

and use the *Moore-Penrose pseudo-inverse*  $(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top$  to determine the solution (2.59) that solves  $\mathbf{A}\mathbf{x} = \mathbf{b}$ , which also corresponds to the minimum norm least-squares solution. A disadvantage of this approach is that it requires many computations for the matrix-matrix product and computing the inverse of  $\mathbf{A}^\top \mathbf{A}$ . Moreover, for reasons of numerical precision it is generally not recommended to compute the inverse or pseudo-inverse. In the following, we therefore briefly discuss alternative approaches to solving systems of linear equations.

Moore-Penrose  
pseudo-inverse

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), computing the rank of a matrix (Section 2.6.2), and determining a basis of a vector space (Section 2.6.1). Gaussian elimination is an intuitive and constructive way to solve a system of linear equations with thousands of variables. However, for systems with millions of variables, it is impractical as the required number of arithmetic operations scales cubically in the number of simultaneous equations.

In practice, systems of many linear equations are solved indirectly, by either stationary iterative methods, such as the Richardson method, the Jacobi method, the Gauß-Seidel method, and the successive over-relaxation method, or Krylov subspace methods, such as conjugate gradients, generalized minimal residual, or biconjugate gradients. We refer to the books by Stoer and Burlirsch (2002), Strang (2003), and Liesen and Mehrmann (2015) for further details.

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

$$\mathbf{x}^{(k+1)} = \mathbf{C}\mathbf{x}^{(k)} + \mathbf{d} \quad (2.60)$$

for suitable  $\mathbf{C}$  and  $\mathbf{d}$  that reduces the residual error  $\|\mathbf{x}^{(k+1)} - \mathbf{x}_*\|$  in every iteration and 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 (Section 2.3). We saw that systems of linear equations can be compactly represented using matrix-vector notation (2.10). 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. 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 2.7 (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:

group

closure

associativity

neutral element

inverse element

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  $y \otimes x = e$ , where  $e$  is the neutral element. We often write  $x^{-1}$  to denote the inverse element of  $x$ .

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

Abelian group

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:

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

- $(\mathbb{Z}, +)$  is an Abelian group.
- $(\mathbb{N}_0, +)$  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.
- $(\mathbb{R} \setminus \{0\}, \cdot)$  is Abelian.
- $(\mathbb{R}^n, +), (\mathbb{Z}^n, +), n \in \mathbb{N}$  are Abelian if  $+$  is defined componentwise, i.e.,

$$(x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1 + y_1, \dots, x_n + y_n). \quad (2.61)$$

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 (2.61)).
- 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 (2.13).
  - 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 “ $\cdot$ ” in  $(\mathbb{R}^{n \times n}, \cdot)$ .

- Inverse element: If the inverse exists ( $\mathbf{A}$  is regular), then  $\mathbf{A}^{-1}$  is the inverse element of  $\mathbf{A} \in \mathbb{R}^{n \times n}$ , and in exactly this case  $(\mathbb{R}^{n \times n}, \cdot)$  is a group, called the *general linear group*.

**Definition 2.8** (General Linear Group). The set of regular (invertible) matrices  $\mathbf{A} \in \mathbb{R}^{n \times n}$  is a group with respect to matrix multiplication as defined in (2.13) and is called *general linear group*  $GL(n, \mathbb{R})$ . However, since matrix multiplication is not commutative, the group is not Abelian.

general linear group

### 2.4.2 Vector Spaces

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  $\mathbf{x} \in \mathcal{G}$  by a scalar  $\lambda \in \mathbb{R}$ . We can think of the inner operation as a form of addition, and the outer operation as a form of scaling. Note that the inner/outer operations have nothing to do with inner/outer products.

**Definition 2.9** (Vector Space). A real-valued *vector space*  $V = (\mathcal{V}, +, \cdot)$  is a set  $\mathcal{V}$  with two operations

vector space

$$+ : \mathcal{V} \times \mathcal{V} \rightarrow \mathcal{V} \quad (2.62)$$

$$\cdot : \mathbb{R} \times \mathcal{V} \rightarrow \mathcal{V} \quad (2.63)$$

where

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

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

vector

vector addition

scalar

multiplication by  
scalars

*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

outer product

(which we usually do), we can use the matrix multiplication as defined in (2.13). However, then the dimensions of the vectors do not match. Only the following multiplications for vectors are defined:  $\mathbf{a}\mathbf{b}^\top \in \mathbb{R}^{n \times n}$  (*outer product*),  $\mathbf{a}^\top \mathbf{b} \in \mathbb{R}$  (*inner/scalar/dot product*).  $\diamond$

### 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.

*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.  $\diamond$

*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*

column vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}. \quad (2.64)$$

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  $\mathbf{x}$ .  $\diamond$

row vector

transpose

### 2.4.3 Vector Subspaces

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 subspaces are a key idea in machine learning. For example, Chapter 10 demonstrates how to use vector subspaces for dimensionality reduction.

**Definition 2.10** (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  $V$ .

vector subspace  
linear subspace

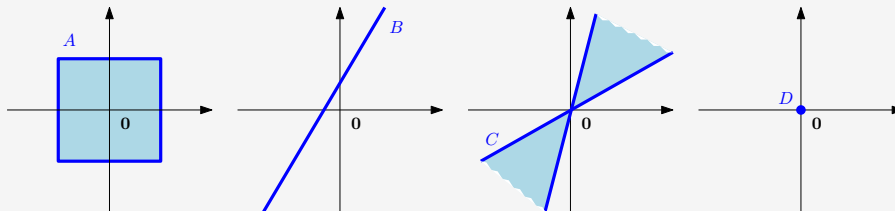
If  $\mathcal{U} \subseteq \mathcal{V}$  and  $V$  is a vector space, then  $U$  naturally inherits many properties directly from  $V$  because they hold for all  $\mathbf{x} \in \mathcal{V}$ , and in particular for all  $\mathbf{x} \in \mathcal{U} \subseteq \mathcal{V}$ . This includes the Abelian group 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

1.  $\mathcal{U} \neq \emptyset$ , in particular:  $\mathbf{0} \in \mathcal{U}$
2. Closure of  $U$ :
  - a. With respect to the outer operation:  $\forall \lambda \in \mathbb{R} \forall \mathbf{x} \in \mathcal{U} : \lambda \mathbf{x} \in \mathcal{U}$ .
  - b. 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 examples:

- For every vector space  $V$ , the trivial subspaces are  $V$  itself and  $\{\mathbf{0}\}$ .
- Only example  $D$  in Figure 2.6 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 system of linear equations  $A\mathbf{x} = \mathbf{0}$  with  $n$  unknowns  $\mathbf{x} = [x_1, \dots, x_n]^\top$  is a subspace of  $\mathbb{R}^n$ .
- The solution of an inhomogeneous system of linear equations  $A\mathbf{x} = \mathbf{b}$ ,  $\mathbf{b} \neq \mathbf{0}$  is not a subspace of  $\mathbb{R}^n$ .
- The intersection of arbitrarily many subspaces is a subspace itself.



