

# Chapter 4

## Matrices

In this chapter we define matrices with their most important operations and we study several groups and rings of matrices. James Joseph Sylvester (1814–1897) coined the term *matrix*<sup>1</sup> in 1850 and described matrices as “an oblong arrangement of terms”. The matrix operations defined in this chapter were introduced by Arthur Cayley (1821–1895) in 1858. His article “A memoir on the theory of matrices” was the first to consider matrices as independent algebraic objects. In our book matrices form the central approach to the theory of Linear Algebra.

### 4.1 Basic Definitions and Operations

We begin with a formal definition of matrices.

**Definition 4.1** Let  $R$  be a commutative ring with unit and let  $n, m \in \mathbb{N}_0$ . An array of the form

$$A = [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}$$

---

<sup>1</sup>The Latin word “matrix” means “womb”. Sylvester considered matrices as objects “out of which we may form various systems of determinants” (cp. Chap. 5). Interestingly, the English writer Charles Lutwidge Dodgson (1832–1898), better known by his pen name Lewis Carroll, objected to Sylvester’s term and wrote in 1867: “I am aware that the word ‘Matrix’ is already in use to express the very meaning for which I use the word ‘Block’; but surely the former word means rather the mould, or form, into which algebraic quantities may be introduced, than an actual assemblage of such quantities”. Dodgson also objected to the notation  $a_{ij}$  for the matrix entries: “...most of the space is occupied by a number of  $a$ ’s, which are wholly superfluous, while the only important part of the notation is reduced to minute subscripts, alike difficult to the writer and the reader.”

with  $a_{ij} \in R, i = 1, \dots, n, j = 1, \dots, m$ , is called a *matrix* of size  $n \times m$  over  $R$ . The  $a_{ij}$  are called the *entries* or *coefficients* of the matrix. The set of all such matrices is denoted by  $R^{n,m}$ .

In the following we usually assume (without explicitly mentioning it) that  $1 \neq 0$  in  $R$ . This excludes the trivial case of the ring that contains only the zero element (cp. Exercise 3.11).

Formally, in Definition 4.1 for  $n = 0$  or  $m = 0$  we obtain “empty matrices” of the size  $0 \times m, n \times 0$  or  $0 \times 0$ . We denote such matrices by  $[ ]$ . They will be used for technical reasons in some of the proofs below. When we analyze algebraic properties of matrices, however, we always consider  $n, m \geq 1$ .

The *zero matrix* in  $R^{n,m}$ , denoted by  $0_{n,m}$  or just  $0$ , is the matrix that has all its entries equal to  $0 \in R$ .

A matrix of size  $n \times n$  is called a *square matrix* or just *square*. The entries  $a_{ii}$  for  $i = 1, \dots, n$  are called the *diagonal entries* of  $A$ . The *identity matrix* in  $R^{n,n}$  is the matrix  $I_n := [\delta_{ij}]$ , where

$$\delta_{ij} := \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases} \quad (4.1)$$

is the *Kronecker delta-function*.<sup>2</sup> If it is clear which  $n$  is considered, then we just write  $I$  instead of  $I_n$ . For  $n = 0$  we set  $I_0 := [ ]$ .

The  $i$ th *row* of  $A \in R^{n,m}$  is  $[a_{i1}, a_{i2}, \dots, a_{im}] \in R^{1,m}, i = 1, \dots, n$ , where we use commas for the optical separation of the entries. The  $j$ th *column* of  $A$  is

$$\begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{bmatrix} \in R^{n,1}, \quad j = 1, \dots, m.$$

Thus, the rows and columns of a matrix are again matrices.

If  $1 \times m$  matrices  $a_i := [a_{i1}, a_{i2}, \dots, a_{im}] \in R^{1,m}, i = 1, \dots, n$ , are given, then we can combine them to the matrix

$$A = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \in R^{n,m}.$$

---

<sup>2</sup>Leopold Kronecker (1823–1891).

We then do not write square brackets around the rows of  $A$ . In the same way we can combine the  $n \times 1$  matrices

$$a_j := \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{bmatrix} \in R^{n,1}, \quad j = 1, \dots, m,$$

to the matrix

$$A = [a_1, a_2, \dots, a_m] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \in R^{n,m}.$$

If  $n_1, n_2, m_1, m_2 \in \mathbb{N}_0$  and  $A_{ij} \in R^{n_i, m_j}$ ,  $i, j = 1, 2$ , then we can combine these four matrices to the matrix

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \in R^{n_1+n_2, m_1+m_2}.$$

The matrices  $A_{ij}$  are then called *blocks* of the *block matrix*  $A$ .

We now introduce four operations for matrices and begin with the *addition*:

$$+ : R^{n,m} \times R^{n,m} \rightarrow R^{n,m}, \quad (A, B) \mapsto A + B := [a_{ij} + b_{ij}].$$

The addition in  $R^{n,m}$  operates *entrywise*, based on the addition in  $R$ . Note that the addition is only defined for matrices of equal size.

The *multiplication* of two matrices is defined as follows:

$$* : R^{n,m} \times R^{m,s} \rightarrow R^{n,s}, \quad (A, B) \mapsto A * B = [c_{ij}], \quad c_{ij} := \sum_{k=1}^m a_{ik} b_{kj}.$$

Thus, the entry  $c_{ij}$  of the product  $A * B$  is constructed by successive multiplication and summing up the entries in the  $i$ th row of  $A$  and the  $j$ th column of  $B$ . Clearly, in order to define the product  $A * B$ , the number of columns of  $A$  must be equal to the number of rows in  $B$ .

