4 Linear Algebra

4.1 Matrices

4.1.1 Notion of Matrix

1. Matrices A of Size (m, n) or Briefly $\mathbf{A}_{(m,n)}$

are systems of m times n elements, e.g., real or complex numbers, or functions, derivatives, vectors, arranged in m rows and n columns:

$$\mathbf{A} = (a_{ij}) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \quad \begin{matrix} \leftarrow & \text{1st row} \\ \leftarrow & \text{2nd row} \\ & \vdots \\ \leftarrow & m\text{-th row} \end{matrix} \qquad (4.1)$$

$\uparrow$ 1st $\uparrow$ 2nd $\uparrow$ n-th column.

With the notion *size of a matrix* matrices are classified according to their number of rows m and number of columns n: $\mathbf{A}$ of size (m, n). A matrix is called a *square matrix* if the number of rows and columns is equal, otherwise it is a *rectangular matrix*.

2. Real and Complex Matrices

Real matrices have real elements, *complex matrices* have complex elements. If a matrix has complex elements

$$a_{\mu\nu} + \mathrm{i} b_{\mu\nu} \qquad (4.2a)$$

it can be decomposed into the form

$$\mathbf{A} + \mathrm{i}\,\mathbf{B} \qquad (4.2b)$$

where $\mathbf{A}$ and $\mathbf{B}$ have real elements only (arithmetical operations see 4.1.4, p. 272).
If a matrix $\mathbf{A}$ has complex elements, then its *conjugate complex matrix* $\mathbf{A}^*$ has the elements

$$a^*_{\mu\nu} = \mathrm{Re}\,(a_{\mu\nu}) - \mathrm{i}\,\mathrm{Im}\,(a_{\mu\nu}). \qquad (4.2c)$$

3. Transposed Matrices $\mathbf{A}^{\mathrm{T}}$

Changing the rows and columns of a matrix $\mathbf{A}$ of size (m, n) gives the *transposed matrix* $\mathbf{A}^{\mathrm{T}}$. This has the size (n, m) and

$$(a_{\nu\mu})^{\mathrm{T}} = (a_{\mu\nu}) \qquad (4.3)$$

is valid.

4. Adjoint Matrices

The *adjoint matrix* $\mathbf{A}^{\mathrm{H}}$ of a complex matrix $\mathbf{A}$ is the transpose of its conjugate complex matrix $\mathbf{A}^*$ (which can not be confused with the adjoint matrix $\mathbf{A}_{adj}$, see 4.2.2, p. 278):

$$\mathbf{A}^{\mathrm{H}} = (\mathbf{A}^*)^{\mathrm{T}}. \qquad (4.4)$$

5. Zero Matrix

A matrix $\mathbf{0}$ is called a *zero matrix* if it has only zero elements:

$$\mathbf{0} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}. \qquad (4.5)$$

I.N. Bronshtein et al., *Handbook of Mathematics*,
DOI 10.1007/978-3-662-46221-8_4

4.1.2 Square Matrices

1. Definition

Square matrices have the same number of rows and columns, i.e., $m = n$:

$$\mathbf{A} = \mathbf{A}_{(n,n)} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}. \tag{4.6}$$

The elements $a_{\mu\nu}$ of a matrix $\mathbf{A}$ in the diagonal from the left upper corner to the right lower one are the *elements of the main diagonal.* They are denoted by $a_{11}, a_{22}, \ldots, a_{nn}$, i.e., they are all the elements $a_{\mu\nu}$ with $\mu = \nu$.

2. Diagonal Matrices

A square matrix $\mathbf{D}$ is called a *diagonal matrix* if all of the non-diagonal elements are equal to zero:

$$a_{\mu\nu} = 0 \quad \text{for} \quad \mu \neq \nu: \quad \mathbf{D} = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix} = \begin{pmatrix} a_{11} & & \mathbf{O} & \\ & a_{22} & & \\ & & \ddots & \\ \mathbf{O} & & & a_{nn} \end{pmatrix}. \tag{4.7}$$

3. Scalar Matrix

A diagonal matrix $\mathbf{S}$ is called a *scalar matrix* if all the diagonal elements are the same real or complex number c:

$$\mathbf{S} = \begin{pmatrix} c & 0 & \cdots & 0 \\ 0 & c & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & c \end{pmatrix}. \tag{4.8}$$

4. Trace or Spur of a Matrix

For a square matrix, the *trace* or *spur* of the matrix is defined as the sum of the main diagonal elements:

$$\mathrm{Tr}\,(\mathbf{A}) = a_{11} + a_{22} + \ldots + a_{nn} = \sum_{\mu=1}^{n} a_{\mu\mu}. \tag{4.9}$$

5. Symmetric Matrices

A square matrix $\mathbf{A}$ is *symmetric* if it is equal to its own transpose:

$$\mathbf{A} = \mathbf{A}^{\mathrm{T}}. \tag{4.10}$$

For the elements lying in symmetric positions with respect to the main diagonal

$$a_{\mu\nu} = a_{\nu\mu} \tag{4.11}$$

is valid.

6. Normal Matrices

satisfy the equality

$$\mathbf{A}^{\mathrm{H}}\mathbf{A} = \mathbf{A}\mathbf{A}^{\mathrm{H}}. \tag{4.12}$$

(For the product of matrices see 4.1.4, p. 272.)

7. Antisymmetric or Skew-Symmetric Matrices

are the square matrices $\mathbf{A}$ with the property:

$$\mathbf{A} = -\mathbf{A}^{\mathrm{T}}. \tag{4.13a}$$

For the elements $a_{\mu\nu}$ of an antisymmetric matrix the equalities

$$a_{\mu\nu} = -a_{\nu\mu}, \quad a_{\mu\mu} = 0 \tag{4.13b}$$

are valid, so the trace of an antisymmetric matrix vanishes:

$$\mathrm{Tr}\,(\mathbf{A}) = 0. \tag{4.13c}$$

The elements lying in symmetric positions with respect to the main diagonal differ from each other only in sign.
Every square matrix $\mathbf{A}$ can be decomposed into the sum of a symmetric matrix $\mathbf{A_s}$ and an antisymmetric matrix $\mathbf{A_{as}}$:

$$\mathbf{A} = \mathbf{A_s} + \mathbf{A_{as}} \qquad \text{with} \qquad \mathbf{A_s} = \frac{1}{2}(\mathbf{A} + \mathbf{A}^{\mathrm{T}});\ \mathbf{A_{as}} = \frac{1}{2}(\mathbf{A} - \mathbf{A}^{\mathrm{T}}). \tag{4.13d}$$

8. Hermitian Matrices or Self-Adjoint Matrices
are square matrices $\mathbf{A}$ equal to their own adjoints :

$$\mathbf{A} = (\mathbf{A}^*)^{\mathrm{T}} = \mathbf{A}^{\mathrm{H}}. \tag{4.14}$$

Over the real numbers the concepts of symmetric and Hermitian matrices are the same. The determinant of a Hermitian matrix is real.

9. Anti-Hermitian or Skew-Hermitian Matrices
are the square matrices equal to their negative adjoints:

$$\mathbf{A} = -(\mathbf{A}^*)^{\mathrm{T}} = -\mathbf{A}^{\mathrm{H}}. \tag{4.15a}$$

For the elements $a_{\mu\nu}$ and the trace of an anti-Hermitian matrix the equalities

$$a_{\mu\nu} = -a^*_{\mu\nu},\ a_{\mu\mu} = 0;\ \mathrm{Tr}\,(\mathbf{A}) = 0 \tag{4.15b}$$

are valid. Every square matrix $\mathbf{A}$ can be decomposed into a sum of a Hermitian matrix $\mathbf{A_h}$ and an anti-Hermitian matrix $\mathbf{A_{ah}}$:

$$\mathbf{A} = \mathbf{A}_{\mathrm{h}} + \mathbf{A}_{\mathrm{ah}} \quad \text{with} \quad \mathbf{A}_{\mathrm{h}} = \frac{1}{2}(\mathbf{A} + \mathbf{A}^{\mathrm{H}}),\ \mathbf{A}_{\mathrm{ah}} = \frac{1}{2}(\mathbf{A} - \mathbf{A}^{\mathrm{H}}). \tag{4.15c}$$

10. Identity Matrix I
is a diagonal matrix such that every diagonal element is equal to 1 and all of the non-diagonal elements are equal to zero:

$$\mathbf{I} = \begin{pmatrix} 1\ 0 \cdots 0 \\ 0\ 1 \cdots 0 \\ \vdots\ \vdots\ \ \vdots\ \ \vdots \\ 0\ 0 \cdots 1 \end{pmatrix} = (\delta_{\mu\nu}) \qquad \text{with} \qquad \delta_{\mu\nu} = \begin{cases} 0 \text{ for } \mu \neq \nu, \\ 1 \text{ for } \mu = \nu. \end{cases} \tag{4.16}$$

The symbol $\delta_{\mu\nu}$ is called the *Kronecker symbol*.

11. Triangular Matrix

1. Upper Triangular Matrix, U, is a square matrix such that all the elements under the main diagonal are equal to zero:

$$\mathbf{R} = (r_{\mu\nu}) \quad \text{with} \quad r_{\mu\nu} = 0 \quad \text{for} \quad \mu > \nu. \tag{4.17}$$

2. Lower Triangular Matrix, L, is a square matrix such that all the elements above the main diagonal are equal to zero:

$$\mathbf{L} = (l_{\mu\nu}) \quad \text{with} \quad l_{\mu\nu} = 0 \quad \text{for} \quad \mu < \nu. \tag{4.18}$$

4.1.3 Vectors

Matrices of size $(n, 1)$ are one-column matrices or *column vectors* of dimension n. Matrices of size $(1, n)$ are one-row matrices or *row vectors* of dimension n:

$$\textbf{Column Vector: } \underline{\mathbf{a}} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}, \tag{4.19a}$$

$$\textbf{Row Vector: } \underline{\mathbf{a}}^{\mathrm{T}} = (a_1, a_2, \ldots, a_n). \tag{4.19b}$$

By transposing, a column vector is changed into a row vector and conversely. A row or column vector

of dimension n can determine a point in the n dimensional Euclidean space $\mathbb{R}^n$.
The zero vector is denoted by $\underline{\mathbf{0}}$ or $\underline{\mathbf{0}}^{\mathrm{T}}$ respectively.

4.1.4 Arithmetical Operations with Matrices

1. Equality of Matrices
Two matrices $\mathbf{A} = (a_{\mu\nu})$ and $\mathbf{B} = (b_{\mu\nu})$ are equal if they have the same size and the corresponding elements are equal:

$$\mathbf{A} = \mathbf{B}, \quad \text{when} \quad a_{\mu\nu} = b_{\mu\nu} \quad \text{for} \quad \mu = 1, \dots, m;\ \nu = 1, \dots, n. \tag{4.20}$$

2. Addition and Subtraction
Matrices can be added or subtracted only if they have the same size. The sum/difference of two matrices is done by adding/subtracting the corresponding elements:

$$\mathbf{A} \pm \mathbf{B} = (a_{\mu\nu}) \pm (b_{\mu\nu}) = (a_{\mu\nu} \pm b_{\mu\nu}). \tag{4.21a}$$

■ $\begin{pmatrix} 1 & 3 & 7 \\ 2 & -1 & 4 \end{pmatrix} + \begin{pmatrix} 3 & -5 & 0 \\ 2 & 1 & 4 \end{pmatrix} = \begin{pmatrix} 4 & -2 & 7 \\ 4 & 0 & 8 \end{pmatrix}$.

For the addition of matrices the commutative law and the associative law are valid:

a) Commutative Law: $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$. (4.21b)

b) Associative Law: $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$. (4.21c)

3. Multiplication of a Matrix by a Number
A matrix $\mathbf{A}$ of size (m, n) is multiplied by a real or complex number α by multiplying every element of $\mathbf{A}$ by α:

$$\alpha\mathbf{A} = \alpha\,(a_{\mu\nu}) = (\alpha a_{\mu\nu}). \tag{4.22a}$$

■ $3\begin{pmatrix} 1 & 3 & 7 \\ 0 & -1 & 4 \end{pmatrix} = \begin{pmatrix} 3 & 9 & 21 \\ 0 & -3 & 12 \end{pmatrix}$.

From (4.22a) it is obvious that one can factor out a constant multiplier contained by every element of a matrix. For the multiplication of a matrix by a scalar the *commutative, associative and distributive laws for multiplication* are valid:

a) Commutative Law: $\alpha\mathbf{A} = \mathbf{A}\alpha$; (4.22b)

b) Associative Law: $\alpha(\beta\mathbf{A}) = (\alpha\beta)\mathbf{A}$; (4.22c)

c) Distributive Law: $(\alpha \pm \beta)\mathbf{A} = \alpha\mathbf{A} \pm \beta\mathbf{A}$; $\quad \alpha(\mathbf{A} \pm \mathbf{B}) = \alpha\mathbf{A} \pm \alpha\mathbf{B}$. (4.22d)

4. Division of a Matrix by a Number
The *division of a matrix by a scalar* $\gamma \neq 0$ is the same as multiplication by $\alpha = 1/\gamma$.

5. Multiplication of Two Matrices
1. The Product A B of two matrices $\mathbf{A}$ and $\mathbf{B}$ can be calculated only if the number of columns of the factor $\mathbf{A}$ on the left-hand side is equal to the number of rows of the factor $\mathbf{B}$ on the right-hand side. If $\mathbf{A}$ is a matrix of size (m, n), then the matrix $\mathbf{B}$ must have size (n, p), and the product $\mathbf{A}\,\mathbf{B}$ is a matrix $\mathbf{C} = (c_{\mu\lambda})$ of size (m, p). The element $c_{\mu\lambda}$ is equal to the scalar product of the μ-th row of the factor $\mathbf{A}$ on the left with the λ-th column of the factor $\mathbf{B}$ on the right:

$$\mathbf{A}\,\mathbf{B} = (\sum_{\nu=1}^{n} a_{\mu\nu} b_{\nu\lambda}) = (c_{\mu\lambda}) = \mathbf{C} \qquad (\mu = 1, 2, \dots, m;\ \lambda = 1, 2, \dots, p). \tag{4.23}$$

■ $\mathbf{A} = \begin{pmatrix} 1 & 3 & 7 \\ \boxed{2 & -1 & 4} \\ -1 & 0 & 1 \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} 3 & \boxed{2} \\ -5 & 1 \\ 0 & 3 \end{pmatrix}$. The element c_{22} of the product matrix C in accordance with (4.23) is $c_{22} = 2 \cdot 2 - 1 \cdot 1 + 4 \cdot 3 = 15$.

2. Inequality of Matrix Products Even if both products $\mathbf{A}\,\mathbf{B}$ and $\mathbf{B}\,\mathbf{A}$ exist, usually $\mathbf{A}\,\mathbf{B} \neq \mathbf{B}\,\mathbf{A}$, i.e., in general the commutative law for multiplication is not valid. If the equality $\mathbf{A}\,\mathbf{B} = \mathbf{B}\,\mathbf{A}$ holds, then one says that the matrices $\mathbf{A}$ and $\mathbf{B}$ are commutable or commute with each other.

3. Falk Scheme Multiplication of matrices $\mathbf{A}\,\mathbf{B} = \mathbf{C}$ can be performed using the *Falk scheme* **(Fig. 4.1)**. The element $c_{\mu\lambda}$ of the product matrix $\mathbf{C}$ appears exactly at the intersection point of the μ-th row of $\mathbf{A}$ with the λ-th column of $\mathbf{B}$.

■ Multiplication of the matrices $\mathbf{A}_{(3,3)}$ and $\mathbf{B}_{(3,2)}$ is shown in **Fig. 4.2** using the Falk scheme.

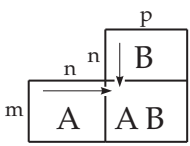

Figure 4.1

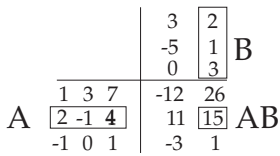

Figure 4.2

4. Multiplication of the Matrices $\mathbf{K_1}$ and $\mathbf{K_2}$ with Complex Elements For multiplication of two matrices with complex elements can be used their decompositions into real and imaginary parts according to (4.2b): $\mathbf{K}_1 = \mathbf{A}_1 + \mathrm{i}\mathbf{B}_1$, $\mathbf{K}_2 = \mathbf{A}_2 + \mathrm{i}\mathbf{B}_2$. Here $\mathbf{A_1}, \mathbf{A_2}, \mathbf{B_1}, \mathbf{B_2}$ are real matrices. After this decomposition, the multiplication results in a sum of matrices whose terms are products of real matrices.

■ $(\mathbf{A} + \mathrm{i}\,\mathbf{B})(\mathbf{A} - \mathrm{i}\,\mathbf{B}) = \mathbf{A}^2 + \mathbf{B}^2 + \mathrm{i}\,(\mathbf{B}\mathbf{A} - \mathbf{A}\mathbf{B})$ (Powers of Matrices see 4.1.5,**8.**, p. 276). Of course when multiplying these matrices it must be considered that the commutative law for multiplication is not valid in general, i.e., the matrices $\mathbf{A}$ and $\mathbf{B}$ do not usually commute with each other.

6. Scalar and Dyadic Product of Two Vectors

If the vectors $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$ are considered as one-row and one-column matrices, respectively, then there are two possibilities to multiply them according to the rules of matrix multiplication:
If $\underline{\mathbf{a}}$ has size $(1, n)$ and $\underline{\mathbf{b}}$ has size $(n, 1)$ then their product has size $(1, 1)$, i.e. it is a number. It is called the *scalar product* of two vectors. If conversely, $\underline{\mathbf{a}}$ has size $(n, 1)$ and $\underline{\mathbf{b}}$ has size $(1, m)$, then the product has size (n, m), i.e., it is a matrix. This matrix is called the *dyadic product* of the two vectors.

1. Scalar Product of Two Vectors The scalar product of a row vector $\underline{\mathbf{a}}^{\mathrm{T}} = (a_1, a_2, \ldots, a_n)$ with a column vector $\underline{\mathbf{b}} = (b_1, b_2, \ldots, b_n)^{\mathrm{T}}$ – both having n elements – is defined as the number

$$\underline{\mathbf{a}}^{\mathrm{T}}\underline{\mathbf{b}} = \underline{\mathbf{b}}^{\mathrm{T}}\underline{\mathbf{a}} = a_1b_1 + a_2b_2 + \cdots + a_nb_n = \sum_{\mu=1}^{n} a_\mu b_\mu. \tag{4.24}$$

The commutative law for multiplication is not valid for a product of vectors in general, so one must keep the exact order of $\underline{\mathbf{a}}^{\mathrm{T}}$ and $\underline{\mathbf{b}}$. If the order of multiplication is reversed, then the product $\underline{\mathbf{b}}\underline{\mathbf{a}}^{\mathrm{T}}$ is a dyadic product.

2. Dyadic Product or Tensor Product of Two Vectors The *dyadic product* of a column vector $\underline{\mathbf{a}} = (a_1, a_2, \ldots, a_n)^{\mathrm{T}}$ of dimension n with a row vector $\underline{\mathbf{b}}^{\mathrm{T}} = (b_1, b_2, \ldots, b_m)$ of dimension m is defined as the following matrix:

$$\underline{\mathbf{a}}\underline{\mathbf{b}}^{\mathrm{T}} = \begin{pmatrix} a_1b_1 & a_1b_2 & \cdots & a_1b_m \\ a_2b_1 & a_2b_2 & \cdots & a_2b_m \\ \vdots & \vdots & \vdots & \vdots \\ a_nb_1 & a_nb_2 & \cdots & a_nb_m \end{pmatrix} \tag{4.25}$$

of size (n, m). Also here the commutative law for multiplication is not valid in general.

3. Hints on the Notion of Vector Products of Two Vectors In the domain of multi-vectors or alternating tensors there is a so-called outer product whose three-dimensional version is the well-

known *vector product* or *cross product* (see 3.5.1.5, **2.**, p. 184 ff). In this book the outer product of multi-vectors of higher rank is not discussed.

7. Rank of a Matrix

1. Definition In a matrix $\mathbf{A}$ the maximum number r of linearly independent column vectors is equal to the maximum number of linearly independent row vectors. This number r is called the *rank of the matrix* and it is denoted by $\operatorname{rank}(\mathbf{A}) = r$.

2. Statements about the Rank of a Matrix

a) Because in a vector space of dimension m there exist no more than m linearly independent m-dimensional row or column vectors (see 5.3.8.2, p. 366), the rank r of a matrix $\mathbf{A}$ of size (m, n) cannot be greater, than the smaller of m and n:

$$\operatorname{rank}(\mathbf{A}_{(m,n)}) = r \leq \min(m, n). \tag{4.26a}$$

b) A square matrix $\mathbf{A}_{(n,n)}$ is called a *regular matrix* if

$$\operatorname{rank}(\mathbf{A}_{(n,n)}) = r = n. \tag{4.26b}$$

A square matrix of size (n, n) is regular if and only if its determinant differs from zero, i.e., $\det \mathbf{A} \neq 0$ (see 4.2.2, **3.**, p. 279). Otherwise it is a *singular matrix*.

c) Consequently for the rank of a singular square matrix $\mathbf{A}_{(n,n)}$, i.e., $\det \mathbf{A} = 0$

$$\operatorname{rank}(\mathbf{A}_{(n,n)}) = r < n \tag{4.26c}$$

is valid.

d) The rank of the zero matrix $\mathbf{0}$ is equal to zero:

$$\operatorname{rank}(\mathbf{0}) = r = 0. \tag{4.26d}$$

e) The rank of the sum and product of matrices satisfies the relations

$$|\operatorname{rank}(\mathbf{A}) - \operatorname{rank}(\mathbf{B})| \leq \operatorname{rank}(\mathbf{A} + \mathbf{B}) \leq \operatorname{rank}(\mathbf{A}) + \operatorname{rank}(\mathbf{B}), \tag{4.26e}$$

$$\operatorname{rank}(\mathbf{AB}) \leq \min(\operatorname{rank}(\mathbf{A}), \operatorname{rank}(\mathbf{B})). \tag{4.26f}$$

3. Rules to Determine the Rank Elementary transformations do not change the rank of matrices. *Elementary transformations* in this relation are:

a) Interchanging two columns or two rows.

b) Multiplication of a row or column by a number.

c) Addition of a row to another row or a column to an other column.

In order to determine their ranks every matrix can be transformed by appropriate linear combinations of rows into a form such that in the μ-th row ($\mu = 2, 3, \ldots, m$), at least the first $\mu - 1$ elements are equal to zero (the principle of Gauss algorithm, see 4.5.2.4, p. 312). The number of row vectors different from the zero vector in the transformed matrix is equal to the rank r of the matrix.

8. Inverse Matrix

For a regular matrix $\mathbf{A} = (a_{\mu\nu})$ there is always an *inverse matrix* $\mathbf{A}^{-1}$ (with respect to multiplication), i.e., the multiplication of a matrix by its inverse yields the identity matrix:

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}. \tag{4.27a}$$

The elements of $\mathbf{A}^{-1} = (\beta_{\mu\nu})$ are

$$\beta_{\mu\nu} = \frac{\mathbf{A}_{\nu\mu}}{\det \mathbf{A}}, \tag{4.27b}$$

where $\mathbf{A}_{\nu\mu}$ is the cofactor belonging to the $a_{\nu\mu}$ element of the matrix $\mathbf{A}$ (see 4.2.1, **1.**, p. 278). For a practical calculation of $\mathbf{A}^{-1}$ the method given in 4.2.2, **2.**, p. 278 should be used. In the case of a matrix of size $(2, 2)$ holds:

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad \mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \tag{4.28}$$

Remark: Why not define division among matrices but instead use the inverse for calculations? This is connected to the fact that division cannot be defined uniquely. The solutions of the equations

$$\begin{array}{l}\mathbf{B}\mathbf{X_1} = \mathbf{A} \\ \mathbf{X_2}\mathbf{B} = \mathbf{A}\end{array} \quad (\mathbf{B}\text{ regular}), \qquad \begin{array}{l}\mathbf{X_1} = \mathbf{B^{-1}}\mathbf{A} \\ \mathbf{X_2} = \mathbf{A}\mathbf{B^{-1}}\end{array} \tag{4.29}$$

are in general different.

9. Orthogonal Matrices

If the relation

$$\mathbf{A}^{\mathrm{T}} = \mathbf{A^{-1}} \quad \text{or} \quad \mathbf{A}\mathbf{A}^{\mathrm{T}} = \mathbf{A}^{\mathrm{T}}\mathbf{A} = \mathbf{I} \tag{4.30}$$

holds for a square matrix $\mathbf{A}$, then it is called an *orthogonal matrix*, i.e., the scalar product of a row and the transpose of another one, or the scalar product of the transpose of a column and another one are zero, while the scalar product of a row with its own transpose or of the transpose of a column with itself are equal to one.

Orthogonal matrices have the following properties:

a) The transpose and the inverse of an orthogonal matrix $\mathbf{A}$ are also orthogonal; furthermore, the determinant is

$$\det \mathbf{A} = \pm 1. \tag{4.31}$$

b) Products of orthogonal matrices are also orthogonal.

■ The *rotation matrix* $\mathbf{D}$, which is used to describe the rotation of a coordinate system, and whose elements are the direction cosines of the new direction of axes (see 3.5.3.3,**2.** p. 212), is also an orthogonal matrix.

10. Unitary Matrix

If for a matrix $\mathbf{A}$ with complex elements

$$(\mathbf{A}^*)^{\mathrm{T}} = \mathbf{A^{-1}} \quad \text{or} \quad \mathbf{A}(\mathbf{A}^*)^{\mathrm{T}} = (\mathbf{A}^*)^{\mathrm{T}}\mathbf{A} = \mathbf{I} \tag{4.32}$$

holds it is called a *unitary matrix*. In the real case unitary and orthogonal matrices are the same.

4.1.5 Rules of Calculation for Matrices

The following rules are valid of course only in the case when the operations can be performed, for instance the identity matrix $\mathbf{I}$ always has a size corresponding to the requirements of the given operation.

1. Multiplication of a Matrix by the Identity Matrix

is also called the *identical transformation*:

$$\mathbf{A}\mathbf{I} = \mathbf{I}\mathbf{A} = \mathbf{A}. \tag{4.33}$$

(This does not mean that the commutative law is valid in general, because the sizes of the matrix $\mathbf{I}$ on the left- and on the right-hand side may be different.)

2. Multiplication of a Square Matrix A by a Scalar Matrix S

or by the identity matrix $\mathbf{I}$ is commutative

$$\mathbf{S}\mathbf{A} = \mathbf{A}\mathbf{S} = c\mathbf{A} \quad \text{with } \mathbf{S} \text{ given in (4.8)}, \quad (4.34a) \qquad \mathbf{A}\mathbf{I} = \mathbf{I}\mathbf{A} = \mathbf{A}. \tag{4.34b}$$

3. Multiplication of a Matrix A by the Zero Matrix 0

results in the zero matrix:

$$\mathbf{A}\mathbf{0} = \mathbf{0}\,, \qquad \mathbf{0}\mathbf{A} = \mathbf{0}. \tag{4.35}$$

(The zero matrices above may have different sizes.) The converse statement is not true in general, i.e., from $\mathbf{A}\mathbf{B} = \mathbf{0}$ it does not follow that $\mathbf{A} = \mathbf{0}$ or $\mathbf{B} = \mathbf{0}$.

4. Vanishing Product of Two Matrices

The product of two matrices $\mathbf{A}$ and $\mathbf{B}$ can be the zero matrix even if neither of them is a zero matrix:

$$\mathbf{A}\,\mathbf{B} = \mathbf{0} \text{ or } \mathbf{B}\,\mathbf{A} = \mathbf{0} \text{ or both, although } \mathbf{A} \neq \mathbf{0},\ \mathbf{B} \neq \mathbf{0}. \tag{4.36}$$

■ $\begin{array}{cc|cc} & & 1 & 1 \\ & & 0 & 0 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array}$.

5. Multiplication of Three Matrices

$$(\mathbf{A}\,\mathbf{B})\mathbf{C} = \mathbf{A}\,(\mathbf{B}\,\mathbf{C}) \tag{4.37}$$

i.e., the associative law of multiplication is valid.

6. Transposing of a Sum or a Product of Two Matrices

$$(\mathbf{A}+\mathbf{B})^{\mathrm{T}} = \mathbf{A}^{\mathrm{T}} + \mathbf{B}^{\mathrm{T}}, \qquad (\mathbf{AB})^{\mathrm{T}} = \mathbf{B}^{\mathrm{T}}\,\mathbf{A}^{\mathrm{T}}, \qquad (\mathbf{A}^{\mathrm{T}})^{\mathrm{T}} = \mathbf{A}. \tag{4.38a}$$

For square invertible matrices $\mathbf{A}_{(n,n)}$:

$$(\mathbf{A}^{\mathrm{T}})^{-\mathbf{1}} = (\mathbf{A}^{-\mathbf{1}})^{\mathrm{T}} \tag{4.38b}$$

holds.