**Figure 2.6** 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.

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

## 2.5 Linear Independence

In the following, we will have a close look at what we can do with vectors (elements of the vector space). In particular, we can add vectors together and multiply them with scalars. The closure property guarantees that we end up with another vector in the same vector space. It is possible to find a set of vectors with which we can represent every vector in the vector space by adding them together and scaling them. This set of vectors is a *basis*, and we will discuss them in Section 2.6.1. Before we get there, we will need to introduce the concepts of linear combinations and linear independence.

**Definition 2.11** (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

$$\mathbf{v} = \lambda_1 \mathbf{x}_1 + \dots + \lambda_k \mathbf{x}_k = \sum_{i=1}^k \lambda_i \mathbf{x}_i \in V \quad (2.65)$$

linear combination

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

The  $\mathbf{0}$ -vector can always be written as the linear combination of  $k$  vectors  $\mathbf{x}_1, \dots, \mathbf{x}_k$  because  $\mathbf{0} = \sum_{i=1}^k 0\mathbf{x}_i$  is always true. In the following, we are interested in non-trivial linear combinations of a set of vectors to represent  $\mathbf{0}$ , i.e., linear combinations of vectors  $\mathbf{x}_1, \dots, \mathbf{x}_k$ , where not all coefficients  $\lambda_i$  in (2.65) are 0.

**Definition 2.12** (Linear (In)dependence). Let us consider a vector space  $V$  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  $\mathbf{x}_1, \dots, \mathbf{x}_k$  are *linearly dependent*. If only the trivial solution exists, i.e.,  $\lambda_1 = \dots = \lambda_k = 0$  the vectors  $\mathbf{x}_1, \dots, \mathbf{x}_k$  are *linearly independent*.

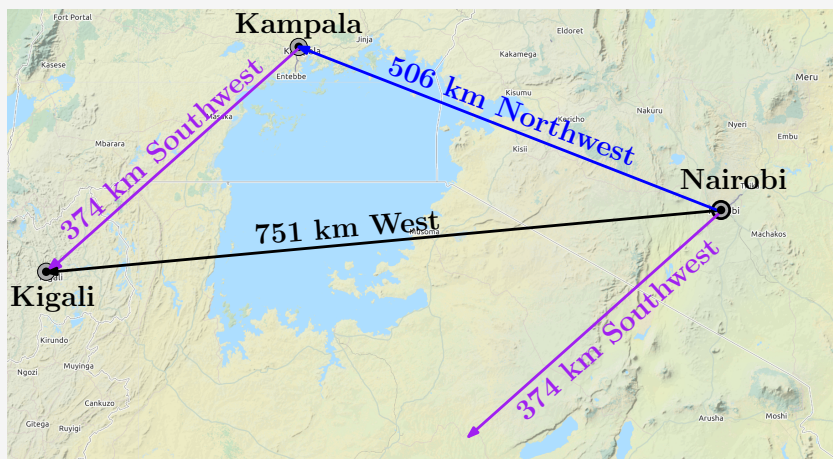
linearly dependent  
linearly  
independent

Linear independence is one of the most important concepts in linear algebra. Intuitively, a set of linearly independent vectors consists of 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.

### 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 curved 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 2.7 for an illustration). 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. Equivalently, given “751 km West” and “374 km Southwest” can be linearly combined to obtain “506 km Northwest”.



**Figure 2.7**  
Geographic example  
(with crude  
approximations to  
cardinal directions)  
of linearly  
dependent vectors  
in a  
two-dimensional  
space (plane).

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

- $k$  vectors are either linearly dependent or linearly independent. There is no third option.
- If at least one of the vectors  $\mathbf{x}_1, \dots, \mathbf{x}_k$  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 unnecessary 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

$$\begin{bmatrix} 1 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix} \quad (2.66)$$

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

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ -3 \\ 4 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 2 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} -1 \\ -2 \\ 1 \\ 1 \end{bmatrix}. \quad (2.67)$$

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

$$\lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2 + \lambda_3 \mathbf{x}_3 = \lambda_1 \begin{bmatrix} 1 \\ 2 \\ -3 \\ 4 \end{bmatrix} + \lambda_2 \begin{bmatrix} 1 \\ 1 \\ 0 \\ 2 \end{bmatrix} + \lambda_3 \begin{bmatrix} -1 \\ -2 \\ 1 \\ 1 \end{bmatrix} = \mathbf{0} \quad (2.68)$$

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:

$$\begin{bmatrix} 1 & 1 & -1 \\ 2 & 1 & -2 \\ -3 & 0 & 1 \\ 4 & 2 & 1 \end{bmatrix} \rightsquigarrow \dots \rightsquigarrow \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}. \quad (2.69)$$

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

$$\begin{aligned} \mathbf{x}_1 &= \sum_{i=1}^k \lambda_{i1} \mathbf{b}_i, \\ &\vdots \\ \mathbf{x}_m &= \sum_{i=1}^k \lambda_{im} \mathbf{b}_i. \end{aligned} \quad (2.70)$$

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

$$\mathbf{x}_j = \mathbf{B} \boldsymbol{\lambda}_j, \quad \boldsymbol{\lambda}_j = \begin{bmatrix} \lambda_{1j} \\ \vdots \\ \lambda_{kj} \end{bmatrix}, \quad j = 1, \dots, m, \quad (2.71)$$

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 (2.71), we obtain

$$\sum_{j=1}^m \psi_j \mathbf{x}_j = \sum_{j=1}^m \psi_j \mathbf{B} \boldsymbol{\lambda}_j = \mathbf{B} \sum_{j=1}^m \psi_j \boldsymbol{\lambda}_j. \quad (2.72)$$

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

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

### 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

$$\begin{aligned} \mathbf{x}_1 &= \mathbf{b}_1 & - & 2\mathbf{b}_2 & + & \mathbf{b}_3 & - & \mathbf{b}_4 \\ \mathbf{x}_2 &= -4\mathbf{b}_1 & - & 2\mathbf{b}_2 & & & + & 4\mathbf{b}_4 \\ \mathbf{x}_3 &= 2\mathbf{b}_1 & + & 3\mathbf{b}_2 & - & \mathbf{b}_3 & - & 3\mathbf{b}_4 \\ \mathbf{x}_4 &= 17\mathbf{b}_1 & - & 10\mathbf{b}_2 & + & 11\mathbf{b}_3 & + & \mathbf{b}_4 \end{aligned} \quad (2.73)$$

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

$$\left\{ \begin{bmatrix} 1 \\ -2 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} -4 \\ -2 \\ 0 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ -1 \\ -3 \end{bmatrix}, \begin{bmatrix} 17 \\ -10 \\ 11 \\ 1 \end{bmatrix} \right\} \quad (2.74)$$

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

$$\mathbf{A} = \begin{bmatrix} 1 & -4 & 2 & 17 \\ -2 & -2 & 3 & -10 \\ 1 & 0 & -1 & 11 \\ -1 & 4 & -3 & 1 \end{bmatrix} \quad (2.75)$$

is given as

$$\begin{bmatrix} 1 & 0 & 0 & -7 \\ 0 & 1 & 0 & -15 \\ 0 & 0 & 1 & -18 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \quad (2.76)$$

We see that the corresponding linear equation system is non-trivially solvable: The last column is not a pivot column, and  $x_4 = -7x_1 - 15x_2 - 18x_3$ . Therefore,  $x_1, \dots, x_4$  are linearly dependent as  $x_4$  can be expressed as a linear combination of  $x_1, \dots, x_3$ .