In the definition of the entries  $c_{ij}$  of the matrix  $A * B$  we have not written the multiplication symbol for the elements in  $R$ . This follows the usual convention of omitting the multiplication sign when it is clear which multiplication is considered. Eventually we will also omit the multiplication sign between matrices.

We can illustrate the multiplication rule “ $c_{ij}$  equals  $i$ th row of  $A$  times  $j$ th column of  $B$ ” as follows:

$$\begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & & \vdots \\ [ a_{i1} & \cdots & a_{im} ] \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix} \begin{bmatrix} b_{11} & \cdots & [ b_{1j} ] & \cdots & b_{1s} \\ \vdots & & \vdots & & \vdots \\ b_{m1} & \cdots & [ b_{mj} ] & \cdots & b_{ms} \end{bmatrix} \rightarrow \begin{bmatrix} \downarrow \\ c_{ij} \end{bmatrix}$$

It is important to note that the matrix multiplication in general is not commutative.

*Example 4.2* For the matrices

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \in \mathbb{Z}^{2,3}, \quad B = \begin{bmatrix} -1 & 1 \\ 0 & 0 \\ 1 & -1 \end{bmatrix} \in \mathbb{Z}^{3,2}$$

we have

$$A * B = \begin{bmatrix} 2 & -2 \\ 2 & -2 \end{bmatrix} \in \mathbb{Z}^{2,2}.$$

On the other hand,  $B * A \in \mathbb{Z}^{3,3}$ . Although  $A * B$  and  $B * A$  are both defined, we obviously have  $A * B \neq B * A$ . In this case one recognizes the non-commutativity of the matrix multiplication from the fact that  $A * B$  and  $B * A$  have different sizes. But even if  $A * B$  and  $B * A$  are both defined and have the same size, in general  $A * B \neq B * A$ . For example,

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \in \mathbb{Z}^{2,2}, \quad B = \begin{bmatrix} 4 & 0 \\ 5 & 6 \end{bmatrix} \in \mathbb{Z}^{2,2}$$

yield the two products

$$A * B = \begin{bmatrix} 14 & 12 \\ 15 & 18 \end{bmatrix} \quad \text{and} \quad B * A = \begin{bmatrix} 4 & 8 \\ 5 & 28 \end{bmatrix}.$$

The matrix multiplication is, however, associative and distributive with respect to the matrix addition.

**Lemma 4.3** For  $A, \widehat{A} \in R^{n,m}$ ,  $B, \widehat{B} \in R^{m,\ell}$  and  $C \in R^{\ell,k}$  the following assertions hold:

- (1)  $A * (B * C) = (A * B) * C$ .
- (2)  $(A + \widehat{A}) * B = A * B + \widehat{A} * B$ .
- (3)  $A * (B + \widehat{B}) = A * B + A * \widehat{B}$ .
- (4)  $I_n * A = A * I_m = A$ .

*Proof* We only show property (1); the others are exercises. Let  $A \in R^{n,m}$ ,  $B \in R^{m,\ell}$ ,  $C \in R^{\ell,k}$  as well as  $(A * B) * C = [d_{ij}]$  and  $A * (B * C) = [\widehat{d}_{ij}]$ . By the definition of the matrix multiplication and using the associative and distributive law in  $R$ , we get

$$\begin{aligned} d_{ij} &= \sum_{s=1}^{\ell} \left( \sum_{t=1}^m a_{it} b_{ts} \right) c_{sj} = \sum_{s=1}^{\ell} \sum_{t=1}^m (a_{it} b_{ts}) c_{sj} = \sum_{s=1}^{\ell} \sum_{t=1}^m a_{it} (b_{ts} c_{sj}) \\ &= \sum_{t=1}^m a_{it} \left( \sum_{s=1}^{\ell} b_{ts} c_{sj} \right) = \widehat{d}_{ij}, \end{aligned}$$

for  $1 \leq i \leq n$  and  $1 \leq j \leq k$ , which implies that  $(A * B) * C = A * (B * C)$ .  $\square$

On the right hand sides of (2) and (3) in Lemma 4.3 we have not written parentheses, since we will use the common convention that the multiplication of matrices binds stronger than the addition.

For  $A \in R^{n,n}$  we define

$$\begin{aligned} A^k &:= \underbrace{A * \dots * A}_{k \text{ times}} \quad \text{for } k \in \mathbb{N}, \\ A^0 &:= I_n. \end{aligned}$$

Another multiplicative operation for matrices is the multiplication with a *scalar*,<sup>3</sup> which is defined as follows:

$$\cdot : R \times R^{n,m} \rightarrow R^{n,m}, \quad (\lambda, A) \mapsto \lambda \cdot A := [\lambda a_{ij}]. \quad (4.2)$$

We easily see that  $0 \cdot A = 0_{n,m}$  and  $1 \cdot A = A$  for all  $A \in R^{n,m}$ . In addition, the scalar multiplication has the following properties.

**Lemma 4.4** For  $A, B \in R^{n,m}$ ,  $C \in R^{m,\ell}$  and  $\lambda, \mu \in R$  the following assertions hold:

- (1)  $(\lambda\mu) \cdot A = \lambda \cdot (\mu \cdot A)$ .
- (2)  $(\lambda + \mu) \cdot A = \lambda \cdot A + \mu \cdot A$ .

<sup>3</sup>The term “scalar” was introduced in 1845 by Sir William Rowan Hamilton (1805–1865). It originates from the Latin word “scale” which means “ladder”.

