Vector Algebra

A

In various sections of this book we referred to the notion of a vector. We assumed the reader to have a basic knowledge on standard school level. In this appendix we recapitulate some basic notions of vector algebra. For a more detailed presentation we refer to [2].

A.1 Cartesian Coordinate Systems

A *Cartesian coordinate system* in the plane (in space) consists of two (three) real lines *(coordinate axes)* which intersect in right angles at the point O (origin). We always assume that the coordinate system is positively (right-handed) oriented. In a planar right-handed system, the positive y-axis lies to the left in viewing direction of the positive x-axis (Fig. A.1). In a positively oriented three-dimensional coordinate system, the direction of the positive z-axis is obtained by turning the x-axis in the direction of the y-axis according to the *right-hand rule*, see Fig. A.2.

The *coordinates* of a point are obtained by parallel projection of the point onto the coordinate axes. In the case of the plane, the point A has the coordinates a_1 and a_2, and we write

$$A = (a_1, a_2) \in \mathbb{R}^2.$$

In an analogous way a point A in space with coordinates a_1, a_2 and a_3 is denoted as

$$A = (a_1, a_2, a_3) \in \mathbb{R}^3.$$

Thus one has a unique representation of points by pairs or triples of real numbers.

M. Oberguggenberger and A. Ostermann, *Analysis for Computer Scientists*, Undergraduate Topics in Computer Science,
https://doi.org/10.1007/978-3-319-91155-7

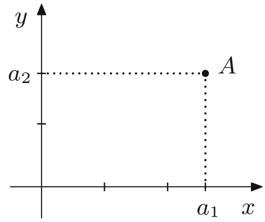

Fig. A.1 Cartesian coordinate system in the plane

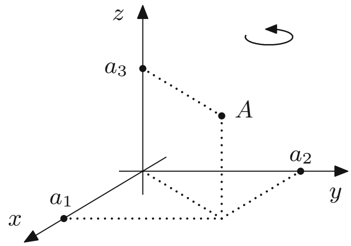

Fig. A.2 Cartesian coordinate system in space

A.2 Vectors

For two points P and Q in the plane (in space) there exists *exactly one* parallel translation which moves P to Q. This translation is called a *vector*. Vectors are thus quantities with *direction and length*. The direction is that from P to Q and the length is the distance between the two points. Vectors are used to model, e.g., forces and velocities. We always write vectors in boldface.

For a vector $\mathbf{a}$, the vector $-\mathbf{a}$ denotes the parallel translation which undoes the action of $\mathbf{a}$; the *zero vector* $\mathbf{0}$ does not cause any translation. The composition of two parallel translations is again a parallel translation. The corresponding operation for vectors is called *addition* and is performed according to the *parallelogram rule*. For a real number $\lambda \geq 0$, the vector $\lambda\,\mathbf{a}$ is the vector which has the same direction as $\mathbf{a}$, but λ times the length of $\mathbf{a}$. This operation is called *scalar multiplication*. For addition and scalar multiplication the usual rules of computation apply.

Let $\mathbf{a}$ be the parallel translation from P to Q. The length of the vector $\mathbf{a}$, i.e. the distance between P and Q, is called *norm* (or *magnitude*) of the vector. We denote it by $\|\mathbf{a}\|$. A vector $\mathbf{e}$ with $\|\mathbf{e}\| = 1$ is called a *unit vector*.

A.3 Vectors in a Cartesian Coordinate System

In a Cartesian coordinate system with origin O, we denote the three unit vectors in direction of the three coordinate axes by $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$, see Fig. A.3. These three vectors are called the *standard basis* of $\mathbb{R}^3$. Here $\mathbf{e}_1$ stands for the parallel translation which moves O to $(1, 0, 0)$, etc.

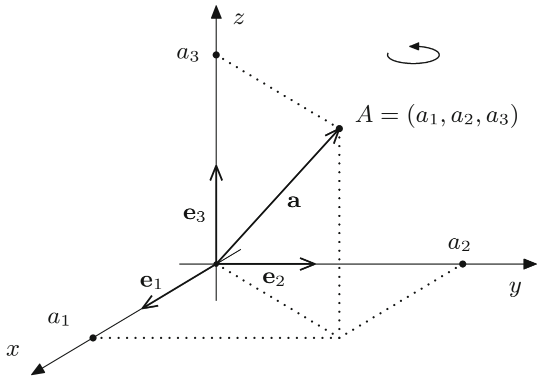

Fig. A.3 Representation of **a** in components

The vector **a** which moves O to A can be decomposed in a unique way as $\mathbf{a} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + a_3\mathbf{e}_3$. We denote it by

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix},$$

where the column on the right-hand side is the so-called *coordinate vector* of **a** with respect to the standard basis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. The vector **a** is also called *position vector* of the point A. Since we are always working with the standard basis, we *identify* a vector with its coordinate vector, i.e.

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

and

$$\mathbf{a} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + a_3\mathbf{e}_3 = \begin{bmatrix} a_1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ a_2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ a_3 \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}.$$

To distinguish between points and vectors we write the coordinates of points in a row, but use column notation for vectors.

For column vectors the usual rules of computation apply:

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \\ a_3 + b_3 \end{bmatrix}, \qquad \lambda \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} \lambda a_1 \\ \lambda a_2 \\ \lambda a_3 \end{bmatrix}.$$

Thus the addition and the scalar multiplication are defined *componentwise*.

The norm of a vector $\mathbf{a} \in \mathbb{R}^2$ with components a_1 and a_2 is computed with Pythagoras' theorem as $\|\mathbf{a}\| = \sqrt{a_1^2 + a_2^2}$. Hence the components of the vector $\mathbf{a}$ have the representation

$$a_1 = \|\mathbf{a}\| \cdot \cos\alpha \qquad \text{and} \qquad a_2 = \|\mathbf{a}\| \cdot \sin\alpha,$$

and we obtain

$$\mathbf{a} = \|\mathbf{a}\| \cdot \begin{bmatrix} \cos\alpha \\ \sin\alpha \end{bmatrix} = \text{length} \cdot \text{direction},$$

see Fig. A.4. For the norm of a vector $\mathbf{a} \in \mathbb{R}^3$ the analogous formula $\|\mathbf{a}\| = \sqrt{a_1^2 + a_2^2 + a_3^2}$ holds.

Remark A.1 The plane $\mathbb{R}^2$ (and likewise the space $\mathbb{R}^3$) appears in two roles: On the one hand as *point space* (its objects are points which cannot be added) and on the other hand as *vector space* (its objects are vectors that can be added). By parallel translation, $\mathbb{R}^2$ (as vector space) can be attached to every point of $\mathbb{R}^2$ (as point space), see Fig. A.5. In general, however, point space and vector space are different sets, as shown in the following example.

Example A.2 (Particle on a circle) Let P be the position of a particle which moves on a circle and $\mathbf{v}$ its velocity vector. Then the point space is the circle and the vector space the tangent to the circle at the point P, see Fig. A.6.

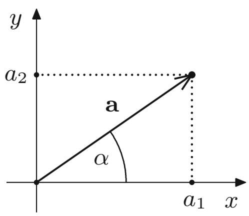

Fig. A.4 A vector **a** with its components a_1 and a_2

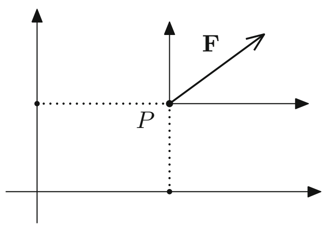

Fig. A.5 Force **F** applied at P

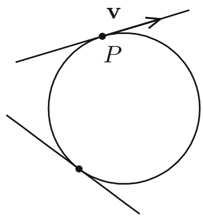

Fig. A.6 Velocity vector is tangential to the circle

A.4 The Inner Product (Dot Product)

The *angle* $\measuredangle(\mathbf{a}, \mathbf{b})$ between two vectors $\mathbf{a}, \mathbf{b}$ is uniquely determined by the condition $0 \le \measuredangle(\mathbf{a}, \mathbf{b}) \le \pi$. One calls a vector $\mathbf{a}$ *orthogonal (perpendicular)* to $\mathbf{b}$ (in symbols: $\mathbf{a} \perp \mathbf{b}$), if $\measuredangle(\mathbf{a}, \mathbf{b}) = \frac{\pi}{2}$. By definition, the zero vector $\mathbf{0}$ is orthogonal to all vectors.

Definition A.3 Let $\mathbf{a}, \mathbf{b}$ be planar (or spatial) vectors. The number

$$\langle \mathbf{a}, \mathbf{b} \rangle = \begin{cases} \|\mathbf{a}\| \cdot \|\mathbf{b}\| \cdot \cos \measuredangle(\mathbf{a}, \mathbf{b}) & \mathbf{a} \neq 0,\ \mathbf{b} \neq 0, \\ 0 & \text{otherwise,} \end{cases}$$

is called the *inner product (dot product)* of $\mathbf{a}$ and $\mathbf{b}$.