## 2.6 Basis and Rank

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

### 2.6.1 Generating Set and Basis

**Definition 2.13** (Generating Set and Span). Consider a vector space  $V = (\mathcal{V}, +, \cdot)$  and set of vectors  $\mathcal{A} = \{x_1, \dots, x_k\} \subseteq \mathcal{V}$ . If every vector  $v \in \mathcal{V}$  can be expressed as a linear combination of  $x_1, \dots, x_k$ ,  $\mathcal{A}$  is called a *generating set* of  $V$ . The set of all linear combinations of vectors in  $\mathcal{A}$  is called the *span* of  $\mathcal{A}$ . If  $\mathcal{A}$  spans the vector space  $V$ , we write  $V = \text{span}[\mathcal{A}]$  or  $V = \text{span}[x_1, \dots, x_k]$ .

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

**Definition 2.14** (Basis). Consider a vector space  $V = (\mathcal{V}, +, \cdot)$  and  $\mathcal{A} \subseteq \mathcal{V}$ . A generating set  $\mathcal{A}$  of  $V$  is called *minimal* if there exists no smaller set  $\tilde{\mathcal{A}} \subsetneq \mathcal{A} \subseteq \mathcal{V}$  that spans  $V$ . Every linearly independent generating set of  $V$  is minimal and is called a *basis* of  $V$ .

Let  $V = (\mathcal{V}, +, \cdot)$  be a vector space and  $\mathcal{B} \subseteq \mathcal{V}, \mathcal{B} \neq \emptyset$ . Then, the following statements are equivalent:

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

$$x = \sum_{i=1}^k \lambda_i b_i = \sum_{i=1}^k \psi_i b_i \quad (2.77)$$

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

A basis is a minimal generating set and a maximal linearly independent set of vectors.

### Example 2.16

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

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}. \quad (2.78)$$

canonical basis

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

$$\mathcal{B}_1 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}, \mathcal{B}_2 = \left\{ \begin{bmatrix} 0.5 \\ 0.8 \\ 0.4 \end{bmatrix}, \begin{bmatrix} 1.8 \\ 0.3 \\ 0.3 \end{bmatrix}, \begin{bmatrix} -2.2 \\ -1.3 \\ 3.5 \end{bmatrix} \right\}. \quad (2.79)$$

- The set

$$\mathcal{A} = \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ -4 \end{bmatrix} \right\} \quad (2.80)$$

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  $\mathcal{A}$ .

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



basis vector

We only consider finite-dimensional vector spaces  $V$ . In this case, the *dimension* of  $V$  is the number of basis vectors of  $V$ , and we write  $\dim(V)$ . If  $U \subseteq V$  is a subspace of  $V$ , then  $\dim(U) \leq \dim(V)$  and  $\dim(U) =$

dimension

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

$\dim(V)$  if and only if  $U = V$ . Intuitively, the dimension of a vector space can be thought of as the number of independent directions in this vector space.

*Remark.* The dimension of a vector space is not necessarily the number of elements in a vector. For instance, the vector space  $V = \text{span}\left[\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right]$  is one-dimensional, although the basis vector possesses two elements.  $\diamond$

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

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

$\diamond$

### Example 2.17 (Determining a Basis)

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

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \\ -1 \\ -1 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 2 \\ -1 \\ 1 \\ 2 \\ -2 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 3 \\ -4 \\ 3 \\ 5 \\ -3 \end{bmatrix}, \quad \mathbf{x}_4 = \begin{bmatrix} -1 \\ 8 \\ -5 \\ -6 \\ 1 \end{bmatrix} \in \mathbb{R}^5, \quad (2.81)$$

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

$$\sum_{i=1}^4 \lambda_i \mathbf{x}_i = \mathbf{0}, \quad (2.82)$$

which leads to a homogeneous system of equations with matrix

$$[\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4] = \begin{bmatrix} 1 & 2 & 3 & -1 \\ 2 & -1 & -4 & 8 \\ -1 & 1 & 3 & -5 \\ -1 & 2 & 5 & -6 \\ -1 & -2 & -3 & 1 \end{bmatrix}. \quad (2.83)$$

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

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

Since the pivot columns indicate which set of vectors is linearly independent, we see from the row-echelon form that  $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_4$  are linearly independent (because the system of linear equations  $\lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2 + \lambda_4 \mathbf{x}_4 = \mathbf{0}$  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

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

*Remark.* The rank of a matrix has some important properties:

- $\text{rk}(\mathbf{A}) = \text{rk}(\mathbf{A}^\top)$ , i.e., the column rank equals the row rank.
- The columns of  $\mathbf{A} \in \mathbb{R}^{m \times n}$  span a subspace  $U \subseteq \mathbb{R}^m$  with  $\dim(U) = \text{rk}(\mathbf{A})$ . Later we will call this subspace the *image* or *range*. A basis of  $U$  can be found by applying Gaussian elimination to  $\mathbf{A}$  to identify the pivot columns.
- The rows of  $\mathbf{A} \in \mathbb{R}^{m \times n}$  span a subspace  $W \subseteq \mathbb{R}^n$  with  $\dim(W) = \text{rk}(\mathbf{A})$ . A basis of  $W$  can be found by applying Gaussian elimination to  $\mathbf{A}^\top$ .
- For all  $\mathbf{A} \in \mathbb{R}^{n \times n}$  it holds that  $\mathbf{A}$  is regular (invertible) if and only if  $\text{rk}(\mathbf{A}) = n$ .
- For all  $\mathbf{A} \in \mathbb{R}^{m \times n}$  and all  $\mathbf{b} \in \mathbb{R}^m$  it holds that the linear equation system  $\mathbf{A}\mathbf{x} = \mathbf{b}$  can be solved if and only if  $\text{rk}(\mathbf{A}) = \text{rk}(\mathbf{A}|\mathbf{b})$ , where  $\mathbf{A}|\mathbf{b}$  denotes the augmented system.
- For  $\mathbf{A} \in \mathbb{R}^{m \times n}$  the subspace of solutions for  $\mathbf{A}\mathbf{x} = \mathbf{0}$  possesses dimension  $n - \text{rk}(\mathbf{A})$ . Later, we will call this subspace the *kernel* or the *null space*. kernel  
null space
- A matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$  has *full rank* if its rank equals the largest possible rank for a matrix of the same dimensions. This means that the rank of a full-rank matrix is the lesser of the number of rows and columns, i.e.,  $\text{rk}(\mathbf{A}) = \min(m, n)$ . A matrix is said to be *rank deficient* if it does not have full rank. full rank  
rank deficient