- (3)  $\lambda \cdot (A + B) = \lambda \cdot A + \lambda \cdot B$ .  
 (4)  $(\lambda \cdot A) * C = \lambda \cdot (A * C) = A * (\lambda \cdot C)$ .

*Proof* Exercise. □

The fourth matrix operation that we introduce is the *transposition*:

$$T : R^{n,m} \rightarrow R^{m,n}, \quad A = [a_{ij}] \mapsto A^T = [b_{ij}], \quad b_{ij} := a_{ji}.$$

For example,

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \in \mathbb{Z}^{2,3}, \quad A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \in \mathbb{Z}^{3,2}.$$

The matrix  $A^T$  is called the *transpose* of  $A$ .

**Definition 4.5** If  $A \in R^{n,n}$  satisfies  $A = A^T$ , then  $A$  is called *symmetric*. If  $A = -A^T$ , then  $A$  is called *skew-symmetric*.

For the transposition we have the following properties.

**Lemma 4.6** For  $A, \hat{A} \in R^{n,m}$ ,  $B \in R^{m,\ell}$  and  $\lambda \in R$  the following assertions hold:

- (1)  $(A^T)^T = A$ .
- (2)  $(A + \hat{A})^T = A^T + \hat{A}^T$ .
- (3)  $(\lambda \cdot A)^T = \lambda \cdot A^T$ .
- (4)  $(A * B)^T = B^T * A^T$ .

*Proof* Properties (1)–(3) are exercises. For the proof of (4) let  $A * B = [c_{ij}]$  with  $c_{ij} = \sum_{k=1}^m a_{ik}b_{kj}$ ,  $A^T = [\tilde{a}_{ij}]$ ,  $B^T = [\tilde{b}_{ij}]$  and  $(A * B)^T = [\tilde{c}_{ij}]$ . Then

$$\tilde{c}_{ij} = c_{ji} = \sum_{k=1}^m a_{jk}b_{ki} = \sum_{k=1}^m \tilde{a}_{kj}\tilde{b}_{ik} = \sum_{k=1}^m \tilde{b}_{ik}\tilde{a}_{kj},$$

from which we see that  $(A * B)^T = B^T * A^T$ . □

### MATLAB-Minute.

Carry out the following commands in order to get used to the matrix operations of this chapter in MATLAB notation: `A=ones(5,2)`, `A+A`, `A-3*A`, `A'`, `A'*A`, `A*A'`.

(In order to see MATLAB's output, do not put a semicolon at the end of the command.)

*Example 4.7* Consider again the example of car insurance premiums from Chap. 1. Recall that  $p_{ij}$  denotes the probability that a customer in class  $C_i$  in this year will move to the class  $C_j$ . Our example consists of four such classes, and the 16 probabilities can be associated with a row-stochastic  $4 \times 4$  matrix (cp. (1.2)), which we denote by  $P$ . Suppose that the insurance company has the following distribution of customers in the four classes: 40 % in class  $C_1$ , 30 % in class  $C_2$ , 20 % in class  $C_3$ , and 10 % in class  $C_4$ . Then the  $1 \times 4$  matrix

$$p_0 := [0.4, 0.3, 0.2, 0.1]$$

describes the initial customer distribution. Using the matrix multiplication we now compute

$$\begin{aligned} p_1 &:= p_0 * P = [0.4, 0.3, 0.2, 0.1] * \begin{bmatrix} 0.15 & 0.85 & 0.00 & 0.00 \\ 0.15 & 0.00 & 0.85 & 0.00 \\ 0.05 & 0.10 & 0.00 & 0.85 \\ 0.05 & 0.00 & 0.10 & 0.85 \end{bmatrix} \\ &= [0.12, 0.36, 0.265, 0.255]. \end{aligned}$$

Then  $p_1$  contains the distribution of the customers in the next year. As an example, consider the entry of  $p_0 * P$  in position (1, 4), which is computed by

$$0.4 \cdot 0.00 + 0.3 \cdot 0.00 + 0.2 \cdot 0.85 + 0.1 \cdot 0.85 = 0.255.$$

A customer in the classes  $C_1$  or  $C_2$  in this year cannot move to the class  $C_4$ . Thus, the respective initial percentages are multiplied by the probabilities  $p_{14} = 0.00$  and  $p_{24} = 0.00$ . A customer in the class  $C_3$  or  $C_4$  will be in the class  $C_4$  with the probabilities  $p_{34} = 0.85$  or  $p_{44} = 0.85$ , respectively. This yields the two products  $0.2 \cdot 0.85$  and  $0.1 \cdot 0.85$ .

Continuing in the same way we obtain after  $k$  years the distribution

$$p_k := p_0 * P^k, \quad k = 0, 1, 2, \dots$$

(This formula also holds for  $k = 0$ , since  $P^0 = I_4$ .) The insurance company can use this formula to compute the revenue from the payments of premium rates in the coming years. Assume that the full premium rate (class  $C_1$ ) is 500 Euros per year. Then the rates in classes  $C_2$ ,  $C_3$ , and  $C_4$  are 450, 400 and 300 Euros (10, 20 and 40 % discount). If there are 1000 customers initially, then the revenue in the first year (in Euros) is

$$1000 \cdot (p_0 * [500, 450, 400, 300]^T) = 445000.$$

If no customer cancels the contract, then this model yields the revenue in year  $k \geq 0$  as

$$1000 \cdot \left( p_k * [500, 450, 400, 300]^T \right) = 1000 \cdot \left( p_0 * (P^k * [500, 450, 400, 300]^T) \right).$$

For example, the revenue in the next 4 years is 404500, 372025, 347340 and 341819 (rounded to full Euros). These numbers decrease annually, but the rate of the decrease seems to slow down. Does there exist a “stationary state”, i.e., a state when the revenue is not changing (significantly) any more? Which properties of the model guarantee the existence of such a state? These are important practical questions for the insurance company. Only the existence of a stationary state guarantees significant revenues in the long-time future. Since the formula depends essentially on the entries of the matrix  $P^k$ , we have reached an interesting problem of Linear Algebra: the analysis of the properties of row-stochastic matrices. We will analyze these properties in Sect. 8.3.

