2.64) are 0.

1131 **Definition 2.11** (Linear (In)dependence). Let us consider a vector space 1132 V with $k \in \mathbb{N}$ and $\mathbf{x}_1, \dots, \mathbf{x}_k \in V$. If there is a non-trivial linear combination, such that $\mathbf{0} = \sum_{i=1}^k \lambda_i \mathbf{x}_i$ with at least one $\lambda_i \neq 0$, the vectors

linearly independent

1134 $\mathbf{x}_1, \dots, \mathbf{x}_k$ are *linearly dependent*. If only the trivial solution exists, i.e.,
1135 $\lambda_1 = \dots = \lambda_k = 0$ the vectors $\mathbf{x}_1, \dots, \mathbf{x}_k$ are *linearly independent*.

linearly dependent

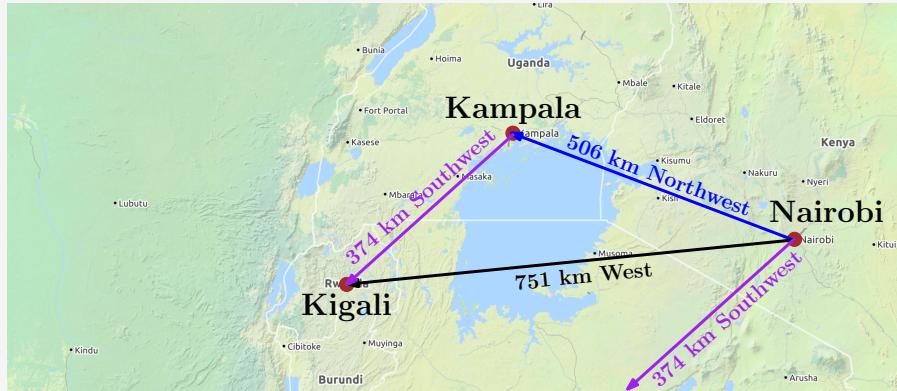
1136 1137 1138 1139 1140 Linear independence is one of the most important concepts in linear algebra. Intuitively, a set of linearly independent vectors are vectors that have no redundancy, i.e., if we remove any of those vectors from the set, we will lose something. Throughout the next sections, we will formalize this intuition more.

In this example, we make crude approximations to cardinal directions.

Example 2.13 (Linearly Dependent Vectors)

A geographic example may help to clarify the concept of linear independence. A person in Nairobi (Kenya) describing where Kigali (Rwanda) is might say “You can get to Kigali by first going 506 km Northwest to Kampala (Uganda) and then 374 km Southwest.”. This is sufficient information to describe the location of Kigali because the geographic coordinate system may be considered a two-dimensional vector space (ignoring altitude and the Earth’s surface). The person may add “It is about 751 km West of here.” Although this last statement is true, it is not necessary to find Kigali given the previous information (see Figure 2.6 for an illustration).

Figure 2.6
Geographic example
(with crude
approximations to
cardinal directions)
of linearly
dependent vectors
in a
two-dimensional
space (plane).



In this example, the “506 km Northwest” vector (blue) and the “374 km Southwest” vector (purple) are linearly independent. This means the Southwest vector cannot be described in terms of the Northwest vector, and vice versa. However, the third “751 km West” vector (black) is a linear combination of the other two vectors, and it makes the set of vectors linearly dependent.

1141 1142 *Remark.* The following properties are useful to find out whether vectors are linearly independent.

1143 1144 • k vectors are either linearly dependent or linearly independent. There is no third option.

- If at least one of the vectors $\mathbf{x}_1, \dots, \mathbf{x}_k$ is $\mathbf{0}$ then they are linearly dependent. The same holds if two vectors are identical.
- The vectors $\{\mathbf{x}_1, \dots, \mathbf{x}_k : \mathbf{x}_i \neq \mathbf{0}, i = 1, \dots, k\}$, $k \geq 2$, are linearly dependent if and only if (at least) one of them is a linear combination of the others. In particular, if one vector is a multiple of another vector, i.e., $\mathbf{x}_i = \lambda \mathbf{x}_j$, $\lambda \in \mathbb{R}$ then the set $\{\mathbf{x}_1, \dots, \mathbf{x}_k : \mathbf{x}_i \neq \mathbf{0}, i = 1, \dots, k\}$ is linearly dependent.
- A practical way of checking whether vectors $\mathbf{x}_1, \dots, \mathbf{x}_k \in V$ are linearly independent is to use Gaussian elimination: Write all vectors as columns of a matrix \mathbf{A} and perform Gaussian elimination until the matrix is in row echelon form (the reduced row echelon form is not necessary here).
 - The pivot columns indicate the vectors, which are linearly independent of the vectors on the left. 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, the row echelon form

$$\begin{bmatrix} 1 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix} \quad (2.65)$$

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

All column vectors are linearly independent if and only if all columns are pivot columns. If there is at least one non-pivot column, the columns (and, therefore, the corresponding vectors) are linearly dependent.

◇

Example 2.14

Consider \mathbb{R}^4 with

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ -3 \\ 4 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 2 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} -1 \\ -2 \\ 1 \\ 1 \end{bmatrix}. \quad (2.66)$$

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

$$\lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2 + \lambda_3 \mathbf{x}_3 = \lambda_1 \begin{bmatrix} 1 \\ 2 \\ -3 \\ 4 \end{bmatrix} + \lambda_2 \begin{bmatrix} 1 \\ 1 \\ 0 \\ 2 \end{bmatrix} + \lambda_3 \begin{bmatrix} -1 \\ -2 \\ 1 \\ 1 \end{bmatrix} = \mathbf{0} \quad (2.67)$$

for $\lambda_1, \dots, \lambda_3$. We write the vectors \mathbf{x}_i , $i = 1, 2, 3$, as the columns of a matrix and apply elementary row operations until we identify the pivot columns:

$$\left[\begin{array}{ccc} 1 & 1 & -1 \\ 2 & 1 & -2 \\ -3 & 0 & 1 \\ 4 & 2 & 1 \end{array} \right] \rightsquigarrow \dots \rightsquigarrow \left[\begin{array}{ccc} 1 & 1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{array} \right] \quad (2.68)$$

Here, every column of the matrix is a pivot column. Therefore, there is no non-trivial solution, and we require $\lambda_1 = 0, \lambda_2 = 0, \lambda_3 = 0$ to solve the equation system. Hence, the vectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ are linearly independent.

Remark. Consider a vector space V with k linearly independent vectors $\mathbf{b}_1, \dots, \mathbf{b}_k$ and m linear combinations

$$\begin{aligned} \mathbf{x}_1 &= \sum_{i=1}^k \lambda_{i1} \mathbf{b}_i, \\ &\vdots \\ \mathbf{x}_m &= \sum_{i=1}^k \lambda_{im} \mathbf{b}_i. \end{aligned} \quad (2.69)$$

Defining $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_k]$ as the matrix whose columns are the linearly independent vectors $\mathbf{b}_1, \dots, \mathbf{b}_k$, we can write

$$\mathbf{x}_j = \mathbf{B} \boldsymbol{\lambda}_j, \quad \boldsymbol{\lambda}_j = \begin{bmatrix} \lambda_{1j} \\ \vdots \\ \lambda_{kj} \end{bmatrix}, \quad j = 1, \dots, m, \quad (2.70)$$

1165 in a more compact form.

We want to test whether $\mathbf{x}_1, \dots, \mathbf{x}_m$ are linearly independent. For this purpose, we follow the general approach of testing when $\sum_{j=1}^m \psi_j \mathbf{x}_j = \mathbf{0}$. With (2.70), we obtain

$$\sum_{j=1}^m \psi_j \mathbf{x}_j = \sum_{j=1}^m \psi_j \mathbf{B} \boldsymbol{\lambda}_j = \mathbf{B} \sum_{j=1}^m \psi_j \boldsymbol{\lambda}_j. \quad (2.71)$$

1166 This means that $\{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ are linearly independent if and only if the 1167 column vectors $\{\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_m\}$ are linearly independent.

◇

1169 *Remark.* In a vector space V , m linear combinations of k vectors $\mathbf{x}_1, \dots, \mathbf{x}_k$ 1170 are linearly dependent if $m > k$.

◇

Example 2.15

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

$$\begin{aligned}\mathbf{x}_1 &= \mathbf{b}_1 - 2\mathbf{b}_2 + \mathbf{b}_3 - \mathbf{b}_4 \\ \mathbf{x}_2 &= -4\mathbf{b}_1 - 2\mathbf{b}_2 + 4\mathbf{b}_4 \\ \mathbf{x}_3 &= 2\mathbf{b}_1 + 3\mathbf{b}_2 - \mathbf{b}_3 - 3\mathbf{b}_4 \\ \mathbf{x}_4 &= 17\mathbf{b}_1 - 10\mathbf{b}_2 + 11\mathbf{b}_3 + \mathbf{b}_4\end{aligned}\quad (2.72)$$

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

$$\left\{ \begin{bmatrix} 1 \\ -2 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} -4 \\ -2 \\ 0 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ -1 \\ -3 \end{bmatrix}, \begin{bmatrix} 17 \\ -10 \\ 11 \\ 1 \end{bmatrix} \right\} \quad (2.73)$$

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

$$\mathbf{A} = \begin{bmatrix} 1 & -4 & 2 & 17 \\ -2 & -2 & 3 & -10 \\ 1 & 0 & -1 & 11 \\ -1 & 4 & -3 & 1 \end{bmatrix} \quad (2.74)$$

is given as

$$\begin{bmatrix} 1 & 0 & 0 & -7 \\ 0 & 1 & 0 & -15 \\ 0 & 0 & 1 & -18 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \quad (2.75)$$

We see that the corresponding linear equation system is non-trivially solvable: The last column is not a pivot column, and $\mathbf{x}_4 = -7\mathbf{x}_1 - 15\mathbf{x}_2 - 18\mathbf{x}_3$. Therefore, $\mathbf{x}_1, \dots, \mathbf{x}_4$ are linearly dependent as \mathbf{x}_4 can be expressed as a linear combination of $\mathbf{x}_1, \dots, \mathbf{x}_3$.