For planar vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$ the inner product is calculated from their components as

$$\langle \mathbf{a}, \mathbf{b} \rangle = \left\langle \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \right\rangle = a_1 b_1 + a_2 b_2.$$

For vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$ the analogous formula holds:

$$\langle \mathbf{a}, \mathbf{b} \rangle = \left\langle \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}, \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \right\rangle = a_1 b_1 + a_2 b_2 + a_3 b_3.$$

Example A.4 The standard basis vectors $\mathbf{e}_i$ have length 1 and are mutually orthogonal, i.e.

$$\langle \mathbf{e}_i, \mathbf{e}_j \rangle = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

For vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$ and a scalar $\lambda \in \mathbb{R}$ the inner product obeys the rules

(a) $\langle \mathbf{a}, \mathbf{b} \rangle = \langle \mathbf{b}, \mathbf{a} \rangle$,

(b) $\langle \mathbf{a}, \mathbf{a} \rangle = \|\mathbf{a}\|^2$,

(c) $\langle \mathbf{a}, \mathbf{b} \rangle = 0 \quad \Leftrightarrow \quad \mathbf{a} \perp \mathbf{b}$,

(d) $\langle \lambda\mathbf{a}, \mathbf{b}\rangle = \langle \mathbf{a}, \lambda\mathbf{b}\rangle = \lambda\langle \mathbf{a}, \mathbf{b}\rangle$,

(e) $\langle \mathbf{a} + \mathbf{b}, \mathbf{c}\rangle = \langle \mathbf{a}, \mathbf{c}\rangle + \langle \mathbf{b}, \mathbf{c}\rangle$.

Example A.5 For the vectors

$$\mathbf{a} = \begin{bmatrix} 2 \\ -4 \\ 0 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 6 \\ 3 \\ 4 \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

we have

$$\|\mathbf{a}\|^2 = 4 + 16 = 20, \quad \|\mathbf{b}\|^2 = 36 + 9 + 16 = 61, \quad \|\mathbf{c}\|^2 = 1 + 1 = 2,$$

and

$$\langle \mathbf{a}, \mathbf{b}\rangle = 12 - 12 = 0, \qquad \langle \mathbf{a}, \mathbf{c}\rangle = 2.$$

From this we conclude that **a** is perpendicular to **b** and

$$\cos\measuredangle(\mathbf{a}, \mathbf{c}) = \frac{\langle \mathbf{a}, \mathbf{c}\rangle}{\|\mathbf{a}\| \cdot \|\mathbf{c}\|} = \frac{2}{\sqrt{20}\sqrt{2}} = \frac{1}{\sqrt{10}}.$$

The value of the angle between **a** and **c** is thus

$$\measuredangle(\mathbf{a}, \mathbf{c}) = \arccos\frac{1}{\sqrt{10}} = 1.249 \text{ rad.}$$

A.5 The Outer Product (Cross Product)

For vectors **a**, **b** in $\mathbb{R}^2$ one defines

$$\mathbf{a} \times \mathbf{b} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \times \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \det\begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} = a_1b_2 - a_2b_1 \in \mathbb{R},$$

the *cross product* of **a** and **b**. An elementary calculation shows that

$$|\mathbf{a} \times \mathbf{b}| = \|\mathbf{a}\| \cdot \|\mathbf{b}\| \cdot \sin\measuredangle(\mathbf{a}, \mathbf{b}).$$

Thus $|\mathbf{a} \times \mathbf{b}|$ is the *area* of the parallelogram spanned by **a** and **b**.

For vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$ one defines the *cross product* as

$$\mathbf{a} \times \mathbf{b} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \times \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} a_2b_3 - a_3b_2 \\ a_3b_1 - a_1b_3 \\ a_1b_2 - a_2b_1 \end{bmatrix} \in \mathbb{R}^3.$$

This product has the following geometric interpretation: If $\mathbf{a} = \mathbf{0}$ or $\mathbf{b} = \mathbf{0}$ or $\mathbf{a} = \lambda\mathbf{b}$ then $\mathbf{a} \times \mathbf{b} = \mathbf{0}$. Otherwise $\mathbf{a} \times \mathbf{b}$ is the vector

(a) which is *perpendicular* to $\mathbf{a}$ and $\mathbf{b}$: $\langle \mathbf{a} \times \mathbf{b}, \mathbf{a} \rangle = \langle \mathbf{a} \times \mathbf{b}, \mathbf{b} \rangle = 0$;

(b) which is directed such that $\mathbf{a}, \mathbf{b}, \mathbf{a} \times \mathbf{b}$ forms a *right-handed system*;

(c) whose length is equal to the *area* F of the *parallelogram* spanned by $\mathbf{a}$ and $\mathbf{b}$: $F = \|\mathbf{a} \times \mathbf{b}\| = \|\mathbf{a}\| \cdot \|\mathbf{b}\| \cdot \sin \measuredangle(\mathbf{a}, \mathbf{b})$.

Example A.6 Let E be the plane spanned by the two vectors

$$\mathbf{a} = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

Then

$$\mathbf{a} \times \mathbf{b} = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} \times \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

is a vector perpendicular to this plane.

For $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$ and $\lambda \in \mathbb{R}$ the following rules apply

(a) $\mathbf{a} \times \mathbf{a} = \mathbf{0}, \quad \mathbf{a} \times \mathbf{b} = -(\mathbf{b} \times \mathbf{a})$,

(b) $\lambda(\mathbf{a} \times \mathbf{b}) = (\lambda \mathbf{a}) \times \mathbf{b} = \mathbf{a} \times (\lambda \mathbf{b})$,

(c) $(\mathbf{a} + \mathbf{b}) \times \mathbf{c} = \mathbf{a} \times \mathbf{c} + \mathbf{b} \times \mathbf{c}$.

However, the cross product is *not associative* and

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) \neq (\mathbf{a} \times \mathbf{b}) \times \mathbf{c}$$

for general $\mathbf{a}, \mathbf{b}, \mathbf{c}$. For instance, the standard basis vectors of the $\mathbb{R}^3$ satisfy the following identities

$$\mathbf{e}_1 \times (\mathbf{e}_1 \times \mathbf{e}_2) = \mathbf{e}_1 \times \mathbf{e}_3 = -\mathbf{e}_2,$$
$$(\mathbf{e}_1 \times \mathbf{e}_1) \times \mathbf{e}_2 = \mathbf{0} \times \mathbf{e}_2 = \mathbf{0}.$$

A.6 Straight Lines in the Plane

The general equation of a straight line in the (x, y)-plane is

$$ax + by = c,$$

where at least one of the coefficients a and b must be different from zero. The straight line consists of all points (x, y) which satisfy the above equation,

$$g = \left\{(x, y) \in \mathbb{R}^2; ax + by = c\right\}.$$

If $b = 0$ (and thus $a \neq 0$) we get

$$x = \frac{c}{a},$$

and thus a line parallel to the y-axis. If $b \neq 0$, one can solve for y and obtains the standard form of a straight line

$$y = -\frac{a}{b}x + \frac{c}{b} = kx + d$$

with *slope* k and *intercept* d.

The *parametric representation* of the straight line is obtained from the general solution of the linear equation

$$ax + by = c.$$

Since this equation is underdetermined, one replaces the independent variable by a parameter and solves for the other variable.

Example A.7 In the equation

$$y = kx + d$$

x is considered as independent variable. One sets $x = \lambda$ and obtains $y = k\lambda + d$ and thus the parametric representation

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ d \end{bmatrix} + \lambda \begin{bmatrix} 1 \\ k \end{bmatrix}, \quad \lambda \in \mathbb{R}.$$

Example A.8 In the equation

$$x = 4$$

y is the independent variable (it does not even appear). This straight line in parametric representation is

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \end{bmatrix} + \lambda \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

In general, the parametric representation of a straight line is of the form

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} p \\ q \end{bmatrix} + \lambda \begin{bmatrix} u \\ v \end{bmatrix}, \quad \lambda \in \mathbb{R}$$

(position vector of a point plus a multiple of a direction vector). A vector perpendicular to this straight line is called a *normal vector*. It is a multiple of

$$\begin{bmatrix} v \\ -u \end{bmatrix}, \quad \text{since} \quad \left\langle \begin{bmatrix} u \\ v \end{bmatrix}, \begin{bmatrix} v \\ -u \end{bmatrix} \right\rangle = 0.$$

The conversion to the nonparametric form is obtained by multiplying the equation in parametric form by a normal vector. Thereby the parameter is eliminated. In the example above one obtains

$$vx - uy = pv - qu.$$

In particular, the coefficients of x and y in the nonparametric form are just the components of a normal vector of the straight line.

A.7 Planes in Space

The general form of a plane in $\mathbb{R}^3$ is

$$ax + by + cz = d,$$

where at least one of the coefficients a, b, c is different from zero. The plane consists of all points which satisfy the above equation, i.e.

$$E = \left\{(x, y, z) \in \mathbb{R}^3; ax + by + cz = d\right\}.$$

Since at least one of the coefficients is nonzero, one can solve the equation for the corresponding unknown.

For example, if $c \neq 0$ one can solve for z to obtain

$$z = -\frac{a}{c}x - \frac{b}{c}y + \frac{d}{c} = kx + ly + e.$$

Here k represents the slope in x-direction, l is the slope in y-direction and e is the intercept on the z-axis (because $z = e$ for $x = y = 0$). By introducing parameters for the independent variables x and y

$$x = \lambda, \quad y = \mu, \quad z = k\lambda + l\mu + e$$

one thus obtains the *parametric representation* of the plane:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ e \end{bmatrix} + \lambda \begin{bmatrix} 1 \\ 0 \\ k \end{bmatrix} + \mu \begin{bmatrix} 0 \\ 1 \\ l \end{bmatrix}, \qquad \lambda, \mu \in \mathbb{R}.$$

In general, the parametric representation of a plane in $\mathbb{R}^3$ is

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} p \\ q \\ r \end{bmatrix} + \lambda \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} + \mu \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$$

with $\mathbf{v} \times \mathbf{w} \neq \mathbf{0}$. If one multiplies this equation with $\mathbf{v} \times \mathbf{w}$ and uses

$$\langle \mathbf{v}, \mathbf{v} \times \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \times \mathbf{w} \rangle = 0,$$

one again obtains the *nonparametric* form

$$\left\langle \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \mathbf{v} \times \mathbf{w} \right\rangle = \left\langle \begin{bmatrix} p \\ q \\ r \end{bmatrix}, \mathbf{v} \times \mathbf{w} \right\rangle.$$

Example A.9 We compute the nonparametric form of the plane

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix} + \lambda \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} + \mu \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

A normal vector to this plane is given by

$$\mathbf{v} \times \mathbf{w} = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} \times \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix},$$

and thus the equation of the plane is

$$-x + y + z = -1.$$

A.8 Straight Lines in Space

A straight line in $\mathbb{R}^3$ can be seen as the *intersection of two planes*:

$$g : \begin{cases} ax + by + cz = d, \\ ex + fy + gz = h. \end{cases}$$

The straight line is the set of all points (x, y, z) which fulfil this system of equations (two equations in three unknowns). Generically, the solution of the above system can be parametrised by one parameter (this is the case of a straight line). However, it may also happen that the planes are parallel. In this situation they either coincide, or they do not intersect at all.