## 4.2 Matrix Groups and Rings

In this section we study algebraic structures that are formed by certain sets of matrices and the matrix operations introduced above. We begin with the addition in  $R^{n,m}$ .

**Theorem 4.8** ( $R^{n,m}, +$ ) is a commutative group. The neutral element is  $0 \in R^{n,m}$  (the zero matrix) and for  $A = [a_{ij}] \in R^{n,m}$  the inverse element is  $-A := [-a_{ij}] \in R^{n,m}$ . (We write  $A - B$  instead of  $A + (-B)$ .)

*Proof* Using the associativity of the addition in  $R$ , for arbitrary  $A, B, C \in R^{n,m}$ , we obtain

$$\begin{aligned} (A + B) + C &= [a_{ij} + b_{ij}] + [c_{ij}] = [(a_{ij} + b_{ij}) + c_{ij}] = [a_{ij} + (b_{ij} + c_{ij})] \\ &= [a_{ij}] + [b_{ij} + c_{ij}] = A + (B + C). \end{aligned}$$

Thus, the addition in  $R^{n,m}$  is associative.

The zero matrix  $0 \in R^{n,m}$  satisfies  $0 + A = [0] + [a_{ij}] = [0 + a_{ij}] = [a_{ij}] = A$ . For a given  $A = [a_{ij}] \in R^{n,m}$  and  $-A := [-a_{ij}] \in R^{n,m}$  we have  $-A + A = [-a_{ij}] + [a_{ij}] = [-a_{ij} + a_{ij}] = [0] = 0$ .

Finally, the commutativity of the addition in  $R$  implies that  $A + B = [a_{ij}] + [b_{ij}] = [a_{ij} + b_{ij}] = [b_{ij} + a_{ij}] = B + A$ .  $\square$

Note that (2) in Lemma 4.6 implies that the transposition is a homomorphism (even an isomorphism) between the groups  $(R^{n,m}, +)$  and  $(R^{m,n}, +)$  (cp. Definition 3.6).

**Theorem 4.9**  $(R^{n,n}, +, *)$  is a ring with unit given by the identity matrix  $I_n$ . This ring is commutative only for  $n = 1$ .

*Proof* We have already shown that  $(R^{n,n}, +)$  is a commutative group (cp. Theorem 4.8). The other properties of a ring (associativity, distributivity and the existence of a unit element) follow from Lemma 4.3. The commutativity for  $n = 1$  holds because of the commutativity of the multiplication in the ring  $R$ . The example

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} * \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \neq \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} * \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

shows that the ring  $R^{n,n}$  is not commutative for  $n \geq 2$ . □

The example in the proof of Theorem 4.9 shows that for  $n \geq 2$  the ring  $R^{n,n}$  has non-trivial zero-divisors, i.e., there exist matrices  $A, B \in R^{n,n} \setminus \{0\}$  with  $A * B = 0$ . These exist even when  $R$  is a field.

Let us now consider the *invertibility* of matrices in the ring  $R^{n,n}$  (with respect to the matrix multiplication). For a given matrix  $A \in R^{n,n}$ , an inverse  $\tilde{A} \in R^{n,n}$  must satisfy the two equations  $\tilde{A} * A = I_n$  and  $A * \tilde{A} = I_n$  (cp. Definition 3.10). If an inverse of  $A \in R^{n,n}$  exists, i.e., if  $A$  is *invertible*, then the inverse is unique and denoted by  $A^{-1}$  (cp. Theorem 3.11). An invertible matrix is sometimes called *non-singular*, while a non-invertible matrix is called *singular*. We will show in Corollary 7.20 that the existence of the inverse already is implied by one of the two equations  $\tilde{A} * A = I_n$  and  $A * \tilde{A} = I_n$ , i.e., if one of them holds, then  $A$  is invertible and  $A^{-1} = \tilde{A}$ . Until then, to be correct, we will have to check the validity of both equations.

Not all matrices  $A \in R^{n,n}$  are invertible. Simple examples are the non-invertible matrices

$$A = [0] \in R^{1,1} \quad \text{and} \quad A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \in R^{2,2}.$$

Another non-invertible matrix is

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} \in \mathbb{Z}^{2,2}.$$

However, considered as an element of  $\mathbb{Q}^{2,2}$ , the (unique) inverse of  $A$  is given by

$$A^{-1} = \begin{bmatrix} 1 & -\frac{1}{2} \\ 0 & \frac{1}{2} \end{bmatrix} \in \mathbb{Q}^{2,2}.$$

**Lemma 4.10** If  $A, B \in R^{n,n}$  are invertible, then the following assertions hold:

- (1)  $A^T$  is invertible with  $(A^T)^{-1} = (A^{-1})^T$ . (We also write this matrix as  $A^{-T}$ .)
- (2)  $A * B$  is invertible with  $(A * B)^{-1} = B^{-1} * A^{-1}$ .