#### Example 2.18 (Rank)

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

$\mathbf{A}$  has two linearly independent rows/columns so that  $\text{rk}(\mathbf{A}) = 2$ .

$$\blacksquare \mathbf{A} = \begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix}.$$

We use Gaussian elimination to determine the rank:

$$\begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix} \rightsquigarrow \dots \rightsquigarrow \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix}. \quad (2.84)$$

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

## 2.7 Linear Mappings

In the following, we will study mappings on vector spaces that preserve their structure, which will allow us to define the concept of a coordinate. 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. We wish to preserve this property 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

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

$$\Phi(\lambda \mathbf{x}) = \lambda \Phi(\mathbf{x}) \quad (2.86)$$

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

**Definition 2.15** (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

$$\forall \mathbf{x}, \mathbf{y} \in V \forall \lambda, \psi \in \mathbb{R} : \Phi(\lambda \mathbf{x} + \psi \mathbf{y}) = \lambda \Phi(\mathbf{x}) + \psi \Phi(\mathbf{y}). \quad (2.87)$$

It turns out that we can represent linear mappings as matrices (Section 2.7.1). Recall that we can also collect a set of vectors as columns of a matrix. When working with matrices, we have to keep in mind what the matrix represents: a linear mapping or a collection of vectors. We will see more about linear mappings in Chapter 4. Before we continue, we will briefly introduce special mappings.

**Definition 2.16** (Injective, Surjective, Bijective). Consider a mapping  $\Phi : \mathcal{V} \rightarrow \mathcal{W}$ , where  $\mathcal{V}, \mathcal{W}$  can be arbitrary sets. Then  $\Phi$  is called

- *Injective* if  $\forall \mathbf{x}, \mathbf{y} \in \mathcal{V} : \Phi(\mathbf{x}) = \Phi(\mathbf{y}) \implies \mathbf{x} = \mathbf{y}$ .
- *Surjective* if  $\Phi(\mathcal{V}) = \mathcal{W}$ .
- *Bijective* if it is injective and surjective.

linear mapping  
vector space  
homomorphism  
linear  
transformation

injective  
surjective  
bijective

If  $\Phi$  is surjective, then every element in  $\mathcal{W}$  can be “reached” from  $\mathcal{V}$  using  $\Phi$ . A bijective  $\Phi$  can be “undone”, i.e., there exists a mapping  $\Psi : \mathcal{W} \rightarrow \mathcal{V}$  so that  $\Psi \circ \Phi(x) = x$ . This mapping  $\Psi$  is then called the inverse of  $\Phi$  and normally denoted by  $\Phi^{-1}$ .

With these definitions, we introduce the following special cases of linear mappings between vector spaces  $V$  and  $W$ :

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

isomorphism  
endomorphism  
automorphism  
  
identity mapping  
identity  
automorphism

### Example 2.19 (Homomorphism)

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

$$\begin{aligned} \Phi \left( \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \right) &= (x_1 + y_1) + i(x_2 + y_2) = x_1 + ix_2 + y_1 + iy_2 \\ &= \Phi \left( \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right) + \Phi \left( \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \right) \\ \Phi \left( \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right) &= \lambda x_1 + \lambda i x_2 = \lambda(x_1 + ix_2) = \lambda \Phi \left( \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right). \end{aligned} \quad (2.88)$$

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.

**Theorem 2.17** (Theorem 3.59 in Axler (2015)). *Finite-dimensional vector spaces  $V$  and  $W$  are isomorphic if and only if  $\dim(V) = \dim(W)$ .*

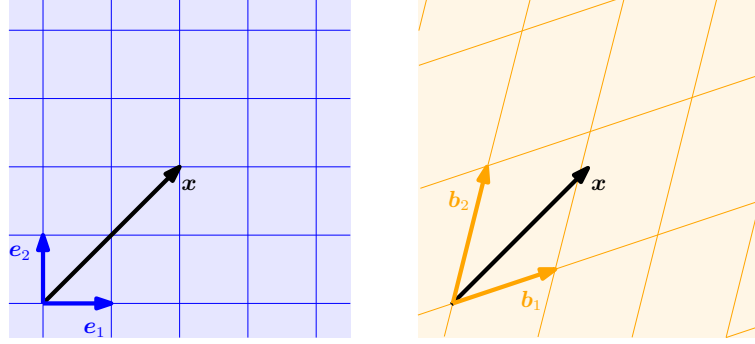
Theorem 2.17 states that there exists a linear, bijective mapping between two vector spaces of the same dimension. Intuitively, this means that vector spaces of the same dimension are kind of the same thing, as they can be transformed into each other without incurring any loss.

Theorem 2.17 also gives us the justification to treat  $\mathbb{R}^{m \times n}$  (the vector space of  $m \times n$ -matrices) and  $\mathbb{R}^{mn}$  (the vector space of vectors of length  $mn$ ) the same, as their dimensions are  $mn$ , and there exists a linear, bijective mapping that transforms one into the other.

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

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

**Figure 2.8** Two different coordinate systems defined by two sets of basis vectors. A vector  $x$  has different coordinate representations depending on which coordinate system is chosen.



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

◇

### 2.7.1 Matrix Representation of Linear Mappings

Any  $n$ -dimensional vector space is isomorphic to  $\mathbb{R}^n$  (Theorem 2.17). We consider a basis  $\{b_1, \dots, b_n\}$  of an  $n$ -dimensional vector space  $V$ . In the following, the order of the basis vectors will be important. Therefore, we write

$$B = (b_1, \dots, b_n) \quad (2.89)$$

ordered basis

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

*Remark* (Notation). We are at the point where notation gets a bit tricky. Therefore, we summarize some parts here.  $B = (b_1, \dots, b_n)$  is an ordered basis,  $\mathcal{B} = \{b_1, \dots, b_n\}$  is an (unordered) basis, and  $\mathbf{B} = [b_1, \dots, b_n]$  is a matrix whose columns are the vectors  $b_1, \dots, b_n$ . ◇

**Definition 2.18** (Coordinates). Consider a vector space  $V$  and an ordered basis  $B = (b_1, \dots, b_n)$  of  $V$ . For any  $x \in V$  we obtain a unique representation (linear combination)

$$x = \alpha_1 b_1 + \dots + \alpha_n b_n \quad (2.90)$$

coordinate

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

$$\alpha = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \in \mathbb{R}^n \quad (2.91)$$

coordinate vector  
coordinate  
representation

is the *coordinate vector/coordinate representation* of  $x$  with respect to the ordered basis  $B$ .