A straight line can also be represented *parametrically* by the position vector of a point and an arbitrary multiple of a direction vector

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} p \\ q \\ r \end{bmatrix} + \lambda \begin{bmatrix} u \\ v \\ w \end{bmatrix}, \quad \lambda \in \mathbb{R}.$$

The direction vector is obtained as difference of the position vectors of two points on the straight line.

Example A.10 We want to determine the straight line through the points $P = (1, 2, 0)$ and $Q = (3, 1, 2)$. A direction vector $\mathbf{a}$ of this line is given by

$$\mathbf{a} = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix}.$$

Thus a parametric representation of the straight line is

$$g : \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + \lambda \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix}, \quad \lambda \in \mathbb{R}.$$

The conversion from parametric to nonparametric form and vice versa is achieved by *elimination* or *introduction* of a parameter λ. In the example above one computes $z = 2\lambda$ from the last equation and inserts it into the first two equations. This yields the nonparametric form

$$\begin{aligned} x - z &= 1, \\ 2y + z &= 4. \end{aligned}$$

Matrices

B

In this book matrix algebra is required in multi-dimensional calculus, for systems of differential equations and for linear regression. This appendix serves to outline the basic notions. A more detailed presentation can be found in [2].

B.1 Matrix Algebra

An $(m \times n)$-*matrix* $\mathbf{A}$ is a rectangular scheme of the form

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}.$$

The *entries (coefficients, elements)* a_{ij}, $i = 1, \dots, m$, $j = 1, \dots, n$ of the matrix $\mathbf{A}$ are real or complex numbers. In this section we restrict ourselves to real numbers. An $(m \times n)$-matrix has m rows and n columns; if $m = n$, and the matrix is called *square*. Vectors of length m can be understood as matrices with one column, i.e as $(m \times 1)$-matrices. In particular, one refers to the columns

$$\mathbf{a}_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}, \quad j = 1, \dots, n$$

of a matrix $\mathbf{A}$ as *column vectors* and accordingly also writes

$$\mathbf{A} = [\mathbf{a}_1 \vdots \mathbf{a}_2 \vdots \dots \vdots \mathbf{a}_n]$$

M. Oberguggenberger and A. Ostermann, *Analysis for Computer Scientists*,
Undergraduate Topics in Computer Science,
https://doi.org/10.1007/978-3-319-91155-7

for the matrix. The rows of the matrix are sometimes called *row vectors*.

The *product* of an $(m \times n)$-matrix $\mathbf{A}$ with a vector $\mathbf{x}$ of length n is defined as

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \qquad \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \end{bmatrix}$$

and results in a vector $\mathbf{y}$ of length m. The kth entry of $\mathbf{y}$ is obtained by the inner product of the kth row vector of the matrix $\mathbf{A}$ (written as a column) with the vector $\mathbf{x}$.

Example B.1 For instance, the product of a (2×3)-matrix with a vector of length 3 is computed as follows:

$$\mathbf{A} = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}, \quad \mathbf{A}\mathbf{x} = \begin{bmatrix} 3a - b + 2c \\ 3d - e + 2f \end{bmatrix}.$$

The assignment $\mathbf{x} \mapsto \mathbf{y} = \mathbf{A}\mathbf{x}$ defines a *linear mapping* from $\mathbb{R}^n$ to $\mathbb{R}^m$. The linearity is characterised by the validity of the relations

$$\mathbf{A}(\mathbf{u} + \mathbf{v}) = \mathbf{A}\mathbf{u} + \mathbf{A}\mathbf{v}, \qquad \mathbf{A}(\lambda\mathbf{u}) = \lambda\mathbf{A}\mathbf{u}$$

for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$, which follow immediately from the definition of matrix multiplication. If $\mathbf{e}_j$ is the jth standard basis vector of $\mathbb{R}^n$, then obviously

$$\mathbf{a}_j = \mathbf{A}\mathbf{e}_j.$$

This means that the columns of the matrix $\mathbf{A}$ are just the images of the standard basis vectors under the linear mapping defined by $\mathbf{A}$.

Matrix arithmetic. Matrices of the same format can be added and subtracted by adding or subtracting their components. Multiplication with a number $\lambda \in \mathbb{R}$ is also defined componentwise. The *transpose* $\mathbf{A}^\mathsf{T}$ of a matrix $\mathbf{A}$ is obtained by swapping rows and columns; i.e. the ith row of the matrix $\mathbf{A}^\mathsf{T}$ consists of the elements of the ith column of $\mathbf{A}$:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}, \quad \mathbf{A}^\mathsf{T} = \begin{bmatrix} a_{11} & a_{21} & \ldots & a_{m1} \\ a_{12} & a_{22} & \ldots & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \ldots & a_{mn} \end{bmatrix}.$$

By transposition an $(m \times n)$-matrix becomes an $(n \times m)$-matrix. In particular, transposition changes a column vector into a row vector and vice versa.

Example B.2 For the matrix **A** and the vector **x** from Example B.1 we have:

$$\mathbf{A}^\mathsf{T} = \begin{bmatrix} a & d \\ b & e \\ c & f \end{bmatrix}, \quad \mathbf{x}^\mathsf{T} = \begin{bmatrix} 3 & -1 & 2 \end{bmatrix}, \quad \mathbf{x} = [3 \ -1 \ 2]^\mathsf{T}.$$

If $\mathbf{a}, \mathbf{b}$ are vectors of length n, then one can regard $\mathbf{a}^\mathsf{T}$ as a $(1 \times n)$-matrix. Its product with the vector $\mathbf{b}$ is defined as above and coincides with the inner product:

$$\mathbf{a}^\mathsf{T}\mathbf{b} = \sum_{i=1}^{n} a_i b_i = \langle \mathbf{a}, \mathbf{b} \rangle.$$

More generally, the *product* of an $(m \times n)$-matrix **A** with an $(n \times l)$-matrix **B** can be defined by forming the inner products of the row vectors of **A** with the column vectors of **B**. This means that the element c_{ij} in the ith row and jth column of $\mathbf{C} = \mathbf{AB}$ is obtained by inner multiplication of the ith row of **A** with the jth column of **B**:

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

The result is an $(m \times l)$-matrix. The product is only defined if the dimensions match, i.e. if the number of columns n of **A** is equal to the number of rows of **B**. The matrix product corresponds to the composition of linear mappings. If **B** is the matrix of a linear mapping $\mathbb{R}^l \to \mathbb{R}^n$ and **A** the matrix of a linear mapping $\mathbb{R}^n \to \mathbb{R}^m$, then **AB** is just the matrix of the composition of the two mappings $\mathbb{R}^l \to \mathbb{R}^n \to \mathbb{R}^m$. The transposition of the product is given by the formula

$$(\mathbf{AB})^\mathsf{T} = \mathbf{B}^\mathsf{T}\mathbf{A}^\mathsf{T},$$

which can easily be deduced from the definitions.

Square matrices. The entries $a_{11}, a_{22}, \ldots, a_{nn}$ of an $(n \times n)$-matrix **A** are called the *diagonal elements*. A square matrix **D** is called *diagonal matrix*, if its entries are all zero with the possible exception of the diagonal elements. Special cases are the *zero matrix* and the *unit matrix* of dimension $n \times n$:

$$\mathbf{O} = \begin{bmatrix} 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{bmatrix}, \quad \mathbf{I} = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix}.$$

The unit matrix is the identity with respect to matrix multiplication. For all $(n \times n)$-matrices **A** it holds that $\mathbf{IA} = \mathbf{AI} = \mathbf{A}$. If for a given matrix **A** there exists a matrix **B** with the property

$$\mathbf{BA} = \mathbf{AB} = \mathbf{I},$$

then one calls $\mathbf{A}$ *invertible* or *regular* and $\mathbf{B}$ the *inverse* of $\mathbf{A}$, denoted by

$$\mathbf{B} = \mathbf{A}^{-1}.$$

Let $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{A}$ an invertible $(n \times n)$-matrix and $\mathbf{y} = \mathbf{Ax}$. Then $\mathbf{x}$ can be computed as $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$; in particular, $\mathbf{A}^{-1}\mathbf{Ax} = \mathbf{x}$ and $\mathbf{AA}^{-1}\mathbf{y} = \mathbf{y}$. This shows that the linear mapping $\mathbb{R}^n \to \mathbb{R}^n$ induced by the matrix $\mathbf{A}$ is bijective and $\mathbf{A}^{-1}$ represents the inverse mapping. The bijectivity of $\mathbf{A}$ can be expressed in yet another way. It means that for every $\mathbf{y} \in \mathbb{R}^n$ there is one and only one $\mathbf{x} \in \mathbb{R}^n$ such that

$$\mathbf{Ax} = \mathbf{y}, \qquad \text{or} \qquad \begin{array}{ccccccc} a_{11}x_1 & + & a_{12}x_2 & + \ldots + & a_{1n}x_n & = & y_1, \\ a_{21}x_1 & + & a_{22}x_2 & + \ldots + & a_{2n}x_n & = & y_2, \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + \ldots + & a_{mn}x_n & = & y_n. \end{array}$$

The latter can be considered as a linear system of equations with right-hand side $\mathbf{y}$ and solution $\mathbf{x} = [x_1 \ x_2 \ \ldots \ x_n]^\mathsf{T}$. In other words, the invertibility of a matrix $\mathbf{A}$ is equivalent with the bijectivity of the corresponding linear mapping and equivalent with the unique solvability of the corresponding linear system of equations (for arbitrary right-hand sides).

For the remainder of this appendix we restrict our attention to (2×2)-matrices. Let $\mathbf{A}$ be a (2×2)-matrix with the corresponding system of equations:

$$\mathbf{A} = [\mathbf{a}_1 \vdots \mathbf{a}_2] = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \qquad \begin{array}{l} a_{11}x_1 + a_{12}x_2 = y_1, \\ a_{21}x_1 + a_{22}x_2 = y_2. \end{array}$$

An important role is played by the *determinant* of the matrix $\mathbf{A}$. In the (2×2)-case it is defined as the cross product of the column vectors:

$$\det \mathbf{A} = \mathbf{a}_1 \times \mathbf{a}_2 = a_{11}a_{22} - a_{21}a_{12}.$$

Since $\mathbf{a}_1 \times \mathbf{a}_2 = \|\mathbf{a}_1\| \|\mathbf{a}_2\| \sin \measuredangle(\mathbf{a}_1, \mathbf{a}_2)$, the column vectors $\mathbf{a}_1, \mathbf{a}_2$ are linearly dependent (so—in $\mathbb{R}^2$—multiples of each other), if and only if $\det \mathbf{A} = 0$. The following theorem characterises invertibility in the (2×2)-case completely.

Proposition B.3 *For (2×2)-matrices $\mathbf{A}$ the following statements are equivalent:*

(a) $\mathbf{A}$ *is invertible.*

(b) The linear mapping $\mathbb{R}^2 \to \mathbb{R}^2$ defined by $\mathbf{A}$ is bijective.

(c) The linear system of equations $\mathbf{Ax} = \mathbf{y}$ has a unique solution $\mathbf{x} \in \mathbb{R}^2$ for arbitrary right-hand sides $\mathbf{y} \in \mathbb{R}^2$.

(d) The column vectors of $\mathbf{A}$ are linearly independent.

(e) The linear mapping $\mathbb{R}^2 \to \mathbb{R}^2$ defined by $\mathbf{A}$ is injective.

(f) *The only solution of the linear system of equations* $\mathbf{Ax} = \mathbf{0}$ *is the zero solution* $\mathbf{x} = \mathbf{0}$.

(g) $\det \mathbf{A} \neq 0$.

Proof The equivalence of the statements (a), (b) and (c) was already observed above. The equivalence of (d), (e) and (f) can easily be seen by negation. Indeed, if the column vectors are linearly dependent, then there exists $\mathbf{x} = [x_1 \ \ x_2]^\mathsf{T} \neq \mathbf{0}$ with $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = \mathbf{0}$. On the one hand, this means that the vector $\mathbf{x}$ is mapped to $\mathbf{0}$ by $\mathbf{A}$; thus this mapping is not injective. On the other hand, $\mathbf{x}$ is a nontrivial solution of the linear system of equations $\mathbf{Ax} = \mathbf{0}$. The converse implications are shown in the same way. Thus (d), (e) and (f) are equivalent. The equivalence of (g) and (d) is obvious from the geometric meaning of the determinant. If the determinant does not vanish then

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

is an inverse to $\mathbf{A}$, as can be verified at once. Thus (g) implies (a). Finally, (e) obviously follows from (b). Hence all statements (a)–(g) are equivalent. □

Proposition B.3 holds for matrices of arbitrary dimension $n \times n$. For $n = 3$ one can still use geometrical arguments. The cross product, however, has to be replaced by the triple product $\langle \mathbf{a}_1 \times \mathbf{a}_2, \mathbf{a}_3 \rangle$ of the three column vectors, which then also defines the determinant of the (3×3)-matrix $\mathbf{A}$. In higher dimensions the proof requires tools from combinatorics, for which we refer to the literature.

B.2 Canonical Form of Matrices

In this subsection we will show that every (2×2)-matrix $\mathbf{A}$ is similar to a matrix of standard type, which means that it can be put into standard form by a basis transformation. We need this fact in Sect. 20.1 for the classification and solution of systems of differential equations. The transformation explained below is a special case of the *Jordan canonical form*[1] for $(n \times n)$-matrices.

If $\mathbf{T}$ is an invertible (2×2)-matrix, then the columns $\mathbf{t}_1, \mathbf{t}_2$ form a basis of $\mathbb{R}^2$. This means that every element $\mathbf{x} \in \mathbb{R}^2$ can be written in a unique way as a *linear combination* $c_1\mathbf{t}_1 + c_2\mathbf{t}_2$; the coefficients $c_1, c_2 \in \mathbb{R}$ are the coordinates of $\mathbf{x}$ with respect to $\mathbf{t}_1$ and $\mathbf{t}_2$. One can regard $\mathbf{T}$ as a linear transformation of $\mathbb{R}^2$ which maps the standard basis $\{[1\ 0]^\mathsf{T}, [0\ 1]^\mathsf{T}\}$ to the basis $\{\mathbf{t}_1, \mathbf{t}_2\}$.

Definition B.4 Two matrices $\mathbf{A}, \mathbf{B}$ are called *similar*, if there exists an invertible matrix $\mathbf{T}$ such that $\mathbf{T}^{-1}\mathbf{AT} = \mathbf{B}$.

[1] C. Jordan, 1838–1922.

The three standard types which will define the similarity classes of (2×2)-matrices are of the following form:

type I	type II	type III
$\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$	$\begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$	$\begin{bmatrix} \mu & -\nu \\ \nu & \mu \end{bmatrix}$

Here the coefficients $\lambda_1, \lambda_2, \lambda, \mu, \nu$ are real numbers.

In what follows, we need the notion of eigenvalues and eigenvectors. If the equation

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$$

has a solution $\mathbf{v} \neq \mathbf{0} \in \mathbb{R}^2$ for some $\lambda \in \mathbb{R}$, then λ is called *eigenvalue* and $\mathbf{v}$ *eigenvector* of $\mathbf{A}$. In other words, $\mathbf{v}$ is the solution of the equation

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0},$$

where $\mathbf{I}$ denotes again the unit matrix. For the existence of a nonzero solution $\mathbf{v}$ it is necessary and sufficient that the matrix $\mathbf{A} - \lambda\mathbf{I}$ is not invertible, i.e.

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0.$$

By writing

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

we see that λ has to be a solution of the *characteristic equation*

$$\det \begin{bmatrix} a - \lambda & b \\ c & d - \lambda \end{bmatrix} = \lambda^2 - (a + d)\lambda + ad - bc = 0.$$

If this equation has a real solution λ, then the system of equations $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$ is underdetermined and thus has a nonzero solution $\mathbf{v} = [v_1 \; v_2]^\mathsf{T}$. Hence one obtains the eigenvectors to the eigenvalue λ by solving the linear system

$$\begin{aligned} (a - \lambda)\, v_1 + b\, v_2 &= 0, \\ c\, v_1 + (d - \lambda)\, v_2 &= 0. \end{aligned}$$

Depending on whether the characteristic equation has two real, a double real or two complex conjugate solutions, we obtain one of the three similarity classes of $\mathbf{A}$.

Proposition B.5 *Every* (2×2)*-matrix* $\mathbf{A}$ *is similar to a matrix of type I, II or III.*

Proof (1) The case of two distinct real eigenvalues $\lambda_1 \neq \lambda_2$. With

$$\mathbf{v}_1 = \begin{bmatrix} v_{11} \\ v_{21} \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} v_{12} \\ v_{22} \end{bmatrix}$$

we denote the corresponding eigenvectors. They are linearly independent and thus form a basis of the $\mathbb{R}^2$. Otherwise they would be multiples of each other and so $c\mathbf{v}_1 = \mathbf{v}_2$ for some nonzero $c \in \mathbb{R}$. Applying $\mathbf{A}$ would result in $c\lambda_1\mathbf{v}_1 = \lambda_2\mathbf{v}_2 = \lambda_2 c\mathbf{v}_1$ and thus $\lambda_1 = \lambda_2$ in contradiction to the hypothesis. According to Proposition B.3 the matrix

$$\mathbf{T} = [\mathbf{v}_1 \vdots \mathbf{v}_2] = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}$$

is invertible. Using

$$\mathbf{A}\mathbf{v}_1 = \lambda_1\mathbf{v}_1, \quad \mathbf{A}\mathbf{v}_2 = \lambda_2\mathbf{v}_2,$$

we obtain the identities

$$\mathbf{T}^{-1}\mathbf{A}\mathbf{T} = \mathbf{T}^{-1}\mathbf{A}\,[\mathbf{v}_1 \vdots \mathbf{v}_2] = \mathbf{T}^{-1}[\lambda_1\mathbf{v}_1 \vdots \lambda_2\mathbf{v}_2]$$

$$= \frac{1}{v_{11}v_{22} - v_{21}v_{12}} \begin{bmatrix} v_{22} & -v_{12} \\ -v_{21} & v_{11} \end{bmatrix} \begin{bmatrix} \lambda_1 v_{11} & \lambda_2 v_{12} \\ \lambda_1 v_{21} & \lambda_2 v_{22} \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}.$$

The matrix $\mathbf{A}$ is similar to a diagonal matrix and thus of type I.
(2) The case of a double real eigenvalue $\lambda = \lambda_1 = \lambda_2$. Since

$$\lambda = \frac{1}{2}\Big(a + d \pm \sqrt{(a-d)^2 + 4bc}\Big)$$

is the solution of the characteristic equation, this case occurs if

$$(a-d)^2 = -4bc, \quad \lambda = \frac{1}{2}(a+d)\,.$$

If $b = 0$ and $c = 0$, then $a = d$ and $\mathbf{A}$ is already a diagonal matrix of the form

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

thus of type I. If $b \neq 0$, we compute c from $(a-d)^2 = -4bc$ and find

$$\mathbf{A} - \lambda\mathbf{I} = \begin{bmatrix} a-\lambda & b \\ c & d-\lambda \end{bmatrix} = \begin{bmatrix} \frac{1}{2}(a-d) & b \\ -\frac{1}{4b}(a-d)^2 & -\frac{1}{2}(a-d) \end{bmatrix}.$$

Note that

$$\begin{bmatrix} \frac{1}{2}(a-d) & b \\ -\frac{1}{4b}(a-d)^2 & -\frac{1}{2}(a-d) \end{bmatrix} \begin{bmatrix} \frac{1}{2}(a-d) & b \\ -\frac{1}{4b}(a-d)^2 & -\frac{1}{2}(a-d) \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

or $(\mathbf{A} - \lambda\mathbf{I})^2 = \mathbf{O}$. In this case, $\mathbf{A} - \lambda\mathbf{I}$ is called a *nilpotent matrix*. A similar calculation shows that $(\mathbf{A} - \lambda\mathbf{I})^2 = \mathbf{O}$ if $c \neq 0$. We now choose a vector $\mathbf{v}_2 \in \mathbb{R}^2$ for which $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v}_2 \neq \mathbf{0}$. Due to the above consideration this vector satisfies

$$(\mathbf{A} - \lambda\mathbf{I})^2\mathbf{v}_2 = \mathbf{0}.$$

If we set

$$\mathbf{v}_1 = (\mathbf{A} - \lambda\mathbf{I})\mathbf{v}_2,$$

then obviously

$$\mathbf{A}\mathbf{v}_1 = \lambda\mathbf{v}_1, \quad \mathbf{A}\mathbf{v}_2 = \mathbf{v}_1 + \lambda\mathbf{v}_2.$$

Further $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent (because if $\mathbf{v}_1$ were a multiple of $\mathbf{v}_2$, then $\mathbf{A}\mathbf{v}_2 = \lambda\mathbf{v}_2$ in contradiction to the construction of $\mathbf{v}_2$). We set

$$\mathbf{T} = [\mathbf{v}_1 \vdots \mathbf{v}_2].$$

The computation

$$\mathbf{T}^{-1}\mathbf{A}\mathbf{T} = \mathbf{T}^{-1}[\lambda\mathbf{v}_1 \vdots \mathbf{v}_1 + \lambda\mathbf{v}_2]$$

$$= \frac{1}{v_{11}v_{22} - v_{21}v_{12}} \begin{bmatrix} v_{22} & -v_{12} \\ -v_{21} & v_{11} \end{bmatrix} \begin{bmatrix} \lambda v_{11} & v_{11} + \lambda v_{12} \\ \lambda v_{21} & v_{21} + \lambda v_{22} \end{bmatrix} = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}.$$

shows that $\mathbf{A}$ is similar to a matrix of type II.

(3) The case of complex conjugate solutions $\lambda_1 = \mu + \mathrm{i}\nu$, $\lambda_2 = \mu - \mathrm{i}\nu$. This case arises if the discriminant $(a - d)^2 + 4bc$ is negative. The most elegant way to deal with this case is to switch to complex variables and to perform the computations in the complex vector space $\mathbb{C}^2$. We first determine complex vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{C}^2$ such that

$$\mathbf{A}\mathbf{v}_1 = \lambda_1\mathbf{v}_1, \quad \mathbf{A}\mathbf{v}_2 = \lambda_2\mathbf{v}_2$$

and then decompose $\mathbf{v}_1 = \mathbf{f} + \mathrm{i}\mathbf{g}$ into real and imaginary parts with vectors $\mathbf{f}$, $\mathbf{g}$ in $\mathbb{R}^2$. Since $\lambda_1 = \mu + \mathrm{i}\nu$, $\lambda_2 = \mu - \mathrm{i}\nu$, it follows that

$$\mathbf{v}_2 = \mathbf{f} - \mathrm{i}\mathbf{g}.$$

Note that $\{\mathbf{v}_1, \mathbf{v}_2\}$ forms a basis of $\mathbb{C}^2$. Thus $\{\mathbf{g}, \mathbf{f}\}$ is a basis of $\mathbb{R}^2$ and

$$\mathbf{A}(\mathbf{f} + \mathrm{i}\mathbf{g}) = (\mu + \mathrm{i}\nu)(\mathbf{f} + \mathrm{i}\mathbf{g}) = \mu\mathbf{f} - \nu\mathbf{g} + \mathrm{i}(\nu\mathbf{f} + \mu\mathbf{g}),$$

consequently

$$\mathbf{A}\mathbf{g} = \nu\mathbf{f} + \mu\mathbf{g}, \quad \mathbf{A}\mathbf{f} = \mu\mathbf{f} - \nu\mathbf{g}.$$

Again we set

$$\mathbf{T} = [\mathbf{g} \vdots \mathbf{f}] = \begin{bmatrix} g_1 & f_1 \\ g_2 & f_2 \end{bmatrix}$$

from which we deduce

$$\mathbf{T}^{-1}\mathbf{A}\mathbf{T} = \mathbf{T}^{-1}[\nu\mathbf{f} + \mu\mathbf{g} \,\vdots\, \mu\mathbf{f} - \nu\mathbf{g}]$$

$$= \frac{1}{g_1 f_2 - g_2 f_1}\begin{bmatrix} f_2 & -f_1 \\ -g_2 & g_1 \end{bmatrix}\begin{bmatrix} \nu f_1 + \mu g_1 & \mu f_1 - \nu g_1 \\ \nu f_2 + \mu g_2 & \mu f_2 - \nu g_2 \end{bmatrix} = \begin{bmatrix} \mu & -\nu \\ \nu & \mu \end{bmatrix}.$$

Thus **A** is similar to a matrix of type III.

Further Results on Continuity

C

This appendix covers further material on continuity which is not central for this book but on the other hand is required in various proofs (like in the chapters on curves and differential equations). It includes assertions about the continuity of the inverse function, the concept of uniform convergence of sequences of functions, the power series expansion of the exponential function and the notions of uniform and Lipschitz continuity.

C.1 Continuity of the Inverse Function

We consider a real-valued function f defined on an interval $I \subset \mathbb{R}$. The interval I can be open, half-open or closed. By $J = f(I)$ we denote the image of f. First, we show that a continuous function $f : I \to J$ is bijective, if and only if it is strictly monotonically increasing or decreasing. Monotonicity was introduced in Definition 8.5. Subsequently, we show that the inverse function is continuous if f is continuous, and we describe the respective ranges.

Proposition C.1 *A real-valued, continuous function $f : I \to J = f(I)$ is bijective if and only if it is strictly monotonically increasing or decreasing.*

Proof We already know that the function $f : I \to f(I)$ is surjective. It is injective if and only if

$$x_1 \neq x_2 \quad \Rightarrow \quad f(x_1) \neq f(x_2).$$

Strict monotonicity thus implies injectivity. To prove the converse implication we start by choosing two points $x_1 < x_2 \in I$. Let $f(x_1) < f(x_2)$, for example. We will show that f is strictly monotonically increasing on the entire interval I. First we observe that for every $x_3 \in (x_1, x_2)$ we must have $f(x_1) < f(x_3) < f(x_2)$. This

M. Oberguggenberger and A. Ostermann, *Analysis for Computer Scientists*,
Undergraduate Topics in Computer Science,
https://doi.org/10.1007/978-3-319-91155-7

is shown by contradiction. Assuming $f(x_3) > f(x_2)$, Proposition 6.14 implies that every intermediate point $f(x_2) < \eta < f(x_3)$ would be the image of a point $\xi_1 \in (x_1, x_3)$ and also the image of a point $\xi_2 \in (x_3, x_2)$, contradicting injectivity.

If we now choose $x_4 \in I$ such that $x_2 < x_4$, then once again $f(x_2) < f(x_4)$. Otherwise we would have $x_1 < x_2 < x_4$ with $f(x_2) > f(x_4)$; this possibility is excluded as in the previous case. Finally, the points to the left of x_1 are inspected in a similar way. It follows that f is strictly monotonically increasing on the entire interval I. In the case $f(x_1) > f(x_2)$, one can deduce similarly that f is monotonically decreasing. □

The function $y = x \cdot \mathbb{1}_{(-1,0]}(x) + (1-x) \cdot \mathbb{1}_{(0,1)}(x)$, where $\mathbb{1}_I$ denotes the indicator function of the interval I (see Sect. 2.2), shows that a discontinuous function can be bijective on an interval without being strictly monotonically increasing or decreasing.

Remark C.2 If I is an open interval and $f : I \to J$ a continuous and bijective function, then J is an open interval as well. Indeed, if J were of the form $[a, b)$, then a would arise as function value of a point $x_1 \in I$, i.e. $a = f(x_1)$. However, since I is open, there are points $x_2 \in I$, $x_2 < x_1$ and $x_3 \in I$ with $x_3 > x_1$. If f is strictly monotonically increasing then we would have $f(x_2) < f(x_1) = a$. If f is strictly monotonically decreasing then $f(x_3) < f(x_1) = a$. Both cases contradict the fact that a was assumed to be the lower boundary of the image $J = f(I)$. In the same way, one excludes the possibilities that $J = (a, b]$ or $J = [a, b]$.

Proposition C.3 *Let $I \subset \mathbb{R}$ be an open interval and $f : I \to J$ continuous and bijective. Then the inverse function $f^{-1} : J \to I$ is continuous as well.*

Proof We take $x \in I$, $y \in J$ with $y = f(x)$, $x = f^{-1}(y)$. For small $\varepsilon > 0$ the ε-neighbourhood $U_\varepsilon(x)$ of x is contained in I. According to Remark C.2 $f(U_\varepsilon(x))$ is an open interval and therefore contains a δ-neighbourhood $U_\delta(y)$ of y for a certain $\delta > 0$. Consider a sequence of values $y_n \in J$ which converges to y as $n \to \infty$. Then there is an index $n(\delta) \in \mathbb{N}$ such that all elements of the sequence y_n with $n \geq n(\delta)$ lie in the δ-neighbourhood $U_\delta(y)$. That, however, means that the values of the function $f^{-1}(y_n)$ from $n(\delta)$ onwards lie in the ε-neighbourhood $U_\varepsilon(x)$ of $x = f^{-1}(y)$. Thus $\lim_{n\to\infty} f^{-1}(y_n) = f^{-1}(y)$ which is the continuity of f^{-1} at y. □

C.2 Limits of Sequences of Functions

We consider a sequence of functions $f_n : I \to \mathbb{R}$, defined on an interval $I \subset \mathbb{R}$. If the function values $f_n(x)$ converge for every fixed $x \in I$, then the sequence $(f_n)_{n\geq 1}$ is called *pointwise convergent*. The pointwise limits define a function $f : I \to \mathbb{R}$ by $f(x) = \lim_{n\to\infty} f_n(x)$, the so-called *limit function*.

Example C.4 Let $I = [0, 1]$ and $f_n(x) = x^n$. Then $\lim_{n\to\infty} f_n(x) = 0$ if $0 \le x < 1$, and $\lim_{n\to\infty} f_n(1) = 1$. The limit function is thus the function

$$f(x) = \begin{cases} 0, & 0 \le x < 1, \\ 1, & x = 1. \end{cases}$$

This example shows that the limit function of a pointwise convergent sequence of continuous functions is not necessarily continuous.

Definition C.5 (*Uniform convergence of sequences of functions*) A sequence of functions $(f_n)_{n\ge 1}$ defined on an interval I is called *uniformly convergent* with *limit function* f, if

$$\forall \varepsilon > 0 \;\; \exists n(\varepsilon) \in \mathbb{N} \;\; \forall n \ge n(\varepsilon) \;\; \forall x \in I \; : \; |f(x) - f_n(x)| < \varepsilon.$$

Uniform convergence means that the index $n(\varepsilon)$ after which the sequence of function values $(f_n(x))_{n\ge 1}$ settles in the ε-neighbourhood $U_\varepsilon(f(x))$ can be chosen independently of $x \in I$.

Proposition C.6 *The limit function f of a uniformly convergent sequence of functions $(f_n)_{n\ge 1}$ is continuous.*

Proof We take $x \in I$ and a sequence of points x_k converging to x as $k \to \infty$. We have to show that $f(x) = \lim_{k\to\infty} f(x_k)$. For this we write

$$f(x) - f(x_k) = \big(f(x) - f_n(x)\big) + \big(f_n(x) - f_n(x_k)\big) + \big(f_n(x_k) - f(x_k)\big)$$

and choose $\varepsilon > 0$. Due to the uniform convergence it is possible to find an index $n \in \mathbb{N}$ such that

$$|f(x) - f_n(x)| < \frac{\varepsilon}{3} \quad \text{and} \quad |f_n(x_k) - f(x_k)| < \frac{\varepsilon}{3}$$

for all $k \in \mathbb{N}$. Since f_n is continuous, there is an index $k(\varepsilon) \in \mathbb{N}$ such that

$$|f_n(x) - f_n(x_k)| < \frac{\varepsilon}{3}$$

for all $k \ge k(\varepsilon)$. For such indices k we have

$$|f(x) - f(x_k)| < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.$$

Thus $f(x_k) \to f(x)$ as $k \to \infty$, which implies the continuity of f. □

Application C.7 The exponential function $f(x) = a^x$ is continuous on $\mathbb{R}$. In Application 5.14 it was shown that the exponential function with base $a > 0$ can be defined for every $x \in \mathbb{R}$ as a limit. Let $r_n(x)$ denote the decimal representation of x, truncated at the nth decimal place. Then

$$r_n(x) \le x < r_n(x) + 10^{-n}.$$

The value of $r_n(x)$ is the same for all real numbers x, which coincide up to the nth decimal place. Thus the mapping $x \mapsto r_n(x)$ is a step function with jumps at a distance of 10^{-n}. We define the function $f_n(x)$ by linear interpolation between the points

$$\big(r_n(x), a^{r_n(x)}\big) \quad \text{and} \quad \big(r_n(x) + 10^{-n}, a^{r_n(x)+10^{-n}}\big),$$

which means

$$f_n(x) = a^{r_n(x)} + \frac{x - r_n(x)}{10^{-n}} \left(a^{r_n(x)+10^{-n}} - a^{r_n(x)}\right).$$

The graph of the function $f_n(x)$ is a polygonal chain (with kinks at the distance of 10^{-n}), and thus f_n is continuous. We show that the sequence of functions $(f_n)_{n\ge1}$ converges uniformly to f on every interval $[-T, T]$, $0 < T \in \mathbb{Q}$. Since $x - r_n(x) \le 10^{-n}$, it follows that

$$|f(x) - f_n(x)| \le \left|a^x - a^{r_n(x)}\right| + \left|a^{r_n(x)+10^{-n}} - a^{r_n(x)}\right|.$$

For $x \in [-T, T]$ we have

$$a^x - a^{r_n(x)} = a^{r_n(x)}\big(a^{x-r_n(x)} - 1\big) \le a^T \big(a^{10^{-n}} - 1\big)$$

and likewise

$$a^{r_n(x)+10^{-n}} - a^{r_n(x)} \le a^T \big(a^{10^{-n}} - 1\big).$$

Consequently

$$|f(x) - f_n(x)| \le 2a^T \big(\sqrt[10^n]{a} - 1\big),$$

and the term on the right-hand side converges to zero independently of x, as was proven in Application 5.15.

The rules of calculation for real exponents can now also be derived by taking limits. Take, for example, $r, s \in \mathbb{R}$ with decimal approximations $(r_n)_{n\ge1}$, $(s_n)_{n\ge1}$. Then Proposition 5.7 and the continuity of the exponential function imply

$$a^r a^s = \lim_{n\to\infty} \big(a^{r_n} a^{s_n}\big) = \lim_{n\to\infty} \big(a^{r_n+s_n}\big) = a^{r+s}.$$

With the help of Proposition C.3 the continuity of the logarithm follows as well.

C.3 The Exponential Series

The aim of this section is to derive the series representation of the exponential function

$$\mathrm{e}^x = \sum_{m=0}^{\infty} \frac{x^m}{m!}$$

by using exclusively the theory of convergent series without resorting to differential calculus. This is important for our exposition because the differentiability of the exponential function is proven with the help of the series representation in Sect. 7.2.

As a tool we need two supplements to the theory of series: Absolute convergence and Cauchy's[2] formula for the product of two series.

Definition C.8 A series $\sum_{k=0}^{\infty} a_k$ is called *absolutely convergent*, if the series $\sum_{k=0}^{\infty} |a_k|$ of the absolute values of its coefficients converges.

Proposition C.9 *Every absolutely convergent series is convergent.*

Proof We define the positive and the negative parts of the coefficient a_k by

$$a_k^+ = \begin{cases} a_k, & a_k \ge 0, \\ 0, & a_k < 0, \end{cases} \qquad a_k^- = \begin{cases} 0, & a_k \ge 0, \\ |a_k|, & a_k < 0. \end{cases}$$

Obviously, we have $0 \le a_k^+ \le |a_k|$ and $0 \le a_k^- \le |a_k|$. Thus the two series $\sum_{k=0}^{\infty} a_k^+$ and $\sum_{k=0}^{\infty} a_k^-$ converge due to the comparison criterion (Proposition 5.21) and the limit

$$\lim_{n\to\infty} \sum_{k=0}^{n} a_k = \lim_{n\to\infty} \sum_{k=0}^{n} a_k^+ - \lim_{n\to\infty} \sum_{k=0}^{n} a_k^-$$

exists. Consequently, the series $\sum_{k=0}^{\infty} a_k$ converges. □

We consider two absolutely convergent series $\sum_{i=0}^{\infty} a_i$ and $\sum_{j=0}^{\infty} b_j$ and ask how their product can be computed. Term-by-term multiplication of the nth partial sums suggests to consider the following scheme:

$$\begin{array}{ccccc}
\boldsymbol{a_0 b_0} & \boldsymbol{a_0 b_1} & \dots & \boldsymbol{a_0 b_{n-1}} & \boldsymbol{a_0 b_n} \\
\boldsymbol{a_1 b_0} & \boldsymbol{a_1 b_1} & \dots & \boldsymbol{a_1 b_{n-1}} & a_1 b_n \\
\vdots & & \iddots & & \vdots \\
\boldsymbol{a_{n-1} b_0} & \boldsymbol{a_{n-1} b_1} & \dots & a_{n-1} b_{n-1} & a_{n-1} b_n \\
\boldsymbol{a_n b_0} & a_n b_1 & \dots & a_n b_{n-1} & a_n b_n
\end{array}$$

[2] A.L. Cauchy, 1789–1857.

Adding all entries of the quadratic scheme one obtains the product of the partial sums

$$P_n = \sum_{i=0}^{n} a_i \sum_{j=0}^{n} b_j .$$

In contrast, adding only the upper triangle containing the bold entries (diagonal by diagonal), one obtains the so-called *Cauchy product formula*

$$S_n = \sum_{m=0}^{n} \left(\sum_{k=0}^{m} a_k b_{m-k} \right) .$$

We want to show that, for absolutely convergent series, the limits are equal:

$$\lim_{n\to\infty} P_n = \lim_{n\to\infty} S_n .$$

Proposition C.10 (Cauchy product) *If the series $\sum_{i=0}^{\infty} a_i$ and $\sum_{j=0}^{\infty} b_j$ converge absolutely then*

$$\sum_{i=0}^{\infty} a_i \sum_{j=0}^{\infty} b_j = \sum_{m=0}^{\infty} \left(\sum_{k=0}^{m} a_k b_{m-k} \right) .$$

The series defined by the Cauchy product formula also converges absolutely.

Proof We set

$$c_m = \sum_{k=0}^{m} a_k b_{m-k}$$

and obtain that the partial sums

$$T_n = \sum_{m=0}^{n} |c_m| \le \sum_{i=0}^{n} |a_i| \sum_{j=0}^{n} |b_j| \le \sum_{i=0}^{\infty} |a_i| \sum_{j=0}^{\infty} |b_j|$$

remain bounded. This follows from the facts that the triangle in the scheme above has fewer entries than the square and the original series converge absolutely. Obviously the sequence T_n is also monotonically increasing; according to Proposition 5.10 it thus has a limit. This means that the series $\sum_{m=0}^{\infty} c_m$ converges absolutely, so the Cauchy product exists. It remains to be shown that it coincides with the product of the series. For the partial sums, we have

$$\Big| P_n - S_n \Big| = \left| \sum_{i=0}^{n} a_i \sum_{j=0}^{n} b_j - \sum_{m=0}^{n} c_m \right| \le \left| \sum_{m=n+1}^{\infty} c_m \right| ,$$

since the difference can obviously be approximated by the sum of the terms below the nth diagonal. The latter sum, however, is just the difference of the partial sum

S_n and the value of the series $\sum_{m=0}^{\infty} c_m$. It thus converges to zero and the desired assertion is proven. □

Let

$$E(x) = \sum_{m=0}^{\infty} \frac{x^m}{m!}, \qquad E_n(x) = \sum_{m=0}^{n} \frac{x^m}{m!}.$$

The convergence of the series for $x = 1$ was shown in Example 5.24 and for $x = 2$ in Exercise 14 of Chap. 5. The absolute convergence for arbitrary $x \in \mathbb{R}$ can either be shown analogously or by using the ratio test (Exercise 15 in Chap. 5). If x varies in a bounded interval $I = [-R, R]$, then the sequence of the partial sums $E_n(x)$ converges uniformly to $E(x)$, due to the uniform estimate

$$\left|E(x) - E_n(x)\right| = \left|\sum_{m=n+1}^{\infty} \frac{x^m}{m!}\right| \le \sum_{m=n+1}^{\infty} \frac{R^m}{m!} \to 0$$

on the interval $[-R, R]$. Proposition C.6 implies that the function $x \mapsto E(x)$ is continuous.

For the derivation of the product formula $E(x)E(y) = E(x + y)$ we recall the *binomial formula*:

$$(x + y)^m = \sum_{k=0}^{m} \binom{m}{k} x^k y^{m-k} \qquad \text{with} \qquad \binom{m}{k} = \frac{m!}{k!(m-k)!},$$

valid for arbitrary $x, y \in \mathbb{R}$ and $n \in \mathbb{N}$, see, for instance, [17, Chap. XIII, Theorem 7.2].

Proposition C.11 *For arbitrary $x, y \in \mathbb{R}$ it holds that*

$$\sum_{i=0}^{\infty} \frac{x^i}{i!} \sum_{j=0}^{\infty} \frac{y^j}{j!} = \sum_{m=0}^{\infty} \frac{(x+y)^m}{m!}.$$

Proof Due to the absolute convergence of the above series, Proposition C.10 yields

$$\sum_{i=0}^{\infty} \frac{x^i}{i!} \sum_{j=0}^{\infty} \frac{y^j}{j!} = \sum_{m=0}^{\infty} \sum_{k=0}^{m} \frac{x^k}{k!} \frac{y^{m-k}}{(m-k)!}.$$

An application of the binomial formula

$$\sum_{k=0}^{m} \frac{x^k}{k!} \frac{y^{m-k}}{(m-k)!} = \frac{1}{m!} \sum_{k=0}^{m} \binom{m}{k} x^k y^{m-k} = \frac{1}{m!}(x+y)^m$$

shows the desired assertion. □

Proposition C.12 (Series representation of the exponential function) *The exponential function possesses the series representation*

$$e^x = \sum_{m=0}^{\infty} \frac{x^m}{m!},$$

valid for arbitrary $x \in \mathbb{R}$.

Proof By definition of the number e (see Example 5.24) we obviously have

$$e^0 = 1 = E(0), \quad e^1 = e = E(1).$$

From Proposition C.11 we get in particular

$$e^2 = e^{1+1} = e^1 e^1 = E(1)E(1) = E(1+1) = E(2)$$

and recursively

$$e^m = E(m) \quad \text{for} \quad m \in \mathbb{N}.$$

The relation $E(m)E(-m) = E(m-m) = E(0) = 1$ shows that

$$e^{-m} = \frac{1}{e^m} = \frac{1}{E(m)} = E(-m).$$

Likewise, one concludes from $\big(E(1/n)\big)^n = E(1)$ that

$$e^{1/n} = \sqrt[n]{e} = \sqrt[n]{E(1)} = E(1/n).$$

So far this shows that $e^x = E(x)$ holds for all rational $x = m/n$. From Application C.7 we know that the exponential function $x \mapsto e^x$ is continuous. The continuity of the function $x \mapsto E(x)$ was shown above. But two continuous functions which coincide for all rational numbers are equal. More precisely, if $x \in \mathbb{R}$ and x_j is the decimal expansion of x truncated at the jth place, then

$$e^x = \lim_{j \to \infty} e^{x_j} = \lim_{j \to \infty} E(x_j) = E(x),$$

which is the desired result. □

Remark C.13 The rigorous introduction of the exponential function is surprisingly involved and is handled differently by different authors. The total effort, however, is approximately the same in all approaches. We took the following route: Introduction of Euler's number e as the value of a convergent series (Example 5.24); definition of the exponential function $x \mapsto e^x$ for $x \in \mathbb{R}$ by using the completeness of the real numbers (Application 5.14); continuity of the exponential function based on uniform convergence (Application C.7); series representation (Proposition C.12); differentiability and calculation of the derivative (Sect. 7.2). Finally, in the course of the computation of the derivative we also obtained the well-known formula $\mathrm{e} = \lim_{n\to\infty}(1 + 1/n)^n$, which Euler himself used to define the number e.

C.4 Lipschitz Continuity and Uniform Continuity

Some results on curves and differential equations require more refined continuity properties. More precisely, methods for quantifying how the function values change in dependence on the arguments are needed.

Definition C.14 A function $f : D \subset \mathbb{R} \to \mathbb{R}$ is called *Lipschitz continuous*, if there exists a constant $L > 0$ such that the inequality

$$|f(x_1) - f(x_2)| \le L|x_1 - x_2|$$

holds for all $x_1, x_2 \in D$. In this case L is called a *Lipschitz constant* of the function f.

If $x \in D$ and $(x_n)_{n\ge 1}$ is a sequence of points in D which converges to x, the inequality $|f(x) - f(x_n)| \le L|x - x_n|$ implies that $f(x_n) \to f(x)$ as $n \to \infty$. Every Lipschitz continuous function is thus continuous. For Lipschitz continuous functions one can quantify how much change in the x-values can be allowed to obtain a change in the function values of $\varepsilon > 0$ at the most:

$$|x_1 - x_2| < \varepsilon/L \quad \Rightarrow \quad |f(x_1) - f(x_2)| < \varepsilon.$$

Occasionally the following weaker quantification is required.

Definition C.15 A function $f : D \subset \mathbb{R} \to \mathbb{R}$ is called *uniformly continuous*, if there exists a mapping $\omega : (0, 1] \to (0, 1] : \varepsilon \mapsto \omega(\varepsilon)$ such that

$$|x_1 - x_2| < \omega(\varepsilon) \quad \Rightarrow \quad |f(x_1) - f(x_2)| < \varepsilon$$

for all $x_1, x_2 \in D$. In this case the mapping ω is called a *modulus of continuity* of the function f.

Every Lipschitz continuous function is uniformly continuous (with $\omega(\varepsilon) = \varepsilon/L$), and every uniformly continuous function is continuous.