*Proof*

(1) Using (4) in Lemma 4.6 we have

$$(A^{-1})^T * A^T = (A * A^{-1})^T = I_n^T = I_n = I_n^T = (A^{-1} * A)^T = A^T * (A^{-1})^T,$$

and thus  $(A^{-1})^T$  is the inverse of  $A^T$ .

(2) This was already shown in Theorem 3.11 for general rings with unit and thus it holds, in particular, for the ring  $(R^{n,n}, +, *)$ .  $\square$

Our next result shows that the invertible matrices form a multiplicative group.

**Theorem 4.11** *The set of invertible  $n \times n$  matrices over  $R$  forms a group with respect to the matrix multiplication. We denote this group by  $GL_n(R)$  (“GL” abbreviates “general linear (group)”).*

*Proof* The associativity of the multiplication in  $GL_n(R)$  is clear. As shown in (2) in Lemma 4.10, the product of two invertible matrices is an invertible matrix. The neutral element in  $GL_n(R)$  is the identity matrix  $I_n$ , and since every  $A \in GL_n(R)$  is assumed to be invertible,  $A^{-1}$  exists with  $(A^{-1})^{-1} = A \in GL_n(R)$ .  $\square$

We now introduce some important classes of matrices.

**Definition 4.12** Let  $A = [a_{ij}] \in R^{n,n}$ .

- (1)  $A$  is called *upper triangular*, if  $a_{ij} = 0$  for all  $i > j$ .  
 $A$  is called *lower triangular*, if  $a_{ij} = 0$  for all  $j > i$  (i.e.,  $A^T$  is upper triangular).
- (2)  $A$  is called *diagonal*, if  $a_{ij} = 0$  for all  $i \neq j$  (i.e.,  $A$  is upper and lower triangular).  
 We write a diagonal matrix as  $A = \text{diag}(a_{11}, \dots, a_{nn})$ .

We next investigate these sets of matrices with respect to their group properties, beginning with the invertible upper and lower triangular matrices.

**Theorem 4.13** *The sets of the invertible upper triangular  $n \times n$  matrices and of the invertible lower triangular  $n \times n$  matrices over  $R$  form subgroups of  $GL_n(R)$ .*

*Proof* We will only show the result for the upper triangular matrices; the proof for the lower triangular matrices is analogous. In order to establish the subgroup property we will prove the three properties from Theorem 3.5.

Since  $I_n$  is an invertible upper triangular matrix, the set of the invertible upper triangular matrices is a nonempty subset of  $GL_n(R)$ .

Next we show that for two invertible upper triangular matrices  $A, B \in R^{n,n}$  the product  $C = A * B$  is again an invertible upper triangular matrix. The invertibility of  $C = [c_{ij}]$  follows from (2) in Lemma 4.10. For  $i > j$  we have

$$\begin{aligned}
c_{ij} &= \sum_{k=1}^n a_{ik}b_{kj} \quad (\text{here } b_{kj} = 0 \text{ for } k > j) \\
&= \sum_{k=1}^j a_{ik}b_{kj} \quad (\text{here } a_{ik} = 0 \text{ for } k = 1, \dots, j, \text{ since } i > j) \\
&= 0.
\end{aligned}$$

Therefore,  $C$  is upper triangular.

It remains to prove that the inverse  $A^{-1}$  of an invertible upper triangular matrix  $A$  is an upper triangular matrix. For  $n = 1$  the assertion holds trivially, so we assume that  $n \geq 2$ . Let  $A^{-1} = [c_{ij}]$ , then the equation  $A * A^{-1} = I_n$  can be written as a system of  $n$  equations

$$\begin{bmatrix} a_{11} & \cdots & \cdots & a_{1n} \\ 0 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{nn} \end{bmatrix} * \begin{bmatrix} c_{1j} \\ \vdots \\ \vdots \\ c_{nj} \end{bmatrix} = \begin{bmatrix} \delta_{1j} \\ \vdots \\ \vdots \\ \delta_{nj} \end{bmatrix}, \quad j = 1, \dots, n. \quad (4.3)$$

Here,  $\delta_{ij}$  is the Kronecker delta-function defined in (4.1).

We will now prove inductively for  $i = n, n-1, \dots, 1$  that the diagonal entry  $a_{ii}$  of  $A$  is invertible with  $a_{ii}^{-1} = c_{ii}$ , and that

$$c_{ij} = a_{ii}^{-1} \left( \delta_{ij} - \sum_{\ell=i+1}^n a_{i\ell}c_{\ell j} \right), \quad j = 1, \dots, n. \quad (4.4)$$

This formula implies, in particular, that  $c_{ij} = 0$  for  $i > j$ .

For  $i = n$  the last row of (4.3) is given by

$$a_{nn}c_{nj} = \delta_{nj}, \quad j = 1, \dots, n.$$

For  $j = n$  we have  $a_{nn}c_{nn} = 1 = c_{nn}a_{nn}$ , where in the second equation we use the commutativity of the multiplication in  $R$ . Therefore,  $a_{nn}$  is invertible with  $a_{nn}^{-1} = c_{nn}$ , and thus

$$c_{nj} = a_{nn}^{-1}\delta_{nj}, \quad j = 1, \dots, n.$$

This is equivalent to (4.4) for  $i = n$ . (Note that for  $i = n$  in (4.4) the sum is empty and thus equal to zero.) In particular,  $c_{nj} = 0$  for  $j = 1, \dots, n-1$ .

Now assume that our assertion holds for  $i = n, \dots, k+1$ , where  $1 \leq k \leq n-1$ . Then, in particular,  $c_{ij} = 0$  for  $k+1 \leq i \leq n$  and  $i > j$ . In words, the rows  $i = n, \dots, k+1$  of  $A^{-1}$  are in ‘‘upper triangular form’’. In order to prove the assertion for  $i = k$ , we consider the  $k$ th row in (4.3), which is given by

$$a_{kk}c_{kj} + a_{k,k+1}c_{k+1,j} + \dots + a_{kn}c_{nj} = \delta_{kj}, \quad j = 1, \dots, n. \quad (4.5)$$

For  $j = k (< n)$  we obtain

$$a_{kk}c_{kk} + a_{k,k+1}c_{k+1,k} + \dots + a_{kn}c_{nk} = 1.$$

By the induction hypothesis, we have  $c_{k+1,k} = \dots = c_{n,k} = 0$ . This implies  $a_{kk}c_{kk} = 1 = c_{kk}a_{kk}$ , where we have used the commutativity of the multiplication in  $R$ . Hence  $a_{kk}$  is invertible with  $a_{kk}^{-1} = c_{kk}$ . From (4.5) we get

$$c_{kj} = a_{kk}^{-1} (\delta_{kj} - a_{k,k+1}c_{k+1,j} - \dots - a_{kn}c_{nj}), \quad j = 1, \dots, n,$$

and hence (4.4) holds for  $i = k$ . If  $k > j$ , then  $\delta_{kj} = 0$  and  $c_{k+1,j} = \dots = c_{nj} = 0$ , which gives  $c_{kj} = 0$ .  $\square$

We point out that (4.4) represents a *recursive formula* for computing the entries of the inverse of an invertible upper triangular matrix. Using this formula the entries are computed “from bottom to top” and “from right to left”. This process is sometimes called *backward substitution*.

In the following we will frequently partition matrices into blocks and make use of the *block multiplication*: For every  $k \in \{1, \dots, n-1\}$ , we can write  $A \in R^{n,n}$  as

$$A = \left[ \begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array} \right] \quad \text{with } A_{11} \in R^{k,k} \text{ and } A_{22} \in R^{n-k,n-k}.$$

If  $A, B \in R^{n,n}$  are both partitioned like this, then the product  $A * B$  can be evaluated blockwise, i.e.,

$$\left[ \begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array} \right] * \left[ \begin{array}{c|c} B_{11} & B_{12} \\ \hline B_{21} & B_{22} \end{array} \right] = \left[ \begin{array}{c|c} A_{11} * B_{11} + A_{12} * B_{21} & A_{11} * B_{12} + A_{12} * B_{22} \\ \hline A_{21} * B_{11} + A_{22} * B_{21} & A_{21} * B_{12} + A_{22} * B_{22} \end{array} \right]. \quad (4.6)$$

In particular, if

$$A = \left[ \begin{array}{c|c} A_{11} & A_{12} \\ \hline 0 & A_{22} \end{array} \right]$$

with  $A_{11} \in GL_k(R)$  and  $A_{22} \in GL_{n-k}(R)$ , then  $A \in GL_n(R)$  and a direct computation shows that

$$A^{-1} = \left[ \begin{array}{c|c} A_{11}^{-1} & -A_{11}^{-1} * A_{12} * A_{22}^{-1} \\ \hline 0 & A_{22}^{-1} \end{array} \right]. \quad (4.7)$$

**MATLAB-Minute.**

Create block matrices in MATLAB by carrying out the following commands:

```
k=5;
A11=gallery('tridiag',-ones(k-1,1),2*ones(k,1),-ones(k-1,1));
A12=zeros(k,2); A12(1,1)=1; A12(2,2)=1;
A22=-eye(2);
A=full([A11 A12; A12' A22])
B=full([A11 A12; zeros(2,k) -A22])
```

Investigate the meaning of the command `full`. Compute the products  $A*B$  and  $B*A$  as well as the inverses  $\text{inv}(A)$  and  $\text{inv}(B)$ . Compute the inverse of  $B$  in MATLAB with the formula (4.7).

**Corollary 4.14** *The set of the invertible diagonal  $n \times n$  matrices over  $R$  forms a commutative subgroup (with respect to the matrix multiplication) of the invertible upper (or lower) triangular  $n \times n$  matrices over  $R$ .*

*Proof* Since  $I_n$  is an invertible diagonal matrix, the invertible diagonal  $n \times n$  matrices form a nonempty subset of the invertible upper (or lower) triangular  $n \times n$  matrices. If  $A = \text{diag}(a_{11}, \dots, a_{nn})$  and  $B = \text{diag}(b_{11}, \dots, b_{nn})$  are invertible, then  $A * B$  is invertible (cp. (2) in Lemma 4.10) and diagonal, since

$$A * B = \text{diag}(a_{11}, \dots, a_{nn}) * \text{diag}(b_{11}, \dots, b_{nn}) = \text{diag}(a_{11}b_{11}, \dots, a_{nn}b_{nn}).$$

Moreover, if  $A = \text{diag}(a_{11}, \dots, a_{nn})$  is invertible, then  $a_{ii} \in R$  is invertible for all  $i = 1, \dots, n$  (cp. the proof of Theorem 4.13). The inverse  $A^{-1}$  is given by the invertible diagonal matrix  $\text{diag}(a_{11}^{-1}, \dots, a_{nn}^{-1})$ . Finally, the commutativity property  $A * B = B * A$  follows directly from the commutativity in  $R$ .  $\square$

**Definition 4.15** A matrix  $P \in R^{n,n}$  is called a *permutation matrix*, if in every row and every column of  $P$  there is exactly one unit and all other entries are zero.