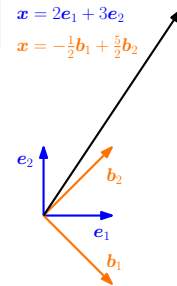
A basis effectively defines a coordinate system. We are familiar with the Cartesian coordinate system in two dimensions, which is spanned by the canonical basis vectors  $e_1, e_2$ . In this coordinate system, a vector  $x \in \mathbb{R}^2$  has a representation that tells us how to linearly combine  $e_1$  and  $e_2$  to obtain  $x$ . However, any basis of  $\mathbb{R}^2$  defines a valid coordinate system, and the same vector  $x$  from before may have a different coordinate representation in the  $(b_1, b_2)$  basis. In Figure 2.8, the coordinates of  $x$  with respect to the standard basis  $(e_1, e_2)$  is  $[2, 2]^\top$ . However, with respect to the basis  $(b_1, b_2)$  the same vector  $x$  is represented as  $[1.09, 0.72]^\top$ , i.e.,  $x = 1.09b_1 + 0.72b_2$ . In the following sections, we will discover how to obtain this representation.

### Example 2.20

Let us have a look at a geometric vector  $x \in \mathbb{R}^2$  with coordinates  $[2, 3]^\top$  with respect to the standard basis  $(e_1, e_2)$  of  $\mathbb{R}^2$ . This means, we can write  $x = 2e_1 + 3e_2$ . However, we do not have to choose the standard basis to represent this vector. If we use the basis vectors  $b_1 = [1, -1]^\top$ ,  $b_2 = [1, 1]^\top$  we will obtain the coordinates  $\frac{1}{2}[-1, 5]^\top$  to represent the same vector with respect to  $(b_1, b_2)$  (see Figure 2.9).

*Remark.* For an  $n$ -dimensional vector space  $V$  and an ordered basis  $B$  of  $V$ , the mapping  $\Phi : \mathbb{R}^n \rightarrow V$ ,  $\Phi(e_i) = b_i$ ,  $i = 1, \dots, n$ , is linear (and because of Theorem 2.17 an isomorphism), where  $(e_1, \dots, e_n)$  is the standard basis of  $\mathbb{R}^n$ .

**Figure 2.9**  
Different coordinate representations of a vector  $x$ , depending on the choice of basis.



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

**Definition 2.19** (Transformation Matrix). Consider vector spaces  $V, W$  with corresponding (ordered) bases  $B = (b_1, \dots, b_n)$  and  $C = (c_1, \dots, c_m)$ . Moreover, we consider a linear mapping  $\Phi : V \rightarrow W$ . For  $j \in \{1, \dots, n\}$ ,

$$\Phi(b_j) = \alpha_{1j}c_1 + \dots + \alpha_{mj}c_m = \sum_{i=1}^m \alpha_{ij}c_i \quad (2.92)$$

is the unique representation of  $\Phi(b_j)$  with respect to  $C$ . Then, we call the  $m \times n$ -matrix  $A_\Phi$ , whose elements are given by

$$A_\Phi(i, j) = \alpha_{ij}, \quad (2.93)$$

the *transformation matrix* of  $\Phi$  (with respect to the ordered bases  $B$  of  $V$  and  $C$  of  $W$ ).

transformation  
matrix

The coordinates of  $\Phi(b_j)$  with respect to the ordered basis  $C$  of  $W$  are the  $j$ -th column of  $A_\Phi$ . Consider (finite-dimensional) vector spaces  $V, W$  with ordered bases  $B, C$  and a linear mapping  $\Phi : V \rightarrow W$  with

transformation matrix  $A_\Phi$ . If  $\hat{x}$  is the coordinate vector of  $x \in V$  with respect to  $B$  and  $\hat{y}$  the coordinate vector of  $y = \Phi(x) \in W$  with respect to  $C$ , then

$$\hat{y} = A_\Phi \hat{x}. \quad (2.94)$$

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

### Example 2.21 (Transformation Matrix)

Consider a homomorphism  $\Phi : V \rightarrow W$  and ordered bases  $B = (b_1, \dots, b_3)$  of  $V$  and  $C = (c_1, \dots, c_4)$  of  $W$ . With

$$\begin{aligned} \Phi(b_1) &= c_1 - c_2 + 3c_3 - c_4 \\ \Phi(b_2) &= 2c_1 + c_2 + 7c_3 + 2c_4 \\ \Phi(b_3) &= 3c_2 + c_3 + 4c_4 \end{aligned} \quad (2.95)$$

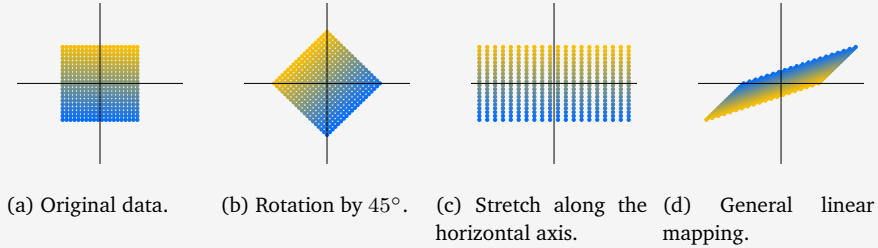
the transformation matrix  $A_\Phi$  with respect to  $B$  and  $C$  satisfies  $\Phi(b_k) = \sum_{i=1}^4 \alpha_{ik} c_i$  for  $k = 1, \dots, 3$  and is given as

$$A_\Phi = [\alpha_1, \alpha_2, \alpha_3] = \begin{bmatrix} 1 & 2 & 0 \\ -1 & 1 & 3 \\ 3 & 7 & 1 \\ -1 & 2 & 4 \end{bmatrix}, \quad (2.96)$$

where the  $\alpha_j$ ,  $j = 1, 2, 3$ , are the coordinate vectors of  $\Phi(b_j)$  with respect to  $C$ .

### Example 2.22 (Linear Transformations of Vectors)

**Figure 2.10** Three examples of linear transformations of the vectors shown as dots in (a); (b) Rotation by  $45^\circ$ ; (c) Stretching of the horizontal coordinates by 2; (d) Combination of reflection, rotation and stretching.