Example C.16 (a) The quadratic function $f(x) = x^2$ is Lipschitz continuous on every bounded interval $[a, b]$. For $x_1 \in [a, b]$ we have $|x_1| \le M = \max(|a|, |b|)$ and likewise for x_2. Thus

$$|f(x_1) - f(x_2)| = |x_1^2 - x_2^2| = |x_1 + x_2||x_1 - x_2| \le 2M|x_1 - x_2|$$

holds for all $x_1, x_2 \in [a, b]$.
(b) The absolute value function $f(x) = |x|$ is Lipschitz continuous on $D = \mathbb{R}$ (with Lipschitz constant $L = 1$). This follows from the inequality

$$\big||x_1| - |x_2|\big| \le |x_1 - x_2|,$$

which is valid for all $x_1, x_2 \in \mathbb{R}$.
(c) The square root function $f(x) = \sqrt{x}$ is uniformly continuous on the interval $[0, 1]$, but not Lipschitz continuous. This follows from the inequality

$$\big|\sqrt{x_1} - \sqrt{x_2}\big| \le \sqrt{|x_1 - x_2|},$$

which is proved immediately by squaring. Thus $\omega(\varepsilon) = \varepsilon^2$ is a modulus of continuity of the square root function on the interval $[0, 1]$. The square root function is not Lipschitz continuous on $[0, 1]$, since otherwise the choice $x_2 = 0$ would imply the relations

$$\sqrt{x_1} \le L|x_1|, \qquad \frac{1}{\sqrt{x_1}} \le L$$

which cannot hold for fixed $L > 0$ and all $x_1 \in (0, 1]$.
(d) The function $f(x) = \frac{1}{x}$ is continuous on the interval $(0, 1)$, but not uniformly continuous. Assume that we could find a modulus of continuity $\varepsilon \mapsto \omega(\varepsilon)$ on $(0, 1)$. Then for $x_1 = 2\varepsilon\omega(\varepsilon)$, $x_2 = \varepsilon\omega(\varepsilon)$ and $\varepsilon < 1$ we would get $|x_1 - x_2| < \omega(\varepsilon)$, but

$$\left|\frac{1}{x_1} - \frac{1}{x_2}\right| = \left|\frac{x_2 - x_1}{x_1 x_2}\right| = \frac{\varepsilon\omega(\varepsilon)}{2\varepsilon^2\omega(\varepsilon)^2} = \frac{1}{2\varepsilon\omega(\varepsilon)}$$

which becomes arbitrarily large as $\varepsilon \to 0$. In particular, it cannot be bounded from above by ε.

From the mean value theorem (Proposition 8.4) it follows that differentiable functions with bounded derivative are Lipschitz continuous. Further it can be shown that every function which is continuous on a closed, bounded interval $[a, b]$ is uniformly continuous there. The proof requires further tools from analysis for which we refer to [4, Theorem 3.13].

Apart from the intermediate value theorem, the *fixed point theorem* is an important tool for proving the existence of solutions of equations. Moreover one obtains an iterative algorithm for approximating the fixed point.

Definition C.17 A Lipschitz continuous mapping f of an interval I to $\mathbb{R}$ is called a *contraction*, if $f(I) \subset I$ and f has a Lipschitz constant $L < 1$. A point $x^* \in I$ with $x^* = f(x^*)$ is called *fixed point* of the function f.

Proposition C.18 (Fixed point theorem) *A contraction f on a closed interval $[a, b]$ has a unique fixed point. The sequence, recursively defined by the iteration*

$$x_{n+1} = f(x_n)$$

converges to the fixed point x^* for arbitrary initial values $x_1 \in [a, b]$.

Proof Since $f([a, b]) \subset [a, b]$ we must have

$$a \le f(a) \quad \text{and} \quad f(b) \le b.$$

If $a = f(a)$ or $b = f(b)$, we are done. Otherwise the intermediate value theorem applied to the function $g(x) = x - f(x)$ yields the existence of a point $x^* \in (a, b)$ with $g(x^*) = 0$. This x^* is a fixed point of f. Due to the contraction property the existence of a further fixed point y^* would result in

$$|x^* - y^*| = |f(x^*) - f(y^*)| \le L|x^* - y^*| < |x^* - y^*|$$

which is impossible for $x^* \ne y^*$. Thus the fixed point is unique.

The convergence of the iteration follows from the inequalities

$$|x^* - x_{n+1}| = |f(x^*) - f(x_n)| \le L|x^* - x_n| \le \ldots \le L^n|x^* - x_1|,$$

since $|x^* - x_1| \le b - a$ and $\lim_{n\to\infty} L^n = 0$. □

Description of the Supplementary Software

D

In our view *using and writing* software forms an essential component of an analysis course for computer scientists. The software that has been developed for this book is available on the website

https://www.springer.com/book/9783319911540

This site contains the Java applets referred to in the text as well as some source files in maple, Python and MATLAB.

For the execution of the maple and MATLAB programs additional licences are needed.

Java applets. The available applets are listed in Table D.1. The applets are executable and only require a current version of Java installed.

Source codes in MATLAB and maple. In addition to the Java applets, you can find maple and MATLAB programs on this website. These programs are numbered according to the individual chapters and are mainly used in experiments and exercises. To run the programs the corresponding software licence is required.

Source codes in Python. For each MATLAB program, an equivalent Python program is provided. To run these programs, a current version of Python has to be installed. We do not specifically refer these programs in the text; the numbering is the same as for the M-files.

M. Oberguggenberger and A. Ostermann, *Analysis for Computer Scientists*,
Undergraduate Topics in Computer Science,
https://doi.org/10.1007/978-3-319-91155-7

Table D.1 List of available Java applets

Sequences
2D-visualisation of complex functions
3D-visualisation of complex functions
Bisection method
Animation of the intermediate value theorem
Newton's method
Riemann sums
Integration
Parametric curves in the plane
Parametric curves in space
Surfaces in space
Dynamical systems in the plane
Dynamical systems in space
Linear regression

References

Textbooks

1. E. Hairer, G. Wanner, *Analysis by Its History* (Springer, New York, 1996)
2. S. Lang, *Introduction to Linear Algebra*, 2nd edn. (Springer, New York, 1986)
3. S. Lang, *Undergraduate Analysis* (Springer, New York, 1983)
4. M.H. Protter, C.B. Morrey, *A First Course in Real Analysis*, 2nd edn. (Springer, New York, 1991)

Further Reading

5. M. Barnsley, *Fractals Everywhere* (Academic Press, Boston, 1988)
6. M. Braun, C.C. Coleman, D.A. Drew (eds.), *Differential Equation Models* (Springer, Berlin, 1983)
7. M. Bronstein, *Symbolic Integration I: Transcendental Functions* (Springer, Berlin, 1997)
8. A. Chevan, M. Sutherland, Hierarchical partitioning. Am. Stat. **45**, 90–96 (1991)
9. J.P. Eckmann, Savez-vous résoudre $z^3 = 1$? La Recherche **14**, 260–262 (1983)
10. N. Fickel, Partition of the coefficient of determination in multiple regression, in *Operations Research Proceedings 1999*, ed. by K. Inderfurth (Springer, Berlin, 2000), pp. 154–159
11. E. Hairer, S.P. Nørsett, G. Wanner, *Solving Ordinary Differential Equations I. Nonstiff Problems*, 2nd edn. (Springer, Berlin, 1993)
12. E. Hairer, G. Wanner, *Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems*, 2nd edn. (Springer, Berlin, 1996)
13. M.W. Hirsch, S. Smale, *Differential Equations, Dynamical Systems, and Linear Algebra* (Academic Press, New York, 1974)
14. R.F. Keeling, S.C. Piper, A.F. Bollenbacher, J.S. Walker, Atmospheric CO_2 records from sites in the SIO air sampling network, in *Trends: A Compendium of Data on Global Change. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy* (Oak Ridge, Tennessy, USA, 2009). https://doi.org/10.3334/CDIAC/atg.035

M. Oberguggenberger and A. Ostermann, *Analysis for Computer Scientists*, Undergraduate Topics in Computer Science,
https://doi.org/10.1007/978-3-319-91155-7

15. E. Kreyszig, *Statistische Methoden und ihre Anwendungen*, 3rd edn. (Vandenhoeck & Ruprecht, Göttingen, 1968)
16. W. Kruskal, Relative importance by averaging over orderings. Am. Stat. **41**, 6–10 (1987)
17. S. Lang, *A First Course in Calculus*, 5th edn. (Springer, New York, 1986)
18. M. Lefebvre, *Basic Probability Theory with Applications* (Springer, New-York, 2009)
19. D.C. Montgomery, E.A. Peck, G.G. Vining, *Introduction to Linear Regression Analysis*, 3rd edn. (Wiley, New York, 2001)
20. M.L. Overton, *Numerical Computing with IEEE Floating Point Arithmetic* (SIAM, Philadelphia, 2001)
21. H.-O. Peitgen, H. Jürgens, D. Saupe, *Fractals for the Classroom. Part One: Introduction to Fractals and Chaos* (Springer, New York, 1992)
22. H.-O. Peitgen, H. Jürgens, D. Saupe, *Fractals for the Classroom. Part Two: Complex Systems and Mandelbrot Set* (Springer, New York, 1992)
23. A. Quarteroni, R. Sacco, F. Saleri, *Numerical Mathematics* (Springer, New York, 2000)
24. H. Rommelfanger, *Differenzen- und Differentialgleichungen* (Bibliographisches Institut, Mannheim, 1977)
25. B. Schuppener, Die Festlegung charakteristischer Bodenkennwerte – Empfehlungen des Eurocodes 7 Teil 1 und die Ergebnisse einer Umfrage. Geotechnik Sonderheft (1999), pp. 32–35
26. STATISTIK AUSTRIA, Statistisches Jahrbuch Österreich. Verlag Österreich GmbH, Wien 2018. http://www.statistik.at
27. M.A. Väth, *Nonstandard Analysis* (Birkhäuser, Basel, 2007)

Index

M. Oberguggenberger and A. Ostermann, *Analysis for Computer Scientists*,
Undergraduate Topics in Computer Science,
https://doi.org/10.1007/978-3-319-91155-7