1171

2.6 Basis and Rank

1172 In a vector space V , we are particularly interested in sets of vectors A that
 1173 possess the property that any vector $\mathbf{v} \in V$ can be obtained by a linear
 1174 combination of vectors in A . These vectors are special vectors, and in the
 1175 following, we will characterize them.

1176

2.6.1 Generating Set and Basis

1177 **Definition 2.12** (Generating Set and Span). Consider a vector space $V =$
 1178 $(\mathcal{V}, +, \cdot)$ and set of vectors $\mathcal{A} = \{\mathbf{x}_1, \dots, \mathbf{x}_k\} \subseteq \mathcal{V}$. If every vector $\mathbf{v} \in$

generating set
span 1179 \mathcal{V} can be expressed as a linear combination of $\mathbf{x}_1, \dots, \mathbf{x}_k$, \mathcal{A} is called a
1180 *generating set* of V . The set of all linear combinations of vectors in \mathcal{A} is
1181 called the *span* of \mathcal{A} . If \mathcal{A} spans the vector space V , we write $V = \text{span}[\mathcal{A}]$
1182 or $V = \text{span}[\mathbf{x}_1, \dots, \mathbf{x}_k]$.

1183 Generating sets are sets of vectors that span vector (sub)spaces, i.e.,
1184 every vector can be represented as a linear combination of the vectors
1185 in the generating set. Now, we will be more specific and characterize the
1186 smallest generating set that spans a vector (sub)space.

minimal 1187 **Definition 2.13 (Basis).** Consider a vector space $V = (\mathcal{V}, +, \cdot)$ and $\mathcal{A} \subseteq$
1188 \mathcal{V} . A generating set \mathcal{A} of V is called *minimal* if there exists no smaller set
1189 $\tilde{\mathcal{A}} \subseteq \mathcal{A} \subseteq \mathcal{V}$ that spans V . Every linearly independent generating set of V
1190 is minimal and is called a *basis* of V .

A basis is a minimal
generating set and^{¶1}
maximal linearly
independent set of
vectors. 1190 Let $V = (\mathcal{V}, +, \cdot)$ be a vector space and $\mathcal{B} \subseteq \mathcal{V}, \mathcal{B} \neq \emptyset$. Then, the
1191 following statements are equivalent:

- 1193 • \mathcal{B} is a basis of V
- 1194 • \mathcal{B} is a minimal generating set
- 1195 • \mathcal{B} is a maximal linearly independent set of vectors in V , i.e., adding any
other vector to this set will make it linearly dependent.
- 1196 • Every vector $\mathbf{x} \in V$ is a linear combination of vectors from \mathcal{B} , and every
linear combination is unique, i.e., with

$$\mathbf{x} = \sum_{i=1}^k \lambda_i \mathbf{b}_i = \sum_{i=1}^k \psi_i \mathbf{b}_i \quad (2.76)$$

1197 and $\lambda_i, \psi_i \in \mathbb{R}$, $\mathbf{b}_i \in \mathcal{B}$ it follows that $\lambda_i = \psi_i$, $i = 1, \dots, k$.

Example 2.16

canonical/standard
basis • In \mathbb{R}^3 , the *canonical/standard basis* is

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}. \quad (2.77)$$

• Different bases in \mathbb{R}^3 are

$$\mathcal{B}_1 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}, \mathcal{B}_2 = \left\{ \begin{bmatrix} 0.5 \\ 0.8 \\ 0.4 \end{bmatrix}, \begin{bmatrix} 1.8 \\ 0.3 \\ 0.3 \end{bmatrix}, \begin{bmatrix} -2.2 \\ -1.3 \\ 3.5 \end{bmatrix} \right\}. \quad (2.78)$$

• The set

$$\mathcal{A} = \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ -4 \end{bmatrix} \right\} \quad (2.79)$$

is linearly independent, but not a generating set (and no basis) of \mathbb{R}^4 : For instance, the vector $[1, 0, 0, 0]^\top$ cannot be obtained by a linear combination of elements in \mathcal{A} .

1198 *Remark.* Every vector space V possesses a basis \mathcal{B} . The examples above
1199 show that there can be many bases of a vector space V , i.e., there is no
1200 unique basis. However, all bases possess the same number of elements,
1201 the *basis vectors*. \diamond

1202 We only consider finite-dimensional vector spaces V . In this case, the
1203 *dimension* of V is the number of basis vectors, and we write $\dim(V)$. If
1204 $U \subseteq V$ is a subspace of V then $\dim(U) \leq \dim(V)$ and $\dim(U) = \dim(V)$
1205 if and only if $U = V$. Intuitively, the dimension of a vector space can be
1206 thought of as the number of independent directions in this vector space.

1207 *Remark.* A basis of a subspace $U = \text{span}[\mathbf{x}_1, \dots, \mathbf{x}_m] \subseteq \mathbb{R}^n$ can be found
1208 by executing the following steps:

- 1209 1. Write the spanning vectors as columns of a matrix \mathbf{A}
- 1210 2. Determine the row echelon form of \mathbf{A} .
- 1211 3. The spanning vectors associated with the pivot columns are a basis of
1212 U .

1213 basis vectors
The dimension of a
vector space
corresponds to the
number of basis
vectors.
dimension

Example 2.17 (Determining a Basis)

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

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \\ -1 \\ -1 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 2 \\ -1 \\ 1 \\ 2 \\ -2 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 3 \\ -4 \\ 3 \\ 5 \\ -3 \end{bmatrix}, \quad \mathbf{x}_4 = \begin{bmatrix} -1 \\ 8 \\ -5 \\ -6 \\ 1 \end{bmatrix} \in \mathbb{R}^5, \quad (2.80)$$

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

$$\sum_{i=1}^4 \lambda_i \mathbf{x}_i = \mathbf{0}, \quad (2.81)$$

which leads to a homogeneous equation system with matrix

$$[\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4] = \begin{bmatrix} 1 & 2 & 3 & -1 \\ 2 & -1 & -4 & 8 \\ -1 & 1 & 3 & -5 \\ -1 & 2 & 5 & -6 \\ -1 & -2 & -3 & 1 \end{bmatrix}. \quad (2.82)$$

With the basic transformation rules for systems of linear equations, we obtain the reduced row echelon form

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

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

2.6.2 Rank

- rank
- 1214
 - 1215 The number of linearly independent columns of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$
 - 1216 equals the number of linearly independent rows and is called the *rank*
 - 1217 of \mathbf{A} and is denoted by $\text{rk}(\mathbf{A})$.
 - 1218 *Remark.* The rank of a matrix has some important properties:
 - 1219 • $\text{rk}(\mathbf{A}) = \text{rk}(\mathbf{A}^\top)$, i.e., the column rank equals the row rank.
 - 1220 • The columns of $\mathbf{A} \in \mathbb{R}^{m \times n}$ span a subspace $U \subseteq \mathbb{R}^m$ with $\dim(U) =$
 - 1221 $\text{rk}(\mathbf{A})$. Later, we will call this subspace the *image* or *range*. A basis of
 - 1222 U can be found by applying Gaussian elimination to \mathbf{A} to identify the
 - 1223 pivot columns.
 - 1224 • The rows of $\mathbf{A} \in \mathbb{R}^{m \times n}$ span a subspace $W \subseteq \mathbb{R}^n$ with $\dim(W) =$
 - 1225 $\text{rk}(\mathbf{A})$. A basis of W can be found by applying Gaussian elimination to
 - 1226 \mathbf{A}^\top .
 - 1227 • For all $\mathbf{A} \in \mathbb{R}^{n \times n}$ holds: \mathbf{A} is regular (invertible) if and only if $\text{rk}(\mathbf{A}) =$
 - 1228 n .
 - 1229 • For all $\mathbf{A} \in \mathbb{R}^{m \times n}$ and all $\mathbf{b} \in \mathbb{R}^m$ it holds that the linear equation
 - 1230 system $\mathbf{Ax} = \mathbf{b}$ can be solved if and only if $\text{rk}(\mathbf{A}) = \text{rk}(\mathbf{A}|\mathbf{b})$, where
 - 1231 $\mathbf{A}|\mathbf{b}$ denotes the augmented system.
 - 1232 • For $\mathbf{A} \in \mathbb{R}^{m \times n}$ the subspace of solutions for $\mathbf{Ax} = \mathbf{0}$ possesses dimen-
 - 1233 sion $n - \text{rk}(\mathbf{A})$. Later, we will call this subspace the *kernel* or the *null*
 - 1234 *space*.

kernel

null space

full rank

rank deficient

 - 1235 • A matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ has *full rank* if its rank equals the largest possible
 - 1236 rank for a matrix of the same dimensions. This means that the rank of
 - 1237 a full-rank matrix is the lesser of the number of rows and columns, i.e.,
 - 1238 $\text{rk}(\mathbf{A}) = \min(m, n)$. A matrix is said to be *rank deficient* if it does not
 - 1239 have full rank.

1240



Example 2.18 (Rank)

- $\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$. \mathbf{A} possesses two linearly independent rows (and columns). Therefore, $\text{rk}(\mathbf{A}) = 2$.
- $\mathbf{A} = \begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix}$ We use Gaussian elimination to determine the rank:

$$\begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix} \rightsquigarrow \dots \rightsquigarrow \begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & 3 \\ 0 & 0 & 0 \end{bmatrix}. \quad (2.83)$$

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

1241

2.7 Linear Mappings

In the following, we will study mappings on vector spaces that preserve their structure. In the beginning of the chapter, we said that vectors are objects that can be added together and multiplied by a scalar, and the resulting object is still a vector. This property we wish to preserve when applying the mapping: Consider two real vector spaces V, W . A mapping $\Phi : V \rightarrow W$ preserves the structure of the vector space if

$$\Phi(\mathbf{x} + \mathbf{y}) = \Phi(\mathbf{x}) + \Phi(\mathbf{y}) \quad (2.84)$$

$$\Phi(\lambda\mathbf{x}) = \lambda\Phi(\mathbf{x}) \quad (2.85)$$

1242 for all $\mathbf{x}, \mathbf{y} \in V$ and $\lambda \in \mathbb{R}$. We can summarize this in the following
1243 definition:

Definition 2.14 (Linear Mapping). For vector spaces V, W , a mapping $\Phi : V \rightarrow W$ is called a *linear mapping* (or *vector space homomorphism/linear transformation*) if

$$\forall \mathbf{x}, \mathbf{y} \in V \forall \lambda, \psi \in \mathbb{R} : \Phi(\lambda\mathbf{x} + \psi\mathbf{y}) = \lambda\Phi(\mathbf{x}) + \psi\Phi(\mathbf{y}). \quad (2.86)$$

1244 Before we continue, we will briefly introduce special mappings.

1245 **Definition 2.15** (Injective, Surjective, Bijective). Consider a mapping $\Phi : 1246 \mathcal{V} \rightarrow \mathcal{W}$, where \mathcal{V}, \mathcal{W} can be arbitrary sets. Then Φ is called

- 1247 • *injective* if for any $\mathbf{x}, \mathbf{y} \in \mathcal{V}$ it follows that $\Phi(\mathbf{x}) \neq \Phi(\mathbf{y})$ if and only if
1248 $\mathbf{x} \neq \mathbf{y}$.
- 1249 • *surjective* if $\Phi(\mathcal{V}) = \mathcal{W}$.

linear mapping
vector space
homomorphism
linear
transformation

injective
surjective

- 1250 • *bijection* if it is injective and surjective. bijective
- 1251 If Φ is injective then it can also be “undone”, i.e., there exists a mapping
- 1252 $\Psi : W \rightarrow V$ so that $\Psi \circ \Phi(\mathbf{x}) = \mathbf{x}$. If Φ is surjective then every element
- 1253 in \mathcal{W} can be “reached” from \mathcal{V} using Φ .
- 1254 With these definitions, we introduce the following special cases of linear
- 1255 mappings between vector spaces V and W :
- Isomorphism • *Isomorphism*: $\Phi : V \rightarrow W$ linear and bijective
- Endomorphism • *Endomorphism*: $\Phi : V \rightarrow V$ linear
- Automorphism • *Automorphism*: $\Phi : V \rightarrow V$ linear and bijective
- identity mapping • We define $\text{id}_V : V \rightarrow V$, $\mathbf{x} \mapsto \mathbf{x}$ as the *identity mapping* in V .

Example 2.19 (Homomorphism)

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

$$\begin{aligned}\Phi\left(\begin{bmatrix}x_1 \\ x_2\end{bmatrix} + \begin{bmatrix}y_1 \\ y_2\end{bmatrix}\right) &= (x_1 + y_1) + i(x_2 + y_2) = x_1 + ix_2 + y_1 + iy_2 \\ &= \Phi\left(\begin{bmatrix}x_1 \\ x_2\end{bmatrix}\right) + \Phi\left(\begin{bmatrix}y_1 \\ y_2\end{bmatrix}\right) \\ \Phi\left(\lambda \begin{bmatrix}x_1 \\ x_2\end{bmatrix}\right) &= \lambda x_1 + \lambda ix_2 = \lambda(x_1 + ix_2) = \lambda\Phi\left(\begin{bmatrix}x_1 \\ x_2\end{bmatrix}\right).\end{aligned}\tag{2.87}$$

This also justifies why complex numbers can be represented as tuples in \mathbb{R}^2 : There is a bijective linear mapping that converts the elementwise addition of tuples in \mathbb{R}^2 into the set of complex numbers with the corresponding addition. Note that we only showed linearity, but not the bijection.

1260 **Theorem 2.16.** *Finite-dimensional vector spaces V and W are isomorphic*

1261 *if and only if $\dim(V) = \dim(W)$.*

1262 Theorem 2.16 states that there exists a linear, bijective mapping be-

1263 tween two vector spaces of the same dimension. Intuitively, this means

1264 that vector spaces of the same dimension are kind of the same thing as

1265 they can be transformed into each other without incurring any loss.

1266 Theorem 2.16 also gives us the justification to treat $\mathbb{R}^{m \times n}$ (the vector

1267 space of $m \times n$ -matrices) and \mathbb{R}^{mn} (the vector space of vectors of length

1268 mn) the same as their dimensions are mn , and there exists a linear, bijec-

1269 tive mapping that transforms one into the other.

1270 **Remark.** Consider vector spaces V, W, X . Then:

- 1271 • For linear mappings $\Phi : V \rightarrow W$ and $\Psi : W \rightarrow X$ the mapping
- 1272 $\Psi \circ \Phi : V \rightarrow X$ is also linear.
- 1273 • If $\Phi : V \rightarrow W$ is an isomorphism then $\Phi^{-1} : W \rightarrow V$ is an isomor-
- 1274 phism, too.

- 1275 • If $\Phi : V \rightarrow W$, $\Psi : V \rightarrow W$ are linear then $\Phi + \Psi$ and $\lambda\Phi$, $\lambda \in \mathbb{R}$, are
1276 linear, too.

◇

2.7.1 Matrix Representation of Linear Mappings

Any n -dimensional vector space is isomorphic to \mathbb{R}^n (Theorem 2.16). We consider a basis $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ of an n -dimensional vector space V . In the following, the order of the basis vectors will be important. Therefore, we write

$$B = (\mathbf{b}_1, \dots, \mathbf{b}_n) \quad (2.88)$$

1279 and call this n -tuple an *ordered basis* of V .

ordered basis

1280 *Remark* (Notation). We are at the point where notation gets a bit tricky.
1281 Therefore, we summarize some parts here. $B = (\mathbf{b}_1, \dots, \mathbf{b}_n)$ is an ordered
1282 basis, $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ is an (unordered) basis, and $B = [\mathbf{b}_1, \dots, \mathbf{b}_n]$ is a
1283 matrix whose columns are the vectors $\mathbf{b}_1, \dots, \mathbf{b}_n$. ◇

Definition 2.17 (Coordinates). Consider a vector space V and an ordered basis $B = (\mathbf{b}_1, \dots, \mathbf{b}_n)$ of V . For any $\mathbf{x} \in V$ we obtain a unique representation (linear combination)

$$\mathbf{x} = \alpha_1 \mathbf{b}_1 + \dots + \alpha_n \mathbf{b}_n \quad (2.89)$$

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

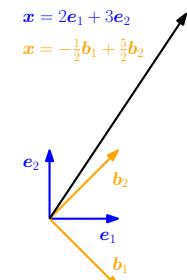
$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \in \mathbb{R}^n \quad (2.90)$$

1284 is the *coordinate vector/coordinate representation* of \mathbf{x} with respect to the
1285 ordered basis B .

coordinate vector
coordinate
representation

1286 *Remark.* Intuitively, the basis vectors can be thought of as being equipped
1287 with units (including common units such as “kilograms” or “seconds”).
1288 Let us have a look at a geometric vector $\mathbf{x} \in \mathbb{R}^2$ with coordinates $[2, 3]^\top$
1289 with respect to the standard basis $\mathbf{e}_1, \mathbf{e}_2$ in \mathbb{R}^2 . This means, we can write
1290 $\mathbf{x} = 2\mathbf{e}_1 + 3\mathbf{e}_2$. However, we do not have to choose the standard basis
1291 to represent this vector. If we use the basis vectors $\mathbf{b}_1 = [1, -1]^\top, \mathbf{b}_2 =$
1292 $[1, 1]^\top$ we will obtain the coordinates $\frac{1}{2}[-1, 5]^\top$ to represent the same
1293 vector (see Figure 2.7). ◇

Figure 2.7
Different coordinate representations of a vector \mathbf{x} , depending on the choice of basis.



1294 *Remark.* For an n -dimensional vector space V and an ordered basis B
1295 of V , the mapping $\Phi : \mathbb{R}^n \rightarrow V$, $\Phi(\mathbf{e}_i) = \mathbf{b}_i$, $i = 1, \dots, n$, is linear
1296 (and because of Theorem 2.16 an isomorphism), where $(\mathbf{e}_1, \dots, \mathbf{e}_n)$ is
1297 the standard basis of \mathbb{R}^n . ◇

1299 Now we are ready to make an explicit connection between matrices and
1300 linear mappings between finite-dimensional vector spaces.

Definition 2.18 (Transformation matrix). Consider vector spaces V, W with corresponding (ordered) bases $B = (\mathbf{b}_1, \dots, \mathbf{b}_n)$ and $C = (\mathbf{c}_1, \dots, \mathbf{c}_m)$. Moreover, we consider a linear mapping $\Phi : V \rightarrow W$. For $j \in \{1, \dots, n\}$

$$\Phi(\mathbf{b}_j) = \alpha_{1j}\mathbf{c}_1 + \dots + \alpha_{mj}\mathbf{c}_m = \sum_{i=1}^m \alpha_{ij}\mathbf{c}_i \quad (2.91)$$

is the unique representation of $\Phi(\mathbf{b}_j)$ with respect to C . Then, we call the $m \times n$ -matrix \mathbf{A}_Φ whose elements are given by

$$A_\Phi(i, j) = \alpha_{ij} \quad (2.92)$$

1301 transformation matrix 1302 the *transformation matrix* of Φ (with respect to the ordered bases B of V and C of W).

The coordinates of $\Phi(\mathbf{b}_j)$ with respect to the ordered basis C of W are the j -th column of \mathbf{A}_Φ . Consider (finite-dimensional) vector spaces V, W with ordered bases B, C and a linear mapping $\Phi : V \rightarrow W$ with transformation matrix \mathbf{A}_Φ . If $\hat{\mathbf{x}}$ is the coordinate vector of $\mathbf{x} \in V$ with respect to B and $\hat{\mathbf{y}}$ the coordinate vector of $\mathbf{y} = \Phi(\mathbf{x}) \in W$ with respect to C , then

$$\hat{\mathbf{y}} = \mathbf{A}_\Phi \hat{\mathbf{x}}. \quad (2.93)$$

1303 This means that the transformation matrix can be used to map coordinates
1304 with respect to an ordered basis in V to coordinates with respect to an
1305 ordered basis in W .

Example 2.20 (Transformation Matrix)

Consider a homomorphism $\Phi : V \rightarrow W$ and ordered bases $B = (\mathbf{b}_1, \dots, \mathbf{b}_3)$ of V and $C = (\mathbf{c}_1, \dots, \mathbf{c}_4)$ of W . With

$$\begin{aligned} \Phi(\mathbf{b}_1) &= \mathbf{c}_1 - \mathbf{c}_2 + 3\mathbf{c}_3 - \mathbf{c}_4 \\ \Phi(\mathbf{b}_2) &= 2\mathbf{c}_1 + \mathbf{c}_2 + 7\mathbf{c}_3 + 2\mathbf{c}_4 \\ \Phi(\mathbf{b}_3) &= 3\mathbf{c}_2 + \mathbf{c}_3 + 4\mathbf{c}_4 \end{aligned} \quad (2.94)$$

the transformation matrix \mathbf{A}_Φ with respect to B and C satisfies $\Phi(\mathbf{b}_k) = \sum_{i=1}^4 \alpha_{ik}\mathbf{c}_i$ for $k = 1, \dots, 3$ and is given as

$$\mathbf{A}_\Phi = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3] = \begin{bmatrix} 1 & 2 & 0 \\ -1 & 1 & 3 \\ 3 & 7 & 1 \\ -1 & 2 & 4 \end{bmatrix}, \quad (2.95)$$

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

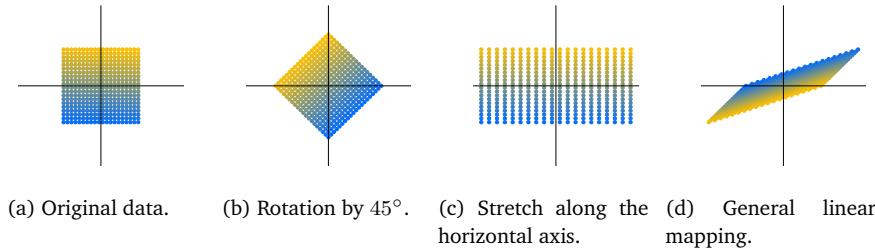


Figure 2.8 Three examples of linear transformations of the vectors shown as dots in (a). (b) Rotation by 45°; (c) Stretching of the horizontal coordinates by 2; (d) Combination of reflection, rotation and stretching.

Example 2.21 (Linear Transformations of Vectors)

We consider three linear transformations of a set of vectors in \mathbb{R}^2 with the transformation matrices

$$\mathbf{A}_1 = \begin{bmatrix} \cos(\frac{\pi}{4}) & -\sin(\frac{\pi}{4}) \\ \sin(\frac{\pi}{4}) & \cos(\frac{\pi}{4}) \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{A}_3 = \frac{1}{2} \begin{bmatrix} 3 & -1 \\ 1 & -1 \end{bmatrix}. \quad (2.96)$$

Figure 2.8 gives three examples of linear transformations of a set of vectors. Figure 2.8(a) shows 400 vectors in \mathbb{R}^2 , each of which is represented by a dot at the corresponding (x_1, x_2) -coordinates. The vectors are arranged in a square. When we use matrix \mathbf{A}_1 in (2.96) to linearly transform each of these vectors, we obtain the rotated square in Figure 2.8(b). If we apply the linear mapping represented by \mathbf{A}_2 , we obtain the rectangle in Figure 2.8(c) where each x_1 -coordinate is stretched by 2. Figure 2.8(d) shows the original square from Figure 2.8(a) when linearly transformed using \mathbf{A}_3 , which is a combination of a reflection, a rotation and a stretch.

2.7.2 Basis Change

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

$$B = (\mathbf{b}_1, \dots, \mathbf{b}_n), \quad \tilde{B} = (\tilde{\mathbf{b}}_1, \dots, \tilde{\mathbf{b}}_n) \quad (2.97)$$

of V and two ordered bases

$$C = (\mathbf{c}_1, \dots, \mathbf{c}_m), \quad \tilde{C} = (\tilde{\mathbf{c}}_1, \dots, \tilde{\mathbf{c}}_m) \quad (2.98)$$

of W . Moreover, $\mathbf{A}_\Phi \in \mathbb{R}^{m \times n}$ is the transformation matrix of the linear mapping $\Phi : V \rightarrow W$ with respect to the bases B and C , and $\tilde{\mathbf{A}}_\Phi \in \mathbb{R}^{m \times n}$ is the corresponding transformation mapping with respect to \tilde{B} and \tilde{C} . In the following, we will investigate how \mathbf{A} and $\tilde{\mathbf{A}}$ are related, i.e., how/whether we can transform \mathbf{A}_Φ into $\tilde{\mathbf{A}}_\Phi$ if we choose to perform a basis change from B, C to \tilde{B}, \tilde{C} .

1313 *Remark.* We effectively get different coordinate representations of the
 1314 identity mapping id_V . In the context of Figure 2.7, this would mean to
 1315 map coordinates with respect to e_1, e_2 onto coordinates with respect to
 1316 b_1, b_2 without changing the vector x . By changing the basis and corre-
 1317 spondingly the representation of vectors, the transformation matrix with
 1318 respect to this new basis can have a particularly simple form that allows
 1319 for straightforward computation. \diamond

Example 2.22 (Basis Change)

Consider a transformation matrix

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \quad (2.99)$$

with respect to the canonical basis in \mathbb{R}^2 . If we define a new basis

$$B = \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right) \quad (2.100)$$

we obtain a diagonal transformation matrix

$$\tilde{\mathbf{A}} = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \quad (2.101)$$

with respect to B , which is easier to work with than \mathbf{A} .

1320 In the following, we will look at mappings that transform coordinate
 1321 vectors with respect to one basis into coordinate vectors with respect to
 1322 a different basis. We will state our main result first and then provide an
 1323 explanation.

Theorem 2.19 (Basis Change). *For a linear mapping $\Phi : V \rightarrow W$, ordered bases*

$$B = (\mathbf{b}_1, \dots, \mathbf{b}_n), \quad \tilde{B} = (\tilde{\mathbf{b}}_1, \dots, \tilde{\mathbf{b}}_n) \quad (2.102)$$

of V and

$$C = (\mathbf{c}_1, \dots, \mathbf{c}_m), \quad \tilde{C} = (\tilde{\mathbf{c}}_1, \dots, \tilde{\mathbf{c}}_m) \quad (2.103)$$

of W , and a transformation matrix \mathbf{A}_Φ of Φ with respect to B and C , the corresponding transformation matrix $\tilde{\mathbf{A}}_\Phi$ with respect to the bases \tilde{B} and \tilde{C} is given as

$$\tilde{\mathbf{A}}_\Phi = \mathbf{T}^{-1} \mathbf{A}_\Phi \mathbf{S}. \quad (2.104)$$

1324 Here, $\mathbf{S} \in \mathbb{R}^{n \times n}$ is the transformation matrix of id_V that maps coordinates
 1325 with respect to B onto coordinates with respect to \tilde{B} , and $\mathbf{T} \in \mathbb{R}^{m \times m}$ is the
 1326 transformation matrix of id_W that maps coordinates with respect to C onto
 1327 coordinates with respect to \tilde{C} .

Proof Following Drumm and Weil (2001) we can write the vectors of the new basis \tilde{B} of V as a linear combination of the basis vectors of B , such that

$$\tilde{\mathbf{b}}_j = s_{1j}\mathbf{b}_1 + \cdots + s_{nj}\mathbf{b}_n = \sum_{i=1}^n s_{ij}\mathbf{b}_i, \quad j = 1, \dots, n. \quad (2.105)$$

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

$$\tilde{\mathbf{c}}_k = t_{1k}\mathbf{c}_1 + \cdots + t_{mk}\mathbf{c}_m = \sum_{l=1}^m t_{lk}\mathbf{c}_l, \quad k = 1, \dots, m. \quad (2.106)$$

We define $\mathbf{S} = ((s_{ij})) \in \mathbb{R}^{n \times n}$ as the transformation matrix that maps coordinates with respect to \tilde{B} onto coordinates with respect to B , and $\mathbf{T} = ((t_{lk})) \in \mathbb{R}^{m \times m}$ as the transformation matrix that maps coordinates with respect to \tilde{C} onto coordinates with respect to C . In particular, the j th column of \mathbf{S} are the coordinate representations of $\tilde{\mathbf{b}}_j$ with respect to B and the j th columns of \mathbf{T} is the coordinate representation of $\tilde{\mathbf{c}}_j$ with respect to C . Note that both \mathbf{S} and \mathbf{T} are regular.

For all $j = 1, \dots, n$, we get

$$\Phi(\tilde{\mathbf{b}}_j) = \sum_{k=1}^m \underbrace{\tilde{a}_{kj} \tilde{\mathbf{c}}_k}_{\in W} \stackrel{(2.106)}{=} \sum_{k=1}^m \tilde{a}_{kj} \sum_{l=1}^m t_{lk} \mathbf{c}_l = \sum_{l=1}^m \left(\sum_{k=1}^m t_{lk} \tilde{a}_{kj} \right) \mathbf{c}_l, \quad (2.107)$$

where we first expressed the new basis vectors $\tilde{\mathbf{c}}_k \in W$ as linear combinations of the basis vectors $\mathbf{c}_l \in W$ and then swapped the order of summation. When we express the $\tilde{\mathbf{b}}_j \in V$ as linear combinations of $\mathbf{b}_j \in V$, we arrive at

$$\Phi(\tilde{\mathbf{b}}_j) \stackrel{(2.105)}{=} \Phi \left(\sum_{i=1}^n s_{ij} \mathbf{b}_i \right) = \sum_{i=1}^n s_{ij} \Phi(\mathbf{b}_i) = \sum_{i=1}^n s_{ij} \sum_{l=1}^m a_{li} \mathbf{c}_l \quad (2.108)$$

$$= \sum_{l=1}^m \left(\sum_{i=1}^n a_{li} s_{ij} \right) \mathbf{c}_l, \quad j = 1, \dots, n, \quad (2.109)$$

where we exploited the linearity of Φ . Comparing (2.107) and (2.109), it follows for all $j = 1, \dots, n$ and $l = 1, \dots, m$ that

$$\sum_{k=1}^m t_{lk} \tilde{a}_{kj} = \sum_{i=1}^n a_{li} s_{ij} \quad (2.110)$$

and, therefore,

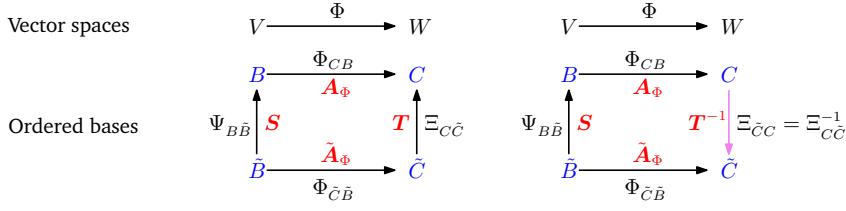
$$\mathbf{T} \tilde{\mathbf{A}}_\Phi = \mathbf{A}_\Phi \mathbf{S} \in \mathbb{R}^{m \times n}, \quad (2.111)$$

such that

$$\tilde{\mathbf{A}}_\Phi = \mathbf{T}^{-1} \mathbf{A}_\Phi \mathbf{S}, \quad (2.112)$$

which proves Theorem 2.19. \square

Figure 2.9 For a homomorphism $\Phi : V \rightarrow W$ and ordered bases B, \tilde{B} of V and C, \tilde{C} of W (marked in blue), we can express the mapping $\Phi_{\tilde{C}\tilde{B}}$ with respect to the bases \tilde{B}, \tilde{C} equivalently as a composition of the homomorphisms $\Phi_{\tilde{C}\tilde{B}} = \Xi_{\tilde{C}C} \circ \Phi_{CB} \circ \Psi_{B\tilde{B}}$ with respect to the bases in the subscripts. The corresponding transformation matrices are in red.



Theorem 2.19 tells us that with a basis change in V (B is replaced with \tilde{B}) and W (C is replaced with \tilde{C}) the transformation matrix A_Φ of a linear mapping $\Phi : V \rightarrow W$ is replaced by an equivalent matrix \tilde{A}_Φ with

$$\tilde{A}_\Phi = T^{-1} A_\Phi S. \quad (2.113)$$

Figure 2.9 illustrates this relation: Consider a homomorphism $\Phi : V \rightarrow W$ and ordered bases B, \tilde{B} of V and C, \tilde{C} of W . The mapping Φ_{CB} is an instantiation of Φ and maps basis vectors of B onto linear combinations of basis vectors of C . Assuming, we know the transformation matrix A_Φ of Φ_{CB} with respect to the ordered bases B, C . When we perform a basis change from B to \tilde{B} in V and from C to \tilde{C} in W , we can determine the corresponding transformation matrix \tilde{A}_Φ as follows: First, we find the matrix representation of the linear mapping $\Psi_{B\tilde{B}} : V \rightarrow V$ that maps coordinates with respect to the new basis \tilde{B} onto the (unique) coordinates with respect to the “old” basis B (in V). Then, we use the transformation matrix A_Φ of $\Phi_{CB} : V \rightarrow W$ to map these coordinates onto the coordinates with respect to C in W . Finally, we use a linear mapping $\Xi_{C\tilde{C}} : W \rightarrow W$ to map the coordinates with respect to C onto coordinates with respect to \tilde{C} . Therefore, we can express the linear mapping $\Phi_{\tilde{C}\tilde{B}}$ as a composition of linear mappings that involve the “old” basis:

$$\Phi_{\tilde{C}\tilde{B}} = \Xi_{C\tilde{C}} \circ \Phi_{CB} \circ \Psi_{B\tilde{B}} = \Xi_{C\tilde{C}}^{-1} \circ \Phi_{CB} \circ \Psi_{B\tilde{B}}. \quad (2.114)$$

Concretely, we use $\Psi_{B\tilde{B}} = \text{id}_V$ and $\Xi_{C\tilde{C}} = \text{id}_W$, i.e., the identity mappings that map vectors onto themselves, but with respect to a different basis.

equivalent
1338 **Definition 2.20** (Equivalence). Two matrices $A, \tilde{A} \in \mathbb{R}^{m \times n}$ are *equivalent* if there exist regular matrices $S \in \mathbb{R}^{n \times n}$ and $T \in \mathbb{R}^{m \times m}$, such that $\tilde{A} = T^{-1}AS$.

similar
1341 **Definition 2.21** (Similarity). Two matrices $A, \tilde{A} \in \mathbb{R}^{n \times n}$ are *similar* if
1342 there exists a regular matrix $S \in \mathbb{R}^{n \times n}$ with $\tilde{A} = S^{-1}AS$

1343 *Remark.* Similar matrices are always equivalent. However, equivalent matrices are not necessarily similar. ◇

1345 *Remark.* Consider vector spaces V, W, X . From the remark on page 48 we
1346 already know that for linear mappings $\Phi : V \rightarrow W$ and $\Psi : W \rightarrow X$ the
1347 mapping $\Psi \circ \Phi : V \rightarrow X$ is also linear. With transformation matrices A_Φ and A_Ψ of the corresponding mappings, the overall transformation matrix
1348 is $A_{\Psi \circ \Phi} = A_\Psi A_\Phi$. ◇

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

- \mathbf{A}_Φ is the transformation matrix of a linear mapping $\Phi_{CB} : V \rightarrow W$ with respect to the bases B, C .
- $\tilde{\mathbf{A}}_\Phi$ is the transformation matrix of the linear mapping $\Phi_{\tilde{C}\tilde{B}} : V \rightarrow W$ with respect to the bases \tilde{B}, \tilde{C} .
- \mathbf{S} is the transformation matrix of a linear mapping $\Psi_{B\tilde{B}} : V \rightarrow V$ (automorphism) that represents \tilde{B} in terms of B . Normally, $\Psi = \text{id}_V$ is the identity mapping in V .
- \mathbf{T} is the transformation matrix of a linear mapping $\Xi_{C\tilde{C}} : W \rightarrow W$ (automorphism) that represents \tilde{C} in terms of C . Normally, $\Xi = \text{id}_W$ is the identity mapping in W .

If we (informally) write down the transformations just in terms of bases then $\mathbf{A}_\Phi : B \rightarrow C$, $\tilde{\mathbf{A}}_\Phi : \tilde{B} \rightarrow \tilde{C}$, $\mathbf{S} : \tilde{B} \rightarrow B$, $\mathbf{T} : \tilde{C} \rightarrow C$ and $\mathbf{T}^{-1} : C \rightarrow \tilde{C}$, and

$$\tilde{B} \rightarrow \tilde{C} = \tilde{B} \rightarrow B \rightarrow C \rightarrow \tilde{C} \quad (2.115)$$

$$\tilde{\mathbf{A}}_\Phi = \mathbf{T}^{-1} \mathbf{A}_\Phi \mathbf{S}. \quad (2.116)$$

Note that the execution order in (2.116) is from right to left because vectors are multiplied at the right-hand side so that $\mathbf{x} \mapsto \mathbf{S}\mathbf{x} \mapsto \mathbf{A}_\Phi(\mathbf{S}\mathbf{x}) \mapsto \mathbf{T}^{-1}(\mathbf{A}_\Phi(\mathbf{S}\mathbf{x})) = \tilde{\mathbf{A}}_\Phi \mathbf{x}$.

Example 2.23 (Basis Change)

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

$$\mathbf{A}_\Phi = \begin{bmatrix} 1 & 2 & 0 \\ -1 & 1 & 3 \\ 3 & 7 & 1 \\ -1 & 2 & 4 \end{bmatrix} \quad (2.117)$$

with respect to the standard bases

$$B = \left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right), \quad C = \left(\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right). \quad (2.118)$$

We seek the transformation matrix $\tilde{\mathbf{A}}_\Phi$ of Φ with respect to the new bases

$$\tilde{B} = \left(\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right) \in \mathbb{R}^3, \quad \tilde{C} = \left(\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right). \quad (2.119)$$

Then,

$$\mathbf{S} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (2.120)$$

where the i th column of \mathbf{S} is the coordinate representation of $\tilde{\mathbf{b}}_i$ in terms of the basis vectors of B . Similarly, the j th column of \mathbf{T} is the coordinate representation of $\tilde{\mathbf{c}}_j$ in terms of the basis vectors of C .

Therefore, we obtain

$$\tilde{\mathbf{A}}_\Phi = \mathbf{T}^{-1} \mathbf{A}_\Phi \mathbf{S} = \frac{1}{2} \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 3 & 2 & 1 \\ 0 & 4 & 2 \\ 10 & 8 & 4 \\ 1 & 6 & 3 \end{bmatrix} \quad (2.121a)$$

$$= \begin{bmatrix} -4 & -4 & -2 \\ 6 & 0 & 0 \\ 4 & 8 & 4 \\ 1 & 6 & 3 \end{bmatrix}. \quad (2.121b)$$

Since B is the standard basis, the coordinate 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 $\sum_{i=1}^3 \lambda_i \mathbf{b}_i = \tilde{\mathbf{b}}_j$, $j = 1, \dots, 3$.

1365 In Chapter 4, we will be able to exploit the concept of a basis change
1366 to find a basis with respect to which the transformation matrix of an endomorphism has a particularly simple (diagonal) form. In Chapter 10, we
1367 will look at a data compression problem and find a convenient basis onto
1368 which we can project the data while minimizing the compression loss.
1369

2.7.3 Image and Kernel

1371 The image and kernel of a linear mapping are vector subspaces with certain important properties. In the following, we will characterize them
1372 more carefully.
1373

1374 **Definition 2.22** (Image and Kernel).

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

$$\ker(\Phi) := \Phi^{-1}(\mathbf{0}_W) = \{\mathbf{v} \in V : \Phi(\mathbf{v}) = \mathbf{0}_W\} \quad (2.122)$$

and the *image/range*

$$\text{Im}(\Phi) := \Phi(V) = \{\mathbf{w} \in W | \exists \mathbf{v} \in V : \Phi(\mathbf{v}) = \mathbf{w}\}. \quad (2.123)$$

kernel
null space

image
range

domain
codomain

1375 We also call V and W also the *domain* and *codomain* of Φ , respectively.

1376 Intuitively, the kernel is the set of vectors in $\mathbf{v} \in V$ that Φ maps onto the neutral element $\mathbf{0}_W \in W$. The image is the set of vectors $\mathbf{w} \in W$ that

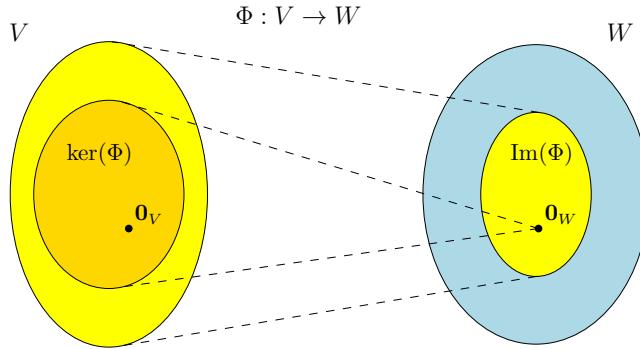


Figure 2.10 Kernel and Image of a linear mapping $\Phi : V \rightarrow W$.

can be “reached” by Φ from any vector in V . An illustration is given in Figure 2.10.

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

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

◇

Remark (Null Space and Column Space). Let us consider $\mathbf{A} \in \mathbb{R}^{m \times n}$ and a linear mapping $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $\mathbf{x} \mapsto \mathbf{Ax}$.

- For $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_n]$, where \mathbf{a}_i are the columns of \mathbf{A} , we obtain

$$\begin{aligned} \text{Im}(\Phi) &= \{\mathbf{Ax} : \mathbf{x} \in \mathbb{R}^n\} = \{x_1\mathbf{a}_1 + \dots + x_n\mathbf{a}_n : x_1, \dots, x_n \in \mathbb{R}\} \\ &\quad (2.124) \end{aligned}$$

$$= \text{span}[\mathbf{a}_1, \dots, \mathbf{a}_n] \subseteq \mathbb{R}^m, \quad (2.125)$$

i.e., the image is the span of the columns of \mathbf{A} , also called the *column space*. Therefore, the column space (image) is a subspace of \mathbb{R}^m , where m is the “height” of the matrix.

column space

- $\text{rk}(\mathbf{A}) = \dim(\text{Im}(\Phi))$
- The kernel/null space $\ker(\Phi)$ is the general solution to the linear homogeneous equation system $\mathbf{Ax} = \mathbf{0}$ and captures all possible linear combinations of the elements in \mathbb{R}^n that produce $\mathbf{0} \in \mathbb{R}^m$.
- The kernel is a subspace of \mathbb{R}^n , where n is the “width” of the matrix.
- The kernel focuses on the relationship among the columns, and we can use it to determine whether/how we can express a column as a linear combination of other columns.
- The purpose of the kernel is to determine whether a solution of the system of linear equations is unique and, if not, to capture all possible solutions.

1403


Example 2.24 (Image and Kernel of a Linear Mapping)

The mapping

$$\Phi : \mathbb{R}^4 \rightarrow \mathbb{R}^2, \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \mapsto \begin{bmatrix} 1 & 2 & -1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_1 + 2x_2 - x_3 \\ x_1 + x_4 \end{bmatrix} \quad (2.126)$$

$$= x_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (2.127)$$

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

$$\text{Im}(\Phi) = \text{span} \left[\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right]. \quad (2.128)$$

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

$$\begin{bmatrix} 1 & 2 & -1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \rightsquigarrow \dots \rightsquigarrow \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix}. \quad (2.129)$$

This matrix is in reduced row echelon form, and we can use the Minus-1 Trick to compute a basis of the kernel (see Section 2.3.3). Alternatively, we can express the non-pivot columns (columns 3 and 4) as linear combinations of the pivot-columns (columns 1 and 2). The third column \mathbf{a}_3 is equivalent to $-\frac{1}{2}$ times the second column \mathbf{a}_2 . Therefore, $\mathbf{0} = \mathbf{a}_3 + \frac{1}{2}\mathbf{a}_2$. In the same way, we see that $\mathbf{a}_4 = \mathbf{a}_1 - \frac{1}{2}\mathbf{a}_2$ and, therefore, $\mathbf{0} = \mathbf{a}_1 - \frac{1}{2}\mathbf{a}_2 - \mathbf{a}_4$. Overall, this gives us the kernel (null space) as

$$\ker(\Phi) = \text{span} \left[\begin{bmatrix} 0 \\ \frac{1}{2} \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ \frac{1}{2} \\ 0 \\ 1 \end{bmatrix} \right]. \quad (2.130)$$

Theorem 2.23 (Rank-Nullity Theorem). *For vector spaces V, W and a linear mapping $\Phi : V \rightarrow W$ it holds that*

$$\dim(\ker(\Phi)) + \dim(\text{Im}(\Phi)) = \dim(V). \quad (2.131)$$

1404

2.8 Affine Spaces

1405

In the following, we will have a closer look at spaces that are offset from the origin, i.e., spaces that are no longer vector subspaces. Moreover, we

1406

1407 will briefly discuss properties of mappings between these affine spaces,
1408 which resemble linear mappings.

1409 **2.8.1 Affine Subspaces**

Definition 2.24 (Affine Subspace). Let V be a vector space, $\mathbf{x}_0 \in V$ and $U \subseteq V$ a subspace. Then the subset

$$L = \mathbf{x}_0 + U := \{\mathbf{x}_0 + \mathbf{u} : \mathbf{u} \in U\} = \{\mathbf{v} \in V \mid \exists \mathbf{u} \in U : \mathbf{v} = \mathbf{x}_0 + \mathbf{u}\} \subseteq V \quad (2.132)$$

1410 is called *affine subspace* or *linear manifold* of V . U is called *direction* or
1411 *direction space*, and \mathbf{x}_0 is called *support point*. In Chapter 12, we refer to
1412 such a subspace as a *hyperplane*.

affine subspace
linear manifold
direction
direction space
support point
hyperplane

1413 Note that the definition of an affine subspace excludes $\mathbf{0}$ if $\mathbf{x}_0 \notin U$.
1414 Therefore, an affine subspace is not a (linear) subspace (vector subspace)
1415 of V for $\mathbf{x}_0 \notin U$.

parameters

1416 Examples of affine subspaces are points, lines and planes in \mathbb{R}^3 , which
1417 do not (necessarily) go through the origin.

1418 *Remark.* Consider two affine subspaces $L = \mathbf{x}_0 + U$ and $\tilde{L} = \tilde{\mathbf{x}}_0 + \tilde{U}$ of a
1419 vector space V . Then, $L \subseteq \tilde{L}$ if and only if $U \subseteq \tilde{U}$ and $\mathbf{x}_0 - \tilde{\mathbf{x}}_0 \in \tilde{U}$.

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

$$\mathbf{x} = \mathbf{x}_0 + \lambda_1 \mathbf{b}_1 + \dots + \lambda_k \mathbf{b}_k, \quad (2.133)$$

1420 where $\lambda_1, \dots, \lambda_k \in \mathbb{R}$. This representation is called *parametric equation*
1421 of L with directional vectors $\mathbf{b}_1, \dots, \mathbf{b}_k$ and *parameters* $\lambda_1, \dots, \lambda_k$. \diamond

parametric equation
parameters

Example 2.25 (Affine Subspaces)

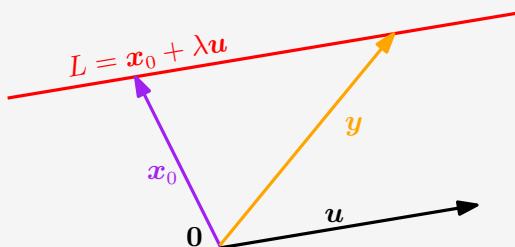


Figure 2.11 Vectors y on a line lie in an affine subspace L with support point x_0 and direction u .

- One-dimensional affine subspaces are called *lines* and can be written as $y = \mathbf{x}_0 + \lambda \mathbf{x}_1$, where $\lambda \in \mathbb{R}$, where $U = \text{span}[\mathbf{x}_1] \subseteq \mathbb{R}^n$ is a one-dimensional subspace of \mathbb{R}^n . This means, a line is defined by a support

lines

planes

point \mathbf{x}_0 and a vector \mathbf{x}_1 that defines the direction. See Figure 2.11 for an illustration.

- Two-dimensional affine subspaces of \mathbb{R}^n are called *planes*. The parametric equation for planes is $\mathbf{y} = \mathbf{x}_0 + \lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2$, where $\lambda_1, \lambda_2 \in \mathbb{R}$ and $U = [\mathbf{x}_1, \mathbf{x}_2] \subseteq \mathbb{R}^n$. This means, a plane is defined by a support point \mathbf{x}_0 and two linearly independent vectors $\mathbf{x}_1, \mathbf{x}_2$ that span the direction space.
- In \mathbb{R}^n , the $(n - 1)$ -dimensional affine subspaces are called *hyperplanes*, and the corresponding parametric equation is $\mathbf{y} = \mathbf{x}_0 + \sum_{i=1}^{n-1} \lambda_i \mathbf{x}_i$, where $\mathbf{x}_1, \dots, \mathbf{x}_{n-1}$ form a basis of an $(n - 1)$ -dimensional subspace U of \mathbb{R}^n . This means, a hyperplane is defined by a support point \mathbf{x}_0 and $(n - 1)$ linearly independent vectors $\mathbf{x}_1, \dots, \mathbf{x}_{n-1}$ that span the direction space. In \mathbb{R}^2 , a line is also a hyperplane. In \mathbb{R}^3 , a plane is also a hyperplane.

hyperplanes

1422 **Remark** (Inhomogeneous linear equation systems and affine subspaces).
1423 For $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$ the solution of the linear equation system
1424 $\mathbf{Ax} = \mathbf{b}$ is either the empty set or an affine subspace of \mathbb{R}^n of dimension
1425 $n - \text{rk}(\mathbf{A})$. In particular, the solution of the linear equation $\lambda_1 \mathbf{x}_1 + \dots +$
1426 $\lambda_n \mathbf{x}_n = \mathbf{b}$, where $(\lambda_1, \dots, \lambda_n) \neq (0, \dots, 0)$, is a hyperplane in \mathbb{R}^n .
1427 In \mathbb{R}^n , every k -dimensional affine subspace is the solution of a linear
1428 inhomogeneous equation system $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$ and
1429 $\text{rk}(\mathbf{A}) = n - k$. Recall that for homogeneous equation systems $\mathbf{Ax} = \mathbf{0}$
1430 the solution was a vector subspace, which we can also think of as a special
1431 affine space with support point $\mathbf{x}_0 = \mathbf{0}$. \diamond

1432

2.8.2 Affine Mappings

1433 Similar to linear mappings between vector spaces, which we discussed
1434 in Section 2.7, we can define affine mappings between two affine spaces.
1435 Linear and affine mappings are closely related. Therefore, many properties
1436 that we already know from linear mappings, e.g., that the composition of
1437 linear mappings is a linear mapping, also hold for affine mappings.

Definition 2.25 (Affine mapping). For two vector spaces V, W and a linear mapping $\Phi : V \rightarrow W$ and $\mathbf{a} \in W$ the mapping

$$\phi : V \rightarrow W \tag{2.134}$$

$$\mathbf{x} \mapsto \mathbf{a} + \Phi(\mathbf{x}) \tag{2.135}$$

affine mapping 1438 is an *affine mapping* from V to W . The vector \mathbf{a} is called the *translation*
 translation vector 1439 *vector* of ϕ .

- Every affine mapping $\phi : V \rightarrow W$ is also the composition of a linear

1441 mapping $\Phi : V \rightarrow W$ and a translation $\tau : W \rightarrow W$ in W , such that
1442 $\phi = \tau \circ \Phi$. The mappings Φ and τ are uniquely determined.

- 1443 • The composition $\phi' \circ \phi$ of affine mappings $\phi : V \rightarrow W$, $\phi' : W \rightarrow X$ is affine.
- 1445 • Affine mappings keep the geometric structure invariant. They also preserve the dimension and parallelism.

1447 Exercises

2.1 We consider $(\mathbb{R} \setminus \{-1\}, \star)$ where

$$a \star b := ab + a + b, \quad a, b \in \mathbb{R} \setminus \{-1\} \quad (2.136)$$

- 1448 1. Show that $(\mathbb{R} \setminus \{-1\}, \star)$ is an Abelian group
2. Solve

$$3 \star x \star x = 15$$

1449 in the Abelian group $(\mathbb{R} \setminus \{-1\}, \star)$, where \star is defined in (2.136).

2.2 Let n be in $\mathbb{N} \setminus \{0\}$. Let k, x be in \mathbb{Z} . We define the congruence class \bar{k} of the integer k as the set

$$\begin{aligned} \bar{k} &= \{x \in \mathbb{Z} \mid x - k = 0 \pmod{n}\} \\ &= \{x \in \mathbb{Z} \mid (\exists a \in \mathbb{Z}) : (x - k = n \cdot a)\}. \end{aligned}$$

We now define $\mathbb{Z}/n\mathbb{Z}$ (sometimes written \mathbb{Z}_n) as the set of all congruence classes modulo n . Euclidean division implies that this set is a finite set containing n elements:

$$\mathbb{Z}_n = \{\bar{0}, \bar{1}, \dots, \bar{n-1}\}$$

For all $\bar{a}, \bar{b} \in \mathbb{Z}_n$, we define

$$\bar{a} \oplus \bar{b} := \overline{a + b}$$

- 1450 1. Show that (\mathbb{Z}_n, \oplus) is a group. Is it Abelian?
2. We now define another operation \otimes for all \bar{a} and \bar{b} in \mathbb{Z}_n as

$$\bar{a} \otimes \bar{b} = \overline{a \times b} \quad (2.137)$$

1451 where $a \times b$ represents the usual multiplication in \mathbb{Z} .

1452 Let $n = 5$. Draw the times table of the elements of $\mathbb{Z}_5 \setminus \{\bar{0}\}$ under \otimes , i.e., calculate the products $\bar{a} \otimes \bar{b}$ for all \bar{a} and \bar{b} in $\mathbb{Z}_5 \setminus \{\bar{0}\}$.

1453 Hence, show that $\mathbb{Z}_5 \setminus \{\bar{0}\}$ is closed under \otimes and possesses a neutral element for \otimes . Display the inverse of all elements in $\mathbb{Z}_5 \setminus \{\bar{0}\}$ under \otimes . Conclude that $(\mathbb{Z}_5 \setminus \{\bar{0}\}, \otimes)$ is an Abelian group.

- 1454 3. Show that $(\mathbb{Z}_8 \setminus \{\bar{0}\}, \otimes)$ is not a group.
- 1455 4. We recall that Bézout theorem states that two integers a and b are relatively prime (i.e., $\gcd(a, b) = 1$, aka. coprime) if and only if there exist two integers u and v such that $au + bv = 1$. Show that $(\mathbb{Z}_n \setminus \{\bar{0}\}, \otimes)$ is a group if and only if $n \in \mathbb{N} \setminus \{0\}$ is prime.

2.3 Consider the set G of 3×3 matrices defined as:

$$G = \left\{ \begin{bmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix} \in \mathbb{R}^{3 \times 3} \mid x, y, z \in \mathbb{R} \right\} \quad (2.138)$$

¹⁴⁶² We define \cdot as the standard matrix multiplication.

¹⁴⁶³ Is (G, \cdot) a group? If yes, is it Abelian? Justify your answer.

¹⁴⁶⁴ 2.4 Compute the following matrix products:

1.

$$\begin{bmatrix} 1 & 2 \\ 4 & 5 \\ 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

2.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

3.

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

4.

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

5.

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

¹⁴⁶⁵ 2.5 Find the set S of all solutions in x of the following inhomogeneous linear systems $\mathbf{A}x = \mathbf{b}$ where \mathbf{A} and \mathbf{b} are defined below:

1.

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 2 & 5 & -7 & -5 \\ 2 & -1 & 1 & 3 \\ 5 & 2 & -4 & 2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ -2 \\ 4 \\ 6 \end{bmatrix}$$

2.

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 & 0 & 1 \\ 1 & 1 & 0 & -3 & 0 \\ 2 & -1 & 0 & 1 & -1 \\ -1 & 2 & 0 & -2 & -1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 3 \\ 6 \\ 5 \\ -1 \end{bmatrix}$$

3. Using Gaussian elimination find all solutions of the inhomogeneous equation system $\mathbf{A}\mathbf{x} = \mathbf{b}$ with

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$$

- 2.6 Find all solutions in $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \in \mathbb{R}^3$ of the equation system $\mathbf{A}\mathbf{x} = 12\mathbf{x}$,

where

$$\mathbf{A} = \begin{bmatrix} 6 & 4 & 3 \\ 6 & 0 & 9 \\ 0 & 8 & 0 \end{bmatrix}$$

and $\sum_{i=1}^3 x_i = 1$.

- 2.7 Determine the inverse of the following matrices if possible:

1.

$$\mathbf{A} = \begin{bmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \\ 4 & 5 & 6 \end{bmatrix}$$

2.

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

- 2.8 Which of the following sets are subspaces of \mathbb{R}^3 ?

1. $A = \{(\lambda, \lambda + \mu^3, \lambda - \mu^3) \mid \lambda, \mu \in \mathbb{R}\}$

2. $B = \{(\lambda^2, -\lambda^2, 0) \mid \lambda \in \mathbb{R}\}$

3. Let γ be in \mathbb{R} .

$C = \{(\xi_1, \xi_2, \xi_3) \in \mathbb{R}^3 \mid \xi_1 - 2\xi_2 + 3\xi_3 = \gamma\}$

4. $D = \{(\xi_1, \xi_2, \xi_3) \in \mathbb{R}^3 \mid \xi_2 \in \mathbb{Z}\}$

- 2.9 Are the following vectors linearly independent?

1.

$$\mathbf{x}_1 = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 3 \\ -3 \\ 8 \end{bmatrix}$$

2.

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

2.10 Write

$$\mathbf{y} = \begin{bmatrix} 1 \\ -2 \\ 5 \end{bmatrix}$$

as linear combination of

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$$

2.11 1. Determine a simple basis of U , where

$$U = \text{span} \left[\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 4 \\ 5 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \\ 5 \\ 3 \end{bmatrix} \right] \subseteq \mathbb{R}^4$$

2. Consider two subspaces of \mathbb{R}^4 :

$$U_1 = \text{span} \left[\begin{bmatrix} 1 \\ 1 \\ -3 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix} \right], \quad U_2 = \text{span} \left[\begin{bmatrix} -1 \\ -2 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 6 \\ -2 \\ -1 \end{bmatrix} \right].$$

¹⁴⁷⁶

Determine a basis of $U_1 \cap U_2$.

3. Consider two subspaces U_1 and U_2 , where U_1 is the solution space of the homogeneous equation system $\mathbf{A}_1 \mathbf{x} = \mathbf{0}$ and U_2 is the solution space of the homogeneous equation system $\mathbf{A}_2 \mathbf{x} = \mathbf{0}$ with

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & -2 & -1 \\ 2 & 1 & 3 \\ 1 & 0 & 1 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 3 & -3 & 0 \\ 1 & 2 & 3 \\ 7 & -5 & 2 \\ 3 & -1 & 2 \end{bmatrix}.$$

¹⁴⁷⁷

1. Determine the dimension of U_1, U_2

¹⁴⁷⁸

2. Determine bases of U_1 and U_2

¹⁴⁷⁹

3. Determine a basis of $U_1 \cap U_2$

2.12 Consider two subspaces U_1 and U_2 , where U_1 is spanned by the columns of \mathbf{A}_1 and U_2 is spanned by the columns of \mathbf{A}_2 with

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & -2 & -1 \\ 2 & 1 & 3 \\ 1 & 0 & 1 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 3 & -3 & 0 \\ 1 & 2 & 3 \\ 7 & -5 & 2 \\ 3 & -1 & 2 \end{bmatrix}.$$

¹⁴⁸⁰

1. Determine the dimension of U_1, U_2

¹⁴⁸¹

2. Determine bases of U_1 and U_2

¹⁴⁸²

3. Determine a basis of $U_1 \cap U_2$

¹⁴⁸³

2.13 Let $F = \{(x, y, z) \in \mathbb{R}^3 \mid x+y-z=0\}$ and $G = \{(a-b, a+b, a-3b) \mid a, b \in \mathbb{R}\}$.

¹⁴⁸⁴

1. Show that F and G are subspaces of \mathbb{R}^3 .

¹⁴⁸⁵

2. Calculate $F \cap G$ without resorting to any basis vector.

- 1486 3. Find one basis for F and one for G , calculate $F \cap G$ using the basis vectors
 1487 previously found and check your result with the previous question.

1488 2.14 Are the following mappings linear?

1. Let a and b be in \mathbb{R} .

$$\Phi : L^1([a, b]) \rightarrow \mathbb{R}$$

$$f \mapsto \Phi(f) = \int_a^b f(x) dx,$$

1489 where $L^1([a, b])$ denotes the set of integrable function on $[a, b]$.

2.

$$\Phi : C^1 \rightarrow C^0$$

$$f \mapsto \Phi(f) = f'.$$

1490 where for $k \geq 1$, C^k denotes the set of k times continuously differentiable
 1491 functions, and C^0 denotes the set of continuous functions.

3.

$$\Phi : \mathbb{R} \rightarrow \mathbb{R}$$

$$x \mapsto \Phi(x) = \cos(x)$$

4.

$$\Phi : \mathbb{R}^3 \rightarrow \mathbb{R}^2$$

$$\mathbf{x} \mapsto \begin{bmatrix} 1 & 2 & 3 \\ 1 & 4 & 3 \end{bmatrix} \mathbf{x}$$

5. Let θ be in $[0, 2\pi[$.

$$\Phi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$$

$$\mathbf{x} \mapsto \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \mathbf{x}$$

2.15 Consider the linear mapping

$$\Phi : \mathbb{R}^3 \rightarrow \mathbb{R}^4$$

$$\Phi \left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \right) = \begin{bmatrix} 3x_1 + 2x_2 + x_3 \\ x_1 + x_2 + x_3 \\ x_1 - 3x_2 \\ 2x_1 + 3x_2 + x_3 \end{bmatrix}$$

- 1492 • Find the transformation matrix A_Φ
- 1493 • Determine $\text{rk}(A_\Phi)$
- 1494 • Compute kernel and image of Φ . What is $\dim(\ker(\Phi))$ and $\dim(\text{Im}(\Phi))$?

1495 2.16 Let E be a vector space. Let f and g be two endomorphisms on E such that
 1496 $f \circ g = \text{id}_E$ (i.e. $f \circ g$ is the identity isomorphism). Show that $\ker f = \ker(g \circ f)$,
 1497 $\text{Img} = \text{Im}(g \circ f)$ and that $\ker(f) \cap \text{Im}(g) = \{\mathbf{0}_E\}$.

- 2.17 Consider an endomorphism $\Phi : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ whose transformation matrix (with respect to the standard basis in \mathbb{R}^3) is

$$\mathbf{A}_\Phi = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 1 & 1 & 1 \end{bmatrix}.$$

- ¹⁴⁹⁸ 1. Determine $\ker(\Phi)$ and $\text{Im}(\Phi)$.
 2. Determine the transformation matrix $\tilde{\mathbf{A}}_\Phi$ with respect to the basis

$$B = \left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right),$$

¹⁴⁹⁹ i.e., perform a basis change toward the new basis B .

- 2.18 Let us consider four vectors $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}'_1, \mathbf{b}'_2$ of \mathbb{R}^2 expressed in the standard basis of \mathbb{R}^2 as

$$\mathbf{b}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \mathbf{b}_2 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \quad \mathbf{b}'_1 = \begin{bmatrix} 2 \\ -2 \end{bmatrix}, \quad \mathbf{b}'_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (2.139)$$

¹⁵⁰⁰ and let us define $B = (\mathbf{b}_1, \mathbf{b}_2)$ and $B' = (\mathbf{b}'_1, \mathbf{b}'_2)$.

- ¹⁵⁰¹ 1. Show that B and B' are two bases of \mathbb{R}^2 and draw those basis vectors.
¹⁵⁰² 2. Compute the matrix \mathbf{P}_1 which performs a basis change from B' to B .

- ¹⁵⁰³ 2.19 We consider three vectors $\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3$ of \mathbb{R}^3 defined in the standard basis of \mathbb{R}
¹⁵⁰⁴ as

$$\mathbf{c}_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \quad \mathbf{c}_2 = \begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix}, \quad \mathbf{c}_3 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \quad (2.140)$$

¹⁵⁰⁵ and we define $C = (\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3)$.

- ¹⁵⁰⁶ 1. Show that C is a basis of \mathbb{R}^3 .
¹⁵⁰⁷ 2. Let us call $C' = (\mathbf{c}'_1, \mathbf{c}'_2, \mathbf{c}'_3)$ the standard basis of \mathbb{R}^3 . Explicit the matrix
¹⁵⁰⁸ \mathbf{P}_2 that performs the basis change from C to C' .

- 2.20 Let us consider $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}'_1, \mathbf{b}'_2$, 4 vectors of \mathbb{R}^2 expressed in the standard basis of \mathbb{R}^2 as

$$\mathbf{b}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \mathbf{b}_2 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \quad \mathbf{b}'_1 = \begin{bmatrix} 2 \\ -2 \end{bmatrix}, \quad \mathbf{b}'_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (2.141)$$

¹⁵⁰⁹ and let us define two ordered bases $B = (\mathbf{b}_1, \mathbf{b}_2)$ and $B' = (\mathbf{b}'_1, \mathbf{b}'_2)$ of \mathbb{R}^2 .

- ¹⁵¹⁰ 1. Show that B and B' are two bases of \mathbb{R}^2 and draw those basis vectors.
¹⁵¹¹ 2. Compute the matrix \mathbf{P}_1 that performs a basis change from B' to B .
¹⁵¹² 3. We consider $\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3$, 3 vectors of \mathbb{R}^3 defined in the standard basis of \mathbb{R}
 as

$$\mathbf{c}_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \quad \mathbf{c}_2 = \begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix}, \quad \mathbf{c}_3 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \quad (2.142)$$

¹⁵¹² and we define $C = (\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3)$.

- 1513 1. Show that C is a basis of \mathbb{R}^3 using determinants
 1514 2. Let us call $C' = (c'_1, c'_2, c'_3)$ the standard basis of \mathbb{R}^3 . Determine the
 1515 matrix P_2 that performs the basis change from C to C' .
4. We consider a homomorphism $\Phi : \mathbb{R}^2 \rightarrow \mathbb{R}^3$, such that

$$\begin{aligned}\Phi(b_1 + b_2) &= c_2 + c_3 \\ \Phi(b_1 - b_2) &= 2c_1 - c_2 + 3c_3\end{aligned}\tag{2.143}$$

1516 where $B = (b_1, b_2)$ and $C = (c_1, c_2, c_3)$ are ordered bases of \mathbb{R}^2 and \mathbb{R}^3 ,
 1517 respectively.

1518 Determine the transformation matrix A_Φ of Φ with respect to the ordered
 1519 bases B and C .

- 1520 5. Determine A' , the transformation matrix of Φ with respect to the bases
 1521 B' and C' .
 1522 6. Let us consider the vector $x \in \mathbb{R}^2$ whose coordinates in B' are $[2, 3]^\top$. In
 1523 other words, $x = 2b'_1 + 3b'_3$.
- 1524 1. Calculate the coordinates of x in B .
 1525 2. Based on that, compute the coordinates of $\Phi(x)$ expressed in C .
 1526 3. Then, write $\Phi(x)$ in terms of c'_1, c'_2, c'_3 .
 1527 4. Use the representation of x in B' and the matrix A' to find this result
 1528 directly.