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

$$A_1 = \begin{bmatrix} \cos(\frac{\pi}{4}) & -\sin(\frac{\pi}{4}) \\ \sin(\frac{\pi}{4}) & \cos(\frac{\pi}{4}) \end{bmatrix}, \quad A_2 = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \quad A_3 = \frac{1}{2} \begin{bmatrix} 3 & -1 \\ 1 & -1 \end{bmatrix}. \quad (2.97)$$

Figure 2.10 gives three examples of linear transformations of a set of vectors. Figure 2.10(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  $A_1$  in (2.97) to linearly transform each of these vectors, we obtain the rotated square in Figure 2.10(b). If we apply the linear mapping represented by  $A_2$ , we obtain the rectangle in Figure 2.10(c) where each  $x_1$ -coordinate is stretched by 2. Figure 2.10(d) shows the original square from Figure 2.10(a) when linearly transformed using  $A_3$ , 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

$$B = (\mathbf{b}_1, \dots, \mathbf{b}_n), \quad \tilde{B} = (\tilde{\mathbf{b}}_1, \dots, \tilde{\mathbf{b}}_n) \quad (2.98)$$

of  $V$  and two ordered bases

$$C = (\mathbf{c}_1, \dots, \mathbf{c}_m), \quad \tilde{C} = (\tilde{\mathbf{c}}_1, \dots, \tilde{\mathbf{c}}_m) \quad (2.99)$$

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

*Remark.* We effectively get different coordinate representations of the identity mapping  $\text{id}_V$ . In the context of Figure 2.9, this would mean to map coordinates with respect to  $(\mathbf{e}_1, \mathbf{e}_2)$  onto coordinates with respect to  $(\mathbf{b}_1, \mathbf{b}_2)$  without changing the vector  $\mathbf{x}$ . By changing the basis and correspondingly the representation of vectors, the transformation matrix with respect to this new basis can have a particularly simple form that allows for straightforward computation.  $\diamond$

#### Example 2.23 (Basis Change)

Consider a transformation matrix

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \quad (2.100)$$

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

$$B = \left( \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right) \quad (2.101)$$

we obtain a diagonal transformation matrix

$$\tilde{\mathbf{A}} = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \quad (2.102)$$

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

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

**Theorem 2.20** (Basis Change). *For a linear mapping  $\Phi : V \rightarrow W$ , ordered bases*

$$B = (\mathbf{b}_1, \dots, \mathbf{b}_n), \quad \tilde{B} = (\tilde{\mathbf{b}}_1, \dots, \tilde{\mathbf{b}}_n) \quad (2.103)$$

of  $V$  and

$$C = (\mathbf{c}_1, \dots, \mathbf{c}_m), \quad \tilde{C} = (\tilde{\mathbf{c}}_1, \dots, \tilde{\mathbf{c}}_m) \quad (2.104)$$

of  $W$ , and a transformation matrix  $\mathbf{A}_\Phi$  of  $\Phi$  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

$$\tilde{\mathbf{A}}_\Phi = \mathbf{T}^{-1} \mathbf{A}_\Phi \mathbf{S}. \quad (2.105)$$

Here,  $\mathbf{S} \in \mathbb{R}^{n \times n}$  is the transformation matrix of  $\text{id}_V$  that maps coordinates with respect to  $\tilde{B}$  onto coordinates with respect to  $B$ , and  $\mathbf{T} \in \mathbb{R}^{m \times m}$  is the transformation matrix of  $\text{id}_W$  that maps coordinates with respect to  $\tilde{C}$  onto coordinates with respect to  $C$ .

*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

$$\tilde{\mathbf{b}}_j = s_{1j}\mathbf{b}_1 + \dots + s_{nj}\mathbf{b}_n = \sum_{i=1}^n s_{ij}\mathbf{b}_i, \quad j = 1, \dots, n. \quad (2.106)$$

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

$$\tilde{\mathbf{c}}_k = t_{1k}\mathbf{c}_1 + \dots + t_{mk}\mathbf{c}_m = \sum_{l=1}^m t_{lk}\mathbf{c}_l, \quad k = 1, \dots, m. \quad (2.107)$$

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  $\mathbf{S}$  is the coordinate representation of  $\tilde{\mathbf{b}}_j$  with respect to  $B$  and

the  $k$ th column of  $T$  is the coordinate representation of  $\tilde{c}_k$  with respect to  $C$ . Note that both  $S$  and  $T$  are regular.

We are going to look at  $\Phi(\tilde{b}_j)$  from two perspectives. First, applying the mapping  $\Phi$ , we get that for all  $j = 1, \dots, n$

$$\Phi(\tilde{b}_j) = \sum_{k=1}^m \underbrace{\tilde{a}_{kj} \tilde{c}_k}_{\in W} \stackrel{(2.107)}{=} \sum_{k=1}^m \tilde{a}_{kj} \sum_{l=1}^m t_{lk} c_l = \sum_{l=1}^m \left( \sum_{k=1}^m t_{lk} \tilde{a}_{kj} \right) c_l, \quad (2.108)$$

where we first expressed the new basis vectors  $\tilde{c}_k \in W$  as linear combinations of the basis vectors  $c_l \in W$  and then swapped the order of summation.

Alternatively, when we express the  $\tilde{b}_j \in V$  as linear combinations of  $b_i \in V$ , we arrive at

$$\Phi(\tilde{b}_j) \stackrel{(2.106)}{=} \Phi \left( \sum_{i=1}^n s_{ij} b_i \right) = \sum_{i=1}^n s_{ij} \Phi(b_i) = \sum_{i=1}^n s_{ij} \sum_{l=1}^m a_{li} c_l \quad (2.109a)$$

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

where we exploited the linearity of  $\Phi$ . Comparing (2.108) and (2.109b), 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,

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

such that

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

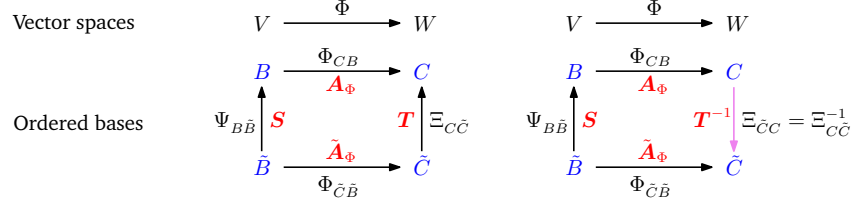
which proves Theorem 2.20.  $\square$

Theorem 2.20 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  $A_\Phi$  of a linear mapping  $\Phi : V \rightarrow W$  is replaced by an equivalent matrix  $\tilde{A}_\Phi$  with

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

Figure 2.11 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  $\Phi_{CB}$  is an instantiation of  $\Phi$  and maps basis vectors of  $B$  onto linear combinations of basis vectors of  $C$ . Assume that we know the transformation matrix  $A_\Phi$  of  $\Phi_{CB}$  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

**Figure 2.11** 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.



corresponding transformation matrix  $\tilde{A}_\Phi$  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  $A_\Phi$  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  $\Xi_{\tilde{C}C} : W \rightarrow W$  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:

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

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

equivalent

**Definition 2.21** (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

**Definition 2.22** (Similarity). Two matrices  $A, \tilde{A} \in \mathbb{R}^{n \times n}$  are *similar* if there exists a regular matrix  $S \in \mathbb{R}^{n \times n}$  with  $\tilde{A} = S^{-1}AS$ .