The term “permutation” means “exchange”. If a matrix  $A \in R^{n,n}$  is multiplied with a permutation matrix from the left or from the right, then its rows or columns, respectively, are exchanged (or permuted). For example, if

$$P = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \in \mathbb{Z}^{3,3},$$

then

$$P * A = \begin{bmatrix} 7 & 8 & 9 \\ 4 & 5 & 6 \\ 1 & 2 & 3 \end{bmatrix} \quad \text{and} \quad A * P = \begin{bmatrix} 3 & 2 & 1 \\ 6 & 5 & 4 \\ 9 & 8 & 7 \end{bmatrix}.$$

**Theorem 4.16** *The set of the  $n \times n$  permutation matrices over  $R$  forms a subgroup of  $GL_n(R)$ . In particular, if  $P \in R^{n,n}$  is a permutation matrix, then  $P$  is invertible with  $P^{-1} = P^T$ .*

*Proof* Exercise. □

From now on we will omit the multiplication sign in the matrix multiplication and write  $AB$  instead of  $A * B$ .

### Exercises

(In the following exercises  $R$  is a commutative ring with unit.)

4.1 Consider the following matrices over  $\mathbb{Z}$ :

$$A = \begin{bmatrix} 1 & -2 & 4 \\ -2 & 3 & -5 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 4 \\ 3 & 6 \\ 1 & -2 \end{bmatrix}, \quad C = \begin{bmatrix} -1 & 0 \\ 1 & 1 \end{bmatrix}.$$

Determine, if possible, the matrices  $CA$ ,  $BC$ ,  $B^T A$ ,  $A^T C$ ,  $(-A)^T C$ ,  $B^T A^T$ ,  $AC$  and  $CB$ .

4.2 Consider the matrices

$$A = [a_{ij}] \in R^{n,m}, \quad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in R^{n,1}, \quad y = [y_1, \dots, y_m] \in R^{1,m}.$$

Which of the following expressions are well defined for  $m \neq n$  or  $m = n$ ?

- (a)  $xy$ , (b)  $x^T y$ , (c)  $yx$ , (d)  $yx^T$ , (e)  $xAy$ , (f)  $x^T Ay$ ,  
 (g)  $xAy^T$ , (h)  $x^T Ay^T$ , (i)  $xyA$ , (j)  $xyA^T$ , (k)  $Axy$ , (l)  $A^T xy$ .

4.3 Show the following computational rules:

$$\mu_1 x_1 + \mu_2 x_2 = [x_1, x_2] \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \text{and} \quad A[x_1, x_2] = [Ax_1, Ax_2]$$

for  $A \in R^{n,m}$ ,  $x_1, x_2 \in R^{m,1}$  and  $\mu_1, \mu_2 \in R$ .

- 4.4 Prove Lemma 4.3 (2)–(4).
- 4.5 Prove Lemma 4.4.
- 4.6 Prove Lemma 4.6 (1)–(3).
- 4.7 Let  $A = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \in \mathbb{Z}^{3,3}$ . Determine  $A^n$  for all  $n \in \mathbb{N} \cup \{0\}$ .
- 4.8 Let  $p = \alpha_n t^n + \dots + \alpha_1 t + \alpha_0 t^0 \in R[t]$  be a polynomial (cp. Example 3.17) and  $A \in R^{m,m}$ . We define  $p(A) \in R^{m,m}$  as  $p(A) := \alpha_n A^n + \dots + \alpha_1 A + \alpha_0 I_m$ .
- (a) Determine  $p(A)$  for  $p = t^2 - 2t + 1 \in \mathbb{Z}[t]$  and  $A = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \in \mathbb{Z}^{2,2}$ .
- (b) For a fixed matrix  $A \in R^{m,m}$  consider the map  $f_A : R[t] \rightarrow R^{m,m}$ ,  $p \mapsto p(A)$ . Show that  $f_A(p + q) = f_A(p) + f_A(q)$  and  $f_A(pq) = f_A(p)f_A(q)$  for all  $p, q \in R[t]$ .  
(The map  $f_A$  is a *ring homomorphism* between the rings  $R[t]$  and  $R^{m,m}$ .)
- (c) Show that  $f_A(R[t]) = \{p(A) \mid p \in R[t]\}$  is a commutative subring of  $R^{m,m}$ , i.e., that  $f_A(R[t])$  is a subring of  $R^{m,m}$  (cp. Exercise 3.14) and that the multiplication in this subring is commutative.
- (d) Is the map  $f_A$  surjective?
- 4.9 Let  $K$  be a field with  $1 + 1 \neq 0$ . Show that every matrix  $A \in K^{n,n}$  can be written as  $A = M + S$  with a symmetric matrix  $M \in K^{n,n}$  (i.e.,  $M^T = M$ ) and a skew-symmetric matrix  $S \in K^{n,n}$  (i.e.,  $S^T = -S$ ).  
Does this also hold in a field with  $1 + 1 = 0$ ? Give a proof or a counterexample.
- 4.10 Show the *binomial formula* for commuting matrices: If  $A, B \in R^{n,n}$  with  $AB = BA$ , then  $(A + B)^k = \sum_{j=0}^k \binom{k}{j} A^j B^{k-j}$ , where  $\binom{k}{j} := \frac{k!}{j!(k-j)!}$ .
- 4.11 Let  $A \in R^{n,n}$  be a matrix for which  $I_n - A$  is invertible. Show that  $(I_n - A)^{-1}(I_n - A^{m+1}) = \sum_{j=0}^m A^j$  holds for every  $m \in \mathbb{N}$ .
- 4.12 Let  $A \in R^{n,n}$  be a matrix for which an  $m \in \mathbb{N}$  with  $A^m = I_n$  exists and let  $m$  be smallest natural number with this property.
- (a) Investigate whether  $A$  is invertible, and if so, give a particularly simple representation of the inverse.
- (b) Determine the cardinality of the set  $\{A^k \mid k \in \mathbb{N}\}$ .
- 4.13 Let  $\mathcal{A} = \{[a_{ij}] \in R^{n,n} \mid a_{nj} = 0 \text{ for } j = 1, \dots, n\}$ .
- (a) Show that  $\mathcal{A}$  is a subring of  $R^{n,n}$ .
- (b) Show that  $AM \in \mathcal{A}$  for all  $M \in R^{n,n}$  and  $A \in \mathcal{A}$ .  
(A subring with this property is called a *left ideal* of  $R^{n,n}$ .)
- (c) Determine an analogous subring  $\mathcal{B}$  of  $R^{n,n}$ , such that  $MB \in \mathcal{B}$  for all  $M \in R^{n,n}$  and  $B \in \mathcal{B}$ .  
(A subring with this property is called a *right ideal* of  $R^{n,n}$ .)
- 4.14 Examine whether  $(G, *)$  with