7. Inverse of a Product of Two Matrices

$$(\mathbf{A}\,\mathbf{B})^{-\mathbf{1}} = \mathbf{B}^{-\mathbf{1}}\,\mathbf{A}^{-\mathbf{1}}. \tag{4.39}$$

8. Powers of Matrices

$$\mathbf{A}^{p} = \underbrace{\mathbf{A}\,\mathbf{A}\ldots\mathbf{A}}_{p\ \text{factors}} \ \text{with}\ p>0,\ \text{integer}, \tag{4.40a}$$

$$\mathbf{A}^{0} = \mathbf{I} \qquad (\det\mathbf{A}\neq 0)\,, \tag{4.40b}$$

$$\mathbf{A}^{-p} = (\mathbf{A}^{-\mathbf{1}})^{p} \qquad (p>0,\ \text{integer}; \det\mathbf{A}\neq 0), \tag{4.40c}$$

$$\mathbf{A}^{p+q} = \mathbf{A}^{p}\mathbf{A}^{q} \qquad (p,q\ \text{integer}). \tag{4.40d}$$

9. Kronecker Product

The *Kronecker product of two matrices* $\mathbf{A}=(a_{\mu\nu})$ of the type (m,n) and $\mathbf{B}=(b_{\mu\nu})$ of the type (p,r) is defined as the rule

$$\mathbf{A}\otimes\mathbf{B} = (a_{\mu\nu}\,\mathbf{B})\,. \tag{4.41}$$

The result is a new matrix of type $(m\cdot p\,,\ n\cdot r)$, arising from the multiplication of every element of $\mathbf{A}$ by the matrix $\mathbf{B}$.

■ $\mathbf{A} = \begin{pmatrix} 3 & -5 & 0 \\ 2 & 1 & 3 \end{pmatrix}$ of type $(2,3)$, $\mathbf{B} = \begin{pmatrix} 1 & 3 \\ 2 & -1 \end{pmatrix}$ of type $(2,2)$.

$$\mathbf{A}\bigotimes\mathbf{B} = \begin{pmatrix} 3\cdot\begin{pmatrix}1&3\\2&-1\end{pmatrix} & -5\cdot\begin{pmatrix}1&3\\2&-1\end{pmatrix} & 0\cdot\begin{pmatrix}1&3\\2&-1\end{pmatrix} \\ 2\cdot\begin{pmatrix}1&3\\2&-1\end{pmatrix} & 1\cdot\begin{pmatrix}1&3\\2&-1\end{pmatrix} & 3\cdot\begin{pmatrix}1&3\\2&-1\end{pmatrix} \end{pmatrix} = \begin{pmatrix} 3 & 9 & -5 & -15 & 0 & 0 \\ 6 & -3 & -10 & 5 & 0 & 0 \\ 2 & 6 & 1 & 3 & 3 & 9 \\ 4 & -2 & 2 & -1 & 6 & -3 \end{pmatrix},$$

gives a matrix of type $(4,6)$.

For the transpose and the trace are valid the equalities:

$$(\mathbf{A}\otimes\mathbf{B})^{\mathrm{T}} = \mathbf{A}^{\mathrm{T}}\otimes\mathbf{B}^{\mathrm{T}}, \tag{4.42}$$

$$\mathrm{Tr}\,(\mathbf{A}\otimes\mathbf{B}) = \mathrm{Tr}\,(\mathbf{A})\cdot\mathrm{Tr}(\mathbf{B})\,. \tag{4.43}$$

10. Differentiation of a Matrix

If a matrix $\mathbf{A}=\mathbf{A}(t)=(a_{\mu\nu}(t))$ has differentiable elements $a_{\mu\nu}(t)$ of a parameter t then its derivative with respect to t is given as

$$\frac{d\mathbf{A}}{dt} = \left(\frac{da_{\mu\nu}(t)}{dt}\right) = (a'_{\mu\nu}(t))\,. \tag{4.44}$$

4.1.6 Vector and Matrix Norms

The norm of a vector or of a matrix can be considered as a generalization of the absolute value (magnitude) of numbers. Therefore a real number is assigned as $||x||$ (*Norm* $\underline{\mathbf{x}}$) to the vector $\underline{\mathbf{x}}$ or as $||\mathbf{A}||$

(*Norm* **A**) to a matrix **A**. These numbers must satisfy the norm axioms (see 12.3.1.1, p. 669). For vectors $\underline{\mathbf{x}} \in \mathbb{R}^n$ they are:

1. $||\underline{\mathbf{x}}|| \geq 0$ for every $\underline{\mathbf{x}}$; $\quad ||\underline{\mathbf{x}}|| = 0 \quad$ if and only if $\underline{\mathbf{x}} = 0$. (4.45)

2. $||\lambda\underline{\mathbf{x}}|| = |\lambda|\,||\underline{\mathbf{x}}||$ for every $\underline{\mathbf{x}}$ and every real number λ. (4.46)

3. $||\underline{\mathbf{x}} + \underline{\mathbf{y}}|| \leq ||\underline{\mathbf{x}}|| + ||\underline{\mathbf{y}}||$ for every $\underline{\mathbf{x}}$ and $\underline{\mathbf{y}}$ (triangle inequality) (see also 3.5.1.1, **1.**, p. 182). (4.47)

There are many different ways to define norms for vectors and matrices. But for practical reasons it is better to define a matrix norm $||\mathbf{A}||$ and a vector norm $||\underline{\mathbf{x}}||$ so that they might satisfy the inequality

$$||\mathbf{A}\underline{\mathbf{x}}|| \leq ||\mathbf{A}||\,||\underline{\mathbf{x}}||. \tag{4.48}$$

This inequality is very useful for error estimations. If the matrix and vector norms satisfy this inequality, then one says that they are *consistent with each other*. If there is a non-zero vector $\underline{\mathbf{x}}$ for every **A** such that the equality holds in (4.48), then one says *the matrix norm* $||\mathbf{A}||$ is the *subordinate to the vector norm* $||\underline{\mathbf{x}}||$.

4.1.6.1 Vector Norms

If $\underline{\mathbf{x}} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$ is a real vector of n dimensions, i.e., $\underline{\mathbf{x}} \in \mathbb{R}^n$, then the most often used vector norms are:

1. Euclidean Norm

$$||\underline{\mathbf{x}}|| = ||\underline{\mathbf{x}}||_2 := \sqrt{\sum_{i=1}^{n} x_i^2}. \tag{4.49}$$

2. Supremum or Uniform Norm

$$||\underline{\mathbf{x}}|| = ||\underline{\mathbf{x}}||_\infty := \max_{1 \leq i \leq n} |x_i|. \tag{4.50}$$

3. Sum Norm

$$||\underline{\mathbf{x}}|| = ||\underline{\mathbf{x}}||_1 := \sum_{i=1}^{n} |\underline{\mathbf{x}}_i|. \tag{4.51}$$

■ In $\mathbb{R}^3$, in elementary vector calculus $||\underline{\mathbf{x}}||_2$ is considered as the magnitude of the vector $\underline{\mathbf{x}}$. The magnitude $|x| = ||\underline{\mathbf{x}}||_2$ gives the length of the vector $\underline{\mathbf{x}}$.

4.1.6.2 Matrix Norms

1. Spectral Norm for Real Matrices

$$||\mathbf{A}|| = ||\mathbf{A}||_2 := \sqrt{\lambda_{\max}(\mathbf{A}^{\mathrm{T}}\mathbf{A})}. \tag{4.52}$$

Here $\lambda_{\max}(\mathbf{A}^{\mathrm{T}}\mathbf{A})$ denotes the greatest eigenvalue (see 4.6.1, p. 314) of the matrix $\mathbf{A}^{\mathrm{T}}\mathbf{A}$.

2. Row-Sum Norm

$$||\mathbf{A}|| = ||\mathbf{A}||_\infty := \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}|. \tag{4.53}$$

3. Column-Sum Norm

$$||\mathbf{A}|| = ||\mathbf{A}||_1 := \max_{1 \leq j \leq n} \sum_{i=1}^{n} |a_{ij}|. \tag{4.54}$$

It can be proved that the matrix norm (4.52) is the subordinate norm to the vector norm (4.49). The same is true for (4.53) and (4.50), and for (4.54) and (4.51).

4.2 Determinants

4.2.1 Definitions

1. Determinants

Determinants are real or complex numbers uniquely associated with square matrices. The *determinant* of order n associated with the (n, n) matrix $\mathbf{A} = (a_{\mu\nu})$,

$$\mathrm{D} = \det \mathbf{A} = \det (a_{\mu\nu}) = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}, \tag{4.55}$$

is calculated in a recursive way using the *Laplace expansion rule*:

$$\det \mathbf{A} = \sum_{\nu=1}^{n} a_{\mu\nu} A_{\mu\nu} \qquad (\mu \text{ fixed, expansion along the } \mu\text{-th row}), \tag{4.56a}$$

$$\det \mathbf{A} = \sum_{\mu=1}^{n} a_{\mu\nu} A_{\mu\nu} \qquad (\nu \text{ fixed, expansion along the } \nu\text{-th column}). \tag{4.56b}$$

Here $A_{\mu\nu}$ is the subdeterminant belonging to the element $a_{\mu\nu}$ multiplied by the sign factor $(-1)^{\mu+\nu}$. $A_{\mu\nu}$ is called the *cofactor* or *algebraic complement*.

2. Subdeterminants

The *subdeterminant* of order $(n-1)$ belonging to the element $a_{\mu\nu}$ of a determinant of order n is the determinant obtained by deleting the μ-th row and the ν-th column.

■ Expansion of a determinant of order four along the third row:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{vmatrix} = a_{31} \begin{vmatrix} a_{12} & a_{13} & a_{14} \\ a_{22} & a_{23} & a_{24} \\ a_{42} & a_{43} & a_{44} \end{vmatrix} - a_{32} \begin{vmatrix} a_{11} & a_{13} & a_{14} \\ a_{21} & a_{23} & a_{24} \\ a_{41} & a_{43} & a_{44} \end{vmatrix} + a_{33} \begin{vmatrix} a_{11} & a_{12} & a_{14} \\ a_{21} & a_{22} & a_{24} \\ a_{41} & a_{42} & a_{44} \end{vmatrix} - a_{34} \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{41} & a_{42} & a_{43} \end{vmatrix}.$$

4.2.2 Rules of Calculation for Determinants

Because of the Laplace expansion the following statements about rows are valid also for columns.

1. Independence of the Value of a Determinant

The value of a determinant does not depend on which row was chosen.

2. Substitution of Cofactors

If during the expansion of a determinant the cofactors of a row are replaced by the cofactors of another one, then one gets zero:

$$\sum_{\nu=1}^{n} a_{\mu\nu} A_{\lambda\nu} = 0 \qquad (\mu, \lambda \text{ fixed}; \lambda \neq \mu)\,. \tag{4.57}$$

This relation and the Laplace expansion result in

$$\mathbf{A}_{\mathrm{adj}}\, \mathbf{A} = \mathbf{A}\, \mathbf{A}_{\mathrm{adj}} = (\det \mathbf{A})\, \mathbf{I}\,. \tag{4.58}$$

The *adjoint matrix* of $\mathbf{A}$, which is the transpose of the matrix made from the cofactors of $\mathbf{A}$, is denoted by $\mathbf{A}_{\mathrm{adj}}$. There must not be a confusion of this adjoint matrix with the transposed conjugate of a complex matrix $\mathbf{A}^{\mathrm{H}}$ (see (4.4), p. 269). From the previous equality one gets the *inverse matrix*

$$\mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}} \mathbf{A}_{\mathrm{adj}}\,, \tag{4.59}$$

3. Zero Value of a Determinant

A determinant is equal to zero if

a) a row contains zero elements only, or

b) two rows are equal to each other, or

c) a row is a linear combination of the others.

4. Changes and Additions

The value of the determinant does not change if

a) its rows are exchanged for its columns, i.e., *reflection* in the main diagonal does not affect the value of it:

$$\det \mathbf{A} = \det \mathbf{A}^{\mathrm{T}}, \tag{4.60}$$

b) any row is added to or subtracted from another one, or

c) a multiple of any row is added to or subtracted from another one, or

d) a linear combination of other rows is added to any row.

5. Sign on Changing Rows

If two rows are interchanged in a determinant, then the sign of the determinant changes.

6. Multiplication of a Determinant by a Number

The value of a determinant will be multiplied by α if the elements of a row are multiplied by this number. The next formula shows the difference between this and the multiplication of a matrix $\mathbf{A}$ of size (n, n) by a number α

$$\det(\alpha \mathbf{A}) = \alpha^n \det \mathbf{A}\,. \tag{4.61}$$

7. Multiplication of Two Determinants

The multiplication of two determinants can be reduced to the multiplication of their matrices:

$$(\det \mathbf{A})(\det \mathbf{B}) = \det(\mathbf{AB}). \tag{4.62}$$

Since $\det \mathbf{A} = \det \mathbf{A}^{\mathrm{T}}$ (see (4.60)), we have the equalities

$$(\det \mathbf{A})(\det \mathbf{B}) = \det(\mathbf{AB}) = \det(\mathbf{AB}^{\mathrm{T}}) = \det(\mathbf{A}^{\mathrm{T}}\mathbf{B}) = \det(\mathbf{A}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}), \tag{4.63}$$

i.e., it is permissible to take the scalar product of rows with columns, rows with rows, columns with rows or columns with columns.

8. Differentiation of a Determinant

Suppose the elements of a determinant of order n are differentiable functions of a parameter t, i.e., $a_{\mu\nu} = a_{\mu\nu}(t)$. In order to differentiate the determinant with respect to t, one differentiates one row at one time and finally one adds the n determinants.

■ For a determinant of size $(3, 3)$ follows:

$$\frac{d}{dt}\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a'_{11} & a'_{12} & a'_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a'_{21} & a'_{22} & a'_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a'_{31} & a'_{32} & a'_{33} \end{vmatrix}.$$

4.2.3 Evaluation of Determinants

1. Value of a Determinant of Second Order

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12}. \tag{4.64}$$

2. Value of a Determinant of Third Order

The *Sarrus rule* gives a convenient scheme for the calculations, but it is valid only for determinants of order three. It is the following:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{matrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32}$$
$$-(a_{31}a_{22}a_{13} + a_{32}a_{23}a_{11} + a_{33}a_{21}a_{12}). \tag{4.65}$$

The first two columns are copied after the determinant, then the sum of the products of the elements along the redrawn declining segments is calculated, then the sum of the products of the elements along the dotted inclining segments is subtracted.

3. Value of a Determinant of n-th Order

By the expansion rule the calculation of the value of a determinant of order n is reduced to the evaluation of n determinants of order $(n-1)$. But for practical reasons (to reduce the number of required operations), first one transforms the determinant with the help of the rules discussed above into a form such that it contains as many zeros as possible.

■ $$\begin{vmatrix} 2 & 9 & 9 & 4 \\ 2 & -3 & 12 & 8 \\ 4 & 8 & 3 & -5 \\ 1 & 2 & 6 & 4 \end{vmatrix} \overset{\text{(rule 4)}}{=} \begin{vmatrix} 2 & 5 & 9 & 4 \\ 2 & -7 & 12 & 8 \\ 4 & 0 & 3 & -5 \\ 1 & 0 & 6 & 4 \end{vmatrix} \overset{\text{(rule 6)}}{=} 3 \begin{vmatrix} 2 & 5 & 3 & 4 \\ 2 & -7 & 4 & 8 \\ 4 & 0 & 1 & -5 \\ 1 & 0 & 2 & 4 \end{vmatrix} = 3 \left(\underbrace{-5 \begin{vmatrix} 2 & 4 & 8 \\ 4 & 1 & -5 \\ 1 & 2 & 4 \end{vmatrix}}_{=0 \text{ (rule 3)}} - 7 \begin{vmatrix} 2 & 3 & 4 \\ 4 & 1 & -5 \\ 1 & 2 & 4 \end{vmatrix} \right)$$
$$= -21 \underbrace{\begin{vmatrix} 1 & 1 & 0 \\ 4 & 1 & -5 \\ 1 & 2 & 4 \end{vmatrix}}_{\text{(rule 4)}} = -21 \left(\begin{vmatrix} 1 & -5 \\ 2 & 4 \end{vmatrix} - \begin{vmatrix} 4 & -5 \\ 1 & 4 \end{vmatrix} \right) = 147.$$

Remark: An especially efficient method to determine the value of a determinant of order n can be obtained by transforming it in the same way as it is done in order to determine the rank of a matrix (see 4.1.4, **7.**, p. 274), i.e., all the elements under the diagonal $a_{11}, a_{22}, \ldots, a_{nn}$ are equal to zero. Then the value of the determinant is the product of the diagonal elements of the transformed determinant.

4.3 Tensors

4.3.1 Transformation of Coordinate Systems

1. Linear Transformation

By the linear transformation

$$\underline{\tilde{\mathbf{x}}} = \mathbf{A}\underline{\mathbf{x}} \qquad \text{or} \qquad \begin{aligned} \tilde{x}_1 &= a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ \tilde{x}_2 &= a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \\ \tilde{x}_3 &= a_{31}x_1 + a_{32}x_2 + a_{33}x_3 \end{aligned} \tag{4.66}$$

a coordinate transformation is defined in the three-dimensional space. Here x_μ and $\tilde{x}_\mu$ $(\mu = 1, 2, 3)$ are the coordinates of the same point but in different coordinate systems K and $\tilde{K}$.

2. Einstein's Summation Convention

Instead of (4.66) one can write

$$\tilde{x}_\mu = \sum_{\nu=1}^{3} a_{\mu\nu}x_\nu \qquad (\mu = 1, 2, 3) \tag{4.67a}$$

or briefly by Einstein

$$x_\mu = a_{\mu\nu}x_\nu, \tag{4.67b}$$

i.e., it is to calculate the sum with respect to the repeated index ν and put down the result for $\mu = 1, 2, 3$. In general, the *summation convention* means that if an index appears twice in an expression, then the expression is added for all values of this index. If an index appears only once in the expressions of an

equation, for instance μ in (4.67b), then it means that the equality is valid for all possible values of this index.

3. Rotation of a Coordinate System

If the Cartesian coordinate system $\tilde{K}$ is given by rotation of the system K, then for the transformation matrix in (4.66) $\mathbf{A} = \mathbf{D}$ is valid. Here $\mathbf{D} = (d_{\mu\nu})$ is the orthogonal *rotation matrix*. The orthogonal rotation matrix $\mathbf{D}$ has the property

$$\mathbf{D}^{-1} = \mathbf{D}^{\mathrm{T}}. \tag{4.68a}$$

The elements $d_{\mu\nu}$ of $\mathbf{D}$ are the direction cosines of the angles between the old and new coordinate axes. From the orthogonality of $\mathbf{D}$, i.e., from

$$\mathbf{D}\mathbf{D}^{\mathrm{T}} = \mathbf{I} \qquad \text{and} \qquad \mathbf{D}^{\mathrm{T}}\mathbf{D} = \mathbf{I}, \tag{4.68b}$$

it follows that

$$\sum_{i=1}^{3} d_{\mu i} d_{\nu i} = \delta_{\mu\nu}, \qquad \sum_{k=1}^{3} d_{k\mu} d_{k\nu} = \delta_{\mu\nu} \qquad (\mu, \nu = 1, 2, 3). \tag{4.68c}$$

The equalities in (4.68c) show that the row and column vectors of the matrix $\mathbf{D}$ are *orthonormalized*, because $\delta_{\mu\nu}$ is the Kronecker symbol (see 4.1.2, **10.**, p. 271).
The elements $d_{\mu\nu}$ of the rotation matrix can be determined by the Cardan angles (see 3.5.3.5, p. 214) or Euler angles (see 3.5.3.6, p. 215). For rotation in the plane see 3.5.2.2,**2.**, p. 191; in space see 3.5.3.3, p. 213.

4.3.2 Tensors in Cartesian Coordinates

1. Definition

A mathematical or a physical quantity $\boldsymbol{T}$ can be described in a Cartesian coordinate system K by 3^n elements $t_{ij\cdots m}$, the so-called translation invariants. Here the number of indices $i, j, \ldots, m$ is exactly equal to n $(n \geq 0)$. The indices are ordered, and every of them takes the values 1, 2 and 3.

If under a coordinate transformation from K to $\tilde{K}$ for the elements $t_{ij\cdots m}$ according to (4.66)

$$\tilde{t}_{\mu\nu\cdots\rho} = \sum_{i=1}^{3}\sum_{j=1}^{3}\cdots\sum_{m=1}^{3} a_{\mu i} a_{\nu j} \cdots a_{\rho m} t_{ij\cdots m}\,, \tag{4.69}$$

is valid, then $\boldsymbol{T}$ is called a *tensor of rank* n, and the elements $t_{ij\cdots m}$ (mostly numbers) with ordered indices are the *components of the tensor* $\boldsymbol{T}$**.**

2. Tensor of Rank 0

A *tensor of rank zero* has only one component, i.e., it is a scalar. Because its value is the same in every coordinate system, one talks about the *invariance of scalars* or about an *invariant scalar*.

3. Tensor of Rank 1

A *tensor of rank* 1 has three components t_1, t_2 and t_3. The transformation law (4.69) is now

$$\tilde{t}_\mu = \sum_{i=1}^{3} a_{\mu i} t_i \qquad (\mu = 1, 2, 3). \tag{4.70}$$

It is the transformation law for vectors, i.e., a vector is a tensor of rank 1.

4. Tensor of Rank 2

If $n = 2$, then the tensor $\boldsymbol{T}$ has nine components t_{ij}, which can be arranged in a matrix

$$\boldsymbol{T} \;= \mathbf{T} = \begin{pmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{pmatrix}. \tag{4.71a}$$

The transformation law (4.70) is now:

$$\tilde{t}_{\mu\nu} = \sum_{i=1}^{3}\sum_{j=1}^{3} a_{\mu i}a_{\nu j}t_{ij} \qquad (\mu,\nu = 1,2,3). \tag{4.71b}$$

So, a tensor of rank 2 can be represented as a matrix.

■ **A:** The moment of inertia Θ_g of a solid with respect to the line g, which goes through the origin and has direction vector $\vec{\mathbf{a}} = \underline{\mathbf{a}}^{\mathrm{T}}$, can be represented in the form

$$\Theta_g = \underline{\mathbf{a}}^{\mathrm{T}}\boldsymbol{\Theta}\underline{\mathbf{a}} \tag{4.72a}$$

$$\text{with} \quad \boldsymbol{\Theta} = (\Theta_{ij}) = \begin{pmatrix} \Theta_x & -\Theta_{xy} & -\Theta_{xz} \\ -\Theta_{xy} & \Theta_y & -\Theta_{yz} \\ -\Theta_{xz} & -\Theta_{yz} & \Theta_z \end{pmatrix}, \tag{4.72b}$$

the so-called *inertia tensor*. Here Θ_x, Θ_y and Θ_z are the moments of inertia with respect to the coordinate axes, and Θ_{xy}, Θ_{xz} and Θ_{yz} are the *deviation moments* with respect to the coordinate axes.

■ **B:** The load-up conditions of an elastically deformed body can be given by the tension tensor

$$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix}. \tag{4.73}$$

The elements σ_{ik} $(i,k = 1,2,3)$ are determined in the following way: At a point P of the elastic body a small plane surface element is chosen whose normal vector points to the direction of the x_1-axis of a right-angle Cartesian coordinate system. The power per surface unit on this element, depending on the material, is a vector with coordinates σ_{11}, σ_{12} and σ_{13}. The other components can be explained similarly.

5. Rules of Calculation

1. Elementary Algebraic Operations The multiplication of a tensor by a number, and addition and subtraction of tensors of the *same rank* are defined componentwise, similarly to the corresponding operations for vectors and matrices.

2. Tensor Product Suppose there are given a tensor $\boldsymbol{A}$ of rank m and a tensor $\boldsymbol{B}$ of rank n with components $a_{ij\ldots}$ and $b_{rs\ldots}$ respectively. Then the 3^{m+n} scalars

$$c_{ij\ldots rs\ldots} = a_{ij\ldots}b_{rs\ldots} \tag{4.74a}$$

give the components of a tensor $\boldsymbol{C}$ of rank $m+n$. It is denoted by $\boldsymbol{C} = \boldsymbol{AB}$ and it is called the *tensor product* of $\boldsymbol{A}$ and $\boldsymbol{B}$. The associative and distributive laws are valid:

$$(\boldsymbol{AB})\,\boldsymbol{C} = \boldsymbol{A}\,(\boldsymbol{BC}), \qquad \boldsymbol{A}\,(\boldsymbol{B}+\boldsymbol{C}) = \boldsymbol{AB}+\boldsymbol{AC}. \tag{4.74b}$$

3. Dyadic Product The product of two tensors of rank 1 $\boldsymbol{A} = (a_1,a_2,a_3)$ and $\boldsymbol{B} = (b_1,b_2,b_3)$ gives a tensor of rank 2 with the elements

$$c_{ij} = a_i b_j \qquad (i,j = 1,2,3), \tag{4.75a}$$

i.e., the tensor product results in the matrix

$$\begin{pmatrix} a_1b_1 & a_1b_2 & a_1b_3 \\ a_2b_1 & a_2b_2 & a_2b_3 \\ a_3b_1 & a_3b_2 & a_3b_3 \end{pmatrix}. \tag{4.75b}$$

This will be denoted as the *dyadic product* of the two vectors $\underline{\mathbf{A}}$ and $\underline{\mathbf{B}}$.

4. Contraction Setting two indices equal to each other in a tensor of rank m $(m \geq 2)$, and summing with respect to them, then one gets a tensor of rank $m-2$, which is called the *contraction* of the tensor.

■ The tensor $\boldsymbol{C}$ of rank 2 of (4.75a) with $c_{ij} = a_i b_j$, which is the tensor product of the vectors $\underline{\mathbf{A}} = (a_1,a_2,a_3)$ and $\underline{\mathbf{B}} = (b_1,b_2,b_3)$, can be contracted by the indices i and j,

$$a_i b_i = a_1b_1 + a_2b_2 + a_3b_3 \tag{4.76}$$

giving a scalar, which is a tensor of rank 0. This gives the scalar product of vectors $\underline{\mathbf{A}}$ and $\underline{\mathbf{B}}$.

4.3.3 Tensors with Special Properties

4.3.3.1 Tensors of Rank 2

1. Rules of Calculation

For tensors of rank 2 the same rules are valid as for matrices. In particular, every tensor $\boldsymbol{T}$ can be decomposed into the sum of a symmetric and a skew-symmetric tensor:

$$\boldsymbol{T} = \frac{1}{2}\left(\boldsymbol{T} + \boldsymbol{T}^{\mathrm{T}}\right) + \frac{1}{2}\left(\boldsymbol{T} - \boldsymbol{T}^{\mathrm{T}}\right). \tag{4.77a}$$

A tensor $\boldsymbol{T} = (t_{ij})$ is called *symmetric* if

$$t_{ij} = t_{ji} \qquad \text{for all } i \text{ and } j \tag{4.77b}$$

holds. In the case

$$t_{ij} = -t_{ji} \qquad \text{for all } i \text{ and } j \tag{4.77c}$$

it is called *skew- or antisymmetric*. Obviously the elements t_{11}, t_{22} and t_{33} of a skew-symmetric tensor are equal to zero. The notion of symmetry and antisymmetry can be extended for tensors of higher rank if refering to certain pairs of elements.

2. Transformation of Principal Axes

For a symmetric tensor $\boldsymbol{T}$, i.e., if $t_{\mu\nu} = t_{\nu\mu}$ holds, there is always an orthogonal transformation $\mathbf{D}$ such that after the transformation the tensor has a diagonal form:

$$\tilde{\boldsymbol{T}} = \begin{pmatrix} \tilde{t}_{11} & 0 & 0 \\ 0 & \tilde{t}_{22} & 0 \\ 0 & 0 & \tilde{t}_{33} \end{pmatrix}. \tag{4.78a}$$

The elements $\tilde{t}_{11}$, $\tilde{t}_{22}$ and $\tilde{t}_{33}$ are called the *eigenvalues of the tensor* $\boldsymbol{T}$. They are equal to the roots λ_1, λ_2 and λ_3 of the algebraic equation of third degree in λ:

$$\begin{vmatrix} t_{11} - \lambda & t_{12} & t_{13} \\ t_{21} & t_{22} - \lambda & t_{23} \\ t_{31} & t_{32} & t_{33} - \lambda \end{vmatrix} = 0. \tag{4.78b}$$

The column vectors $\underline{\mathbf{d}}_1$, $\underline{\mathbf{d}}_2$ and $\underline{\mathbf{d}}_3$ of the transformation matrix $\mathbf{D}$ are called the *eigenvectors* corresponding to the eigenvalues, and they satisfy the equations

$$\boldsymbol{T}\underline{\mathbf{d}}_\nu = \lambda_\nu \underline{\mathbf{d}}_\nu \qquad (\nu = 1, 2, 3). \tag{4.78c}$$

Their directions are called the *directions of the principal axes*, and the transformation $\boldsymbol{T}$ to diagonal form is called the *transformation of the principal axes*.

4.3.3.2 Invariant Tensors

1. Definition

A Cartesian tensor is called *invariant* if its components are the same in all Cartesian coordinate systems. Physical quantities such as scalars and vectors, which are special tensors, do not depend on the coordinate system in which they are determined; they must not change their value either under translation of the origin or rotation of a coordinate system K. One talks about *translation invariance* and about *rotation invariance* or in general about *transformation invariance*.

2. Generalized Kronecker Delta or Delta Tensor

If the elements t_{ij} of a tensor of rank 2 are the Kronecker symbols, i.e.,

$$t_{ij} = \delta_{ij} = \begin{cases} 1 & \text{for } i = j, \\ 0 & \text{otherwise,} \end{cases} \tag{4.79a}$$

then from the transformation law (4.71b) in the case of a rotation of the coordinate system considering (4.68c) follows

$$\tilde{t}_{\mu\nu} = d_{\mu i} d_{\nu j} = \delta_{\mu\nu}\,, \tag{4.79b}$$

i.e., the elements are *rotation invariant*. Putting them into a coordinate system so that they are independent of the choice of the origin, i.e., they will be *translation invariant*, then the numbers δ_{ij} form a tensor of rank 2, the so-called *generalized Kronecker delta* or *delta tensor*.