*Remark.* Similar matrices are always equivalent. However, equivalent matrices are not necessarily similar.  $\diamond$

*Remark.* Consider vector spaces  $V, W, X$ . From the remark that follows Theorem 2.17, we already know that for linear mappings  $\Phi : V \rightarrow W$  and  $\Psi : W \rightarrow X$  the mapping  $\Psi \circ \Phi : V \rightarrow X$  is also linear. With transformation matrices  $A_\Phi$  and  $A_\Psi$  of the corresponding mappings, the overall transformation matrix is  $A_{\Psi \circ \Phi} = A_\Psi A_\Phi$ .  $\diamond$

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

- $A_\Phi$  is the transformation matrix of a linear mapping  $\Phi_{CB} : V \rightarrow W$  with respect to the bases  $B, C$ .
- $\tilde{A}_\Phi$  is the transformation matrix of the linear mapping  $\Phi_{\tilde{C}\tilde{B}} : V \rightarrow W$  with respect to the bases  $\tilde{B}, \tilde{C}$ .
- $S$  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  $V$ .

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

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

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

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

### Example 2.24 (Basis Change)

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

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

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  $\tilde{A}_\Phi$  of  $\Phi$  with respect to the new bases

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

Then,

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

where the  $i$ th column of  $S$  is the coordinate representation of  $\tilde{b}_i$  in terms of the basis vectors of  $B$ . 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  $\lambda_i$  such that

$\sum_{i=1}^3 \lambda_i \mathbf{b}_i = \tilde{\mathbf{b}}_j, j = 1, \dots, 3$ . 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

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

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

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

### 2.7.3 Image and Kernel

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

**Definition 2.23** (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*

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

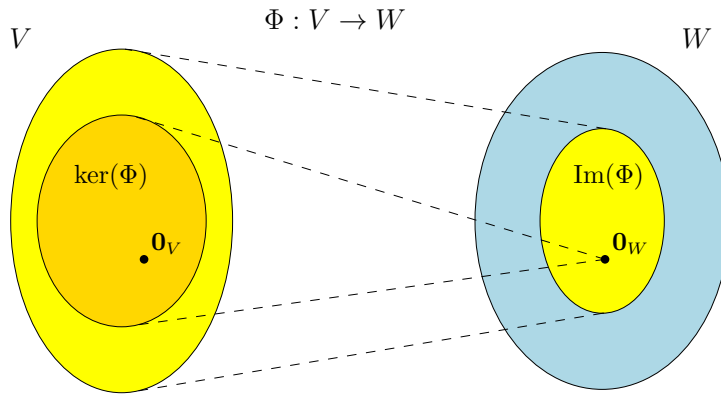
We also call  $V$  and  $W$  the *domain* and *codomain* of  $\Phi$ , respectively.

Intuitively, the kernel is the set of vectors  $\mathbf{v} \in V$  that  $\Phi$  maps onto the neutral element  $\mathbf{0}_W \in W$ . The image is the set of vectors  $\mathbf{w} \in W$  that can be “reached” by  $\Phi$  from any vector in  $V$ . An illustration is given in Figure 2.12.

*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  $V$ .





**Figure 2.12** Kernel and image of a linear mapping  $\Phi : V \rightarrow W$ .

- $\Phi$  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  $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$ .

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

$$\text{Im}(\Phi) = \{\mathbf{A}\mathbf{x} : \mathbf{x} \in \mathbb{R}^n\} = \left\{ \sum_{i=1}^n x_i \mathbf{a}_i : x_1, \dots, x_n \in \mathbb{R} \right\} \quad (2.124a)$$

$$= \text{span}[\mathbf{a}_1, \dots, \mathbf{a}_n] \subseteq \mathbb{R}^m, \quad (2.124b)$$

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.

- $\text{rk}(\mathbf{A}) = \dim(\text{Im}(\Phi))$ .
- The kernel/null space  $\ker(\Phi)$  is the general solution to the homogeneous system of linear equations  $\mathbf{A}\mathbf{x} = \mathbf{0}$  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.

◇

### Example 2.25 (Image and Kernel of a Linear Mapping)

The mapping

$$\Phi : \mathbb{R}^4 \rightarrow \mathbb{R}^2, \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \mapsto \begin{bmatrix} 1 & 2 & -1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_1 + 2x_2 - x_3 \\ x_1 + x_4 \end{bmatrix} \quad (2.125a)$$

$$= x_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (2.125b)$$

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

$$\text{Im}(\Phi) = \text{span}\left[\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right]. \quad (2.126)$$

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

$$\begin{bmatrix} 1 & 2 & -1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \rightsquigarrow \cdots \rightsquigarrow \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix}. \quad (2.127)$$

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

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

rank-nullity  
theorem

**Theorem 2.24** (Rank-Nullity Theorem). *For vector spaces  $V, W$  and a linear mapping  $\Phi : V \rightarrow W$  it holds that*

$$\dim(\ker(\Phi)) + \dim(\text{Im}(\Phi)) = \dim(V). \quad (2.129)$$

fundamental  
theorem of linear  
mappings

The rank-nullity theorem is also referred to as the *fundamental theorem of linear mappings* (Axler, 2015, theorem 3.22). The following are direct consequences of Theorem 2.24:

- If  $\dim(\text{Im}(\Phi)) < \dim(V)$ , then  $\ker(\Phi)$  is non-trivial, i.e., the kernel contains more than  $\mathbf{0}_V$  and  $\dim(\ker(\Phi)) \geq 1$ .
- If  $\mathbf{A}_\Phi$  is the transformation matrix of  $\Phi$  with respect to an ordered basis and  $\dim(\text{Im}(\Phi)) < \dim(V)$ , then the system of linear equations  $\mathbf{A}_\Phi \mathbf{x} = \mathbf{0}$  has infinitely many solutions.
- If  $\dim(V) = \dim(W)$ , then the following three-way equivalence holds:
  - $\Phi$  is injective
  - $\Phi$  is surjective
  - $\Phi$  is bijective
 since  $\text{Im}(\Phi) \subseteq W$ .