$$G = \left\{ \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix} \mid \alpha \in \mathbb{R} \right\}$$

is a subgroup of  $GL_2(\mathbb{R})$ .

4.15 Generalize the block multiplication (4.6) to matrices  $A \in R^{n,m}$  and  $B \in R^{m,\ell}$ .

4.16 Determine all invertible upper triangular matrices  $A \in R^{n,n}$  with  $A^{-1} = A^T$ .

4.17 Let  $A_{11} \in R^{n_1,n_1}$ ,  $A_{12} \in R^{n_1,n_2}$ ,  $A_{21} \in R^{n_2,n_1}$ ,  $A_{22} \in R^{n_2,n_2}$  and

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \in R^{n_1+n_2,n_1+n_2}.$$

(a) Let  $A_{11} \in GL_{n_1}(R)$ . Show that  $A$  is invertible if and only if  $A_{22} - A_{21}A_{11}^{-1}A_{12}$  is invertible and derive in this case a formula for  $A^{-1}$ .

(b) Let  $A_{22} \in GL_{n_2}(R)$ . Show that  $A$  is invertible if and only if  $A_{11} - A_{12}A_{22}^{-1}A_{21}$  is invertible and derive in this case a formula for  $A^{-1}$ .

4.18 Let  $A \in GL_n(R)$ ,  $U \in R^{n,m}$  and  $V \in R^{m,n}$ . Show the following assertions:

(a)  $A + UV \in GL_n(R)$  holds if and only if  $I_m + VA^{-1}U \in GL_m(R)$ .

(b) If  $I_m + VA^{-1}U \in GL_m(R)$ , then

$$(A + UV)^{-1} = A^{-1} - A^{-1}U(I_m + VA^{-1}U)^{-1}VA^{-1}.$$

(This last equation is called the *Sherman-Morrison-Woodbury formula*; named after Jack Sherman, Winifred J. Morrison and Max A. Woodbury.)

4.19 Show that the set of block upper triangular matrices with invertible  $2 \times 2$  diagonal blocks, i.e., the set of matrices

$$\begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1m} \\ 0 & A_{22} & \cdots & A_{2m} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & A_{mm} \end{bmatrix}, \quad A_{ii} \in GL_2(R), \quad i = 1, \dots, m,$$

is a group with respect to the matrix multiplication.

4.20 Prove Theorem 4.16. Is the group of permutation matrices commutative?

4.21 Show that the following is an equivalence relation on  $R^{n,n}$ :

$$A \sim B \Leftrightarrow \text{There exists a permutation matrix } P \text{ with } A = P^T B P.$$

4.22 A company produces from four raw materials  $R_1, R_2, R_3, R_4$  five intermediate products  $Z_1, Z_2, Z_3, Z_4, Z_5$ , and from these three final products  $E_1, E_2, E_3$ . The following tables show how many units of  $R_i$  and  $Z_j$  are required for producing one unit of  $Z_k$  and  $E_\ell$ , respectively:

	$Z_1$	$Z_2$	$Z_3$	$Z_4$	$Z_5$
$R_1$	0	1	1	1	2
$R_2$	5	0	1	2	1
$R_3$	1	1	1	1	0
$R_4$	0	2	0	1	0

	$E_1$	$E_2$	$E_3$
$Z_1$	1	1	1
$Z_2$	1	2	0
$Z_3$	0	1	1
$Z_4$	4	1	1
$Z_5$	3	1	1

For instance, five units of  $R_2$  and one unit of  $R_3$  are required for producing one unit of  $Z_1$ .

- (a) Determine, with the help of matrix operations, a corresponding table which shows how many units of  $R_i$  are required for producing one unit of  $E_\ell$ .
- (b) Determine how many units of the four raw materials are required for producing 100 units of  $E_1$ , 200 units of  $E_2$  and 300 units of  $E_3$ .