3. Alternating Tensor

If $\vec{e}_i$, $\vec{e}_j$ and $\vec{e}_k$ are unit vectors in the directions of the axes of a right-angle coordinate system, then for the mixed product (see 3.5.1.6, **2.**, p. 185) holds

$$\epsilon_{ijk} = \vec{e}_i\,(\vec{e}_j \times \vec{e}_k) = \begin{cases} 1, & \text{if } i,j,k \text{ cyclic (right-hand rule)},\\ -1, & \text{if } i,j,k \text{ anticyclic},\\ 0, & \text{otherwise}\,. \end{cases} \tag{4.80a}$$

Altogether there are $3^3 = 27$ elements, which are the elements of a tensor of rank 3. In the case of a rotation of the coordinate system from the transformation law (4.69) it follows that

$$\tilde{t}_{\mu\nu\rho} = d_{\mu i} d_{\nu j} d_{\rho k} \epsilon_{ijk} = \begin{vmatrix} d_{\mu 1} & d_{\nu 1} & d_{\rho 1} \\ d_{\mu 2} & d_{\nu 2} & d_{\rho 2} \\ d_{\mu 3} & d_{\nu 3} & d_{\rho 3} \end{vmatrix} = \epsilon_{\mu\nu\rho}, \tag{4.80b}$$

i.e., the elements are *rotation invariant*. Putting them into a coordinate system so that they are independent of the choice of the origin, i.e., they are *translation invariant*, then the numbers ϵ_{ijk} form a tensor of rank 3, the so-called *alternating tensor*.

4. Tensor Invariants

There must not be a confusion between *tensor invariants* and invariant tensors. Tensor invariants are functions of the components of tensors whose forms and values do not change during the rotation of the coordinate system.

■ **A:** If for instance the tensor $\boldsymbol{T} = (t_{ij})$ is transformed in $\tilde{\boldsymbol{T}} = (\tilde{t}_{ij})$ by a rotation, then the *trace* (*spur*) of it does not change:

$$\mathrm{Tr}(\boldsymbol{T}) = t_{11} + t_{22} + t_{33} = \tilde{t}_{11} + \tilde{t}_{22} + \tilde{t}_{33}\,. \tag{4.81}$$

The trace of the tensor $\boldsymbol{T}$ is equal to the sum of the eigenvalues (see 4.1.2, **4.**, p. 270).

■ **B:** For the determinant of the tensor $\boldsymbol{T} = (t_{ij})$

$$\begin{vmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{vmatrix} = \begin{vmatrix} \tilde{t}_{11} & \tilde{t}_{12} & \tilde{t}_{13} \\ \tilde{t}_{21} & \tilde{t}_{22} & \tilde{t}_{23} \\ \tilde{t}_{31} & \tilde{t}_{32} & \tilde{t}_{33} \end{vmatrix} \tag{4.82}$$

is valid. The determinant of the tensor is equal to the product of the eigenvalues.

4.3.4 Tensors in Curvilinear Coordinate Systems

4.3.4.1 Covariant and Contravariant Basis Vectors

1. Covariant Basis

By the help of the variable position vector are introduced the general *curvilinear coordinates* u, v, w:

$$\vec{r} = \vec{r}\,(u,v,w) = x(u,v,w)\vec{e}_x + y(u,v,w)\vec{e}_y + z(u,v,w)\vec{e}_z. \tag{4.83a}$$

The *coordinate surfaces* corresponding to this system can be got by fixing the independent variables u, v, w in $\vec{r}\,(u,v,w)$, one at a time. There are three coordinate surfaces passing through every point of the considered region of space, and any two of them intersect each other in a *coordinate line*, and of course these curves pass through the considered point, too. The three vectors

$$\frac{\partial \vec{r}}{\partial u}\,, \quad \frac{\partial \vec{r}}{\partial v}\,, \quad \frac{\partial \vec{r}}{\partial w} \tag{4.83b}$$

point along the directions of the coordinate lines in the considered point. They form the *covariant basis* of the curvilinear coordinate system.

2. Contravariant Basis

The three vectors

$$\frac{1}{D}\left(\frac{\partial\vec{r}}{\partial v}\times\frac{\partial\vec{r}}{\partial w}\right),\quad \frac{1}{D}\left(\frac{\partial\vec{r}}{\partial w}\times\frac{\partial\vec{r}}{\partial u}\right),\quad \frac{1}{D}\left(\frac{\partial\vec{r}}{\partial u}\times\frac{\partial\vec{r}}{\partial v}\right) \tag{4.84a}$$

with the functional determinant (Jacobian determinant see 2.18.2.6,**3.**, p. 123)

$$D=\frac{D(x,y,z)}{D(u,v,w)}=\begin{vmatrix} x_u & x_v & x_w\\ y_u & y_v & y_w\\ z_u & z_v & z_w\end{vmatrix} \tag{4.84b}$$

are always perpendicular to the coordinate surfaces at the considered surface element and they form the so-called *contravariant basis* of the curvilinear coordinate system.

Remark: In the case of orthogonal curvilinear coordinates, i.e., if

$$\frac{D(x,y,z)}{D(u,v,w)}=\begin{vmatrix} t_{11} & t_{12} & t_{13}\\ t_{21} & t_{22} & t_{23}\\ t_{31} & t_{32} & t_{33}\end{vmatrix}=\begin{vmatrix} \tilde{t}_{11} & \tilde{t}_{12} & \tilde{t}_{13}\\ \tilde{t}_{21} & \tilde{t}_{22} & \tilde{t}_{23}\\ \tilde{t}_{31} & \tilde{t}_{32} & \tilde{t}_{33}\end{vmatrix}\quad \frac{\partial\vec{r}}{\partial u}\cdot\frac{\partial\vec{r}}{\partial v}=0,\quad \frac{\partial\vec{r}}{\partial u}\cdot\frac{\partial\vec{r}}{\partial w}=0,\quad \frac{\partial\vec{r}}{\partial v}\cdot\frac{\partial\vec{r}}{\partial w}=0, \tag{4.85}$$

then the directions of the covariant and contravariant basis are coincident.

4.3.4.2 Covariant and Contravariant Coordinates of Tensors of Rank 1

In order to be able to apply the summation convention of Einstein the following notation is introduced for the covariant and contravariant basis:

$$\begin{aligned}&\frac{\partial\vec{r}}{\partial u}=\vec{g}_1,\quad \frac{\partial\vec{r}}{\partial v}=\vec{g}_2,\quad \frac{\partial\vec{r}}{\partial w}=\vec{g}_3\quad\text{and}\\ &\frac{1}{D}\left(\frac{\partial\vec{r}}{\partial v}\times\frac{\partial\vec{r}}{\partial w}\right)=\vec{g}^{\,1},\quad \frac{1}{D}\left(\frac{\partial\vec{r}}{\partial w}\times\frac{\partial\vec{r}}{\partial u}\right)=\vec{g}^{\,2},\quad \frac{1}{D}\left(\frac{\partial\vec{r}}{\partial u}\times\frac{\partial\vec{r}}{\partial v}\right)=\vec{g}^{\,3}.\end{aligned} \tag{4.86}$$

Then the following representations hold for $\vec{v}$:

$$\vec{v}=V^1\vec{g}_1+V^2\vec{g}_2+V^3\vec{g}_3=V^k\vec{g}_k\quad\text{or}\quad \vec{v}=V_1\vec{g}^{\,1}+V_2\vec{g}^{\,2}+V_3\vec{g}^{\,3}. \tag{4.87}$$

The components V^k are the contravariant coordinates, the components V_k are the covariant coordinates of the vector $\vec{v}$. For these coordinates the equalities

$$V^k=g^{kl}V_l\qquad\text{and}\qquad V_k=g_{kl}V^l \tag{4.88a}$$

are valid, where

$$g_{kl}=g_{lk}=\vec{g}_k\cdot\vec{g}_l\qquad\text{and}\qquad g^{kl}=g^{lk}=\vec{g}^{\,k}\cdot\vec{g}^{\,l} \tag{4.88b}$$

respectively. Furthermore using the Kronecker symbol the equality

$$\vec{g}_k\cdot\vec{g}^{\,l}=\delta_{kl}, \tag{4.89a}$$

holds, and consequently

$$g^{kl}g_{lm}=\delta_{km}. \tag{4.89b}$$

The transition from V^k to V_k or from V_k to V^k according to (4.88b) is described by raising or lowering the indices by *oversliding*.

Remark: In Cartesian coordinate systems covariant and contravariant coordinates are equal to each

other.

4.3.4.3 Covariant, Contravariant and Mixed Coordinates of Tensors of Rank 2

1. Coordinate Transformation

In a Cartesian coordinate system with basis vectors $\vec{e}_1$, $\vec{e}_2$ and $\vec{e}_3$ a tensor $\boldsymbol{T}$ of rank 2 can be represented as a matrix

$$\mathbf{T} = \begin{pmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{pmatrix}. \tag{4.90}$$

To introduce curvilinear coordinates u_1, u_2, u_3 the following vector is used:

$$\vec{r} = x_1(u_1, u_2, u_3)\vec{e}_1 + x_2(u_1, u_2, u_3)\vec{e}_2 + x_3(u_1, u_2, u_3)\vec{e}_3 \,. \tag{4.91}$$

The new basis is denoted by the vectors $\vec{g}_1$, $\vec{g}_2$ and $\vec{g}_3$. Now it holds:

$$\vec{g}_l = \frac{\partial \vec{r}}{\partial u_l} = \frac{\partial x_1}{\partial u_l}\vec{e}_1 + \frac{\partial x_2}{\partial u_l}\vec{e}_2 + \frac{\partial x_3}{\partial u_l}\vec{e}_3 = \frac{\partial x_k}{\partial u_l}\vec{e}_k. \tag{4.92}$$

Substituting $\vec{e}_l = \vec{g}^{\,l}$, then follows $\vec{g}_l$ and $\vec{g}^{\,l}$ as covariant and contravariant basis vectors.

2. Linear Vector Function

In a fixed coordinate system with the tensor $\boldsymbol{T}$ given as in (4.90) by the equality

$$\vec{w} = \boldsymbol{T}\vec{v} \tag{4.93a}$$

the following vector representations

$$\vec{v} = V_k\vec{g}^{\,k} = V^k\vec{g}_k, \quad \vec{w} = W_k\vec{g}^{\,k} = W^k\vec{g}_k \tag{4.93b}$$

define a linear relation between the vectors $\vec{v}$ and $\vec{w}$. So (4.93a) is to be considered as a *linear vector function*.

3. Mixed Coordinates

Changing the coordinate system, the equality (4.93a) will have the form

$$\tilde{\vec{w}} = \tilde{\boldsymbol{T}}\tilde{\vec{v}} \,. \tag{4.94a}$$

The relation between the components of $\boldsymbol{T}$ and $\tilde{\boldsymbol{T}}$ is the following:

$$\tilde{t}_{kl} = \frac{\partial u_k}{\partial x_m}\frac{\partial x_n}{\partial u_l}t_{mn} \,. \tag{4.94b}$$

Introducing the notation

$$\tilde{t}_{kl} = T^{k}_{\cdot l} \tag{4.94c}$$

one talks about *mixed coordinates* of the tensor; k contravariant index, l covariant index. For the components of vectors $\vec{v}$ and $\vec{w}$ holds

$$W^k = T^{k}_{\cdot l}V^l \,. \tag{4.94d}$$

If the covariant basis $\vec{g}_k$ is replaced by the contravariant basis $\vec{g}^{\,k}$, then one gets similarly to (4.94b) and (4.94c)

$$T_k^{\cdot l} = \frac{\partial x_m}{\partial u_k}\frac{\partial u_l}{\partial x_n}t_{mn} \,, \tag{4.95a}$$

and (4.94d) is transformed into

$$W_k = T_k^{\cdot l}V_l. \tag{4.95b}$$

For the mixed coordinates $T_k^{\cdot l}$ and $T^{k}_{\cdot l}$ holds the formula

$$T^{k}_{\cdot l} = g^{km}g_{ln}T_m^{\cdot n} \,. \tag{4.95c}$$

4. Pure Covariant and Pure Contravariant Coordinates

Substituting in (4.95b) for V_l the relation $V_l = g_{lm}V^m$, then one gets

$$W_k = T_k^{\cdot l} g_{lm} V^m = T_{km} V^m \,, \tag{4.96a}$$

also considering that

$$T_k^{\cdot l} g_{lm} = T_{km} \,. \tag{4.96b}$$

The T_{km} are called the covariant coordinates of the tensor $\boldsymbol{T}$, because both indices are covariant. Similarly one gets the contravariant coordinates

$$T_l^{km} = g^{ml} T_{\cdot l}^{k} \,. \tag{4.97}$$

The explicit forms are:

$$T_{kl} = \frac{\partial x_m}{\partial u_k} \frac{\partial x_n}{\partial u_l} t_{mn} \,, \tag{4.98a}$$

$$T^{kl} = \frac{\partial u_k}{\partial x_m} \frac{\partial u_l}{\partial x_n} t_{mn} \,. \tag{4.98b}$$

4.3.4.4 Rules of Calculation

In addition to the rules described on 4.3.2, **5.**, p. 283, the following rules of calculations are valid:

1. Addition, Subtraction Tensors of the same rank whose corresponding indices are both covariant or contravariant can be added or subtracted elementwise, and the result is a tensor of the same rank.

2. Multiplication The multiplication of the coordinates of a tensor of rank n by the coordinates of a tensor of rank m results in a tensor of rank $m + n$.

3. Contraction If making the indices of a covariant and a contravariant coordinates of a tensor of rank n ($n \geq 2$) equal, then can be used the Einstein summation convention for this index, and one gets a tensor of rank $n - 2$. This operation is called *contraction*.

4. Oversliding *Oversliding* of two tensors is the following operation: Multiply both, then making a contraction so that the indices by which the contraction is made belong to different factors.

5. Symmetry A tensor is called symmetric with respect to two covariant or two contravariant standing indices if when exchanging them the tensor does not change.

6. Skew-Symmety A tensor is called skew-symmetric with respect to two covariant or two contravariant standing indices if when exchanging them the tensor is multiplied by -1.

■ The alternating tensor (see 4.3.3.2, **3.**,p. 284) is skew-symmetric with respect to two arbitrary covariant or contravariant indices.

4.3.5 Pseudotensors

The reflection of a tensor plays a special role in physics. Because of their different behavior with respect to reflection *polar* and *axial vectors* are distinguished (see 3.5.1.1, **2.**, p. 181), although mathematically they can be handled in the same way. Axial and polar vectors differ from each other in their determination, because axial vectors can be represented by an orientation in addition to length and direction. Axial vectors are also called *pseudovectors*. Since vectors can be considered as tensors, the general notion of pseudotensors is introduced.

4.3.5.1 Symmetry with Respect to the Origin

1. Behavior of Tensors under Space Inversion

1. Notion of Space Inversion The *reflection of the position coordinates* of points in space with respect to the origin is called *space inversion* or *coordinate inversion*. In a three-dimensional Cartesian coordinate system space inversion means the change of the sign of the coordinates:

$$(x, y, z) \to (-x, -y, -z). \tag{4.99}$$

By this a right-hand coordinate system becomes a left-hand system. Similar rules are valid for other coordinate systems. In the spherical coordinate system holds:

$$(r, \vartheta, \varphi) \to (-r, \pi - \vartheta, \varphi + \pi). \tag{4.100}$$

Under this type of reflection the length of the vectors and the angles between them do not change. The transition can be given by a linear transformation.

2. Transformation Matrix According to (4.66), the transformation matrix $\mathbf{A} = (a_{\mu\nu})$ of a linear transformation of three-dimensional space has the following properties in the case of space inversion:

$$a_{\mu\nu} = -\delta_{\mu\nu}, \quad \det \mathbf{A} = -1. \tag{4.101a}$$

For the components of a tensor of rank n (4.69)

$$\tilde{t}_{\mu\nu\cdots\rho} = (-1)^n t_{\mu\nu\cdots\rho} \tag{4.101b}$$

holds. That is: In the case of point symmetry with respect to the origin a tensor of rank 0 remains a scalar, unchanged; a tensor of rank 1 remains a vector with a change of sign; a tensor of rank 2 remains unchanged, etc.

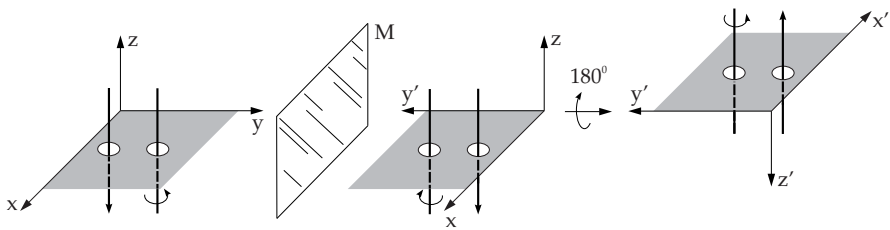

Figure 4.3

2. Geometric Representation

The inversion of space in a three-dimensional Cartesian coordinate system can be realized in two steps **(Fig.4.3)**:

1. By reflection with respect to the coordinate plane, for instance the x, z plane, the coordinate system x, y, z turns into the coordinate system $x, -y, z$. A right-hand system becomes a left-hand system (see 3.5.3.1,**2.**, p. 209).

2. By a rotation of the system x, y, z around the y-axis by 180° we have the complete coordinate system x, y, z reflected with respect to the origin. This coordinate system stays left-handed, as it was after the first step.

Conclusion: Space inversion changes the orientation of a polar vector by 180°, while an axial vector keeps its orientation.

4.3.5.2 Introduction to the Notion of Pseudotensors

1. Vector Product under Space Inversion Under space inversion two polar vectors $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$ are transformed into the vectors $-\underline{\mathbf{a}}$ and $-\underline{\mathbf{b}}$, i.e., their components satisfy the transformation formula (4.101b) for tensors of rank 1. However, if considering the vector product $\underline{\mathbf{c}} = \underline{\mathbf{a}} \times \underline{\mathbf{b}}$ as an example of an axial vector, then one gets $\underline{\mathbf{c}} = \underline{\mathbf{c}}$ under reflection with respect to the origin. This is a violation of the transformation formula (4.101a) for tensors of rank 1. Therefore the axial vector $\underline{\mathbf{c}}$ is called a *pseudovector* or generally a *pseudotensor*.

■ The vector products $\vec{r} \times \vec{v}$, $\vec{r} \times \vec{F}$, $\nabla \times \vec{v} = \operatorname{rot} \vec{v}$ with the position vector $\vec{r}$, the speed vector $\vec{v}$, the power vector $\vec{F}$ and the nabla operator ∇ are examples of axial vectors, which have "false" behavior under reflection.

2. Scalar Product under Space Inversion If using space inversion for a scalar product of a polar and an axial vector, then again there is a case of violation of the transformation formula (4.101b) for tensors of rank 1. Because the result of a scalar product is a scalar, and a scalar should be the

same in every coordinate system, here it is a very special scalar, which is called a *pseudoscalar*. It has the property that it changes its sign under space inversion. Pseudoscalars do not have the *rotation invariance property* of scalars.

■ The scalar product of the polar vectors $\vec{r}$ (position vector) and $\vec{v}$ (speed vector) by the axial vector $\vec{\omega}$ (angular velocity vector) results in the scalars $\vec{r}\cdot\vec{\omega}$ and $\vec{v}\cdot\vec{\omega}$, which have the "false" behavior under reflection, so they are pseudoscalars.

3. Mixed Product under Space Inversion The mixed product $(\underline{\mathbf{a}}\times\underline{\mathbf{b}})\cdot\underline{\mathbf{c}}$ (see 3.5.1.6, **2.**, p. 185) of polar vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, and $\underline{\mathbf{c}}$ is a *pseudoscalar* according to (2.), because the factor $(\underline{\mathbf{a}}\times\underline{\mathbf{b}})$ is an axial vector. The sign of the mixed product changes under space inversion.

4. Pseudovector and Skew-Symmetric Tensor of Rank 2 The tensor product of axial vectors $\underline{\mathbf{a}} = (a_1, a_2, a_3)^{\mathrm{T}}$ and $\underline{\mathbf{b}} = (b_1, b_2, b_3)^{\mathrm{T}}$ results in a tensor of rank 2 with components $t_{ij} = a_i b_j \quad (i, j = 1, 2, 3)$ according to (4.74a). Since every tensor of rank 2 can be decomposed into a sum of a symmetric and a skew-symmetric tensor of rank 2, according to (4.81)

$$t_{ij} = \frac{1}{2}(a_i b_j + a_j b_i) + \frac{1}{2}(a_i b_j - a_j b_i) \quad (i, j = 1, 2, 3) \tag{4.102}$$

holds. The skew-symmetric part of (4.102) contains exactly the components of the vector product $(\underline{\mathbf{a}}\times\underline{\mathbf{b}})$ multiplied by $\frac{1}{2}$, so the axial vector $\underline{\mathbf{c}} = (\underline{\mathbf{a}}\times\underline{\mathbf{b}})$ with components c_1, c_2, c_3 can be considered as a skew-symmetric tensor of rank 2

$$\boldsymbol{C} = \underline{\mathbf{c}} = \begin{pmatrix} 0 & c_{12} & c_{13} \\ -c_{12} & 0 & c_{23} \\ -c_{13} & -c_{23} & 0 \end{pmatrix} \tag{4.103a}$$

$$\text{where} \quad \begin{aligned} c_{23} &= a_2 b_3 - a_3 b_2 = c_1, \\ c_{31} &= a_3 b_1 - a_1 b_3 = c_2, \\ c_{12} &= a_1 b_2 - a_2 b_1 = c_3, \end{aligned} \tag{4.103b}$$

whose components satisfy the transformation formula (4.101b) for tensors of rank 2.
Consequently every axial vector (pseudovector or pseudotensor of rank 1) $\underline{\mathbf{c}} = (c_1, c_2, c_3)^{\mathrm{T}}$ can be considered as a skew-symmetric tensor $\boldsymbol{C}$ of rank 2:

$$\boldsymbol{C} = \underline{\mathbf{c}} = \begin{pmatrix} 0 & c_3 & -c_2 \\ -c_3 & 0 & c_1 \\ c_2 & -c_1 & 0 \end{pmatrix}. \tag{4.104}$$

5. Pseudotensors of Rank n The generalization of the notion of pseudoscalar and pseudovector is a pseudotensor of rank n. It has the same property under rotation as a tensor of rank n (rotation matrix $\mathbf{D}$ with $\det\mathbf{D} = 1$) but it has a (-1) factor under reflection through the origin. Examples of pseudotensors of higher rank can be found in the literature, e.g., [4.1].

4.4 Quaternions and Applications

Quaternions were defined by Sir William Rowan Hamilton in 1843. The basic question which resulted the discovering of quaternions was that how could division of vectors in the three dimensional Euclidian space be defined. It is possible by embedding them into $\mathbb{R}^4$, and introducing the quaternion multiplication, what leads to the *division ring* of quaternions.

Quaternions, like complex numbers, both are special cases of a Clifford-algebra of order n, with 2^n generalized numbers as basis (see [4.5], [22.22]):

$$A = \sum_{l=1}^{2^n} \mathbf{i}_l\, a_l \quad (\mathbf{i}_l \text{ hyper-complex elements},\ a_l \text{ complex numbers}). \tag{4.105a}$$

The following special cases have practical importance:

$n = 1$**:** 2-dimensional complex numbers with

$$\mathbf{i}_1 = 1, \mathbf{i}_2 = \mathrm{i} \ \text{(i imaginary unit)},\ a_1, a_2 \ \text{(real numbers)}. \tag{4.105b}$$

$n = 2$**:** Quaternions as 4-dimensional numbers with hyper-complex elements

$$\mathbf{i}_1 = 1,\ \mathbf{i}_2 = \mathbf{i},\ \mathbf{i}_3 = \mathbf{j},\ \mathbf{i}_4 = \mathbf{k}\ \text{(hyper-complex elements)},\ a_1, a_2, a_3, a_4\ \text{(real numbers)} \tag{4.105c}$$

and the rules of multiplication (4.107). In physics the PAULI's spin matrices and spinors are represented as quaternions.

$n = 3$**:** Biquaternions (s. 4.4.3.6, **1.**, p. 306)

$n = 4$**:** Clifford-numbers are known in physics as DIRAC-matrices.

Quaternions are used most often to describe rotations. The advantages of the quaternions are:
• the rotation is performed directly around the required axis,
• the gimbal-lock problem does not occur. Gimbal is a pivoted support allowing the rotation around a single axis (e.g. gyrocompass), and gimbal lock means that the axes of two of the three gimbals are driven into parallel configuration.

The disadvantage of quaternions is that only rotations can be described with them. To represent translations, scaling or projections matrices are needed. This disadvantage can be overcome by biquaternions, by which every motion of rigid bodies can be represented.

Quaternion are used in computer-graphics, satellite-navigation, in vector analysis, in physics and in mechanics.

4.4.1 Quaternions

4.4.1.1 Definition and Representation

1. Imaginary Units

Quaternions are generalized complex numbers in the form

$$w + \mathbf{i}x + \mathbf{j}y + \mathbf{k}z, \tag{4.106}$$

where w, x, y, z are real numbers, and the generalized imaginary units are $\mathbf{i}$, $\mathbf{j}$,$\mathbf{k}$, which satisfy the following rules of multiplication:

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1, \qquad \mathbf{ij} = \mathbf{k} = -\mathbf{ji}, \quad \mathbf{jk} = \mathbf{i} = -\mathbf{kj}, \quad \mathbf{ki} = \mathbf{j} = -\mathbf{ik}. \tag{4.107}$$

	$\mathbf{i}$	$\mathbf{j}$	$\mathbf{k}$
$\mathbf{i}$	-1	$\mathbf{k}$	$-\mathbf{j}$
$\mathbf{j}$	$-\mathbf{k}$	-1	$\mathbf{i}$
$\mathbf{k}$	$\mathbf{j}$	$-\mathbf{i}$	-1

Multiplication table

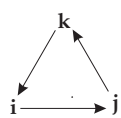

Figure 4.4

The multiplication of the generalized units is shown in the accompanying table of multiplication. Alternatively the multiplication rules can be represented by the cycle shown in **Fig.4.4**. Multiplication in the direction of an arrow results in a positive sign, opposite to the arrow direction results in a negative sign.

Consequently, the multiplication is not commutative but associative. The four-dimensional Euclidian vector space $\mathbb{R}^4$ provided with quaternion-multiplication is denoted by $\mathbb{H}$ in honour of R.W. HAMILTON. Quaternions form an algebra, namely the *division ring of quaternions*. .

2. Representation of Quaternions

There are different representation of quaternions:

• as hyper-complex numbers $q = w + \mathbf{i}x + \mathbf{j}y + \mathbf{k}z = q_0 + \underline{\mathbf{q}}$ with scalar part $q_0 = \mathrm{Sc}q$ and vector part $\underline{\mathbf{q}} = \mathrm{Vec}q$,

• as four dimensional vector $q = (w, x, y, z)^T = (q_0, \underline{\mathbf{q}})^T$ consisting of the number $w \in \mathbb{R}$ and the vector $(x, y, z)^T \in \mathbb{R}^3$,

• in trigonometric form $q = r(\cos\varphi + \underline{\mathbf{n}}_{\underline{\mathbf{q}}} \sin\varphi)$, where $r = |q| = \sqrt{w^2 + x^2 + y^2 + z^2}$ is the length of the four dimensional vector in $\mathbb{R}^4$, and $\cos\varphi = \dfrac{w}{|q|}$, and $\underline{\mathbf{n}}_{\underline{\mathbf{q}}} = \dfrac{1}{|(x, y, z)^T|}(x, y, z)^T$. $\underline{\mathbf{n}}_{\underline{\mathbf{q}}}$ is a unit vector in $\mathbb{R}^3$, depending on q.

Remark: The multiplication rules for quaternions differ from the usually introduced rules in $\mathbb{R}^3$ and

$\mathbb{R}^4$ (see (4.109b), (4.114), (4.115)).

3. Relation between Hyper-complex Number and Trigonometric Form

$$q = q_0 + \underline{\mathbf{q}} = |q| \left(\frac{q_0}{|q|} + \frac{\underline{\mathbf{q}}}{|q|} \right) = |q| \left(\frac{q_0}{|q|} + \frac{\underline{\mathbf{q}}}{|\underline{\mathbf{q}}|} \frac{|\underline{\mathbf{q}}|}{|q|} \right) = r(\cos\varphi + \underline{\mathbf{n}}_{\underline{\mathbf{q}}} \sin\varphi)\,, \tag{4.108a}$$

if $|\underline{\mathbf{q}}| \neq 0$. If $|\underline{\mathbf{q}}| = 0$, then there is

$$q = q_0 = |q_0| \frac{q_0}{|q_0|} = \begin{cases} |q_0| = |q_0| \cos 0 & \text{for } q_0 > 0, \\ |q_0|(-1) = |q_0| \cos\pi & \text{for } q_0 < 0, \end{cases} \tag{4.108b}$$

if $q_0 \neq 0$.

4. Pure Quaternions

A pure quaternion has a zero scalar part: $q_0 = 0$. The set of pure quaternions is denoted by $\mathbb{H}_0$. It is often useful to identify the pure quaternions $\underline{\mathbf{q}}$ with the geometric vector $\vec{\mathbf{q}} \in \mathbb{R}^3$, i.e.

$$q = q_0 + \begin{cases} \underline{\mathbf{q}}, & \text{if } \underline{\mathbf{q}} \text{ represents a pure quaternion,} \\ \vec{\mathbf{q}}, & \text{if } \underline{\mathbf{q}} \text{ is interpreted as a geometric vector.} \end{cases} \tag{4.109a}$$

The *multiplication rule* for $\underline{\mathbf{p}}$, $\underline{\mathbf{q}} \in \mathbb{H}_0$ is

$$\underline{\mathbf{p}}\,\underline{\mathbf{q}} = -\vec{\mathbf{p}} \cdot \vec{\mathbf{q}} + \vec{\mathbf{p}} \times \vec{\mathbf{q}}\,, \tag{4.109b}$$

where $\cdot$ and $\times$ denote the dot-product and cross-product in $\mathbb{R}^3$, respectively. The result of (4.109b) is to interpret as a quaternion.

■ Let $\nabla = \dfrac{\partial}{\partial x}\vec{\mathbf{i}} + \dfrac{\partial}{\partial y}\vec{\mathbf{j}} + \dfrac{\partial}{\partial z}\vec{\mathbf{k}}$ be the nabla-operator (see 13.2.6.1, p. 715), and let $\vec{\mathbf{v}} = v_1(x,y,z)\vec{\mathbf{i}} + v_2(x,y,z)\vec{\mathbf{j}} + v_3(x,y,z)\vec{\mathbf{k}}$ be a vector-field. Here $\vec{\mathbf{i}}, \vec{\mathbf{j}}, \vec{\mathbf{k}}$ are unit-vectors being parallel to the coordinate axes in a Cartesian coordinate system. If ∇ and $\vec{\mathbf{v}}$ are interpreted as pure quaternions then according to (4.107) their product is:

$$\nabla\vec{\mathbf{v}} = -\frac{\partial v_1}{\partial x} - \frac{\partial v_2}{\partial y} - \frac{\partial v_3}{\partial z} + \vec{\mathbf{i}}\left(\frac{\partial v_3}{\partial y} - \frac{\partial v_2}{\partial z}\right) + \vec{\mathbf{j}}\left(\frac{\partial v_1}{\partial z} - \frac{\partial v_3}{\partial x}\right) + \vec{\mathbf{k}}\left(\frac{\partial v_2}{\partial x} - \frac{\partial v_1}{\partial y}\right).$$

This quaternion can be written in vector interpretation:

$$\nabla\vec{\mathbf{v}} = -\operatorname{div}\vec{\mathbf{v}} + \operatorname{rot}\vec{\mathbf{v}}\,,$$

but the result should be considered as a quaternion.

5. Unit Quaternions

A quaternion q is a unit quaternion if $|q| = 1$. The set of unit quaternions is denoted by $\mathbb{H}_1$. $\mathbb{H}_1$ is a so-called multiplicative Lie-group. The set $\mathbb{H}_1$ can be identified with the three dimensional sphere $S^3 = \{\underline{\mathbf{x}} \in \mathbb{R}^4 : |\underline{\mathbf{x}}| = 1\}$.

4.4.1.2 Matrix Representation of Quaternions

1. Real Matrices

If the number 1 is identified with the identity matrix

$$1 \triangleq \begin{pmatrix} 1\,0\,0\,0 \\ 0\,1\,0\,0 \\ 0\,0\,1\,0 \\ 0\,0\,0\,1 \end{pmatrix} \tag{4.110a}$$

furthermore

$$\mathbf{i} \triangleq \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}, \quad \mathbf{j} \triangleq \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \quad \mathbf{k} \triangleq \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}, \tag{4.110b}$$

then a quaternion $q = w + \mathbf{i}x + \mathbf{j}y + \mathbf{k}z$ can be represented as a matrix

$$q \triangleq \begin{pmatrix} w & -x & -y & z \\ x & w & -z & -y \\ y & z & w & x \\ -z & y & -x & w \end{pmatrix}. \tag{4.110c}$$

2. Complex Matrices

Quaternions can be represented by complex matrices:

$$\mathbf{i} \triangleq \begin{pmatrix} 0 & -\mathrm{i} \\ -\mathrm{i} & 0 \end{pmatrix}, \quad \mathbf{j} \triangleq \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{k} \triangleq \begin{pmatrix} -\mathrm{i} & 0 \\ 0 & \mathrm{i} \end{pmatrix}. \tag{4.111a}$$

So

$$q = w + \mathbf{i}x + \mathbf{j}y + \mathbf{k}z \triangleq \begin{pmatrix} w - \mathrm{i}z & -\mathrm{i}x - y \\ -\mathrm{i}x + y & w + \mathrm{i}z \end{pmatrix}. \tag{4.111b}$$

Remarks:
1. On the right hand side of equations (4.111a,b) i represents the imaginary unit of complex numbers.
2. Matrix representation of quaternions is not unique, i.e. it is possible to give representations different from the ones in (4.110b,c) and (4.111a,b).

3. Conjugate and inverse Element

The conjugate of quaternion $q = w + \mathbf{i}x + \mathbf{j}y + \mathbf{k}z$ is the quaternion

$$\bar{q} = w - \mathbf{i}x - \mathbf{j}y - \mathbf{k}z\,. \tag{4.112a}$$

Obviously

$$|q|^2 = q\,\bar{q} = \bar{q}\,q = w^2 + x^2 + y^2 + z^2\,. \tag{4.112b}$$

Consequently every quaternion $q \in \mathbb{H} \setminus \{0\}$ has an inverse element

$$q^{-1} = \frac{\bar{q}}{|q|^2}\,. \tag{4.112c}$$

4.4.1.3 Calculation Rules

1. Addition and Subtraction

Addition and subtraction of two or more quaternions are defined as

$$\begin{aligned} & q_1 + q_2 - q_3 + \ldots \\ &= (w_1 + \mathbf{i}x_1 + \mathbf{j}y_1 + \mathbf{k}z_1) + (w_2 + \mathbf{i}x_2 + \mathbf{j}y_2 + \mathbf{k}z_2) - (w_3 + \mathbf{i}x_3 + \mathbf{j}y_3 + \mathbf{k}z_3) + \ldots \\ &= (w_1 + w_2 - w_3 + \ldots) + \mathbf{i}(x_1 + x_2 - x_3 + \ldots) + \mathbf{j}(y_1 + y_2 - y_3 + \ldots) \\ &\quad + \mathbf{k}(z_1 + z_2 - z_3 + \ldots)\,. \end{aligned} \tag{4.113}$$

Quaternions are added and subtracted as vectors in $\mathbb{R}^4$, or as matrices.

2. Multiplication

The multiplication is associative, so

$$\begin{aligned} q_1 q_2 = \; & (\; w_1 + \mathbf{i}x_1 + \mathbf{j}y_1 + \mathbf{k}z_1)(w_2 + \mathbf{i}x_2 + \mathbf{j}y_2 + \mathbf{k}z_2) \\ & = (w_1 w_2 - x_1 x_2 - y_1 y_2 - z_1 z_2) + \mathbf{i}(w_1 x_2 + w_2 x_1 + y_1 z_2 - z_1 y_2) + \\ & + \mathbf{j}(w_1 y_2 + w_2 y_1 + z_1 x_2 - z_2 x_1) + \mathbf{k}(w_1 z_2 + w_2 z_1 + x_1 y_2 - x_2 y_1). \end{aligned} \tag{4.114}$$

Using the usual vector products in $\mathbb{R}^3$ (see 3.5.1.5, p. 184) it can be written in the form

$$q_1\,q_2 = (q_{01} + \underline{\mathbf{q}}_1)(q_{02} + \underline{\mathbf{q}}_2) = q_{01}\,q_{02} - \vec{\mathbf{q}}_1 \cdot \vec{\mathbf{q}}_2 + \vec{\mathbf{q}}_1 \times \vec{\mathbf{q}}_2\,, \tag{4.115}$$

where $\vec{\mathbf{q}}_1 \cdot \vec{\mathbf{q}}_2$ is the dot-product, and $\vec{\mathbf{q}}_1 \times \vec{\mathbf{q}}_2$ is the cross-product of the vectors $\vec{\mathbf{q}}_1$, $\vec{\mathbf{q}}_2 \in \mathbb{R}^3$. Next is to identify the $\mathbb{R}^3$ with the space $\mathbb{H}_0$ of the pure quaternions.

Remark: Multiplication of quaternions is not commutative!

The *product* $q_1 q_2$ corresponds to the matrix multiplication of matrix $\mathbf{L}_{q_1}$ with vector q_2, and it is equal to the product of matrix $\mathbf{R}_{q_2}$ with q_1:

$$q_1 q_2 = \mathbf{L}_{q_1} q_2 = \begin{pmatrix} w_1 & -x_1 & -y_1 & -z_1 \\ x_1 & w_1 & -z_1 & y_1 \\ y_1 & z_1 & w_1 & -x_1 \\ z_1 & -y_1 & x_1 & w_1 \end{pmatrix} \begin{pmatrix} w_2 \\ x_2 \\ y_2 \\ z_2 \end{pmatrix} = \mathbf{R}_{q_2} q_1 = \begin{pmatrix} w_2 & -x_2 & -y_2 & -z_2 \\ x_2 & w_2 & z_2 & -y_2 \\ y_2 & -z_2 & w_2 & x_2 \\ z_2 & y_2 & -x_2 & w_2 \end{pmatrix} \begin{pmatrix} w_1 \\ x_1 \\ y_1 \\ z_1 \end{pmatrix}. \tag{4.116}$$

3. Division

The definition of division of two quaternions is based on the multiplication: $q_1, q_2 \in \mathbb{H}$, $q_2 \neq 0$,

$$\frac{q_1}{q_2} := q_1\, q_2^{-1} = q_1 \frac{\overline{q}_2}{|q_2|^2}\,. \tag{4.117}$$

The order of the factors is important.

■ Let $q_1 = 1 + \mathbf{j}$, $q_2 = \frac{1}{\sqrt{2}}(1 - \mathbf{k})$, then $|q_2| = 1$, $\overline{q}_2 = \dfrac{1}{\sqrt{2}}(1 + \mathbf{k})$ and so

$$\frac{q_1}{q_2} := q_1 \frac{\overline{q}_2}{|q_2|^2} = \frac{1}{\sqrt{2}}(1 + \mathbf{i} + \mathbf{j} + \mathbf{k}) \neq \frac{\overline{q}_2}{|q_2|^2} q_1 = \frac{1}{\sqrt{2}}(1 - \mathbf{i} + \mathbf{j} + \mathbf{k})\,.$$

4. Generalized Moivre Formula

Let $q \in \mathbb{H}$, whrer $q = q_0 + \underline{\mathbf{q}} = r(\cos\varphi + \underline{\mathbf{n}}_{\underline{\mathbf{q}}} \sin\varphi)$ with $r = |q|$ and $\varphi = \arccos \dfrac{q_0}{|q|}$, $\cos\varphi = \dfrac{q_0}{|q|}$, $\sin\varphi = \dfrac{|\underline{\mathbf{q}}|}{|q|}$, then for arbitrary $k \in \mathbb{N}$:

$$q^k = r^k e^{\underline{\mathbf{n}}_{\underline{\mathbf{q}}} k\varphi} = r^k(\cos(k\varphi) + \underline{\mathbf{n}}_{\underline{\mathbf{q}}} \sin(k\varphi))\,. \tag{4.118}$$

5. Exponential Function

For $q = q_0 + \underline{\mathbf{q}} \in \mathbb{H}$ its exponential expression is defined as

$$e^q = \sum_{k=0}^{\infty} \frac{q^k}{k!} = e^{q_0}(\cos|\underline{\mathbf{q}}| + \underline{\mathbf{n}}_{\underline{\mathbf{q}}} \sin|\underline{\mathbf{q}}|). \tag{4.119}$$

Properties of the exponential function:
For any $q \in \mathbb{H}$ holds:

$$e^{-q} e^q = 1\,, \quad \text{(4.120a)} \qquad e^q \neq 0\,, \quad \text{(4.120b)} \qquad e^q = e^{q_0 + \underline{\mathbf{q}}} = e^{q_0} e^{\underline{\mathbf{q}}}\,, \quad \text{(4.120c)}$$

$$e^{\underline{\mathbf{n}}_{\underline{\mathbf{q}}}\,\pi} = -1, \text{ especially } e^{\mathbf{i}\pi} = e^{\mathbf{j}\pi} = e^{\mathbf{k}\pi} = -1\,. \tag{4.120d}$$

$$\text{Unit quaternion } u \text{ and } \vartheta \in \mathbb{R} : e^{\vartheta\, u} = \cos\vartheta + u \sin\vartheta\,. \tag{4.120e}$$

If $q_1\, q_2 = q_2\, q_1$ then $e^{q_1 + q_2} = e^{q_1}\, e^{q_2}$. But it does not follow from $e^{q_1+q_2} = e^{q_1}\, e^{q_2}$ that $q_1\, q_2 = q_2\, q_1$.

■ Since $(\mathbf{i}\pi)(\mathbf{j}\pi) = \mathbf{k}\pi^2 \neq -\mathbf{k}\pi^2 = (\mathbf{j}\pi)(\mathbf{i}\pi)$ therefore also holds

$$e^{\mathbf{i}\pi} e^{\mathbf{j}\pi} = (\cos\pi)(\cos\pi) = (-1)(-1) = 1\,, \text{ but } e^{\mathbf{i}\pi + \mathbf{j}\pi} = \left(\cos\left(\sqrt{2}\pi\right) + \frac{\mathbf{i} + \mathbf{j}}{\sqrt{2}} \sin\left(\sqrt{2}\pi\right)\right) \neq 1\,.$$

6. Trigonometric Functions

For $q \in \mathbb{H}$ let

$$\cos q := \frac{1}{2}\left(e^{\underline{\mathbf{n}}_{\underline{\mathbf{q}}} q} + e^{-\underline{\mathbf{n}}_{\underline{\mathbf{q}}} q}\right), \qquad \sin q := -\underline{\mathbf{n}}_{\underline{\mathbf{q}}}\left(e^{\underline{\mathbf{n}}_{\underline{\mathbf{q}}} q} - e^{-\underline{\mathbf{n}}_{\underline{\mathbf{q}}} q}\right). \tag{4.121}$$

$\cos q$ is an even function, against which $\sin q$ is an odd function.

Addition formula: It is valid for any $q = q_0 + \underline{\mathbf{q}} \in \mathbb{H}$

$$\cos q = \cos q_0 \cos \underline{\mathbf{q}} - \sin q_0 \sin \underline{\mathbf{q}}\,, \quad \sin q = \sin q_0 \cos \underline{\mathbf{q}} + \cos q_0 \sin \underline{\mathbf{q}}\,. \tag{4.122}$$

7. Hyperbolic Functions

For $q \in \mathbb{H}$ let

$$\cosh q := \frac{1}{2}\left(e^q + e^{-q}\right), \qquad \sinh q := -\underline{\mathbf{n}}_{\underline{\mathbf{q}}}\left(e^q - e^{-q}\right). \tag{4.123}$$

$\cosh q$ is an even function, against which $\sinh q$ is an odd function.

Addition formula: It is valid for any $q = q_0 + \underline{\mathbf{q}} \in \mathbb{H}$

$$\cosh q = \cosh q_0 \cos \underline{\mathbf{q}} - \sinh q_0 \sinh \underline{\mathbf{q}}\,, \quad \sinh q = \sinh q_0 \cos \underline{\mathbf{q}} + \cosh q_0 \sinh \underline{\mathbf{q}}\,. \tag{4.124}$$

8. Logarithmic Function

For $q = q_0 + \underline{\mathbf{q}} = r(\cos\varphi + \underline{\mathbf{n}}_{\underline{\mathbf{q}}} \sin\varphi) \in \mathbb{H}$ and $k \in \mathbb{Z}$ the k-th branch of the logarithmic function is defined as

$$\log_k q := \begin{cases} \ln r + \underline{\mathbf{n}}_{\underline{\mathbf{q}}}(\varphi + 2k\pi), & |\underline{\mathbf{q}}| \neq 0 \quad \text{or} \quad |\underline{\mathbf{q}}| = 0 \text{ and } q_0 > 0\,, \\ \text{not defined for} & |\underline{\mathbf{q}}| = 0 \quad \text{and} \quad q_0 < 0. \end{cases} \tag{4.125}$$

Properties of the logarithmic function:

$$e^{\log_k q} = q \text{ for any } q \in \mathbb{H}\,, \text{ for which } \log_k q \text{ is defined}\,, \tag{4.126a}$$

$$\log_0 e^q = q \text{ for any } q \in \mathbb{H} \text{ with } |\underline{\mathbf{q}}| \neq (2l+1)\pi\,,\ l \in \mathbb{Z}\,, \tag{4.126b}$$

$$\log_k 1 = 0\,, \tag{4.126c}$$

$$\log_0 \mathbf{i} = \frac{\pi}{2}\mathbf{i}\,, \quad \log_0 \mathbf{j} = \frac{\pi}{2}\mathbf{j}, \quad \log_0 \mathbf{k} = \frac{\pi}{2}\mathbf{k}\,. \tag{4.126d}$$

In the case when $\log q_1$ and $\log q_2$ or q_1 and q_2 commute, then, if k is suitable defined, the following known equality (4.127) holds:

$$\log(q_1\, q_2) = \log q_1 + \log q_2\,. \tag{4.127}$$

For the unit quaternions $q \in \mathbb{H}_1$ holds $|q| = 1$ and $q = \cos\varphi + \underline{\mathbf{n}}_{\underline{\mathbf{q}}} \sin\varphi$ and so

$$\log q := \log_0 q = \underline{\mathbf{n}}_{\underline{\mathbf{q}}}\, \varphi \text{ for } q \neq -1\,, \tag{4.128}$$

9. Power function

Let $q \in \mathbb{H}$ and $\alpha \in \mathbb{R}$, then

$$q^\alpha := e^{\alpha \log q}. \tag{4.129}$$

4.4.2 Representation of Rotations in $\mathbb{R}^3$

Spatial rotations are performed around an axis, the so-called rotation axis. It goes through the origin. It is oriented by a direction vector $\vec{\mathbf{a}} \neq \vec{\mathbf{0}}$ (on the axis). The positive direction on this axis is chosen by $\vec{\mathbf{a}}$. The positive rotation (rotation angle $\varphi \geq 0$) is a counterclockwise rotation with respect to the positive direction. The direction vector is usually given normed, i.e. $|\vec{\mathbf{a}}| = 1$.

The equality

$$\vec{\mathbf{w}} = \mathbf{R}\vec{\mathbf{v}}\,, \tag{4.130a}$$

means vector $\vec{\mathbf{w}}$ arises from vector $\vec{\mathbf{v}}$ by the rotation matrix $\mathbf{R}$, i.e., the *rotation matrix* $\mathbf{R}$ transforms vector $\vec{\mathbf{v}}$ into $\vec{\mathbf{w}}$. Since rotation matrices are orthogonal matrices holds

$$\mathbf{R}^{-1} = \mathbf{R}^{\mathrm{T}} \tag{4.130b}$$

and (4.130a) is equivalent to

$$\vec{\mathbf{v}} = \mathbf{R}^{-1}\vec{\mathbf{w}} = \mathbf{R}^{\mathrm{T}}\vec{\mathbf{w}}. \tag{4.130c}$$

Remark: At spatial transformations it is necessary to distinguish between the followings:
a) *geometric transformations*, i.e. when geometric objects are transformed with respect to a fixed coordinate system, and
b) *coordinate transformations*, i.e. the object is fixed while the coordinate system is transformed with respect to the object (see 3.5.4, p. 229).
Now, geometric transformations are handled with quaternions.

4.4.2.1 Rotations of an Object About the Coordinate Axes

In a Cartesian coordinate system the axes are oriented by the basis vectors. The rotation around the x axis is given by matrix $\mathbf{R}_x$, around the y axis by $\mathbf{R}_y$ and around the z axis by $\mathbf{R}_z$, where:

$$\mathbf{R}_x(\alpha) := \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix}, \quad \mathbf{R}_y(\beta) := \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix}, \quad \mathbf{R}_z(\gamma) := \begin{pmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{4.131}$$

The relation between a rotation of an object and the rotation of the coordinate system (see 3.5.3.3,**3.**, p. 213)) is

$$\mathbf{R}_x(\alpha) = \mathbf{D}_x^{\mathrm{T}}(\alpha)\,, \mathbf{R}_y(\beta) = \mathbf{D}_y^{\mathrm{T}}(\beta)\,, \mathbf{R}_z(\gamma) = \mathbf{D}_z^{\mathrm{T}}(\gamma)\,. \tag{4.132}$$

Remark: The representation of the rotation matrices in homogenous coordinates is given in 3.5.4.5,**1.,2.**, p. 234.

4.4.2.2 Cardan-Angles

Every rotation $\mathbf{R}$ around an axis passing through the origin can be given as a sequence of rotations around the coordinate axes in a given coordinate system (see also 3.5.3.5, p. 214), where here
• the first rotation is around the x axis by an angle α_C, then
• the second rotation is around the y axis by an angle β_C, then
• the third rotation is around the z axis by an angle γ_C.
The angles α_C, β_C, γ_C are called Cardan-angles. Then the rotation matrix is

$$\mathbf{R} = \mathbf{R}_C := \mathbf{R}_z(\gamma_C)\mathbf{R}_y(\beta_C)\mathbf{R}_x(\alpha_C) \tag{4.133a}$$

$$= \begin{pmatrix} \cos\beta_C\cos\gamma_C & \sin\alpha_C\sin\beta_C\cos\gamma_C - \cos\alpha_C\sin\gamma_C & \cos\alpha_C\sin\beta_C\cos\gamma_C + \sin\alpha_C\sin\gamma_C \\ \cos\beta_C\sin\gamma_C & \sin\alpha_C\sin\beta_C\sin\gamma_C + \cos\alpha_C\cos\gamma_C & \cos\alpha_C\sin\beta_C\sin\gamma_C - \sin\alpha_C\cos\gamma_C \\ -\sin\beta_C & \sin\alpha_C\cos\beta_C & \cos\alpha_C\cos\beta_C \end{pmatrix}. \tag{4.133b}$$

Adventages:
• very popular representation of rotations,
• clear structure.

Disadventages:
• the order of rotations is important i.e. in general holds

$$\mathbf{R}_x(\alpha_C)\mathbf{R}_y(\beta_C)\mathbf{R}_z(\gamma_C) \neq \mathbf{R}_z(\gamma_C)\mathbf{R}_y(\beta_C)\mathbf{R}_x(\alpha_C)\,, \tag{4.133c}$$

• the representation is not unique since $\mathbf{R}(\alpha_C,\, \beta_C,\, \gamma_C) = \mathbf{R}(-\alpha_C \pm 180°,\, \beta_C \pm 180°,\, \gamma_C \pm 180°)$,
• not suitable for rotations after each other (e.g. at animations),
• gimbal lock can happen (rotation of an axis by 90° goes into an other axis)

■ Gimbal Lock case: rotation around the y axis by 90°

$$\mathbf{R}(\alpha_C, 90°, \gamma_C) = \begin{pmatrix} 0 & \sin(\alpha_C - \gamma_C) & \cos(\alpha_C - \gamma_C) \\ 0 & \cos(\alpha_C - \gamma_C) & -\sin(\alpha_C - \gamma_C) \\ -1 & 0 & 0 \end{pmatrix}. \tag{4.133d}$$

It can be seen that one degree of freedom is lost. In practical applications it can lead to unpredictable motions.

Remark: It can be realized that Cardan-angles are called sometimes as Euler-angles, in the literature however their definitions can be different (see 3.5.3.6, p. 215).

4.4.2.3 Euler Angles

The Euler-angles ψ, ϑ, φ are often introduced as follows (see 3.5.3.6, p. 215):
• the first rotation around the z axis by angle ψ,
• the second rotation around the image of x axis by angle ϑ,
• the third rotation around the image of z axis by angle φ.
The rotation matrix is

$$\mathbf{R} = \mathbf{R}_{\mathrm{E}} := \mathbf{R}_z(\varphi)\mathbf{R}_x(\vartheta)\mathbf{R}_z(\psi) \tag{4.134a}$$

$$= \begin{pmatrix} \cos\psi\cos\varphi - \sin\psi\cos\vartheta\sin\varphi & -\cos\psi\sin\varphi - \sin\psi\cos\vartheta\cos\varphi & \sin\psi\sin\vartheta \\ \sin\psi\cos\varphi + \cos\psi\cos\vartheta\sin\varphi & -\sin\psi\sin\varphi + \cos\psi\cos\vartheta\cos\varphi & -\cos\psi\sin\vartheta \\ \sin\vartheta\sin\varphi & \sin\vartheta\cos\varphi & \cos\vartheta \end{pmatrix}. \tag{4.134b}$$

4.4.2.4 Rotation Around an Arbitrary Zero Point Axis

The counterclockwise rotation around a normed vector $\vec{\mathbf{a}} = (a_x,\, a_y,\, a_z)$ with $|\vec{\mathbf{a}}| = 1$ by an angle φ is made in 5 steps:

1. Rotation of $\vec{\mathbf{a}}$ around the y axis until reaching the x, y plane: $\vec{\mathbf{a}}' = \mathbf{R}_1\vec{\mathbf{a}}$ using $\mathbf{R}_1$ according to (4.135a). The result: vector $\vec{\mathbf{a}}'$ is in the x, y plane.
2. Rotation of $\vec{\mathbf{a}}'$ around the z axis until becoming parallel to the x axis $\vec{\mathbf{a}}'' = \mathbf{R}_2\vec{\mathbf{a}}'$ by $\mathbf{R}_2$ according to (4.135b). The result: vector $\vec{\mathbf{a}}''$ is parallel to the x axis.

$$\mathbf{R}_1 = \begin{pmatrix} \frac{a_x}{\sqrt{a_x^2+a_z^2}} & 0 & \frac{a_z}{\sqrt{a_x^2+a_z^2}} \\ 0 & 1 & 0 \\ \frac{-a_z}{\sqrt{a_x^2+a_z^2}} & 0 & \frac{a_x}{\sqrt{a_x^2+a_z^2}} \end{pmatrix}, \tag{4.135a}$$

$$\mathbf{R}_2 = \begin{pmatrix} \sqrt{a_x^2+a_z^2} & a_y & 0 \\ -a_y & \sqrt{a_x^2+a_z^2}\,0 & \\ 0 & 0 & 1 \end{pmatrix}. \tag{4.135b}$$

3. Rotation around the x axis by an angle φ by $\mathbf{R}_3$:

$$\mathbf{R}_3 = \mathbf{R}_x(\varphi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{pmatrix}. \tag{4.135c}$$

Rotations $\mathbf{R}_1$ and $\mathbf{R}_2$ are inverted in the following two steps.
4. Inverse rotation of $\mathbf{R}_2$, i.e. rotation around the z axis by the angle β where $\sin\beta = a_y$, $\cos\beta = \sqrt{a_x^2+a_z^2}$ according to (4.135d).
5. Inverse rotation of $\mathbf{R}_1$, i.e. rotation around the y axis by $-\alpha$ where $\sin(-\alpha) = \dfrac{-a_z}{\sqrt{a_x^2+a_z^2}}$, $\cos(-\alpha) = \dfrac{a_x}{\sqrt{a_x^2+a_z^2}}$ around the y axis according to (4.135e).

$$\mathbf{R}_2^{-1} = \begin{pmatrix} \sqrt{a_x^2+a_z^2} & -a_y & 0 \\ a_y & \sqrt{a_x^2+a_z^2}\,0 & \\ 0 & 0 & 1 \end{pmatrix}, \tag{4.135d}$$

$$\mathbf{R}_1^{-1} = \begin{pmatrix} \frac{a_x}{\sqrt{a_x^2+a_z^2}} & 0 & \frac{-a_z}{\sqrt{a_x^2+a_z^2}} \\ 0 & 1 & 0 \\ \frac{a_z}{\sqrt{a_x^2+a_z^2}} & 0 & \frac{a_x}{\sqrt{a_x^2+a_z^2}} \end{pmatrix}. \tag{4.135e}$$

Finally the composition matrix is:

$$\mathbf{R}(\vec{\mathbf{a}},\, \varphi) = \mathbf{R}_1^{-1}\mathbf{R}_2^{-1}\mathbf{R}_3\mathbf{R}_2\mathbf{R}_1 = \tag{4.135f}$$

$$\begin{pmatrix} \cos\varphi + a_x^2(1-\cos\varphi) & a_x a_y(1-\cos\varphi) - a_z\sin\varphi & a_x a_z(1-\cos\varphi) + a_y\sin\varphi \\ a_y a_x(1-\cos\varphi) + a_z\sin\varphi & \cos\varphi + a_y^2(1-\cos\varphi) & a_y a_z(1-\cos\varphi) - a_x\sin\varphi \\ a_z a_x(1-\cos\varphi) - a_y\sin\varphi & a_z a_y(1-\cos\varphi) + a_x\sin\varphi & \cos\varphi + a_z^2(1-\cos\varphi) \end{pmatrix}. \quad (4.135g)$$

Matrix $\mathbf{R}(\vec{\mathbf{a}},\,\varphi)$ is an orthogonal matrix, i.e. its inverse is equal to its transpose: $\mathbf{R}^{-1}(\vec{\mathbf{a}},\,\varphi) = \mathbf{R}^{\mathrm{T}}(\vec{\mathbf{a}},\,\varphi)$. The following formulas are also valid:

$$\begin{aligned} \mathbf{R}\vec{\mathbf{x}} &= \mathbf{R}(\vec{\mathbf{a}},\,\varphi)\vec{\mathbf{x}} \\ &= (\cos\varphi)\vec{\mathbf{x}} + (1-\cos\varphi)\frac{\vec{\mathbf{x}}\cdot\vec{a}}{|\vec{\mathbf{a}}|^2}\vec{\mathbf{a}} + \frac{\sin\varphi}{|a|}\vec{\mathbf{a}}\times\vec{\mathbf{x}} && (4.136a) \\ &= (\cos\varphi)\vec{\mathbf{x}} + (1-\cos\varphi)\vec{\mathbf{x}}_{\vec{\mathbf{a}}} + (\sin\varphi)\frac{\vec{\mathbf{a}}}{|\vec{\mathbf{a}}|}\times\vec{\mathbf{x}}. && (4.136b) \end{aligned}$$

In these formulas the vector $\vec{\mathbf{x}}$ is decomposed into two components, one is parallel, the other is perpendicular to $\vec{\mathbf{a}}$. The parallel part is $\vec{\mathbf{x}}_{\vec{\mathbf{a}}} = \dfrac{\vec{\mathbf{x}}\cdot\vec{\mathbf{a}}}{|\vec{\mathbf{a}}|^2}\vec{\mathbf{a}}$, the perpendicular part is $\vec{\mathbf{r}} = \vec{\mathbf{x}} - \vec{\mathbf{x}}_{\vec{\mathbf{a}}}$. The orthogonal part is in a plane whose normal vector is $\vec{\mathbf{a}}$, so its image is $\cos\varphi\vec{\mathbf{r}} + \sin\varphi\vec{\mathbf{r}}^*$, where $\vec{\mathbf{r}}^*$ is obtained from $\vec{\mathbf{r}}$ by a 90° rotation in positive direction: $\vec{\mathbf{r}}^* = \dfrac{1}{|\vec{\mathbf{a}}|}\vec{\mathbf{a}}\times\vec{\mathbf{r}}$. The result of the rotation of vector $\vec{\mathbf{x}}$ is

$$\vec{\mathbf{x}}_{\vec{\mathbf{a}}} + \cos\varphi\vec{\mathbf{r}} + \sin\varphi\vec{\mathbf{r}}^* = \frac{\vec{\mathbf{x}}\cdot\vec{\mathbf{a}}}{|\vec{\mathbf{a}}|^2}\vec{\mathbf{a}} + (\cos\varphi)\left(\vec{\mathbf{x}} - \frac{\vec{\mathbf{x}}\cdot\vec{\mathbf{a}}}{|\vec{\mathbf{a}}|^2}\vec{\mathbf{a}}\right) + (\sin\varphi)\frac{1}{|\vec{\mathbf{a}}|}\vec{\mathbf{a}}\times\vec{\mathbf{r}} \quad (4.136c)$$

$$\text{with}\quad \vec{\mathbf{a}}\times\vec{\mathbf{r}} = \vec{\mathbf{a}}\times(\vec{\mathbf{x}} - \vec{\mathbf{x}}_a) = \vec{\mathbf{a}}\times\vec{\mathbf{x}}\,. \quad (4.136d)$$

Advantages:
- „Standard representation" in computer-graphics,
- CARDAN-angles should not be determined,
- no gimbal lock.

Disadvantage: Not suitable for animation, i.e. for interpolation of rotations.

4.4.2.5 Rotation and Quaternions

If the unit vector $\vec{\mathbf{a}}$ in (4.135f) is identified as the pure quaternion $\underline{\mathbf{a}}$ (while the angle of rotation φ remains the same), then one gets:

$$R(\underline{\mathbf{a}},\varphi) = \begin{pmatrix} q_0^2+q_1^2-q_2^2-q_3^2 & 2q_1q_2 - 2q_0q_3 & 2q_1q_3 + 2q_0q_2 \\ 2q_1q_2 + 2q_0q_3 & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2q_2q_3 - 2q_0q_1 \\ 2q_1q_3 - 2q_0q_2 & 2q_2q_3 + 2q_0q_1 & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{pmatrix} =: \mathbf{R}(q) \quad (4.137a)$$

where $q_0 = \cos\dfrac{\varphi}{2}$ and $\underline{\mathbf{q}} = (q_1,\,q_2,\,q_3)^{\mathrm{T}} = (a_x,\,a_y,\,a_z)^{\mathrm{T}}\sin\dfrac{\varphi}{2}$, i.e. q is the unit quaternion $q = q(\underline{\mathbf{a}},\varphi) = \cos\dfrac{\varphi}{2} + \underline{\mathbf{a}}\sin\dfrac{\varphi}{2} \in \mathbb{H}_1$. If vector $\vec{\mathbf{x}}$ is considered as $\mathbb{R}^3 \ni \vec{\mathbf{x}} = x_1\mathbf{i} + x_2\mathbf{j} + x_3\mathbf{k} \in \mathbb{H}_0$, then

$$\mathbf{R}(\underline{\mathbf{a}},\varphi)\,\underline{\mathbf{x}} = \mathbf{R}(q)\,\underline{\mathbf{x}} = q\,\underline{\mathbf{x}}\,\bar{q}\,. \quad (4.137b)$$

Especially the columns of the rotation matrix are vectors $q\,\underline{\mathbf{e}}_k\,\bar{q}$:

$$\mathbf{R}\,(\underline{\mathbf{a}},\varphi) = \left(\, q\begin{pmatrix}1\\0\\0\end{pmatrix}\bar{q} \quad q\begin{pmatrix}0\\1\\0\end{pmatrix}\bar{q} \quad q\begin{pmatrix}0\\0\\1\end{pmatrix}\bar{q} \,\right) = (\, q\,\mathbf{i}\,\bar{q} \quad q\,\mathbf{j}\,\bar{q} \quad q\,\mathbf{k}\,\bar{q}\,)\,. \quad (4.137c)$$

Consequences:
- The matrix of rotation can be determined with the help of quaternion $q = \cos\dfrac{\varphi}{2} + \underline{\mathbf{a}}\sin\dfrac{\varphi}{2}$.

• For the rotated vector $\mathbf{R}(\underline{\mathbf{a}}, \varphi)\underline{\mathbf{x}}$ holds $\mathbf{R}(\underline{\mathbf{a}}, \varphi)\underline{\mathbf{x}} = q\,\underline{\mathbf{x}}\,\overline{q}$ in the sense of quaternion multiplication and identifying $\mathbf{R}^3$ with the set of pure quaternions $\mathbb{H}_0$.

For every unit quaternion $q \in \mathbb{H}_1$ q and $-q$ determine the same rotation, so $\mathbb{H}_1$ is a double covering of $SO(3)$. Performing rotations one after the other corresponds to multiplication of quaternions, i.e.

$$\mathbf{R}(q_2)\,\mathbf{R}(q_1) = \mathbf{R}(q_2\,q_1)\,; \tag{4.138}$$

and the *conjugate quaternion* corresponds to the inverse rotation:

$$\mathbf{R}^{-1}(q) = \mathbf{R}(\overline{q}). \tag{4.139}$$

■ Rotation by 60° around the axis defined by $(1,1,1)^{\mathrm{T}}$. First the direction vector should be normed: $\underline{\mathbf{a}} = \dfrac{1}{\sqrt{3}}(1,1,1)^{\mathrm{T}}$. Then with $\sin\varphi = \sin 60° = \dfrac{\sqrt{3}}{2}$ and $\cos\varphi = \cos 60° = \dfrac{1}{2}$ the rotation matrix becomes

$$\mathbf{R}\left(\frac{1}{\sqrt{3}}(1,1,1)^{\mathrm{T}},\, 60°\right) = \frac{1}{3}\begin{pmatrix} 2 & -1 & 2 \\ 2 & 2 & -1 \\ -1 & 2 & 2 \end{pmatrix}.$$

The quaternion describing the rotation is:

$$\begin{aligned} q &= q\left(\frac{1}{\sqrt{3}}(1,1,1)^{\mathrm{T}},\, 60°\right) = \cos 30° + \frac{1}{\sqrt{3}}(\mathbf{i}+\mathbf{j}+\mathbf{k})\sin 30° \\ &= \frac{\sqrt{3}}{2} + \frac{1}{\sqrt{3}}(\mathbf{i}+\mathbf{j}+\mathbf{k})\frac{1}{2} = \frac{\sqrt{3}}{2} + \frac{\sqrt{3}}{6}(\mathbf{i}+\mathbf{j}+\mathbf{k}). \end{aligned}$$

Furthermore

$$\begin{aligned} q\begin{pmatrix}1\\0\\0\end{pmatrix}\overline{q} &= \left(\frac{\sqrt{3}}{2} + \frac{\sqrt{3}}{6}(\mathbf{i}+\mathbf{j}+\mathbf{k})\right)\mathbf{i}\left(\frac{\sqrt{3}}{2} - \frac{\sqrt{3}}{6}(\mathbf{i}+\mathbf{j}+\mathbf{k})\right) \\ &= \left(\frac{\sqrt{3}}{2} + \frac{\sqrt{3}}{6}(\mathbf{i}+\mathbf{j}+\mathbf{k})\right)\left(\frac{\sqrt{3}}{2}\mathbf{i} + \frac{\sqrt{3}}{6} - \frac{\sqrt{3}}{6}\mathbf{k} + \frac{\sqrt{3}}{6}\mathbf{j}\right) \\ &= \frac{24}{36}\mathbf{i} + \frac{24}{36}\mathbf{j} - \frac{12}{36}\mathbf{k} = \frac{1}{3}(2\mathbf{i}+2\mathbf{j}-\mathbf{k}) \mathrel{\hat{=}} \frac{1}{3}\begin{pmatrix}2\\2\\-1\end{pmatrix}. \end{aligned}$$

The two other columns are determined analogously:

$$q\begin{pmatrix}0\\1\\0\end{pmatrix}\overline{q} = \left(\frac{\sqrt{3}}{2} + \frac{\sqrt{3}}{6}(\mathbf{i}+\mathbf{j}+\mathbf{k})\right)\mathbf{j}\left(\frac{\sqrt{3}}{2} - \frac{\sqrt{3}}{6}(\mathbf{i}+\mathbf{j}+\mathbf{k})\right) = \frac{1}{3}(-\mathbf{i}+2\mathbf{j}+2\mathbf{k}) \mathrel{\hat{=}} \frac{1}{3}\begin{pmatrix}-1\\2\\2\end{pmatrix},$$

$$q\begin{pmatrix}0\\0\\1\end{pmatrix}\overline{q} = \left(\frac{\sqrt{3}}{2} + \frac{\sqrt{3}}{6}(\mathbf{i}+\mathbf{j}+\mathbf{k})\right)\mathbf{k}\left(\frac{\sqrt{3}}{2} - \frac{\sqrt{3}}{6}(\mathbf{i}+\mathbf{j}+\mathbf{k})\right) = \frac{1}{3}(2\mathbf{i}-\mathbf{j}+2\mathbf{k}) \mathrel{\hat{=}} \frac{1}{3}\begin{pmatrix}2\\-1\\2\end{pmatrix},$$

$$\mathbf{R}\left(\frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix},\, 60°\right) = \left(q\begin{pmatrix}1\\0\\0\end{pmatrix}\overline{q} \quad q\begin{pmatrix}0\\1\\0\end{pmatrix}\overline{q} \quad q\begin{pmatrix}0\\0\\1\end{pmatrix}\overline{q}\right) = \frac{1}{3}\begin{pmatrix}2&-1&2\\2&2&-1\\-1&2&2\end{pmatrix}.$$

4.4.2.6 Quaternions and Cardan Angles

The rotation matrix in Cardan angles (see (4.133a,b), p. 295) is exactly a matrix of rotation with a unit quaternion $q \in \mathbb{H}_1$.

$$\mathbf{R}_C(\alpha_C,\, \beta_C,\, \gamma_C) = \mathbf{R}_z(\gamma_C)\mathbf{R}_y(\beta_C)\mathbf{R}_x(\alpha_C) \tag{4.140a}$$

$$= \begin{pmatrix} \cos\beta_C \cos\gamma_C & \sin\alpha_C \sin\beta_C \cos\gamma_C - \cos\alpha_C \sin\gamma_C & \cos\alpha_C \sin\beta_C \cos\gamma_C + \sin\alpha_C \sin\gamma_C \\ \cos\beta_C \sin\gamma_C & \sin\alpha_C \sin\beta_C \sin\gamma_C + \cos\alpha_C \cos\gamma_C & \cos\alpha_C \sin\beta_C \sin\gamma_C - \sin\alpha_C \cos\gamma_C \\ -\sin\beta_C & \sin\alpha_C \cos\beta_C & \cos\alpha_C \cos\beta_C \end{pmatrix} \quad (4.140b)$$

$$= [r_{ij}]_{i,j=1}^3 = \begin{pmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2q_1q_2 - 2q_0q_3 & 2q_1q_3 + 2q_0q_2 \\ 2q_1q_2 + 2q_0q_3 & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2q_2q_3 - 2q_0q_1 \\ 2q_1q_3 - 2q_0q_2 & 2q_2q_3 + 2q_0q_1 & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{pmatrix} = \mathbf{R}(q) \quad (4.140c)$$

$$= \left(q \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \overline{q} \quad q \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \overline{q} \quad q \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \overline{q} \right). \quad (4.140d)$$

Comparing the matrix elements one gets

$$\tan\gamma_C = \frac{r_{21}}{r_{11}}, \quad \sin\beta_C = -r_{31}, \quad \tan\alpha_C = \frac{r_{32}}{r_{33}}. \quad (4.141a)$$

In general, the solution is not unique, which is typical in trigonometric problems. However the uniqueness of angles can be reached by discussion of the defined domains.

Reversed, it is easy to get the unit quaternion from the rotation matrix.

$$4q_0q_1 = r_{32} - r_{23}, \quad 4q_0q_2 = r_{13} - r_{31}, \quad 4q_0q_3 = r_{21} - r_{12}\,, \quad (4.141b)$$

$$4q_0^2 - 1 = 4q_0^2 - q_0^2 - q_1^2 - q_2^2 - q_3^2 = r_{11} + r_{22} + r_{33}. \quad (4.141c)$$

Since q and $-q$ define the same rotation, q_0 can be determined as

$$q_0 = \frac{1}{2}\sqrt{r_{11} + r_{22} + r_{33} + 1}\,. \quad (4.141d)$$

The other components are

$$q_1 = \frac{r_{32} - r_{23}}{4q_0}, \quad q_2 = \frac{r_{13} - r_{31}}{4q_0}, \quad q_3 = \frac{r_{21} - r_{12}}{4q_0}. \quad (4.141e)$$

■ Let the rotation matrix be the following:

$$\mathbf{R} = \frac{1}{2} \begin{pmatrix} \sqrt{2} & -\frac{1}{2}\sqrt{6} & \frac{1}{2}\sqrt{2} \\ \sqrt{2} & \frac{1}{2}\sqrt{6} & -\frac{1}{2}\sqrt{2} \\ 0 & 1 & \sqrt{3} \end{pmatrix}.$$

1. Determination of the Cardan angles: Based on the above formulas $\sin\beta_C = -r_{31} = 0$, so $\beta_C = k\pi$, $k \in \mathbb{Z}$. Furthermore $\tan\gamma_C = \dfrac{r_{21}}{r_{11}} = 1$, so $\gamma_C = \dfrac{\pi}{4} + k\pi$, $k \in \mathbb{Z}$, and from $\tan\alpha_C = \dfrac{r_{32}}{r_{33}} = \dfrac{1}{\sqrt{3}}$ it follows that $\alpha_C = \dfrac{\pi}{6} + k\pi$, $k \in \mathbb{Z}$. The angles are unique, if they are determined as the ,,possible smallest" ones, i.e. the rotation whose angles have an absolute value $\leq \dfrac{\pi}{2}$. So the angles are

$$\alpha_C = \frac{\pi}{6}, \quad \beta_C = 0, \quad \gamma_C = \frac{\pi}{4}.$$

2. Determination of the unit quaternion which results this rotation:

$$4q_0^2 - 1 = \frac{1}{2}\left(\sqrt{2} + \frac{1}{2}\sqrt{6} + \sqrt{3}\right) \quad \text{also} \quad q_0 = \frac{1}{2}\sqrt{1 + \frac{1}{2}(\sqrt{2} + \frac{1}{2}\sqrt{6} + \sqrt{3})} \approx 0,8924 = \cos\frac{\varphi}{2}\,.$$

The (possible smallest) angle is $\varphi = 53.6474°$, so $\sin\dfrac{\varphi}{2} = 0.4512$.

3. Determination of the further components of q and the direction of the axis of rotation

$\underline{\mathbf{a}} = (a_x,\, a_y,\, a_z)^{\mathrm{T}}$:

$$q_1 = \frac{r_{32} - r_{23}}{4q_0} = \frac{\left(\frac{1}{2} + \frac{1}{4}\sqrt{2}\right)}{4q_0} \approx 0,2391 \qquad \text{so } a_x = \frac{q_1}{\sin\frac{\varphi}{2}} \approx 0,5299,$$

$$q_2 = \frac{r_{13} - r_{31}}{4q_0} = \frac{\frac{1}{2}\cdot\frac{1}{2}\sqrt{2}}{4q_0} \approx 0,0991 \qquad \text{so } a_y = \frac{q_2}{\sin\frac{\varphi}{2}} \approx 0,2195,$$

$$q_3 = \frac{r_{21} - r_{12}}{4q_0} = \frac{\frac{1}{2}\left(\sqrt{2} + \frac{1}{2}\sqrt{6}\right)}{4q_0} \approx 0,3696 \text{ so } a_z = \frac{q_3}{\sin\frac{\varphi}{2}} \approx 0,8192.$$

Remark: At the calculation of the components in (4.141e) it can be a problem when q_0 is zero or close to zero. In this case the unit quaternion can not be determined by the formulas in (4.141e). To understand this situation one discusses the trace of the rotation matrix:

$$\mathrm{Tr}\,\mathbf{R} = r_{11} + r_{22} + r_{33} = 4q_0^2 - 1\,. \tag{4.142a}$$

If $\mathrm{Tr}\,\mathbf{R} > 0$, then $q_0 = \frac{1}{2}\sqrt{\mathrm{Tr}\,\mathbf{R} + 1} > 0$, and the formulas (4.141e) can be used without any problem. If $\mathrm{Tr}\mathbf{R} \le 0$, then q_0 can be close to zero. In this case the greatest element of the main diagonal is considered. Assume, it is r_{11}. Then $|q_1|$ is greater than $|q_2|$ or $|q_3|$. The components q_1, q_2, q_3 also can be determined from the elements of the main diagonal of the rotation matrix. Choosing the positive sign for the square-roots follows:

$$q_1 = \frac{1}{2}\sqrt{1 + r_{11} - r_{22} - r_{33}}\,,\quad q_2 = \frac{1}{2}\sqrt{1 + r_{22} - r_{11} - r_{33}}\,,\quad q_3 = \frac{1}{2}\sqrt{1 + r_{33} - r_{11} - r_{22}}\,. \tag{4.142b}$$

Calculation rules: From this facts the following calculation rules are derived:

- If $\mathrm{Tr}\,\mathbf{R} \le 0$ and $r_{11} \ge r_{22}$ and $r_{11} \ge r_{33}$, then q_1 has the greatest absolute value, so

$$q_0 = \frac{r_{32} - r_{23}}{4q_1}, \quad q_2 = \frac{r_{21} + r_{12}}{4q_1}, \quad q_3 = \frac{r_{13} + r_{31}}{4q_1}\,. \tag{4.142c}$$

- If $\mathrm{Tr}\,\mathbf{R} \le 0$ and $r_{22} \ge r_{11}$ and $r_{22} \ge r_{33}$, then q_2 has the greatest absolute value, so

$$q_0 = \frac{r_{13} - r_{31}}{4q_2}, \quad q_1 = \frac{r_{21} + r_{12}}{4q_2}, \quad q_3 = \frac{r_{23} + r_{32}}{4q_2}\,. \tag{4.142d}$$

- If $\mathrm{Tr}\,\mathbf{R} \le 0$ and $r_{33} \ge r_{11}$ and $r_{33} \ge r_{22}$, then q_3 has the greatest absolute value, so

$$q_0 = \frac{r_{21} - r_{12}}{4q_3}, \quad q_1 = \frac{r_{31} + r_{13}}{4q_3}, \quad q_2 = \frac{r_{23} + r_{32}}{4q_3}\,. \tag{4.142e}$$

Since the CARDAN-angles define the rotations around the corresponding axes, one can find the assignments given in the following table. Then the rotation

$$\mathbf{R}(\alpha,\, \beta,\, \gamma) = \mathbf{R}((0,0,1)^{\mathrm{T}},\, \gamma)\mathbf{R}((0,1,0)^{\mathrm{T}},\, \beta)\mathbf{R}((1,0,0)^{\mathrm{T}},\, \alpha) \tag{4.142f}$$

corresponds to the unit quaternion

$$q = Q_z\, Q_y\, Q_x\,. \tag{4.142g}$$

rotation	Cardan angle	around	quaternion
$\mathbf{R}_{\mathrm{C}}((1,0,0)^{\mathrm{T}},\, \alpha_{\mathrm{C}})$	α_{C}	x axis	$Q_x := \cos\frac{\alpha_{\mathrm{C}}}{2} + \mathbf{i}\sin\frac{\alpha_{\mathrm{C}}}{2}$
$\mathbf{R}_{\mathrm{C}}((0,1,0)^{\mathrm{T}},\, \beta_{\mathrm{C}})$	β_{C}	y axis	$Q_y := \cos\frac{\beta_{\mathrm{C}}}{2} + \mathbf{j}\sin\frac{\beta_{\mathrm{C}}}{2}$
$\mathbf{R}_{\mathrm{C}}((0,0,1)^{\mathrm{T}},\, \gamma_{\mathrm{C}})$	γ_{C}	z axis	$Q_z := \cos\frac{\gamma_{\mathrm{C}}}{2} + \mathbf{k}\sin\frac{\gamma_{\mathrm{C}}}{2}$

■ Knowing the Cardan angles $\alpha_{\mathrm{C}} = \frac{\pi}{6}$, $\beta_{\mathrm{C}} = 0$, $\gamma_{\mathrm{C}} = \frac{\pi}{4}$, the quaternion describing this rotation can be determined in the following way:

$$Q_x = \cos\frac{\alpha_{\mathrm{C}}}{2} + \mathbf{i}\sin\frac{\alpha_{\mathrm{C}}}{2} = \cos\frac{\pi}{12} + \mathbf{i}\sin\frac{\pi}{12}\,,$$

$$Q_y = \cos\frac{\beta_{\mathrm{C}}}{2} + \mathbf{j}\sin\frac{\beta_{\mathrm{C}}}{2} = \cos 0 + \mathbf{j}\sin 0 = 1\,,$$

$$Q_z = \cos\frac{\gamma_{\mathrm{C}}}{2} + \mathbf{k}\sin\frac{\gamma_{\mathrm{C}}}{2} = \cos\frac{\pi}{8} + \mathbf{k}\sin\frac{\pi}{8}\,.$$

The final result coincides with that given on page 299:

$$\begin{aligned} q := Q_z\, Q_y\, Q_x &= \left(\cos\frac{\pi}{8} + \mathbf{k}\sin\frac{\pi}{8}\right) 1 \left(\cos\frac{\pi}{12} + \mathbf{i}\sin\frac{\pi}{12}\right) \\ &= \cos\frac{\pi}{8}\cdot\cos\frac{\pi}{12} + \mathbf{i}\cos\frac{\pi}{8}\cdot\sin\frac{\pi}{12} + \mathbf{j}\sin\frac{\pi}{8}\cdot\sin\frac{\pi}{12} + \mathbf{k}\sin\frac{\pi}{8}\cdot\cos\frac{\pi}{12} \\ &= 0,8924 + 0,2391\mathbf{i} + 0,0991\mathbf{j} + 0,3696\mathbf{k}\,. \end{aligned}$$

4.4.2.7 Efficiency of the Algorithms

To estimate the efficiency of the algorithms standard operations are defined from which the more complex operations are originated. For complicated comparisons with other methods see [4.12].

Let

- M: number of **mu**ltiplications,
- A: number of **a**dditions and subtractions,
- D: number of **d**ivisions,
- S: number of **s**tandard functions calls, e.g. trigonometric functions, which are composed of a considerable number of multiplications, divisions and additions,
- C: number of **c**omparisons of expressions, which increase the computing time by interrupting the algorithm.

Operation	A	M	D	S	C
Quaternion into Matrix	12	12			
Matrix into Quaternion ($\mathrm{Tr}\,\mathbf{R} > \mathbf{0}$)	6	5	1	1	1
Matrix into Quaternion ($\mathrm{Tr}\,\mathbf{R} \leq \mathbf{0}$)	6	5	1	1	3

Rotation of a vector	A	M	Remarks
with rotation matrix	6	9	
with unit quaternion	24	32	normal quaternion multiplication
with unit quaternion	17	24	fast quaternion multiplication
with unit quaternion	18	21	changing into rotation matrix

Rotation of n vectors	A	M	Remarks
with rotation matrix	$6n$	$9n$	
with unit quaternion	$24n$	$32n$	normal quaternion multiplication
with unit quaternion	$17n$	$24n$	fast quaternion multiplication
with unit quaternion	$12{+}6n$	$12{+}9n$	changing into rotation matrix

composition of two rotations	A	M
with rotation matrix	18	27
with unit quaternion	12	16

Summary: An algorithm based on quaternions is faster only when rotations are performed after each other. It occurs mainly in computer graphics at animations, i.e. at approximations of rotations.

4.4.3 Applications of Quaternions

4.4.3.1 3D Rotations in Computer Graphics

To describe motion flows interpolation of rotations are used. Since the 3D-rotations can be represented by unit quaternions, algorithms are developed for interpolation of rotations in computer graphics. The easiest idea is to start analogously to the definition of linear interpolation in Euclidian-spaces. Basic algorithms are Lerp, Slerp and Squad.

1. Lerp (linear interpolation)

Let $p, q \in \mathbb{H}_1$ and $t \in [0, 1]$, then

$$\mathrm{Lerp}(p,\, q,\, t) = p(1-t) + q\,t\,. \tag{4.143}$$

- This is a linear segment in $\mathbb{R}^4$, connecting $p \in \mathbb{H}_1 \sim S^3 \subset \mathbb{R}^4$ with $q \in \mathbb{H}_1 \sim S^3 \subset \mathbb{R}^4$.
- This segment is inside of the unit sphere in $\mathbb{R}^4$, and does not represent any connecting curve on the unit sphere $S^3 \sim \mathbb{H}_1$.
- Therefore the rotation is determined by normalizing the found quaternion.

This simple algorithm is almost perfect. The only problem is that after finding the interpolation points on the secant between the given points and normalizing the found quaternions, the resulted unit quaternions are not equidistant quaternions. This problem is solved by the following algorithm.

2. Slerp (Spherical linear interpolation)

Let $p, q \in \mathbb{H}_1,\ t \in [0,\, 1]$ and φ $(0 < \varphi < \pi)$ the angle between p and q. Then

$$\mathrm{Slerp}(p,\, q,\, t) = p(\overline{p}\,q)^t = p^{1-t}q^t = p\left[\frac{\sin((1-t)\varphi)}{\sin\varphi}\right] + q\left[\frac{\sin(t\,\varphi)}{\sin\varphi}\right]. \tag{4.144}$$

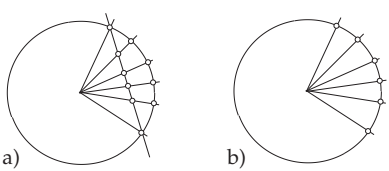

Figure 4.5

- Interpolation along the great circle on the unit sphere $S^3 \subset \mathbb{R}^4$, p and q are connected;
- The shortest connection is chosen, $-\mathrm{Sc}\,(p\,q) = \langle p,\, q\rangle > 0$ must hold (where $\langle\ ,\ \rangle$ denotes the dot product of p and q in $\mathbb{R}^4$).

In **Fig.4.5** the interpolations according to Lerp(a) and Slerp(b) are compared.

Special case $p = 1$: Let $p = 1 = (1,\, 0,\, 0,\, 0)^{\mathrm{T}}$ and $q = \cos\varphi + \underline{\mathbf{n}}_{\underline{\mathbf{q}}}\sin\varphi$, then

$$\mathrm{Slerp}(p,\, q,\, t) = \cos(t\,\varphi) + \underline{\mathbf{n}}_{\underline{\mathbf{q}}}\sin(t\,\varphi)\,. \tag{4.145}$$

Special case equidistant grids: Let $\psi = \dfrac{\varphi}{n}$, then

$$q_k := \mathrm{Slerp}(p,\, q,\, \frac{k}{n}) = \frac{1}{\sin\varphi}\left(\sin((\varphi - k\,\psi)\,p + \sin(k\,\psi)q\right), \quad k = 0,\, 1,\, \ldots,\, n\,. \tag{4.146}$$

Interpretation of the Slerp interpolation: To show the equivalence of the two expressions in (4.144) first $Q = p^{-1}q = \dfrac{\overline{p}}{|p|^2}q = \overline{p}\,q$ is calculated. Since $p, q \in \mathbb{H}_1$ the scalar part is

$$Q_0 = \mathrm{Sc}\,Q = \mathrm{Sc}\,(\overline{p}\,q) = \langle p,\, q\rangle = \cos\varphi\,. \tag{4.147}$$

Since $p = p \cdot 1$, and $q = p\,p^{-1}q = p\,Q$, the interpolation between 1 and Q is multiplied by p to keep the interpolation between p and q .

$$\begin{aligned} Q(t) &= \frac{\sin((1-t)\varphi)}{\sin\varphi} + Q\frac{\sin(t\varphi)}{\sin\varphi} = \frac{\sin((1-t)\varphi)}{\sin\varphi} + \cos\varphi\frac{\sin(t\varphi)}{\sin\varphi} + \vec{n}_{\vec{Q}}\frac{\sin(t\varphi)}{\sin\varphi} \\ &= \frac{\sin\varphi\cos(t\varphi) - \sin(t\varphi)\cos\varphi + \sin(t\varphi)\cos\varphi}{\sin\varphi} + \vec{n}_{\vec{Q}}\frac{\sin(t\varphi)\sin\varphi}{\sin\varphi} \\ &= \cos(t\varphi) + \vec{n}_{\vec{Q}}\sin(t\varphi) = e^{t\,\vec{n}_{\vec{Q}}\varphi} = e^{t\log Q} = Q^t\,. \end{aligned} \tag{4.148}$$

It follows that

$$q(t) = p\,Q(t) = p\frac{\sin((1-t)\varphi)}{\sin\varphi} + q\frac{\sin(t\varphi)}{\sin\varphi} = p\,Q^t = p(p^{-1}q)^t = p^{1-t}q^t\,. \tag{4.149}$$

3. Squad (spherical and quadrangle)

For $q_i,\, q_{i+1} \in \mathbb{H}_1$ and $t \in [0,1]$ the rule is

$$\mathrm{Squad}(q_i\; q_{i+1},\, s_i,\, s_{i+1},\, t) = \mathrm{Slerp}(\mathrm{Slerp}(q_i,\, q_{i+1},\, t), \mathrm{Slerp}(s_i,\, s_{i+1},\, t),\, 2t(1-t)) \tag{4.150}$$

$$\text{with} \quad s_i = q_i \exp\left(-\frac{\log(q_i^{-1}q_{i+1}) + \log(q_i^{-1}q_{i-1})}{4}\right)\,.$$

- The resulted curve is similar to a Bézier curve, but it keeps the spherical instead of the linear interpolation.
- The algorithm produces an interpolation curve for a sequence of quaternions $q_0,\, q_1,\, \ldots,\, q_N$.
- The expression is not defined in the first and last interval, since q_{-1} is necessary to calculate s_0 and q_{N+1} to calculate s_N . A possible way out is to choose $s_0 = q_0$ and $s_N = q_N$, (or to define q_{-1} and q_{N+1}).

There are additional algorithms based on quaternions: nlerp, log-lerp, islerp, quaternion de Casteljau-splines.

4.4.3.2 Interpolation by Rotation matrices

The Slerp-algorithm can be described completely analogously with the help of rotation matrices. The logarithm of a 3×3 rotation matrix $\mathbf{R}$ is needed (i.e. an element of group $SO(3)$) and it is defined by a group-theoretical context as the skew-symmetric matrix $\mathbf{r}$ (i.e. an element of the Lie group so(3)), for which $e^{\mathbf{r}} = \mathbf{R}$. Then the Slerp-algorithm can be used to interpolate between rotation matrices $\mathbf{R}_0$ and $\mathbf{R}_1$,which is described as

$$\mathbf{R}(t) = \mathbf{R}_0(\mathbf{R}_0^{-1}\mathbf{R}_1)^t = \mathbf{R}_0\exp(t\log(\mathbf{R}_0^{-1}\mathbf{R}_1))\,. \tag{4.151}$$

In general it is more simple to use the quaternions based algorithm and to determine $\mathbf{R}(t)$ from $q(t)$ according to the calculations of the rotation matrix representing the unit quaternion.

4.4.3.3 Stereographic Projection

If $1 \in \mathbb{H}_1 \sim S^3$ is taken as the North pole of the three dimensional sphere S^3 , then unit quaternions or elements of the three dimensional sphere can be mapped by the stereographic projection $\mathbb{H}_1 \ni q \mapsto$

$(1+q)(1-q)^{-1} \in \mathbb{H}_0 \sim \mathbb{R}^3$ into pure quaternions or into $\mathbb{R}^3$ respectively. The corresponding inverse mapping is

$$\mathbb{R}^3 \sim \mathbb{H}_0 \ni p \mapsto (p-1)(p+1)^{-1} \in \mathbb{H}_1 \sim S^3 \,. \tag{4.152}$$

4.4.3.4 Satellite navigation

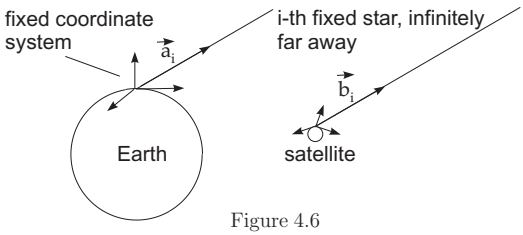

Figure 4.6

The orientation of an artificial satellite circulating around the Earth is to be determined. The fixed stars are considered to be infinitely far away, so their direction with respect to the Earth and the satellite are identical (see **Fig.4.6**). Any difference in measurements can be deduced from the different coordinate systems and so from the relative rotation of the coordinate systems.

Let $\vec{\mathbf{a}}_i$ be the unit vector pointing into the direction of the i-th fixed star from in the Earth's fixed coordinate system, and $\vec{\mathbf{b}}_i$ be the unit vector pointing into the direction of the i-th fixed star in the satellite's fixed coordinate system. The relative rotation of both coordinate systems can be described by a unit quaternion $h \in \mathbb{H}_1$:

$$\vec{\mathbf{b}}_i = h\,\vec{\mathbf{a}}_i\,\overline{h}\,. \tag{4.153}$$

If more fixed stars are considered, and the data are overlapped by measuring errors, then the solution is determined by the least squares method, i.e. as the minimum of (4.154), where h is a unit quaternion and $\underline{\mathbf{a}}_i = \vec{\mathbf{a}}_i$ and $\underline{\mathbf{b}}_i = \vec{\mathbf{b}}_i$ are unit vectors:

$$\begin{aligned} Q^2 &= \sum_{i=1}^{n} |\vec{\mathbf{b}}_i - h\,\vec{\mathbf{a}}_i\,\overline{h}|^2 = \sum_{i=1}^{n} (\vec{\mathbf{b}}_i - h\,\vec{\mathbf{a}}_i\,\overline{h}) \cdot (\vec{\mathbf{b}}_i - h\,\vec{\mathbf{a}}_i\,\vec{\mathbf{h}}) \\ &= \sum_{i=1}^{n} (\underline{\mathbf{b}}_i - h\,\underline{\mathbf{a}}_i\,\overline{h})\overline{(\underline{\mathbf{b}}_i - h\,\underline{\mathbf{a}}_i\,\overline{h})} = \sum_{i=1}^{n} (2 - \underline{\mathbf{b}}_i\,h\,\overline{\underline{\mathbf{a}}}_i\,\overline{h} - h\,\underline{\mathbf{a}}_i\,\overline{h}\,\overline{\underline{\mathbf{b}}}_i) \,. \end{aligned} \tag{4.154}$$

Since the group $\mathbb{H}_1$ of the unit quaternions form a Lie group, the critical points of Q^2 can be determined by the help of derivative

$$\partial_v h = \lim_{\vartheta \to 0} \frac{e^{\vartheta\, v} h - h}{\vartheta} = vh \quad (v, h \text{ quaternions}, \vartheta \text{ real}) \tag{4.155}$$

from

$$\partial_v Q^2 = -\sum_{i=1}^{n} \left(\underline{\mathbf{b}}_i\,\underline{\mathbf{v}}\,h\,\overline{\underline{\mathbf{a}}}_i\,\overline{h} + \underline{\mathbf{b}}_i\,h\,\overline{\underline{\mathbf{a}}}_i\,\overline{(\underline{\mathbf{v}} h)} + \underline{\mathbf{v}}\,h\,\underline{\mathbf{a}}_i\,\overline{h}\,\overline{\underline{\mathbf{b}}}_i + h\,\underline{\mathbf{a}}_i \overline{(\underline{\mathbf{v}}\,h)}\,\overline{\underline{\mathbf{b}}}_i \right) = 0 \,. \tag{4.156}$$

Here $\underline{\mathbf{v}}$, $\underline{\mathbf{b}}_i$ and $h\,\underline{\mathbf{a}}_i\,\overline{h}$ are pure quaternions, so $\overline{\underline{\mathbf{v}}} = -\underline{\mathbf{v}}$, and therefore (4.156) can be simplified:

$$\partial_{\underline{\mathbf{v}}} Q^2 = -4\underline{\mathbf{v}} \cdot \left(\sum_{i=1}^{n} h\,\vec{\mathbf{a}}_i\,\overline{h} \times \underline{\mathbf{b}}_i \right) = 0 \,. \tag{4.157}$$

Since here $\underline{\mathbf{v}}$ is arbitrary, this expression vanishes if

$$\sum_{i=1}^{n} h\,\vec{\mathbf{a}}_i\,\overline{h} \times \underline{\mathbf{b}}_i = \underline{\mathbf{0}} \,. \tag{4.158}$$

Let $\mathbf{R}$ be the rotation matrix represented by the unit quaternion h, i.e. $h\,\vec{\mathbf{a}}_i\overline{h} = \mathbf{R}\,\vec{\mathbf{a}}_i$. With the 3×3

matrix

$$\mathbf{K}(\vec{\mathbf{a}}) = \begin{pmatrix} 0 & -a_z & a_y \\ a_z & 0 & -a_x \\ -a_y & a_x & 0 \end{pmatrix} \tag{4.159}$$

defined by vector $\vec{\mathbf{a}} = (a_x,\, a_y,\, a_z) \in \mathbb{R}^3$ for any vector $\vec{\mathbf{b}} \in \mathbb{R}^3$ one gets:

$$\mathbf{K}(\vec{\mathbf{a}})\,\vec{\mathbf{b}} = \vec{\mathbf{a}} \times \vec{\mathbf{b}}, \qquad \mathbf{K}(\mathbf{K}(\vec{\mathbf{a}})\,\vec{\mathbf{b}}) = \vec{\mathbf{b}}\,\vec{\mathbf{a}}^{\mathrm{T}} - \vec{\mathbf{a}}\,\vec{\mathbf{b}}^{\mathrm{T}}. \tag{4.160}$$

From this relation the critical points of the minimum problems are determined:

$$\sum_{i=1}^{n} \mathbf{K}\,(\mathbf{R}\,\vec{\mathbf{a}}_i \times \vec{\mathbf{b}}_i) = \mathbb{0} \iff \sum_{i=1}^{n} (\vec{\mathbf{b}}_i\,\vec{\mathbf{a}}_i^{\mathrm{T}}\,\mathbf{R}^{\mathrm{T}} - \mathbf{R}\,\vec{\mathbf{a}}_i\,\vec{\mathbf{b}}_i^{\mathrm{T}}) = \mathbb{0} \iff \mathbf{R}\,\mathbf{P} = \mathbf{P}^{\mathrm{T}}\mathbf{R}^{\mathrm{T}} \tag{4.161}$$

where $\mathbf{P} = \sum_{i=1}^{n} \vec{\mathbf{a}}_i\,\vec{\mathbf{b}}_i^{\mathrm{T}}$. If $\mathbf{P}$ is decomposed into the product $\mathbf{P} = \mathbf{R}_p^{\mathrm{T}}\mathbf{S}$, where matrix S is symmetric and $\mathbf{P} = \mathbf{R}_p$ is a rotation matrix, then from (4.161) follows

$$\mathbf{R}\mathbf{R}_p^{\mathrm{T}}\mathbf{S} = \mathbf{S}\mathbf{R}_p\mathbf{R}^{\mathrm{T}}\,, \tag{4.162}$$

and

$$\mathbf{R} = \mathbf{R}_p \tag{4.163}$$

is obviously a solution, since in this case $\mathbf{R}_p\mathbf{R}_p^{\mathrm{T}}\mathbf{S} = \mathbf{S} = \mathbf{S}\mathbf{R}_p\mathbf{R}_p^{\mathrm{T}}$, because $\mathbf{R}_p\mathbf{R}_p^{\mathrm{T}} = \mathbf{E}$. However there are three other solutions, namely

$$\mathbf{R} = \mathbf{R}_j\mathbf{R}_p \quad (j = 1, 2, 3)\,, \tag{4.164}$$

where $\mathbf{R}_j$ denotes the rotation by π around the j-th eigenvector of $\mathbf{S}$, i.e., there is $\mathbf{R_j}\mathbf{S}\mathbf{R_j} = \mathbf{S}$. That $\mathbf{R} = \mathbf{R}_j\mathbf{R}_\nu$ is a solution of (4.162) which can be seen from $\mathbf{R}_j\mathbf{R}_p\mathbf{R}_p^{\mathrm{T}}\mathbf{S} = \mathbf{S}\mathbf{R}_p\mathbf{R}_p^{\mathrm{T}}\mathbf{R}_j^{\mathrm{T}} \iff \mathbf{R}_j\mathbf{S} = \mathbf{S}\mathbf{R}_j^{\mathrm{T}} \iff \mathbf{R}_j\mathbf{S}\mathbf{R}_j = \mathbf{S}$.

The solution for which Q^2 is minimal is

$$\mathbf{R} = \begin{cases} \mathbf{R}_p\,, & \text{falls} \quad \det\mathbf{P} > 0\,, \\ \mathbf{R}_{j_0}\mathbf{R}_p\,, & \text{falls} \quad \det\mathbf{P} < 0\,, \end{cases} \tag{4.165}$$

where $\mathbf{R}_{j_0}$ is the rotation by π around the eigenvector of $\mathbf{S}$ associated with the eigenvalue of the smallest absolute value.

4.4.3.5 Vector Analysis

If the ∇ operator (see (13.67), 13.2.6.1, p. 715) and a vector $\vec{\mathbf{v}}$ (see 13.1.3, p. 704) are identified with ∇_Q and $\underline{\mathbf{v}}$ in quaternion calculus, i.e.

$$\nabla_Q = \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z}\,, \tag{4.166}$$

$$\underline{\mathbf{v}}(x,\, y,\, z) = v_1(x,\, y,\, z)\mathbf{i} + v_2(x,\, y,\, z)\mathbf{j} + v_3(x,\, y,\, z)\mathbf{k} \tag{4.167}$$

with $\mathbf{i}\,, \mathbf{j}\,, \mathbf{k}$ (according to (4.107), p. 290), then the multiplication rule for quaternions (see (4.109b), p. 291) gives

$$\nabla_Q\underline{\mathbf{v}} = -\nabla\cdot\vec{\mathbf{v}} + \nabla\times\vec{\mathbf{v}} = -\mathrm{div}\,\vec{\mathbf{v}} + \mathrm{rot}\;\vec{\mathbf{v}}, \tag{4.168}$$

(see also ■ in 4.4.1.1, **4.** p. 291).

Substituting

$$D = \frac{\partial}{\partial t} + \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z} \quad \text{and} \tag{4.169a}$$

$$\begin{aligned} w(t,\, x,\, y,\, z) &= w_0(t,\, x,\, y,\, z) + w_1(t,\, x,\, y,\, z)\mathbf{i} + w_2(t,\, x,\, y,\, z)\mathbf{j} + w_3(t,\, x,\, y,\, z)\mathbf{k} \\ &= w_0(t,\, x,\, y,\, z) + \underline{\mathbf{w}}(t,\, x,\, y,\, z), \end{aligned} \tag{4.169b}$$

then

$$D\,w = \frac{\partial}{\partial t}\,w_0 - \operatorname{div}\underline{\mathbf{w}} + \operatorname{rot}\,\underline{\mathbf{w}} + \operatorname{grad}\,w_0\,. \tag{4.169c}$$

Especially, for an arbitrary twice continuously differentiable function $f(t,x,y,z)$

$$\nabla_Q\overline{\nabla}_Q\,f = \overline{\nabla}_Q\nabla_Q\,f = \nabla\nabla\,f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} = \Delta_3\,f \quad \text{und} \tag{4.170a}$$

$$\nabla_Q\nabla_Q\,f = -\nabla\nabla\,f = -\frac{\partial^2 f}{\partial x^2} - \frac{\partial^2 f}{\partial y^2} - \frac{\partial^2 f}{\partial z^2} = -\Delta_3\,f\,, \tag{4.170b}$$

where Δ_3 denotes the Laplace operator in $\mathbb{R}^3$ (see (13.75) in 13.2.6.5, p. 716).

$$D\overline{D}f = \overline{D}Df = \frac{\partial^2 f}{\partial t^2} + \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} = \Delta_4 f \tag{4.170c}$$

where Δ_4 denotes the Laplace-operator in $\mathbb{R}^4$. ∇_Q, just as D is often called the Dirac- or Cauchy-Riemann operator.

4.4.3.6 Normalized Quaternions and Rigid Body Motion

1. Biquaternions

A biquaternion $\check{h}$ has the form

$$\check{h} = h_0 + \epsilon\,h_1, \quad \text{with} \quad h_0,\,h_1 \in \mathbf{H} \tag{4.171}$$

Here ϵ is the dual unit, that commutes with every quaternion, furthermore $\epsilon^2 = 0$. The multiplication is the usual quaternion multiplication (see 4.115, p. 292).

2. Rigid Body Motion

By the help of unit biquaternions, i.e. biquaternions with

$$\check{h}\overline{\check{h}} = (h_0 + \epsilon h_1)(\overline{h}_0 + \epsilon\overline{h}_1) = 1 \iff \begin{cases} h_0\,\overline{h}_0 = 1\,, \\ h_0\,\overline{h}_1 + h_1\overline{h}_0 = 0\,, \end{cases} \tag{4.172}$$

rigid-body motion (rotation and translation after each other) can be described in $\mathbb{R}^3$.

Table 4.1 Rigid body motions with biquaternions

Element	Representation by
point $\underline{\mathbf{p}} = (p_x,\,p_y,\,p_z)$ in space	$\check{p} = 1 + \underline{\mathbf{p}}\,\epsilon$ with $\underline{\mathbf{p}} = p_x\,\mathbf{i} + p_y\,\mathbf{j} + p_z\,\mathbf{k}$
rotations	$r \in \mathbb{H}_1$ unit quaternions
translations $\underline{\mathbf{t}} = (t_x,\,t_y,\,t_z)$	$1 + \frac{1}{2}\underline{\mathbf{t}}\,\epsilon$ with $\underline{\mathbf{t}} = t_x\,\mathbf{i} + t_y\,\mathbf{j} + t_z\,\mathbf{k}$

The unit biquaternions

$$\check{h} = h_0 + h_1\,\epsilon = \left(1 + \frac{1}{2}\,\underline{\mathbf{t}}\,\epsilon\right)\,r = r + \frac{1}{2}\underline{\mathbf{t}}\,r\,\epsilon, \quad \underline{\mathbf{t}} \in \mathbb{H}_0\,, \quad r \in \mathbb{H}_1\,, \tag{4.173}$$

give a double covering over the group $\mathbf{SE}(\mathbf{3})$ of the rigid-body motion in $\mathbb{R}^3$ since $\check{h}$ and $-\check{h}$ describe the same rigid-body motion.

4.5 Systems of Linear Equations

4.5.1 Linear Systems, Pivoting

4.5.1.1 Linear Systems

A *linear system* contains m linear forms

$$\begin{array}{l} y_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n + a_1 \\ y_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n + a_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ y_m = a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n + a_m \end{array} \quad \text{or} \quad \underline{\mathbf{y}} = \mathbf{A}\underline{\mathbf{x}} + \underline{\mathbf{a}} \tag{4.174a}$$

with

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad \underline{\mathbf{a}} = \begin{pmatrix} a_1, \\ a_2, \\ \vdots \\ a_m \end{pmatrix}, \quad \underline{\mathbf{x}} = \begin{pmatrix} x_1, \\ x_2, \\ \vdots \\ x_n \end{pmatrix}, \quad \underline{\mathbf{y}} = \begin{pmatrix} y_1, \\ y_2, \\ \vdots \\ y_m \end{pmatrix}. \tag{4.174b}$$

The elements $a_{\mu\nu}$ of the matrix $\mathbf{A}$ of size (m, n) and the components a_μ $(\mu = 1, 2, \ldots, m)$ of the column vector $\underline{\mathbf{a}}$ are constants. The components x_ν $(m = 1, 2, \ldots, n)$ of the column vector $\underline{\mathbf{x}}$ are the *independent variables*, the components y_μ $(\mu = 1, 2, \ldots, m)$ of the column vector $\underline{\mathbf{y}}$ are the *dependent variables*.

4.5.1.2 Pivoting

1. Scheme of Pivoting

If an element a_{ik} is not equal to zero in (4.174a), the variable y_i can be exchanged for an independent one and the variable x_k for a dependent one in a so-called *pivoting step*. The pivoting step is the basic element of pivoting, the method by which for instance systems of linear equations and linear optimization problems can be solved. The pivoting step is achieved by the schemes

$$\begin{array}{c|cccccc} & x_1 & x_2 & \cdots & x_k & \cdots & x_n & 1 \\ \hline y_1 & a_{11} & a_{12} & \cdots & a_{1k} & \cdots & a_{1n} & a_1 \\ y_2 & a_{21} & a_{22} & \cdots & a_{2k} & \cdots & a_{2n} & a_2 \\ \vdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ y_i & a_{i1} & a_{i2} & \cdots & \boxed{a_{ik}} & \cdots & a_{in} & a_i \\ \vdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \underline{y_m} & a_{m1} & a_{m2} & \cdots & a_{mk} & \cdots & a_{mn} & a_m \\ \hline x_k & \alpha_{i1} & \alpha_{i2} & \cdots & \alpha_{ik} & \cdots & \alpha_{in} & \alpha_i \end{array}, \qquad \begin{array}{c|cccccc} & x_1 & x_2 & \cdots & y_i & \cdots & x_n & 1 \\ \hline y_1 & \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1k} & \cdots & \alpha_{1n} & \alpha_1 \\ y_2 & \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2k} & \cdots & \alpha_{2n} & \alpha_2 \\ \vdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_k & \alpha_{i1} & \alpha_{i2} & \cdots & \alpha_{ik} & \cdots & \alpha_{in} & \alpha_i \\ \vdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ y_m & \alpha_{m1} & \alpha_{m2} & \cdots & \alpha_{mk} & \cdots & \alpha_{mn} & \alpha_m \end{array}, \tag{4.175}$$

where the left scheme corresponds to the system (4.174a).

2. Pivoting Rules

The framed element in the scheme a_{ik} $(a_{ik} \neq 0)$ is called the *pivot element*; it is at the intersection of the *pivot column* and *pivot row*. The elements $\alpha_{\mu\nu}$ and α_μ of the new scheme on the right-hand side will be calculated by the following *pivoting rules*:

1. $$\alpha_{ik} = \frac{1}{a_{ik}}, \tag{4.176a}$$

2. $$\alpha_{\mu k} = \frac{a_{\mu k}}{a_{ik}} \quad (\mu = 1, \ldots, m; \mu \neq i), \tag{4.176b}$$

3. $$\alpha_{i\nu} = -\frac{a_{i\nu}}{a_{ik}}, \quad \alpha_i = -\frac{a_i}{a_{ik}} \quad (\nu = 1, 2, \ldots, n; \nu \neq k), \tag{4.176c}$$

4. $$\alpha_{\mu\nu} = a_{\mu\nu} - a_{\mu k}\frac{a_{i\nu}}{a_{ik}} = a_{\mu\nu} + a_{\mu k}\alpha_{i\nu}, \quad \alpha_\mu = a_\mu + a_{\mu k}\alpha_i$$

(for every $\mu \neq i$ and every $\nu \neq k$). (4.176d)

To make the calculations easier (rule 4) one writes the elements $\alpha_{i\nu}$ in the $(m+1)$-th row of the pivoting scheme (cellar row). With this pivoting rule one can change further variables.

4.5.1.3 Linear Dependence

The linear forms (4.174a) are linearly independent (see 9.1.2.3, **2.**, p. 553), if all y_μ can be changed for an independent variable x_ν . The linear independence will be used, for instance to determine the rank of a matrix. Otherwise, the dependence relation can be found directly from the scheme.

■

	x_1	x_2	x_3	x_4	1
y_1	2	1	1	0	−2
y_2	1	−1	0	0	2
y_3	1	5	2	0	0
y_4	0	2	0	1	0

After three pivoting steps (for instance $y_4 \to x_4$, $y_2 \to x_1$, $y_1 \to x_3$) the table becomes:

	y_2	x_2	y_1	y_4	1
x_3	−2	−3	1	0	6
x_1	1	1	0	0	−2
y_3	−3	[0]	2	0	10
x_4	0	−2	0	1	0

No further change is possible because $\alpha_{32} = 0$, and one can see the dependence relation $y_3 = 2y_1 - 3y_2 + 10$. Also for another sequence of pivoting, there remains one pair of not exchangeable variables.

4.5.1.4 Calculation of the Inverse of a Matrix

If $\mathbf{A}$ is a regular matrix of size (n, n), then the inverse matrix $\mathbf{A}^{-1}$ can be obtained after n steps using the pivoting procedure for the system $\underline{\mathbf{y}} = \mathbf{A}\underline{\mathbf{x}}$.

■ $\mathbf{A} = \begin{pmatrix} 3 & 5 & 1 \\ 2 & 4 & 5 \\ 1 & 2 & 2 \end{pmatrix} \implies$

	x_1	x_2	x_3
y_1	3	5	1
y_2	2	4	5
y_3	[1]	2	2

,

	y_3	x_2	x_3
y_1	3	−1	−5
y_2	2	0	[1]
x_1	1	−2	−2

,

	y_3	x_2	y_2
y_1	13	[−1]	−5
x_3	−2	0	1
x_1	5	−2	−2

,

	y_3	y_1	y_2
x_2	13	−1	−5
x_3	−2	0	1
x_1	−21	2	8

.

After rearranging the elements one gets $\mathbf{A}^{-1} = \begin{pmatrix} 2 & 8 & -21 \\ -1 & -5 & 13 \\ 0 & 1 & -2 \end{pmatrix}$. (The columns are to be arranged with respect to the indices of y_i, the rows with respect to the indices of x_k.)

4.5.2 Solution of Systems of Linear Equations

4.5.2.1 Definition and Solvability

1. System of Linear Equations

A system of m linear equations with n unknowns $x_1, x_2, \ldots, x_n$

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= a_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= a_2 \\ \vdots \qquad\qquad & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= a_m \end{aligned} \qquad \text{or briefly} \quad \mathbf{A}\underline{\mathbf{x}} = \underline{\mathbf{a}}, \tag{4.177a}$$

is called a *linear equation system.* Here the following designations are used:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \qquad \underline{\mathbf{a}} = \begin{pmatrix} a_1, \\ a_2, \\ \vdots \\ a_m \end{pmatrix}, \qquad \underline{\mathbf{x}} = \begin{pmatrix} x_1, \\ x_2, \\ \vdots \\ x_n \end{pmatrix}. \tag{4.177b}$$

If the column vector $\underline{\mathbf{a}}$ is the zero vector ($\underline{\mathbf{a}} = \underline{\mathbf{0}}$), then the system of equations is called a *homogeneous system*, otherwise ($\underline{\mathbf{a}} \neq \underline{\mathbf{0}}$) it is called an *inhomogeneous system of equations.* The coefficients $a_{\mu\nu}$ of the system are the elements of the so-called *matrix of coefficients* $\mathbf{A}$, and the components a_μ of the column vector $\underline{\mathbf{a}}$ are the *constant terms (absolute terms).*

2. Solvability of a Linear System of Equations

A linear system of equations is called *solvable* or *consistent* or *compatible* if it has a solution, i.e., there exists at least one vector $\underline{\mathbf{x}} = \underline{\boldsymbol{\alpha}}$ such that (4.177a) is an identity. Otherwise, it is called inconsistent. The existence and uniqueness of the solution depend on the rank of the *augmented matrix* $(\mathbf{A}, \underline{\mathbf{a}})$. One gets the augmented matrix by attaching the vector $\underline{\mathbf{a}}$ to the matrix $\mathbf{A}$ as its $(n+1)$-th column.

1. General Rules for Inhomogeneous Linear Systems of Equations An inhomogeneous linear system of equations $\mathbf{A}\underline{\mathbf{x}} = \underline{\mathbf{a}}$ has at least one solution if

$$\operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{A}, \underline{\mathbf{a}}), \tag{4.178a}$$

is valid. Furthermore, if r denotes the rank of $\mathbf{A}$, i.e., $r = \operatorname{rank}(\mathbf{A})$, then

a) for $r = n$ the system has a unique solution, (4.178b)

b) for $r < n$ the system has infinitely many solutions, (4.178c)

i.e., the values of $n - r$ unknowns as parameters can be chosen freely.

■ **A:**

$$\begin{aligned} x_1 - 2x_2 + 3x_3 - x_4 + 2x_5 &= 2 \\ 3x_1 - x_2 + 5x_3 - 3x_4 - x_5 &= 6 \\ 2x_1 + x_2 + 2x_3 - 2x_4 - 3x_5 &= 8 \end{aligned}$$

The rank of $\mathbf{A}$ is 2, the rank of the *augmented matrix of coefficients* $(\mathbf{A}, \underline{\mathbf{a}})$ is 3, i.e., the system is inconsistent.

■ **B:**

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

Both the matrices $\mathbf{A}$ and $(\mathbf{A}, \underline{\mathbf{a}})$ have rank equal to 3. Because $r = n = 3$ the system has a unique solution: $x_1 = \frac{10}{7}$, $x_2 = -\frac{1}{7}$, $x_3 = -\frac{2}{7}$.

■ **C:**

$$\begin{aligned} x_1 - x_2 + x_3 - x_4 &= 1 \\ x_1 - x_2 - x_3 + x_4 &= 0 \\ x_1 - x_2 - 2x_3 + 2x_4 &= -\frac{1}{2} \end{aligned}$$

Both the matrices $\mathbf{A}$ and $(\mathbf{A}, \underline{\mathbf{a}})$ have rank equal to 2. The system is consistent but because $r < n$ it does not have a unique solution. Therefore $n - r = 2$ unknowns can be considered as free parameters: $x_2 = x_1 - \frac{1}{2}$, $x_3 = x_4 + \frac{1}{2}$, $(x_1, x_4$ arbitrary values).

■ **D:**

$$\begin{aligned} x_1 + 2x_2 - x_3 + x_4 &= 1 \\ 2x_1 - x_2 + 2x_3 + 2x_4 &= 2 \\ 3x_1 + x_2 + x_3 + 3x_4 &= 3 \\ x_1 - 3x_2 + 3x_3 + x_4 &= 0 \end{aligned}$$

There is the same number of equations as unknowns but the system has no solution because $\operatorname{rank}(\mathbf{A}) = 2$, and $\operatorname{rank}(\mathbf{A}, \underline{\mathbf{a}}) = 3$.

2. Trivial Solution and Fundamental System of Homogeneous Systems

a) The homogeneous system of equations $\mathbf{A}\underline{\mathbf{x}} = \underline{\mathbf{0}}$ always has a solution, the so-called trivial solution

$$x_1 = x_2 = \ldots = x_n = 0. \tag{4.179a}$$

(The equality $\operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{A}, \underline{\mathbf{0}})$ always holds.)

b) If the homogeneous system has the non-trivial solutions $\underline{\boldsymbol{\alpha}} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ and $\underline{\boldsymbol{\beta}} = (\beta_1, \beta_2, \ldots, \beta_n)$, i.e., $\underline{\boldsymbol{\alpha}} \neq \underline{\mathbf{0}}$ and $\underline{\boldsymbol{\beta}} \neq \underline{\mathbf{0}}$, then $\underline{\mathbf{x}} = s\,\underline{\boldsymbol{\alpha}} + t\,\underline{\boldsymbol{\beta}}$ is also a solution with arbitrary constants s and t, i.e., any linear combination of the solutions is a solution as well.

Suppose, the system has exactly l non-trivial linearly independent solutions $\underline{\boldsymbol{\alpha}}_1, \underline{\boldsymbol{\alpha}}_2, \ldots, \underline{\boldsymbol{\alpha}}_l$. Then these solutions form a so-called *fundamental system* of solutions (see 9.1.2.3, **2.**, p. 553), and the general solution of the homogeneous system of equations has the form

$$\underline{\mathbf{x}} = k_1\underline{\boldsymbol{\alpha}}_1 + k_2\underline{\boldsymbol{\alpha}}_2 + \cdots + k_l\underline{\boldsymbol{\alpha}}_l \qquad (k_1, k_2, \ldots, k_l \text{ arbitrary constants}). \tag{4.179b}$$

If the rank r of the coefficient matrix $\mathbf{A}$ of the homogeneous system of equations is less than the number of unknowns n, i.e., $r < n$, then the system of equations has $l = n - r$ linearly independent non-trivial solutions. If $r = n$, then the solution is unique, i.e., the homogeneous system has only the trivial

solution.

To determine a fundamental system in the case $r < n$ one chooses $n - r$ unknowns as free parameters, and expresses the remaining unknowns in terms of them. If reordering the equations and the unknowns so that the subdeterminant of order r in the left upper corner is not equal to zero, then one gets for instance:

$$\begin{array}{l} x_1 = x_1(x_{r+1}, x_{r+2}, \ldots, x_n) \\ x_2 = x_2(x_{r+1}, x_{r+2}, \ldots, x_n) \\ \vdots \quad \vdots \qquad\qquad \vdots \\ x_r = x_r(x_{r+1}, x_{r+2}, \ldots, x_n). \end{array} \tag{4.180}$$

Then one can get a fundamental system of solutions choosing the free parameters, for instance in the following way:

$$\begin{array}{lccccc} & x_{r+1} & x_{r+2} & x_{r+3} & \cdots & x_n \\ \hline \text{1. fundamental solution:} & 1 & 0 & 0 & \cdots & 0 \\ \text{2. fundamental solution:} & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ (n-r)\text{-th fundamental solution:} & 0 & 0 & 0 & \cdots & 1 \end{array} . \tag{4.181}$$

■ **E:**

$$\begin{array}{r} x_1 - x_2 + 5x_3 - x_4 = 0 \\ x_1 + x_2 - 2x_3 + 3x_4 = 0 \\ 3x_1 - x_2 + 8x_3 + x_4 = 0 \\ x_1 + 3x_2 - 9x_3 + 7x_4 = 0 \end{array}$$

The rank of the matrix $\mathbf{A}$ is equal to 2. The system can be solved for x_1 and x_2 resulting to: $x_1 = -\frac{3}{2}x_3 - x_4$, $x_2 = \frac{7}{2}x_3 - 2x_4$ (x_3, x_4 arbitrary). Fundamental solutions are $\underline{\boldsymbol{\alpha}_1} = (-\frac{3}{2}, \frac{7}{2}, 1, 0)^{\mathrm{T}}$ and $\underline{\boldsymbol{\alpha}_2} = (-1, -2, 0, 1)^{\mathrm{T}}$.

4.5.2.2 Application of Pivoting

1. System of Linear Functions Corresponding to a Linear System of Equations

In order to solve (4.177a), a system of linear functions $\underline{\mathbf{y}} = \mathbf{A}\underline{\mathbf{x}} - \underline{\mathbf{a}}$ is assigned to the system of equations $\mathbf{A}\underline{\mathbf{x}} = \underline{\mathbf{a}}$ so the use of pivoting (see 4.5.1.2, p. 307) is possible:

$$\mathbf{A}\underline{\mathbf{x}} = \underline{\mathbf{a}} \tag{4.182a}$$

is equivalent to

$$\underline{\mathbf{y}} = \mathbf{A}\underline{\mathbf{x}} - \underline{\mathbf{a}} = \underline{\mathbf{0}}. \tag{4.182b}$$

The matrix $\mathbf{A}$ is of size (m, n), $\underline{\mathbf{a}}$ is a column vector with m components, i.e., the number of equations m must not be equal to the number of unknowns n. After finishing the pivoting one substitutes $\underline{\mathbf{y}} = \underline{\mathbf{0}}$. The existence and uniqueness of the solution of $\mathbf{A}\underline{\mathbf{x}} = \underline{\mathbf{a}}$ can be seen directly from the last pivoting scheme.

2. Solvability of Linear Systems of Equations

The linear system of equations (4.182a) has a solution if one of the following two cases holds for the corresponding linear functions (4.182b):

Case 1: All y_μ ($\mu = 1, 2, \ldots, m$) can be exchanged for some x_ν. This means the corresponding system of linear functions is linearly independent.

Case 2: At least one y_σ cannot be exchanged for any x_ν, i.e.,

$$y_\sigma = \lambda_1 y_1 + \lambda_2 y_2 + \cdots + \lambda_m y_m + \lambda_0 \tag{4.183}$$

holds and also $\lambda_0 = 0$. This means the corresponding system of linear functions is linearly dependent.

3. Inconsistency of Linear Systems of Equations

The linear system of equations has no solution if in case 2 above $\lambda_0 \neq 0$ holds. In this case the system has contradictory equations.

$$\begin{aligned} x_1 - 2x_2 + 4x_3 - x_4 &= 2 \\ -3x_1 + 3x_2 - 3x_3 + 4x_4 &= 3 \\ 2x_1 - 3x_2 + 5x_3 - 3x_4 &= -1 \end{aligned}$$

■

	x_1	x_2	x_3	x_4	1
y_1	1	−2	4	−1	−2
y_2	−3	3	−3	4	−3
y_3	2	−3	5	−3	1

After three pivoting steps (for instance $y_1 \to x_1$, $y_3 \to x_4$, $y_2 \to x_2$) follows:

	y_1	y_2	x_3	y_3	1
x_1	$\frac{3}{2}$	$-\frac{3}{2}$	2	$-\frac{5}{2}$	1
x_2	$-\frac{1}{2}$	$-\frac{1}{2}$	3	$-\frac{1}{2}$	−2
x_4	$\frac{3}{2}$	$-\frac{1}{2}$	0	$-\frac{3}{2}$	3.

This calculation ends with case 1: y_1, y_2, y_3 and x_3 are independent variables. Substituting $y_1 = y_2 = y_3 = 0$, and $x_3 = t$ ($-\infty < t < \infty$ is a parameter) consequently, the solution is: $x_1 = 2t+1$, $x_2 = 3t-2$, $x_3 = t$, $x_4 = 3$.

4.5.2.3 Cramer's Rule

There is the very important special case when the number of equations is equal to the number of unknowns

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= a_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= a_2 \\ \vdots \qquad\qquad\qquad\quad \vdots \quad &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= a_n \end{aligned} \tag{4.184a}$$

and the determinant of the coefficients does not vanish, i.e.,

$$\mathrm{D} = \det \mathbf{A} \neq 0\,. \tag{4.184b}$$

In this case the unique solution of the system of equations (4.184a) can be given in an explicit and unique form:

$$x_1 = \frac{\mathrm{D}_1}{\mathrm{D}}, \quad x_2 = \frac{\mathrm{D}_2}{\mathrm{D}}, \quad \ldots, \quad x_n = \frac{\mathrm{D}_n}{\mathrm{D}}. \tag{4.184c}$$

D_ν denotes the determinant, which is obtained from D by replacing the elements $a_{\mu\nu}$ of the ν-th column of D by the constant terms a_μ, for instance

$$\mathrm{D}_2 = \begin{vmatrix} a_{11} & a_1 & a_{13} & \cdots & a_{1n} \\ a_{21} & a_2 & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_n & a_{n3} & \cdots & a_{nn} \end{vmatrix}. \tag{4.184d}$$

If $\mathrm{D} = 0$ and there is at least one $\mathrm{D}_\nu \neq 0$, then the system (4.184a) has no solution.
In the case $\mathrm{D} = 0$ and $\mathrm{D}_\nu = 0$ for all $\nu = 1, 2, \ldots, n$, then it is possible that the system has a solution but it is not unique. (see Remark p. 311).

$$\begin{aligned} 2x_1 + x_2 + 3x_3 &= 9 \\ x_1 - 2x_2 + x_3 &= -2 \\ 3x_1 + 2x_2 + 2x_3 &= 7. \end{aligned} \qquad \mathrm{D} = \begin{vmatrix} 2 & 1 & 3 \\ 1 & -2 & 1 \\ 3 & 2 & 2 \end{vmatrix} = 13,$$

■

$$\mathrm{D}_1 = \begin{vmatrix} 9 & 1 & 3 \\ -2 & -2 & 1 \\ 7 & 2 & 2 \end{vmatrix} = -13, \quad \mathrm{D}_2 = \begin{vmatrix} 2 & 9 & 3 \\ 1 & -2 & 1 \\ 3 & 7 & 2 \end{vmatrix} = 26, \quad \mathrm{D}_3 = \begin{vmatrix} 2 & 1 & 9 \\ 1 & -2 & -2 \\ 3 & 2 & 7 \end{vmatrix} = 39.$$

The system has the unique solution $x_1 = \frac{\mathrm{D}_1}{\mathrm{D}} = -1$, $x_2 = \frac{\mathrm{D}_2}{\mathrm{D}} = 2$, $x_3 = \frac{\mathrm{D}_3}{\mathrm{D}} = 3$.

Remark: From practical consideration the Cramer rule is not useful for higher-dimensional problems. As the dimension of the problem increases, the number of required operations increases very fast, so, for numerical solutions of linear systems of equations one uses the Gauss algorithm or pivoting or an

iteration procedure (see 19.1.1, p. 949).

4.5.2.4 Gauss's Algorithm

1. Gauss Elimination Method In order to solve the linear system of equations $\mathbf{A}\underline{\mathbf{x}} = \underline{\mathbf{a}}$ (4.177a) of m equations with n unknowns one can use the *Gauss elimination method*. With the help of an equation one unknown is to be eliminated from all the other equations. So one gets a system of $m-1$ equations and $n-1$ unknowns. This method will be repeated until the result is a system of equations in *row echelon form*, and from this form one can determine the existence and uniqueness of the solution easily, and the solution itself can be found if it exists.

2. Gauss Steps The first Gauss step is to be demonstrated on the augmented matrix of coefficients $(\mathbf{A}, \underline{\mathbf{a}})$ (see examples on p. 309):

Supposing $a_{11} \neq 0$, otherwise exchanging the first equation for another one. In the matrix

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & a_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & a_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & a_m \end{array}\right) \tag{4.185a}$$

the appropriate multiple of the first row is to be added to the others in order to make the coefficients of x_1 equal to zero, i.e., multiply the first row by $-\frac{a_{21}}{a_{11}}, -\frac{a_{31}}{a_{11}}, \ldots, -\frac{a_{m1}}{a_{11}}$ then add them to the second, third,..., m-th row. The transformed matrix has the form

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & a_1 \\ 0 & a'_{22} & \cdots & a'_{2n} & a'_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & a'_{m2} & \cdots & a'_{mn} & a'_m \end{array}\right). \tag{4.185b}$$

After applying this Gauss step $(r-1)$ times the result is a matrix in row echelon form

$$\left(\begin{array}{cccccccc|c} a_{11} & a_{12} & a_{13} & \ldots & & a_{1,r+1} & \ldots & a_{1n} & a_1 \\ 0 & a'_{22} & a'_{23} & \ldots & & a'_{2,r+1} & \ldots & a'_{2n} & a'_2 \\ 0 & 0 & a''_{33} & \ldots & & a''_{3,r+1} & \ldots & a''_{3n} & a''_3 \\ \ldots & \ldots & & \ldots & & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & & a_{r,r}^{(r-1)} & a_{r,r+1}^{(r-1)} & \ldots & a_{rn}^{(r-1)} & a_r^{(r-1)} \\ 0 & 0 & \ldots & \ldots & 0 & 0 & \ldots & 0 & a_{r+1}^{(r-1)} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \ldots & 0 & 0 & \ldots & 0 & a_m^{(r-1)} \end{array}\right). \tag{4.186}$$

3. Existence and Uniqueness of the Solution The Gauss steps are elementary row operations so they do not affect the rank of the matrix $(\mathbf{A}, \underline{\mathbf{a}})$, consequently the existence and uniqueness of the solution and the solution itself do not change. Formula (4.186) implies that the following cases may occure concerning the solutions of the inhomogeneous linear system of equations:

Case 1: The system has no solution if any of the numbers $a_{r+1}^{(r-1)}, a_{r+2}^{(r-1)}, \ldots, a_m^{(r-1)}$ differs from zero.

Case 2: The system has a solution if $a_{r+1}^{(r-1)} = a_{r+2}^{(r-1)} = \ldots = a_m^{(r-1)} = 0$ is valid. Then there are two cases:

a) $r = n$: The solution is unique.

b) $r < n$: The solution is not unique; $n - r$ unknowns can be chosen as free parameters.

If the system has a solution, then the unknowns are to be determined in a successive way starting with the last row of the system of equations with the matrix in row echelon form (4.186).

■ **A:** $\begin{aligned} x_1 + 2x_2 + 3x_3 + 4x_4 &= -2 \\ 2x_1 + 3x_2 + 4x_3 + x_4 &= 2 \\ 3x_1 + 4x_2 + x_3 + 2x_4 &= 2 \\ 4x_1 + x_2 + 2x_3 + 3x_4 &= -2 \end{aligned}$ After three Gauss steps the augmented matrix of coefficients has the form $\left(\begin{array}{cccc|c} 1 & 2 & 3 & 4 & -2 \\ 0 & -1 & -2 & -7 & 6 \\ 0 & 0 & -4 & 4 & -4 \\ 0 & 0 & 0 & 40 & -40 \end{array}\right)$.

The solution is unique and from the corresponding system of equations with a triangular matrix follows: $x_4 = -1$, $x_3 = 0$, $x_2 = 1$, $x_1 = 0$.

■ **B:** $\begin{aligned} -x_1 - 3x_2 - 12x_3 &= -5 \\ -x_1 + 2x_2 + 5x_3 &= 2 \\ 5x_2 + 17x_3 &= 7 \\ 3x_1 - x_2 + 2x_3 &= 1 \\ 7x_1 - 4x_2 - x_3 &= 0 \end{aligned}$ After two Gauss steps the augmented matrix of coefficients has the form $\left(\begin{array}{ccc|c} -1 & -3 & -12 & -5 \\ 0 & 5 & 17 & 7 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$.

There is a solution but it is not unique. Choosing one unknown as a free parameter, for instance $x_3 = t$ $(-\infty < t < \infty)$, and one gets $x_3 = t$, $x_2 = \frac{7}{5} - \frac{17}{5}t$, $x_1 = \frac{4}{5} - \frac{9}{5}t$.

4.5.3 Overdetermined Linear Systems of Equations

4.5.3.1 Overdetermined Linear Systems of Equations and Linear Least Squares Problems

1. Overdetermined System of Equations

Consider the linear system of equations

$$\mathbf{A}\underline{\mathbf{x}} = \underline{\mathbf{b}} \tag{4.187}$$

with the rectangular matrix of coefficients $\mathbf{A} = (a_{ij})$ $(i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n;\ m > n)$. The matrix $\mathbf{A}$ and the vector $\underline{\mathbf{b}} = (b_1, b_2, \ldots, b_m)^{\mathrm{T}}$ on the right-hand side are given, and the vector $\underline{\mathbf{x}} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$ is unknown. Because $m > n$ holds this system is called an *over-determined system*. One can tell the existence and uniqueness of the solution and sometimes also the solution, for instance by pivoting.

2. Linear Least Squares Problem

If (4.187) is the mathematical model representing a practical problem (i.e., $\mathbf{A}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{x}}$ are reals), then because of measuring or other errors it is impossible to find an exact solution of (4.187) such that it satisfies all of the equations. Substituting any vector $\underline{\mathbf{x}}$ there will be a *residual vector* $\underline{\mathbf{r}} = (r_1, r_2, \ldots, r_m)^{\mathrm{T}}$ given as

$$\underline{\mathbf{r}} = \mathbf{A}\underline{\mathbf{x}} - \underline{\mathbf{b}}, \quad \underline{\mathbf{r}} \neq \underline{\mathbf{0}}. \tag{4.188}$$

In this case $\underline{\mathbf{x}}$ is to be determined to make the norm of the residual vector $\underline{\mathbf{r}}$ as small as possible. Suppose now $\mathbf{A}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{x}}$ are real. If considering the Euclidean norm, then

$$\sum_{i=1}^{m} r_i^2 = \underline{\mathbf{r}}^{\mathrm{T}}\underline{\mathbf{r}} = (\mathbf{A}\underline{\mathbf{x}} - \underline{\mathbf{b}})^{\mathrm{T}}(\mathbf{A}\underline{\mathbf{x}} - \underline{\mathbf{b}}) = \min \tag{4.189}$$

must be valid, i.e., the *residual sum of squares* must be minimal. Gauss already had this idea. The formula (4.189) is called a *linear least squares problem*. The *norm* $||\underline{\mathbf{r}}|| = \sqrt{\underline{\mathbf{r}}^{\mathrm{T}}\underline{\mathbf{r}}}$ of the residual vector $\underline{\mathbf{r}}$ is called the *residue*.

3. Gauss Transformation

The vector $\underline{\mathbf{x}}$ is the solution of (4.189) if the residual vector $\underline{\mathbf{r}}$ is orthogonal to every column of $\mathbf{A}$. That is:

$$\mathbf{A}^{\mathrm{T}}\underline{\mathbf{r}} = \mathbf{A}^{\mathrm{T}}(\mathbf{A}\underline{\mathbf{x}} - \underline{\mathbf{b}}) = \underline{\mathbf{0}} \qquad \text{or} \qquad \mathbf{A}^{\mathrm{T}}\mathbf{A}\underline{\mathbf{x}} = \mathbf{A}^{\mathrm{T}}\underline{\mathbf{b}}. \tag{4.190}$$

Equation (4.190) is actually a linear system of equations with a square matrix of coefficients. One refers to it as the *system of normal equations*. It has dimension n. The transition from (4.187) to (4.190) is called *Gauss transformation*. The matrix $\mathbf{A}^T\mathbf{A}$ is symmetric.

If the matrix $\mathbf{A}$ has the rank n (because $m > n$ all columns of $\mathbf{A}$ are independent), then the matrix $\mathbf{A}^T\mathbf{A}$ is positive definite and also regular, i.e., the system of normal equations has a unique solution if the rank of $\mathbf{A}$ is equal to the number of unknowns.

4.5.3.2 Suggestions for Numerical Solutions of Least Squares Problems

1. Cholesky Method

Because the matrix $\mathbf{A}^T\mathbf{A}$ is symmetric and positive definite in the case rank $(\mathbf{A}) = n$, in order to solve the normal system of equations one can use the Cholesky method (see 19.2.1.2, p. 958). Unfortunately this algorithm is numerically fairly unstable although it works fairly well in the cases of a "big" residue $||\underline{\mathbf{r}}||$ and a "small" solution $||\underline{\mathbf{x}}||$.

2. Householder Method

Numerically useful procedures in order to solve the least squares problem are the *orthogonalization methods* which are based on the decomposition $\mathbf{A} = \mathbf{QR}$. Especially useful is the *Householder method*, where $\mathbf{Q}$ is an orthogonal matrix of size (m, m) and $\mathbf{R}$ is a triangular matrix of size (m, n) (see 4.1.2, **11.**, p. 271).

3. Regularized Problem

In the case of *rank deficiency*, i.e., if rank $(\mathbf{A}) < n$ holds, then the normal system of equations no longer has a unique solution, and the orthogonalization method gives useless results. Then instead of (4.189) the so-called *regularized problem* is considered:

$$\underline{\mathbf{r}}^T\underline{\mathbf{r}} + \alpha\underline{\mathbf{x}}^T\underline{\mathbf{x}} = \min! \tag{4.191}$$

Here $\alpha > 0$ is a *regularization parameter*. The normal equations for (4.191) are:

$$(\mathbf{A}^T\mathbf{A} + \alpha\mathbf{I})\underline{\mathbf{x}} = \mathbf{A}^T\underline{\mathbf{b}}. \tag{4.192}$$

The matrix of coefficients of this linear system of equations is positive definite and regular for $\alpha > 0$, but the appropriate choice of the regularization parameter α is a difficult problem (see [4.7]).

4.6 Eigenvalue Problems for Matrices

4.6.1 General Eigenvalue Problem

Let $\mathbf{A}$ and $\mathbf{B}$ be two square matrices of size (n, n). Their elements can be real or complex numbers. The *general eigenvalue problem* is to determine the numbers λ and the corresponding vectors $\underline{\mathbf{x}} \neq \underline{\mathbf{0}}$ satisfying the equation

$$\mathbf{A}\underline{\mathbf{x}} = \lambda\mathbf{B}\underline{\mathbf{x}}. \tag{4.193}$$

The number λ is called an *eigenvalue*, the vector $\underline{\mathbf{x}}$ an *eigenvector* corresponding to λ. An eigenvector is determined up to a constant factor, because if $\underline{\mathbf{x}}$ is an eigenvector corresponding to λ, so is $c\underline{\mathbf{x}}$ ($c =$ constant) as well. In the special case when $\mathbf{B} = \mathbf{I}$ holds, where $\mathbf{I}$ is the unit matrix of order n, i.e.,

$$\mathbf{A}\underline{\mathbf{x}} = \lambda\underline{\mathbf{x}} \qquad \text{or} \qquad (\mathbf{A} - \lambda\mathbf{I})\underline{\mathbf{x}} = \underline{\mathbf{0}}, \tag{4.194}$$

the problem is called the *special eigenvalue problem*. It occures very often in practical problems, especially with a symmetric matrix $\mathbf{A}$, and so it is to be discussed later in detail. More information about the general eigenvalue problem can be found in the literature (see [4.16]).

4.6.2 Special Eigenvalue Problem

4.6.2.1 Characteristic Polynomial

The eigenvalue equation (4.194) yields a homogeneous system of equations which has non-trivial solutions $\underline{\mathbf{x}} \neq \underline{\mathbf{0}}$ only if

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0. \tag{4.195a}$$

By the expansion of $\det(\mathbf{A} - \lambda\,\mathbf{I})$ one gets

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \begin{vmatrix} a_{11}-\lambda & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22}-\lambda & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn}-\lambda \end{vmatrix}$$

$$= P_n(\lambda) = (-1)^n\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0 = 0. \tag{4.195b}$$

So the determination of the eigenvalues is equivalent to the solution of a polynomial equation. This equation is called the *characteristic equation*; the polynomial $P_n(\lambda)$ is the *characteristic polynomial*. Its roots are the eigenvalues of the matrix $\mathbf{A}$. For an arbitrary square matrix $\mathbf{A}$ of size (n, n) the following statements hold:

Case 1: The matrix $\mathbf{A}_{(n,n)}$ has exactly n eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, because a polynomial of degree n has n roots if they are considered with their multiplicity. The eigenvalues of a real symmetric matrix are real numbers, in other cases the eigenvalues can also be complex.

Case 2: If all the n eigenvalues are different, then the matrix $\mathbf{A}_{(n,n)}$ has exactly n linearly independent eigenvectors $\underline{\mathbf{x}}_i$ as the solutions of the equation system (4.194) with $\lambda = \lambda_i$.

Case 3: If λ_i has multiplicity n_i among the eigenvalues, and the rank of the matrix $\mathbf{A}_{(n,n)} - \lambda_i\mathbf{I}$ is equal to r_i, then the number of linearly independent eigenvectors corresponding to λ_i is equal to the so-called *nullity* $n - r_i$ of the matrix of coefficients. The inequality $1 \leq n - r_i \leq n_i$ holds, i.e., for a real or complex quadratic matrix $\mathbf{A}_{(n,n)}$ there are at least one and at most n real or complex linearly independent eigenvectors.

■ **A:** $\begin{pmatrix} 2 & -3 & 1 \\ 3 & 1 & 3 \\ -5 & 2 & -4 \end{pmatrix}$, $\det(\mathbf{A} - \lambda\mathbf{I}) = \begin{vmatrix} 2-\lambda & -3 & 1 \\ 3 & 1-\lambda & 3 \\ -5 & 2 & -4-\lambda \end{vmatrix} = -\lambda^3 - \lambda^2 + 2\lambda = 0.$

The eigenvalues are $\lambda_1 = 0, \lambda_2 = 1, \lambda_3 = -2$. The eigenvectors are determined from the corresponding homogeneous linear system of equations.

• $\lambda_1 = 0$:
$$\begin{aligned} 2x_1 - 3x_2 + x_3 &= 0 \\ 3x_1 + x_2 + 3x_3 &= 0 \\ -5x_1 + 2x_2 - 4x_3 &= 0. \end{aligned}$$

One gets for instance by pivoting: x_1 arbitrary, $x_2 = \dfrac{3}{10}x_1, x_3 = -2x_1 + 3x_2 = -\dfrac{11}{10}x_1$. Choosing $x_1 = 10$ the eigenvector is $\underline{\mathbf{x_1}} = C_1 \begin{pmatrix} 10 \\ 3 \\ -11 \end{pmatrix}$, where $C_1 \neq 0$ is an arbitrary constant.

• $\lambda_2 = 1$: The corresponding homogeneous system yields: x_3 is arbitrary, $x_2 = 0, x_1 = 3x_2 - x_3 = -x_3$. Choosing $x_3 = 1$ the eigenvector is $\underline{\mathbf{x_2}} = C_2 \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$, where $C_2 \neq 0$ is an arbitrary constant.

• $\lambda_3 = -2$: The corresponding homogeneous system yields: x_2 is arbitrary, $x_1 = \dfrac{4}{3}x_2$, $x_3 = -4x_1 + 3x_2 = -\dfrac{7}{3}x_2$. Choosing $x_2 = 3$ the eigenvector is $\underline{\mathbf{x_3}} = C_3 \begin{pmatrix} 4 \\ 3 \\ -7 \end{pmatrix}$, where $C_3 \neq 0$ is an arbitrary

constant.

■ **B:** $\begin{pmatrix} 3 & 0 & -1 \\ 1 & 4 & 1 \\ -1 & 0 & 3 \end{pmatrix}$, $\det(\mathbf{A}-\lambda\mathbf{I}) = \begin{vmatrix} 3-\lambda & 0 & -1 \\ 1 & 4-\lambda & 1 \\ -1 & 0 & 3-\lambda \end{vmatrix} = -\lambda^3+10\lambda^2-32\lambda+32=0.$

The eigenvalues are $\lambda_1=2$, $\lambda_2=\lambda_3=4$.

• $\lambda_1=2$: One obtains x_3 is arbitrary, $x_2=-x_3, x_1=x_3$ and chooses for instance $x_3=1$. So the corresponding eigenvector is $\underline{\mathbf{x_1}} = C_1\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$, where $C_1\neq 0$ is an arbitrary constant.

• $\lambda_2=\lambda_3=4$: One obtains x_2, x_3 are arbitrary, $x_1=-x_3$. There are two linearly independent eigenvectors, e.g., for $x_2=1, x_3=0$ and $x_2=0$, $x_3=1$: $\underline{\mathbf{x_2}} = C_2\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, $\underline{\mathbf{x_3}} = C_3\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$, where $C_2\neq 0$, $C_3\neq 0$ are arbitrary constants.

4.6.2.2 Real Symmetric Matrices, Similarity Transformations

In the case of the special eigenvalue problem (4.194) for a real symmetric matrix $\mathbf{A}$ the following statements hold:

1. Properties Concerning the Eigenvalue Problem

1. Number of Eigenvalues The matrix $\mathbf{A}$ has exactly n real eigenvalues λ_i $(i=1,2,\ldots,n)$, counting them by their multiplicity.

2. Orthogonality of the Eigenvectors The eigenvectors $\underline{\mathbf{x}}_i$ and $\underline{\mathbf{x}}_j$ corresponding to different eigenvalues $\lambda_i\neq\lambda_j$ are orthogonal to each other, i.e., for the scalar product of $\underline{\mathbf{x}}_i$ and $\underline{\mathbf{x}}_j$

$$\underline{\mathbf{x}}_i^{\mathrm{T}}\,\underline{\mathbf{x}}_j = (\underline{\mathbf{x}}_i, \underline{\mathbf{x}}_j) = 0 \tag{4.196}$$

is valid.

3. Matrix with an Eigenvalue of Multiplicity *p* For an eigenvalue which has multiplicity p $(\lambda=\lambda_1=\lambda_2=\ldots=\lambda_p)$, there exist p linearly independent eigenvectors $\underline{\mathbf{x}}_1, \underline{\mathbf{x}}_2, \ldots, \underline{\mathbf{x}}_p$. Because of (4.194) all the non-trivial linear combinations of them are also eigenvectors corresponding to λ. Using the Gram–Schmidt orthogonalization process one can choose p of them such that they are orthogonal to each other.

Summarizing: The matrix $\mathbf{A}$ has exactly n real orthogonal eigenvectors.

■ $\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$, $\det(\mathbf{A}-\lambda\,\mathbf{I}) = -\lambda^3+3\lambda+2=0$. The eigenvalues are $\lambda_1=\lambda_2=-1$ and $\lambda_3=2$.

• $\lambda_1=\lambda_2=-1$: From the corresponding homogenous system of equations one gets: x_1 is arbitrary, x_2 is arbitrary, $x_3=-x_1-x_2$. Choosing first $x_1=1$, $x_2=0$ then $x_1=0$, $x_2=1$ one gets the linearly independent eigenvectors $\underline{\mathbf{x}}_1 = C_1\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$ and $\underline{\mathbf{x}}_2 = C_2\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$, where $C_1\neq 0$ and $C_2\neq 0$ are arbitrary constants.

• $\lambda_3=2$: One gets: x_1 is arbitrary, $x_2=x_1$, $x_3=x_1$, and choosing for instance $x_1=1$ one gets the eigenvector $\underline{\mathbf{x}}_3 = C_3\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, where $C_3\neq 0$ is an arbitrary constant. The matrix $\mathbf{A}$ is symmetric, so the eigenvectors corresponding to different eigenvalues are orthogonal.

4. Gram-Schmidt Orthogonalization Process Let V_n be an arbitrary n-dimensional Euclidean vector space. Let the vectors $\underline{\mathbf{x}}_1, \underline{\mathbf{x}}_2, \ldots, \underline{\mathbf{x}}_n \in V_n$ be linearly independent. Then there exists an or-

thogonal system of vectors $\underline{\mathbf{y}}_1, \underline{\mathbf{y}}_2, \ldots, \underline{\mathbf{y}}_n \in V_n$ which can be obtained by the vectors $\underline{\mathbf{x}}_i$ as follows:

$$\underline{\mathbf{y}}_1 = \underline{\mathbf{x}}_1,\ \underline{\mathbf{y}}_k = \underline{\mathbf{x}}_k - \sum_{i=1}^{k-1} \frac{(\underline{\mathbf{x}}_k, \underline{\mathbf{y}}_i)}{(\underline{\mathbf{y}}_i, \underline{\mathbf{y}}_i)}\, \underline{\mathbf{y}}_i \qquad (k = 2, 3, \ldots, n). \tag{4.197}$$

Remarks:
1. Here $(\underline{\mathbf{x}}_k,\ \underline{\mathbf{y}}_i) = \underline{\mathbf{x}}_k^{\mathrm{T}}\, \underline{\mathbf{y}}_i$ is the scalar product of the vectors $\underline{\mathbf{x}}_k$ and $\underline{\mathbf{y}}_i$.
2. Corresponding to the orthogonal system of the vectors $\underline{\mathbf{y}}_1, \underline{\mathbf{y}}_2, \ldots, \underline{\mathbf{y}}_n$ one gets the orthonormal system $\tilde{\underline{\mathbf{x}}}_1, \tilde{\underline{\mathbf{x}}}_2, \ldots, \tilde{\underline{\mathbf{x}}}_n$ with $\tilde{\underline{\mathbf{x}}}_1 = \dfrac{\underline{\mathbf{y}}_1}{||\underline{\mathbf{y}}_1||}$, $\tilde{\underline{\mathbf{x}}}_2 = \dfrac{\underline{\mathbf{y}}_2}{||\underline{\mathbf{y}}_2||}, \ldots,$ $\tilde{\underline{\mathbf{x}}}_n = \dfrac{\underline{\mathbf{y}}_n}{||\underline{\mathbf{y}}_n||}$, where $||\underline{\mathbf{y}}_i|| = \sqrt{(\underline{\mathbf{y}}_i, \underline{\mathbf{y}}_i)}$ is the Euclidean norm of the vector $\underline{\mathbf{y}}_i$.

■ $\underline{\mathbf{x}}_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$, $\underline{\mathbf{x}}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$, $\underline{\mathbf{x}}_3 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$. From here it follows:

$\underline{\mathbf{y}}_1 = \underline{\mathbf{x}}_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$ and $\tilde{\underline{\mathbf{x}}}_1 = \dfrac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$; $\underline{\mathbf{y}}_2 = \underline{\mathbf{x}}_2 - \dfrac{\left(\underline{\mathbf{x}}_2\,,\, \underline{\mathbf{y}}_1\right)}{\left(\underline{\mathbf{y}}_1\,,\, \underline{\mathbf{y}}_1\right)} \underline{\mathbf{y}}_1 = \begin{pmatrix} 1 \\ -1/2 \\ 1/2 \end{pmatrix}$ and $\tilde{\underline{\mathbf{x}}}_2 = \dfrac{1}{\sqrt{6}} \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}$;

$\underline{\mathbf{y}}_3 = \underline{\mathbf{x}}_3 - \dfrac{(\underline{\mathbf{x}}_3, \underline{\mathbf{y}}_1)}{(\underline{\mathbf{y}}_1, \underline{\mathbf{y}}_1)} \underline{\mathbf{y}}_1 - \dfrac{(\underline{\mathbf{x}}_3, \underline{\mathbf{y}}_2)}{(\underline{\mathbf{y}}_2, \underline{\mathbf{y}}_2)} \underline{\mathbf{y}}_2 = \begin{pmatrix} 2/3 \\ 2/3 \\ -2/3 \end{pmatrix}$ and $\tilde{\underline{\mathbf{x}}}_3 = \dfrac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$.

2. Transformation of Principal Axes, Similarity Transformation

For every real symmetric matrix $\mathbf{A}$, there is an orthogonal matrix $\mathbf{U}$ and a diagonal matrix $\mathbf{D}$ such that

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{T}}. \tag{4.198}$$

The diagonal elements of $\mathbf{D}$ are the eigenvalues of $\mathbf{A}$, and the columns of $\mathbf{U}$ are the corresponding normed eigenvectors. From (4.198) it is obvious that

$$\mathbf{D} = \mathbf{U}^{\mathrm{T}}\mathbf{A}\mathbf{U}. \tag{4.199}$$

Transformation (4.199) is called the *transformation of principal axes*. In this way $\mathbf{A}$ is reduced to a diagonal matrix (see also 4.1.2, **2.**, p. 270).
If the square matrix $\mathbf{A}$ (not necessarily symmetric) is transformed by a square regular matrix $\mathbf{G}$ such a way that

$$\mathbf{G}^{-1}\mathbf{A}\,\mathbf{G} = \tilde{\mathbf{A}} \tag{4.200}$$

then it is called a *similarity transformation*. The matrices $\mathbf{A}$ and $\tilde{\mathbf{A}}$ are called *similar* and they have the following properties:

1. The matrices $\mathbf{A}$ and $\tilde{\mathbf{A}}$ have the same eigenvalues, i.e., the similarity transformation does not affect the eigenvalues.

2. If $\mathbf{A}$ is symmetric and $\mathbf{G}$ is orthogonal, then $\tilde{\mathbf{A}}$ is symmetric, too:

$$\tilde{\mathbf{A}} = \mathbf{G}^{\mathrm{T}}\mathbf{A}\,\mathbf{G} \qquad \text{with} \qquad \mathbf{G}^{\mathrm{T}}\mathbf{G} = \mathbf{I}. \tag{4.201}$$

The relation (4.201) is called an *orthogonal-similarity transformation*. In this context (4.199) means that a real symmetric matrix $\mathbf{A}$ can be transformed orthogonally similar to a real diagonal form $\mathbf{D}$.

4.6.2.3 Transformation of Principal Axes of Quadratic Forms

1. Real Quadratic Form, Definition

A real quadratic form Q of variables $x_1, x_2, \ldots, x_n$ has the form

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j = \underline{\mathbf{x}}^{\mathrm{T}} \mathbf{A} \underline{\mathbf{x}}, \tag{4.202}$$

where $\underline{\mathbf{x}} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$ is the vector of real variables and the matrix $\mathbf{A} = (a_{ij})$ is a real symmetric matrix.
The form Q is called *positive definite* or *negative definite*, if it takes only positive or only negative values respectively, and it takes the zero value only in the case $x_1 = x_2 = \ldots = x_n = 0$.
The form Q is called *positive* or *negative semidefinite*, if it takes non-zero values only with the sign according to its name, but it can take the zero value for non-zero vectors, too.
A real quadratic form is called *indefinite* if it takes both positive and negative values. According to the behavior of Q the associated real symmetric matrix $\mathbf{A}$ is called positive or negative definite, semidefinite or indefinite.

2. Real Positive Definite Quadratic Form, Properties

1. In a real positive definite quadratic form Q all elements of the main diagonal of the corresponding real symmetric matrix $\mathbf{A}$ are positive, i.e.,

$$a_{ii} > 0 \quad (i = 1, 2, \ldots, n) \tag{4.203}$$

holds. (4.203) represents a very important property of positive definite matrices.

2. A real quadratic form Q is positive definite if and only if all eigenvalues of the corresponding matrix $\mathbf{A}$ are positive.

3. Suppose the rank of the matrix $\mathbf{A}$ corresponding to the real quadratic form $Q = \underline{\mathbf{x}}^{\mathrm{T}}\mathbf{A}\underline{\mathbf{x}}$ is equal to r. Then the quadratic form can be transformed by a linear transformation

$$\underline{\mathbf{x}} = \mathbf{C}\underline{\tilde{\mathbf{x}}} \tag{4.204}$$

into a sum of pure quadratic terms, into the so-called *normal form*

$$Q = \underline{\tilde{\mathbf{x}}}^{\mathrm{T}}\mathbf{K}\underline{\tilde{\mathbf{x}}} = \sum_{i=1}^{r} p_i \tilde{x}_i^{\,2} \tag{4.205}$$

where $p_i = (\operatorname{sign} \lambda_i)k_i$ and $k_1, k_2, \ldots, k_r$ are arbitrary, previously given, positive constants.

Remark: Regardless of the non-singular transformation (4.204) that transforms the real quadratic form of rank r into the normal form (4.205), the number p of positive coefficients and the number $q = r - p$ of negative coefficients among the p_i of the normal form are invariant (the *inertia theorem of Sylvester*). The value p is called the *index of inertia of the quadratic form*.

3. Generation of the Normal Form

A practical method to use the transformation (4.205) follows from the transformation of principal axes (4.199). First it is to perform a rotation on the coordinate system by the orthogonal matrix $\mathbf{U}$, whose columns are the eigenvectors of $\mathbf{A}$ (i.e., the directions of the axes of the new coordinate system are the directions of the eigenvectors). This gives the form

$$Q = \underline{\tilde{\mathbf{x}}}^{\mathrm{T}}\mathbf{L}\underline{\tilde{\mathbf{x}}} = \sum_{i=1}^{r} \lambda_i \tilde{x}_i^{\,2}. \tag{4.206}$$

Here $\mathbf{L}$ is a diagonal matrix with the eigenvalues of $\mathbf{A}$ in the diagonal. Then a dilatation is performed by the diagonal matrix $\mathbf{D}$ whose diagonal elements are $d_i = \sqrt{\frac{k_i}{|\lambda_i|}}$. The whole transformation is now given by the matrix

$$\mathbf{C} = \mathbf{U}\,\mathbf{D}, \tag{4.207}$$

and one gets:

$$\begin{aligned} Q &= \underline{\tilde{\mathbf{x}}}^{\mathrm{T}}\mathbf{A}\underline{\tilde{\mathbf{x}}} = (\mathbf{U}\,\mathbf{D}\underline{\tilde{\mathbf{x}}})^{\mathrm{T}}\mathbf{A}(\mathbf{U}\,\mathbf{D}\underline{\tilde{\mathbf{x}}}) = \tilde{\underline{\mathbf{x}}}^{\mathrm{T}}(\mathbf{D}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}\mathbf{A}\,\mathbf{U}\,\mathbf{D})\tilde{\underline{\mathbf{x}}} \\ &= \underline{\tilde{\mathbf{x}}}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{L}\,\mathbf{D}\underline{\tilde{\mathbf{x}}} = \underline{\tilde{\mathbf{x}}}^{\mathrm{T}}\mathbf{K}\underline{\tilde{\mathbf{x}}}. \end{aligned} \tag{4.208}$$

Remark: The transformation of principal axes of quadratic forms plays an essential role at the classification of curves and surfaces of second order (see 3.5.2.11, p. 206 and 3.5.3.14, p. 228).

4. Jordan Normal Form

Let $\mathbf{A}$ be an arbitrary real or complex (n, n) matrix. Then there exists a non-singular matrix $\mathbf{T}$ such that

$$\mathbf{T}^{-1}\mathbf{A}\mathbf{T} = \mathbf{J} \tag{4.209}$$

holds, where $\mathbf{J}$ is called the *Jordan matrix* or *Jordan normal form* of $\mathbf{A}$. The Jordan matrix has a block diagonal structure of the form (4.210), where the elemnts $\mathbf{J}_j$ of $\mathbf{J}$ are called *Jordan blocks*:

$$\mathbf{J} = \begin{pmatrix} \mathbf{J}_1 & & & \\ & \mathbf{J}_2 & & \mathbf{O} \\ & & \ddots & \\ & \mathbf{O} & & \mathbf{J}_{k-1} \\ & & & & \mathbf{J}_k \end{pmatrix}. \tag{4.210}$$

$$\mathbf{J} = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \mathbf{O} \\ & & \ddots & \\ & \mathbf{O} & & \lambda_{n-1} \\ & & & & \lambda_n \end{pmatrix}. \tag{4.211}$$

They have the following structure:

1. If $\mathbf{A}$ has only single eigenvalues λ_j, then $\mathbf{J}_j = \lambda_j$ and $k = n$, i.e., $\mathbf{J}$ is a diagonal matrix (4.211).

2. If λ_j is an eigenvalue of multiplicity p_j, then there are one or more blocks of the form (4.212) where the sum of the sizes of all such blocks is equal to p_j and $\sum_{j=1}^{k} p_j = n$ holds. The exact structure of a Jordan block depends on the structure of the elementary divisors of the characteristic matrix $\mathbf{A} - \lambda\mathbf{I}$.

$$\mathbf{J}_j = \begin{pmatrix} \lambda_j & 1 & & \\ & \lambda_j & 1 & \mathbf{O} \\ & & \ddots & \ddots \\ & & & \lambda_j & 1 \\ & \mathbf{O} & & & \lambda_j \end{pmatrix}, \tag{4.212}$$

For further information see [4.15], [19.16] vol. 1.

4.6.2.4 Suggestions for the Numerical Calculations of Eigenvalues

1. Eigenvalues can be calculated as the roots of the characteristic equation (4.195b) (see examples on p. 315). In order to get them the coefficients a_i $(i = 0, 1, 2, \ldots, n-1)$ of the characteristic polynomial of the matrix $\mathbf{A}$ must be determined. However, one should avoid this method of calculation, because this procedure is extremely unstable, i.e., small changes in the coefficients a_i of the polynomial result in big changes in the roots λ_j.

2. There are many algorithms for the solution of the eigenvalue problem of symmetric matrices. Two types can be distinguished (see [4.7]):

a) Transformation methods, for instance the Jacobi method, Householder tridiagonalization, QR algorithm.

b) Iterative methods, for instance vector iteration, the Rayleigh-Ritz algorithm, inverse iteration, the Lanczos method, the bisection method. As an example the *power method of Mises* is discussed here.

3. The Power Method of Mises Assume that $\mathbf{A}$ is real and symmetric and has a unique dominant eigenvalue. This iteration method determines this eigenvalue and the associated eigenvector. Let the dominant eigenvalue be denoted by λ_1, that is,

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_n|. \tag{4.213}$$

Let $\underline{\mathbf{x}}_1, \underline{\mathbf{x}}_2, \ldots, \underline{\mathbf{x}}_n$ be the associated linearly independent eigenvectors. Then:

1. $\mathbf{A}\underline{\mathbf{x}}_i = \lambda_i \underline{\mathbf{x}}_i \quad (i = 1, 2, \ldots, n).$ (4.214)

2. Each element $\underline{\mathbf{x}} \in \mathbb{R}^n$ can be expressed as a linear combination of these eigenvectors $\underline{\mathbf{x}}_i$:

$$\underline{\mathbf{x}} = c_1\underline{\mathbf{x}}_1 + c_2\underline{\mathbf{x}}_2 + \cdots + c_n\underline{\mathbf{x}}_n \quad (c_i \text{ const};\ \ i = 1, 2, \ldots, n). \tag{4.215}$$

Multiplying both sides of (4.215) by $\mathbf{A}$ k times, then using (4.214) follows

$$\mathbf{A}^k\underline{\mathbf{x}} = c_1\lambda_1^k\underline{\mathbf{x}}_1 + c_2\lambda_2^k\underline{\mathbf{x}}_2 + \cdots + c_n\lambda_n^k\underline{\mathbf{x}}_n = \lambda_1^k[c_1\underline{\mathbf{x}}_1 + c_2\left(\frac{\lambda_2}{\lambda_1}\right)^k\underline{\mathbf{x}}_2 + \cdots + c_n\left(\frac{\lambda_n}{\lambda_1}\right)^k\underline{\mathbf{x}}_n]. \tag{4.216}$$

From this relation and (4.213) one can see that

$$\frac{\mathbf{A}^k\underline{\mathbf{x}}}{\lambda_1^k c_1} \longrightarrow \underline{\mathbf{x}}_1 \text{ as } k \to \infty\,, \quad \text{that is, } \mathbf{A}^k\underline{\mathbf{x}} \approx c_1\lambda_1^k\underline{\mathbf{x}}_1. \tag{4.217}$$

This is the basis of the following iteration procedure:
Step 1: Select an arbitrary starting vector $\underline{\mathbf{x}}^{(0)} \in \mathbb{R}^n$.
Step 2: Iterative computation of

$$\underline{\mathbf{x}}^{(k+1)} = \mathbf{A}\underline{\mathbf{x}}^{(k)} \quad (k = 0, 1, 2, \ldots; \underline{\mathbf{x}}^{(0)} \text{ is given}). \tag{4.218}$$

From (4.218) and keeping in mind (4.217) follows:

$$\underline{\mathbf{x}}^{(k)} = \mathbf{A}^k\underline{\mathbf{x}}^{(0)} \approx c_1\lambda_1^k\underline{\mathbf{x}}_1. \tag{4.219}$$

Step 3: From (4.218) and (4.219) it follows that

$$\begin{aligned}
\underline{\mathbf{x}}^{(k+1)} &= \mathbf{A}\underline{\mathbf{x}}^{(k)} = \mathbf{A}(\mathbf{A}^k\underline{\mathbf{x}}^{(0)}),\\
\mathbf{A}(\mathbf{A}^k\underline{\mathbf{x}}^{(0)}) &\approx \mathbf{A}(c_1\lambda_1^k\underline{\mathbf{x}}_1) = c_1\lambda_1^k(\mathbf{A}\underline{\mathbf{x}}_1),\\
c_1(\lambda_1^k\mathbf{A}\underline{\mathbf{x}}_1) &= \lambda_1(c_1\lambda_1^k\underline{\mathbf{x}}_1) \approx \lambda_1\underline{\mathbf{x}}^{(k)}, \text{ therefore}\\
\underline{\mathbf{x}}^{(k+1)} &\approx \lambda_1\underline{\mathbf{x}}^{(k)},
\end{aligned} \tag{4.220}$$

that is, for large values of k the consecutive vectors $\underline{\mathbf{x}}^{(k+1)}$ and $\underline{\mathbf{x}}^{(k)}$ differ approximately by a factor λ_1.
Step 4: Relations (4.219) and (4.220) imply for $\underline{\mathbf{x}}_1$ and λ_1:

$$\underline{\mathbf{x}}_1 \approx \underline{\mathbf{x}}^{(k+1)}, \quad \lambda_1 \approx \frac{\left(\underline{\mathbf{x}}^{(k)}, \underline{\mathbf{x}}^{(k+1)}\right)}{\left(\underline{\mathbf{x}}^{(k)}, \underline{\mathbf{x}}^{(k)}\right)}. \tag{4.221}$$

■ For example, let

$$\mathbf{A} = \begin{pmatrix} 2.23 & -1.15 & 1.77\\ -1.15 & 9.25 & -2.13\\ 1.77 & -2.13 & 1.56 \end{pmatrix}, \quad \underline{\mathbf{x}}^{(0)} = \begin{pmatrix} 1\\ 0\\ 0 \end{pmatrix}.$$

$\underline{\mathbf{x}}^{(0)}$	$\underline{\mathbf{x}}^{(1)}$	$\underline{\mathbf{x}}^{(2)}$	$\underline{\mathbf{x}}^{(3)}$	normalization	$\underline{\mathbf{x}}^{(4)}$	$\underline{\mathbf{x}}^{(5)}$	normalization
1	3.23	14.89	88.27	1	7.58	67.75	1
0	−1.15	−18.12	−208.03	−2.36	−24.93	−256.85	−3.79
0	1.77	10.93	82.00	0.93	8.24	79.37	1.17
λ_1				9.964			10.177

$\underline{\mathbf{x}}^{(6)}$	$\underline{\mathbf{x}}^{(7)}$	normalization	$\underline{\mathbf{x}}^{(8)}$	$\underline{\mathbf{x}}^{(9)}$	normalization
9.66	96.40	1	10.09	102.33	$\begin{pmatrix} 1\\ -4.129\\ 1.229 \end{pmatrix} \approx \underline{\mathbf{x}}_1$
− 38.78	−394.09	−4.09	−41.58	−422.49	
11.67	117.78	1.22	12.38	125.73	
		10.16			$10.161 \approx \lambda_1$

Remarks:

1. Since eigenvectors are unique only up to a constant multiplier, it is preferable to normalize the vectors $\underline{x}^{(k)}$ as shown in the example.

2. The eigenvalue with the smallest absolute value and the associated eigenvector can be obtained by using the power method of Mises for $\mathbf{A}^{-1}$. If $\mathbf{A}^{-1}$ does not exist, then 0 is this eigenvalue and any vector from the null-space of $\mathbf{A}$ can be selected as an associated eigenvector.

3. The other eigenvalues and the associated eigenvectors of $\mathbf{A}$ can be obtained by repeated application

of the following idea. Select a starting vector which is orthogonal to the known vector $\underline{\mathbf{x}}_1$, and in this subspace λ_2 becomes the dominant eigenvalue that can be obtained by using the power method. In order to obtain λ_3, the starting vector has to be orthogonal to both $\underline{\mathbf{x}}_1$ and $\underline{\mathbf{x}}_2$, and so on. This procedure is known as *matrix deflation*.

4. Based on (4.218) the power method is sometimes called *vector iteration*.

4.6.3 Singular Value Decomposition

1. Singular Values and Singular Vectors Let $\mathbf{A}$ be a real matrix of size (m,n) and its rank be equal to r. The matrices $\mathbf{A}\mathbf{A}^{\mathrm{T}}$ and $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ have r non-zero eigenvalues λ_ν, and they are the same for both of the matrices. The positive square roots $d_\nu = \sqrt{\lambda_\nu}$ $(\nu = 1,2,\ldots,r)$ of the eigenvalues λ_ν of the matrix $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ are called the *singular values* of the matrix $\mathbf{A}$. The corresponding eigenvectors $\underline{\mathbf{u}}_\nu$ of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ are called *right-singular vectors* of $\mathbf{A}$, the corresponding eigenvectors $\underline{\mathbf{v}}_\nu$ of $\mathbf{A}\mathbf{A}^{\mathrm{T}}$ *left-singular vectors*:

$$\mathbf{A}^{\mathrm{T}}\mathbf{A}\underline{\mathbf{u}}_\nu = \lambda_\nu \underline{\mathbf{u}}_\nu, \quad \mathbf{A}\mathbf{A}^{\mathrm{T}}\underline{\mathbf{v}}_\nu = \lambda_\nu \underline{\mathbf{v}}_\nu \quad (\nu = 1,2,\ldots,r). \tag{4.222a}$$

The relations between the right and left-singular vectors are:

$$\mathbf{A}\underline{\mathbf{u}}_\nu = d_\nu \underline{\mathbf{v}}_\nu, \quad \mathbf{A}^{\mathrm{T}}\underline{\mathbf{v}}_\nu = d_\nu \underline{\mathbf{u}}_\nu. \tag{4.222b}$$

A matrix $\mathbf{A}$ of size (m,n) with rank r has r positive-singular values d_ν $(\nu = 1,2,\ldots,r)$. There exist r orthonormalized right-singular vectors $\underline{\mathbf{u}}_\nu$ and r orthonormalized left-singular vectors $\underline{\mathbf{v}}_\nu$. Furthermore, there exist to the zero-singular value $n-r$ orthonormalized right-singular vectors $\underline{\mathbf{u}}_\nu (\nu = r+1,\ldots,n)$ and $m-r$ orthonormalized left-singular vectors $\underline{\mathbf{v}}_\nu$ $(\nu = r+1,\ldots,m)$. Consequently, a matrix of size (m,n) has n right-singular vectors and m left-singular vectors, and two orthogonal matrices can be made from them (see 4.1.4, **9.**, p. 275):

$$\mathbf{U} = (\underline{\mathbf{u_1}}, \underline{\mathbf{u_2}}, \ldots, \underline{\mathbf{u_n}}), \quad \mathbf{V} = (\underline{\mathbf{v_1}}, \underline{\mathbf{v_2}}, \ldots, \underline{\mathbf{v_m}}). \tag{4.223}$$

2. Singular-Value Decomposition The representation

$$\mathbf{A} = \mathbf{V}\hat{\mathbf{A}}\mathbf{U}^{\mathrm{T}} \quad (4.224a) \qquad \text{with } \hat{\mathbf{A}} = \left(\begin{array}{cccc|c} d_1 & 0 & \cdots & 0 & 0\cdots 0 \\ 0 & d_2 & & 0 & 0 \quad \vdots \\ \vdots & & \ddots & \vdots & \vdots \\ & & & 0 & \\ 0 & \cdots & & d_r & 0\cdots 0 \\ \hline 0 & 0 & \cdots & 0 & 0\cdots 0 \\ 0 & & & 0 & \vdots \\ \vdots & & & \vdots & \\ 0 & 0 & \cdots & 0 & 0\cdots 0 \end{array}\right) \begin{array}{l} \left.\begin{array}{l}\\\\\\\\\end{array}\right\} r \text{ rows} \\ \left.\begin{array}{l}\\\\\\\end{array}\right\} m-r \text{ rows} \end{array} \tag{4.224b}$$

$$\underbrace{\qquad\qquad}_{r \text{ columns}} \quad \underbrace{\qquad}_{n-r \text{ columns}}$$

is called the *singular-value decomposition* of the matrix $\mathbf{A}$. The matrix $\hat{\mathbf{A}}$, as the matrix $\mathbf{A}$, is of size (m,n) and has only zero elements except the first r diagonal elements $a_{\nu\nu} = d_\nu$ $(\nu = 1,2,\ldots,r)$. The values d_ν are the singular values of $\mathbf{A}$.

Remark: If substituting A^{H} instead of A^{T} and consider unitary matrices U and V instead of orthogonals, then all the statements about singular-value decomposition are valid also for matrices with complex elements.

3. Application Singular-value decomposition can be used to determine the rank of the matrix $\mathbf{A}$ of size (m,n) and to calculate an approximate solution of the over-determined system of equations

$\mathbf{A}\underline{\mathbf{x}} = \underline{\mathbf{b}}$ (see 4.5.3.1, p. 313) with the so-called *regularization method*, i.e., to solve the problem

$$||\mathbf{A}\underline{\mathbf{x}} - \underline{\mathbf{b}}||^2 + \alpha||\underline{\mathbf{x}}||^2 = \sum_{i=1}^{m}\left[\sum_{k=1}^{n} a_{ik}x_k - b_i\right]^2 + \alpha\sum_{k=1}^{n} x_k^2 = \text{min!}\,, \tag{4.225}$$

where $\alpha > 0$ is a regularization parameter.