## 2.8 Affine Spaces

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

*Remark.* In the machine learning literature, the distinction between linear and affine is sometimes not clear so that we can find references to affine spaces/mappings as linear spaces/mappings.  $\diamond$

### 2.8.1 Affine Subspaces

**Definition 2.25** (Affine Subspace). Let  $V$  be a vector space,  $\mathbf{x}_0 \in V$  and  $U \subseteq V$  a subspace. Then the subset

$$L = \mathbf{x}_0 + U := \{\mathbf{x}_0 + \mathbf{u} : \mathbf{u} \in U\} \quad (2.130a)$$

$$= \{\mathbf{v} \in V \mid \exists \mathbf{u} \in U : \mathbf{v} = \mathbf{x}_0 + \mathbf{u}\} \subseteq V \quad (2.130b)$$

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

affine subspace  
linear manifold  
direction  
direction space  
support point  
hyperplane

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

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

*Remark.* Consider two affine subspaces  $L = \mathbf{x}_0 + U$  and  $\tilde{L} = \tilde{\mathbf{x}}_0 + \tilde{U}$  of a 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

$$\mathbf{x} = \mathbf{x}_0 + \lambda_1 \mathbf{b}_1 + \dots + \lambda_k \mathbf{b}_k, \quad (2.131)$$

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

parametric equation  
parameters

#### Example 2.26 (Affine Subspaces)

- One-dimensional affine subspaces are called *lines* and can be written as  $\mathbf{y} = \mathbf{x}_0 + \lambda \mathbf{b}_1$ , where  $\lambda \in \mathbb{R}$  and  $U = \text{span}[\mathbf{b}_1] \subseteq \mathbb{R}^n$  is a one-dimensional subspace of  $\mathbb{R}^n$ . This means that a line is defined by a support point  $\mathbf{x}_0$  and a vector  $\mathbf{b}_1$  that defines the direction. See Figure 2.13 for an illustration.

line

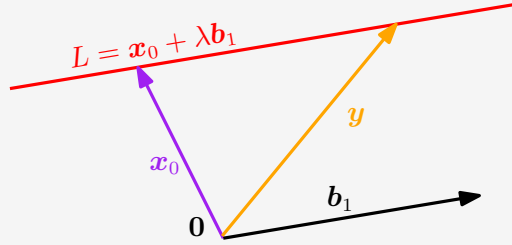
plane

- Two-dimensional affine subspaces of  $\mathbb{R}^n$  are called *planes*. The parametric equation for planes is  $\mathbf{y} = \mathbf{x}_0 + \lambda_1 \mathbf{b}_1 + \lambda_2 \mathbf{b}_2$ , where  $\lambda_1, \lambda_2 \in \mathbb{R}$  and  $U = \text{span}[\mathbf{b}_1, \mathbf{b}_2] \subseteq \mathbb{R}^n$ . This means that a plane is defined by a support point  $\mathbf{x}_0$  and two linearly independent vectors  $\mathbf{b}_1, \mathbf{b}_2$  that span the direction space.

hyperplane

- In  $\mathbb{R}^n$ , the  $(n - 1)$ -dimensional affine subspaces are called *hyperplanes*, and the corresponding parametric equation is  $\mathbf{y} = \mathbf{x}_0 + \sum_{i=1}^{n-1} \lambda_i \mathbf{b}_i$ , where  $\mathbf{b}_1, \dots, \mathbf{b}_{n-1}$  form a basis of an  $(n - 1)$ -dimensional subspace  $U$  of  $\mathbb{R}^n$ . This means that a hyperplane is defined by a support point  $\mathbf{x}_0$  and  $(n - 1)$  linearly independent vectors  $\mathbf{b}_1, \dots, \mathbf{b}_{n-1}$  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.

**Figure 2.13** Lines are affine subspaces. Vectors  $\mathbf{y}$  on a line  $\mathbf{x}_0 + \lambda \mathbf{b}_1$  lie in an affine subspace  $L$  with support point  $\mathbf{x}_0$  and direction  $\mathbf{b}_1$ .



*Remark* (Inhomogeneous systems of linear equations and affine subspaces). For  $\mathbf{A} \in \mathbb{R}^{m \times n}$  and  $\mathbf{x} \in \mathbb{R}^m$ , the solution of the system of linear equations  $\mathbf{A}\boldsymbol{\lambda} = \mathbf{x}$  is either the empty set or an affine subspace of  $\mathbb{R}^n$  of dimension  $n - \text{rk}(\mathbf{A})$ . In particular, the solution of the linear equation  $\lambda_1 \mathbf{b}_1 + \dots + \lambda_n \mathbf{b}_n = \mathbf{x}$ , where  $(\lambda_1, \dots, \lambda_n) \neq (0, \dots, 0)$ , is a hyperplane in  $\mathbb{R}^n$ .

In  $\mathbb{R}^n$ , every  $k$ -dimensional affine subspace is the solution of an inhomogeneous system of linear equations  $\mathbf{A}\mathbf{x} = \mathbf{b}$ , where  $\mathbf{A} \in \mathbb{R}^{m \times n}$ ,  $\mathbf{b} \in \mathbb{R}^m$  and  $\text{rk}(\mathbf{A}) = n - k$ . Recall that for homogeneous equation systems  $\mathbf{A}\mathbf{x} = \mathbf{0}$  the solution was a vector subspace, which we can also think of as a special affine space with support point  $\mathbf{x}_0 = \mathbf{0}$ .  $\diamond$

### 2.8.2 Affine Mappings

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

**Definition 2.26** (Affine Mapping). For two vector spaces  $V, W$ , a linear

mapping  $\Phi : V \rightarrow W$ , and  $\mathbf{a} \in W$ , the mapping

$$\phi : V \rightarrow W \quad (2.132)$$

$$\mathbf{x} \mapsto \mathbf{a} + \Phi(\mathbf{x}) \quad (2.133)$$

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

affine mapping  
translation vector

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

## 2.9 Further Reading

There are many resources for learning linear algebra, including the textbooks by Strang (2003), Golan (2007), Axler (2015), and Liesen and Mehrmann (2015). There are also several online resources that we mentioned in the introduction to this chapter. We only covered Gaussian elimination here, but there are many other approaches for solving systems of linear equations, and we refer to numerical linear algebra textbooks by Stoer and Burlirsch (2002), Golub and Van Loan (2012), and Horn and Johnson (2013) for an in-depth discussion.

In this book, we distinguish between the topics of linear algebra (e.g., vectors, matrices, linear independence, basis) and topics related to the geometry of a vector space. In Chapter 3, we will introduce the inner product, which induces a norm. These concepts allow us to define angles, lengths and distances, which we will use for orthogonal projections. Projections turn out to be key in many machine learning algorithms, such as linear regression and principal component analysis, both of which we will cover in Chapters 9 and 10, respectively.

### Exercises

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

$$a \star b := ab + a + b, \quad a, b \in \mathbb{R} \setminus \{-1\} \quad (2.134)$$

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

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

in the Abelian group  $(\mathbb{R} \setminus \{-1\}, \star)$ , where  $\star$  is defined in (2.134).

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:

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

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

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

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

$$\bar{a} \otimes \bar{b} = \overline{a \times b}, \quad (2.135)$$

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

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  $\mathbb{Z}_5 \setminus \{\bar{0}\}$ .

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.

- Show that  $(\mathbb{Z}_8 \setminus \{\bar{0}\}, \otimes)$  is not a group.
- We recall that the Bézout theorem states that two integers  $a$  and  $b$  are relatively prime (i.e.,  $\gcd(a, b) = 1$ ) 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  $\mathcal{G}$  of  $3 \times 3$  matrices defined as follows:

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

We define  $\cdot$  as the standard matrix multiplication.

Is  $(\mathcal{G}, \cdot)$  a group? If yes, is it Abelian? Justify your answer.

2.4 Compute the following matrix products, if possible: