A Compact Course on Mathematical Methods

Pascal Philipp

November 26, 2020

The notes at hand cover the following topics:

1. Vectors and Matrices

2. Functions of Several Variables

3. Integration

4. Differential Equations

The main prerequiste that is required for being able to work through the text is familiarity with functions of a single variable and differentiation. Chapter 2 is then the natural follow-up to that prerequisite. Differentiation for functions of several variables and applications such as finding minima or maxima are covered. Chapter 3 introduces integration; first for functions of one variable and then for functions of several variables. That second part, multivariate integration, is rather brief and more of a quick taster and introduction of concepts rather than a systematic treatment. Chapter 4 then does the same for differential equations. You could call these three chapters 'advanced calculus'.

The material in chapter 1, vectors and matrices, isn't usually taught together with calculus – it is covered here because it is the other important basic mathematical methods topic needed in STEM (with calculus being the first; actually, there is a third such topic: probability and statistics, which is not covered, but I'd be open to starting collaborations to add a compact chapter on it). Chapter 1 isn't a prerequisite for the other chapters, but it makes some of the multivariate notation prettier.

These notes are application-oriented and the philosophy is to cover the basic concepts quickly, go through a good amount of examples and exercises, and then move on. No need to go down rabbit holes. That's where the 'compact' in the title comes from. Despite focus on methods for applied computations and solving problems, a few more theoretical remarks and exercises can be found in the text. Elements of abstract mathematical rigour can be found as well – I tried to do give a taste of this way of thinking in a friendly way.

This document belongs to the GitHub repository

$$\texttt{https://github.com/pasc85/MathematicalMethods}$$

You can find the license there as well as information on how to contribute.

Contents

List of Applications

```
Customise the format:  drop ''Application:  ''.
```

Chapter 1

Vectors and Matrices

Matrices (singular: *matrix*) are arrays of numbers, for example,

$$M = \begin{bmatrix} 4 & -\sqrt{3} & \pi \\ -5.2 & 0 & 13 \end{bmatrix},$$

and *vectors* are matrices that have only one column. The use of vectors and matrices makes the notation and handling of data and variables in large computations clearer and more compact, and their study has also led to new concepts and theories. Vectors and matrices are fundamental in mathematics and for applications of mathematical analysis.

In the example below, we will work with a matrix that represents the connections in a network of cities, and this matrix can therefore be considered the link that makes the road network accessible to mathematical analysis. Other networks that are frequently analysed using matrices include social networks (e.g., to study the spread of news), the world wide web (e.g., Google PageRank), contact networks (e.g., to minimise contagion in a hospital).

Application (Random walks)**.** A very large number of hikers is travelling randomly around the network of cities
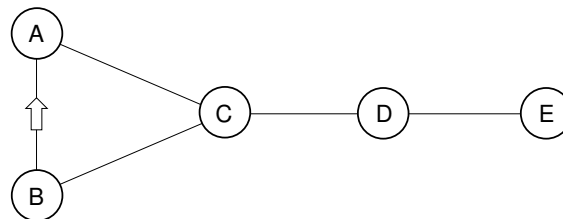
where the road between A and B can only be travelled in one direction. A hiker would arrive in a city, stay for the day, randomly pick one of the outgoing roads – each with equal probability; possibly the city he or she came from the previous day – and then travel there the next morning. For example, if there are 100 hikers in D today, then an average of 50 of them will hike to C the next day. Besides those new arrivals from D, the city C will further receive new hikers from A and from B.

The hikers in this example are called *random walkers* in mathematical jargon, and an important and applicable task is to find the *steady state* of the system –

that is, the distribution of hikers so that the total number of hikers in each city does not change from one day to the next. For example, looking at the above map, you might expect that after a long time there should be a larger concentration of hikers in C than in E.

The steady state can be found by balancing the number of incoming and outgoing hikers for each city. Let d and e be the steady-state proportion of hikers in cities D and E, and consider all travelling to and from E: outgoing $= e$, incoming $= \frac{1}{2}d$, which gives $e = \frac{1}{2}d$. Repeating this for the other cities leads to a collection of five equations.

Alternatively, one can choose a matrix approach and use vectors to describe the distribution of hikers over the network. The first entry of that vector stands for the proportion of hikers in A, the second for the proportion in B, etc. For example,

$$v_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \end{bmatrix}$$

means in the first case that all hikers are in C, and in the second case that they are evenly distributed over all five cities. Next we describe their movements using a matrix. After the review of matrix multiplication in the next section, you will be able to convince yourself that

$$v_{\text{tomorrow}} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 & 0 \\ 1 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} & 0 & 1 \\ 0 & 0 & 0 & \frac{1}{2} & 0 \end{bmatrix} v_{\text{today}} = P\,v_{\text{today}}$$

reflects the movement of the hikers around the network of cities[1]. The task of finding a steady state now corresponds to finding a vector (i.e., a distribution of hikers) that does not change through application of P (i.e., from one day to the next). Note that individual hikers keep moving – the steady state is the distribution of hikers such that their *total* number in each city stays the same. Denoting this vector by v^* and its entries a, b, c, d, e, we obtain $v^* = Pv^*$, or

$$\begin{bmatrix} a \\ b \\ c \\ d \\ e \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 & 0 \\ 1 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} & 0 & 1 \\ 0 & 0 & 0 & \frac{1}{2} & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \\ e \end{bmatrix}.$$

This is an equation of 5-vectors (the result of the matrix multiplication on the right-hand side is a 5-vector as well) and therefore corresponds to a collection of five ordinary equations – can you locate the equation $e = \frac{1}{2}d$ from the previous paragraph in it?

The equation $v^* = Pv^*$ is in fact an eigenvalue equation for the matrix P. We will learn how to solve it in this chapter. The solution is

$$a = 0.177, \;\; b = 0.118, \;\; c = 0.353, \;\; d = 0.235, \;\; e = 0.118\,.$$

1.1 Review of Matrix Arithmetic

Definition 1.1 (Matrices).

(i) A $m \times n$ matrix is an array

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix},$$

where m is the number of rows and n the number of columns. We also refer to A as a matrix of *size* $m \times n$.

(ii) An equation of the form $A = B$, where A and B are matrices of the same size $m \times n$, means that $a_{ij} = b_{ij}$ for all $i \in \{1, 2, \ldots, m\}, j \in \{1, 2, \ldots, n\}$. Matrices of different sizes can never be equal.

(iii) Addition and subtraction can only be carried out for matrices of the same size, and then the operation is carried out elementwise:

$$A \pm B = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \pm \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} \pm b_{11} & a_{12} \pm b_{12} & \cdots & a_{1n} \pm b_{1n} \\ a_{21} \pm b_{21} & a_{22} \pm b_{22} & \cdots & a_{2n} \pm b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} \pm b_{m1} & a_{m2} \pm b_{m2} & \cdots & a_{mn} \pm b_{mn} \end{bmatrix}.$$

(iv) Multiplying a number with a matrix is called *scalar multiplication*:

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

(v) Multiplying two matrices A and B is called *matrix multiplication* – it can be carried out only if the number of columns of A agrees with the number of rows of B:

$$AB = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1k} \\ b_{21} & b_{22} & \cdots & b_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nk} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1k} \\ c_{21} & c_{22} & \cdots & c_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mk} \end{bmatrix},$$

where $c_{ij} = \sum_{s=1}^{n} a_{is} b_{sj}$. The resulting matrix is of size $m \times k$.

(vi) A *square matrix* is a matrix with $m = n$. The *identity matrix*, which has entries 1 on the diagonal and is zero everywhere else,

$$I = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix},$$

is an important example of a square matrix.

Remark 1.2. (i) Regarding the condition on when two matrices can be multiplied, it should be useful to remember that "$m \times n \cdot n \times k$ works and gives a $m \times k$ matrix." Therefore, the result of a $1 \times n$ with a $n \times 1$ is a 1×1, which is just a single real number, also called a *scalar*.

(ii) The rule for matrix multiplication may seem quite complicated – it becomes easier to remember once one breaks down the procedure as follows. The case $1 \times n \cdot n \times 1$ just mentioned is the building block for matrix products:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \end{bmatrix} \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \\ \vdots \\ b_{n1} \end{bmatrix} = a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} + \cdots + a_{1n}b_{n1}$$

$$= \sum_{s=1}^{n} a_{1s}b_{s1} = c_{11} .$$

Here, both matrices have the same number of elements, and those entries are multiplied pairwise and then added up. For larger matrices, one just repeats this procedure $m \cdot k$ times:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1k} \\ b_{21} & b_{22} & \cdots & b_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nk} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1k} \\ c_{21} & c_{22} & \cdots & c_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mk} \end{bmatrix} .$$

(iii) Another important definition for matrices is the *transpose* of a matrix, which leaves the entries $a_{11}, a_{22}, a_{33}, \ldots$ alone and swaps all other entries across the diagonal. This operation is denoted with a "$^{\top}$" and changes the size of the matrix unless it is square, cf. the examples below.

Example 1.3. (i) The 2×2 identity matrix is $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$

(ii) The matrix

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

is of size 3×2 and its transpose is the 2×3 matrix

$$A^\top = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}.$$

(iii) For the matrix A from (ii), we have

$$A + A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} = \begin{bmatrix} 1+1 & 2+2 \\ 3+3 & 4+4 \\ 5+5 & 6+6 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 6 & 8 \\ 10 & 12 \end{bmatrix},$$

which agrees with

$$2\,A = 2 \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} = \begin{bmatrix} 2\cdot 1 & 2\cdot 2 \\ 2\cdot 3 & 2\cdot 4 \\ 2\cdot 5 & 2\cdot 6 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 6 & 8 \\ 10 & 12 \end{bmatrix}.$$

(iv) For A and its transpose $A^\top$, we obtain the products

$$A\,A^\top = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$$

$$= \begin{bmatrix} 1\cdot 1+2\cdot 2 & 1\cdot 3+2\cdot 4 & 1\cdot 5+2\cdot 6 \\ 3\cdot 1+4\cdot 2 & 3\cdot 3+4\cdot 4 & 3\cdot 5+4\cdot 6 \\ 5\cdot 1+6\cdot 2 & 5\cdot 3+6\cdot 4 & 5\cdot 5+6\cdot 6 \end{bmatrix} = \begin{bmatrix} 5 & 11 & 17 \\ 11 & 25 & 39 \\ 17 & 39 & 61 \end{bmatrix}$$

and

$$A^\top A = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

$$= \begin{bmatrix} 1\cdot 1+3\cdot 3+5\cdot 5 & 1\cdot 2+3\cdot 4+5\cdot 6 \\ 2\cdot 1+4\cdot 3+6\cdot 5 & 2\cdot 2+4\cdot 4+6\cdot 6 \end{bmatrix} = \begin{bmatrix} 35 & 44 \\ 44 & 56 \end{bmatrix}.$$

Properties 1.4. (i) Matrix addition is commutative and associative:

$$A + B = B + A\,, \qquad A + B + C = (A + B) + C = A + (B + C)\,,$$

where all three matrices are of the same size.

(ii) Matrix multiplication is associative as well, but it is not commutative. That is, AB is not always the same as BA. However, scalars can be swapped with matrices,

$$A(\lambda B) = \lambda AB\,.$$

(iii) For combinations of the two operations, distributive laws hold:

$$A(B + \widetilde{B}) = AB + A\widetilde{B}\,, \qquad (A + \widetilde{A})B = AB + \widetilde{A}B\,,$$

where $A, \widetilde{A}$ are $m \times n$ and $B, \widetilde{B}$ are $n \times k$.

(iv) For the transpose of a matrix multiplication, we have

$$(AB)^\top = B^\top A^\top,$$

where the sizes of A and B are as in the previous property.

Definition 1.5 (Vectors).

(i) A n-vector is a $n \times 1$ matrix:

$$v = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_n \end{bmatrix}.$$

We also refer to v as a vector of size n.

(ii) As for matrices, $v = w$ means that $v_i = w_i$ for all $i \in \{1, 2, \ldots, n\}$, and vectors of different sizes can never be equal. In equations like

$$v = 0,$$

the 0 on the right-hand side is understood to stand for the $n \times 1$ vector of all zeros rather than the number zero.

(iii) Addition and subtraction can only be carried out for vectors of the same size, and then the operation is carried out elementwise:

$$v \pm w = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \pm \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} v_1 \pm w_1 \\ v_2 \pm w_2 \\ \vdots \\ v_n \pm w_n \end{bmatrix}.$$

(iv) Multiplying a number with a vector is called scalar multiplication:

$$\lambda v = \lambda \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} \lambda v_1 \\ \lambda v_2 \\ \vdots \\ \lambda v_n \end{bmatrix}.$$

(v) Multiplication of a matrix and a vector is carried out according to the rules of matrix multiplication: If the number of columns of A agrees with the size of v, then

$$A\,v = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix},$$

where $c_i = \sum_{s=1}^{n} a_{is} v_s$. The result is a vector of size m.

(vi) The *scalar product* (or *dot product*) of two vectors v and w of the same size is

$$v \circ w = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \circ \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \sum_{s=1}^{n} v_s w_s \,.$$

(vii) A *row vector* of size n is a $1 \times n$ matrix:

$$v = \begin{bmatrix} v_1 & v_2 & v_3 & \cdots & v_n \end{bmatrix} .$$

(viii) The *norm* of a vector is

$$\|v\| = \sqrt{v_1^2 + v_2^2 + \ldots v_n^2} = \sqrt{v \circ v} \,.$$

Remark 1.6. (i) The term "vector" always refers to a one-column matrix as in (i) of Definition 1.5, and it is sometimes called *column vector* to distinguish it more explicitly from a row vector. Note that parts (i)-(v) of 1.5 are inherited from the corresponding matrix definitions, and only (vi)-(viii) are new and specific to vectors. Regarding multiplication and products: a simple dot "·" stands either for regular multiplication of two numbers, for scalar multiplication (multiplication of a number with a vector or matrix), or for matrix multiplication (multiplication of matrices and vectors of appropriate size). However, the "·" may be omitted. For scalar products, we always write the "∘".

(ii) The matrix transpose allows to write the scalar product of two vectors as a matrix product: Let v, w be two vectors of the same size, then

$$v \circ w = v^\top w \,.$$

(iii) When defining a vector in-line, it is more convenient to write down its transpose. For example, $v = \begin{bmatrix} 2 & -1 & 3 \end{bmatrix}^\top$ is the vector

$$v = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} .$$

(iv) We write $v \in \mathbb{R}^n$ for a vector of size n, meaning that v has n entries and each of them is a real number. For $n = 2$ or $n = 3$: Picturing a vector as an arrow in the two-dimensional plane $\mathbb{R}^2$ or in the three-dimensional space $\mathbb{R}^3$, the norm $\|v\|$ is simply its length. The angle θ between two vectors v and w of the same size is

$$\cos \theta = \frac{v \circ w}{\|v\| \, \|w\|} \,.$$

Example 1.7. (i)
$$\begin{bmatrix} 11 \\ -6 \\ 6 \end{bmatrix} - \begin{bmatrix} -7 \\ -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 18 \\ -5 \\ 3 \end{bmatrix}$$

(ii)
$$\begin{bmatrix} 4 & -1 \\ 0 & -3 \\ 7 & 1 \end{bmatrix} \begin{bmatrix} -9 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 \cdot (-9) + (-1) \cdot 3 \\ 0 \cdot (-9) + (-3) \cdot 3 \\ 7 \cdot (-9) + 1 \cdot 3 \end{bmatrix} = \begin{bmatrix} -39 \\ -9 \\ -60 \end{bmatrix}$$

(iii) The scalar product of $v = \begin{bmatrix} 4 & -4 & 1 \end{bmatrix}^\top$ and $w = \begin{bmatrix} 3 & -2 & -5 \end{bmatrix}^\top$ is

$$v \circ w = \begin{bmatrix} 4 \\ -4 \\ 1 \end{bmatrix} \circ \begin{bmatrix} 3 \\ -2 \\ -5 \end{bmatrix} = 4 \cdot 3 + (-4) \cdot (-2) + 1 \cdot (-5) = 15 \, .$$

(iv) The scalar product in the previous example can also be computed as a matrix multiplication, namely $v^\top w = 15$. Transposing the second vector instead gives a 3×3 matrix:

$$vw^\top = \begin{bmatrix} 4 \\ -4 \\ 1 \end{bmatrix} \begin{bmatrix} 3 & -2 & -5 \end{bmatrix}$$

$$= \begin{bmatrix} 4 \cdot 3 & 4 \cdot (-2) & 4 \cdot (-5) \\ (-4) \cdot 3 & (-4) \cdot (-2) & (-4) \cdot (-5) \\ 1 \cdot 3 & 1 \cdot (-2) & 1 \cdot (-5) \end{bmatrix} = \begin{bmatrix} 12 & -8 & -20 \\ -12 & 8 & 20 \\ 3 & -2 & -5 \end{bmatrix} \, .$$

(v)
$$\begin{bmatrix} 0 & -2 & 13 \end{bmatrix} \begin{bmatrix} 3 & 2 & -4 \\ 5 & -3 & 0 \\ 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 7 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 & -2 & 13 \end{bmatrix} \begin{bmatrix} 27 \\ -6 \\ -12 \end{bmatrix} = -144$$

(vi) The norm of $v = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}^\top$ is

$$||v|| = \sqrt{1^2 + 2^2 + 3^2} = \sqrt{14} \, .$$

For its angle with $\begin{bmatrix} 3 & 2 & 1 \end{bmatrix}^\top$, we find

$$\cos \theta = \frac{3 + 4 + 3}{\sqrt{14}\sqrt{14}} = \frac{5}{7} \, ,$$

which gives $\theta = \arccos\left(\frac{5}{7}\right) \approx 0.7752$ (this angle is in radians and corresponds to 44.42 degrees; here, arccos is the inverse function of cos).

Definition 1.8 (Determinants). The *determinant* of a square matrix is a scalar and defined as follows for matries of sizes up to 3×3.

11

(i) The determinant of a 1×1 matrix is

$$\det \begin{bmatrix} a_{11} \end{bmatrix} = a_{11} \,.$$

(ii) The determinant of a 2×2 matrix is

$$\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{21}a_{12} \,.$$

(iii) The determinant of a 3×3 matrix is

$$\det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32}$$

$$- a_{31}a_{22}a_{13} - a_{32}a_{23}a_{11} - a_{33}a_{21}a_{12} \,.$$

Remark 1.9. (i) One can write $|\ldots|$ as an abbreviation for $\det [\ldots]$.

(ii) Determinants of 1×1 matrices are quite straightforward; for 2×2, subtract the product of the off-diagonal entries from the product of the diagonal entries; see the following two points for 3×3. Determinants are also defined for larger square matrices, but the formulas become quite bulky then and one would not write them out as in the definition above.

(iii) Sarrus' rule should help you memorise the formula for the determinant of a 3×3 matrix: For A with entries as in (iii) of the definition above, write down A, copy the first two columns and append them to the right of the matrix; draw (or imagine) the lines

add up the products of the entries in each blue line, and subtract the three red products.

(iv) Laplace's formula:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \,.$$

Here we have "developed" the determinant along the first row (a_{11}, a_{12}, a_{13}), but it is also possible to develop along a different row or even along a column. The general formulas are

$$\det A = \sum_{j=1}^{3} (-1)^{i+j} a_{ij} \det \widetilde{A}_{ij} \qquad \text{(dev. along } i\text{-th row)},$$

$$\det A = \sum_{i=1}^{3} (-1)^{i+j} a_{ij} \det \widetilde{A}_{ij} \qquad \text{(dev. along } j\text{-th column)},$$

(1.1)

where $\widetilde{A}_{ij}$ is the 2×2 matrix we obtain by removing the i-th row and the j-th column from A. Choose the row/column wisely! For example, if there is a row or a column with only one nonzero entry, developing along it will shorten the computation.

(v) Laplace's formula generalises to larger square matrices – just replace the maximum index 3 of the sums in (1.1) with n. Note that Laplace's formula does not directly compute determinants – it reduces the problem of finding a $n \times n$ determinant to finding the determinants of n matrices of size $(n-1) \times (n-1)$. Sarrus' rule works only for 3×3's, it can not be applied to larger matrices.

Example 1.10. (i)
$$\begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} = 1 \cdot 4 - 2 \cdot 3 = -2$$

(ii)
$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} = 1 \cdot 5 \cdot 9 + 2 \cdot 6 \cdot 7 + 3 \cdot 4 \cdot 8 - 7 \cdot 5 \cdot 3 - 8 \cdot 6 \cdot 1 - 9 \cdot 4 \cdot 2 = 0$$

(iii)
$$\begin{vmatrix} 1 & -3 & 2 \\ 0 & 7 & -1 \\ 0 & 0 & -3 \end{vmatrix} = 1 \cdot 7 \cdot (-3) = -21$$

(iv) To compute the determinant of
$$A = \begin{bmatrix} 4 & -1 & 11 & 3 \\ -2 & 3 & -3 & -5 \\ 0 & 1 & -8 & 0 \\ 1 & 0 & -6 & -2 \end{bmatrix},$$

we use Laplace's formula, developing along the row with the most zero entries:

$$\det A = (-1)^{3+1} \cdot 0 \begin{vmatrix} \ldots \end{vmatrix} + (-1)^{3+2} \cdot 1 \begin{vmatrix} \ldots \end{vmatrix} + (-1)^{3+3} \cdot (-8) \begin{vmatrix} \ldots \end{vmatrix}$$

$$+ (-1)^{3+4} \cdot 0 \begin{vmatrix} \ldots \end{vmatrix} = - \begin{vmatrix} 4 & 11 & 3 \\ -2 & -3 & -5 \\ 1 & -6 & -2 \end{vmatrix} - 8 \begin{vmatrix} 4 & -1 & 3 \\ -2 & 3 & -5 \\ 1 & 0 & -2 \end{vmatrix}$$

$$= -4 \begin{vmatrix} -3 & -5 \\ -6 & -2 \end{vmatrix} + 11 \begin{vmatrix} -2 & -5 \\ 1 & -2 \end{vmatrix} - 3 \begin{vmatrix} -2 & -3 \\ 1 & -6 \end{vmatrix} - 8 \cdot \begin{vmatrix} \ldots \end{vmatrix} = \ldots = 342.$$

Properties 1.11. Let A be a $n \times n$ matrix and let λ be a scalar.

(i) If A has a row or a column of zeros, then $\det A = 0$.

(ii) If a row of A is a multiple of some other row of A, then $\det A = 0$. Similar for columns.

(iii) If A is an upper triangular matrix,

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & & a_{2n} \\ 0 & 0 & a_{33} & & a_{3n} \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & a_{nn} \end{bmatrix},$$

then the determinant is just the product along the diagonal,

$$\det A = a_{11}a_{22}a_{33}\ldots a_{nn} \,.$$

Application (Orthogonal projection). A data scientist is working on a large data set of the form

$$\{(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_n, y_n)\} \,.$$

She begins her analysis by plotting all data points in the xy-plane:

Data in the xy-plane:
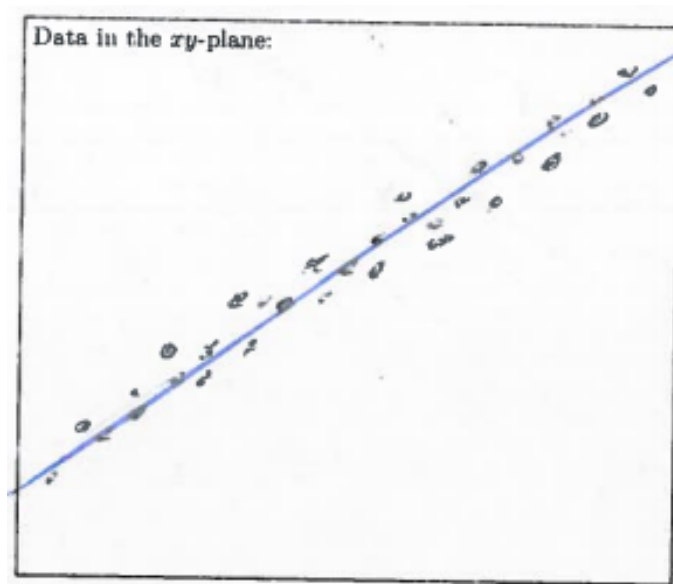
Here, n individuals or events were observed, and two pieces of information were recorded for each individual/event (e.g., n patients, x_k stands for the score on a certain medical test, and y_k stands for the score on a different test). Note that the data points are concentrated around a line. The data scientist has computed the slope of the line – we will learn how fit lines through data later – and now wants to project the data points onto it. One of the main advantages of this step is that the projected data will then be one-dimensional. That is, each point can then be described by one real number, namely its position along the line. This reduction can make a significant difference for the overall computational complexity of big data projects.

We now derive how to project points *orthogonally* onto a line. Let $L \subseteq \mathbb{R}^2$ be a line and let $w \in \mathbb{R}^2$ be a point that is not an element of L. Denote the projection of w onto L by $\widetilde{w}$. To project orthogonally means that the line segment connecting

w and $\widetilde{w}$ is orthogonal to L. Let v be a vector that spans L, i.e. L is the set of all scalar multiples of V,

$$L = \{\lambda v \mid \lambda \in \mathbb{R}\} \subseteq \mathbb{R}^2 \,.$$

This looks as follows, and the goal is now to find a formula for $\widetilde{w}$ in terms of v and w.

Orth. proj. onto a line:
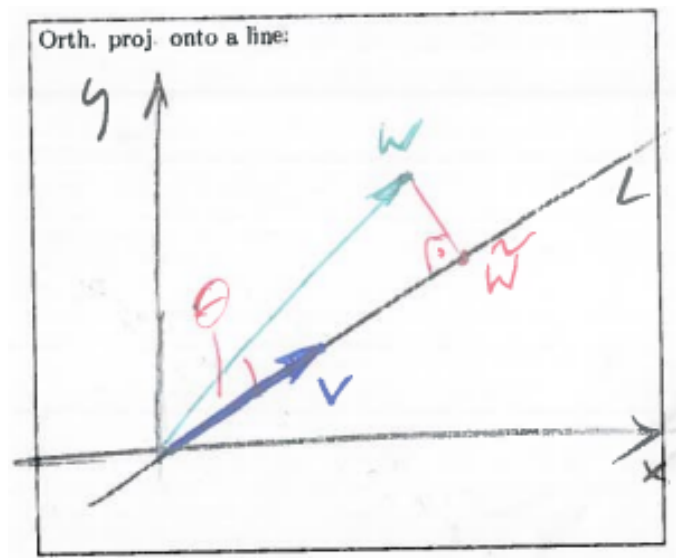
To visualise elements of $\mathbb{R}^2$, we are using two interpretations interchangeably: w can be interpreted as a point, e.g.,

$$w = (1, 2) \,,$$

or as a vector,

$$w = \begin{bmatrix} 1 \\ 2 \end{bmatrix} ,$$

where the latter is often drawn as an arrow. Basing the arrow at the origin $(0,0)$ of the xy-plane, and interpreting it as "move 1 in the x direction and 2 in the y direction", we see that the vector w points to the point w. This correspondence justifies switching freely between the two interpretations.

Back to the task at hand, let us collect a few formulas for the right-angled triangle in the sketch:

1.
$$\cos \theta = \frac{v \circ w}{\|v\|\,\|w\|}$$

(cf. Remark 1.6 (iv))

2.
$$\cos \theta = \frac{\text{length adjacent side}}{\text{length hypotenuse}} = \frac{\|\widetilde{w}\|}{\|w\|}$$

(trig. identity for right-angled triangle)

15

3.

$$\widetilde{w} = \frac{||\widetilde{w}||}{||v||} v$$

Regarding the last formula: $\widetilde{w}$ can be written as $\widetilde{w} = \lambda v$ with $\lambda > 0$, since it lies on the side of L in which v points. To verify that the factor $\lambda = {||\widetilde{w}||}/{||v||}$ is correct, we use the fact that $||\lambda v|| = |\lambda|\,||v||$:

$$\left|\left|\frac{||\widetilde{w}||}{||v||}v\right|\right| = \left|\frac{||\widetilde{w}||}{||v||}\right|\,||v|| = \frac{||\widetilde{w}||}{||v||}||v|| = ||\widetilde{w}|| \quad \checkmark$$

Combining these formulas gives

$$\widetilde{w} \overset{(3)}{=} v\cdot\frac{||\widetilde{w}||}{||v||} \overset{(2)}{=} v\cdot\frac{\cos\theta\,||w||}{||v||} \overset{(1)}{=} v\cdot\frac{\frac{v\circ w}{||v||\,||w||}||w||}{||v||} = \frac{v\cdot(v\circ w)}{||v||^2} \ .$$

Both $v\circ w$ and $||v||^2 = v\circ v$ are scalars that can be written as matrix products via transposition of the first factor. Once all operations are regular products, associativity of matrix multiplication can be used. We find that the projection $w\mapsto\widetilde{w}$ is carried out by multiplication with a 2×2 matrix:

$$\widetilde{w} = \frac{v\cdot(v^\top\cdot w)}{v^\top\cdot v} = \frac{v\,v^\top}{v^\top v}\,w = P_L\,w\ .$$

Exercise 1.12. (i) Practise matrix multiplication by finding AB and BA for the matrices[2]

$$A = \begin{bmatrix} 4 & 2 & -1 \\ -3 & -3 & 9 \\ 1 & 0 & 6 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & 3 & -1 \\ 0 & -4 & 2 \\ 0 & 0 & 2 \end{bmatrix}.$$

(ii) Compute the matrix product

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

The multiplication can be carried out in two different ways: $(AB)C$ and $A(BC)$ – try both approaches and make sure that you obtain the same result.

(iii) Consider the function $f(x) = x^2 - 2x - 3$. This expression can also be written as $x^2 - 2\cdot x^1 - 3\cdot x^0$, since $x^0 = 1$. For matrices, the zeroth power is defined similarly: $A^0 = I$, where I is the identity matrix. Hence find $f(A)$ for the matrix[3]

$$A = \begin{bmatrix} -1 & 0 \\ 4 & 3 \end{bmatrix}.$$

(iv) Find the angle between the two vectors[4]

$$v = \begin{bmatrix} 2 & -1 & 0 & 1 \end{bmatrix}^\top, \quad w = \begin{bmatrix} 4 & 2 & 2 & 0 \end{bmatrix}^\top.$$

(v) Pick a 2×3 and a 3×1 matrix and check that property (iv) in 1.4 holds. Similarly, verify the statements made in 1.11 for a few examples. (Be aware though that verifying examples does not constitute a proof!)

16

(vi) Compute the determinants of[5]

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

(vii) Let A, B, C be $n \times n$ matrices with the properties $AB = I$, $BC = I$, $CA = I$. Find $\frac{1}{3}\left(A^2 + B^2 + C^2\right)$.

(viii) Let I be the 2×2 identity matrix. Show that for any 2×2 matrix[6] A,

$$A\,I = I\,A = A\,.$$

(ix) The product AB is not always the same as BA, but for some matrices we do have $AB = BA$ – e.g., if one of the two is the zero matrix or the identity matrix. Perhaps there are more matrices that commute with a given matrix A.

Consider

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

For which 2×2 matrices X do we have[7] $AX = XA$?

(x) Show that[8]

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}^n = \begin{bmatrix} \cos n\theta & -\sin n\theta \\ \sin n\theta & \cos n\theta \end{bmatrix}.$$

(xi) Suppose you have to find the determinant of a 7×7 matrix that has no zero entries. You plan to reduce the problem to finding 3×3's by applying Laplace's formula several times. How many 3×3 determinants do you have to compute?

(xii) Find the two scalars λ_1 and λ_2 for which the determinant of

$$A_\lambda = \begin{bmatrix} \lambda & 1 & 1 \\ 1 & \lambda & 1 \\ 1 & 1 & \lambda \end{bmatrix}$$

is equal to zero[9].

(xiii) Project the point $p = (^1\!/_2, 10)$ orthogonally onto the line[10] $L : y = 5x$.

1.2 Systems of Linear Equations: Gaussian Elimination

Suppose that the two equations

$$\begin{cases} 2\,x & - & 3\,y & = & 7 \\ -2\,x & + & y & = & 5 \end{cases}$$

need to be satisfied simultaneously. Adding them together gives $-2y = 12$ and therefore $y = -6$. Now that y is known, the first of the original equations reads $2x - 3(-6) = 7$ and hence the solution is

$$x = -\frac{11}{2}, \; y = -6 \, . \tag{1.2}$$

Remark 1.13. (i) The first equation above defines a line in the xy-plane, namely $y = \frac{2}{3}x - \frac{7}{3}$. Similarly, the second equation describes the line $y = 2x + 5$. With the computation above, we have found the intersection of those two lines.

(ii) In order to convince yourself that it is permissible to add two equations together, think of two libra scales. Suppose that each of them is in balance. Then, taking the two objects from one scale and putting them onto the two arms of the other, the scale will still be in balance.

(iii) The solution (1.2) above – which consists of two equations – can also be written as one vector equation:
$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -\frac{11}{2} \\ -6 \end{bmatrix} .$$

Definition 1.14 (Linear systems of equations). A collection of equations of the form

$$\begin{cases} a_{11}\, x_1 & + & a_{12}\, x_2 & + & a_{13}\, x_3 & \ldots & + & a_{1n}\, x_n & = & b_1 \\ a_{21}\, x_1 & + & a_{22}\, x_2 & + & a_{23}\, x_3 & \ldots & + & a_{2n}\, x_n & = & b_2 \\ \;\vdots & & & & \;\vdots & & & & & \;\vdots \\ a_{m1}\, x_1 & + & a_{m2}\, x_2 & + & a_{m3}\, x_3 & \ldots & + & a_{mn}\, x_n & = & b_m \end{cases}$$

is called a *linear system of equations*. Here, $(x_1, x_2, \ldots, x_n)$ are the variables, the a_{ij} are the coefficients of the system, and $(b_1, b_2, \ldots, b_m)$ is the right-hand side (RHS). If the RHS is zero, $b_1 = b_2 = \cdots = b_m = 0$, then the system is called *homogeneous*. A combination of variables $(x_1^*, x_2^*, \ldots, x_n^*)$ that satisfies all m equations simultaneously is called a *solution* of the system.

Remark 1.15. (i) For $n = 2$ or $n = 3$ the variables are often called (x, y) or (x, y, z).

(ii) Each equation demands that some "linear combination" of the variables – that is, the variables multiplied by some coefficients and then added up – be equal to some given value. If powers or square roots or other functions of the variables appear in an equation, then it is not linear and can not be solved with the theory developed in this chapter.

(iii) Such a system can have no solution, a unique solution, or many solutions.

(iv) We refer to the equations in the system as *rows*, and the following definition lists modifications of a system that do not change the set of solutions. We have already used one such modification in the computation above: adding one row to another.

Definition 1.16 (Elementary row operations)**.** The following *row operations* on a system of equations do not change the set of its solutions.

(i) Add one row to another.

(ii) Multiply a row by a scalar different from zero.

(iii) Add the multiple of a row to another row.

(iv) Swap rows.

Example 1.17. Solve the system

$$\begin{cases} x & - & y & & & = & 3 \\ -3x & + & 4y & + & z & = & -1 \\ 2x & & & + & 7z & = & -3 \,. \end{cases}$$

Sol.: We "eliminate" the terms $-3x$ and $2x$ by adding 3 times the first row to the second and by subtracting 2 times the first row from the third. This transformation is denoted "$R2 \to R2 + 3 \cdot R1$" and "$R3 \to R3 - 2 \cdot R1$", i.e. the arrow is to be read as "is replaced with":

$$
\begin{aligned}
& \quad\quad\quad\quad \to \quad\quad \begin{cases} x & - & y & & & = & 3 \\ & & y & + & z & = & 8 \\ & & 2y & + & 7z & = & -9 \end{cases} \\
R2 &\to R2 + 3 \cdot R1 \\
R3 &\to R3 - 2 \cdot R1 \\
& \quad\quad\quad\quad \to \quad\quad \begin{cases} x & - & y & & & = & 3 \\ & & y & + & z & = & 8 \\ & & & & 5z & = & -25 \,. \end{cases} \\
R3 &\to R3 - 2 \cdot R2
\end{aligned}
\tag{1.3}
$$

The last line yields $z = -5$, and one can then find y and x by back-substituting into the first two equations,

$$x = 16, \; y = 13, \; z = -5 \,.$$

Remark 1.18. (i) The system in Definition 1.14 can also be written in matrix form. Keeping in mind that an equation of two n-vectors amounts to n equations, convince yourself that the following is equivalent to the system in 1.14.

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \,.$$

(ii) The transformation (1.3) via elementary row operations in the previous example is called *Gaussian elimination.* We now define two standard matrix forms that are the goal of this process.

Definition 1.19 (Augmented matrix, row echelon form, rank).

(i) The *augmented matrix* for the system in Definition 1.14 is

$$[A \mid b] = \left[\begin{array}{ccccc|c} a_{11} & a_{12} & a_{13} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & a_{23} & \ldots & a_{2n} & b_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \ldots & a_{mn} & b_m \end{array} \right],$$

and it can be subjected to elementary row operations the same way as fully written-out systems of linear equations.

(ii) The first nonzero entry in a row of a matrix is called the *leading entry*. A matrix (augmented or not) of the form

where all entries in the red (light) part are zero and the $(*)$ stand for leading entries, is said to be in *row echelon form* (REF). That is, zero rows are at the bottom and for the remaining rows – say there are r of them – we have

$$j_1 < j_2 < j_3 < \cdots < j_r \,,$$

where j_i is the column index of the leading entry in row i (e.g., in the schematic example above, $j_1 = 1, j_2 = 3, j_3 = 4, j_4 = 7$).

(iii) A matrix in REF is further said to be in *reduced row echelon form* (RREF), if all leading entries are 1 and if each leading entry is the only nonzero entry in its column.

(iv) The *rank* of a matrix is the number of nonzero rows in its REF.

Example 1.20. (i) The matrices

$$\begin{bmatrix} -3 & 0 & 1 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 2 & -3 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \\ 0 & 0 & 0 \end{bmatrix}$$

all are in REF and have ranks 2, 2, and 3. If those matrices are the outcome of Gaussian elimination, then the original matrices have the same rank – the rank does not change under elementary row operations, but it can be read off directly only from the REF. The matrices

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

are not in REF.

(ii) The matrices

$$I, \begin{bmatrix} 1 & -2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & -3 \end{bmatrix}$$

all are in RREF, but the matrices in (i) are not.

(iii) In Example 1.17, we have transformed

$$\begin{bmatrix} 1 & -1 & 0 & 3 \\ -3 & 4 & 1 & -1 \\ 2 & 0 & 7 & -3 \end{bmatrix} \overset{(1.3)}{\longrightarrow} \begin{bmatrix} 1 & -1 & 0 & 3 \\ 0 & 1 & 1 & 8 \\ 0 & 0 & 5 & -25 \end{bmatrix}.$$

The transformed augmented matrix is in REF and has rank 3. Therefore, the original augmented matrix has rank 3 as well. Including the vector of variables that is dropped when writing systems as augmented matrices, the last line reads

$$\begin{bmatrix} 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = -25 \,,$$

i.e., $0 \cdot x + 0 \cdot y + 5 \cdot z = 5z = -25$. With back-substitution, one can then find x and y as before. Another option is to bring the augmented matrix into reduced row echelon form:

$$\begin{bmatrix} 1 & -1 & 0 & 3 \\ 0 & 1 & 1 & 8 \\ 0 & 0 & 5 & -25 \end{bmatrix} \overset{R3 \to 1/5 \cdot R3}{\longrightarrow} \begin{bmatrix} 1 & -1 & 0 & 3 \\ 0 & 1 & 1 & 8 \\ 0 & 0 & 1 & -5 \end{bmatrix}$$

$$\overset{R2 \to R2 - R3}{\longrightarrow} \begin{bmatrix} 1 & -1 & 0 & 3 \\ 0 & 1 & 0 & 13 \\ 0 & 0 & 1 & -5 \end{bmatrix}$$

$$\overset{R1 \to R1 + R2}{\longrightarrow} \begin{bmatrix} 1 & 0 & 0 & 16 \\ 0 & 1 & 0 & 13 \\ 0 & 0 & 1 & -5 \end{bmatrix}.$$

The augmented matrix now corresponds to

$$\begin{bmatrix} 16 \\ 13 \\ -5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix},$$

from which – for this example – the solution can be read off directly.

(iv) To solve the system

$$\begin{cases} -3x & & + & 3z & = & 4 \\ 3x & + & 5y & + & z & = & 0 \\ -x & + & 5y & + & 5z & = & 3 \,, \end{cases}$$

we construct the augmented matrix and transform it with elementary row operations. It is preferable to have an entry 1 in the upper left corner – therefore, we start by swapping row 3 to the top and multiply it by -1 :

$$\begin{bmatrix} -3 & 0 & 3 & 4 \\ 3 & 5 & 1 & 0 \\ -1 & 5 & 5 & 3 \end{bmatrix} \xrightarrow{R3 \leftrightarrow R1} \begin{bmatrix} -1 & 5 & 5 & 3 \\ 3 & 5 & 1 & 0 \\ -3 & 0 & 3 & 4 \end{bmatrix}$$

$$\xrightarrow{R1 \to -R1} \begin{bmatrix} 1 & -5 & -5 & -3 \\ 3 & 5 & 1 & 0 \\ -3 & 0 & 3 & 4 \end{bmatrix}$$

$$\xrightarrow{R2 \to R2 - 3R1} \begin{bmatrix} 1 & -5 & -5 & -3 \\ 0 & 20 & 16 & 9 \\ -3 & 0 & 3 & 4 \end{bmatrix}$$

$$\xrightarrow{R3 \to R3 + 3R1} \begin{bmatrix} 1 & -5 & -5 & -3 \\ 0 & 20 & 16 & 9 \\ 0 & -15 & -12 & -5 \end{bmatrix}$$

$$\xrightarrow{R3 \to R3 + {}^3/_4 R2} \begin{bmatrix} 1 & -5 & -5 & -3 \\ 0 & 20 & 16 & 9 \\ 0 & 0 & 0 & {}^7/_4 \end{bmatrix}.$$

The last line reads

$$ {}^7/_4 = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 0 \cdot x + 0 \cdot y + 0 \cdot z = 0 \,,$$

which is never true – for no combination (x, y, z). Hence the system of linear equations does not have a solution.

Remark 1.21. As the previous example has shown, if the REF of the augmented matrix has a row of zeros on the left-hand side, but the corresponding entry on the right is different from zero, then the system does not have a solution. A formal way of expressing this is

$$\operatorname{rank} A < \operatorname{rank} [A \mid b]$$

– the rank of the coefficient matrix alone is smaller than the rank of the augmented matrix.

Theorem 1.22 (Solutions of systems of linear equations). Suppose a system of m linear equations in n variables is given. That is, the coefficient matrix A of the system is of size $m \times n$. Denote the right-hand side of the system by b. Then:

(i) If

$$\operatorname{rank} A < \operatorname{rank} [A \mid b] \,,$$

then the system does not have a solution.

(ii) If
$$\operatorname{rank} A = \operatorname{rank}\left[A \mid b\right] = n\,,$$

then the system has a unique solution.

(iii) If
$$\operatorname{rank} A = \operatorname{rank}\left[A \mid b\right] < n\,,$$

then the system has many solutions.

(iv) Those cases cover all possibilities, as rank A can not be greater than n or rank $[A \mid b]$.

Example 1.23. (i) Example (iii) of 1.20 corresponds to the unique-solution case of Theorem 1.22, and example (iv) to the no-solution case. The third case is demonstrated in the following examples.

(ii) Let us re-do Example 1.20 (iv) with the right-hand side $b = \begin{bmatrix} -9 & 8 & -4 \end{bmatrix}^{\top}$. The same row operations as before lead to

$$\left[\begin{array}{ccc|c} -3 & 0 & 3 & -9 \\ 3 & 5 & 1 & 8 \\ -1 & 5 & 5 & -4 \end{array}\right] \quad \dashrightarrow \quad \left[\begin{array}{ccc|c} 1 & -5 & -5 & 4 \\ 0 & 20 & 16 & -4 \\ 0 & -15 & -12 & 3 \end{array}\right].$$

The third row is a multiple of the second, and the next operation, $R3 \to R3 + {}^{3}/_{4}R2$, eliminates the third row altogether. We then continue to bring the system into RREF:

$$\left[\begin{array}{ccc|c} 1 & -5 & -5 & 4 \\ 0 & 20 & 16 & -4 \\ 0 & 0 & 0 & 0 \end{array}\right] \quad \overset{R2 \to {}^{1}/_{20}R2}{\to} \quad \left[\begin{array}{ccc|c} 1 & -5 & -5 & 4 \\ 0 & 1 & {}^{4}/_{5} & -{}^{1}/_{5} \\ 0 & 0 & 0 & 0 \end{array}\right]$$

$$\overset{R1 \to R1 + 5R2}{\to} \quad \left[\begin{array}{ccc|c} 1 & 0 & -1 & 3 \\ 0 & 1 & {}^{4}/_{5} & -{}^{1}/_{5} \\ 0 & 0 & 0 & 0 \end{array}\right].$$

The solution $\begin{bmatrix} x & y & z \end{bmatrix}^{\top}$ is now obtained as follows. Variables corresponding to columns that do not have a leading entry can be chosen freely. This is expressed using a parameter,

$$z = t \qquad (t \in \mathbb{R})\,.$$

To obtain an expression for y, we use the row whose leading entry is in the column corresponding to y:

$$y + \tfrac{4}{5}z = -\tfrac{1}{5} \quad \to \quad y = -\tfrac{1}{5} - \tfrac{4}{5}t\,.$$

Similarly for x:

$$x - z = 3 \quad \to \quad x = 3 + t\,.$$

Therefore, the answer is

$$\begin{cases} x = 3 + t \\ y = -\frac{1}{5} - \frac{4}{5}t \\ z = t \, , \end{cases} \qquad (t \in \mathbb{R}),$$

which can also be written as

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3 \\ -\frac{1}{5} \\ 0 \end{bmatrix} + t \cdot \begin{bmatrix} 1 \\ -\frac{4}{5} \\ 1 \end{bmatrix}, \qquad (t \in \mathbb{R}).$$

The latter form is the equation of a line – namely the line of common points of the three planes $-3x + 3z = -9$, $3x + 5y + z = 8$, $-x + 5y + 5z = -4$.

(iii) For the system

$$\begin{cases} -6x_1 & +6x_2 & +2x_3 & -2x_4 & = & 2 \\ -9x_1 & +8x_2 & +3x_3 & -2x_4 & = & 3 \\ -3x_1 & +2x_2 & + x_3 & & = & 1 \\ -15x_1 & +14x_2 & +5x_3 & -4x_4 & = & 5 \, , \end{cases}$$

Gaussian elimination leads to the REF

$$\left[\begin{array}{cccc|c} -3 & 2 & 1 & 0 & 1 \\ 0 & 2 & 0 & -2 & 0 \end{array} \right]$$

(dropping two rows of zeros) and to the RREF

$$\left[\begin{array}{cccc|c} 1 & 0 & -\frac{1}{3} & -\frac{2}{3} & -\frac{1}{3} \\ 0 & 1 & 0 & -1 & 0 \end{array} \right].$$

The solution is

$$\begin{cases} x_1 = \frac{1}{3}(t + 2s - 1) \\ x_2 = s \\ x_3 = t \\ x_4 = s \, , \end{cases}$$

where $t, s \in \mathbb{R}$.

Exercise 1.24. (i) Fill in the Gaussian elimination steps for Example 1.23 (iii) (careful: the RREF of a matrix is unique, but the REF is not – therefore, the REF you find may differ from the one given above).

(ii) Find rank A and rank $[A \,|\, b]$ for the augmented matrices

$$\left[\begin{array}{cc|c} 1 & 0 & \frac{1}{3} \\ 1 & 1 & \frac{1}{3} \end{array} \right], \quad \left[\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 0 & 5 & 6 & 7 \\ 0 & 0 & 8 & 9 \\ 2 & 5 & 7 & 3 \end{array} \right], \quad \left[\begin{array}{cccc|c} 3 & 2 & 0 & 5 & 0 \\ 3 & -2 & 3 & 6 & -1 \\ 2 & 0 & 1 & 5 & -3 \\ 1 & 6 & -4 & -1 & 4 \end{array} \right],$$

and state for each case whether we have no solution, a unique solution, or infinitely many solutions[11].

(iii) Find all solutions of the homogeneous system[12]

$$\begin{cases} 4x & -2y & +4z & = & 0 \\ & +y & +2z & = & 0 \\ 3x & -y & +4z & = & 0 \,. \end{cases}$$

(iv) Convince yourself of the fact that homogeneous systems always have a solution, i.e. at least one solution[13].

(v) Find all solutions of the inhomogeneous system

$$\begin{cases} 4x & -2y & +4z & = & -3 \\ -x & +y & +2z & = & 1 \\ 3x & -y & +4z & = & 7 \,. \end{cases}$$

(vi) An equation $a_1 x + a_2 y + a_3 z = b$ describes a plane in $\mathbb{R}^3$ (just as $a_1 x + a_2 y = b$ describes a line in $\mathbb{R}^2$; here, $b \in \mathbb{R}$). For example, $y = 0$ describes the xz-plane. Think about this interpretation for a moment and connect it to solving systems of equations, i.e., finding simultaneous solutions of several equations. For example, try to picture different arrangements of three planes such that they have no point in common, exactly one point in common, and infinitely many points in common.

(vii) Find α such that
$$\begin{cases} 5x & -3y & = & 2 \\ -x & +2y & = & 1 \\ -4x & +4y & = & \alpha \end{cases}$$

has a solution[14].

(viii) Find α, β such that
$$\begin{bmatrix} \alpha & 1 & 1 \\ 1 & \beta & 1 \\ 1 & 3\beta & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \\ 4 \end{bmatrix}$$

has no solution, a unique solution, infinitely many solutions[15].

(ix) In the last exercise of the previous section, you were asked to orthogonally project the point $p = (^1/_2, 10)$ onto the line $L \,:\, y = 5x$. Now consider the following two different types of projections.

(a) $P_{L,x} \,:\, p = (^1/_2, 10) \mapsto \tilde{p} = (2, 10)$ (projection in the x direction)
(b) $P_{L,y} \,:\, p = (^1/_2, 10) \mapsto \tilde{p} = (^1/_2, {}^5/_2)$ (projection in the y direction)

Sketch the action of the three different types of projections and find the matrices $P_{L,x}, P_{L,y}$. What is the advantage of orthogonal projection over projection along the coordinate axes[16]?

25

1.3 Eigenvalues and Eigenvectors

Definition 1.25 (Eigenvalues and eigenvectors)**.** Let A be a $n \times n$ square matrix. A scalar $\lambda \in \mathbb{R}$ is called an *eigenvalue* of A, if there exists a n-vector $v \neq 0$ with

$$Av = \lambda v \, .$$

Such a vector v is called an *eigenvector* of A.

Remark 1.26.   (i) In this section, all matrices are square, of size $n \times n$, and vectors are of the corresponding size n.

 (ii) The vector on the left-hand side of the eigenvalue equation above is obtained via matrix multiplication. The vector on the right is found by scalar multiplication, which is much easier to compute. This observation already suggests that eigenvalues and eigenvectors are useful for simplifying matrix computations.

(iii) The condition $v \neq 0$ for eigenvectors is crucial. Indeed, for the zero vector, we have
$$A0 = 0 = \lambda 0$$
for any square matrix A and for any scalar λ. Therefore, eigenvectors are nonzero by definition (otherwise, any $\lambda \in \mathbb{R}$ would be an eigenvalue). However, eigen*values* can be zero: If there exists $v \neq 0$ such that
$$Av = 0 = 0 \cdot v \, ,$$
then v is an eigenvector of A with eigenvalue $\lambda = 0$ (where the first 0 in the equation stands for zero vector and the second 0 for the number zero).

(iv) If v is an eigenvector of A, then any nonzero multiple of v is an eigenvector with the same eigenvalue: If we scalar-multiply v by $\mu \neq 0$, then
$$A(\mu v) = \mu(Av) = \mu(\lambda v) = \lambda(\mu v) \, .$$

 (v) An approach to computing eigenvalues will be given by Theorem 1.28 below. That theorem and the results in the preceding lemma provide good opportunities to present a few proofs.

Lemma 1.27. For a $n \times n$ matrix M, we have:

 (i) There exists $v \neq 0$ with $Mv = 0 \quad \Longleftrightarrow \quad \text{rank } M < n$.

 (ii) The determinant of M does not change when M is subjected to elementary row operation (iii) in Definition 1.16 – adding the multiple of a row to another row.

(iii) Each time two rows are swapped, the determinant changes by a factor of -1.

(iv) $\text{rank } M < n \quad \Longleftrightarrow \quad \det M = 0$.

Proof. (i) Note that rank M = rank $[M \mid 0]$, and we are therefore either in case (ii) or in case (iii) of Theorem 1.22. Note further that the zero vector $w = 0$ solves $Mw = 0$. If rank $M = n$, this solution is unique by 1.22 – that is, there are no other solutions and hence no nonzero solutions. If rank $M < n$, there are other solutions – that is, nonzero solutions do exist in that case.

(ii) Denote the rows of M by $r^{(1)}, r^{(2)}, \ldots, r^{(n)}$, and let B be the matrix obtained by adding μ times the q-th row to the p-th row, where $p \neq q$:

$$
M = \begin{bmatrix} -- \, r^{(1)} \, -- \\ -- \, r^{(2)} \, -- \\ \vdots \\ -- \, r^{(n)} \, -- \end{bmatrix}, \qquad
B = \begin{bmatrix} -- \, r^{(1)} \, -- \\ \vdots \\ -- \, r^{(p-1)} \, -- \\ - \, r^{(p)} + \mu r^{(q)} \, - \\ -- \, r^{(p+1)} \, -- \\ \vdots \\ -- \, r^{(n)} \, -- \end{bmatrix} .
$$

To show that $\det B = \det M$, we apply Laplace's formula developing along the p-th row of B. As in the previous section, the matrix $\widetilde{A}_{ij}$ is the $(n-1)\times(n-1)$ matrix obtained by deleting the i-th row and the j-th column from A:

$$
\det B = \sum_{j=1}^{n} (-1)^{p+j} \left(r_j^{(p)} + \mu r_j^{(q)} \right) \det \widetilde{B}_{pj}
$$

$$
= \sum_{j=1}^{n} (-1)^{p+j} \left(r_j^{(p)} + \mu r_j^{(q)} \right) \det \widetilde{M}_{pj}
$$

$$
= \sum_{j=1}^{n} (-1)^{p+j} r_j^{(p)} \det \widetilde{M}_{pj} + \sum_{j=1}^{n} (-1)^{p+j} \mu r_j^{(q)} \det \widetilde{M}_{pj}
$$

$$
= \det M + \det \left(\begin{bmatrix} r^{(1)} \\ \vdots \\ r^{(q)} \\ \vdots \\ \mu r^{(q)} \\ \vdots \\ r^{(n)} \end{bmatrix} \right) = \det M + 0 = \det M \, ,
$$

where the simplification in the last line is due to property (ii) of 1.11. We have shown that the determinant does not change when a multiple of a row is added to another row.

(iii) By induction over n :

n = 2 :
$$
\begin{vmatrix} c & d \\ a & b \end{vmatrix} = bc - ad = - \begin{vmatrix} a & b \\ c & d \end{vmatrix} \qquad \checkmark
$$

27

n = 3 : Suppose two rows of a 3×3 matrix have been swapped. Apply Laplace's formula in the row that has not changed. In each of the three 2×2 matrices, the rows have been swapped, and we therefore – cf. the $n = 2$ case above – obtain an overall factor of -1.

n = 4 : Suppose two rows of a 4×4 matrix have been swapped. Apply Laplace's formula in one of the rows that has not changed. In each of the four 3×3 matrices, two rows have been swapped, and we therefore – cf. the $n = 3$ case above – obtain an overall factor of -1.

n > 4 : Via repetition of the step $n \rightsquigarrow n + 1$, the statement follows for all $n \geq 2$.

(iv) First suppose that M is in REF. Square matrices in REF have only zero entries below the diagonal. Therefore, M is a upper diagonal matrix and its determinant is the product of its diagonal entries. If rank $M < n$, then at least one diagonal entry is equal to zero, and thus $\det M = 0$. If rank $M = n$, then none of the diagonal entries is zero, and therefore $\det M \neq 0$. Hence statement (iv) is true for REF matrices. Since all matrices can be brought into REF by swapping rows and adding multiples of rows to other rows, the general case follows from (ii) and (iii).

□

Theorem 1.28 (Eigenvalue equation)**.** If λ is an eigenvalue of A, then

$$\det(A - \lambda I) = 0 \,.$$

The converse is also true: If $\det(A - \lambda I) = 0$, then λ is an eigenvalue of A. The above equation – a condition on the parameter λ – is called the *eigenvalue equation* of A.

Proof.

$$
\begin{array}{rlrll}
\lambda \text{ eigenvalue of } A & \overset{\text{def.}}{\Longleftrightarrow} & Av & = \lambda v & \text{(for some } v \neq 0) \\
& \Longleftrightarrow & Av - \lambda v & = 0 & \text{(for some } v \neq 0) \\
& \Longleftrightarrow & (A - \lambda I)v & = 0 & \text{(for some } v \neq 0) \\
& \overset{\text{Lemma 1.27 (i)}}{\Longleftrightarrow} & \text{rank}\,(A - \lambda I) & < n & \\
& \overset{\text{1.27 (iv)}}{\Longleftrightarrow} & \det(A - \lambda I) & = 0 &
\end{array}
$$

□

Example 1.29. (i) Find the eigenvalues and eigenvectors of

$$A = \begin{bmatrix} 1 & 5 \\ 10 & -4 \end{bmatrix}.$$

Sol.: The eigenvalue equation is

$$
\begin{aligned}
0 = \det(A - \lambda I) &= \det\left(\begin{bmatrix} 1 & 5 \\ 10 & -4 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \\
&= \begin{vmatrix} 1 - \lambda & 5 \\ 10 & -4 - \lambda \end{vmatrix} = (1 - \lambda)(-4 - \lambda) - 10 \cdot 5 \\
&= \lambda^2 + 3\lambda - 54 = (\lambda + 9)(\lambda - 6) \,,
\end{aligned}
$$

and its roots

$$\begin{cases} \lambda_1 = -9 \\ \lambda_2 = 6 \, . \end{cases}$$

are the eigenvalues of A.

The eigenvector v_1 is found by solving $Av_1 = \lambda_1 v_1 \leftrightarrow (A - \lambda_1 I)v_1 = 0$:

$$\left[\begin{array}{cc|c} 1 - \lambda_1 & 5 & 0 \\ 10 & -4 - \lambda_1 & 0 \end{array} \right] \qquad = \qquad \left[\begin{array}{cc|c} 10 & 5 & 0 \\ 10 & 5 & 0 \end{array} \right]$$

$$\overset{R2 \to R2 - R1}{\longrightarrow} \left[\begin{array}{cc|c} 10 & 5 & 0 \\ 0 & 0 & 0 \end{array} \right].$$

The first row of that system corresponds to

$$10x + 5y = 0 \, ,$$

which gives $y = -2x$. As was pointed out in Remark 1.26, multiples of eigenvectors are eigenvectors as well – this explains why the above system does not determine x and y uniquely; for any solution (x, y) and any constant factor $c \neq 0$, the scalar multiple (cx, cy) is a solution as well. Choosing $x = 1$, we obtain

$$v_1 = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \qquad (\lambda_1 = -9) \, .$$

For the second eigenvector, we solve $(A - \lambda_2 I)v_2 = 0$:

$$\left[\begin{array}{cc|c} 1 - \lambda_2 & 5 & 0 \\ 10 & -4 - \lambda_2 & 0 \end{array} \right] \qquad = \qquad \left[\begin{array}{cc|c} -5 & 5 & 0 \\ 10 & -10 & 0 \end{array} \right]$$

$$\overset{R2 \to R2 + 2R1}{\longrightarrow} \left[\begin{array}{cc|c} -5 & 5 & 0 \\ 0 & 0 & 0 \end{array} \right].$$

Here, the first row reads

$$-5x + 5y = 0$$

and leads to

$$v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad (\lambda_2 = 6) \, .$$

(ii) Verify the results from the previous example.
   Sol.:

$$Av_1 = \begin{bmatrix} 1 & 5 \\ 10 & -4 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 1 - 10 \\ 10 + 8 \end{bmatrix} = \begin{bmatrix} -9 \\ 18 \end{bmatrix} = -9 \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \lambda_1 v_1 \qquad \checkmark$$

$$Av_2 = \begin{bmatrix} 1 & 5 \\ 10 & -4 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 + 5 \\ 10 - 4 \end{bmatrix} = \begin{bmatrix} 6 \\ 6 \end{bmatrix} = 6 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \lambda_2 v_2 \qquad \checkmark$$

(iii) Find the eigenvalues and eigenvectors of

$$M = \begin{bmatrix} -5 & 0 & 7 \\ 6 & 2 & -6 \\ -4 & 0 & 6 \end{bmatrix}.$$

Sol.:

$$
\begin{aligned}
0 = \det(M - \lambda I) &= \begin{vmatrix} -5 - \lambda & 0 & 7 \\ 6 & 2 - \lambda & -6 \\ -4 & 0 & 6 - \lambda \end{vmatrix} \\
&= (-5 - \lambda)(2 - \lambda)(6 - \lambda) + 0 + 0 - (-4)(2 - \lambda)(7) - 0 - 0 \\
&= (2 - \lambda)\left[(-5 - \lambda)(6 - \lambda) + 28\right] = (2 - \lambda)\left[(\lambda + 5)(\lambda - 6) + 28\right] \\
&= (2 - \lambda)\left[\lambda^2 - \lambda - 2\right] = -(\lambda + 1)(\lambda - 2)(\lambda - 2),
\end{aligned}
$$

which gives eigenvalues

$$\begin{cases} \lambda_1 = -1 \\ \lambda_2 = 2 \\ \lambda_3 = 2. \end{cases}$$

For v_1 :

$$\begin{bmatrix} -5 - \lambda_1 & 0 & 7 & \big| & 0 \\ 6 & 2 - \lambda_1 & -6 & \big| & 0 \\ -4 & 0 & 6 - \lambda_1 & \big| & 0 \end{bmatrix} = \begin{bmatrix} -4 & 0 & 7 & \big| & 0 \\ 6 & 3 & -6 & \big| & 0 \\ -4 & 0 & 7 & \big| & 0 \end{bmatrix}$$

$$\xrightarrow{R3 \to R3 - R1} \begin{bmatrix} -4 & 0 & 7 & \big| & 0 \\ 6 & 3 & -6 & \big| & 0 \\ 0 & 0 & 0 & \big| & 0 \end{bmatrix}$$

$$\xrightarrow{R2 \to R2 + {}^3/_2 R1} \begin{bmatrix} -4 & 0 & 7 & \big| & 0 \\ 0 & 3 & {}^9/_2 & \big| & 0 \\ 0 & 0 & 0 & \big| & 0 \end{bmatrix}$$

$$\xrightarrow{R2 \to {}^1/_3 R2,\ R1 \to -{}^1/_4 R1} \begin{bmatrix} 1 & 0 & -{}^7/_4 & \big| & 0 \\ 0 & 1 & {}^3/_2 & \big| & 0 \\ 0 & 0 & 0 & \big| & 0 \end{bmatrix}.$$

We now need to find the components x, y, z of v_1. One of them can be chosen freely, say $z = 4$. Then the first equation reads $x - \frac{7}{4}4 = 0$, yielding $x = 7$. The second equation reads $y + \frac{3}{2}4 = 0$ and we obtain $y = -6$ and

$$v_1 = \begin{bmatrix} 7 \\ -6 \\ 4 \end{bmatrix} \qquad (\lambda_1 = -1).$$

For a matrix with three distinct eigenvalues, the computations of v_2 and v_3 would be analogous to the computation of v_1. The matrix in this example

has a double eigenvalue though, $\lambda_2 = \lambda_3 = 2$. In this case, one has to solve only one system, $(A - 2I)v = 0$. Due to $\lambda = 2$ being a double eigenvalue, it has a larger set of solutions from which one then has to choose two *different* solutions v_2 and v_3. Here, "different" means that v_3 is not simply a scalar multiple of v_2.

For $v_{2/3}$:

$$
\left[\begin{array}{ccc|c} -5 - \lambda_{2/3} & 0 & 7 & 0 \\ 6 & 2 - \lambda_{2/3} & -6 & 0 \\ -4 & 0 & 6 - \lambda_{2/3} & 0 \end{array} \right]
=
\left[\begin{array}{ccc|c} -7 & 0 & 7 & 0 \\ 6 & 0 & -6 & 0 \\ -4 & 0 & 4 & 0 \end{array} \right]
$$

$$
\overset{\cdots}{\Rightarrow}
\left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right].
$$

We are left with one equation that imposes a condition on three variables. A simple choice to represent the solution is

$$
v_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad v_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \qquad (\lambda_2 = \lambda_3 = 2) .
$$

(Check that these two vectors solve the system of equations represented by the augmented matrix above.)

(iv) Find the eigenvalues and eigenvectors of

$$
L = \begin{bmatrix} 0 & 0.75 \\ 0.75 & 0.4375 \end{bmatrix} .
$$

Sol.: The eigenvalue equation is

$$
0 = \det(L - \lambda I) = \begin{vmatrix} 0 - \lambda & 0.75 \\ 0.75 & 0.4375 - \lambda \end{vmatrix} = \lambda^2 - 0.4375\lambda - 0.5625 ,
$$

which gives eigenvalues (none of the decimals appearing here is rounded)

$$
\begin{cases} \lambda_1 = -0.5625 \\ \lambda_2 = 1 . \end{cases}
$$

We find v_1 by solving $(L - \lambda_1 I)v_1 = 0$:

$$
\left[\begin{array}{cc|c} 0 - \lambda_1 & 0.75 & 0 \\ 0.75 & 0.4375 - \lambda_1 & 0 \end{array} \right]
=
\left[\begin{array}{cc|c} 0.5625 & 0.75 & 0 \\ 0.75 & 1 & 0 \end{array} \right]
$$

$$
\overset{R2 \to R2 - \frac{0.75}{0.5625} R1}{\Longrightarrow}
\left[\begin{array}{cc|c} 0.5625 & 0.75 & 0 \\ 0 & 0 & 0 \end{array} \right].
$$

Letting x and y stand for the components of v_1, the first equation reads

$$\left(\frac{3}{4}\right)^2 x + \frac{3}{4} y = 0 \,,$$

which gives $x = -\frac{4}{3}y$. As in the previous examples, y can be chosen freely and then x is determined accordingly. We choose

$$v_1 = \begin{bmatrix} 0.8 \\ -0.6 \end{bmatrix} \qquad \text{(for } \lambda_1 = -0.5625) \,,$$

and obtain the second eigenvector similarly:

$$v_2 = \begin{bmatrix} 0.6 \\ 0.8 \end{bmatrix} \qquad \text{(for } \lambda_2 = 1) \,.$$

Definition 1.30 (Linear independence). Consider a set

$$S = \{v^{(1)}, v^{(2)}, v^{(3)}, \ldots v^{(m)}\} \subseteq \mathbb{R}^n$$

of m vectors of size n.

(i) A vector $w \in \mathbb{R}^n$ is said to be a *linear combination* of the vectors in S if it can be written in the form

$$w = c_1 v^{(1)} + c_2 v^{(2)} + c_3 v^{(3)} + \cdots + c_m v^{(m)}$$

for some coefficients $c_1, c_2, c_3, \ldots, c_m$.

(ii) The vectors in S are said to be *linearly independent* if

$$c_1 v^{(1)} + c_2 v^{(2)} + c_3 v^{(3)} + \cdots + c_m v^{(m)} = 0$$

implies

$$c_1 = 0, \ c_2 = 0, \ c_3 = 0, \ \ldots, \ c_m = 0 \,.$$

That is, the vectors of S are called linearly independent if the zero vector can only be written as the trivial linear combination of vectors in S (the linear combination where all coefficients are zero). Otherwise, the vectors are called *linearly dependent*.

Remark 1.31. (i) Suppose we have a linearly dependent set of vectors,

$$c_1 v^{(1)} + c_2 v^{(2)} + c_3 v^{(3)} + \cdots + c_m v^{(m)} = 0 \,.$$

Not all of the coefficients are equal to zero; say $c_m \neq 0$. Then

$$v^{(m)} = -\frac{c_1}{c_m} v^{(1)} - \frac{c_2}{c_m} v^{(2)} - \frac{c_3}{c_m} v^{(3)} - \cdots - \frac{c_{m-1}}{c_m} v^{(m-1)} \,,$$

which shows that in a set of linearly dependent vectors, at least one of the vectors is a linear combination of the others.

(ii) For 3×3 matrices, the eigenvalue equation is a third-order equation, which can not be solved as readily as a second-order equation. One therefore should be careful not to give away any information when computing the determinant: In Example 1.29 (iii), a factor of $(2 - \lambda)$ was kept rather than multiplied out – multiplying it out would have given an expression of the form

$$-\lambda^3 + a\lambda^2 + b\lambda + c \,,$$

of which one then would have to guess a root before being able to continue with polynomial division and the quadratic formula.

Example 1.32. (i) Show that $v^{(1)} = \begin{bmatrix} 1 & 0 \end{bmatrix}^\top$, $v^{(2)} = \begin{bmatrix} 0 & 1 \end{bmatrix}^\top$ are linearly independent.
Sol.: Suppose we have coefficients c_1 and c_2 such that the corresponding linear combination of the two vectors gives the zero vector. That is,

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = c_1 v^{(1)} + c_2 v^{(2)} = \begin{bmatrix} v_1^{(1)} & v_1^{(2)} \\ v_2^{(1)} & v_2^{(2)} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix},$$

which reads $c_1 = 0$ and $c_2 = 0$. Therefore, the vectors $v^{(1)}, v^{(2)}$ are linearly independent. This computation was simplified by the fact that the matrix obtained from the vectors is the identity matrix. In less straight-forward cases and also for larger n, one checks whether the system

$$\begin{bmatrix} v_1^{(1)} & v_1^{(2)} & \cdots & v_1^{(m)} \\ v_2^{(1)} & v_2^{(2)} & \cdots & v_2^{(m)} \\ \vdots & \vdots & \ddots & \vdots \\ v_n^{(1)} & v_n^{(2)} & \cdots & v_n^{(m)} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{1.4}$$

has solutions other than $c_1 = c_2 = \ldots = c_m = 0$.

(ii) Show that $v^{(1)} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^\top$, $v^{(2)} = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^\top$, $v^{(3)} = \begin{bmatrix} -2 & 0 & 7 \end{bmatrix}^\top$ are linearly dependent.
Sol.: Solving the system (1.4) is an approach that always works, but here, noticing that $v^{(3)}$ is a linear combination of the other two,

$$v^{(3)} = -2v^{(1)} + 7v^{(2)} \,,$$

leads more quickly to the conclusion that the set of vectors is linearly dependent.

Application (Leslie matrices)**.** We now apply matrices to the study of population dynamics. In addition to providing an application in the natural sciences, it will also demonstrate the usefulness of eigenvectors.

A biologist has been monitoring a certain population of birds over a number of years. Each bird is either a hatchling or an adult, and vectors are used to describe the state of the population. For example, a population of 25 hatchlings and 89 adults is denoted $p = \begin{bmatrix} 25 & 89 \end{bmatrix}^\top$. The biologist has found that

- the survival rate of the hatchlings is 35%, and hatchlings that survive their first year become adults;

- the reproduction rate of the adults is 26%, meaning that 100 adults contribute an average of 26 hatchlings to next year's population;

- hatchlings do not reproduce; and

- the survival rate of adults is 67%.

It would now be useful to find a 2×2 matrix L that describes the growth/decline of the population from one year to the next:

$$p_{\text{next year}} = L \, p_{\text{this year}} \, .$$

The properties above show what happens to 100 hatchlings,

$$\begin{bmatrix} 100 & 0 \end{bmatrix}^\top \mapsto \begin{bmatrix} 0 & 35 \end{bmatrix}^\top ,$$

and what happens to 100 adults,

$$\begin{bmatrix} 0 & 100 \end{bmatrix}^\top \mapsto \begin{bmatrix} 26 & 67 \end{bmatrix}^\top .$$

The matrix that carries out these mappings is

$$L = \begin{bmatrix} 0 & 0.26 \\ 0.35 & 0.67 \end{bmatrix} ,$$

which is called the *Leslie matrix* and maps a population vector to next year's state.

What we have achieved so far is merely a systematic way of writing out the biologist's observations – no progress that adds value to his research has been made yet. This changes once an analysis of eigenvectors is included into our study: The eigenvalues of L are

$$\lambda_1 = -0.116 \, , \quad \lambda_2 = 0.786 \, .$$

Denote the corresponding eigenvectors v_1, v_2 as usual and let the current state of the population be p_0. Writing p_0 as a linear combination of the two eigenvectors,

$$p_0 = c_1 v_1 + c_2 v_2$$

(the coefficients can be found by solving a linear system), we find for the population k years from now:

$$
\begin{aligned}
p_k = L^k p_0 = L^{k-1} L p_0 &= L^{k-1} \left[L(c_1 v_1 + c_2 v_2) \right] \\
&= L^{k-1} \left[c_1 L(v_1) + c_2 L(v_2) \right] = L^{k-1} \left[c_1 \lambda_1 v_1 + c_2 \lambda_2 v_2 \right] \\
&= L^{k-2} \left[c_1 \lambda_1 L(v_1) + c_2 \lambda_2 L(v_2) \right] = \ldots = c_1 \lambda_1^k v_1 + c_2 \lambda_2^k v_2 \, .
\end{aligned}
$$

Note that

$$\lambda^k \to 0 \quad \text{for } k \to \infty$$

for both eigenvalues, since both have absolute value less than 1. This means, unfortunately, that the birds will go extinct unless preservation measures are put in place. Besides identifying a need for preservation, the eigenvalue analysis of Leslie matrices can also indicate which measures would be most effective.

Exercise 1.33. (i) Verify that $v_1 = \begin{bmatrix} -17 & 2 & 34 \end{bmatrix}^\top$ is an eigenvector with eigenvalue $\lambda_1 = 13$ of

$$A = \begin{bmatrix} 1 & 0 & -6 \\ 0 & -4 & 1 \\ -2 & 0 & 12 \end{bmatrix}.$$

Further verify that $v_2 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^\top$ and $v_3 = \begin{bmatrix} 24 & 1 & 4 \end{bmatrix}^\top$ are eigenvectors as well. What are the corresponding eigenvalues, λ_2 and λ_3?

(ii) For the matrices

$$\begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix}, \quad \begin{bmatrix} 9 & -1 \\ 3 & 5 \end{bmatrix},$$

find all eigenvalues as well as the corresponding eigenvectors[17].

(iii) For the matrices

$$\begin{bmatrix} 3 & 2 & 4 \\ 2 & 0 & 2 \\ 4 & 2 & 3 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix},$$

find all eigenvalues as well as the corresponding eigenvectors[18].

(iv) Find all eigenvalues of

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

Can you make a general statement for upper triangular square matrices of any size? Hence argue that two matrices with the same set of eigenvalues need not be the same.

(v) For each of the following sets of vectors, decide whether it is linearly independent or not[19].

$$\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}, \quad \left\{ \begin{bmatrix} 2 \\ -3 \end{bmatrix}, \begin{bmatrix} -6 \\ 9 \end{bmatrix} \right\}, \quad \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} -4 \\ 5 \end{bmatrix} \right\}.$$

(vi) Write $w = \begin{bmatrix} 2 & -3 & 9 & 1 \end{bmatrix}^\top$ as a linear combination of[20]

$$v_1 = \begin{bmatrix} 1 \\ 3 \\ 0 \\ 5 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 4 \end{bmatrix}, \quad v_3 = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 3 \end{bmatrix}, \quad v_4 = \begin{bmatrix} 1 \\ -3 \\ 6 \\ -1 \end{bmatrix}.$$

(vii) Find $\alpha \in \mathbb{R}$ such that $\lambda = 0$ is an eigenvalue of

$$\begin{bmatrix} 1 & -1 & 1 \\ 1 & 2 & \alpha \\ 3 & 0 & 6 \end{bmatrix}.$$

For this value of α, find the other two eigenvalues of the matrix[21].

35

(viii) Show that the matrix

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

has no *real* eigenvalues. There are complex solutions to the eigenvalue equation though – what are they?

(ix) Let A be a square matrix such that there exists $k \in \mathbb{N}$ such that

$$A^k = 0 \,,$$

where the zero on the right stands for the zero matrix, not the number zero. Show that $\lambda = 0$ is an eigenvalue of A and that it is the only eigenvalue.

(x) Consider a set S of n vectors of size n,

$$S = \{v^{(1)}, v^{(2)}, \ldots, v^{(n)}\} \,,$$

and show that

$$S \text{ linearly independent} \quad \Longleftrightarrow \quad \det V \neq 0 \,,$$

where V is the matrix whose columns are the vectors of S, i.e., the matrix in (1.4) of Example 1.32 (with $m = n$)[22].

(xi) Let A be a $n \times n$ matrix, v an n-vector, and $k \in \mathbb{N}$ such that

$$\begin{cases} A^k v & \neq & 0 \\ A^{k+1} v & = & 0 \,. \end{cases}$$

Show that the set
$$S = \{v, Av, v, A^2 v, \ldots, A^k v\}$$
is linearly independent[23].

(xii) Consider a bird population that has the matrix L from Example 1.29 as its Leslie matrix. Suppose the current state is $p_0 = \begin{bmatrix} 200 & 1100 \end{bmatrix}^\top$. Find the long-term behaviour of the population[24].

1.4 Inverse Matrices

Definition 1.34. If for a square matrix A there exists a matrix B of the same size with
$$AB = BA = I \,,$$
then we say that A is *invertible* and B is called its *inverse*, written $B = A^{-1}$.

Remark 1.35. Not all square matrices are invertible.

Theorem 1.36.
$$A \text{ invertible} \quad \Longleftrightarrow \quad \det A \neq 0$$

Example 1.37. (i) In order to invert the matrix

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix},$$

define a 2×2 matrix with general entries $a, b, c, d \in \mathbb{R}$ and check whether it is possible to obtain the identity matrix as a product:

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a+2c & b+2d \\ 2a+4c & 2b+4d \end{bmatrix} \overset{!}{=} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

where the exclamation mark means that we want the matrix containing combinations of a, b, c, d to be equal to the identity matrix on the right. However, no choice of a, b, c, d achieves this – for example, setting the bottom left entry, $2a + 4c$, equal to zero implies that the upper left entry is zero as well.

That A is not invertible could have been found more easily by looking at its determinant.

(ii) The determinant of

$$B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

is $\det B = 1 \cdot 4 - 3 \cdot 2 = -2 \neq 0$, and therefore B is invertible. Its inverse is

$$B^{-1} = \begin{bmatrix} -2 & 1 \\ 3/2 & -1/2 \end{bmatrix}.$$

One can find the inverse with the naive approach in (i) or with the the more systematic approaches below. Either way, you can verify that the stated B^{-1} is indeed the inverse of B by finding their matrix products.

(iii) A 1×1 matrix is just a number, $C = \begin{bmatrix} c_{11} \end{bmatrix}$. The determinant of that trivial matrix is its entry, $\det C = c_{11}$. The above theorem states that this number has an inverse if and only if it is different from zero. You knew that already – for $x = 0$ there is no $y \in \mathbb{R}$ with $xy = 1$, but for all other x, there is.

Remark 1.38. We now derive a systematic way to find inverses. Let A be an invertible $n \times n$ matrix. To find A^{-1}, denote the vectors $\begin{bmatrix} 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{bmatrix}^\top$ of size n, where the entry 1 is in the j-th place, by e_j. Now convince yourself of the following observations.

(i) For any $n \times n$ matrix B, the j-th column of B is Be_j.

(ii) If A maps v to w, then A^{-1} maps w to v :

$$Av = w \quad \Longrightarrow \quad A^{-1}w = A^{-1}Av = Iv = v.$$

(iii) Solving the augmented matrix $[A \mid w]$ gives the vector v that is mapped to w by $A : Av = w$.

(iv) Now combine those three points to find the columns of A^{-1} :

$$\text{denote the } j\text{-th column of } A^{-1} \text{ by } c_j \xrightarrow{\text{(i)}} c_j = A^{-1}e_j$$

$$\xrightarrow{\text{(ii)}} Ac_j = e_j$$

$$\xrightarrow{\text{(iii)}} c_j \text{ is the solution of } [A \mid e_j] \, .$$

(v) We therefore need to solve $[A \mid e_j]$ for all $j \in \{1, 2, 3, \ldots, n\}$ to obtain the columns c_j of A^{-1}. This can be done with a single Gaussian elimination by augmenting all n vectors at once:

- Augmenting $e_1, e_2, e_3, \ldots, e_n$ side by side amounts to augmenting the identity matrix.

- The RREF of A is the identity matrix, since A has a nonzero determinant and therefore rank n (cf. Lemma 1.27). That is, A can be transformed into I.

- If the left-hand side of an augmented matrix is I, then the solution can be read off directly: b is the solution of $[I \mid b]$ (as in Example 1.20 (iii)).

We have derived the following algorithm for finding inverse matrices.

(vi) To find the inverse of the invertible matrix A, augment the identity matrix I and bring A in reduced row echelon form using elementary row operations. The matrix on the right is then the inverse of A :

$$[A \mid I] \quad \xrightsquigarrow{\text{elementary row operations}} \quad [I \mid A^{-1}] \, .$$

Example 1.39. Check whether

$$A = \begin{bmatrix} -5 & 0 & 7 \\ 6 & 2 & -6 \\ -4 & 0 & 6 \end{bmatrix}$$

is invertible and find its inverse in case it is.

Sol.: Developing along the second column, Laplace's method gives

$$\det A = (-1)^{2+1} \cdot 0 \cdot |\ldots| + (-1)^{2+2} \cdot 2 \cdot |\ldots| + (-1)^{2+3} \cdot 0 \cdot |\ldots|$$

$$= 2 \cdot \begin{vmatrix} -5 & 7 \\ -4 & 6 \end{vmatrix} = 2 \cdot (-30 + 28) = -4 \neq 0 \, .$$

Hence A is invertible. The method from the previous remark,

$$\left[\begin{array}{ccc|ccc} -5 & 0 & 7 & 1 & 0 & 0 \\ 6 & 2 & -6 & 0 & 1 & 0 \\ -4 & 0 & 6 & 0 & 0 & 1 \end{array}\right] \overset{\text{(i)}}{\rightsquigarrow} \left[\begin{array}{ccc|ccc} 120 & 0 & -168 & -24 & 0 & 0 \\ -120 & -40 & 120 & 0 & -20 & 0 \\ -120 & 0 & 180 & 0 & 0 & 30 \end{array}\right]$$

$$\overset{\text{(ii)}}{\rightsquigarrow} \left[\begin{array}{ccc|ccc} 120 & 0 & -168 & -24 & 0 & 0 \\ 0 & -40 & -48 & -24 & -20 & 0 \\ 0 & 0 & 12 & -24 & 0 & 30 \end{array}\right]$$

$$\overset{\text{(iii)}}{\rightsquigarrow} \left[\begin{array}{ccc|ccc} 120 & 0 & 0 & -15\cdot 24 & 0 & 14\cdot 30 \\ 0 & -40 & 0 & -5\cdot 24 & -20 & 4\cdot 30 \\ 0 & 0 & 12 & -24 & 0 & 30 \end{array}\right]$$

$$\overset{\text{(iv)}}{\rightsquigarrow} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -3 & 0 & 7/2 \\ 0 & 1 & 0 & 3 & 1/2 & -3 \\ 0 & 0 & 1 & -2 & 0 & 5/2 \end{array}\right],$$

leads to

$$A^{-1} = \begin{bmatrix} -3 & 0 & 7/2 \\ 3 & 1/2 & -3 \\ -2 & 0 & 5/2 \end{bmatrix}.$$

The Gaussian elimination can of course be carried out in a different way. Here, the following steps were used.

(i) Multiply $R1, R2, R3$ by $-24, -20, 30$, respectively, to prepare the elimination of the bottom two entries of the first column.

(ii) Add $R1$ to both $R2$ and $R3$ to carry out that elimination.

(iii) In general, the middle entry of the third row would have to be eliminated next, but here it is already zero. That is, we already have REF. Now use the entry 12 to eliminate the entries -168 and -48.

(iv) Again, the middle entry of the first row would have to be eliminated next, but here it is already zero. We have the left-hand side in diagonal form, and it remains to divide the three rows by $120, -40, 12$ to obtain the identity matrix.

Remark 1.40. As a first simple example to do on your own, you are encouraged to find B^{-1} from Remark 1.37 (ii) with that method. Once you are comfortable with that computation, you may use the formula for inverses of 2×2 matrices:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad\Longrightarrow\quad A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

You can verify this formula by multiplication. Of course, one can apply it only when the given matrix A is invertible – what happens if you try use it for a matrix that is not invertible (e.g. for the matrix in Example 1.37 (i))?

Application (Approximate solutions). We now derive a general technique as an application of the theory covered so far: finding approximate solutions of linear systems that do not have solutions. This is of general importance in many different areas of mathematics and other STEM disciplines.

Let A be a $m \times n$ matrix and define the *range* of A as the set of vectors of $\mathbb{R}^m$ that A maps to,

$$\text{range}\, A = \{w \in \mathbb{R}^m \mid \text{there exists } v \in \mathbb{R}^n \text{ such that } Av = w\}\,.$$

The connection to solving linear systems is

$$Av = b \text{ has a solution} \quad \Longleftrightarrow \quad b \in \text{range}\, A\,.$$

Now suppose that $Av = b$ that does not have a solution. An idea for finding an approximate solution is to find an element $\widetilde{b}$ of the range of A that is close to b. Since $\widetilde{b} \in \text{range}\, A$, the system $A\widetilde{v} = \widetilde{b}$ is solvable, and its solution can then be considered an approximate solution of $Av = b$. The idea for obtaining $\widetilde{b}$ is to project the vector b onto the range of A. In Section 1.1, we have seen how to project a point onto the line spanned by a vector $v = \begin{bmatrix} a & b \end{bmatrix}^\top$, and that is in fact the same as projecting onto the range of the matrix

$$A = \begin{bmatrix} 0 & a \\ 0 & b \end{bmatrix}.$$

This observation suggests to generalise the formula

$$P_L = \frac{v\, v^\top}{v^\top v} \tag{1.5}$$

from the application in Section 1.1 to matrices. The notation $\frac{1}{x} = x^{-1}$ for real numbers suggests to replace the fraction in (1.5) with a matrix product that contains an inverse. The product in the numerator, $AA^\top$, is of the form $m \times n \cdot n \times m = m \times m$. The denominator, $A^\top A$, has size $n \times m \cdot m \times n = n \times n$, and its inverse is $n \times n$ as well. Therefore the matrices $AA^\top$ and $A^\top A$ can not multiply in either order, as the sizes do not match. However, it is possible to insert the matrix coming from the denominator between the two factors of the numerator:

$$P_A = A\, (A^\top A)^{-1} A^\top.$$

We have derived the following approach to finding approximate solutions $\widetilde{v}$: if $Av = b$ does not have a solution, then

$$Av = b \quad \rightsquigarrow \quad A\widetilde{v} = \widetilde{b} = P_A\, b = A\, (A^\top A)^{-1} A^\top b\,,$$

of which

$$\widetilde{v} = (A^\top A)^{-1} A^\top b \tag{1.6}$$

is a solution. Note that we have not proven the validity of this method – we have merely identified the only consistent way to generalise the projection formula from Section 1.1 to matrices. It will be shown in the next chapter that (1.6) is correct.

Application (Linear regression). Next we derive an important technique that is used for statistical modelling and machine learning. Suppose data of the following form has been collected.

$$X = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_n^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \dots & x_n^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(m)} & x_2^{(m)} & \dots & x_n^{(m)} \end{bmatrix}, \quad y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix}.$$

In the matrix X, which is called the *design matrix*, each row stands for an individual or event for which data has been collected, and the elements of that row,

$$\begin{bmatrix} x_1^{(k)} & x_2^{(k)} & \dots & x_n^{(k)} \end{bmatrix},$$

describe different attributes. There further is an additional observed value, $y^{(k)}$, for that individual/event, which we suspect to be related to the corresponding attributes $x_j^{(k)}$ $(1 \leq j \leq n)$. If we can find the relation $X \leftrightarrow y$, it can be used to predict future y values. For example, the data X above could be information extracted from last year's annual reports of m companies. The attributes could be company size, net income, total amount of debt, etc., and the values y the change in price of the companies' shares on the stock market.

The most simple relationship between X and y is

$$y^{(k)} = w_1 \cdot x_1^{(k)} + w_2 \cdot x_2^{(k)} + \dots + w_n \cdot x_n^{(k)}.$$

While this equation is certainly solvable for an individual row, it is very unlikely that there is a single vector w that works for all rows – for the model to be statistically sound, the amount of data, m, needs to be larger than n. Then there are more equations than variables, and the existence of a solution is unlikely. Fortunately, we know how to find the approximate solution of $Xw = y$:

$$\widetilde{w} = (X^\top X)^{-1} X^\top y.$$

Returning to the financial application mentioned above: Once you have found the relation w between last year's reported data and the subsequent performance on the stock market, and once this year's annual reports have been published, you could produce the matrix X for a selection of listed companies and then compute the corresponding y-values as

$$y^{\text{this year}} = X^{\text{this year}} w^{\text{last year}}.$$

However, such "predictions" can be wholly inappropriate; there is a large number of caveats, e.g.: perhaps this approach is not able to capture the true relationship $X \leftrightarrow y$, perhaps that relationship depends on other factors as well, or perhaps there is no true relationship and the model is built on coincidence. One therefore has to be very careful when interpreting the predictions given by machine learning algorithms. Statisticians and data scientists always take a range of measures to test and validate their model before they deploy their findings.

Exercise 1.41. (i) Find the inverses of

$$\begin{bmatrix} 2 & -2 \\ 0 & 8 \end{bmatrix}, \begin{bmatrix} 9 & -4 \\ 7 & 8 \end{bmatrix}.$$

Next, pick a vector, compute the vector that it gets mapped to by the matrix, and then check that the inverse undoes that transformation.

(ii) Find the inverses of

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 6 & 7 \end{bmatrix}, \begin{bmatrix} -5/18 & 1/18 & 7/18 \\ 1/18 & 7/18 & -5/18 \\ 7/18 & -5/18 & 1/18 \end{bmatrix}.$$

(iii) Find the inverse of[25]

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

(iv) Consider the system

$$\begin{cases} x & + & 5y & = & 4 \\ -2x & + & y & = & 3, \end{cases}$$

and solve it in two ways: First, as in Section 1.2. Secondly, find the inverse of the matrix A of coefficients and obtain the solution $v = \begin{bmatrix} x & y \end{bmatrix}^{-1}$ as

$$v = A^{-1} \begin{bmatrix} 4 \\ 3 \end{bmatrix}.$$

Which approach do you find faster?

(v) Convince yourself that Theorem 1.36 is true[26].

(vi) Let λ be an eigenvalue of an invertible matrix A. Show that $\lambda \neq 0$ and that A^{-1} has λ^{-1} as an eigenvalue.

(vii) Given that the set of vectors $\{v_1, v_2, v_3\}$ in $\mathbb{R}^3$ is linearly independent, show that $\{w_1, w_2, w_3\}$, where

$$\begin{cases} w_1 = v_1 + v_2 \\ w_2 = 3v_2 + 2v_3 \\ w_3 = v_1 - 2v_2 + v_3, \end{cases}$$

is linearly independent as well[27].

(viii) For the bird population from the last problem of exercise set 1.33, find the state of the population in the previous year.

(ix) Suppose you are taking a module which is assessed via a mid-term exam, coursework assignments, and a final exam. You know three of last year's students, and they had the following marks.

$$\begin{cases} S_1 = (55, 43, 61) \\ S_2 = (72, 60, 82) \\ S_3 = (64, 63, 68) \,, \end{cases}$$

where the marks are stated in the format "(mid-term exam, coursework, final exam)". Given that you have scored 73 and 74 on the mid-term exam and on the coursework assignment, predict your score on the final exam using linear regression. However, this "prediction" is not an appropriate use of linear regression! Why not[28]?

Chapter 2

Functions of Several Variables

We have seen that matrices make real-life problems accessible to mathematical analysis. The single-variable functions you know from school are another such link. However, real-life quantities – for example, the expected profit of a business strategy – rarely depend on one variable only, and it is therefore important to also study *functions of several variables*. The basic ideas for doing that are the same as for the single-variable theory, but we will also encounter new concepts, new difficulties, and new potential in this chapter.

Application (Stabilisation of mechanical processes)**.** Alice is practising archery. She can control the velocity of the arrow and the angle at which she shoots. Note that there are many different ways to hit the target: she could aim a very strong shot directly at the bullseye or she could shoot with less force and aim higher to compensate for the reduced velocity of the arrow. Alice's maths skills are excellent – for any given angle α, she can compute the velocity v required to hit the bullseye. Of all such choices (α, v), Alice wants to use the most stable one. That is, the configuration (α^*, v^*) that hits the bullseye and is the least sensitive to trembling, misjudging of the angle or the velocity, etc., needs to be found.

For our modelling, we assume that the tip of the arrow is exactly level with the bullseye when it is loosed, cf. the sketch above. The target is 30 metres away. Then the height of the arrow when it hits the target is

$$h(\alpha, v) = -\frac{4414.5}{\cos^2(\alpha)\, v^2} + 30 \tan \alpha \,.$$

Here, $h = 0$ corresponds to hitting the bullseye. There are different possibilities for defining a measurement of "instability", and Alice decides to use

$$f(\alpha, v) = h_\alpha^2 + 3h_v^2$$

since she finds controlling the velocity more difficult than controlling the angle. The symbols h_α and h_v are the partial derivatives of h and will be defined in the next section. After familiarising yourself with that concept, convince yourself that f as defined above can be considered a measure of instability of[29] h. In this chapter, we will learn how to minimise functions like f (instability) under constraints like $h = 0$ (hitting the bullseye). The optimal angle and velocity for Alice's target practice are

$$\alpha^* = 44.70^\circ\,, \quad v^* = 17.16\tfrac{m}{s}\,.$$

The fact that those numbers do not agree with how archers usually shoot is due to our assumptions and our modelling: wind was neglected, it was assumed that initially the tip of the arrow and the bullseye are exactly level, that the archer can do the maths and control the angle and velocity exactly, etc. – in the absence of these assumptions, a very strong shot almost directly aimed at the bullseye is the most straightforward option. However, understanding the above methods is a first step towards more complex real-life applications such as the design of mechanical machines.

2.1 Multivariate Functions and Partial Derivatives

Definition 2.1 (Functions of Several Variables). A *function of two variables* is a rule f that assigns to each pair (x, y) in a set $D \subseteq \mathbb{R}^2$ a unique real number $f(x, y)$. The set D is called the *domain* of f and also denoted $D(f)$. The *range* of f is the set of values it maps to,

$$R(f) = \{z \in \mathbb{R} \mid z = f(x, y) \text{ for some } (x, y) \in D\}\,.$$

Similarly, a *function of n variables* is a rule f that assigns a unique real number $f(x_1, x_2, \ldots, x_n)$ to each $(x_1, x_2, \ldots, x_n) \in D \subseteq \mathbb{R}^n$.

Remark 2.2. (i) Writing $z = f(x, y)$, we call x and y the independent variables, and z the dependent variable. Functions of several variables are also called *multivariate functions.*

(ii) Unless specified otherwise, the domain is always the largest set of points (x, y) at which the expression that defines f can be evaluated.

(iii) Functions of two variables can be visualised as surfaces/graphs in $\mathbb{R}^3$: The domain is a subset of the xy-plane and for (x_0, y_0) in that domain, we mark a point of height $f(x_0, y_0)$ over/under (x_0, y_0). That is, the point of the graph with (x, y)-coordinates (x_0, y_0) will have z-coordinate $z_0 = f(x_0, y_0)$.

(iv) Restriction to lines in the domain is an important technique for working with multivariate functions, cf. part (iv) of the following set of examples.

Example 2.3. (i) Draw the graph of $f(x, y) = x^2 + y^2$.
    Sol.:

Graph of f (paraboloid) with curves from parts (iv) and (v) (3D):
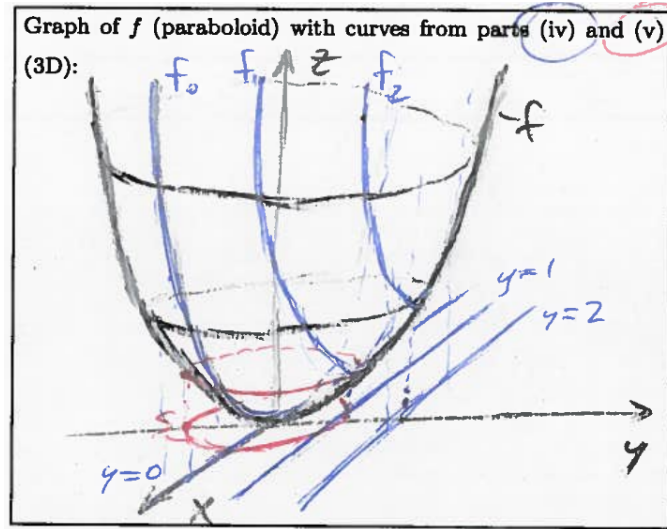
(ii) Find the domain and range of $g(x, y) = 1 + \sqrt{y - x^2}$.

Sol.: The function g is defined on the set

$$D(g) = \left\{ (x, y) \in \mathbb{R}^2 \mid y \geq x^2 \right\} = \left\{ y \geq x^2 \right\},$$

since we need $y - x^2 \geq 0$ to be able to evaluate the square root. On this domain, the argument of the square root takes all non-negative real values $t \in [0, \infty)$, producing non-negative real numbers $\sqrt{t} \in [0, \infty)$. Therefore, the range is $R(g) = [1, \infty)$.

(iii) Find the domain and range of $h(x_1, x_2, x_3) = \sin(x_1 x_2 + x_3)$.

Sol.: The domain is $D = \mathbb{R}^3$ since h can be evaluated at any (x_1, x_2, x_3). The range of h is $R = [-1, 1]$.

(iv) Draw the graphs of the following one-variable functions

$$f_0(x) = x^2,$$
$$f_1(x) = x^2 + 1,$$
$$f_2(x) = x^2 + 4.$$

Where do these curves appear in the graph in (i)?

Sol.: The functions f_0, f_1, f_2 are restrictions of $f(x, y)$ in (i) to the lines $y = 0, y = 1, y = 2$. This can also be expressed as

$$f_0(x) = f(x, 0) = x^2 + 0^2,$$
$$f_1(x) = f(x, 1) = x^2 + 1^2,$$
$$f_2(x) = f(x, 2) = x^2 + 2^2.$$

Graphs of f_0, f_1, f_2 (compare to graph (i); 2D):
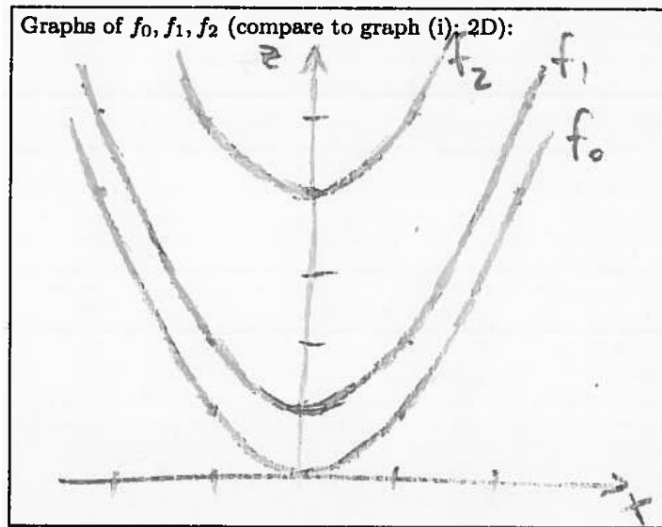
(v) The domain of the function $f(x,y) = x^2 + y^2$ from (i) is $D(f) = \mathbb{R}^2$. Describe the subset S of points in D with $f(x,y) = 1$.

Sol.: The equation

$$x^2 + y^2 = 1$$

defines a circle in the xy-plane:

Curve S in xy-plane (compare to graph (i); 2D):
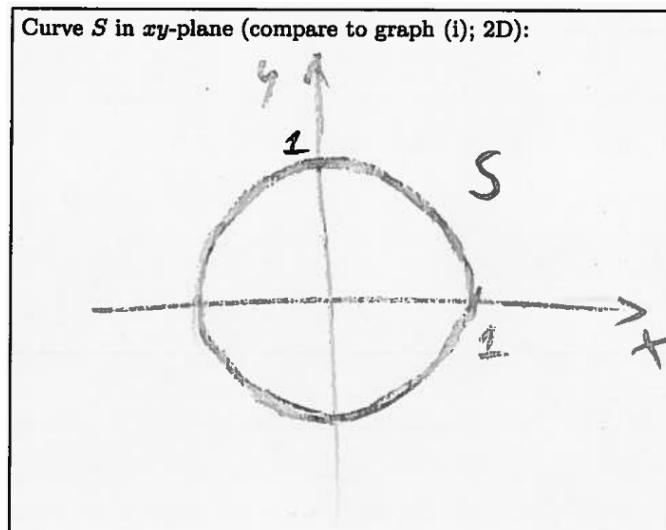
Definition 2.4 (Partial Derivatives)**.** For a function f of two variables, the *partial derivatives* with respect to x, respectively y, are the functions

$$\frac{\partial f}{\partial x}(x,y) = \lim_{h \to 0} \frac{f(x+h, y) - f(x,y)}{h} \ ,$$

$$\frac{\partial f}{\partial y}(x,y) = \lim_{h \to 0} \frac{f(x, y+h) - f(x,y)}{h} \ .$$

We also write f_x for $\dfrac{\partial f}{\partial x}$ and f_y for $\dfrac{\partial f}{\partial y}$. Similarly for functions of more than two variables:

$$\frac{\partial f}{\partial x_j}(x_1, x_2, \ldots, x_n) = \lim_{h \to 0} \frac{f(x_1, x_2, \ldots, x_{j-1}, x_j + h, x_{j+1}, \ldots, x_n) - f(x_1, x_2, \ldots, x_n)}{h} \ .$$

Example 2.5. Find f_x for the function $f(x, y) = 5x^2y^7$.
Sol.:

$$\frac{\partial f}{\partial x}(x, y) = \lim_{h \to 0} \frac{f(x + h, y) - f(x, y)}{h}$$

$$= \lim_{h \to 0} \frac{5(x + h)^2 y^7 - 5x^2 y^7}{h}$$

$$= 5y^7 \lim_{h \to 0} \frac{(x + h)^2 - x^2}{h}$$

$$= 5y^7 \lim_{h \to 0} \frac{x^2 + 2hx + h^2 - x^2}{h}$$

$$= 5y^7 \lim_{h \to 0}(2x + h) = 5(2x)y^7 = 10xy^7 \; .$$

The factor $2x$ in the last line is just the ordinary single-variable derivative of the factor x^2 of f. That is, the operator $\partial/\partial x$ differentiates the x-dependent part of f in the usual way and treats terms that do not depend on x as constants:

$$\frac{\partial f}{\partial x}(x, y) = \frac{\partial}{\partial x}\left(5x^2 y^7\right) = 5\frac{\partial}{\partial x}\left(x^2\right)y^7 = 5(2x)y^7 = 10xy^7 \; .$$

Example 2.6.   (i) For $f = x^3 + y^5 - 2$, find f_x and f_y.
   Sol.:

$$f_x = \frac{\partial}{\partial x}\left(x^3 + y^5 - 2\right) = 3x^2 + 0 + 0 = 3x^2 \; ,$$

$$f_y = \frac{\partial}{\partial y}\left(x^3 + y^5 - 2\right) = 0 + 5y^4 + 0 = 5y^4 \; .$$

(ii) For $f = x^3 + x^2y^3 - 2y^2$, find f_x and f_y.
   Sol.:

$$f_x = \frac{\partial}{\partial x}\left(x^3\right) + \frac{\partial}{\partial x}\left(x^2\right)y^3 - 0 = 3x^2 + 2xy^3 \; ,$$

$$f_y = 0 + x^2\frac{\partial}{\partial y}\left(y^3\right) - 2\frac{\partial}{\partial y}\left(y^2\right) = 3x^2y^2 - 4y \; .$$

Remark 2.7.   (i) For a geometric interpretation, we consider the graph of a function f of two variables, i.e. $z = f(x, y)$. Then $\partial z/\partial x$ is the slope on the surface "when moving in the x direction". That is, x is varied, y is held constant, and we take note of the change of function values. Similarly, $\partial z/\partial y$ is the slope in the y direction.

As an intuitive example, consider a hiker climbing a mountain of the shape $z = 4 - (x^2 + y^2)$.

"Steep mountain" with hiker's position and paths in x and y direction (3D):

slope 0 in yz plane

slope -2 in xz- plane
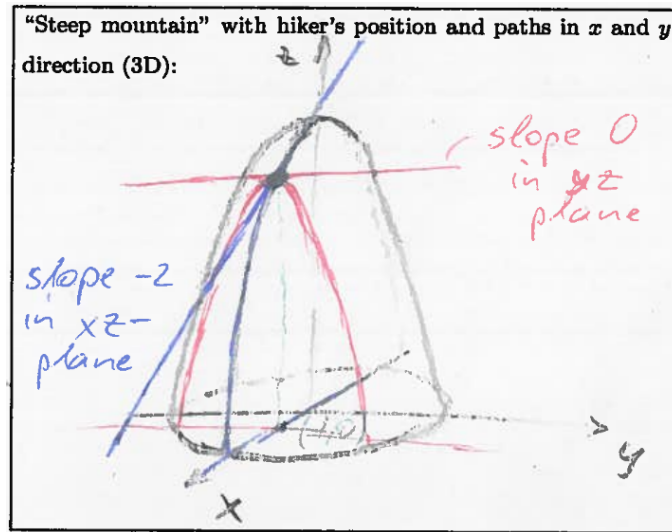
Suppose she checks her GPS and finds that her coordinates are $(x, y) = (1, 0)$. The slopes at this point in both directions are

$$z_x(1,0) = \frac{\partial}{\partial x} \left(4 - (x^2 + y^2) \right)_{|(x,y)=(1,0)} = (-2x)_{|(x,y)=(1,0)} = -2 \,,$$

$$z_y(1,0) = \frac{\partial}{\partial y} \left(4 - (x^2 + y^2) \right)_{|(x,y)=(1,0)} = (-2y)_{|(x,y)=(1,0)} = 0 \,.$$

That is, there is no slope in the y direction. The slope in the x direction is negative, because the height decreases as x increases (from where she stands, increasing x would be moving away from the peak).

(ii) In order to be able to work with a larger set of functions – not just sums of products and powers of variables – we extend the single-variable differentiation rules to the multivariate setting: The x-derivative of a product of functions is

$$(fg)_x = f_x g + f g_x \,,$$

that is, the product rule extends to partial derivatives without any changes. Similarly for $\partial/\partial y$. The quotient rule for a y-derivative is

$$\left(\frac{f}{g} \right)_y = \frac{f_y g - f g_y}{g^2} \,.$$

(iii) In order to find partial derivatives of functions like

$$f(x, y) = \sin(x + y^2) \,,$$

we also need to combine the chain rule for single-variable functions with partial derivatives. Note that the outer function, $h(t) = \sin(t)$, is a single-variable function. Let c be a constant and review the following applications of the single-variable chain rule.

$$\frac{\mathrm{d}}{\mathrm{d}t} \left(\sin(t + c) \right) = \cos(t + c) \frac{\mathrm{d}}{\mathrm{d}t} (t + c) = \cos(t + c) \,,$$

$$\frac{\mathrm{d}}{\mathrm{d}t} \left(\sin(c + t^2) \right) = \cos(c + t^2) \frac{\mathrm{d}}{\mathrm{d}t} (c + t^2) = 2t \cos(c + t^2) \,. \tag{2.1}$$

49

Now, the term y^2 looks like a constant to the operator $\partial/\partial x$. Similarly for x and $\partial/\partial y$. We can thus apply the single-variable computations (2.1) to the two-variable function f (in the first case, just replace y^2 with c):

$$f_x = \frac{\partial}{\partial x}\left(\sin(x + y^2)\right) = \cos(x + y^2)\frac{\partial}{\partial x}(x + y^2) = \cos(x + y^2),$$

$$f_y = \frac{\partial}{\partial y}\left(\sin(x + y^2)\right) = \cos(x + y^2)\frac{\partial}{\partial y}(x + y^2) = 2y\cos(x + y^2).$$

This can be stated as

$$\frac{\partial}{\partial x}h(g(x, y)) = h'(g(x, y)) \cdot g_x(x, y),$$

where h is a single-variable function. Similarly for $\partial/\partial y$. Note that the $'$ is not ambiguous – h has only one variable, and h' is the derivative with respect to (abbreviated: "w.r.t") that variable.

Example 2.8. (i) The y-derivative of $f(x, y) = e^{x^2 + y^3}$ is found easily,

$$\frac{\partial}{\partial y}\left(e^{x^2 + y^3}\right) = e^{x^2 + y^3} \cdot \frac{\partial}{\partial y}(x^2 + y^3) = 3y^2\, e^{x^2 + y^3},$$

and we can verify this result with the following alternative approach.

$$\frac{\partial}{\partial y}\left(e^{x^2 + y^3}\right) = \frac{\partial}{\partial y}\left(e^{x^2}e^{y^3}\right) = e^{x^2}\frac{\partial}{\partial y}\left(e^{y^3}\right)$$

$$= e^{x^2}\frac{d}{dy}\left(e^{y^3}\right) = e^{x^2}e^{y^3}3y^2 = 3y^2\, e^{x^2 + y^3}.$$

Here, the "∂" was changed to a "d" to emphasise that in this approach, an ordinary derivative was taken, i.e., a derivative of a single-variable expression.

(ii)

$$\frac{\partial}{\partial x}\left(\frac{\sin x \cos y}{x + y}\right) = \frac{\frac{\partial}{\partial x}\left(\sin x \cos y\right)(x + y) - (\sin x \cos y)\frac{\partial}{\partial x}\left(x + y\right)}{(x + y)^2}$$

$$= \frac{\cos y\left[\cos x \cdot (x + y) - \sin x\right]}{(x + y)^2}$$

Definition 2.9 (Higher-Order Partial Derivatives)**.** The *second-order partial derivatives* with respect to x and y of a function f of two variables are written

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right),$$

$$\frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right),$$

and can also be denoted f_{xx} and f_{yy}. We further have *mixed* second-order derivatives:

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right),$$
$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right).$$

Similarly for functions of more than two variables and for higher-order derivatives; e.g.,

$$f_{xyyz} = \frac{\partial^4 f}{\partial x \partial y \partial y \partial z} = \frac{\partial}{\partial x}\left(\frac{\partial}{\partial y}\left(\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial z}\right)\right)\right)$$

is a fourth-order derivative of a three-variable function $f = f(x, y, z)$.

Example 2.10. (i) For
$$f(x, y) = 2x^9 y^5,$$

find all partial derivatives of order up to 2.
Sol.:

$$f = 2x^9 y^5,$$
$$f_x = 2(9x^8)y^5 = 18x^8 y^5,$$
$$f_y = 2x^9(5y^4) = 10x^9 y^4,$$
$$f_{xx} = \frac{\partial}{\partial x}\left(18x^8 y^5\right) = 18(8x^7)y^5 = 144x^7 y^5,$$
$$f_{yy} = \frac{\partial}{\partial y}\left(10x^9 y^4\right) = 10x^9(4y^3) = 40x^9 y^3,$$
$$f_{xy} = \frac{\partial}{\partial x}\left(10x^9 y^4\right) = 10(9x^8)y^4 = 90x^8 y^4,$$
$$f_{yx} = \frac{\partial}{\partial y}\left(18x^8 y^5\right) = 18x^8(5y^4) = 90x^8 y^4.$$

(ii) For
$$f(x, y) = x^2 + 3xy - y^2 + 7x - 4y + 2,$$

find *all* partial derivatives.
Sol.:

$$f = x^2 + 3xy - y^2 + 7x - 4y + 2,$$
$$f_x = 2x + 3y + 7,$$
$$f_y = 3x - 2y - 4,$$
$$f_{xx} = 2,$$
$$f_{yy} = -2,$$
$$f_{xy} = 3,$$
$$f_{yx} = 3.$$

Since all the second-order derivatives are constant, all derivatives of higher order are zero,
$$\frac{\partial^k f}{\partial \ldots} = 0 \qquad \text{for } k > 2.$$

(iii) For
$$f(x, y) = x \cos(xy^2) \,,$$

find all partial derivatives of order up to 2.

Sol.:

$$\begin{aligned}
f &= x \cos(xy^2) \,, \\
f_x &= \cos(xy^2) - xy^2 \sin(xy^2) \,, \\
f_y &= -2x^2 y \sin(xy^2) \,, \\
f_{xx} &= -2y^2 \sin(xy^2) - xy^4 \cos(xy^2) \,, \\
f_{yy} &= -2x^2 \sin(xy^2) - 4x^3 y^2 \cos(xy^2) \,, \\
f_{xy} &= -4xy \sin(xy^2) - 2x^2 y^3 \cos(xy^2) = f_{yx} \,.
\end{aligned}$$

Remark 2.11. In all of the previous examples, we had $f_{xy} = f_{yx}$. The following theorem states that this is not a coincidence – it is always true as long as certain technical conditions are satisfied. Those conditions are met for the functions considered in this book, hence we can always work with the conclusion of the following theorem. However, be aware that there are functions that do not meet its requirements.

Theorem 2.12 (Schwarz's Theorem or Clairaut's Theorem). Consider a function $f = f(x, y)$ and a point (a, b) in its domain. If the functions f_{xy} and f_{yx} are both continuous at (a, b), then
$$f_{xy}(a, b) = f_{yx}(a, b) \,.$$

Remark 2.13. Short remark on continuity for multivariate functions; or leave it -- depending whether 1D continuity is covered in chapter 1.

Exercise 2.14. (i) Restrict the function $f(x, y) = xy$ to the lines $x = 0$, $y = 0$, $y = x$, $y = -x$. Hence draw the graph of[30] f.

(ii) Find $f_x, f_y, f_{xx}, f_{xy}, f_{yy}, f_{yx}$ for the functions[31]

    (a) $f(x, y) = x^2 y + x^3 y^4 + 7y$;

    (b) $f(x, y) = y \, e^x$;

    (c) $f(x, y) = \cos x \ln y$.

(iii) Find $f_x, f_y, f_{xx}, f_{xy}, f_{yy}, f_{yx}$ for the functions[32]

    (a) $f(x, y) = \sin(xy + y^3)$;

    (b) $f(x, y) = \frac{x^2 y^2}{2} - x^2 y^2 \ln(xy)$;

    (c) $f(x, y) = \frac{x}{1 + x^2 + y^2}$.

(iv) Find the domain and the range of[33] $f(x, y) = \sqrt{2 \sin x \sin y}$.

(v) Let $D \subseteq \mathbb{R}^2$ be the open disk of radius 1, $D = \{x^2 + y^2 < 1\}$. Define a function that is defined for $(x, y) \in D$, and another function that is defined on the *complement* of D,

$$D^c = \mathbb{R}^2 \setminus D = \{x^2 + y^2 \geq 1\}$$

(read this as "all points outside of D")[34].

(vi) Verify that

(a) $y(t, x) = e^{-kn^2 t} \sin(nx)$ satisfies $y_t = k y_{xx}$;
(b) $r(x, y, z) = \sqrt{x^2 + y^2 + z^2}$ satisfies $r_{xx} + r_{yy} + r_{zz} = \frac{2}{r}$.

(vii) Find *all* partial derivatives of[35] $f(x, y) = e^{2x+3y}$.

(viii) Consider the function $f(x, y) = x^2(1 - x^2) - y^2$. Find all points for which both the x- and the y-derivative is equal to zero. That is, find (x_0, y_0) with[36] $f_x(x_0, y_0) = f_y(x_0, y_0) = 0$.

2.2 Chain Rule and Implicit Differentiation

Remark 2.15. *Curves* in the xy-plane can be parametrised by writing x and y as single-variable functions of a third variable, usually t (think of it as "time"):

$$(x, y) = (x(t), y(t)) .$$

We denote such curves γ,

$$\gamma : \mathbb{R} \to \mathbb{R}^2$$
$$t \mapsto \gamma(t) = (x(t), y(t)) .$$

A given function $f = f(x, y)$ can then be restricted to γ – this is carried out formally via a composition : define the single-variable function F as

$$F(t) = (f \circ \gamma)(t) = f(\gamma(t)) = f(x(t), y(t)) .$$

Finding the derivative of F is now an important and applicable task.

The distinction $f \leftrightarrow F$ is often not made to simplify notation. That is, the single-variable function of t may be called f as well. This reduces the number of function names needed to write out a computation, but it also carries potential for confusion since the symbol "f" would then used for two different mathematical objects.

Example 2.16. (i) Draw the curves

$$\gamma_1 : \quad (x(t), y(t)) = (t, t) ,$$
$$\gamma_2 : \quad (x(t), y(t)) = (3 \cos t, 3 \sin t) ,$$
$$\gamma_3 : \quad (x(t), y(t)) = (t, 1 - t^2) ,$$

and highlight the point corresponding to $t = 0$ in each of them.
Sol.:

The curves $\gamma_1, \gamma_2, \gamma_3$ (2D):
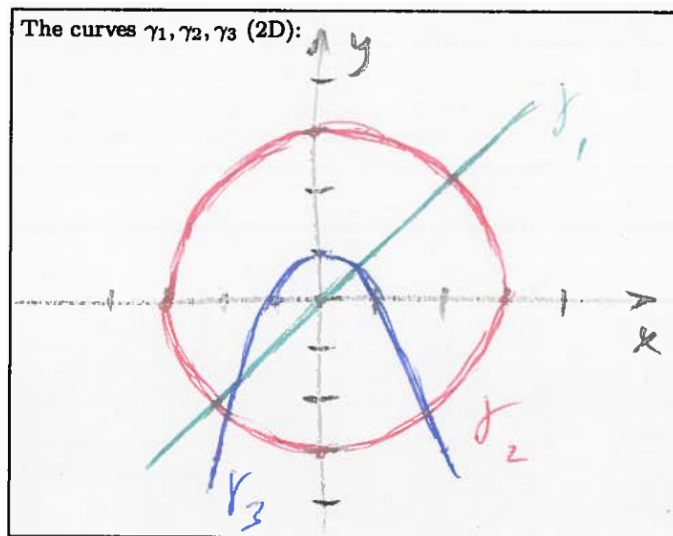
(ii) Consider the functions

$$
\begin{aligned}
f_1 &: & f_1(x, y) &= x - y\,, \\
f_2 &: & f_2(x, y) &= x^2 + y^2\,, \\
f_3 &: & f_3(x, y) &= \ln(3 + x^2 - y)\,,
\end{aligned}
$$

and find the three single-variable functions $F_i = f_i \circ \gamma_i$.

Sol.:

$$
\begin{aligned}
F_1(t) &= (f_1 \circ \gamma_1)(t) = f_1(t, t) = t - t = 0\,, \\
F_2(t) &= (f_2 \circ \gamma_2)(t) = f_2(3\cos t, 3\sin t) = (3\cos t)^2 + (3\sin t)^2 = 9\,, \qquad (2.2) \\
F_3(t) &= \ln\left(3 + t^2 - (1 - t^2)\right) = \ln\left(2 + 2t^2\right) = \ln 2 + \ln\left(1 + t^2\right).
\end{aligned}
$$

Theorem 2.17 (Chain Rule I)**.** Let f and F be as in Remark 2.15. Then

$$
\frac{\mathrm{d}F}{\mathrm{d}t} = \frac{\partial f}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial f}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t}\,.
$$

Example 2.18. (i) The derivatives of the functions F_i in 2.16 are

$$
\frac{\mathrm{d}F_1}{\mathrm{d}t} = \frac{\partial f_1}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial f_1}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t} = \frac{\partial(x - y)}{\partial x}\frac{\mathrm{d}t}{\mathrm{d}t} + \frac{\partial(x - y)}{\partial y}\frac{\mathrm{d}t}{\mathrm{d}t} = 1 \cdot 1 + (-1) \cdot 1 = 0\,,
$$

$$
\frac{\mathrm{d}F_2}{\mathrm{d}t} = 2x(t) \cdot (-3\sin t) + 2y(t) \cdot 3\cos t = -6\cos t \sin t + 6\sin t \cos t = 0\,,
$$

$$
\frac{\mathrm{d}F_3}{\mathrm{d}t} = \frac{2x}{3 + x^2 - y} \cdot 1 + \frac{-1}{3 + x^2 - y} \cdot (-2t) = \frac{4t}{2 + 2t^2} = \frac{2t}{1 + t^2}\,.
$$

Alternatively, these derivatives could have been found by directly differentiating the functions $F_i = F_i(t)$ in (2.2). However, the chain rule allowed us to find the derivatives without using the explicit expressions of t on the right-hand sides of (2.2), which is often very helpful.

54

(ii) We now present an example in the simplified notation mentioned at the end of Remark 2.15: For $f = x^2y + 3xy^4$, $x = \sin 2t$, $y = \cos t$, find $\mathrm{d}f/\mathrm{d}t$ at $t = 0$.
Sol.:

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \frac{\partial f}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial f}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t} = (2xy + 3y^4)(2\cos 2t) + (x^2 + 12xy^3)(-\sin t)\,.$$

The expressions containing the variable t can be evaluated at $t = 0$ directly. The other expressions we evaluate at the x and y values at $t = 0$: $x(0) = 0$, $y(0) = 1$. Therefore, the t-derivative of f at $t = 0$ is

$$\frac{\mathrm{d}f}{\mathrm{d}t}(0) = 3 \cdot 2 + 0 \cdot 0 = 6\,.$$

To give this derivative a geometric interpretation, think of a hiker walking along the path $(x(t), y(t))$ in the "mountain range" that is given by the graph of f. Then our computations show that at time $t = 0$, he will be climbing at a rate of six height units per time unit.

Remark 2.19. *Changes of variables* in the xy-plane can be realised by writing x and y as functions of *two* other variables,

$$(x, y) = (x(s, t), y(s, t))\,.$$

Given a function $f = f(x, y)$, one can then define a new function F that has the same values as f, but is written out in term of s and t :

$$F(s, t) = f(x(s, t), y(s, t))\,.$$

These changes of variables are quite useful – for example, we will use them to solve differential equations in chapter 4 – and it is important to be able to understand the relation of the partial derivatives of f and F.

As in Remark 2.15, the distinction $f \leftrightarrow F$ is sometimes not made, and the function of the new variables may be called f as well. Again, this simplifies the notation but also carries potential for confusion.

Theorem 2.20 (Chain Rule II). For s, t, x, y, f, F as in Remark 2.19, we have

$$\frac{\partial F}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s}\,,$$
$$\frac{\partial F}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t}\,.$$

Example 2.21. (i) Consider the function $f(x, y) = x - y$ and the change of variables

$$x(s, t) = s + t\,, \qquad y(s, t) = s - t\,.$$

For f in the new variables, i.e. $F(s, t) = f(x(s, t), y(s, t))$, we have partial derivatives

$$\frac{\partial F}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s} = 1 \cdot 1 + (-1) \cdot 1 = 0\,,$$
$$\frac{\partial F}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t} = 1 \cdot 1 + (-1) \cdot (-1) = 2\,.$$

As in the examples for derivatives along curves, the derivatives can be checked by explicitly writing out the function F in terms of s and t :

$$F(s,t) = f(x,y) = x - y = (s+t) - (s-t) = 2t ,$$

which has the partial derivatives that were obtained above.

(ii) Consider the paraboloid $f(x,y) = x^2 + y^2$ and the change of variables

$$x(r,\theta) = r\cos\theta, \qquad y(r,\theta) = r\sin\theta . \tag{2.3}$$

The partial derivatives of $F(r,\theta) = f(x,y)$ are

$$\frac{\partial F}{\partial r} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial r} = 2x \cdot \cos\theta + 2y \cdot \sin\theta = 2r\left(\cos^2\theta + \sin^2\theta\right) = 2r ,$$

$$\frac{\partial F}{\partial \theta} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial \theta} = 2r\cos\theta \cdot (-r\sin\theta) + 2r\sin\theta \cdot r\cos\theta = 0 .$$

Again, we check these derivatives by explicitly writing out the function F :

$$F(r,\theta) = f(x,y) = x^2 + y^2 = r^2\left(\cos^2\theta + \sin^2\theta\right) = r^2 ,$$

which confirms the partial derivatives that were found via chain rule II.

The change of variables (2.3) is the very important change to *polar coordinates.*
`Adjust this paragraph depending on whether polar coordinates are reviewed in the foundations chapter.`

(iii) Let $g(s,t) = f(s^2 - t^2, t^2 - s^2)$, where f is an arbitrary differentiable function. Show that g satisfies

$$t\frac{\partial g}{\partial s} + s\frac{\partial g}{\partial t} = 0 .$$

Sol.: Let $x = s^2 - t^2, y = t^2 - s^2$. Then,

$$\frac{\partial g}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s} = f_x \cdot 2s + f_y \cdot (-2s) ,$$

$$\frac{\partial g}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t} = f_x \cdot (-2t) + f_y \cdot 2t .$$

Therefore

$$t\frac{\partial g}{\partial s} + s\frac{\partial g}{\partial t} = 0 .$$

(iv) As in the examples 2.18 for chain rule I, we now present an example in simplified notation, where the function in the new variables is not given a new name. Consider the expression

$$3\frac{\partial^2 u}{\partial x^2} + \frac{1}{2}\frac{\partial^2 u}{\partial x \partial y} - \frac{1}{2}\frac{\partial^2 u}{\partial y^2} \tag{2.4}$$

for the function $u = u(x, y)$, and rewrite it in terms of the new variables ξ ("xi") and η ("eta") defined as

$$\xi = x + 3y, \quad \eta = x - 2y.$$

Sol.: Chain rule II allows to rewrite the derivatives of u with respect to x and y as derivatives w.r.t. ξ and η :

$$\frac{\partial u}{\partial x} = \frac{\partial u}{\partial \xi}\frac{\partial \xi}{\partial x} + \frac{\partial u}{\partial \eta}\frac{\partial \eta}{\partial x} = \frac{\partial u}{\partial \xi} \cdot 1 + \frac{\partial u}{\partial \eta} \cdot 1 = u_\xi + u_\eta,$$

$$\frac{\partial u}{\partial y} = \frac{\partial u}{\partial \xi}\frac{\partial \xi}{\partial y} + \frac{\partial u}{\partial \eta}\frac{\partial \eta}{\partial y} = \frac{\partial u}{\partial \xi} \cdot 3 + \frac{\partial u}{\partial \eta} \cdot (-2) = 3u_\xi - 2u_\eta.$$

These formulas for the first-order derivatives of u are only intermediate steps in our computation – we need the second-order derivatives appearing in (2.4):

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x}\right) = \frac{\partial}{\partial x}\left(u_\xi + u_\eta\right) = \frac{\partial u_\xi}{\partial x} + \frac{\partial u_\eta}{\partial x}$$

$$= \frac{\partial u_\xi}{\partial \xi}\frac{\partial \xi}{\partial x} + \frac{\partial u_\xi}{\partial \eta}\frac{\partial \eta}{\partial x} + \frac{\partial u_\eta}{\partial \xi}\frac{\partial \xi}{\partial x} + \frac{\partial u_\eta}{\partial \eta}\frac{\partial \eta}{\partial x} = u_{\xi\xi} + u_{\eta\xi} + u_{\xi\eta} + u_{\eta\eta}.$$

Recalling Theorem 2.12, we obtain

$$\frac{\partial^2 u}{\partial x^2} = u_{\xi\xi} + 2u_{\xi\eta} + u_{\eta\eta}.$$

The second-order derivative w.r.t. y is

$$\frac{\partial^2 u}{\partial y^2} = \frac{\partial}{\partial y}\left(3u_\xi - 2u_\eta\right) = 3\frac{\partial u_\xi}{\partial y} - 2\frac{\partial u_\eta}{\partial y}$$

$$= 3\left(\frac{\partial u_\xi}{\partial \xi}\frac{\partial \xi}{\partial y} + \frac{\partial u_\xi}{\partial \eta}\frac{\partial \eta}{\partial y}\right) - 2\left(\frac{\partial u_\eta}{\partial \xi}\frac{\partial \xi}{\partial y} + \frac{\partial u_\eta}{\partial \eta}\frac{\partial \eta}{\partial y}\right)$$

$$= 3\frac{\partial u_\xi}{\partial \xi} \cdot 3 + 3\frac{\partial u_\xi}{\partial \eta} \cdot (-2) - 2\frac{\partial u_\eta}{\partial \xi} \cdot 3 - 2\frac{\partial u_\eta}{\partial \eta} \cdot (-2)$$

$$= 9u_{\xi\xi} - 12u_{\xi\eta} + 4u_{\eta\eta}.$$

The mixed second-order derivative,

$$u_{xy} = 3u_{\xi\xi} + u_{\xi\eta} - 2u_{\eta\eta},$$

is obtained similarly. This allows to write out (2.4) in terms of ξ and η :

$$3\frac{\partial^2 u}{\partial x^2} + \frac{1}{2}\frac{\partial^2 u}{\partial x \partial y} - \frac{1}{2}\frac{\partial^2 u}{\partial y^2}$$

$$= 3\left[u_{\xi\xi} + 2u_{\xi\eta} + u_{\eta\eta}\right] + \frac{1}{2}\left[3u_{\xi\xi} + u_{\xi\eta} - 2u_{\eta\eta}\right] - \frac{1}{2}\left[9u_{\xi\xi} - 12u_{\xi\eta} + 4u_{\eta\eta}\right]$$

$$= 0 \cdot u_{\xi\xi} + \left(6 + \frac{1}{2} + 6\right) \cdot u_{\xi\eta} + 0 \cdot u_{\eta\eta} = \frac{25}{2}u_{\xi\eta}.$$

Later, in Section 4.3, we will use computations of this kind to solve a certain type of differential equation.

Definition 2.22 (Level Sets). Given a function f of n variables and a constant c in the range of f, the set of points in the domain of f with

$$f = c$$

is called a *level set* of f. That is, a level set of f is of the form

$$\left\{ (x, y) \in \mathbb{R}^2 \mid f(x, y) = c \right\} .$$

Remark 2.23.   (i) Level sets usually have one dimension less than the domain of the function. For example, the function $f(x, y) = x^2 + y^2$ is defined on the xy-plane, $\mathbb{R}^2$, which is two-dimensional. We have seen in Example 2.3 that the condition

$$f(x, y) = x^2 + y^2 = 1$$

describes a curve, i.e. a one-dimensional object, namely the circle of radius 1. In fact, $f(x, y) = x^2 + y^2 = c$ describes a one-dimensional object for any $c > 0$. This may not work for all values of c though: the only solution to $f = 0$ is the point $(x, y) = (0, 0)$, and points are considered zero-dimensional. The equation $f = c$, where $c < 0$, has no solutions at all.

Similarly, consider the function $f(x, y, z) = x^2 + y^2 + z^2$. It is defined in the three-dimensional space $\mathbb{R}^3$, and its level sets $f = c$ for $c > 0$ are two-dimensional objects: spheres of radius $\sqrt{c}$ (the surfaces of the balls of radius $\sqrt{c}$ centred at the origin of the coordinate system).

(ii) Considering x and y independent variables, we have

$$\frac{\partial y}{\partial x} = 0$$

(the x-derivative of the function $g(x, y) = y$ is zero, since g does not depend on x).

Now consider a level set $f(x, y) = c$. Under the constraint $f(x, y) = c$, the variables cannot both move freely any more. Such a level set is one-dimensional by the previous remark, i.e. a curve, and we can use x as a parameter for it. That is, we let x vary freely (independent variable) and then y is determined (not independent any more) by the choice of x.

For example, requiring $f_1 = 0$ in 2.16 gives $y(x) = x$, and $f_3 = \ln(2)$ gives $y(x) = 1 + x^2$.

Theorem 2.24 (Implicit Differentiation). If the independent variable y of the xy-plane $(x, y) \in \mathbb{R}^2$ is turned into a dependent variable $y = y(x)$ via a condition $f(x, y) = c$ (as in (ii) of the previous remark), then

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{f_x}{f_y} .$$

Similarly in $\mathbb{R}^3$: If the independent variable z of $(x, y, z) \in \mathbb{R}^3$ is turned into a dependent variable $z = z(x, y)$ via a condition $f(x, y, z) = c$, then

$$\frac{\partial z}{\partial x} = -\frac{f_x}{f_z} \,,$$

$$\frac{\partial z}{\partial y} = -\frac{f_y}{f_z} \,.$$

Proof. Differentiating the condition $f(x, y(x)) = c$ with respect to x gives

$$\frac{\mathrm{d}c}{\mathrm{d}x} = 0 \tag{2.5}$$

on the right-hand side. On the left-hand side, we use chain rule I to obtain

$$\frac{\mathrm{d}f(x, y(x))}{\mathrm{d}x} = \frac{\partial f(x, y)}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}x} + \frac{\partial f(x, y)}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}x} \,. \tag{2.6}$$

Note that here, f appears as a function of one independent variable on the left, and as a two-variable function on the right. Since $\mathrm{d}x/\mathrm{d}x = 1$, combining (2.5) and (2.6) leads to the claimed formula for $\mathrm{d}y/\mathrm{d}x$.

Similarly, we differentiate the condition $f(x, y, z(x, y)) = c$ with respect to x and y,

$$\frac{\partial f(x, y, z(x, y))}{\partial x} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial x} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial x} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial x} = f_x \cdot 1 + f_y \cdot 0 + f_z \cdot \frac{\partial z}{\partial x} = 0 \,,$$

$$\frac{\partial f(x, y, z(x, y))}{\partial y} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial y} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial y} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial y} = f_x \cdot 0 + f_y \cdot 1 + f_z \cdot \frac{\partial z}{\partial y} = 0 \,,$$

to derive the other formulas stated in the theorem. The terms with $\partial y/\partial x$ and $\partial x/\partial y$ drop out by the previous remark. Here, f appears as a two-variable function on the left and as a three-variable function in the other expressions. $\square$

Example 2.25. (i) Find $\mathrm{d}y/\mathrm{d}x$ if $x^3 + y^3 = 6xy$.
   Sol.:

$$f(x, y) = x^3 + y^3 - 6xy = 0 \quad\Longrightarrow\quad \frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{f_x}{f_y} = -\frac{3x^2 - 6y}{3y^2 - 6x} \,.$$

(ii) Find $\frac{\partial z}{\partial x}$, $\frac{\partial z}{\partial y}$, $\frac{\partial^2 z}{\partial x \partial y}$, where $z = z(x, y)$ is defined by the equation

$$x + y - z = \mathrm{e}^z \,.$$

Method 1: For $f(x, y, z) = x + y - z - \mathrm{e}^z$, we have

$$f_x = \frac{\partial f}{\partial x} = 1, \quad f_y = \frac{\partial f}{\partial y} = 1, \quad f_z = \frac{\partial f}{\partial z} = -1 - \mathrm{e}^z \,.$$

By Theorem 2.24,

$$\frac{\partial z}{\partial x} = -\frac{f_x}{f_z} = \frac{1}{1 + \mathrm{e}^z} \,,$$

$$\frac{\partial z}{\partial y} = -\frac{f_y}{f_z} = \frac{1}{1 + \mathrm{e}^z} \,,$$

and then, using the single-variable chain rule,

$$\frac{\partial^2 z}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{1}{1 + e^z}\right) = -\frac{1}{(1 + e^z)^2}\frac{\partial}{\partial x}(e^z)$$

$$= -\frac{e^z}{(1 + e^z)^2}\frac{\partial}{\partial x}(z) = -\frac{e^z}{(1 + e^z)^3} .$$

Method 2: Differentiating both sides of $x + y - z = e^z$ with respect to x gives

$$1 - \frac{\partial z}{\partial x} = e^z\frac{\partial z}{\partial x} .$$

Solving this for z_x, we find the same derivative as before.

The two computations are very closely related – Method 2 was presented to show that a more flexible approach can be used as well.

(iii) For $z = u^v$, where $u = \sin x$ and $v = \cos x$, find $\mathrm{d}z/\mathrm{d}x$.
   Sol.: The chain rule for $z = z(x)$ reads

$$\frac{\mathrm{d}z}{\mathrm{d}x} = \frac{\partial z}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial z}{\partial v}\frac{\partial v}{\partial x} .$$

The partials of $z = z(u, v)$ are

$$\frac{\partial z}{\partial u} = vu^{v-1}, \quad \frac{\partial z}{\partial v} = u^v \ln u ,$$

and therefore

$$\frac{\mathrm{d}z}{\mathrm{d}x} = vu^{v-1}\cos x - u^v \ln u \cdot \sin x = (\sin x)^{\cos x - 1}\cos^2 x - \ln(\sin x)\cdot(\sin x)^{\cos x + 1} .$$

(iv) (Logarithmic Differentiation) Find $\partial z/\partial x$, $\partial z/\partial y$ for

$$z = (x^2 y + xy^2)^{\cos(xy)} . \tag{2.7}$$

Sol.: The idea for dealing with the function in the power on the right is to use the rule $\ln(a^p) = p\ln(a)$ for logarithms. If $a = b$, then $\ln(a) = \ln(b)$ – applying this principle to (2.7) gives

$$\ln z = \ln\left((x^2 y + xy^2)^{\cos(xy)}\right) = \cos(xy)\ln(x^2 y + xy^2) .$$

Differentiating both sides with respect to x, we obtain

$$\frac{1}{z}\frac{\partial z}{\partial x} = -y\sin(xy)\ln(x^2 y + xy^2) + \cos(xy)\frac{2xy + y^2}{x^2 y + xy^2}$$

and

$$\frac{\partial z}{\partial x} = (x^2 y + xy^2)^{\cos(xy)}\left[-y\sin(xy)\ln(x^2 y + xy^2) + \cos(xy)\frac{2x + y}{x^2 + xy}\right] .$$

The y-derivative of z will be given as an exercise. It can be found analoguously, but it is much faster to point to the symmetry in (2.7) and hence write down $\partial z/\partial y$ immediately.

Application (Trajectories in phase space). `Simple and quick example of a 2D phases space, e.g. for 1D motion; define energy on that phase space and find level curves.`

Exercise 2.26.   (i) Find the rate of change of the functions

$$f(x,y) = x + x^2 y - 5y^3 , \quad g(x,y) = \sqrt{1 + x^2 + y^2}$$

along the curve
$$\gamma(t) = (x(t), y(t)) = (t, t^2) .$$

That is, find the t-derivatives of[37] $F(t) = f(\gamma(t))$ and $G(t) = g(\gamma(t))$.

(ii) Write the function $f(x,y) = x + y^2$ in polar coordinates, i.e. find an expression for $F = F(r, \theta) = f(x,y)$ in terms of r and θ. Then differentiate F directly and compare to the r and θ derivatives obtained via chain rule II (as in example 2.21 (ii)).

(iii) Find $\partial z / \partial x$ and $\partial z / \partial x$ for $z = z(x,y)$ defined by $z^3 - 3xyz = 4$.

(iv) In Example 2.25 (iv), the x-derivative of z in (2.7) was found by applying $\partial / \partial x$ to the logarithm of (2.7). Now find the derivative with respect to y by repeating those steps. Comparing the two partial derivatives, try to interpret the reference to "symmetry" at the end of the example[38].

(v) Consider $u = u(x,y)$ and define the new variables $\xi = x - y$, $\eta = x + by$. Find b so that the equation

$$\frac{\partial^2 u}{\partial x^2} + 4 \frac{\partial^2 u}{\partial x \partial y} + 3 \frac{\partial^2 u}{\partial y^2} = 0$$

transforms to[39] $u_{\xi\eta} = 0$.

(vi) Consider $u = u(x,y)$ and the new variables s, t defined by $x = e^s \cos t$, $y = e^2 \sin t$. Show that[40]
$$u_{xx} + u_{yy} = e^{-2s}(u_{ss} + u_{tt}) .$$

(vii) Find functions f and corresponding constants c such that the level sets $f = c$ are: a straight line of slope $+1$, a circle, a straight line of slope different from ± 1, a parabola, a hyperbola, an ellipse, two concentric circles, etc. Check your examples using software[41].

(viii) Have the level set
$$x^2(1 - x^2) - y^2 = 0$$

plotted and find the values of x for which $\partial y / \partial x = 0$. What is the connection between those x-values and the plot of the level set[42]?

(ix) Find the first-order partial derivatives of[43] $z(x,y) = (\cos(2x))^{\ln y}$.

(x) `Trajectory in phase space`

61

2.3 Directional Derivatives and the Gradient Vector

Definition 2.27 (Gradient Vectors). The *gradient vector* ∇f of a function $f(x,y)$ is the row vector containing its partial derivatives,

$$\nabla f = \begin{bmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} \end{bmatrix}.$$

Similarly for $f = f(x,y,z)$,

$$\nabla f = \begin{bmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} & \dfrac{\partial f}{\partial z} \end{bmatrix}.$$

Definition 2.28 (Directional Derivatives). For a vector $v = \begin{bmatrix} a & b \end{bmatrix}^\top$ and a function f of two variables, the *directional derivative* of f along v is

$$D_v f = a\frac{\partial f}{\partial x} + b\frac{\partial f}{\partial y} = \nabla f \begin{bmatrix} a \\ b \end{bmatrix},$$

and similarly for three-variable functions and directions $v = \begin{bmatrix} a & b & c \end{bmatrix}^\top$.

Example 2.29. Find the gradient of the function $f(x,y) = x\cos(y) - \sin(xy)$, and further its directional derivative along $v = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \end{bmatrix}^\top$ at the point $(x_0, y_0) = (\pi, 0)$.
Sol.:

$$f(x,y) = x\cos(y) - \sin(xy),$$
$$\nabla f(x,y) = \begin{bmatrix} \cos(y) - \cos(xy)y & -x\sin(y) - \cos(xy)x \end{bmatrix},$$
$$\nabla f(x_0, y_0) = \begin{bmatrix} \cos(0) - \cos(\pi 0)0 & -\pi\sin(0) - \cos(\pi 0)\pi \end{bmatrix} = \begin{bmatrix} 1 & -\pi \end{bmatrix},$$
$$D_v f(0, \pi) = \nabla f(x_0, y_0) \cdot v = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -\pi \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1 - \pi}{\sqrt{2}}.$$

Remark 2.30. (i) Consider a function $f = f(x,y)$ and a curve

$$\gamma(t) = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix}$$

in the xy-plane. (In the previous section, we had used the notation "$(x(t), y(t))$" for curves. Here, we simply switched from point notation (x, y) for the points of γ to vector notation $\begin{bmatrix} x & y \end{bmatrix}^\top$, cf. the application in Section 1.1.) Define the function $F(t) = f(\gamma(t)) = f(x(t), y(t))$. Then, comparing the definition of directional derivatives to chain rule I, we find

$$\frac{\mathrm{d}F}{\mathrm{d}t} = \frac{\partial f}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial f}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t} = \begin{bmatrix} f_x & f_y \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} = D_{\gamma'}f,$$

where prime dentotes the t-derivative and $\gamma' = \begin{bmatrix} x' & y' \end{bmatrix}^\top$.

(ii) We introduce the following notation for the statement of the next theorem:

$$v \parallel w$$

means that the vectors v and w are *parallel*, i.e. one of them can be obtained from the other by scalar multiplication. The expression

$$v \perp w$$

means that the two vectors are *perpendicular* (or *orthogonal*), i.e. the angle between them is $\pi/2$ ($90°$ degrees). Reviewing section 1.1, in particular remark 1.6 and the definition before that, you will find:

$$v \perp w \quad \Longleftrightarrow \quad v \circ w = 0 \,.$$

An important operation for vectors is to *normalise* them: Given a vector v, we scalar-multiply it by the inverse of its length,

$$\bar{v} = \frac{1}{||v||} \cdot v = \frac{v}{||v||} \,,$$

to obtain a vector $\bar{v}$ that (a) is parallel to v, i.e. points in the same direction, and (b) has unit length, $||\bar{v}|| = 1$. This is where the factor $1/\sqrt{2}$ of v in the previous example comes from – convince yourself that this vector has unit length!

Theorem 2.31 (Geometric Interpretation of Directional Derivatives). For vectors v of length $||v|| = 1$, the absolute value of the directional derivative is maximal if $\nabla f^\top \parallel v$, and zero if $\nabla f^\top \perp v$.

In particular, level sets of a function are always perpendicular to the gradient of the function.

Proof. The directional derivative of f along v at the point (x_0, y_0) is

$$|\mathrm{D}_v f(x_0, y_0)| = |\nabla f(x_0, y_0) \cdot v| = \left| (\nabla f(x_0, y_0))^\top \circ v \right|$$

$$\overset{\text{Remark 1.6}}{=} \left| \cos(\theta) \cdot || (\nabla f(x_0, y_0))^\top || \cdot ||v|| \right| = |\cos(\theta)| \cdot || (\nabla f(x_0, y_0))^\top || \,,$$

where θ is the angle between the vectors v and $(\nabla f(x_0, y_0))^\top$. The length of the latter is fixed, and only the cosine term can be changed by choosing a different vector v of length 1. If the two vectors are parallel, then $|\cos \theta| = 1$; if they are perpendicular, then $|\cos \theta| = 0$; and for all other angles, $|\cos \theta|$ will be between 0 and 1. This proves the first part of the statement.

For the second part, let the curve $\gamma(t)$ traverse a level set of f and consider a fixed t-value t_0. Then $F(t) = f(\gamma(t))$ is constant, and therefore

$$0 = \frac{\mathrm{d}F}{\mathrm{d}t}(t_0) = \nabla f(x_0, y_0) \cdot \gamma'(t_0) = (\nabla f(x_0, y_0))^\top \circ \gamma'(t_0) \,,$$

where $\begin{bmatrix} x'(t_0) & y'(t_0) \end{bmatrix}^\top = \gamma'(t_0)$. This dot product being 0 means that the two vectors are perpendicular. Since $\gamma'(t_0)$ is the tangent vector of the curve $\gamma(t)$ at t_0, we have completed the proof of the theorem. $\qquad\square$

Remark 2.32. (i) The same works in three dimensions: Then, a level set $f = c$ is a surface (two-dimensional), and the vector $(\nabla f)^\top$ will always be perpendicular to it.

(ii) We now introduce a matrix of second-order derivatives that will be useful in the next section. For that, review Theorem 2.12.

Definition 2.33 (Hessian Matrix). The *Hessian matrix* or *Hessian* of $f = f(x, y)$ is the matrix
$$\text{Hess} f = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} .$$
Similarly for functions of three or more variables.

Example 2.34. Find the gradient and the Hessian of $f(x, y, z) = xy^2 + z^3$.
Sol.:
$$\nabla f = \begin{bmatrix} y^2 & 2xy & 3z^2 \end{bmatrix},$$
$$\text{Hess} f = \begin{bmatrix} 0 & 2y & 0 \\ 2y & 2x & 0 \\ 0 & 0 & 6z \end{bmatrix} .$$

Exercise 2.35. (i) For $f(x, y) = \ln(x + y^2)$, find the gradient and the Hessian matrix at[44] $(x_0, y_0) = (3, 1)$.

(ii) Find the directional derivatives $D_v f(x_0, y_0)$ for[45]:

(a) $f(x, y) = y + {x^2}/{y}$, $(x_0, y_0) = (1, 2)$, $v = \begin{bmatrix} 12/13 & 5/13 \end{bmatrix}^\top$;

(b) $f(x, y) = x^2 - 3xy + 2y^2$, $(x_0, y_0) = (1, 1)$, $v = \begin{bmatrix} 3/5 & -4/5 \end{bmatrix}^\top$;

(c) $f(x, y) = x \arctan(y/x)$, $(x_0, y_0) = (1, -1)$, $v = \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix}^\top$.

(iii) Find the Hessian of $f(x, y) = \frac{1}{6}\left((x - 1)^3 + (y + 2)^3\right)$ at the point where[46] $\nabla f = 0$.

(iv) For $f(x, y) = (e^{3x} + \sin(4y))^5$, find the unit vector v (i.e., $||v|| = 1$) of steepest descent at[47] $(x, y) = (0, \pi/4)$.

(v) In remark 2.30, we have written out chain rule I as a matrix multiplication of a row vector and a column vector. Do the same for chain rule II: Find the matrix M such that for the functions $f = f(x, y)$ and $F = F(s, t)$ in Theorem 2.20 we have[48]
$$\nabla F = \nabla f \cdot M .$$

2.4   Taylor Approximations

Remark 2.36. Recall single-variable Taylor approximations: For $F = F(t)$, we have the approximation
$$F(t) \approx T_a^{(2)} F(t) = F(a) + F'(a) \cdot (t - a) + \frac{1}{2} F''(a) \cdot (t - a)^2$$

close to the point $t = a$. In this section, we will derive the corresponding formula for functions of several variables.

Theorem 2.37 (Taylor Approximation). Let $f(x, y)$ have continuous partial derivatives of order up to 2, and let the point (a, b) be in its domain. Then the *first- and second-order Taylor approximations* of f at (a, b) are

$$T^{(1)}_{(a,b)} f(x, y) = f(a, b) + \nabla f(a, b) \begin{bmatrix} x - a \\ y - b \end{bmatrix},$$

$$T^{(2)}_{(a,b)} f(x, y) = f(a, b) + \nabla f(a, b) \begin{bmatrix} x - a \\ y - b \end{bmatrix} \tag{2.8}$$

$$+ \frac{1}{2} \begin{bmatrix} x - a & y - b \end{bmatrix} \operatorname{Hess} f(a, b) \begin{bmatrix} x - a \\ y - b \end{bmatrix}.$$

Close to (a, b), we have

$$T^{(1)}_{(a,b)} f \approx f, \qquad T^{(2)}_{(a,b)} f \approx f,$$

where the second-order approximation is better in the sense that the error

$$\left| T^{(2)}_{(a,b)} f(x, y) - f(x, y) \right|$$

decays faster than $\left| T^{(1)}_{(a,b)} f(x, y) - f(x, y) \right|$ as $(x, y) \to (a, b)$. Similarly for $f = f(x, y, z)$.

Proof. Let (x, y) be close to the point (a, b), so that both

$$h = x - a \quad \text{and} \quad k = y - b$$

are small. Define the curve

$$\gamma(t) = (a + ht, b + kt),$$

and let $F(t) = f(\gamma(t))$. Note that $\gamma(0) = (a, b)$ and $\gamma(1) = (x, y)$. We now find the second-order Taylor approximation of the single-variable function F at $t = 0$:

$$F(t) \approx T_0^{(2)} F(t) = F(0) + F'(0) \cdot t + \frac{1}{2} F''(0) \cdot t^2$$

$$= F(0) + t \cdot \frac{\mathrm{d}F}{\mathrm{d}t}(0) + \frac{1}{2} t^2 \cdot \frac{\mathrm{d}^2 F}{\mathrm{d}t^2}(0)$$

$$= f(a, b) + t \cdot \left[\frac{\partial f}{\partial x} \frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial f}{\partial y} \frac{\mathrm{d}y}{\mathrm{d}t} \right]_{t=0} + \frac{1}{2} t^2 \cdot \frac{\mathrm{d}}{\mathrm{d}t} \left[\frac{\partial f}{\partial x} \frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial f}{\partial y} \frac{\mathrm{d}y}{\mathrm{d}t} \right]_{t=0}$$

$$= f(a, b) + t \cdot \nabla f(a, b) \begin{bmatrix} h \\ k \end{bmatrix} + \frac{1}{2} t^2 \cdot \frac{\mathrm{d}}{\mathrm{d}t} \left[f_x h + f_y k \right]_{t=0}.$$

The remaining t-derivative is that of a composition with γ (the evaluation at $t = 0$ indicated as $[\ldots]_{t=0}$ does not mean that f_x and f_y are evaluated at $t = 0$ – they

are to be evaluated at $\gamma(0) = (a, b)$), and therefore, chain rule I needs to be applied one more time. Note that h and k do not depend on t.

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[f_x\,h + f_y\,k\right]_{t=0} = h \cdot \left[\frac{\partial f_x}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial f_x}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t}\right] + k \cdot \left[\frac{\partial f_y}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial f_y}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t}\right]$$

$$= h^2 f_{xx} + hk f_{yx} + kh f_{xy} + k^2 f_{yy} \,,$$

where all the partials are evaluated at (a, b). The last expression can be written as a matrix multiplication of the row vector containing h and k, a matrix containing the second-order derivatives of f, and the column vector containing h and k. This gives

$$F(t) \approx f(a, b) + t\,\nabla f(a, b)\begin{bmatrix} h \\ k \end{bmatrix} + \frac{t^2}{2}\begin{bmatrix} h & k \end{bmatrix}\operatorname{Hess}f(a, b)\begin{bmatrix} h \\ k \end{bmatrix}\,.$$

Evaluating this approximation at $t = 1$, we obtain the stated second-order approximation of f,

$$f(x, y) \approx f(a, b) + \nabla f(a, b)\begin{bmatrix} h \\ k \end{bmatrix} + \frac{1}{2}\begin{bmatrix} h & k \end{bmatrix}\operatorname{Hess}f(a, b)\begin{bmatrix} h \\ k \end{bmatrix}\,.$$

Note that $t = 1$ is not small, i.e. that it is not close to the point $t = 0$ about which F was developed, and that in general it may be too big for a good approximation. However, in the above situation, this can be controlled by requiring h and k to be small, that is, by letting (x, y) be close to the point (a, b) from which the approximation of f is carried out. $\qquad\square$

Example 2.38. (i) Find the first- and second-order Taylor approximation of $f(x, y) = \mathrm{e}^x \ln(1 + y)$ at $(a, b) = (0, 0)$.

Sol.: We begin by finding all the values, vectors, and matrices at (a, b) that are needed for writing out the Taylor approximations (2.8):

$$f(x, y) = \mathrm{e}^x \ln(1 + y)\,,$$

$$f(0, 0) = 1 \cdot 0 = 0\,,$$

$$\nabla f(x, y) = \begin{bmatrix} \mathrm{e}^x \ln(1 + y) & \frac{\mathrm{e}^x}{1+y} \end{bmatrix}\,,$$

$$\nabla f(0, 0) = \begin{bmatrix} 0 & 1 \end{bmatrix}\,,$$

$$\operatorname{Hess}f(x, y) = \begin{bmatrix} \mathrm{e}^x \ln(1 + y) & \frac{\mathrm{e}^x}{1+y} \\ \frac{\mathrm{e}^x}{1+y} & -\mathrm{e}^x \frac{1}{(1+y)^2} \end{bmatrix}\,,$$

$$\operatorname{Hess}f(0, 0) = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix}\,.$$

This gives

$$T^{(1)}_{(0,0)}f(x, y) = f(0, 0) + \nabla f(0, 0)\begin{bmatrix} x - 0 \\ y - 0 \end{bmatrix} = 0 + \begin{bmatrix} 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = y$$

and

$$T^{(2)}_{(0,0)}f(x, y) = 0 + \begin{bmatrix} 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \frac{1}{2}\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$

$$= y + \frac{1}{2}\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} y \\ x - y \end{bmatrix} = y + xy - \frac{y^2}{2}\,.$$

(ii) Find the first- and second-order Taylor approximation of $f(x, y) = 2x^2 + y^2$ about the point $(a, b) = (1, 1)$.

Sol.:

$$f(x, y) = 2x^2 + y^2 \,,$$
$$f(1, 1) = 3 \,,$$
$$\nabla f(x, y) = \begin{bmatrix} 4x & 2y \end{bmatrix} \,,$$
$$\nabla f(1, 1) = \begin{bmatrix} 4 & 2 \end{bmatrix} \,,$$
$$\mathrm{Hess} f(x, y) = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} \,,$$
$$\mathrm{Hess} f(1, 1) = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} \,,$$

$$T_{(1,1)}^{(2)} f(x, y) = 3 + \begin{bmatrix} 4 & 2 \end{bmatrix} \begin{bmatrix} x - 1 \\ y - 1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} x - 1 & y - 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x - 1 \\ y - 1 \end{bmatrix}$$

$$= -3 + 4x + 2y + \begin{bmatrix} x - 1 & y - 1 \end{bmatrix} \begin{bmatrix} 2x - 2 \\ y - 1 \end{bmatrix} = 2x^2 + y^2 \,.$$

Note that we have obtained $T_{(a,b)}^{(2)} f = f$. This is because $f(x, y)$ is already a very simple function – an expression of x and y containing terms of order at most 2. The analogous situation in the corresponding single-variable theory is: The m-th-order Taylor approximation of a polynomial of order $d \leq m$ is the original polynomial itself.

It is important not to forget about one of the parts we were asked about – the first-order approximation! It is contained in the above computation:

$$T_{(1,1)}^{(1)} f(x, y) = -3 + 4x + 2y \,.$$

(iii) Find the second-order Taylor approximation of $f(x, y) = \sin(x) \sin(y)$ about the point $(a, b) = (\pi/4, \pi/4)$.

Sol.: Note that $\sin(\pi/4) = \cos(\pi/4) = \sqrt{2}/2$. The required quantities at (a, b) are

$$f(x, y) = \sin(x) \sin(y) \,,$$
$$f(\pi/4, \pi/4) = \frac{1}{2} \,,$$
$$\nabla f(x, y) = \begin{bmatrix} \cos(x) \sin(y) & \sin(x) \cos(y) \end{bmatrix} \,,$$
$$\nabla f(\pi/4, \pi/4) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix} \,,$$
$$\mathrm{Hess} f(x, y) = \begin{bmatrix} -\sin(x) \sin(y) & \cos(x) \cos(y) \\ \cos(x) \cos(y) & -\sin(x) \sin(y) \end{bmatrix} \,,$$
$$\mathrm{Hess} f(\pi/4, \pi/4) = \frac{1}{2} \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \,.$$

For simplicity, we write out the approximation in terms of $h = x - \pi/4$ and

67

$k = y - \pi/4$:

$$T^{(2)}_{(\pi/4,\pi/4)} f(x,y) = \frac{1}{2} + \frac{1}{2} \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} h \\ k \end{bmatrix} + \frac{1}{4} \begin{bmatrix} h & k \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} h \\ k \end{bmatrix}$$

$$= \frac{1}{2} \left[1 + h + k + hk - \frac{h^2}{2} - \frac{k^2}{2} \right] ,$$

which can then be translated back to an expression of x and y by replacing h and k with $x - \pi/4$ and $y - \pi/4$ respectively.

Remark 2.39. (i) The second-order Taylor approximation of $f(x,y)$ at (a,b) has the following properties (denote it by T for simplicity: $T = T^{(2)}_{(a,b)} f$).

(a) $T(a,b) = f(a,b)$ (same function value at (a,b))

(b) $\nabla T(a,b) = \nabla f(a,b)$ (same first-order derivatives at (a,b))

(c) $\mathrm{Hess}\, T(a,b) = \mathrm{Hess}\, f(a,b)$ (same second-order derivatives at (a,b))

(d) T is a second-order expression, i.e., a linear combination of $1, x, y, x^2, xy, y^2$.

The last point should help you identify mistakes. For example, if you obtain terms like $x^2 y$ or y^5 or $x \sin y$ in the approximation, you should review your computation – forgetting to evaluate the gradient or the Hessian at the given point can lead to such expressions.

(ii) Similarly, $T^{(1)}_{(a,b)} f$ is a first-order expression,

$$T^{(1)}_{(a,b)} f = c_1 + c_2 x + c_3 y ,$$

It further is a function of x and y, of course, and we plot such functions in $\mathbb{R}^3$ by drawing the function values as the z-coordinate:

$$z = c_1 + c_2 x + c_3 y .$$

This is the equation of a plane in $\mathbb{R}^3$, namely the plane tangent to the graph of f at the point $(x_0, y_0, z_0) = (a, b, f(a,b))$. This is similar to the corresponding 1D theory: The first-order Taylor approximation gives the equation of the tangent line.

Graphs of f and $T^{(1)}_{(a,b)} f$ (3D):
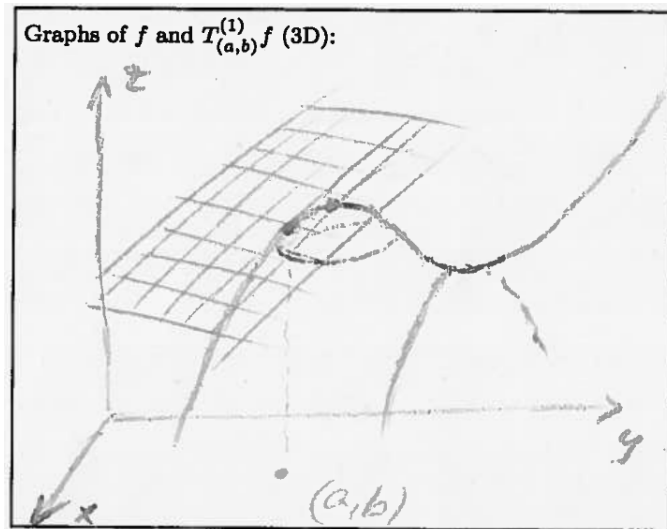
Exercise 2.40. (i) Find the second-order Taylor approximation of $f(x,y) = x^2(1-y^3)$ at the point[49] $(a,b) = (2,1)$.

(ii) Find the second-order Taylor approximation of $f(x,y,z) = xy^2 + z^3$ at[50] $(a,b,c) = (1,1,1)$.

(iii) Find the second-order Taylor approximation of $f(x,y) = \arctan(x+2y)$ at[51] $(a,b) = (5,-2)$.

(iv) Re-do example (ii) above, $f(x,y) = 2x^2 + y^2$, but now at a general point $(a,b) \in \mathbb{R}^2$. That is, use parameters a and b for the point at which the approximation is carried out, rather than specific numbers.

(v) Find the intersection of the xy-plane with the tangent plane of $f(x,y) = 2x^2 + y^2$ at[52] $(1,-2)$.

2.5 Local Extrema and Saddle Points

Definition 2.41 (Local Extrema and Critical Points)**.** For a function $f(x,y)$ and a point (a,b) in its domain (similarly in three variables):

(i) (a,b) is a *local maximum* of f if

$$f(x,y) \le f(a,b)$$

in some neighbourhood of (a,b).

(ii) (a,b) is a *local minimum* of f if there exists an $\delta > 0$ such that

$$f(x,y) \ge f(a,b)$$

for all (x,y) with $\left\| \begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} a \\ b \end{bmatrix} \right\| < \delta$.

(iii) (a,b) is a *critical point* of f if

$$\nabla f(a,b) = \begin{bmatrix} 0 & 0 \end{bmatrix}.$$

Remark 2.42. (i) Neighbourhoods are small sets surrounding the point in question. For example, in one dimension: For $t_0 = 0.001$, there exists a small neighbourhood of t_0 that contains only positive numbers, e.g. the open interval $(t_0 - 0.0005, t_0 + 0.0005)$. For $t_0 = 0$, however, this is not possible: every neighbourhood – no matter how small – contains both positive and negative numbers. In the higher dimensional domains of multivariate functions, one can work with disks (in 2D) or balls (in 3D). This notion is very important in Mathematics and used informally in part (i) of the definition above. In part (ii), the same condition is written out more formally using a parameter δ and the corresponding disk of radius δ around the point (a,b). It does not matter how small δ is, but $\delta > 0$ is essential.

(ii) A local extremum is either a local minimum or a local maximum. Local maxima are not surpassed in some small neighbourhood, while for local minima there exists a neighbourhood in which none of the other function values is smaller. There is no requirement for surrounding function values to be strictly smaller, respectively strictly larger – that means, for example, that for a constant function $f(x, y) = c$, every point in its domain is both a local maximum and a local minimum.

(iii) For the material in this section, it is useful to recall the corresponding single-variable theory. The following result should look familiar and its proof will be left as an exercise.

Theorem 2.43. If (a, b) is a local extremum of f, then (a, b) is a critical point of f.

Example 2.44. (i) Find the critical points of $f(x, y) = x^2 y - y + 7$.
Sol.: The gradient of f is

$$\nabla f(x, y) = \begin{bmatrix} 2xy & x^2 - 1 \end{bmatrix} .$$

Setting its second component, i.e., the y-derivative of f, equal to zero, we obtain $x^2 - 1 = 0$, which has solutions $x = \pm 1$. The second equation is $2xy = 0$, which, since we know $x \neq 0$, has the solution $y = 0$. Therefore, the critical points are $(-1, 0)$ and $(+1, 0)$.

(ii) Find the critical points of $f(x, y) = xy$.
Sol.: Setting the gradient of f,

$$\nabla f(x, y) = \begin{bmatrix} y & x \end{bmatrix} ,$$

equal to zero gives the critical point $(a, b) = (0, 0)$.

Note, however, that this is not a local extremum: The function value at that point is 0, and there are points with both positive and negative function values arbitrarily close to it. For example, consider the points $(x, y) = (\varepsilon, \varepsilon)$ and $(x, y) = (\varepsilon, -\varepsilon)$, where ε is a small positive constant.

(iii) Find the critical points of $f(x, y) = x^3 - 3x + 3xy^2$.
Sol.: After computing the gradient of f, we find that $f_x = 0$ for all points on the unit circle, and $f_y = 0$ for all points that lie on one of the two axes. This gives the critical points $(+1, 0)$, $(0, +1)$, $(-1, 0)$, $(0, -1)$.

Theorem 2.45 (Second-Derivative Test)**.** Let (a, b) be a critical point of $f(x, y)$ and let
$$\Delta = \det \left(\mathrm{Hess} f(a, b) \right) = \left(f_{xx} f_{yy} - f_{xy}^2 \right)_{|(a,b)} .$$
Then (a, b) can be classified as follows.

(i) If $\Delta > 0$ and $f_{xx}(a, b) > 0$, then (a, b) is a local minimum.

(ii) If $\Delta > 0$ and $f_{xx}(a, b) < 0$, then (a, b) is a local maximum.

(iii) If $\Delta < 0$, then (a, b) is not a local extremum.

(iv) If $\Delta = 0$, then the test is inconclusive and (a, b) could be any of the three possibilities above (local maximum, local minimum, or neither).

Example 2.46. (i) Find and classify all critical points of $f(x, y) = x^2 y - y + 7$.
Sol.: In the previous example, we have found the critical points $P_1 = (-1, 0)$ and $P_2 = (+1, 0)$. The Hessian of f is

$$\operatorname{Hess} f(x, y) = \begin{bmatrix} 2y & 2x \\ 2x & 0 \end{bmatrix}.$$

At the first point, we have

$$\operatorname{Hess} f(P_1) = \begin{bmatrix} 0 & -2 \\ -2 & 0 \end{bmatrix},$$

which has determinant $\Delta = 0 \cdot 0 - (-2) \cdot (-2) = -4 < 0$. Hence, P_1 is neither a local maximum nor a local minimum. Points like P_1 are called *saddle points*, cf. Remark 2.47. At the second critical point, we have

$$\operatorname{Hess} f(P_2) = \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix},$$

which has determinant $\Delta = 0 \cdot 0 - 2 \cdot 2 = -4 < 0$. Hence, P_2 is a saddle point as well and therefore not a local extremum.

(ii) Find and classify all critical points of $f(x, y) = x^3 - 3x + 3xy$.
Sol.: Earlier we had found the critical points $P_1 = (+1, 0)$, $P_2 = (0, +1)$, $P_3 = (-1, 0)$, and $P_4 = (0, -1)$. The Hessian of f is

$$\operatorname{Hess} f(x, y) = \begin{bmatrix} 6x & 6y \\ 6y & 6x \end{bmatrix} = 6 \begin{bmatrix} x & y \\ y & x \end{bmatrix}.$$

This gives
$$\operatorname{Hess} f(P_1) = 6I, \qquad \operatorname{Hess} f(P_3) = -6I,$$

which both have determinant $\Delta = +36 > 0$. Inspecting the signs of the entries in the upper left of those Hessians, we find that P_1 is a local minimum and that P_3 is a local maximum. At P_2 and P_4, we have

$$\operatorname{Hess} f(P_{2/4}) = \begin{bmatrix} 0 & \pm 6 \\ \pm 6 & 0 \end{bmatrix},$$

which both have determinant $\Delta = -36 < 0$. Therefore, P_2 and P_4 are not local extrema.

(iii) Find all critical points of $f(x, y) = x - \ln x + 2y^2 + xy$ and classify them using the second derivative test.

Sol.:

$$f(x, y) = x - \ln x + 2y^2 + xy\,,$$
$$\nabla f(x, y) = \begin{bmatrix} 1 - \frac{1}{x} + y & 4y + x \end{bmatrix}\,,$$
$$\operatorname{Hess} f(x, y) = \begin{bmatrix} \frac{1}{x^2} & 1 \\ 1 & 4 \end{bmatrix}\,,$$
$$P = (2, -\tfrac{1}{2})\,,$$
$$\Delta = 0\,.$$

Therefore, the second derivative test is inconclusive, and more work is required to classify the critical point $P = (2, -\frac{1}{2})$ of f.

Remark 2.47. It is useful to understand the workings behind the second derivative test. We therefore sketch its derivation:

Since (a, b) is a critical point of f, the second-order Taylor approximation at that point is

$$T^{(2)}_{(a,b)} f(x, y) = f(a, b) + 0 + \frac{1}{2} \begin{bmatrix} x - a & y - b \end{bmatrix} \operatorname{Hess} f(a, b) \begin{bmatrix} x - a \\ y - b \end{bmatrix}\,,$$

and it seems plausible that studying the approximation $T^{(2)}_{(a,b)} f$ should allow to classify (a, b). Now, for the question whether (a, b) is an extremum or not, overall additive constants such as the term $f(a, b)$ above do not matter. Also, we may shift our coordinate system so that $(a, b) \to (0, 0)$. That is, we let $\widetilde{x} = x - a$, $\widetilde{y} = y - b$, as in the proof of the Taylor approximation, and we then write $\widetilde{x}$ and $\widetilde{y}$ as x and y again to simplify the notation. Thirdly, we denote the matrix $\frac{1}{2} \cdot \operatorname{Hess} f$ at the critical point by M. Note that M is symmetric. With those steps, we have reduced classifying the critical point (a, b) of the original function f to classifying the critical point $(0, 0)$ of the auxiliary function

$$h(x, y) = \begin{bmatrix} x & y \end{bmatrix} M \begin{bmatrix} x \\ y \end{bmatrix}\,.$$

More advanced theory of matrices implies that it suffices to restrict one's attention to the case when M is a diagonal matrix. The basic idea for this argument is: if M is not diagonal, then there is a change of variables so that M written out in the new variables is diagonal. This change of variables is a rotation of the xy-plane, which does not affect the property of being a minimum, maximum, or neither. The function h is now of the form

$$h(x, y) = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} \lambda & 0 \\ 0 & \mu \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}\,,$$

where $\lambda, \mu \in \mathbb{R}$. For example, consider the case $\lambda = 4$, $\mu = 1$. Then

$$h(x, y) = 4x^2 + y^2 = (2x)^2 + y^2 = \widetilde{x}^2 + y^2\,,$$

where we rescaled the x axis, $x \to \widetilde{x} = 2x$. The expression on the right is a paraboloid, for which $(0, 0)$ is a minimum. The rescaling causes a deformation

– imagine a paraboloid, then squeeze it so that its level sets are ellipses rather than circles – but, again, this does not affect the property of having a minimum, maximum, or neither at the point $(0,0)$. We have therefore found that h with $\lambda = 4$ and $\mu = 1$ has a local minimum at $(0,0)$. By this argument, we find that all the cases with $\lambda > 0$ and $\mu > 0$ are qualitatively the same and have a minimum. Let us choose the identity matrix as representing those cases,

$$M_1 = I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

for which $h(x,y) = x^2 + y^2$ and the graph of h is a paraboloid:

Example for case (i) (h induced by M_1, paraboloid; 3D):
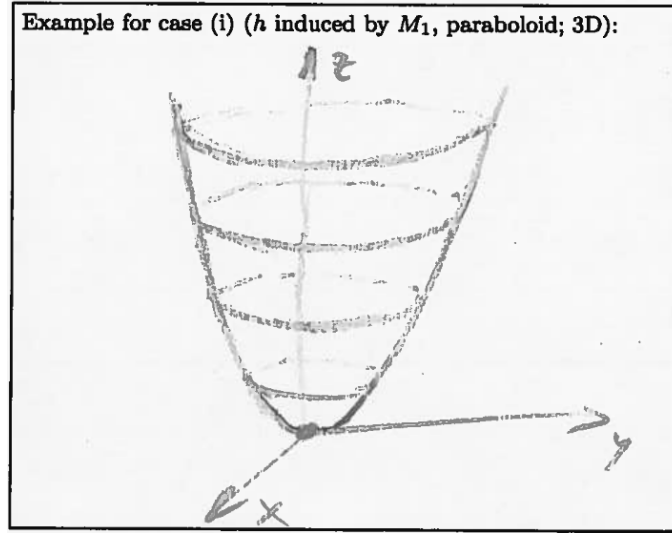
This is case (i) of the second derivative test, a local minimum. Note that the chosen representative $M_1 = I$ for this class satisfies the assumptions of (i): $\det M_1 > 0$ and its upper left entry is positive.

Continuing this line of reasoning, one finds that the other cases to consider are:

$$M_2 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, M_3 = \begin{bmatrix} +1 & 0 \\ 0 & -1 \end{bmatrix}, M_4 = \begin{bmatrix} +1 & 0 \\ 0 & 0 \end{bmatrix}, M_5 = \begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix}, M_6 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

The matrix M_2 generates the upside-down paraboloid $h(x,y) = -(x^2 + y^2)$,

Example for case (ii) (h induced by M_2, upside-down paraboloid; 3D):
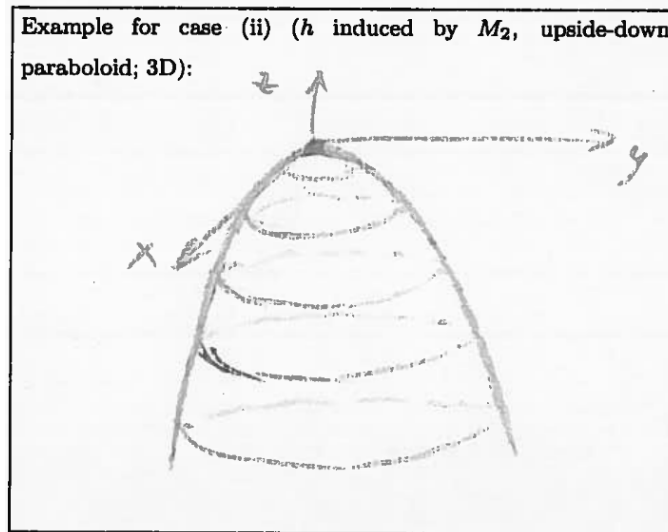
which has a maximum at $(0,0)$. Again, note that M_2 satisfies the assumptions of (ii) of the second derivative test. The matrix M_3 gives

$$h(x,y) = x^2 - y^2 \,,$$

which does not have an extremum at $(0,0)$: we have $h(0,0) = 0$, but both positive and negative function values arbitrarily close (e.g. $h(\varepsilon,0) > 0$ and $h(0,\varepsilon) < 0$). This is case (iii) of the second derivative test, and both the graph of $h(x,y) = x^2 - y^2$,

Example for case (iii), (h induced by M_3, saddle; 3D):
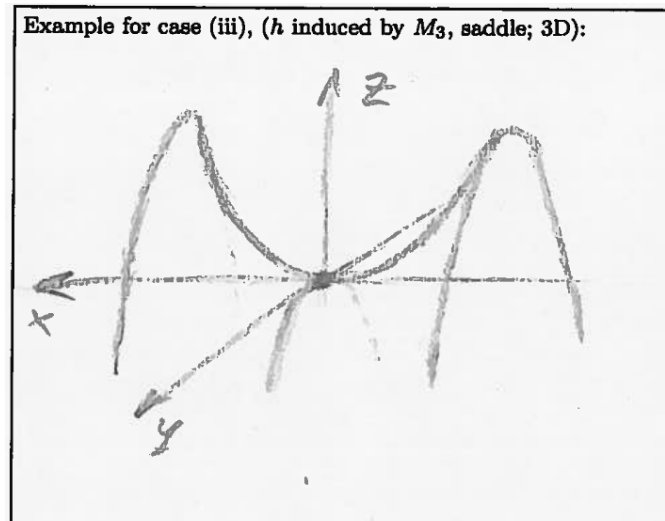
and the form of M_3 agree with the statements in (iii).

Finally, we argue that the cases represented by M_4, M_5, M_6 do not allow us to draw a conclusion, i.e. they correspond to case (iv) of the second derivative test. They all do satisfy the assumption $\det M = 0$. The choice M_5 gives $h(x,y) = -x^2$, which has the following graph.

Example for case (iv) (h induced by M_5; 3D):
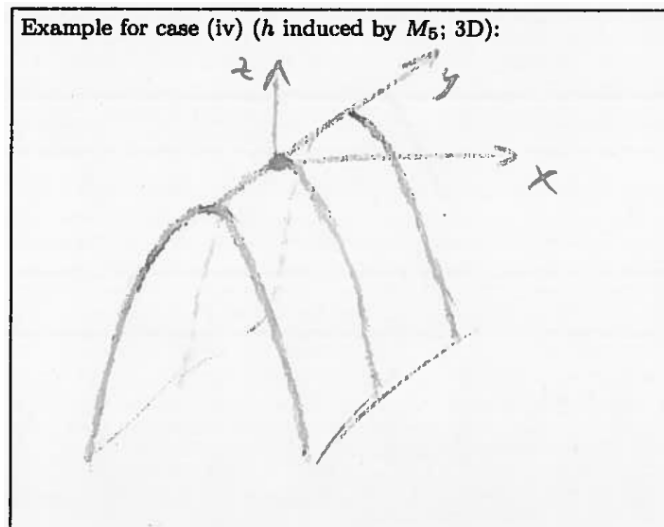
The matrix M_4 produces the same graph, but upside-down, and the graph obtained by choosing M_6 is the plane $h(x, y) = 0$. In each of these cases, there is at least one straight line passing through the critical point $(0, 0)$ in question. On this line, small contributions of the original function f, that are not captured by the second-order Taylor approximation, could tip the balance to different conclusions. Some guidance for understanding this will be provided in the exercises below.

The purpose of this remark was to explain the workings behind the second derivative test, to outline its proof, and, perhaps most importantly, to help avoid confusion of the cases (i) and (ii): For example, suppose you are classifying a critical point, you have obtained a Hessian matrix with positive determinant and positive entry in the upper left corner, but you have forgotten whether this implies a minimum or a maximum. Then think of the representative $M = I$ of this situation – this gives the function $h(x, y) = x^2 + y^2$, which is the well-known paraboloid, which has a *minimum* at $(0, 0)$.

Application (Mean squared error for linear regression). Define mean squared error for the linear regression application from the previous chapter; apply theory from the current chapter to basic matrix functions and hence show that linear regression minimises the MSE.

Exercise 2.48. (i) Fill in the gaps for examples 2.44 (iii) and 2.46 (iii), which were only sketched.

(ii) Find an classify all critical points of[53] $f(x, y) = x^2 + xy + 2y - 1$.

(iii) Find and classify all critical points of[54] $f(x, y) = x^3 + 3xy^2 - 3x^2 - 3y^2 - 2$.

(iv) Convince yourself that Theorem 2.43 is true[55].

(v) Give examples of a local minimum, a local maximum, and a critical point that is neither, that cannot be classified with the second derivative test[56].

(vi) Use graphing software to explore different functions of the form

$$h(x, y) = \begin{bmatrix} x & y \end{bmatrix} M \begin{bmatrix} x \\ y \end{bmatrix},$$

where M is a symmetric 2×2 matrix (i.e., $a_{12} = a_{21}$), and compare the different shapes you find to Remark 2.47.

(vii) For 2.46 (iii), use graphing software to find out what the point P is[57].

2.6 Extrema under Constraints: Lagrange Multipliers

Example 2.49. Find the maximum of the function $z = f(x, y) = 2x + y$ on the circle $S = \{(x, y) \in \mathbb{R}^2 \,|\, x^2 + y^2 = 1\}$, as well as the value that f takes at that point. *Sol.:* First note that the graph of f is a plane, and therefore f does not have any extrema at all on its full domain $\mathbb{R}^2$. However, a maximum does exist when restricting to the circle:
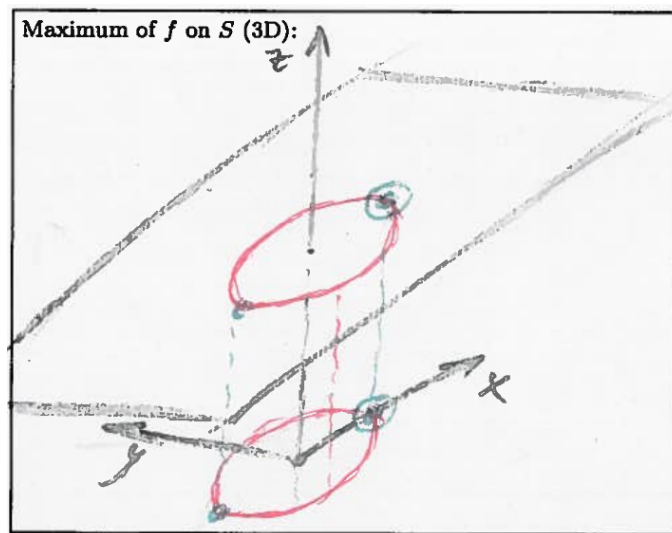
To find this point, we parametrise the circle as

$$(x(t), y(t)) = (\cos(t), \sin(t)) \,,$$

and then define the function $F(t) = f(x(t), y(t))$. This gives

$$F(t) = 2\cos(t) + \sin(t) \,,$$
$$F'(t) = -2\sin(t) + \cos(t) \,.$$

Setting $F'(t)$ equal to zero, we obtain

$$\tan(t_0) = \frac{1}{2} \,,$$

which has solutions $t_0 \approx 0.464 + k\pi$. Taking $k = 0$ and $k = 1$ leads to the points

$$(x_1, y_1) \approx (\cos(0.464), \sin(0.464)) \approx (0.894, 0.448) \,,$$
$$(x_2, y_2) \approx (\cos(3.606), \sin(3.606)) \approx (-0.894, -0.448) \,,$$

and all other k will yield one of those two points again. Evaluating, we see that f has a larger value at (x_1, y_1), and hence we obtain the answer: f takes the maximum value 2.236 on the circle S at the point $(x, y) = (0.894, 0.448)$.

Remark 2.50. Now suppose the set S is given as a level set of a function g, and is less regular and can not be parametrised easily. The following figure shows such a set S in the domain of f, and further a level set $f(x,y) = c$, where the constant $c = c_1$ is chosen to be larger than any of the values f takes on S.

The level sets $g = 0$ and $f = c$ for large c (2D):
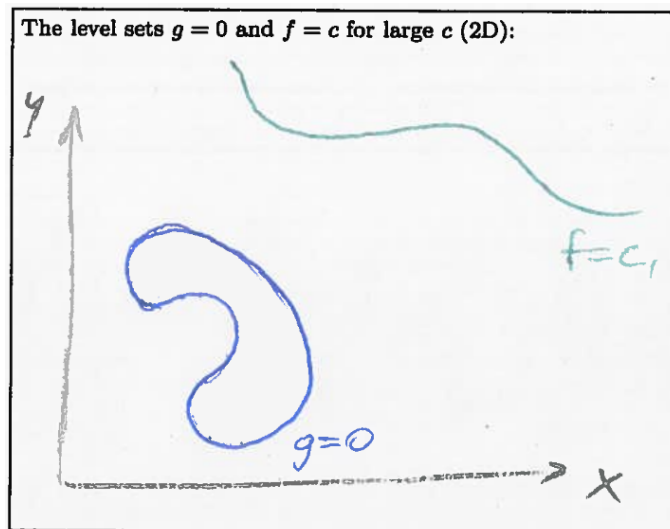
Now, choosing a slightly smaller constant c_2, the curve $f = c$ will move towards $g = 0$. We continue this process until the first contact between the two curves is made, say for $c = c_3$. This contact point is a local maximum – convince yourself of that[58]!

The level sets $g = 0$ and $f = c$ for smaller values of c (2D):
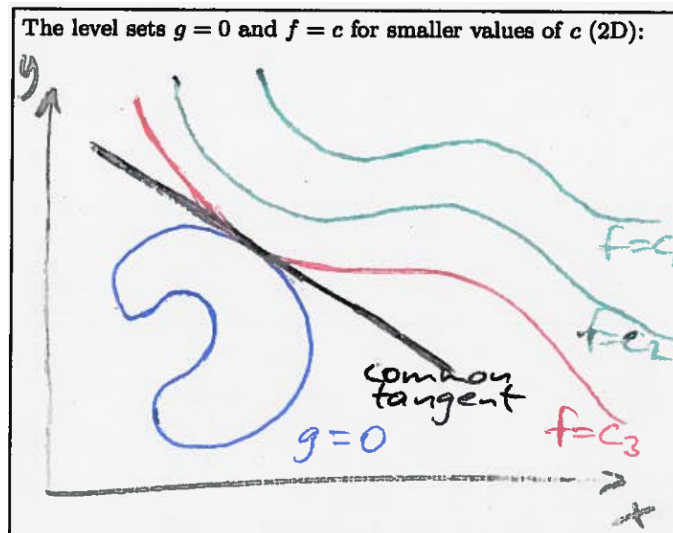
Note that the level curves $g = 0$ and $f = c_3$ have a common tangent line – convince yourself that this is always the case[59]. Recall that gradient vectors are orthogonal to level sets and their tangent lines or planes. It must therefore be the case that the gradients of f and g are parallel:

$$\nabla f = \lambda \nabla g, \qquad \text{for some } \lambda \in \mathbb{R}.$$

Theorem 2.51 (Lagrange Multipliers). To maximise or minimise a function $f(x, y)$ on a set $S = \{(x, y) \in \mathbb{R}^2 \mid g(x, y) = 0\}$, define the function

$$F(\lambda, x, y) = f(x, y) - \lambda g(x, y) ,$$

and then solve

$$\nabla F = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} .$$

The function f can then be evaluated at the solutions (x, y) of that system to identify the extreme value. The parameter λ is called *Lagrange multiplier*.

Example 2.52. (i) Re-do example 2.49 using Lagrange multipliers.
    Sol.: We have $f(x, y) = 2x + y$ and we let

$$g(x, y) = x^2 + y^2 - 1 ,$$

so that

$$S = \{g = 0\} .$$

Then we define

$$F(\lambda, x, y) = f(x, y) - \lambda g(x, y) = 2x + y - \lambda(x^2 + y^2 - 1) ,$$

which has partial derivatives

$$\begin{aligned} F_\lambda(\lambda, x, y) &= 1 - (x^2 + y^2) , \\ F_x(\lambda, x, y) &= 2 - 2\lambda x , \\ F_y(\lambda, x, y) &= 1 - 2\lambda y . \end{aligned}$$

Setting ∇F equal to 0 leads to a system of three equations (the equations are not linear, and hence the techniques from chapter 1 can not be used to solve it). The equations $F_x = 0$ and $F_y = 0$ lead to

$$x = \frac{1}{\lambda}, \qquad y = \frac{1}{2\lambda} ,$$

which we can then substitute into $F_\lambda = 0$:

$$0 = 1 - \left(\left(\frac{1}{\lambda} \right)^2 + \left(\frac{1}{2\lambda} \right)^2 \right) = 1 - \frac{5}{4\lambda^2} \quad \rightarrow \quad \lambda = \pm \frac{\sqrt{5}}{2} .$$

Using the equations for x and y above, we obtain the points

$$(x_1, y_1) = \left(\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}} \right) \approx (0.894, 0.447) ,$$

$$(x_2, y_2) = \left(-\frac{2}{\sqrt{5}}, -\frac{1}{\sqrt{5}} \right) \approx (-0.894, -0.447) .$$

Now, f has to have both a maximum and a minimum on S – this is similar to the extreme value theorem for continuous single-variable functions – and the

two points above are the only candidates for those extrema. Evaluating, we find

$$f(x_1, y_1) = 2\frac{2}{\sqrt{5}} + \frac{1}{\sqrt{5}} = \sqrt{5}$$

and $f(x_2, y_2) = -\sqrt{5}$. Hence, (x_1, y_1) is the maximum.

Note that, contrary to the computation in example 2.49, we did not have to numerically evaluate inverse trigonometric functions, and we were able to obtain exact symbolic expressions for the maximum function value and the coordinates of the point where it is taken. Also, comparing the two different approaches, we have proven that

$$\cos\left(\arctan\left(\frac{1}{2}\right)\right) = \frac{2}{\sqrt{5}} \ .$$

(ii) Find the point on the surface $(x - y)^2 - z^2 = 1$ that is closest to the origin $(x, y, z) = (0, 0, 0)$ of $\mathbb{R}^3$.

Sol.: The function

$$d(x, y, z) = \sqrt{x^2 + y^2 + z^2} \ ,$$

which assigns to every point in $\mathbb{R}^3$ its distance from the origin, needs to be minimised on the surface

$$S = \{x, y, z \in \mathbb{R} \mid (x - y)^2 - z^2 = 1\} \subseteq \mathbb{R}^3 \ .$$

Suppose we have found a point $P \in S$ with minimal $d(P)$. Then $(d(P))^2$ will be minimal as well (the formal justification for this step is that $x \mapsto x^2$ is monotonously increasing for $x \geq 0$). Hence we can simplify our computations by minimising

$$f(x, y, z) = x^2 + y^2 + z^2$$

instead of d. The function $g(x, y, z) = (x - y)^2 - z^2 - 1$ defines S and gives

$$F(\lambda, x, y, z) = x^2 + y^2 + z^2 - \lambda((x - y)^2 - z^2 - 1) \ .$$

Setting the gradient of F equal to zero, we obtain the system

$$\begin{cases} 0 & = 2x - \lambda \cdot 2(x - y) \ , \\ 0 & = 2y - \lambda \cdot 2(x - y)(-1) \ , \\ 0 & = 2z - \lambda \cdot (-2z) \ , \\ 0 & = 1 + z^2 - (x - y)^2 \ . \end{cases}$$

There are different ways to solve this, and it is important to gain experience with computations of that kind in order to be able to carry them out efficiently and with confidence. Combining the first two equations yields $y = -x$ and the new system

$$\begin{cases} 0 & = 2x(1 - 2\lambda) \ , \\ 0 & = 2z(1 + \lambda) \ , \\ 0 & = 1 + z^2 - 4x^2 \ . \end{cases}$$

In R1, one of the two factors has to be 0. Taking $x = 0$, R3 becomes $1 + z^2 = 0$, which does not have solutions. Taking $\lambda = \frac{1}{2}$ in R1 leads to the points

$$P_{1/2} \; : \; (x, y, z) = \left(\pm \frac{1}{2}, \mp \frac{1}{2}, 0 \right) .$$

The function values at those critical points are $f(P_1) = f(P_2) = \frac{1}{2}$, corresponding to distances $d(P_1) = d(P_2) = \frac{\sqrt{2}}{2}$ from the origin. Now, there have to be points of minimal distance to the origin on S, and since the two critical points we have found are the only candidates for such extrema, we see that P_1 and P_2 are minima. Note that f does not have any maxima – this is possible because S is unbounded.

(iii) Bob wants to sell free-range eggs and needs to decide how many chicken to keep ("x") and how much land to lease ("y"). In order to be able to label his eggs as free-range, every chicken needs to have at least $4\,m^2$ of space. Leasing land costs $f_2(y) = y/10$ and selling eggs brings in an income $f_1(x) = 8\sqrt{x}$. Help Bob optimise his profits. (The numbers in this problem are fictional.)
Sol.:

$$\text{Profit} \; : \quad f(x, y) = f_1(x) - f_2(y) \,,$$
$$\text{Function } F \; : \quad F(\lambda, x, y) = 8\sqrt{x} - \frac{y}{10} - \lambda \left(\frac{y}{x} - 4 \right) ,$$
$$\text{Answer} \; : \quad x = 100, \quad y = 400, \quad f_{\max} = 40 \,.$$

Exercise 2.53. (i) Find the maximum of $f(x, y) = x$ on[60] $x^3 + y^2 = 1$.

(ii) Just to make sure we are not sending Bob down the wrong path: also solve example 2.52 (iii) with single-variable theory, as in example 2.49.

(iii) Find the maximum and minimum value of $f(x, y) = x^3 y^5$ on the ellipse[61] $3x^2 + y^2 = 8$.

(iv) Find the point on $S \; : \; z = xy$ that is closest to the sphere of radius 1 centred at[62] $(x, y, z) = (0, 0, 5)$.

(v) A single-variable problem similar to example 2.52 (ii): For $c \in \{1, -4\}$, consider the curve $y = x^2 + c$ and the distances of its points to the origin $(0, 0)$ of the xy-plane. Sketch the curves and think about what kind of extrema exist in each case. Then compute all local and global extrema as well as the corresponding distances to the origin[63].

(vi) Find the maximum value of

$$f(x, y, z) = \ln x + \ln y + 3 \ln z \,, \quad x > 0, y > 0, z > 0$$

on the sphere $x^2 + y^2 + z^2 = 5R^2$.

Chapter 3

Integration
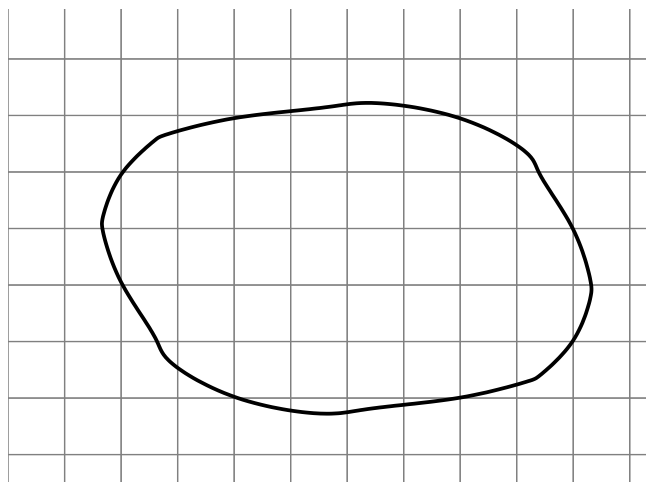
Suppose you have an oddly shaped room with square tiles, and you want to estimate its square footage. Let n be the number of full tiles you count, and let m stand for the number of broken tiles. Denote the area of a single tile – which can be computed: it is the square of the side length – by A_0. Then we have the lower and upper bounds

$$n \cdot A_0 \leq A \leq (n + m) \cdot A_0$$

for the area A of the room. Now, convince yourself that this estimate would be better (tighter), if the tiles were smaller[64]!

Finding areas under curves or volumes under surfaces is called *integration*. It is one of the most fundamental and important concepts of mathematics, and it has applications in every scientific discipline. The above approach of approximating areas with large numbers of simple pieces can be quite cumbersome. Fortunately, there is very powerful connection to differentiation, which allows to compute integrals in a more efficient way.

Application (Centre of mass)**.** A large metal plate of uniform thickness and the shape in the sketch (let us call it S; the area under $f(x) = \sqrt{x}$ for x from 0 to 9) needs to be balanced on a single point.
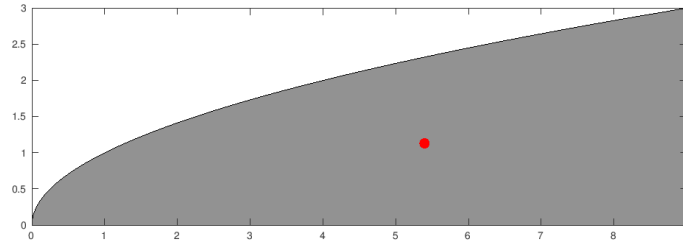
That is, the centre of mass of S needs to be found.

This can be solved with integration: The x and y coordinates of the centre of mass of S are

$$x_c = \frac{\iint_S x \, \mathrm{d}A}{\iint_S 1 \, \mathrm{d}A} = 5.4,$$

$$y_c = \frac{\iint_S y \, \mathrm{d}A}{\iint_S 1 \, \mathrm{d}A} = 1.125.$$

We will learn how to compute those integrals in this chapter.

3.1 Theory of Integration in One Dimension

Definition 3.1 (Definite Integral, Riemann Sum)**.** Let $f(x)$ be continuous on $a \leq x \leq b$ and divide the interval $[a, b]$ into n subintervals of equal width $\Delta x = {}^{(b-a)}\!/_n$. Let

$$x_0 = a, \quad x_1 = a + \Delta x, \quad x_2 = x_1 + \Delta x = x_0 + 2\Delta x, \quad \ldots, \quad x_n = x_0 + n\Delta x = b$$

be the endpoints of these subintervals.

(i) Then the *definite integral* of f from a to b is defined as

$$\int_a^b f(x) \, \mathrm{d}x := \lim_{n \to \infty} \sum_{j=1}^n f(c_j)\Delta x, \tag{3.1}$$

where $x_{j-1} \leq c_j \leq x_j$. Here $f(x)$ is called the *integrand*, a the lower boundary, and b the upper boundary.

(ii) The sum appearing in the above definition is called a *Riemann sum* of f over $[a, b]$. If the points c_j are always chosen to be the right-hand end point of their subinterval, that is $c_j = x_j$, we obtain the right-hand Riemann sum

$$\sum_{j=1}^n f(x_j)\Delta x.$$

Using the left end points of the subintervals, $c_j = x_{j-1}$, gives the left-hand Riemann sum

$$\sum_{j=1}^n f(x_{j-1})\Delta x = \sum_{j=0}^{n-1} f(x_j)\Delta x.$$

Remark 3.2. (i) Equation (3.1) introduces a new quantity – the definite integral of f over $[a, b]$ – and *defines* (that is what the ":=" means) it to be equal to the limit for $n \to \infty$ of the Riemann sums on the right. Note that on the right-hand side of (3.1), different choices could be made for the points c_j. One should therefore show now that all these possible different choices lead to the same limit. That is, one should show that $\int_a^b f(x)\,dx$ as defined above is *well-defined*! This will be given as an exercise at the end of this section.

(ii) The following figure illustrates the right-hand Riemann sum for $n = 4$,

$$\sum_{j=1}^{4} f(x_j)\Delta x,$$

of some function f.

Right-hand Riemann sum of f for $n = 4$:
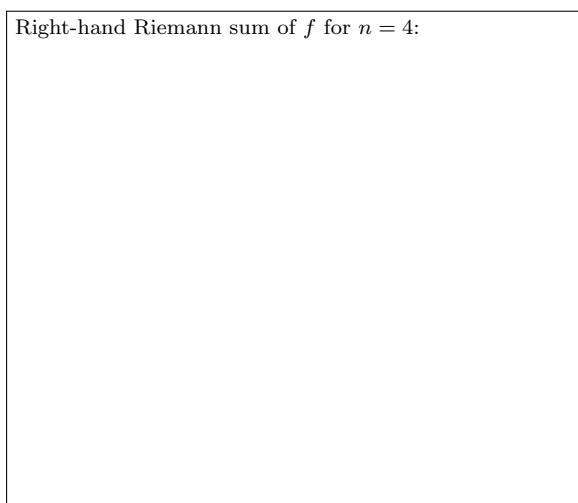
The rectangles in this figure have areas "height" $\cdot$ "width" $= f(x_j) \cdot \Delta x$, and they are added up to obtain the Riemann sum. For comparison, the left-hand Riemann sum is

Left-hand Riemann sum of f for $n = 4$:
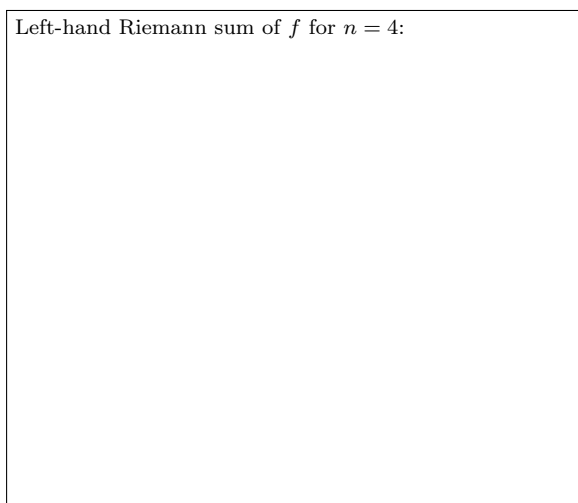
We see that in both cases, the Riemann sums approximate the area under the curve $y = f(x)$, and as in the example of the oddly-shaped room in the introduction, those approximations will get better if we choose n to be larger. We can therefore interpret their limit, i.e., $\int_a^b f(x)\,\mathrm{d}x$, as the signed area under the curve $y = f(x)$:

Definite integral of f over the integral $[a, b]$:

Here, the word "signed" was added since areas below the x-axis contribute negatively to the definite integral.

Example 3.3. Consider $f(x) = {}^3\!/_5\, x$, $a = 0$, $b = 5$. First, let us find the left-hand Riemann sum for $n = 5$ of f. The partition of $[a, b] = [0, 5]$ is

$$\Delta x = \frac{b - a}{n} = \frac{5 - 0}{5} = 1,$$
$$x_0 = 0,\ x_1 = 1,\ x_2 = 2,\ \ldots, x_5 = 5,$$

and therefore

$$\int_0^5 \frac{3}{5} x\,\mathrm{d}x \approx \sum_{j=0}^{4} f(x_j)\Delta x = [f(0) + f(1) + f(2) + f(3) + f(4)] \cdot 1$$
$$= \frac{3}{5}[0 + 1 + 2 + 3 + 4] = 6.$$

We now generalise this computation to find the definite integral f over the interval $[a, b]$. The partition of the interval is

$$\Delta x = \frac{b - a}{n} = \frac{5 - 0}{n} = \frac{5}{n},$$
$$x_0 = 0,\ x_1 = \frac{5}{n},\ x_2 = 2\frac{5}{n},\ \ldots, x_n = n\frac{5}{n} = 5,$$

84

and therefore

$$\int_0^5 \frac{3}{5} x \, \mathrm{d}x = \lim_{n \to \infty} \sum_{j=0}^{n-1} f(x_j) \Delta x = \lim_{n \to \infty} \frac{5}{n} \sum_{j=0}^{n-1} \frac{3}{5} x_j$$

$$= \lim_{n \to \infty} \frac{5}{n} \sum_{j=0}^{n-1} \frac{3}{5} \frac{5}{n} j = \lim_{n \to \infty} \frac{15}{n^2} \sum_{j=0}^{n-1} j$$

$$= \lim_{n \to \infty} \frac{15}{n^2} \frac{(n-1)n}{2} = \lim_{n \to \infty} \frac{15}{2} \frac{n-1}{n} = 7.5,$$

Note that this result agrees with the area

$$A = \frac{3 \cdot 5}{2},$$

found by noticing that the region under the curve $y = f(x)$ is a right-angled triangle.

Properties 3.4. Let $a, b, \alpha \in \mathbb{R}$, $a < b$, and let $f(x), g(x)$ be continuous functions on $[a, b]$. Then:

(1) $\qquad \int_a^a f(x) \, \mathrm{d}x = 0,$

(2) $\qquad \int_a^b f(x) \, \mathrm{d}x = - \int_b^a f(x) \, \mathrm{d}x,$

(3) $\qquad \int_a^b [f(x) + g(x)] \, \mathrm{d}x = \int_a^b f(x) \, \mathrm{d}x + \int_a^b g(x) \, \mathrm{d}x,$

(4) $\qquad \int_a^b \alpha \cdot f(x) \, \mathrm{d}x = \alpha \cdot \int_a^b f(x) \, \mathrm{d}x,$

(5) $\qquad \int_a^b f(x) \, \mathrm{d}x = \int_a^c f(x) \, \mathrm{d}x + \int_c^b f(x) \, \mathrm{d}x$

(6) $\qquad$ If $f(x) \geq 0$ on $[a, b]$, then $\int_a^b f(x) \, \mathrm{d}x \geq 0.$

Now suppose that $f(x)$ is continuous on the interval $[-a, a]$, where $a > 0$. Then we have

(7) $\qquad$ If $f(x)$ is even, i.e. $f(-x) = f(x)$, then $\int_{-a}^a f(x) \, \mathrm{d}x = 2 \int_0^a f(x) \, \mathrm{d}x,$

(8) $\qquad$ If $f(x)$ is odd, i.e. $f(-x) = -f(x)$, then $\int_{-a}^a f(x) \, \mathrm{d}x = 0.$

Remark 3.5. (i) All those properties are proven starting from the definition 3.1 of the definite integral. (There is nothing else to start from, is there?) Only the proof of the last property is written out below.

(ii) Properties (3) and (4) state that integration is *linear*, and they can by proven by using the corresponding laws for sums and limits (e.g., pull the constant α out of the sum and then out of the limit in definition 3.1).

85

Proof. For the proof of (8), we consider an odd function and we use only even n for the limit $n \to \infty$. Since all choices for the points c_j lead to the same limit, we make a choice that suits our purpose well: let the c_j be the midpoints of their subintervals. The Riemann sums are

$$\sum_{j=1}^{n} f(c_j)\Delta x = \Delta x \cdot [f(c_1) + f(c_2) + \ldots f(c_{n-1}) + f(c_n)],$$

and comparing to

An odd function and its Riemann sum, using midpoints:
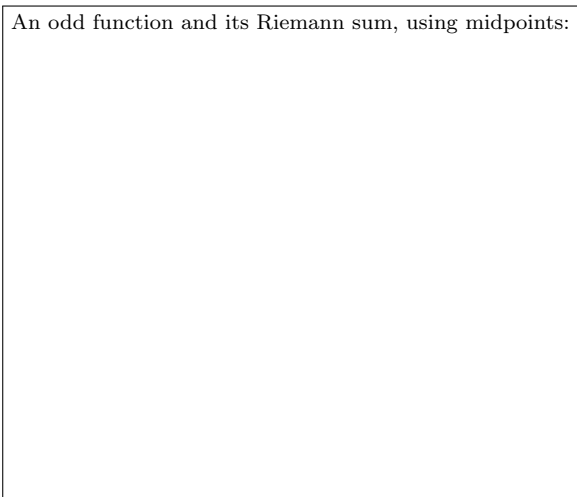
we see that this Riemann sum is zero since $f(c_1) + f(c_n) = 0$, $f(c_2) + f(c_{n-1}) = 0$, ... This gives

$$\int_{-a}^{a} f(x)\,\mathrm{d}x = \lim_{n \to \infty} 0 = 0.$$

Now, the above assertion that $f(c_j) + f(c_{n-j+1}) = 0$ seems to be confirmed by the sketch, but we should prove it properly: After working out formulas for the partition points x_j, we find

$$c_j = -a + \Delta x \left(j - \frac{1}{2}\right) = -a + \frac{2a}{n}\left(j - \frac{1}{2}\right)$$

for the midpoints c_j of the subintervals. This gives

$$c_{n-j+1} = -a + \frac{2a}{n}\left(n - j + 1 - \frac{1}{2}\right)$$

$$= -a + 2a + \frac{2a}{n}\left(-j + \frac{1}{2}\right) = -c_j,$$

and therefore, since f is odd,

$$f(c_j) + f(c_{n-j+1}) = f(c_j) + f(-c_j) = f(c_j) - f(c_j) = 0.$$

$\square$

Definition 3.6 (Mean). For continuous $f : [a, b] \to \mathbb{R}$, the real number

$$\bar{f} := \frac{1}{b - a} \int_a^b f(x) \, dx$$

is called the *mean* or *average* of f over the interval $[a, b]$.

Remark 3.7. We have

$$\bar{f} \cdot (b - a) = \int_a^b f(x) \, dx,$$

that is, the area under the constant function $\bar{f}$ over $[a, b]$ is the same as the area under the graph of f:

Definite integrals of f and of the constant function $\bar{f}$:
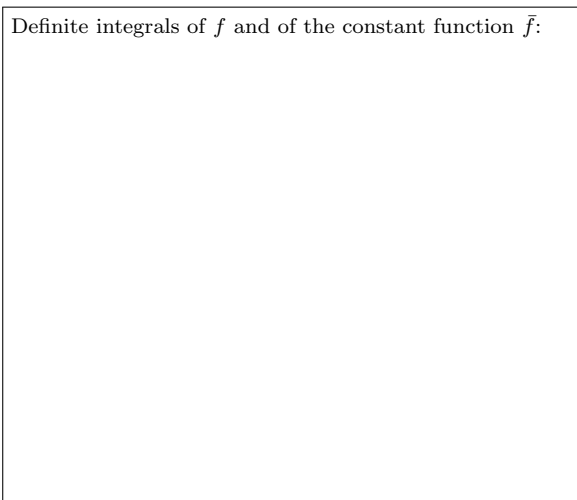
Note that this is similar for the average of numbers – if Alice has an average of $\bar{m}$ on all the tests in a module, and Bob scored exactly $\bar{m}$ each time, then they have the same total mark.

Theorem 3.8 (Mean Value Theorem (MVT)). Let $f : [a, b] \to \mathbb{R}$ be continuous. Then there exists a point $c \in (a, b)$ such that

$$f(c) = \frac{1}{b - a} \int_a^b f(x) \, dx.$$

That is, f attains its mean value at some point in (a, b).

Proof. Let

$$M := \max_{x \in [a,b]} f(x), \qquad m := \min_{x \in [a,b]} f(x).$$

Since f is continuous, it takes every value between m and M. Therefore, we need to show that

$$m \leq \bar{f} \leq M. \tag{3.2}$$

We have

$$M - f(x) \geq 0 \qquad \text{on } [a, b]$$

$$\stackrel{\text{prop. (6)}}{\Longrightarrow} \quad \int_a^b M - f(x)\,\mathrm{d}x \geq 0$$

$$\stackrel{\text{prop. (3)}}{\Longrightarrow} \quad \int_a^b M\,\mathrm{d}x - \int_a^b f(x)\,\mathrm{d}x \geq 0$$

$$\Longrightarrow \quad \int_a^b f(x)\,\mathrm{d}x \leq \int_a^b M\,\mathrm{d}x = M(b - a).$$

Combining this with the outcome of the analogous computation for $f(x) - m$ gives

$$m(b - a) \leq \int_a^b f(x)\,\mathrm{d}x \leq M(b - a).$$

Division by $(b - a)$ gives the inequality 3.2, and hence completes the proof. $\qquad\square$

Definition 3.9 (Area Function). For continuous $f : [a, b] \to \mathbb{R}$, the *area function*

$$A : [a, b] \to \mathbb{R}$$

of f is

$$A(x) := \int_a^x f(t)\,\mathrm{d}t.$$

Remark 3.10. We use the variable t in the integrand to avoid confusion with the variable of A.

The area function for a continuous function f:

Theorem 3.11 (Fundamental Theorem of Calculus (FTC)). Let $f : [a, b] \to \mathbb{R}$ be continuous.

(i) The derivative of the area function is

$$A'(x) = \frac{\mathrm{d}}{\mathrm{d}x}A(x) = f(x).$$

That is,

$$\frac{\mathrm{d}}{\mathrm{d}x}\int_a^x f(t)\,\mathrm{d}t = f(x).$$

88

(ii) If F is any *antiderivative* of f, i.e. a function with $F'(x) = f(x)$, then

$$\int_a^b f(x)\,\mathrm{d}x = F(b) - F(a).$$

Example 3.12. We solve the following definite integral by guessing a function whose derivative is the integrand. Once this guessed antiderivative is written down, one should always check whether it really differentiates to the original integrand.

$$\int_1^3 (x^3 - 6x)\,\mathrm{d}x = \left(\frac{x^4}{4} - 3x^2 \,\bigg|_1^3\right) \qquad \left[\text{check}: \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{x^4}{4} - 3x^2\right) = x^3 - 6x \quad \checkmark\right]$$

$$= \left(\frac{3^4}{4} - 3 \cdot 3^2\right) - \left(\frac{1^4}{4} - 3 \cdot 1^2\right) = \frac{81}{4} - 27 - \frac{1}{4} + 3 = -4$$

The vertical line in the second expression stands for "evaluate at $x = 3$ and then subtract the evaluation at $x = 1$" – as on the right-hand side of FTC (ii). To foreshadow the proof of the FTC, note that the choice of antiderivative is not unique, but different choices seem to be leading to the same result; e.g.,

$$\int_1^3 x^3 - 6x\,\mathrm{d}x = \left(\frac{x^4}{4} - 3x^2 + 42 \,\bigg|_1^3\right) = \frac{81}{4} - 27 + 42 - \frac{1}{4} + 3 - 42 = -4.$$

Proof. (i) Using the definition of the derivative, the definition of the area function, and property (5) of the definite integral, we obtain

$$A'(x) = \lim_{h \to 0} \frac{A(x+h) - A(x)}{h} = \lim_{h \to 0} \frac{1}{h} \int_x^{x+h} f(t)\,\mathrm{d}t.$$

The expression within the limit on the right is the average of f over the interval $[x, x+h]$. By the MVT, there exist $c_h \in [x, x+h]$ such that

$$f(c_h) = \frac{1}{h} \int_x^{x+h} f(t)\,\mathrm{d}t.$$

This allows to continue the computation of $A'(x)$ above as follows:

$$A'(x) = \lim_{h \to 0} f(c_h) \qquad \text{where } c_h \in [x, x+h]$$

$$= f\left(\lim_{h \to 0} c_h\right) = f(x).$$

(ii) The area function $A(x)$ is an antiderivative of f – by (i) – and for it, the claim is true by definition:

$$A(b) - A(a) = A(b) = \int_a^b f(t)\,\mathrm{d}t.$$

Now let $F(t)$ be a different antiderivative of f. Then

$$\frac{\mathrm{d}}{\mathrm{d}x}(F(x) - A(x)) = f(x) - f(x) = 0,$$

and therefore F and A differ only by a constant,

$$F(x) = A(x) + c.$$

This gives

$$F(b) - F(a) = (A(b) + c) - (A(a) + c) = A(b) - A(a) = \int_a^b f(t)\,\mathrm{d}t.$$

$\square$

Corollary 3.13. (i)
$$\int_a^b F'(t)\,\mathrm{d}t = F(b) - F(a).$$

(ii)
$$\frac{\mathrm{d}}{\mathrm{d}x} \int_{a(x)}^{b(x)} f(t)\,\mathrm{d}t = f(b(x))\,b'(x) - f(a(x))\,a'(x).$$

Remark 3.14. (i) The word "corollary" is usually used for results that follow from important theorems.

(ii) The first identity is also called the "Total Change Theorem", and it follows immediately from FTC (ii). It reads: "Integrating the rate of change of a function over an interval gives the total change over that interval."

(iii) Make sure to take note of the second identity – it is important for differentiating more complicated integral expressions. It follows from the chain rule:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x} \int_c^{b(x)} f(t)\,\mathrm{d}t &= \frac{\mathrm{d}}{\mathrm{d}x} \int_c^y f(t)\,\mathrm{d}t && (y = b(x)) \\
&= \frac{\mathrm{d}}{\mathrm{d}y} \int_c^y f(t)\,\mathrm{d}t \cdot \frac{\mathrm{d}y}{\mathrm{d}x} && \text{(chain rule)} \\
&= f(y) \cdot \frac{\mathrm{d}y}{\mathrm{d}x} = f(b(x))\,b'(x),
\end{aligned}$$

and using property (5) of the definite integral, we can prove the formula in full generality, that is, for the case when the lower bound is a function of x as well.

Definition 3.15 (Indefinite Integral)**.** The *indefinite integral*

$$\int f(x)\,\mathrm{d}x$$

of $f(x)$ is the collection of all antiderivatives of the function f.

Example 3.16. (i) For
$$F(x) = \int_1^{x^3} t^5\,\mathrm{d}t,$$

we find, identifying $f(t) = t^5$ and using the formula from theorem 3.13,

$$F'(x) = \frac{\mathrm{d}}{\mathrm{d}x} \int_1^{x^3} f(t)\,\mathrm{d}t = f(x^3)\cdot\frac{\mathrm{d}}{\mathrm{d}x}\left(x^3\right) - f(1)\cdot\frac{\mathrm{d}}{\mathrm{d}x}(1) = \left(x^3\right)^5\cdot 3x^2 + 0 = 3x^{17}.$$

We can check this by finding $F(x)$ explicitly,

$$F(x) = \int_1^{x^3} t^5\,\mathrm{d}t = \left(\frac{t^6}{6}\Big|_1^{x^3}\right) = \frac{(x^3)^6}{6} - \frac{1}{6},$$

and then differentiating:

$$F'(x) = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{x^{18}}{6} - \frac{1}{6}\right) = 18\cdot\frac{x^{17}}{6} - 0 = 3x^{17} \quad \checkmark$$

(ii)

$$\int e^{3x}\,\mathrm{d}x = \frac{1}{3}e^{3x} + c.$$

Here, we use the constant c as a place holder to cover all the infinitely many possible antiderivatives, e.g.

$$\frac{1}{3}e^{3x}, \quad \frac{1}{3}e^{3x} - 5, \quad \frac{1}{3}e^{3x} + 117.38, \quad \ldots,$$

of $f(x) = e^{3x}$.

Remark 3.17. We conclude this section with the following remarks:

(i) Integration is "the opposite of differentiation".

(ii) Integrals with boundaries are definite integrals, and then the result is a real number.

(iii) Integrals without boundaries are indefinite integrals, and then the result is a function! Make sure to not forget the constant of integration!

Exercise 3.18. (i) Work through all the material from the lectures and the practice problems on the tutorial sheet that corresponding to this section. Interpret the properties 3.4 – perhaps add a few sketches to your notes – and prove one or two of them.

(ii) Find the area function A_0 of $f(x) = x^3 - 6x$ with $a = 0$. Then set $a = 1$ and find the corresponding area function A_1. Compare A_0 and A_1 to each other and to the indefinite integral of[65] $f(x)$.

(iii) Find the definite integral

$$\int_0^1 e^x\,\mathrm{d}x$$

using Riemann sums[66], and then compare to the value found via direct integration.

91

(iv) Show that the definite integral, (3.1), is well-defined[67].

(v) Find the derivative of[68]

$$F(x) = \int_0^{x^2} x \, dt.$$

3.2 Methods of Integration

3.2.1 Basic Integrals

Properties 3.19 (Basic Integrals).

$$(1) \qquad \int x^r \, dx = \frac{x^{r+1}}{r+1} + c \qquad \text{for } r \neq -1,$$

$$(2) \qquad \int \frac{1}{x} \, dx = \ln|x| + c,$$

$$(3) \qquad \int e^x \, dx = e^x + c,$$

$$(4) \qquad \int a^x \, dx = \frac{a^x}{\ln a} + c \qquad \text{for } a > 0,$$

$$(5) \qquad \int \cos x \, dx = \sin x + c,$$

$$(6) \qquad \int \sin x \, dx = -\cos x + c,$$

$$(7) \qquad \int \frac{1}{\cos^2 x} \, dx = \tan x + c,$$

$$(8) \qquad \int \frac{1}{\sin^2 x} \, dx = -\cot x + c,$$

$$(9) \qquad \int \frac{1}{1 + x^2} \, dx = \arctan x + c,$$

$$(10) \qquad \int \frac{1}{\sqrt{1 - x^2}} \, dx = \arcsin x + c,$$

$$(11) \qquad \int \cosh x \, dx = \sinh x + c,$$

$$(12) \qquad \int \sinh x \, dx = \cosh x + c.$$

Remark 3.20. (i) The function $\arctan x$ is the inverse function of $\tan x$. It is sometimes denoted $\tan^{-1}$, but this carries potential for confusion, as $\arctan x \neq {}^1/_{\tan x}$! Similarly, for $\cos x$ and $\sin x$ and their inverse functions. The expressions appearing in (7) and (8) could be rewritten using the conventions

$$\sec x = \frac{1}{\cos x}, \quad \csc x = \frac{1}{\sin x}, \quad \cot x = \frac{\cos x}{\sin x}.$$

(ii) The linearity properties (3) and (4) of 3.4 remain true for indefinite integrals. Therefore, we can now integrate all linear combinations of functions appearing in the integrands of properties 3.19.

(iii) All the integrals above are verified by differentiating the antiderivative on the right and comparing to the integrand on the left. For example,

$$\frac{\mathrm{d}}{\mathrm{d}x}(-\cos x + c) = \sin x,$$

and hence (6) is correct.

(iv) Formula (2) can be used to evaluate definite integrals such as $\int_1^2 \frac{1}{x}\,\mathrm{d}x$ or $\int_{-23}^{-5} \frac{1}{x}\,\mathrm{d}x$, but definite integrals across $x = 0$ are not permitted since $f(x) = \frac{1}{x}$ is not defined at $x = 0$. This means that in this context, x is either always positive or always negative. This allows to verify (2) by considering two cases:

Case (a): $x > 0$
If x is positive, then $|x| = x$ and

$$\frac{\mathrm{d}}{\mathrm{d}x}(\ln|x| + c) = \frac{\mathrm{d}}{\mathrm{d}x}\ln x = \frac{1}{x} \qquad \checkmark$$

Case (b): $x < 0$
If x is negative, then $|x| = -x$ and we obtain

$$\frac{\mathrm{d}}{\mathrm{d}x}(\ln|x| + c) = \frac{\mathrm{d}}{\mathrm{d}x}\ln(-x) = \frac{1}{-x} \cdot (-1) = \frac{1}{x}$$

as well.

Example 3.21. (i)

$$\int x^{11} + 5\cos x - 7^x\,\mathrm{d}x = \int x^{11}\,\mathrm{d}x + 5\int \cos x\,\mathrm{d}x - \int 7^x\,\mathrm{d}x$$
$$= \frac{x^{12}}{12} + 5\sin x - \frac{7^x}{\ln 7} + c.$$

(ii)

$$\int_1^e \frac{1}{x}\,\mathrm{d}x = \ln|x|\,\big|_1^e = \ln e - \ln 1 = 1,$$

which some calculus text books use to define the constant $e \approx 2.718$.

Remark 3.22. In the following example, recognising an expression coming from the chain rule allows us to "guess" an integral:

$$\int 2x\,\cosh(x^2)\,\mathrm{d}x = \sinh(x^2) + c.$$

The next theorem makes the application of this idea more systematic.

3.2.2 Substitution

Theorem 3.23 (Substitution)**.** If $u = u(x)$ is differentiable and f continuous, then

(i)
$$\int f(u(x)) \cdot u'(x) \, dx = \int f(u) \, du.$$

(ii)
$$\int_a^b f(u(x)) \cdot u'(x) \, dx = \int_{u(a)}^{u(b)} f(u) \, du.$$

Example 3.24. (i)

$$\int x \left(1 + x^2\right)^{13} \, dx \qquad \left[\begin{array}{rcl} u & = & 1 + x^2 \\ 1 \cdot du & = & 2x \cdot dx \\ \to \ x\, dx & = & {}^1\!/_2 \, du \end{array}\right]$$

$$= \int u^{13} \frac{1}{2} \, du$$

$$= \frac{1}{2} \frac{u^{14}}{14} + c = \frac{\left(1 + x^2\right)^{14}}{28} + c.$$

Check:

$$\frac{d}{dx}\left(\frac{\left(1 + x^2\right)^{14}}{28} + c\right) = \frac{14}{28}\left(1 + x^2\right)^{14-1} \cdot 2x = x\left(1 + x^2\right)^{13} \qquad \checkmark$$

(ii)

$$\int_0^{\pi/2} \cos(\cos x) \sin x \, dx \qquad \left[\begin{array}{rcl} u & = & \cos x \\ du & = & -\sin x \, dx \\ \to \ \sin x \, dx & = & -du \\ u = 0 & \leftrightarrow & x = {}^\pi\!/_2 \\ u = 1 & \leftrightarrow & x = 0 \end{array}\right]$$

$$= \int_1^0 \cos u \cdot (-1) \, du$$

$$= \int_0^1 \cos u \, du = \sin u \,\big|_0^1 = \sin 1.$$

(iii)

$$I = \int_0^{\sqrt{3}} e^{\sqrt{1+x^2}} \frac{x}{\sqrt{1 + x^2}} \, dx \qquad \left[\begin{array}{rcl} u & = & \sqrt{1 + x^2} \\ du & = & \frac{x}{\sqrt{1+x^2}} dx \\ u = 2 & \leftrightarrow & x = \sqrt{3} \\ u = 1 & \leftrightarrow & x = 0 \end{array}\right]$$

$$= \int_1^2 e^u \, du$$

$$= e^u \,\big|_1^2 = e^2 - e^1 = e(e - 1).$$

Alternatively, one could leave the boundaries in terms of x and evaluate the antiderivative after substituting back to x:

$$I = \mathrm{e}^u \, \big| \,\vdots\, = \mathrm{e}^{\sqrt{1+x^2}} \, \Big|_0^{\sqrt{3}} = \mathrm{e}^{\sqrt{1+3}} - \mathrm{e}^{\sqrt{1}} = \mathrm{e}(\mathrm{e} - 1).$$

Remark 3.25. In the integrands of the previous examples, we have seen the inner functions

$$1 + x^2, \quad \cos x, \quad \sqrt{1 + x^2},$$

to which the outer functions

$$x^{13}, \quad \cos x, \quad \mathrm{e}^x,$$

respectively, are applied. A first attempt for solving integrals of that kind should always be to substitute for the identified inner function. The question is then whether the other factors in the integrand are absorbed by the transformation of the differential $\mathrm{d}\cdot$. If not, then more work or a different approach is required.

Example 3.26. Compute the indefinite integral

$$I = \int \frac{1}{4 + x^2} \, \mathrm{d}x.$$

Note that we could read off the answer immediately from the list in properties 3.19 if the 4 in the denominator was a 1. We therefore bring the integral in that form using substitution:

$$
\begin{aligned}
I = \int \frac{1}{4 + x^2} \, \mathrm{d}x &= \int \frac{1}{4\left(1 + {x^2}/{4}\right)} \, \mathrm{d}x \\
&= \frac{1}{4} \int \frac{1}{1 + \left({x}/{2}\right)^2} \, \mathrm{d}x
&& \begin{bmatrix} u &=& {x}/{2} \\ \mathrm{d}u &=& {1}/{2}\,\mathrm{d}x \end{bmatrix} \\
&= \frac{1}{2} \int \frac{1}{1 + u^2} \, \mathrm{d}u \\
&= \frac{1}{2} \arctan u + c = \frac{1}{2} \arctan\left(\frac{x}{2}\right) + c.
\end{aligned}
$$

3.2.3 Trigonometric Identities

Properties 3.27 (Trigonometric Identities).

$$
\begin{aligned}
&(1) && \cos^2 x + \sin^2 x = 1, \\
&(2) && \cos(x + y) = \cos x \cos y - \sin x \sin y, \\
&(3) && \sin(x + y) = \sin x \cos y + \cos x \sin y, \\
&(4) && \cos(2x) = 2\cos^2 x - 1 = 1 - 2\sin^2 x, \\
&(5) && \sin(2x) = 2 \sin x \cos x, \\
&(6) && 2 \sin x \cos y = \sin(x - y) + \sin(x + y), \\
&(7) && 2 \cos x \cos y = \cos(x - y) + \cos(x + y), \\
&(8) && 2 \sin x \sin y = \cos(x - y) - \cos(x + y), \\
&(9) && \cosh^2 x - \sinh^2 x = 1,
\end{aligned}
$$

Example 3.28.   (i)

$$\int_0^\pi \sin^2 x \, dx = \int_0^\pi \sin x \cdot \sin x \, dx \overset{(8)}{=} \int_0^\pi \frac{1}{2} \left[\cos(x - x) - \cos(x + x) \right] \, dx$$

$$= \int_0^\pi \frac{1}{2} \left[\cos(0) - \cos(2x) \right] \, dx = \frac{1}{2} \int_0^\pi 1 - \cos(2x) \, dx$$

$$= \frac{1}{2} \left(x - \tfrac{1}{2} \sin(2x) \, \big|_0^\pi \right) = \frac{\pi}{2}.$$

A better way to compute this integral would be to use the identity (4) – this allows to go directly from the first expression to the fifth.

(ii)

$$\int \cos^4 x \, dx = \int \left(\cos^2 x \right)^2 \, dx \overset{(4)}{=} \int \left(\frac{\cos(2x) + 1}{2} \right)^2 \, dx$$

$$= \int \frac{1}{4} \left(\cos^2(2x) + 2\cos(2x) + 1 \right) \, dx$$

$$\overset{(4)}{=} \int \frac{1}{4} \left(\frac{\cos(4x) + 1}{2} + 2\cos(2x) + 1 \right) \, dx$$

$$= \frac{1}{32} \sin(4x) + \frac{1}{4} \sin(2x) + \frac{3}{8} x + c.$$

(iii)

$$\int \sin(2x) \sin(7x) \, dx \overset{(8)}{=} \int \frac{1}{2} \left[\cos(-5x) - \cos(9x) \right] \, dx = \frac{\sin(5x)}{10} - \frac{\sin(9x)}{18} + c.$$

(iv) The following example contains a typical substitution with trigonometric functions, and some comments on it will be made in the next remark.

$$\int_{2.5}^5 \frac{\sqrt{25 - x^2}}{x^2} \, dx \qquad \left[\begin{array}{rcl} x & = & 5\sin\theta \\ dx & = & 5\cos\theta \, d\theta \\ x = 5 & \leftrightarrow & \theta = \pi/2 \\ x = 2.5 & \leftrightarrow & \theta = \pi/6 \end{array} \right]$$

$$= \int_{\pi/6}^{\pi/2} \frac{\sqrt{25 - (5\sin\theta)^2}}{(5\sin\theta)^2} \, 5\cos\theta \, d\theta$$

$$= \int_{\pi/6}^{\pi/2} \frac{\sqrt{1 - \sin^2\theta}}{\sin^2\theta} \cos\theta \, d\theta \overset{(1)}{=} \int_{\pi/6}^{\pi/2} \frac{\cos^2\theta}{\sin^2\theta} \, d\theta$$

$$\overset{(1)}{=} \int_{\pi/6}^{\pi/2} \frac{1 - \sin^2\theta}{\sin^2\theta} \, d\theta = \int_{\pi/6}^{\pi/2} \frac{1}{\sin^2\theta} - 1 \, d\theta$$

$$= -\cot\theta - \theta \, \big|_{\pi/6}^{\pi/2} = -0 - \frac{\pi}{2} + \frac{\sqrt{3}/2}{1/2} + \frac{\pi}{6} = \sqrt{3} - \frac{\pi}{3}.$$

(v)

$$\int \frac{1}{\sqrt{x^2 - a^2}} \, \mathrm{d}x \qquad \left[\begin{array}{ccc} x & = & a \cosh t \\ \mathrm{d}x & = & a \sinh t \, \mathrm{d}t \end{array} \right]$$

$$= \int \frac{a \sinh t}{\sqrt{(a \cosh t)^2 - a^2}} \, \mathrm{d}t$$

$$\overset{(9)}{=} \int 1 \, \mathrm{d}t = t + c = \operatorname{arccosh}\left(\frac{x}{a}\right) + c.$$

Remark 3.29. (i) Note the similarities in the approaches for (iv) and (v) above, and also note the differences: In both cases, we have square roots of expressions $\pm(a^2 - x^2)$ in the integrand, and we use the trigonometric Pythagoras formula (1) of properties 3.27 and the corresponding hyperbolic version (9) to simplify them. The following thoughts might help you to avoid confusion of the two. For $\sqrt{a^2 - x^2}$, we need x to be small, namely in the range $[0, a]$, and the trigonometric functions have a bounded range. For $\sqrt{x^2 - a^2}$, we need x to be large, and the function $g(t) = a \cosh t$ has the correct range, $R(g) = [a, \infty)$.

(ii) For our next integration method, we apply the product rule to differentiate the product fg, and then bring one of the terms on the other side,

$$f'g = (fg)' - fg'.$$

Integrating this expression with respect to x, we obtain the following result.

3.2.4 Integration by Parts

Theorem 3.30 (Integration by Parts)**.**

$$\int f'(x)g(x) \, \mathrm{d}x = f(x)g(x) - \int f(x)g'(x) \, \mathrm{d}x.$$

Example 3.31. (i) The integration by parts formula allows us to find an import integral that was missing in properties 3.19 – the integral of the logarithm:

$$\int \ln x \, \mathrm{d}x = \int \underbrace{1}_{\text{int.}} \cdot \underbrace{\ln x}_{\text{diff.}} \, \mathrm{d}x = x \ln x - \int x \cdot \frac{\mathrm{d}}{\mathrm{d}x}(\ln x) \, \mathrm{d}x$$

$$= x \ln x - \int x \cdot \frac{1}{x} \, \mathrm{d}x = x \ln x - \int 1 \, \mathrm{d}x = x \ln x - x + c.$$

We have integrated $\ln x$ by differentiating it!

(ii)

$$\int \underbrace{x^2}_{\text{diff.}} \underbrace{e^{3x}}_{\text{int.}} \, \mathrm{d}x = \frac{1}{3}e^{3x}x^2 - \int \frac{1}{3}e^{3x}2x \, \mathrm{d}x = \frac{1}{3}e^{3x}x^2 - \frac{2}{3}\left[\int \underbrace{e^{3x}}_{\text{int.}} \underbrace{x}_{\text{diff.}} \, \mathrm{d}x\right]$$

$$= \frac{1}{3}e^{3x}x^2 - \frac{2}{3}\left[\frac{1}{3}e^{3x}x - \int \frac{1}{3}e^{3x} \, \mathrm{d}x\right]$$

$$= \frac{1}{3}e^{3x}x^2 - \frac{2}{3}\left[\frac{1}{3}e^{3x}x - \frac{1}{9}e^{3x} + \tilde{c}\right] = \frac{1}{3}e^{3x}\left[x^2 - \frac{2}{3}x + \frac{2}{9}\right] + c.$$

(iii) When applying integration by parts to definite integrals, all terms need to be evaluated over the interval; e.g.,

$$\int_0^1 \underbrace{x}_{\text{diff.}} \underbrace{(1+x)^{17}}_{\text{int.}} \, \mathrm{d}x = x \frac{(1+x)^{18}}{18} \Big|_0^1 - \int_0^1 \frac{(1+x)^{18}}{18} \, \mathrm{d}x$$

$$= 1 \cdot \frac{2^{18}}{18} - 0 - \frac{1}{18 \cdot 19} \left((1+x)^{19} \Big|_0^1 \right)$$

$$= \frac{1}{18 \cdot 19} \left(19 \cdot 2^{18} - 2^{19} + 1 \right) = \frac{495161}{38}.$$

(iv)

$$I = \int \underbrace{\mathrm{e}^x}_{\text{int.}} \underbrace{\cos x}_{\text{diff.}} \, \mathrm{d}x = \mathrm{e}^x \cos x - \int \mathrm{e}^x (-\sin x) \, \mathrm{d}x$$

$$= \mathrm{e}^x \cos x + \int \underbrace{\mathrm{e}^x}_{\text{int.}} \underbrace{\sin x}_{\text{diff.}} \, \mathrm{d}x$$

$$= \mathrm{e}^x \cos x + \mathrm{e}^x \sin x - \int \mathrm{e}^x \cos x \, \mathrm{d}x$$

$$= \mathrm{e}^x \cos x + \mathrm{e}^x \sin x - I$$

This computation does not seem to lead anywhere, as the integral we need to compute reappears after two applications of the integration by parts rule. Differentiating e^x and integrating the trigonometric term instead, leads to a similar situation. However, we can solve the equation for I ! This gives

$$I = \int \mathrm{e}^x \cos x \, \mathrm{d}x = \frac{1}{2} \mathrm{e}^x (\cos x + \sin x) + c.$$

Remark 3.32. (i) When faced with an integrand that contains a power of x and other, more complicated terms, one should first try substitution. If this does not work, integration by parts is the next best option. In this case, one would usually choose the power of x as the term that is to be differentiated – the reason for this is that differentiation makes powers simpler and repeated integration by parts will eventually dispose of them. However, there are no firm rules for integration and one should remain open-minded to non-standard approaches.

(ii) Many of the examples above are specifically designed so that the computations work out relatively smoothly. In general, integration is quite hard and sometimes even impossible. It is therefore important to keep in mind that differentiation provides a straight-forward way to verify the integrals you have found. Here is an integral that cannot be solved:

$$\int \mathrm{e}^{-x^2} \, \mathrm{d}x.$$

(iii) Consider the computation

$$\frac{2}{x+4} - \frac{3}{x-1} = \frac{2(x-1) - 3(x+4)}{(x+4)(x-1)} = \frac{-(x+14)}{x^2+3x-4},$$

and note that, while we can integrate the expression on the left-hand side – e.g.,

$$\int \frac{2}{x+4} \, dx = 2\ln(x+4) + c$$

– the expression on the right is not covered by the integration methods we have developed so far. Reversing finding the common denominator therefore extends the set of functions we can integrate. This is called integration by partial fractions.

3.2.5 Partial Fractions

Example 3.33. (i) Find the integral

$$I = \int \frac{2x+16}{x^2+2x-35} \, dx.$$

Sol.:

$$\frac{2x+16}{x^2+2x-35} = \frac{2x+16}{(x-5)(x+7)} = \frac{A}{x-5} + \frac{B}{x+7}$$

$$= \frac{A(x+7) + B(x-5)}{(x-5)(x+7)} = \frac{(A+B)x + (7A-5B)}{(x-5)(x+7)}$$

$$\implies \begin{cases} A+B &= 2 \\ 7A - 5B &= 16 \end{cases} \implies \begin{cases} A &= {}^{13}/_6 \\ B &= {}^{-1}/_6 \end{cases}$$

$$\implies \quad I = \int \frac{2x+16}{x^2+2x-35} \, dx = \int \frac{{}^{13}/_6}{x-5} + \frac{{}^{-1}/_6}{x+7} \, dx$$

$$= \frac{13}{6} \int \frac{1}{x-5} \, dx - \frac{1}{6} \int \frac{1}{x+7} \, dx = \frac{13}{6} \ln|x-5| - \frac{1}{6} \ln|x+7| + c.$$

(ii) Integrate

$$f(x) = \frac{x^4 - x^3 - 3x^2 + x - 2}{x^3 + 9x}$$

over the interval $[1,2]$.

Sol.: Before we can "split up" the fraction as above, we have to bring it in a form in which the degree of the numerator is smaller than the degree of the denominator. The idea for doing that is

$$\frac{x^4 - x^3 - 3x^2 + x - 2}{x^3 + 9x} = \frac{x^4 \boxed{+9x^2 - 9x^2} - x^3 - 3x^2 + x - 2}{x^3 + 9x}$$

$$= \frac{x^4 + 9x^2}{x^3 + 9x} + \frac{-9x^2 - x^3 - 3x^2 + x - 2}{x^3 + 9x}$$

$$= x + \frac{-x^3 - 12x^2 + x - 2}{x^3 + 9x} = \dots$$

99

This is long division of polynomials, and it is usually written out more systematically as

$$
\begin{array}{rrrrrrl}
(& x^4 & -x^3 & -3x^2 & +x & -2 &) & : & (x^3 + 9x) & = & x - 1 + \frac{-12x^2 + 10x - 2}{x^3 + 9x}. \\
-(& x^4 & & +9x^2 & & &) \\
& (& -x^3 & -12x^2 & +x & -2 &) \\
& -(& -x^3 & & -9x & &) \\
& & (& -12x^2 & +10x & -2 &)
\end{array}
$$

Next, we split up the remaining fraction, as in the example above:

$$
\begin{aligned}
\frac{-12x^2 + 10x - 2}{x^3 + 9x} &= \frac{-12x^2 + 10x - 2}{x(x^2 + 9)} = \frac{A}{x} + \frac{Bx + C}{x^2 + 9} \\
&= \frac{A(x^2 + 9) + (Bx + C)x}{x(x^2 + 9)} = \frac{(A + B)x^2 + Cx + 9A}{x(x^2 + 9)}.
\end{aligned}
$$

General guidelines for how to make the ansatz involving the constants A, B, C in the middle expression will be given in the remark below. The next step is to compare coefficients in the numerator on the left and on the right:

$$
\begin{cases}
A + B & = -12 \\
C & = 10 \\
9A & = -2
\end{cases}
\quad \Longrightarrow \quad
\begin{cases}
A & = {}^{-2}\!/_9 \\
B & = {}^{-106}\!/_9 \\
C & = 10.
\end{cases}
$$

We can now write the definite integral as a sum of simpler integrals:

$$
\begin{aligned}
I &= \int_1^2 \frac{x^4 - x^3 - 3x^2 + x - 2}{x^3 + 9x} \, \mathrm{d}x = \int_1^2 x - 1 + \frac{{}^{-2}\!/_9}{x} + \frac{{}^{-106}\!/_9 \, x + 10}{x^2 + 9} \, \mathrm{d}x \\
&= \int_1^2 x \, \mathrm{d}x - \int_1^2 1 \, \mathrm{d}x - \frac{2}{9} \int_1^2 \frac{1}{x} \, \mathrm{d}x + \int_1^2 \frac{{}^{-106}\!/_9 \, x + 10}{x^2 + 9} \, \mathrm{d}x.
\end{aligned}
$$

The first integrals are basic integrals, denote them I_1, I_2, I_3. The fourth we split up further as

$$
I = I_1 + I_2 + I_3 - \frac{53}{9} \int_1^2 \frac{2x}{x^2 + 9} \, \mathrm{d}x + 10 \int_1^2 \frac{1}{x^2 + 9} \, \mathrm{d}x,
$$

so that the fourth integral can be computed with a straightforward substitution, and the last one as in example 3.26. Computing the five definite integrals and adding them together gives the answer,

$$
I = \frac{1}{2} - \frac{2}{9} \ln(2) - \frac{53}{9} \ln(1.3) + \frac{10}{3} \left(\arctan({}^2\!/_3) - \arctan({}^1\!/_3) \right) \approx -0.312.
$$

(iii) Find the integral

$$
I = \int \frac{x - {}^3\!/_2}{x^2 - 3x + 7} \, \mathrm{d}x.
$$

Sol.: Flexibility is key for integration – this is a substitution problem:

$$I = \frac{1}{2} \int \frac{2x - 3}{x^2 - 3x + 7} \, \mathrm{d}x = \frac{1}{2} \int \frac{1}{u} \, \mathrm{d}u$$
$$= \frac{1}{2} \ln |u| + c = \frac{1}{2} \ln \left| x^2 - 3x + 7 \right| + c.$$

There are examples where both approaches – the substitution here in (iii), and the partial fraction approach in (i) – can be used, cf. the exercise below. However, for the problem at hand, the partial fraction approach would not work, as the denominator $x^2 - 3x + 7$ can not be factored.

Remark 3.34. The steps for integration by partial fractions are:

(1) If necessary, use polynomial division to transform to a polynomial plus a fraction in which the degree of the numerator is smaller than degree of the denominator.

(2) Write the denominator as a product of irreducible factors. Each of them will have degree at most two. In this module, we only consider the case where none of these factors is repeated.

(3) For each factor, define a fraction that has that factor in the denominator and a general polynomial of degree one less in the numerator. That is, if the denominator is a linear expression (polynomial of degree 1), then the numerator is just a constant c (polynomial of degree 0). If the denominator is a quadratic expression (polynomial of degree 2), then the numerator should be of the form $c_1 x + c_2$ (polynomial of degree 1).

(4) Make the ansatz that the sum of the fractions from the previous step is equal to the original fraction. This will lead to a system of equations for all the constants appearing in the numerators defined in the previous step. Solve it.

(5) In the computation of the integral, replace the original integrand with the sum of partial fractions you have found, plus possibly the polynomial that was obtained in step (1). Now use the linearity of the integral to obtain a combination of simple integrals that can be solved with the methods we have seen in this section. (In general, it is possible to obtain integrals at this point that can not be solved with the methods we have learned so far – but there will not be any such examples in the exam.)

Application (Signal conversion)**.** . . .

Exercise 3.35. (i) We recommend to practise integration intensively by working through a large number of examples. This is important as integration is a very fundamental skill for all branches of mathematics, engineering, and other sciences. Besides the material from the lectures and the tutorial sheet, you can look for more practice examples in books and online. You can even make up you own examples and check your results using differentiation. Mathematics software such as WolframAlpha can compute many integrals, but *do not rely on this.*

(ii) Make sure to complete all the steps in this section that were only sketched, e.g. the last lines of 3.33 (ii). Following the comment at the end of 3.33 (iii), compute

$$\int \frac{x+1}{x^2 + 2x - 3} \, \mathrm{d}x$$

in two different ways.

(iii) Using integration, find the area inside the circle of radius[69] R.

(iv) Find[70]

$$\int_{\pi/6}^{\pi/4} \frac{\cos 2x}{\cos^2 x \sin^2 x} \, \mathrm{d}x.$$

(v) Find[71]

$$\int (1 + |x|)^2 \, \mathrm{d}x.$$

3.3   Improper Integrals

Definition 3.36 (Improper Integral). A definite integral $\int_a^b f(x) \, \mathrm{d}x$ is called *improper* if

(I) the interval of integration is infinite, i.e. $a = -\infty$, or $b = \infty$, or both; or if

(II) the interval of integration contains a *singularity* of f, that is, a point where f is not defined, e.g. a zero of the denominator.

If $b = \infty$ (type I) or b is a singularity (type II), then

$$\int_a^b f(x) \, \mathrm{d}x := \lim_{\substack{t \to b \\ t < b}} \int_a^t f(x) \, \mathrm{d}x,$$

and the improper integral is said to exist / not exist depending on whether the limit on the right-hand side exists.

   Similarly if a is the boundary that causes the integral to be improper. If both boundaries are infinite (type I) or if the singularity lies inside the interval (type II), then the improper integral should be split up, cf. example 3.39 (i) below.

Example 3.37 (Type I).   (i)

$$\int_{-\infty}^0 \mathrm{e}^{2x} \, \mathrm{d}x = \lim_{R \to -\infty} \int_R^0 \mathrm{e}^{2x} \, \mathrm{d}x = \lim_{R \to -\infty} \frac{1}{2} \mathrm{e}^{2x} \Big|_R^0 = \lim_{R \to -\infty} \frac{\mathrm{e}^{2 \cdot 0} - \mathrm{e}^{2 \cdot R}}{2}$$

$$= \frac{1}{2} \qquad \Longrightarrow \text{the impr. int. does exist and is equal to } {}^1/_2.$$

(ii)

$$\int_1^\infty \frac{1}{x} \, \mathrm{d}x = \lim_{R \to \infty} \int_1^R \frac{1}{x} \, \mathrm{d}x = \lim_{R \to \infty} \ln x \Big|_1^R = \lim_{R \to \infty} \ln R - \ln 0$$

$$= \infty \qquad \Longrightarrow \text{the impr. int. does not exist.}$$

Remark 3.38. (i) The first example states that the area under the curve

$$f(x) = e^{2x} \qquad x \leq 0$$

is finite! Here, the word "area" refers to the total amount of area – as in "square footage" – rather than the set of points under that curve. The latter is an infinitely long "spike".

(ii) Improper integrals of type I are similar to series, and (ii) corresponds to the fact that the harmonic series $\sum \frac{1}{n}$ diverges.

(iii) Recalling the definition of the area function $A(x)$, we see that the limit that is to be found for an improper integral of a continuous function on $[a, \infty)$ is that of the area function:

$$\int_a^\infty f(x)\,\mathrm{d}x = \lim_{R \to \infty} A(R).$$

Example 3.39 (Type II). (i) Integrate $f(x) = \frac{1}{\sqrt[3]{x^2}}$ over the interval $[-1, +1]$.
Sol.:

$$\int_{-1}^{+1} \frac{1}{\sqrt[3]{x^2}}\,\mathrm{d}x = \int_{-1}^0 \frac{1}{\sqrt[3]{x^2}}\,\mathrm{d}x + \int_0^{+1} \frac{1}{\sqrt[3]{x^2}}\,\mathrm{d}x$$

$$= 2 \int_0^1 \frac{1}{\sqrt[3]{x^2}}\,\mathrm{d}x \qquad \text{(since } f \text{ is an even function)}$$

$$= 2 \lim_{\varepsilon \to 0} \int_\varepsilon^1 x^{-2/3}\,\mathrm{d}x = 2 \lim_{\varepsilon \to 0} 3x^{1/3}\Big|_\varepsilon^1 = 6 - 6 \lim_{\varepsilon \to 0} \varepsilon^{1/3} = 6.$$

The improper integral exists and is equal to 6.

(ii)

$$\int_0^{\pi/2} \frac{1}{\sin^2 x}\,\mathrm{d}x = \lim_{\varepsilon \to 0} \int_\varepsilon^{\pi/2} \frac{1}{\sin^2 x}\,\mathrm{d}x = \lim_{\varepsilon \to 0} \left(-\cot x\Big|_\varepsilon^{\pi/2} \right)$$

$$= \lim_{\varepsilon \to 0} \left(-\frac{\cos \pi/2}{\sin \pi/2} + \frac{\cos \varepsilon}{\sin \varepsilon} \right) = -0 + \lim_{\varepsilon \to 0} \frac{\cos \varepsilon}{\sin \varepsilon} = \infty.$$

Therefore, the improper integral does not exist.

Exercise 3.40. (i) As usual: Work through the material from the lectures and the improper integrals on the tutorial sheet.

(ii) For which powers $p \in \mathbb{R}$ do the integrals

$$\int_0^1 x^p\,\mathrm{d}x, \qquad \int_1^\infty x^p\,\mathrm{d}x$$

exist[72]?

(iii) Evaluate the improper integral[73]

$$I_n = \int_0^\infty x^n \mathrm{e}^{-x}\,\mathrm{d}x.$$

(iv) The goal of this exercise is to go through a quite difficult integration that consists of a number of steps: Evaluate the improper integral

$$\int_{1/2}^{3/2} \frac{1}{\sqrt{|x - x^2|}}\,\mathrm{d}x$$

by (1) considering two cases, (2) using the identities

$$x - x^2 = \frac{1}{4}\left[1 - (2x - 1)^2\right], \quad x^2 - x = \frac{1}{4}\left[(2x - 1)^2 - 1\right],$$

(3) substituting $u = 2x - 1$, and (4) using a basic integral from the table 3.19 and[74]

$$\int \frac{1}{\sqrt{t^2 - 1}}\,\mathrm{d}t = \ln\left(\sqrt{t^2 - 1} + t\right) + c.$$

3.4   Integrals of Functions of Several Variables

Example 3.41. As an introduction to this section, we compute a simple double-integral. That is done systematically starting from the innermost integral. In the following example, a $\mathrm{d}x$ integral is to be found first. For this step, the other variable, y, is treated like a constant, as it is not the variable with respect to which the current integration is carried out:

$$I = \int_0^1 \int_0^1 x + y \,\mathrm{d}x\,\mathrm{d}y = \int_0^1 \left[\int_0^1 x + y \,\mathrm{d}x\right]\mathrm{d}y$$

$$= \int_0^1 \left[\frac{x^2}{2} + xy\,\Big|_{x=0}^{x=1}\right]\mathrm{d}y = \int_0^1 \left[\frac{1}{2} + 1\cdot y - \frac{0^2}{2} - 0\cdot y\right]\mathrm{d}y$$

$$= \int_0^1 \frac{1}{2} + y \,\mathrm{d}y = \frac{y}{2} + \frac{y^2}{2}\,\Big|_0^1 = \frac{1}{2} + \frac{1}{2} - 0 - 0 = 1.$$

Remark 3.42.   (i) Just as the integral $\int_a^b f(x)\,\mathrm{d}x$ is the area between the curve $y = f(x)$ and the interval $[a, b]$ of the x-axis, the double integral

$$\int_a^b \int_c^d f(x, y)\,\mathrm{d}y\,\mathrm{d}x$$

is the volume between the surface $z = f(x, y)$ and the rectangle

$$[a, b] \times [c, d] = \{(x, y) \in \mathbb{R}^2 \mid a \le x \le b, c \le y \le d\}$$

of the xy-plane.

(ii) To check whether this interpretation agrees with the result $I = 1$ we have found above, note that the graph of the integrand $f(x, y) = x + y$ is a plane and has heights

$$f(0, 0) = 0, \quad f(0, 1) = 1, \quad f(1, 0) = 1, \quad f(1, 1) = 2,$$

over the corner points of the domain of integration, $D = [0, 1] \times [0, 1]$. This means that the volume that is to be found is that of a cuboid of size $1 \times 1 \times 2$ that is diagonally cut in half:

Sketch corresponding to example 3.41:
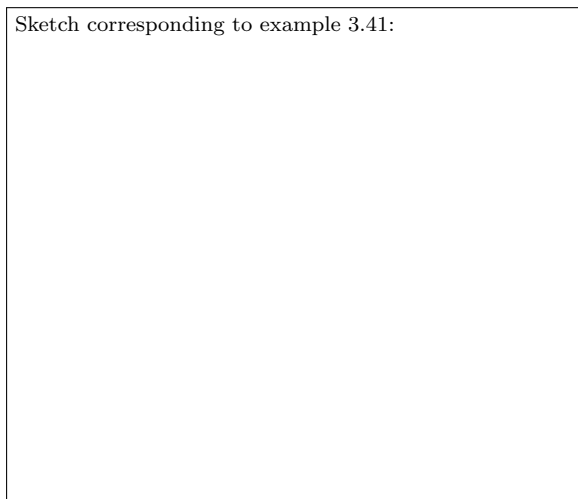
This solid has the volume

$$V = \frac{1 \cdot 1 \cdot 2}{2} = 1,$$

which we had also obtained with the integration above.

(iii) We write $\mathrm{d}A$ for $\mathrm{d}x\mathrm{d}y$,

$$\mathrm{d}A = \mathrm{d}x\mathrm{d}y,$$

meaning, roughly, that the change of area is equal to the change of x times the change of y. $\mathrm{d}A$ is called the *area element*. The order of the integrations $\mathrm{d}x$ and $\mathrm{d}y$ can be swapped, but one needs to be careful about the boundaries, cf. later examples.

(iv) There also is a formulation of Riemann sums for functions of several variables. We will not address this further in MTH1002, but the idea is as follows:

Riemann Sum of a 2D function:

Example 3.43. Integrate the function $f(x, y) = y/x$ over the domain $D = [3, 6] \times [1, 2]$.

Sol.:

$$I = \iint_D f \, dA = \int_1^2 \int_3^6 \frac{y}{x} \, dx \, dy = \int_1^2 \left[\int_3^6 \frac{y}{x} \, dx \right] \, dy.$$

The inside integral is with respect to x, and it therefore treats y like a constant – it is therefore permissible to pull out y,

$$I = \int_1^2 y \left[\int_3^6 \frac{1}{x} \, dx \right] \, dy = \int_1^2 y \left[\ln x |_3^6 \right] \, dy = \ln 2 \int_1^2 y \, dy = \ln \sqrt{8}.$$

Remark 3.44. All domains of integration so far have been rectangles. In this case, the x and y boundaries are constant, and the order of integration can be swapped easily – convince yourself of that by re-doing one of the problems above integrating with respect to y and then w.r.t. x. Next we study non-rectangular domains of integration, for which the boundaries of the inner integral depend on the outer variable.

Example 3.45.

$$I = \int_0^1 \int_0^{x^2} 1 \, dy \, dx = \int_0^1 \left[\int_0^{x^2} 1 \, dy \right] \, dx$$

$$= \int_0^1 \left(y|_0^{x^2} \right) \, dx = \int_0^1 \left(x^2 - 0 \right) \, dx = \int_0^1 x^2 \, dx = \frac{1}{3}.$$

Remark 3.46. (i) In the previous example, the domain of integration was

$$D = \{(x, y) \mid 0 \leq x \leq 1, \, 0 \leq y \leq x^2\}.$$

Integrating the function $f(x, y) = 1$ over a subset of $\mathbb{R}^2$ gives the area of that set. This is similar to 1D integration: integrating $f(x) = 1$ over an interval gives the length of that interval.

(ii) Our choice to integrate first w.r.t. y and then w.r.t. x in the previous example corresponds to the following steps: (1) for each $x \in [0, 1]$, integrate over each vertical line in the sketch, then (2) collect those values in the x direction.

106

Order of integration for example 3.45:

We could also find $I = \iint_D 1 \, \mathrm{d}A$ by integrating the other way around. For this, one has to solve the equations that define the boundaries for the other variable:

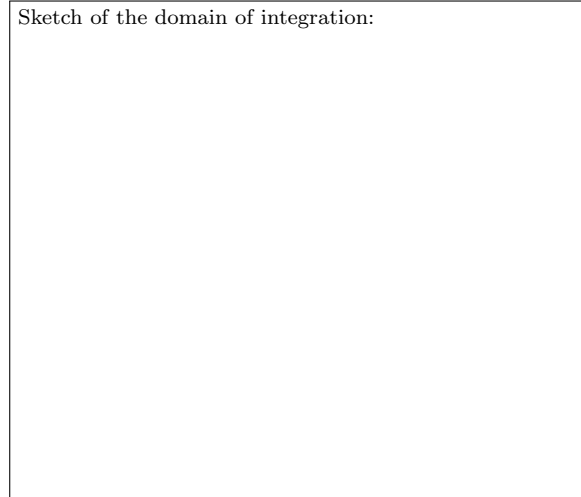$$I = \int_0^1 \int_{\sqrt{y}}^1 1 \, \mathrm{d}x \, \mathrm{d}y$$

– check that this gives the same result.

Illustration of integration in the other order:

Example 3.47. Integrate $f(x, y) = y \sin x$ over the triangle D with corner points $(0, 0)$, $(\pi, 0)$, and $(\pi, 1)$.

Sol.:

Sketch of the domain of integration:

$$\iint f \mathrm{d}A = \int_0^\pi \int_0^{x/\pi} y \sin x \, \mathrm{d}y \, \mathrm{d}x = \int_0^\pi \sin x \left[\int_0^{x/\pi} y \, \mathrm{d}y \right] \mathrm{d}x$$

$$= \int_0^\pi \sin x \left(\frac{y^2}{2} \Big|_0^{x/\pi} \right) \mathrm{d}x = \int_0^\pi \sin x \frac{x^2}{2\pi^2} \, \mathrm{d}x$$

$$= \frac{1}{2\pi^2} \int_0^\pi x^2 \cdot \sin x \, \mathrm{d}x = \cdots = \frac{\pi^2 - 4}{2\pi^2}.$$

Remark 3.48. It is helpful to always start a computation of a 2D integral with a sketch of the domain of integration. Labelling its boundaries with the formulas that describe them helps to correctly write out $\iint_D f(x, y) \, \mathrm{d}A$ as $\int \int f(x, y) \, \mathrm{d}y \, \mathrm{d}x$. The integration can be carried out in either order, but the next example and the corresponding exercise below show that for some domains, one order of integration is easier than the other.

Example 3.49. (i) Let D be the region bounded by the line $y = x + 1$ and by the parabola $y = x^2 - 1$. Find

$$I = \iint_D xy + 2 \, \mathrm{d}A.$$

Sol.: The domain of integration is the region between the line and the parabola below. We find the intersection points as follows:

$$x + 1 \overset{!}{=} x^2 - 1$$
$$\rightarrow \quad 0 = x^2 - x - 2$$
$$\implies \quad x_1 = -1, \quad x_2 = 2.$$

The domain of integration D:

Let us integrate with respect to x first and then w.r.t. y; that is, the outside integral is w.r.t. y. The sketch shows that y ranges from $y = -1$ to $y = 3$. Note that the right x boundary is $x = +\sqrt{y+1}$ for any $y \in [-1, 3]$. However, for the left x boundary, we have to distinguish two cases: $x = -\sqrt{y+1}$ for any $y \in [-1, 0]$, and $y = x + 1 \leftrightarrow x = y - 1$ for any $y \in [0, 3]$. We therefore split D along $y = 0$ to obtain

$$I = \iint_{D_-} xy + 2 \, dA + \iint_{D_+} xy + 2 \, dA$$

$$= \int_{-1}^{0} \int_{-\sqrt{y+1}}^{+\sqrt{y+1}} xy + 2 \, dx \, dy + \int_{0}^{3} \int_{y-1}^{+\sqrt{y+1}} xy + 2 \, dx \, dy$$

$$= \int_{-1}^{0} \left(\frac{x^2 y}{2} + 2x \Big|_{-\sqrt{y+1}}^{+\sqrt{y+1}} \right) dy + \int_{0}^{3} \left(\frac{x^2 y}{2} + 2x \Big|_{y-1}^{\sqrt{y+1}} \right) dy$$

$$= \int_{-1}^{0} \frac{(y+1)y}{2} + 2\sqrt{y+1} - \frac{(y+1)y}{2} + 2\sqrt{y+1} \, dy$$

$$+ \int_{0}^{3} \frac{(y+1)y}{2} + 2\sqrt{y+1} - \frac{(y-1)^2 y}{2} - 2y + 2 \, dy$$

$$= \int_{-1}^{0} 4\sqrt{y+1} \, dy + \int_{0}^{3} \frac{-y^3 + 3y^2 - 4y + 4}{2} + 2\sqrt{y+1} \, dy$$

$$= \frac{8}{3} \left((y+1)^{3/2} \Big|_{-1}^{0} \right) + \frac{1}{2} \left(-\frac{y^4}{4} + \frac{3y^3}{3} - \frac{4y^2}{2} + 4y \Big|_{0}^{3} \right) + \frac{4}{3} \left((y+1)^{3/2} \Big|_{0}^{3} \right)$$

$$= \left(\frac{8}{3} - 0 \right) + \frac{1}{2} \left(-\frac{3^4}{4} + 3^3 - 2 \cdot 3^2 + 4 \cdot 3 - 0 \right) + \frac{4}{3} \left(4^{3/2} - 1 \right)$$

$$= \frac{8}{3} + \frac{3}{8} + \frac{28}{3} = \frac{99}{8}.$$

(ii) For $D = [0, 2] \times [0, 1]$, find

$$I = \iint_{D} \left| (x + y)^2 - 1 \right| \, dA.$$

109

Sol.: Let $f(x, y) = (x + y)^2 - 1$. In order to integrate its *absolute value* over D, one first has to split D into the subset D_+ on which f is positive and the subset D_- where it is negative, and then use the definition of the absolute value,

$$|f| = \begin{cases} f, & \text{when } f \geq 0, \\ -f, & \text{when } f < 0. \end{cases}$$

The boundary between these subsets is the level set $f = 0$:

$$(x + y)^2 - 1 = 0 \quad \Longrightarrow \quad \begin{cases} y_1 & = 1 - x, \\ y_2 & = -1 - x, \end{cases}$$

but only y_1 intersects D. This gives

$$
\begin{aligned}
I &= \iint_{D_-} -\left[(x + y)^2 - 1\right] \, dA + \iint_{D_+} +\left[(x + y)^2 - 1\right] \, dA \\
&= \int_0^1 \int_0^{1-y} 1 - (x + y)^2 \, dx \, dy + \int_0^1 \int_{1-y}^2 (x + y)^2 - 1 \, dx \, dy \\
&= \int_0^1 \left(x - \frac{(x + y)^3}{3} \Big|_0^{1-y} \right) dy + \int_0^1 \left(\frac{(x + y)^3}{3} - x \Big|_{1-y}^2 \right) dy \\
&= \int_0^1 1 - y - \frac{1^3}{3} - 0 + \frac{y^3}{3} + \frac{(2 + y)^3}{3} - 2 - \frac{1^3}{3} + 1 - y \, dy \\
&= \int_0^1 -\frac{2}{3} - 2y + \frac{y^3}{3} + \frac{(2 + y)^3}{3} \, dy = \cdots = \frac{23}{6}.
\end{aligned}
$$

Note that you do not need to multiply out $(2+y)^3/3$ to integrate it: similar to the integration of $(x + y)^2$ earlier in the computation, its antiderivative is $(2+y)^4/4 \cdot 3$ – check using differentiation in case you have doubts!

Exercise 3.50. (i) Work through the material from the lectures and the integrals on the tutorial sheet that correspond to this section.

(ii) Complete the computation in 3.47. Also re-do example 3.49 (i), now integrating the other way around: first w.r.t. y and then w.r.t. x – this is a very important exercise, as it provides crucial insight on how to best choose the order of integration.

(iii) Using integration, find the volume of the pyramid with base $[-2, 2] \times [-2, 2]$ and height[75] $h = 3$. Compare your result to the volume obtained with geometry formulas.

(iv) Let $f(x)$ be continuous on $[0, 1]$ with

$$\int_0^1 f(x) \, dx = \alpha.$$

Find[76]

$$I = \int_0^1 \int_x^1 f(x) \cdot f(y) \, dy \, dx.$$

3.5 Change of Variables and Integration in Polar Coordinates

Remark 3.51. (i) The theory and computations in this section require familiarity with polar coordinates. You can review the material from the foundations module MTH1000 to refresh your memory on that.

(ii) The next theorem states a change-of-variables formula for two-dimensional integrals, and it is very similar to theorem 3.23. It is first stated in full generality, (i), and then for the special case when the transformation of variables is that from polar coordinates to Cartesian coordinates, (ii). The derivation of (ii) from (i) is given as an exercise, but both this derivation and (i) itself are not examinable. You do have to be able to integrate in polar coordinates though, i.e., you do have to know (ii).

(iii) Suppose we need to find the volume between the paraboloid $z = f(x, y) = x^2 + y^2$ and the disk $D = \{x^2 + y^2 \le a^2\}$ in the xy-plane. Note that both the function and the domain of integration are simpler in polar coordinates, $(x, y) = (r \cos \theta, r \sin \theta)$:

$$f(x, y) = x^2 + y^2 = (r \cos \theta)^2 + (r \sin \theta)^2 = r^2 \left(\cos^2 \theta + \sin^2 \theta \right) = r^2,$$

and D corresponds to a rectangular domain in polar coordinates,

$$(x, y) \in D \quad \leftrightarrow \quad (r, \theta) \in [0, a] \times [0, 2\pi].$$

You know from the previous section that 2D integrals over rectangular domains are the easier ones, and the next theorem allows us to "pull back" the integrand f and the domain of integration D into that setting.

Theorem 3.52 (Substitution in Higher Dimensions). (i) Let a differentiable and injective (or "one-to-one") transformation

$$\Phi \ : \quad \begin{matrix} U & \to & \mathbb{R}^2 \\ (u, v) & \mapsto & (x, y) \end{matrix} \qquad (U \subseteq \mathbb{R}^2)$$

and a continuous function $f = f(x, y)$ be given. Then we have

$$\iint_{\Phi(U)} f(x, y) \, \mathrm{d}x\mathrm{d}y = \iint_U f(\Phi(u, v)) \left| \det J_\phi(u, v) \right| \, \mathrm{d}u\mathrm{d}v,$$

where J_Φ is the *Jacobian* of the transformation $(u, v) \rightsquigarrow (x, y)$,

$$J_\Phi = \begin{bmatrix} \partial x / \partial u & \partial x / \partial v \\ \partial y / \partial u & \partial y / \partial v \end{bmatrix}.$$

(ii) Consider the transformation

$$\Phi \ : \quad \begin{matrix} \mathbb{R}_0^+ \times [0, 2\pi] & \to & \mathbb{R}^2 \\ (r, \theta) & \mapsto & (x, y) & = & (r \cos \theta, r \sin \theta) \end{matrix}$$

from polar coordinates to Cartesian coordinates, and let a continuous function $f = f(x,y)$ be given. Then we have for some set U in the domain of Φ that

$$\iint_{\Phi(U)} f(x,y)\, \mathrm{d}x\mathrm{d}y = \iint_U f(r\cos\theta, r\sin\theta)\, r\, \mathrm{d}r\mathrm{d}\theta.$$

That is, the area element is

$$\mathrm{d}A = r\, \mathrm{d}r\mathrm{d}\theta$$

in polar coordinates.

Remark 3.53. You may have noticed that the transformation $\Phi : \mathbb{R}_0^+ \times [0, 2\pi] \to \mathbb{R}^2$ in (ii) above is not injective. For example, all points $(r, \theta) = (0, \theta)$ are mapped to $(x, y) = (0, 0)$, and $(r, 0), (r, 2\pi)$ are mapped to the same points $(x, y) = (r, 0)$. This is not problematic, as the sets on which this non-injectiveness happens are of lower dimension and therefore do not contribute to the integral. Also, the choice of domain for the angle θ is flexible and can more generally be taken as $[\theta_0, \theta_0 + 2\pi]$, where $\theta_0 \in \mathbb{R}$, cf. example (ii) below.

Example 3.54. (i) Find

$$I = \iint_D x^2 + y^2\, \mathrm{d}A,$$

where D is the disk of radius a, $D = \{x^2 + y^2 \le a^2\}$.
Sol.: This is the integral from the remark at the beginning of this section. We have

$$f(r\cos\theta, r\sin\theta) = r^2$$

and $D = \Phi(U)$, where

$$U = [0, a] \times [0, 2\pi]$$

and Φ is the transformation from polar to Cartesian coordinates. This gives

$$I = \int_0^{2\pi} \int_0^a r^2 \cdot r\, \mathrm{d}r\, \mathrm{d}\theta = \int_0^{2\pi} \left(\frac{r^4}{4}\Big|_0^a\right) \mathrm{d}\theta = \frac{a^4\pi}{2}.$$

You could now also carry out that integration in Cartesian coordinates x, y – as an additional exercise for the previous section and to convince yourself of the benefits of integrating in polar coordinates. Of course, not all 2D integrals are easier to solve in polar coordinates.
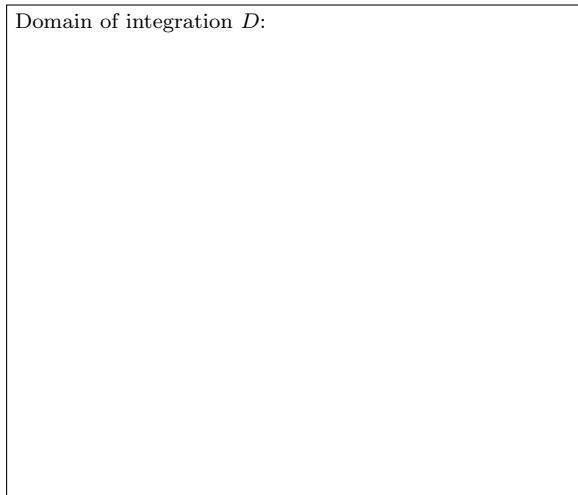
(ii) Let

$$D = \{(x, y) \in \mathbb{R}^2 \mid x \ge 0,\ |y| \le x,\ 9 \le x^2 + y^2 \le 25\},$$

and find $\iint_D x\, \mathrm{d}A$.
Sol.: The domain of integration is a segment of the ring with outer radius $r = 5$ and inner radius $r = 3$:

Domain of integration D:

We see that r ranges from 3 to 5 and θ from $-\pi/4$ to $\pi/4$ – again, a rectangle! That is, for
$$U = [3, 5] \times \left[-\frac{\pi}{4}, \frac{\pi}{4} \right],$$
we have $D = \Phi(U)$. This gives

$$\iint_D x \, \mathrm{d}A = \int_3^5 \int_{-\pi/4}^{\pi/4} r \cos \theta \, r \, \mathrm{d}\theta \, \mathrm{d}r = \int_3^5 r^2 \int_{-\pi/4}^{\pi/4} \cos \theta \, \mathrm{d}\theta \, \mathrm{d}r$$

$$= \int_3^5 r^2 \left(\sin \theta \big|_{-\pi/4}^{\pi/4} \right) \, \mathrm{d}r = \sqrt{2} \int_3^5 r^2 \, \mathrm{d}r = \sqrt{2} \cdot \frac{98}{3} \approx 46.198.$$

This result, 46.2, has a physical meaning – compare to the application in the introduction to this chapter to see what it is[77]. Note that integration in Cartesian coordinates would have been more laborious, as one would have to split up D into at least three pieces[78].

(iii) Find the area A enclosed by one loop of the four-leaved rose $r = \cos(2\theta)$.
Sol.: Here, a curve in the xy-plane is not given by an explicit (e.g. $y = x^2$) or implicit (e.g. $x^2 + y^2 = 1$) formula in x and y, but via an equation in polar coordinates. Those are *polar curves*. For example, $r = 1$ is the unit circle centred at the origin, and $\theta = \pi/3$ describes the ray that leaves the origin at an angle of $60°$. Examples of polar curves – including the four-leaved rose – can be viewed in the interactive MTH1000 worksheet "Polar coordinates". You can also enter commands like `polar plot r=cos(2t)` into WolframAlpha. To find the area within one loop of the four-leaved rose,

113

The four-leaved rose:
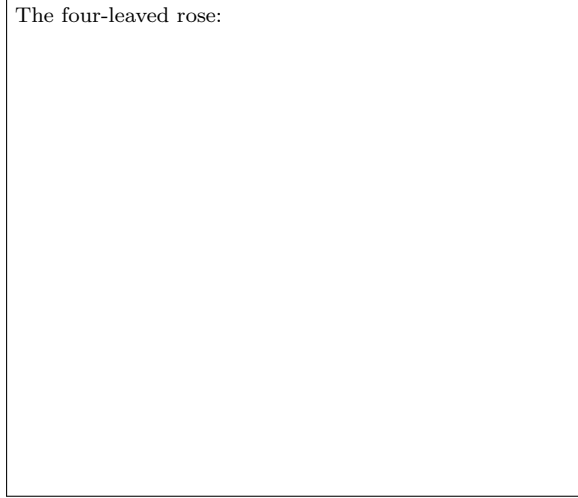
we integrate the function $f = 1$ over $\theta \in [\pi/4, {}^{3\pi}/4]$ and $r \in [0, \cos 2\theta]$:

$$A = \iint_D 1 \, \mathrm{d}A = \int_{\pi/4}^{3\pi/4} \int_0^{\cos 2\theta} r \, \mathrm{d}r \, \mathrm{d}\theta = \frac{1}{2} \int_{\pi/4}^{3\pi/4} \cos^2(2\theta) \, \mathrm{d}\theta$$

$$= \frac{1}{4} \int_{\pi/4}^{3\pi/4} 1 + \cos(4\theta) \, \mathrm{d}\theta = \frac{\pi}{8} + \frac{1}{16} \left(\sin(4\theta)|_{\pi/4}^{3\pi/4} \right) = \frac{\pi}{8}.$$

Example 3.55. Find the volume within the ball of radius R in $\mathbb{R}^3$ (centred at the origin).

Sol.: Denote that ball by B_R,

$$B_R = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 \leq R^2\} \subseteq \mathbb{R}^3.$$

Besides using a basic geometry formula, there are two approaches to finding the volume of B_R: (1) compute the volume of the upper hemisphere by integrating the 2D function whose graph forms the surface of the upper hemisphere over the the disk D_R of radius R, and then multiply by 2; or (2) find the integral of the 3D function $f(x, y, z) = 1$ over B_R. In maths terms:

$$(1) \quad V = 2 \iint_{D_R} \sqrt{R^2 - x^2 - y^2} \, \mathrm{d}A,$$

$$(2) \quad V = \iiint_{B_R} 1 \, \mathrm{d}V.$$

(1) can be carried out nicely in polar coordinates, but we will take the second approach now, as it is a good opportunity to apply the general form of theorem 3.52.

For this, we need *spherical polar coordinates,* i.e. a form of polar coordinates for $\mathbb{R}^3$. You do not need to know this transformation for the MTH1002 exam, but it is good to preview it for your second year in the programme:

$$\Phi \; : \; \mathbb{R}_0^+ \times [0, 2\pi] \times [0, \pi] \; \rightarrow \; \mathbb{R}^3$$
$$\begin{pmatrix} r \\ \phi \\ \theta \end{pmatrix} \quad \mapsto \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r \cos \phi \sin \theta \\ r \sin \phi \sin \theta \\ r \cos \theta \end{pmatrix}$$

For example, for fixed $r = r_0$, the points (x, y, z) will traverse the surface of the ball of radius r_0, with $\theta = 0$ corresponding to the north pole, $\theta = \pi$ to the south pole, and $\theta = \pi/2$ to the equator.

Spherical polar coordinates in $\mathbb{R}^3$:
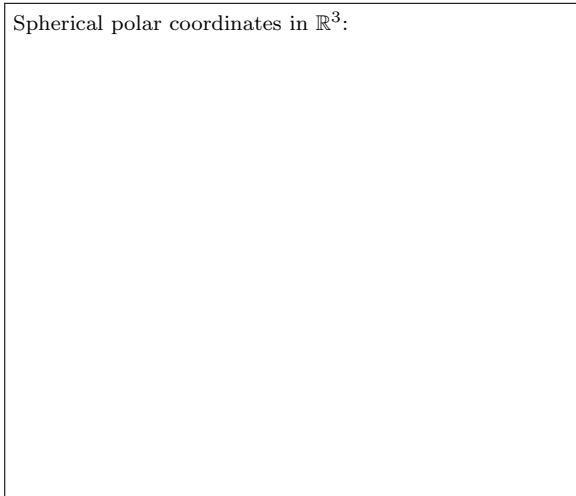
The Jacobian for this transformation is

$$J_\Phi = \begin{bmatrix} \cos\phi\sin\theta & -r\sin\phi\sin\theta & r\cos\phi\cos\theta \\ \sin\phi\sin\theta & r\cos\phi\sin\theta & r\sin\phi\cos\theta \\ \cos\theta & 0 & -r\sin\theta \end{bmatrix},$$

whose determinant has an absolute value of $|\det J_\Phi| = r^2 \sin\theta$. Hence the *volume element* (the 3D version of the area element) is

$$dV = r^2 \sin\theta \, drd\phi d\theta.$$

This allows to find the volume of the ball of radius R with 3D integration:

$$\begin{aligned} V &= \iiint_{B_R} 1 \, dV \\ &= \int_0^\pi \int_0^{2\pi} \int_0^R r^2 \sin\theta \, dr \, d\phi \, d\theta \\ &= \int_0^\pi \int_0^{2\pi} \sin\theta \left(\frac{r^3}{3} \Big|_0^R \right) d\phi \, d\theta \\ &= \frac{R^3}{3} \int_0^\pi \sin\theta \int_0^{2\pi} 1 \, d\phi \, d\theta \\ &= \frac{2\pi R^3}{3} \int_0^\pi \sin\theta \, d\theta = \frac{2\pi R^3}{3} \left(-\cos\theta \big|_0^\pi \right) = \frac{4\pi}{3} R^3. \end{aligned}$$

Application (One more application of integration). ...

Exercise 3.56. (i) As usual, work through the material from the lectures and the corresponding problems on the tutorial sheet. You might also want to review polar coordinates using the material from MTH1000 – e.g., you can practise converting to and from polar coordinates, working with polar curves, and identifying polar domains using the interactive worksheet "Polar coordinates".

(ii) Evaluate

$$\iint_D \left| \sin \left(\sqrt{x^2 + y^2} \right) \right| \, \mathrm{d}A,$$

where D is the upper half of the disk of radius[79] 2π,

$$D = \{ y \geq 0, \ x^2 + y^2 \leq (2\pi)^2 \}.$$

(iii) In theorem 3.52, derive the special case (ii) from the general formula (i).

(iv) Evaluate the integral

$$\iiint_V r \, \mathrm{d}V,$$

where V is the volume between the the two spheres of radius 1 and 2 in $\mathbb{R}^3$ (both centred at the origin)[80].

Chapter 4

Differential Equations

You are perfectly familiar with solving equations involving an unknown variable x and its powers. For example, $x = 2$ solves $x^2 - x - 2 = 0$. We now study equations involving an unknown function and its derivatives. That is, we will be studying *differential equations* (DEs). For example, $f(x) = e^{2x}$ solves $f''(x) - f'(x) - 2f(x) = 0$. Differential equations are difficult to solve, but it is important to keep in mind that checking whether a given function is a solution is more straightforward: you can verify that $f(x) = e^{2x}$ solves $f''(x) - f'(x) - 2f(x) = 0$ solely with your knowledge of differentiation. Which of $f(x) = e^x$ and $f(x) = e^{-x}$ is another solution of that differential equation?

Differential equations for functions of one variable are called *ordinary differential equations* (ODEs), and for functions of several variables, they are called *partial differential equations* (PDEs). Almost all natural processes obey differential equations, and they are therefore very important for physics and in engineering. However, besides classical physics and engineering applications such as describing the motion of objects, many other phenomena can be modelled with differential equations as well: for example, the growth of populations and the spread of diseases.

Application (Draining a tank)**.**
A water tank is being drained, and we are asked to find a formula for the height $h(t)$ of the water level as a function of time.
The behaviour of the height of the water level over time depends on how quickly water is leaving the tank, i.e. on the rate of change $h'(t)$. The problem now is the following: $h'(t)$ in turn depends on $h(t)$ – for example, if the level is very high, the water pressure at the bottom of the tank will be very high and water will be forced out at a very high velocity.
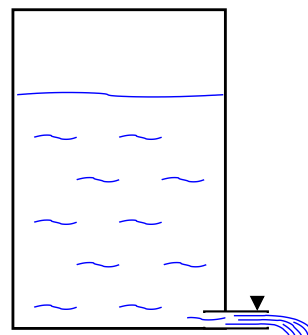
From physical principles, one can derive the differential equation

$$h'(t) = -\alpha \cdot \sqrt{h(t)}$$

for the height of the water level, where α is some positive constant that depends

on the size of the tank and other parameters. This differential equation is known as Torricelli's law, and we will now learn how to solve it!

4.1 First-Order Ordinary Differential Equations

Definition 4.1 (ODEs, order, IVPs, solutions). (i) An *ordinary differential equation (ODE)* is an equation involving an unknown function $y(x)$ and its derivatives

$$y'(x), \, y''(x), \, y'''(x), \, y^{(4)}(x), \, \ldots, \, y^{(n)}(x)$$

and other known functions of x.

(ii) The order of the highest derivative appearing in an ordinary differential equation is called the *order* of the ODE. For example, a first-order ODE is of the form

$$y' = f(x, y),$$

where f is a known function. The argument x of y is often omitted, but it is important to be aware that y is a function.

(iii) A first-order ODE together with an *initial condition* of the form

$$y(x_0) = y_0,$$

where the numbers x_0, y_0 are given, is called an *initial-value problem (IVP)*.

(iv) A function $y = y(x)$ that satisfies a given ODE is called a *solution*. If y contains parameters (constants) such that varying them covers all possible solutions to the ODE, then y is called the *general solution*. A solution that also satisfies given initial conditions is called a *particular solution*. For now, i.e. until section 4.3, we can omit the specification 'ordinary'.

Example 4.2. (i) The DE

$$y''(x) + 4y(x) = 0$$

has $y(x) = \sin(2x)$ as a solution. Indeed, we can check

$$
\begin{aligned}
y(x) &= \sin(2x) \\
\rightarrow \quad y'(x) &= 2\cos(2x) \\
\rightarrow \quad y''(x) &= -4\sin(2x) \\
\rightarrow \quad y''(x) + 4y(x) &= -4\sin(2x) + 4\sin(2x) = 0 \quad \checkmark.
\end{aligned}
$$

We will see later that the general solution is

$$y(x) = c_1 \sin(2x) + c_2 \cos(2x),$$

which gives our guessed solution via the choice $(c_1, c_2) = (1, 0)$ of constants.

(ii) For the DE
$$y^{(4)} = 0,$$
one can guess solutions $y(x) = 1$, $y(x) = x^3$, $y(x) = 2x^2 - 3x + 5$, and hence the general solution
$$y(x) = Ax^3 + Bx^2 + Cx + D$$
– the polynomials of degree less than 4.

Remark 4.3. (i) As for integration: While solving DEs can be tricky, verifying solutions using differentiation is straightforward.

(ii) Differential equations of the form
$$y' = f(x),$$
i.e. when the function $f(x, y)$ in definition 4.1 (ii) does not depend on y, are the easiest to solve – by integration:
$$y(x) = \int f(x) \, \mathrm{d}x.$$
For example, the function $y(x) = \ln(x) + c$ solves the DE $y' = {}^1/_x$. We will refer to this type of DE as *directly integrable*.

(iii) The general solution of a nth-order DE usually contains n constants. You can see that in the examples above. For directly integrable DEs, one can argue as follows: To solve
$$y^{(n)} = f(x),$$
we integrate $f(x)$ n times, obtaining a constant of integration each time. For this module, you can always work with the assumption that the general solution of a nth-order DE contains n constants. Proving this for certain types of DEs – it is not true for all DEs – is beyond the scope of MTH1002.

(iv) Our first non-trivial technique for solving ODEs is for those of the form
$$y' = h(x) \cdot g(y).$$

4.1.1 Separable DEs

Example 4.4. (i) To solve the DE
$$y' = x \cdot y,$$
we first write it as
$$\frac{\mathrm{d}y}{\mathrm{d}x} = x \cdot y$$

and then *separate variables.* That is, bring all expressions with an x on one side and all expressions with y on the other side. For that, we treat dx like a number and 'multiply' the DE by it. Integrating gives

$$\int \frac{1}{y} \, dy = \int x \, dx, \qquad (4.1)$$

which leads to

$$\ln |y| = \frac{x^2}{2} + c_1 \qquad (c_1 \in \mathbb{R}).$$

Applying $\exp(.)$ on both sides gives

$$|y| = e^{x^2/2 + c_1} = e^{x^2/2} \, e^{c_1} = c_2 \, e^{x^2/2} \qquad (c_2 \in \mathbb{R}^+),$$

where we let $c_2 = e^{c_1}$. Now, the absolute value on the left is either a factor of $+1$ or -1, depending on whether y is positive or negative. Absorbing this potential negative sign into the constant, we have

$$y = c_3 \, e^{x^2/2} \qquad (c_3 \in \mathbb{R}\backslash\{0\}).$$

Even though $c_3 \neq 0$ is excluded, we note that a constant of 0 works as well: Then we have $y(x) = 0$ and

$$y'(x) = 0 = x \cdot y.$$

Hence we include the possibility $c = 0$ and find the general solution to be

$$y(x) = c \, e^{x^2/2} \qquad (c \in \mathbb{R}).$$

Note that after (4.1), one should have stated that the function $y(x)$ can not have any zeros, since otherwise the integrand on the left-hand side would not be defined. Hence one solves the DE only on intervals on which y is either always positive or always negative, and considers the case of y having a zero separately. If $y(x_0) = 0$, then $y'(x_0) = 0$ by the original DE, and this suggests that the constant function $y(x) = 0$ is a solution. Now, for the remaining cases, y is either always positive or always negative, which allows to absorb the potential negative from $|y|$ into the constant c_2. However, we will not always discuss the solution of DEs and the ranges of any constants involved that carefully. For example, in the next computation we might change the meaning of constants from one line to the next without explicitly renaming them.

(ii) Solve the IVP

$$\begin{cases} 1 + xyy' = y^2 + yy' \\ \qquad y(0) = -^1/_2. \end{cases}$$

Sol.:

$$\rightarrow \quad (xy - y)\frac{\mathrm{d}y}{\mathrm{d}x} = y^2 - 1 \quad \overset{\text{sep. of var.}}{\rightarrow} \quad \int \frac{y}{y^2 - 1} \,\mathrm{d}y = \int \frac{1}{x - 1} \,\mathrm{d}x$$

$$\overset{u=y^2-1}{\rightarrow} \quad \frac{1}{2}\int \frac{1}{u}\,\mathrm{d}u = \ln|x - 1| + c \quad \overset{\text{integrate}}{\rightarrow} \quad \frac{1}{2}\ln\left|y^2 - 1\right| = \ln|x - 1| + c$$

$$\overset{\cdot 2}{\rightarrow} \quad \ln\left|y^2 - 1\right| = \ln(x - 1)^2 + c \quad \overset{\exp(\cdot)}{\rightarrow} \quad \left|y^2 - 1\right| = e^{\ln(x-1)^2 + c}$$

$$\overset{\text{as in (i)}}{\rightarrow} \quad y^2 - 1 = c\,(x - 1)^2 \quad \rightarrow \quad y^2 + c\,(x - 1)^2 = 1.$$

This is the general solution of the DE, but it is in implicit form, i.e. not solved for y (it is not an explicit formula for $y(x)$). Next, we use the given initial data $(x_0, y_0) = (0, -{}^1\!/_2)$:

$$\left(-\frac{1}{2}\right)^2 + c\,(0 - 1)^2 = \frac{1}{4} + c \overset{!}{=} 1,$$

which gives $c = {}^3\!/_4$ and the explicit solution

$$y = \pm\sqrt{1 - {}^3\!/_4\,(x - 1^2)}.$$

For the particular solution to the given IVP, it remains to decide between $+$ and $-$. We choose

$$y(x) = -\sqrt{1 - {}^3\!/_4\,(x - 1^2)}$$

– why not the positive solution[81]?

Remark 4.5. (i) The implicit general solution $y^2 + c\,(x - 1)^2 = 1$ in (ii) describes an ellipse for $c > 0$ (centred at $(1, 0)$; a circle for $c = 1$), a hyperbola for $c < 0$, and a pair of horizontal lines for $c = 0$.

(ii) Next we study DEs of the form

$$\frac{\mathrm{d}y}{\mathrm{d}x} = f(x, y) = F\left({}^y\!/_x\right),$$

which we call *homogeneous-type* and which can be transformed into separable DEs. A DE is homogeneous-type if

$$f(\lambda x, \lambda y) = f(x, y) \qquad \text{for all } \lambda \neq 0.$$

4.1.2 Homogeneous-Type

Example 4.6. Solve

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{x^2 + y^2}{xy}.$$

Sol.: This DE is not separable. The expression on the right is invariant under replacing x and y with λx and λy, respectively, so the DE is of homogeneous type. We solve it as follows:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{x^2\left(1 + {}^{y^2}\!/_{x^2}\right)}{x^2\,{}^y\!/_x} = \frac{1 + ({}^y\!/_x)^2}{{}^y\!/_x} = F\left({}^y\!/_x\right), \tag{4.2}$$

where F is the single-variable function defined by

$$F(t) = \frac{1 + t^2}{t}.$$

Set $t = y/x$, and note that it is a function of x: $t = t(x)$. Next, we find the relation between y' and t':

$$\frac{dy}{dx} = \frac{d}{dx}(x \cdot t) = xt' + t. \tag{4.3}$$

Combining (4.2) and (4.3), we obtain a DE for $t(x)$:

$$x\,t' + t = \frac{1 + t^2}{t} \quad \rightarrow \quad x\,t' = \frac{1 + t^2}{t} - t = \frac{1}{t}.$$

Writing this as

$$x\,\frac{dt}{dx} = \frac{1}{t},$$

we see that it is a separable DE, which we solve with the steps from the previous section:

$$\int t\,dt = \int \frac{1}{x}\,dx \quad \rightarrow \quad t^2 = \ln x^2 + c = \left(\frac{y}{x}\right)^2,$$

which gives the general solution

$$y(x) = \pm x\,\sqrt{\ln x^2 + c}\,.$$

4.1.3 Linear DEs

Definition 4.7 (Linear First-Order ODEs, Homogeneity).   (i) A first-order ODE of the form

$$y' + p(x)\,y = r(x)$$

is called *linear*.

(ii) It is further called *homogeneous* if $r(x) = 0$, and *inhomogeneous* otherwise.

Remark 4.8. Multiplying this DE by its *integrating factor*

$$h(x) := e^{P(x)},$$

where P is any antiderivative of p, we obtain

$$h\,r = h\,y' + h\,p\,y = h\,y' + h'\,y = (h\,y)',$$

which can be solved by integrating both sides:

$$h(x) \cdot y(x) = \int h(x) \cdot r(x)\,dx.$$

Example 4.9. (i) Solve
$$y' + \frac{5}{x} y = \cos\left(x^6\right).$$

Sol.: The DE is linear, and the integrating factor
$$h(x) = e^{\int 5/x\, dx} = e^{5\ln x} = x^5.$$

Note that the constant of integration was omitted in the integration of $p(x)$ – that is because the method using the integrating factor works for any antiderivative, and hence one can choose the easiest one: the antiderivative with $c = 0$. Next, we find
$$x^5 y = \int x^5 \cos\left(x^6\right)\, dx \overset{u=x^6}{=} \cdots = \frac{1}{6}\left(\sin\left(x^6\right) + c\right).$$

Finally, dividing by the integrating factor gives the solution:
$$y(x) = \frac{\sin\left(x^6\right) + c}{6x^5}.$$

(ii) Solve the IVP
$$\begin{cases} (1 + x^2)y' - 2xy = 1 + x^2 \\ \qquad\qquad y(1) = 6. \end{cases}$$

Sol.: First bring the DE in the correct form:
$$y' - \frac{2x}{1 + x^2} y = 1.$$

Now one can read off the antiderivative $P(x) = -\ln(1 + x^2)$ of $p(x)$. This gives the integrating factor
$$h(x) = e^{-\ln(1+x^2)} = \frac{1}{1 + x^2},$$

and therefore the general solution
$$y(x) = (1 + x^2) \int \frac{1}{1 + x^2} \cdot 1\, dx = (1 + x^2)(\arctan x + c).$$

The initial condition gives
$$6 = (1 + 1^2)\left(\frac{\pi}{4} + c\right) \quad \to \quad y(x) = (1 + x^2)\left(\arctan x + 3 - \frac{\pi}{4}\right).$$

Application (Logistic growth). ...

Exercise 4.10. (i) Work through the first-order DEs on the tutorial sheet for chapter 4.

(ii) Letting $\alpha = 1$ and $h(0) = 100$, solve the differential equation for the water tank from the beginning of this chapter. Plot your solution – does it make sense[82]? How long does it take for the tank to drain[83]?

(iii) Solve[84]
$$xy' = y + x \sec\left(\frac{y}{x}\right), \quad y(1) = \frac{\pi}{3}.$$

(iv) Solve[85]
$$y' = y \tan x + \sin x, \quad y(0) = 0.$$

(v) Prove the statement from remark 4.5 (ii): Let $f(x, y)$ be a continuous function on $\mathbb{R}^2 \backslash \{(0, 0)\}$, i.e. f is defined on the xy-plane without the origin $(x, y) = (0, 0)$. There exists a single-variable function $F(t)$ such that

$$f(x, y) = F(y/x)$$

for any $x \neq 0$, if and only if[86]

$$f(x, y) = f(\lambda x, \lambda y) \qquad \text{for all } \lambda \neq 0.$$

4.2 Linear Ordinary Differential Equations of Higher Order

Definition 4.11 (Second-Order Linear ODEs). The standard form for a *second-order linear ODE is*
$$y'' + p(x)y' + q(x)y = r(x).$$
A LODE is called homogeneous if $r(x) = 0$, inhomogeneous otherwise.

Remark 4.12.   (i) If we set

$$L[y] = y'' + p(x)y' + q(x)y,$$

the DE reads $L[y] = r(x)$. Here, L is a *differential operator* – it takes a function as input and outputs a function. For example, the differential operator

$$\widetilde{L} = 7\frac{\mathrm{d}}{\mathrm{d}x}$$

would map the input x^2 to $14x$.

(ii) For LODEs: If y_1, y_2 are functions and α, β constants, then

$$L[\alpha y_1 + \beta y_2] = \alpha L[y_1] + \beta L[y_2].$$

(iii) If y_1 and y_2 are solutions to a homogeneous s-o LODE and they are *linearly independent* (i.e. $y_2 \neq c \cdot y_1$), then

$$y(x) = c_1 y_1(x) + c_2 y_2(x)$$

is the general solution.

(iv) However, the previous statement, (iii), is not true in the inhomogeneous case: If

$$L[y_1] = r(x) \quad \text{and} \quad L[y_2] = r(x),$$

then

$$L[y_1 + y_2] = L[y_1] + L[y_2] = 2r(x),$$

that is, $y_1 + y_2$ does not satisfy $L[y] = r(x)$ (unless $r(x) = 0$, which is the homogeneous case (iii)).

(v) The following technique allows us to find the general solution if one solution is known or can be guessed.

4.2.1 Reduction of Order

Example 4.13. Given the solution $y_1(x) = x$ of

$$x^2 y'' + x y' - y = 0,$$

find y_2 and hence the general solution.

Sol.: First, one should verify that y_1 really is a solution:

$$x^2 y_1'' + x y_1' - y_1 = x^2 \cdot 0 + x \cdot 1 - x = x - x = 0 \quad \checkmark.$$

Now make the ansatz $y(x) = v(x) \cdot y_1(x)$, or, abbreviated,

$$y = v y_1.$$

The idea behind this ansatz is: the other solutions should be structurally similar to y_1, as they satisfy the same differential equation. From this viewpoint, it seems possible that using the knowledge of y_1 and analysing its ratio with other solutions might reduce the complexity of the problem. We have

$$y = v y_1, \quad y' = v' y_1 + v y_1', \quad y'' = v'' y_1 + 2 v' y_1' + v y_1'',$$

and therefore

$$
\begin{aligned}
L[y] = L[v y_1] &= x^2 \cdot [v'' y_1 + 2 v' y_1' + v y_1''] + x \cdot [v' y_1 + v y_1'] - [v y_1] \\
&= x^2 v'' y_1 + 2 x^2 v' y_1' + x v' y_1 + v \cdot \underbrace{\left[x^2 y_1'' + x y_1' - y_1 \right]}_{=L[y_1]=0} \\
&= x^2 v'' y_1 + 2 x^2 v' y_1' + x v' y_1 + v \cdot 0 \quad \text{(now use } y_1 = x, \ y_1' = 1, \ y_1'' = 0) \\
&= x^3 v'' + 3 x^2 v' \overset{w=v'}{=} x^3 w' + 3 x^2 w \overset{!}{=} 0.
\end{aligned}
$$

That is, if $w = w(x)$ satisfies the first-order LODE

$$x^3 w' + 3 x^2 w = 0,$$

then $y = vy_1$ will satisfy the original second-order LODE. We find

$$w(x) = \frac{\alpha}{x^3},$$

$$v(x) = -\frac{\alpha}{2}x^{-2} + \beta,$$

$$y(x) = \beta x - \frac{\alpha}{2}\frac{1}{x},$$

where $\alpha, \beta \in \mathbb{R}$, and, after renaming the constants,

$$y(x) = c_1 \cdot x + c_2 \cdot \frac{1}{x}.$$

This is the general solution of the DE, and it is in the form mentioned in remark 4.12 (iii), with $y_2 = {}^1/_x$.

Remark 4.14. We recommend writing out $L[y] = L[vy_1] = \ldots$ as above, rather than using the formula for y_1, which is already known at this point. Then, *after* the term $v \cdot L[y_1]$ has dropped out of the computation, one can substitute in the formula for y_1 and its derivatives. Using the formula for y_1 from the beginning, i.e. $L[y] = L[v \cdot x] = \ldots$ in this case, would often make the computation more complicated, e.g. if y_1 is a rational function with bulky derivatives.

4.2.2 Constant-Coefficient Homogeneous

Remark 4.15. Consider the second-order LODE

$$L[y] = y'' + py' + qy = 0, \tag{4.4}$$

where p, q are *constants*, not functions of x. Making the ansatz

$$y(x) = e^{\lambda x},$$

we find

$$0 \overset{!}{=} L[e^{\lambda x}] = \frac{d^2}{dx^2}\left(e^{\lambda x}\right) + p\frac{d}{dx}\left(e^{\lambda x}\right) + q\,e^{\lambda x} = e^{\lambda x}\left(\lambda^2 + p\,\lambda + q\right).$$

That is, if λ satisfies the *characteristic equation* (or *auxiliary equation*)

$$\lambda^2 + p\,\lambda + q = 0, \tag{4.5}$$

then $y = e^{\lambda x}$ is a solution of (4.4). The solutions of (4.5) are

$$\lambda_{1/2} = \frac{-p \pm \sqrt{p^2 - 4q}}{2},$$

and we now have to consider three cases:

(I) If (4.5) has two distinct real solutions λ_1 and λ_2 – this happens when $p^2 - 4q > 0$ – then we have solutions

$$y_1(x) = e^{\lambda_1 x}, \qquad y_2(x) = e^{\lambda_2 x},$$

of (4.4). From earlier theory, we conclude that the general solution is

$$y(x) = c_1 e^{\lambda_1 x} + c_2 e^{\lambda_2 x}.$$

(II) If (4.5) has one real solution λ (i.e. a double root, $\lambda_1 = \lambda_2$) – this happens when $p^2 - 4q = 0$ – then we have solutions

$$y_1(x) = e^{\lambda x}, \qquad y_2(x) = x e^{\lambda x}.$$

That y_1 is a solution follows again from λ solving the characteristic equation, and y_2 is obtained via reduction of order. Note that $\lambda = -p/2$ in this case. The general solution is

$$y(x) = c_1 e^{\lambda x} + c_2 x e^{\lambda x}.$$

(III) If (4.5) has no real solutions – this happens when $p^2 - 4q < 0$ – then it has complex solutions

$$\lambda_{1/2} = a \pm ib \qquad (a, b \in \mathbb{R}),$$

and the solutions to the differential equation are

$$y_1(x) = e^{ax} \cos(bx), \qquad y_2(x) = e^{ax} \sin(bx).$$

The derivation of this is outlined in the exercises below. The general solution is

$$y(x) = e^{ax} \left[c_1 \cos(bx) + c_2 \sin(bx) \right].$$

Definition 4.16 (IVPs and BVPs). For a second-order ODE, $L[y] = r(x)$, the system of equations

$$\begin{cases} L[y] = r(x) \\ y(a) = A \\ y'(a) = B \end{cases}$$

is called an *initial-value problem* (IVP). A system of the form

$$\begin{cases} L[y] = r(x) \\ y(a) = A \\ y(b) = B \end{cases}$$

is called an *boundary-value problem* (BVP).

Example 4.17. (i)
$$\begin{cases} y'' - 3\,y' + {}^9\!/_4\,y = 0 \\ \qquad\qquad y(0) = 0 \\ \qquad\quad y\left({}^2\!/_3\right) = 1 \end{cases}$$

Sol.: We first consider the DE only, which is second-order, linear, and homogeneous with constant coefficients. Its characteristic equation is

$$0 = \lambda^2 - 3\lambda + \frac{9}{4} = (\lambda - {}^3\!/_2)(\lambda - {}^3\!/_2),$$

which has the double root $\lambda = {}^3\!/_2$, i.e. we have case (II) from the discussion above. This value of λ implies that the general solution is

$$y(x) = c_1\, \mathrm{e}^{{}^3\!/_2\,x} + c_2\, x\, \mathrm{e}^{{}^3\!/_2\,x},$$

and it remains to determine the constants so that y satisfies the given boundary conditions:

$$0 \overset{!}{=} y(0) = c_1 \cdot 1 + c_2 \cdot 0 \cdot 1 = c_1$$
$$1 \overset{!}{=} y\left({}^2\!/_3\right) = c_1 \cdot \mathrm{e} + c_2 \cdot {}^2\!/_3 \cdot \mathrm{e} = {}^{2\mathrm{e}}\!/_3\, c_2,$$

which leads to the particular solution

$$y(x) = \frac{3x}{2\mathrm{e}}\, \mathrm{e}^{{}^2\!/_3\,x}.$$

(ii)
$$\begin{cases} y'' - y' + {}^{17}\!/_4\,y = 0 \\ \qquad\qquad y(0) = 0 \\ \qquad\quad y'(0) = 10 \end{cases}$$

Sol.: The characteristic equation is

$$\lambda^2 - \lambda + \frac{17}{4} = 0,$$

which has solutions

$$\lambda_{1/2} = \frac{-(-1) \pm \sqrt{(-1)^2 - 4 \cdot {}^{17}\!/_4}}{2} = \frac{1}{2} \pm \frac{\sqrt{-16}}{2} = \frac{1}{2} \pm i \cdot 2,$$

and therefore leads to the general solution

$$y(x) = \mathrm{e}^{x/2} \left(c_1 \cos(2x) + c_2 \sin(2x)\right).$$

In order to find the solution of the IVP, we need to set the values of y and y' at $x = 0$ equal to the given values. Therefore, $y'(x)$ needs to be found:

$$y'(x) = \frac{1}{2}\mathrm{e}^{x/2} \left(c_1 \cos(2x) + c_2 \sin(2x)\right) + \mathrm{e}^{x/2} \left(-2c_1 \sin(2x) + 2c_2 \cos(2x)\right).$$

Now we can solve for the constants:

$$\begin{cases} 0 = y(0) = e^0 (c_1 \cos 0 + c_2 \sin 0) = c_1 \\ 10 = y'(0) = {}^1\!/_2\, e^0 (c_1 \cos 0 + c_2 \sin 0) + e^0 (-2c_1 \sin 0 + 2c_2 \cos 0) = {}^1\!/_2\, c_1 + 2c_2, \end{cases}$$

which gives $c_1 = 0$, $c_2 = 5$ and hence

$$y(x) = 5\, e^{x/2} \sin(2x).$$

(iii)

$$\begin{cases} y'' + y' - 6y = 0 \\ y(0) = 1 \\ y'(0) = 12 \end{cases}$$

Sol.:

$$y(x) = 3\, e^{2x} - 2\, e^{-3x}$$

Remark 4.18. (i) At the beginning of this chapter, in remark 4.3, it was stated that the general solution of a nth-order DE contains n constants. From this, one can deduce that it takes n conditions to specify a particular solution – this agrees with the examples we have seen so far: one condition in section 4.1 and two conditions (either IVP or BVP) for the second-order DEs in the previous example.

(ii) The difference between an IVP and an BVP can be illustrated as follows. Many natural processes obey second-order differential equations, and the motion of objects is a prominent example. Suppose a ball is kicked vertically in the air, and you are shown a snapshot of it. That is, you are given the height at a certain time, $h(t_1)$. From this, you cannot say what will happen next – for example, you do not know whether the ball was on its way up or down when the picture was taken. However, if you are given a second piece of information, you can – with your knowledge from this module – determine the exact trajectory $h(t)$ of the ball. This second piece of information could be the velocity $h'(t_1)$ of the ball ($\to$ IVP; how fast and in which direction was it moving when the picture was taken) or the position at some other specified time ($\to$ BVP; e.g., another snapshot, taken exactly one second later).

(iii) The method of using the characteristic equation remains applicable for LODEs of higher orders – however, then the characteristic equation will be of higher order as well, and it might therefore not be readily solvable. Sometimes, one is able to guess one solution and then reduce its order, cf. the next example.

Example 4.19.

$$y''' - \frac{1}{2}\, y'' - \frac{41}{2}\, y' + 35\, y = 0$$

Sol.: The characteristic equation is

$$\lambda^3 - \frac{1}{2}\,\lambda^2 - \frac{41}{2}\,\lambda + 35 = 0,$$

and after some experimentation, one finds the solution $\lambda = 2$, since $8 - 2 - 41 + 35 = 0$. This means that $(\lambda - 2)$ is a factor of the cubic expression on the left – dividing by it (long division of polynomials), we obtain

$$0 = \lambda^3 - \frac{1}{2}\lambda^2 - \frac{41}{2}\lambda + 35 = (\lambda - 2)\left(\lambda^2 + \frac{3}{2}\lambda - \frac{35}{2}\right) = (\lambda - 2)(\lambda - \sqrt[7]{2})(\lambda + 5),$$

and therefore the general solution

$$y(x) = c_1 \, e^{2x} + c_2 \, e^{7/2x} + c_3 \, e^{-5x}.$$

4.2.3   Constant-Coefficient Inhomogeneous

Remark 4.20.   (i) The steps for solving an *inhomogeneous* linear second-order DE with constant coefficients,

$$L[y] = y'' + py' + qy = r(x), \qquad\qquad p, q \in \mathbb{R},$$

are:

(1) Solve the corresponding homogeneous DE, $L[y] = 0$. Its general solution is called the *complementary function*:

$$y_{CF} = c_1 y_1 + c_2 y_2.$$

(2) Find one solution of the inhomogeneous DE, $L[y] = r(x)$. This solution is denoted y_{PI} and called the *particular integral*. Finding y_{PI} will be discussed below.

(3) The general solution of the inhomogeneous DE is

$$y = y_{CF} + y_{PI} = c_1 y_1 + c_2 y_2 + y_{PI}.$$

(4) If the DE was given as an IVP or a BVP, i.e. with two additional conditions, then determine the constants in the general solution (3) so that the resulting function satisfies those conditions. This function is the solution to the IVP/BVP.

(ii) Let us check whether the function from step (3) above really is the general solution of $L[y] = r(x)$: The function

$$y = y_{CF} + y_{PI} = c_1 y_1 + c_2 y_2 + y_{PI}$$

does contain two constants – as we expect for the general solution of a second-order DE – and

$$L[y] = L[c_1 y_1 + c_2 y_2 + y_{PI}] = c_1 L[y_1] + c_2 L[y_2] + L[y_{PI}]$$
$$= c_1 \cdot 0 + c_2 \cdot 0 + r(x) = r(x) \quad \checkmark.$$

Example 4.21. Solve the IVP

$$\begin{cases} y'' - y' - 12y = e^{2x} \\ \quad\quad\quad\quad y(0) = -2 \;. \\ \quad\quad\quad\quad y'(0) = 2 \end{cases}$$

Sol.:

(1) First we consider the homogeneous version of the DE,

$$y'' - y' - 12y = 0.$$

Its characteristic equation is

$$0 = \lambda^2 - \lambda - 12 = (\lambda - 4)(\lambda + 3),$$

and it leads to the two solutions

$$y_1(x) = e^{4x}, \qquad y_2(x) = e^{-3x},$$

of the homogeneous DE. Hence the complimentary function is

$$y_{CF}(x) = c_1\, e^{4x} + c_2\, e^{-3x}.$$

(2) Next, we need one solution of the inhomogeneous DE, the particular integral, y_{PI}. The idea is to make an ansatz containing parameters for it and then substitute it into $L[y] = e^{2x}$ to determine those parameters. Choosing the correct form for y_{PI} is key:

Attempt 1: Take $y_{PI}(x) = \sqrt{Ax^2 + Bx + C}$. However, thinking about the action of the operator L, we see that

$$L[y_{PI}] = L\left[\sqrt{Ax^2 + Bx + C}\right]$$
$$= \frac{\mathrm{d}^2}{\mathrm{d}x^2}\left[\sqrt{Ax^2 + Bx + C}\right] - \frac{\mathrm{d}}{\mathrm{d}x}\left[\sqrt{Ax^2 + Bx + C}\right]$$
$$- 12\left[\sqrt{Ax^2 + Bx + C}\right] = \dots$$

will not lead to an expression that can be made equal to e^{2x} by making suitable choices for A, B, C.

Attempt 2: For $y_{PI}(x) = A\cos x + B\sin x$, application of L will give a combination of trigonometric functions $\cos x$ and $\sin x$, not e^{2x}.

Attempt 3: The previous attempts suggest that one should make an ansatz of the same form as the inhomogeneity $r(x)$. Hence try $y_{PI}(x) = Ae^{2x}$:

$$L[y_{PI}] = L\left[Ae^{2x}\right] = A\left[4e^{2x} - 2e^{2x} - 12e^{2x}\right] = A(-10)e^{2x} \stackrel{!}{=} e^{2x},$$

and therefore $A = {}^{-1}/10$,

$$y_{PI}(x) = \frac{-1}{10}e^{2x}.$$

General advice on how to make the ansatz for the particular integral will be given below.

(3) The general solution of the original inhomogeneous DE is

$$y = y_{CF} + y_{PI} = c_1 \, e^{4x} + c_2 \, e^{-3x} - \frac{1}{10} \, e^{2x}.$$

(4) The initial conditions give rise to the system

$$\begin{cases} -2 = y(0) = c_1 + c_2 - {}^1\!/_{10} \\ 2 = y'(0) = 4c_1 - 3c_2 - {}^2\!/_{10}, \end{cases}$$

which has solutions $c_1 = -{}^1\!/_2$, $c_2 = -{}^7\!/_5$. This gives the solution to the IVP as

$$y = -\frac{1}{2} \, e^{4x} - \frac{7}{5} \, e^{-3x} - \frac{1}{10} \, e^{2x}.$$

Remark 4.22. (i) The following table and remarks should help you make the ansatz for the particular integral:

Right-hand side $r(x)$	Ansatz for y_{PI}
polynomial of degree m: $\alpha_m x^m + \ldots \alpha_1 x + \alpha_0$	polynomial of degree m: $A_m x^m + \ldots A_1 x + A_0$
exponential: $\alpha e^{\lambda x}$	exponential: $A e^{\lambda x}$
comb. of trig. functions: $\alpha \cos(\lambda x) + \beta \sin(\lambda x)$	comb. of trig. functions: $A \cos(\lambda x) + B \sin(\lambda x)$
$\ldots$ (e.g.: comb. of hyp. trig. fcns	$\ldots$ comb. of hyp. trig. fcns)
a sum of any of the above: $r(x) = r_1(x) + \cdots + r_k(x)$	find $y_{PI,j}$ for each r_j and then: $y_{PI} = y_{PI,1} + \cdots + y_{PI,k}$

Here it is important to always have full generality on the ansatz side: E.g., if $r(x) = 2\sin(3x)$, i.e. there is no $\cos(3x)$ term in the inhomogeneity, one still needs $y_{PI} = A\cos(3x) + B\sin(3x)$ rather than an ansatz with a sine term only. If, for example, $r(x) = \sin(5x) + \cos(2x)$, the rule in the middle row of the table needs to be applied twice – once for $\lambda = 5$ and once for $\lambda = 2$ – and the outcomes are then to be combined as in the last row. If you have written out the ansatz y_{PI} based on $r(x)$, but this y_{PI} happens to already be a solution to the homogeneous version of the DE, then multiply it by x.

(ii) The linearity of the ODE is crucial for the above approach of finding all solutions of the homogeneous DE and then adding one solution of the inhomogeneous DE. Let us compare this to another linear situation: An inhomogeneous linear equation for the variables x, y, z: Solve

$$x + 2y - 5z = 2. \tag{4.6}$$

Sol.:

(1) For the homogeneous version of the equation,

$$x + 2y - 5z = 0,$$

we can find the two linearly independent (cf. 1.30 for the general definition of this term) solutions

$$v_1 = \begin{bmatrix} 5 \\ 0 \\ 1 \end{bmatrix}, \qquad v_2 = \begin{bmatrix} 0 \\ 5 \\ 2 \end{bmatrix}$$

and hence the general solution

$$v_{CF} = c_1 \begin{bmatrix} 5 \\ 0 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 5 \\ 2 \end{bmatrix}.$$

(2) We now just need a single vector that satisfies the inhomogeneous equation (4.6). For example,

$$v_{PI} = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix},$$

as

$$\begin{bmatrix} 1 & 2 & -5 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} = 2 \quad \checkmark$$

(the entries of the row vector are the coefficients of (4.6)).

(3) The general solution of the inhomogeneous equation is

$$v = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} + c_1 \begin{bmatrix} 5 \\ 0 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 5 \\ 2 \end{bmatrix}.$$

This is a parametrisation of the plane (4.6).

Example 4.23. Find the general solution of

$$L[y] = y'' - 3\,y' + \tfrac{9}{4}\,y = e^{3x/2}.$$

Sol.:

(1) We have already found the solution to the homogeneous version of that DE in example 4.17 (i):

$$y_{CF}(x) = c_1 \, e^{3/2\,x} + c_2 \, x \, e^{3/2\,x}.$$

(2) Based on the form of the right-hand side, $r(x) = e^{3x/2}$, one would make the ansatz $y_{PI} = A e^{3x/2}$. However, this will be mapped to 0 by L, as it is a solution of $L[y] = 0$ – after all, the function $A e^{3x/2}$ is already contained in the complementary function; for $(c_1, c_2) = (A, 0)$. According to the last sentence of

133

remark 4.22 (i), we multiply by x to obtain $y_{PI} = Ax\mathrm{e}^{3x/2}$. Again, this already solves the homogeneous DE, and we therefore multiply by x one more time:

$$y_{PI}(x) = Ax^2\mathrm{e}^{3x/2}.$$

Substituting this into $L[y] = \mathrm{e}^{3x/2}$ gives

$$
\begin{aligned}
L\left[Ax^2\mathrm{e}^{3x/2}\right] &= A\left[2\mathrm{e}^{3x/2} + 6x\mathrm{e}^{3x/2} + \frac{9}{4}x^2\mathrm{e}^{3x/2}\right] \\
&\quad - 3A\left[2x\mathrm{e}^{3x/2} + \frac{3}{2}x^2\mathrm{e}^{3x/2}\right] + \frac{9}{4}A\left[x^2\mathrm{e}^{3x/2}\right] \\
&= A\left[2 + 6x + \frac{9}{4}x^2 - 6x - \frac{9}{2}x^2 + \frac{9}{4}x^2\right]\mathrm{e}^{3x/2} = 2A\mathrm{e}^{3x/2} \overset{!}{=} \mathrm{e}^{3x/2},
\end{aligned}
$$

and hence

$$y_{PI}(x) = \frac{x^2}{2}\mathrm{e}^{3x/2}.$$

(3) Steps (1) and (2) lead to the general solution

$$y(x) = \mathrm{e}^{3x/2}\left(c_1 + c_2x + \frac{x^2}{2}\right).$$

Exercise 4.24. (i) Solve the higher-order ODEs on the tutorial sheet.

(ii) Consider the s-o LODE
$$y'' + 4y = 0$$
from example 4.2 and the solution $y_1(x) = \sin(2x)$ we had already guessed and verified. Using reduction of order, find the general solution. Then also solve the DE using the theory from section 4.2.2.

(iii) Solve the BVP[87]
$$
\begin{cases}
y'' - 2y' + 2y = 10 \\
\quad\quad\quad y(0) = 7 \\
\quad\quad\quad y(\pi/2) = 4.
\end{cases}
$$

(iv) Find the general solution of[88]
$$y'' + 2y' + y = \cosh(3x) + \mathrm{e}^{-x}.$$

(v) Consider the s-o LODE
$$L[y] = x^3y'' + 3x^2y' + xy = 0. \tag{4.7}$$
Find one solution via the ansatz $y_1(x) = x^r$, and then find the general solution via reduction of order.

(vi) Solve the DE (4.7) again, this time by transforming to a DE for the function $u = u(t)$, where $u = x \cdot y$ and $x = \mathrm{e}^t$.

(vii) Show that the function y_2 in (II) of remark 4.15 and both functions $y_{1/2}$ in (III) really are solutions of (4.4) in the respective cases. The former can be done with reduction of order, and the latter by carrying out (I) for complex λ, e.g. $y_1 = e^{(a+ib)x}$ – the usual argument involving the characteristic equation remains applicable, and the laws of exponentials are true in the complex case as well – and then using *Euler's identity*,

$$e^{ix} = \cos(x) + i\sin(x)$$

and renaming the constants. A different approach would be, of course, to simply verify that the functions that are given satisfy the constant-coefficient homogeneous second-order LODE (4.4). Euler's identity is very significant and perhaps surprising: for $x = \pi$ we obtain

$$e^{i\pi} = -1,$$

combining three important mathematical constants on the left – that are quite non-trivial (irrational/imaginary) and seemed unrelated until now – to a rather simple number on the right.

4.3   Partial Differential Equations

Remark 4.25. *Partial differential equations* (PDEs) are differential equations for multivariate functions, i.e. equations involving the partial derivatives of an unknown function. Many natural processes obey partial differential equations, for example: The temperature in a room as a function of time can be modelled with

$$u = u(t, x, y, z)$$

and satisfies the *heat equation*

$$u_t - \alpha\left(u_{xx} + u_{yy} + u_{zz}\right) = 0.$$

The heat equation is a very important PDE. We will not be able to discuss it in MTH1002, but you will learn how to solve it in MTH2003.

Example 4.26.   (i) Solve the PDE

$$\frac{\partial u}{\partial x} = 0,$$

where $u = u(x, y)$.

Sol.: Let us just integrate that equation with respect to x:

$$\int \frac{\partial u}{\partial x}\,\mathrm{d}x = u = \int 0\,\mathrm{d}x = c(y).$$

Here, the constant of integration can be a function of y, since any expression that does not depend on x has 0 as its x-derivative. Renaming the "constant", we find the solution

$$u(x, y) = w(y),$$

where w is an arbitrary differentiable single-variable function. Since this is a new type of computation, and since it might be surprising to have so much freedom of choice in the general solution – for first-order ODEs, that would be a constant; here it is a function that can be chosen freely – let us check:

$$\frac{\partial u}{\partial x} = \frac{\partial}{\partial x} \left(w(y) \right) = 0 \quad \checkmark.$$

(ii)

$$\frac{\partial^2 u}{\partial x \partial y} = 0$$

Sol.:

$$\frac{\partial}{\partial x} \left(\frac{\partial u}{\partial y} \right) = 0 \quad \overset{\int \cdot \, \mathrm{d}x}{\to} \quad \frac{\partial u}{\partial y} = \int 0 \, \mathrm{d}x = w(y)$$

$$\overset{\int \cdot \, \mathrm{d}y}{\to} \quad u = \int w(y) \, \mathrm{d}y \overset{W'=w}{=} W(y) + v(x)$$

$$\overset{\text{rename}}{\to} \quad u(x,y) = f(x) + g(y),$$

where f and g are arbitrary differentiable single-variable functions. Check this result using partial differentiation!

Remark 4.27. (i) As hinted on above, the set of solution is *much* bigger for PDEs than for ODEs of the same order. You will see that when solving the exercises below and the PDEs on the tutorial sheet.

(ii) The previous examples are analogous to directly integrable ODEs – they can be solved directly using integration. The next example can not be solved as readily, but we will bring it into directly integrable form using theory from chapter 2.

Example 4.28. Find $u = u(x,y)$ that satisfies

$$L[u] = 6\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial x \partial y} - \frac{\partial^2 u}{\partial y^2} = 0.$$

Sol.: Note that we can not simply integrate the PDE, as $L[u]$ is a sum of different partial derivatives of u. The idea is to define new variables,

$$\xi = x + ay, \qquad \eta = x + by, \qquad (\xi \to \text{"xi"}, \, \eta \to \text{"eta"})$$

and then choose the parameters a and b so that the PDE can be integrated directly in the new variables (we had already carried out such a transformation in chapter 2: example 2.21 (iv)). The partials transform as follows:

$$\frac{\partial u}{\partial x} = u_\xi + u_\eta, \qquad \frac{\partial u}{\partial y} = a\,u_\xi + b\,u_\eta,$$

$$\frac{\partial^2 u}{\partial x^2} = u_{\xi\xi} + 2\,u_{\xi\eta} + u_{\eta\eta},$$

$$\frac{\partial^2 u}{\partial y^2} = a^2\,u_{\xi\xi} + 2ab\,u_{\xi\eta} + b^2\,u_{\eta\eta},$$

$$\frac{\partial^2 u}{\partial x \partial y} = a\,u_{\xi\xi} + (a+b)\,u_{\xi\eta} + b\,u_{\eta\eta},$$

and hence $L[u]$ transforms to

$$
\begin{aligned}
L[u] &= 6\,u_{\xi\xi} + 12\,u_{\xi\eta} + 6\,u_{\eta\eta} + a\,u_{\xi\xi} + (a+b)\,u_{\xi\eta} + b\,u_{\eta\eta} - a^2\,u_{\xi\xi} - 2ab\,u_{\xi\eta} - b^2\,u_{\eta\eta} \\
&= \left[6 + a - a^2\right] u_{\xi\xi} + \left[12 + a + b - 2ab\right] u_{\xi\eta} + \left[6 + b - b^2\right] u_{\eta\eta} \\
&=: c_1\,u_{\xi\xi} + c_2\,u_{\xi\eta} + c_3\,u_{\eta\eta} \overset{!}{=} c \cdot u_{\xi\eta}.
\end{aligned}
$$

The last line states that we assigned the names c_1, c_2, c_3 to the expressions in square brackets in the previous line. We would then like this combination of second-order partials with respect to ξ and η to be of the form some constants times $u_{\xi\eta}$, as this is a PDE that we can solve, cf. the previous example. For $L[u] = 0$ to be equivalent to $u_{\xi\eta} = 0$, we need $c_1 = 0$, $c_3 = 0$, and $c_2 \neq 0$, because then

$$
L[u] = 0 \quad \Longleftrightarrow \quad c_2\,u_{\xi\eta} = 0 \quad \Longleftrightarrow \quad u_{\xi\eta} = 0.
$$

Setting $c_1 = 0$ and $c_3 = 0$ leads to

$$
a \in \{-2, 3\} \quad \text{and} \quad b \in \{-2, 3\},
$$

which a-priori leaves four choices for the combination of parameters (a, b). However, if $a = b$, then

$$
c_2 = 12 + 2a - 2a^2 = 2[6 + a - a^2] = 2 \cdot 0 = 0,
$$

which shows that $a = b$ is not a suitable choice. (If we chose $a = b$, we would have $\xi = \eta$, and we would therefore be looking for functions that can effectively be written as functions of one variable – this would not give us full generality for solutions of $L[u] = 0$.) We are left with two equivalent choices for (a, b). Take

$$
\xi = x + 3y, \qquad \eta = x - 2y.
$$

The new DE is

$$
u_{\xi\eta} = 0,
$$

which we have already solved in the previous example (well, the variables were called x and y then). Its solution is $f(\xi) + g(\eta)$, and we therefore obtain the solution

$$
u(x, y) = f(x + 3y) + g(x - 2y)
$$

in the original variables. Here, f and g are arbitrary single-variable functions (twice differentiable).

Application (Wave equation). `Stone dropped in a swimming pool, do quickly, use Fourier theory from the previous chapter to find correct superposition of modes; mention heat equation and Black-Scholes as other applications.`

Exercise 4.29. (i) Work through the PDEs on the tutorial sheet. Check the solution from the previous example by finding its second-order partial derivatives and check whether they add to 0 when combined in way prescribed by the PDE. This will require the 1D chain rule, e.g.:

$$
\frac{\partial}{\partial y}\left(f(x + 3y)\right) = f'(x + 3y)\,\frac{\partial}{\partial y}\,(x + 3y) = 3f'(x + 3y).
$$

(ii) Find $u = u(x, y)$ that satisfies[89]

$$2\frac{\partial^2 u}{\partial x^2} + 5\frac{\partial^2 u}{\partial x \partial y} + 2\frac{\partial^2 u}{\partial y^2} = 0.$$

(iii) Find the function $u = u(x, y)$ with[90]

$$\begin{cases} \dfrac{\partial^2 u}{\partial y^2} = 0 \\ u(x, 0) = \cosh(x) \\ u(x, -1) = 0. \end{cases}$$

(iv) Solve the first-order PDE[91]

$$\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} = 0.$$

4.4   Systems of Differential Equations

Application (Coupled oscillator). The two applications in this section connect differential equation to matrices. We first briefly derive some general theory in a context that is very important and fundamental in physics. This is then followed by an application to biology.

Suppose we have two oscillators (e.g., pendulums, weights on springs) described by functions $x(t)$ and $y(t)$. Each of those two functions obeys a second-order ODE, for example an ODE of the form $x''(t) = c \cdot x(t)$. Now suppose we couple the two (e.g., connect them with a spring) – then the position of y will affect the motion of x and vice versa. This can be described by a system of ODEs. An example would be

$$\begin{cases} x'' &= -2x + y \\ y'' &= x - 2y, \end{cases} \tag{4.8}$$

where $'$ stands for the derivative with respect to t.

We approach solving this system by writing it in matrix form:

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

and

$$w'' = Aw, \tag{4.9}$$

where $w(t) = \begin{bmatrix} x(t) & y(t) \end{bmatrix}^\top$ and A is the matrix from the previous equation. Next, we compute the eigenvalues and eigenvectors of A:

$$\lambda_1 = -3, \quad v_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

$$\lambda_2 = -1, \quad v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

The ansatz

$$w(t) = f_1(t)v_1 = f_1(t)\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} f_1(t) \\ -f_1(t) \end{bmatrix},$$

where $f_1(t)$ is a function to be determined, is now very useful for solving (4.9) because (a) the derivatives in (4.9) affect only $f_1(t)$, as v_1 is constant; (b) the matrix product affects only v_1, as $f_1(t)$ is a scalar factor that can be swapped with the matrix multiplication; and (c) this matrix multiplication is simplified by v_1 being an eigenvector. We obtain

$$A(w(t)) = A(f_1(t)v_1) = f_1(t)Av_1 = f_1(t)\lambda_1 v_1 = \lambda_1 f_1(t) \cdot v_1$$

and

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} w(t) = \frac{\mathrm{d}^2}{\mathrm{d}t^2}(f_1(t)v_1) = \frac{\mathrm{d}^2}{\mathrm{d}t^2}(f_1(t)) \cdot v_1. \tag{4.10}$$

Equating the the two leads to

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}(f_1(t))v_1 = \lambda_1 f_1(t)v_1 \implies \begin{cases} f_1''(t) \cdot 1 & = -3f_1(t) \cdot 1 \\ f_1''(t) \cdot (-1) & = -3f_1(t) \cdot (-1) \end{cases} \implies f_1'' = -3f_1.$$

This gives

$$f_1(t) = A\cos\left(\sqrt{3}t\right) + B\sin\left(\sqrt{3}t\right),$$

and therefore a first solution

$$w_1(t) = \left[A\cos\left(\sqrt{3}t\right) + B\sin\left(\sqrt{3}t\right)\right] \cdot v_1.$$

Similarly, one finds

$$f_2(t) = C\cos(t) + D\sin(t)$$

for $f_2(t)$ in the ansatz $w_2(t) = f_2(t)v_2$.

We have therefore found the general solution:

$$w(t) = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = A\cos\left(\sqrt{3}t\right)v_1 + B\sin\left(\sqrt{3}t\right)v_1 + C\cos(t)v_2 + D\sin(t)v_2$$

$$= A\begin{bmatrix} \cos(\sqrt{3}t) \\ -\cos(\sqrt{3}t) \end{bmatrix} + B\begin{bmatrix} \sin(\sqrt{3}t) \\ -\sin(\sqrt{3}t) \end{bmatrix} + C\begin{bmatrix} \cos(t) \\ \cos(t) \end{bmatrix} + D\begin{bmatrix} \sin(t) \\ \sin(t) \end{bmatrix}. \tag{4.11}$$

The constants A, B, C, D may now be chosen so that $w(t)$ satisfies any initial conditions the system (4.8) might have come with. If in doubt about any of the vector steps above – e.g. taking the derivative of a vector – one could always re-write them as pairs of ordinary equations. For example, one could write out (4.10) as

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}w = \frac{\mathrm{d}^2}{\mathrm{d}t^2}\begin{bmatrix} f_1 \\ -f_1 \end{bmatrix} = \begin{bmatrix} f_1'' \\ -f_1'' \end{bmatrix} = f_1'' \cdot \begin{bmatrix} 1 \\ -1 \end{bmatrix} = f_1'' \cdot v_1.$$

Application (Competing species.). Do quickly, also eigenvalue analysis to classify the equilibria; mention predator-prey and SIR as similar models.

Exercise 4.30. (i) Suppose the motion of the first oscillator in (4.8) is given as

$$x(t) = 2\cos\left(\sqrt{3}t\right) - \sin\left(\sqrt{3}t\right) + \cos(t) - 3\sin(t).$$

Differentiate $x(t)$ twice, and then find $y(t)$ using the first equation of (4.8). (This approach to find y does not use any of the theory above.) Could you have read off $y(t)$ immediately from the general solution (4.11) and the given $x(t)$?

(ii) `Simple system without context.`

(iii) `Simple system with context.`

(iv) `...and perhaps a more theoretical/conceptual problem.`

Hints and Answers

1. For example:

$$\text{all hikers in } C \leftrightarrow v = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \xrightarrow{P} Pv = \begin{bmatrix} 1/3 \\ 1/3 \\ 0 \\ 1/3 \\ 0 \end{bmatrix} \leftrightarrow \text{hikers evenly distributed over } A, B, D \qquad \checkmark$$

   Now check this for the other four "concentrated" configurations. If that works, then P is correct.

2. For the difference $AB - BA$, you should obtain

$$\begin{bmatrix} 10 & 7 & -20 \\ -23 & -9 & 39 \\ 1 & 3 & -1 \end{bmatrix}.$$

3. $f(A) = 0$.

4. $\theta = \frac{\pi}{3}$.

5. The determinants are $8, 18, 30$.

6. Consider a general matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$; multiply with I in both orders; compare.

7. Matrices of the form

$$X = \begin{bmatrix} a & 0 \\ b & a+b \end{bmatrix},$$

   where $a, b \in \mathbb{R}$, commute with A.

8. Prove by induction; use angle sum identities for trigonometric functions.

9. Computing this determinant with Sarrus' rule leads to a third-order algebraic expression in λ. To find its zeros, you first need to guess one solution – property (ii) of 1.11 should be helpful. The second value for λ is $\lambda_2 = -2$.

10.

$$\widetilde{p} = \frac{1}{26} \begin{bmatrix} 1 & 5 \\ 5 & 25 \end{bmatrix} \begin{bmatrix} 1/2 \\ 10 \end{bmatrix} = \frac{50.5}{26} \begin{bmatrix} 1 \\ 5 \end{bmatrix} \approx \begin{bmatrix} 1.942 \\ 9.712 \end{bmatrix}$$

11.

	n	rank A	rank $[A \mid b]$	# sol.
(i)	2	2	2	1
(ii)	3	3	4	0
(iii)	4	3	3	∞

12. Using the parameter $t \in \mathbb{R}$, we obtain $(x, y, z) = (-2t, -2t, t)$.

13. The vector

$$\begin{bmatrix} x & y & z \end{bmatrix}^\top = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^\top,$$

satisfies all three equations. One can also argue more abstractly that the no-solution case in Theorem 1.22 never happens if $b = 0$.

14. Bring the system in REF using Gaussian elimination as usual. Then choose α such that we are not in the no-solution case of Theorem 1.22.

15. Unique solution if $\alpha \neq 1$ and $\beta \neq 0$. Infinitely many solutions if $\alpha = 1$ and $\beta = \frac{1}{3}$. Otherwise, no solutions.

16.
$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix} \overset{!}{=} P_{L,x} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \frac{1}{5}y \\ y \end{bmatrix} \quad \Rightarrow \quad P_{L,x} = \begin{bmatrix} 0 & \frac{1}{5} \\ 0 & 1 \end{bmatrix}$$

Alternatively, one can think about where the standard vectors $\begin{bmatrix} 1 & 0 \end{bmatrix}^\top$ and $\begin{bmatrix} 0 & 1 \end{bmatrix}^\top$ get mapped to:
$$P_{L,y} : \begin{bmatrix} 1 \\ 0 \end{bmatrix} \mapsto \begin{bmatrix} 1 \\ 5 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Rightarrow \quad P_{L,y} = \begin{bmatrix} 1 & 0 \\ 5 & 0 \end{bmatrix}.$$

17. The eigenvalues of the first matrix are $\lambda_1 = 5$, $\lambda_2 = -1$, and the second matrix has eigenvalues 8 and 6. To verify the eigenvectors you have found, make sure that multiplying them against the matrix gives the correct scalar multiple.

18. $(\lambda_1, \lambda_2, \lambda_3) = (-1, -1, 8); (-1, \frac{1+\sqrt{5}}{2}, \frac{1-\sqrt{5}}{2})$.

19. Yes; no; no.

20. The discussion in Example 1.32 shows how to set up the system for finding the coefficients.

21. $(\lambda_2, \lambda_3) = (3, 6)$.

22. Lemma 1.27.

23. The set S is linearly independent if
$$c_0 \, v + c_1 \, A \, v + c_2 \, A^2 \, v + \ldots + c_k \, A^k \, v = 0 \qquad (\star)$$
implies that all coefficients are zero. Hence show that all coefficients in $(\star)$ are zero – apply powers of A to $(\star)$ to do that.

24. The population converges to
$$p_\infty = \begin{bmatrix} 600 & 800 \end{bmatrix}^\top.$$

25. Matrices like the given one are called *block diagonal*. Here we have two 1×1 blocks and one 2×2 block on the diagonal. What do you notice for the inverse?

26. Lemma 1.27 should be helpful here. The following is an alternative approach. When $\det A = 0$, we have a certain eigenvalue, $\lambda = \ldots$ Which one? Now try to apply the $v \leftrightarrow w$ argument from Remark 1.38 to the corresponding eigenvector. Does that work?

27. Let V and W be the 3×3 matrices consisting of the vectors v_j and w_j. The observation
$$V \begin{bmatrix} 1 & 0 & 1 \\ 1 & 3 & -2 \\ 0 & 2 & 1 \end{bmatrix} = W$$
is now very useful. Check whether the matrix C of coefficients for the transformation $v_j \rightsquigarrow w_j$ is invertible. If so, you can write $V = WC^{-1}$. Compare to Example 1.32 for inspiration on how to complete the proof.

28. Your predicted score on the final exam is
$$y = \begin{bmatrix} 73 & 74 \end{bmatrix} \left(\begin{bmatrix} 55 & 72 & 64 \\ 43 & 60 & 63 \end{bmatrix} \begin{bmatrix} 55 & 43 \\ 72 & 60 \\ 64 & 63 \end{bmatrix} \right)^{-1} \begin{bmatrix} 55 & 72 & 64 \\ 43 & 60 & 63 \end{bmatrix} \begin{bmatrix} 61 \\ 82 \\ 68 \end{bmatrix} = \cdots \approx 77.3 \,.$$

You should not rely on this prediction though, as it is problematic in a number of ways. Most notably:

- there is not enough data;

- the lecturer may have changed the level of difficulty of the assessments;

- and, most importantly, your performance on the final exam depends above all on how well you study!

29. What does it mean for a derivative to be large at a point? It means that a small change of the variable will have a large effect!

30. We have
$$f(x, x) = x^2 , \quad f(x, -x) = -x^2 .$$

That is, there is a parabola "sitting" on the line $y = x$ of the xy-plane and an upside-down parabola "hanging" under the line $y = -x$. The full graph looks like a saddle. You can use software to help visualising such functions; e.g., enter `plot x*y` in WolframAlpha. Note that, even though this is a curved surface, you can form it with straight lines (e.g., with mikado sticks) since the restriction of the graph to any $y = c$ is a line.

31. The xy derivatives of the three functions are
$$2x + 12x^2y^3 , \quad e^x , \quad -\sin x \, \tfrac{1}{y} ,$$

and we always have $f_{xy} = f_{yx}$. You can use software to check your results; e.g., `partial derivatives of ...` in WolframAlpha. Use such commands only to support your studies though – do not rely on software!

32. The xy derivatives of the three functions are
$$\cos(xy + y^3) - y(x + 3x^2)\sin(xy + y^3) , \quad -4xy - 2xy \ln xy , \quad \tfrac{2y(1+x^2+y^2)^2 - 4xy(1+x^2+y^2)(2x)}{(1+x^2+y^2)^4} .$$

(If you don't quite get those expressions, you may have forgotten to apply the product rule.)

33. You can check your domain with a WolframAlpha command of the form `plot ... >= 0`.

34. For example: $\ln(1 - x^2 - y^2)$ and $\sqrt{x^2 + x^2 - 1}$.

35. Use Theorem 2.12 to justify that your answer can be given in the form
$$\frac{\partial^n}{\partial y^n} \frac{\partial^m f}{\partial x^m} = \dots ,$$

i.e. it can be assumed that x-derivatives are taken first.

36. Setting both partial derivatives equal to zero gives you a (non-linear) system of two equations; solving it should lead to three points.

37.
$$\begin{aligned}
\frac{dF}{dt} &= \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} = (1 + 2xy) \cdot 1 + (x^2 - 15y^2) \cdot 2t \\
&= 1 + 2tt^2 + (t^2 - 15t^4)2t = 1 + 4t^3 + 30t^5 \\
\frac{dG}{dt} &= \frac{t + 4t^3}{\sqrt{1 + t^2 + t^4}}
\end{aligned}$$

38. In (2.7), the variable names x and y can be swapped, can't they? Having that in mind, what is the difference between the given z_x and the z_y you have found (they should be very similar – if not, you have made a mistake)?

143

39. Carrying out the transformation as in Example 2.21 (iv) leads to

$$0 = 0 \cdot \frac{\partial^2 u}{\partial \xi^2} - 2(1 + b) \cdot \frac{\partial^2 u}{\partial \xi \partial \eta} + (1 + 4b + 3b^2) \cdot \frac{\partial^2 u}{\partial \eta^2} \, .$$

The equation $u_{\xi\eta} = 0$ we need to transform to has only a mixed derivatives. Hence b needs to be chosen such that $1 + 4b + 3b^2 = 0$. We choose $b = -1/3$, since the other solution would also make the mixed derivative disappear. This gives $-4/3 u_{u_{\xi\eta}} = 0$; now multiply by $-3/4$.

40. This requires quite a bit of work. Start with

$$u_s = \frac{\partial}{\partial s}(u) = \frac{\partial u}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial s} = e^s \left(u_x \cos t + u_y \sin t \right) \, .$$

We need the product rule for the second s-derivative:

$$u_{ss} = \frac{\partial}{\partial s}(u_s) = \frac{\partial}{\partial s}(e^s) \cdot (u_x \cos t + u_y \sin t) + e^s \left(\frac{\partial u_x}{\partial s} \cos t + \frac{\partial u_y}{\partial s} \sin t \right) .$$

Now find the s-derivatives of the functions u_x, u_y by applying the chain rule – as in the first step. Then simplify, repeat for u_{tt}, find the sum $u_{ss} + u_{tt}$.

41. For the hyperbola:

$$y = \frac{1}{x} \quad \longrightarrow \quad xy = 1 \quad \Longrightarrow \quad f(x, y) = xy \, , \ c = 1 \, .$$

Here is an example for the WolframAlpha syntax: `plot sin(x)*sin(y) = 0.1` .

42. The level set is a curve of the shape of an infinity sign: ∞. The y-coordinate of its points can not be written as a function of x since for most x, there are two corresponding y-values. Most individual segments of the curve can be written as $y(x)$ though (careful: this is not possible at $x = -1, 0, 1$), and then $y'(x_0) = 0$ has the usual interpretation: a horizontal tangent line.

43. You can *check* (not "find"!) your result with `differentiate z(x,y) = ...`

44.
$$\nabla f(3, 1) = \frac{1}{4} \begin{bmatrix} 1 & 2 \end{bmatrix} \, , \quad \text{Hess} f(3, 1) = \frac{1}{16} \begin{bmatrix} -1 & -2 \\ -2 & +4 \end{bmatrix}$$

45. $D_v f(x_0, y_0) = \frac{63}{52}; \ -\frac{7}{5}; \ \frac{3-\pi}{2\sqrt{5}}.$

46. This quite a bit of work if you expand f, and a quick computation if you work with the given form. Answer: I.

47. The steps for solving this are: Compare to theorem 2.31, find the gradient of f at the given point, normalise, choose one of the two possible directions. Answer: $v = \begin{bmatrix} -0.6 & 0.8 \end{bmatrix}^\top.$

48. The matrix M contains the derivatives of the transformation $(s, t) \rightsquigarrow (x, y)$, cf. Theorem 3.52.

49.
$$T^{(2)}_{(2,1)} f(x, y) = -12y^2 - 12xy + 12x - 36y - 24$$

50. You can verify your answer by checking the conditions in Remark 2.39. To make sure that your second derivatives are correct, compare to the example at the end of the previous section.

51.
$$T^{(2)}_{(5,-2)} f(x, y) = -\frac{x^2}{4} - y^2 - xy + x + 2y + \frac{\pi - 3}{4}$$

52. $y = x - \frac{3}{2}.$

53. Only one critical point: a saddle point at $(-2, 4)$.

54. The critical points are
$$(0,0), \quad (2,0), \quad (1,1), \quad (1,-1),$$
and they are a maximum, minimum, saddle point, saddle point, respectively.

55. You could contrapose that statement and/or look at Theorem 2.31 for inspiration.

56. $f(x,y) = x^4 + y^4$ covers one of those cases.

57. Classifying P is quite difficult. Zooming in by appending `for x from 1.99 to 2.01 and y from -0.51 to -0.49` to the WolframAlpha plot command provides some insight.

58. Can any of the other points in S have larger function values?

59. Try to sketch a *first-contact* scenario where the tangent lines at the contact point intersect (you will not succeed; be aware that level curves of smooth functions do not have ends – they are either closed curves or they go out to infinity).

60. Maximising that function amounts to finding the point of the curve that has the largest x-coordinate: $x_{\max} = 1$, taken at $(x,y) = (1,0)$.

61. The maximum and minimum values of f are $\pm 5^{5/2}$, taken at the points $(\pm 1, \pm\sqrt{5})$.

62. A point on S, that is closest to the sphere, is also closest to its centre; considering different cases helps solving $\nabla F = 0$; to check your answer: you should obtain a minimum distance of 2.

63. You should obtain four extrema with distances 1, 1.936, and 4 to the origin. The minima are both local and global, the maximum is only local. Make sure to state the extrema (the points at which the minimal/maximal distances are realised), as they are part of what you were asked for.

64. Mark the areas $n \cdot A_0$ and $(n+m) \cdot A_0$ in the sketch above, and then think about what would happen if we had tiles of half the side length

65. A $A_0(x)$ and $A_1(x)$ should differ by the constant $^{-11}/_4$ – where does this number come from? By choosing the constant of integration suitably, you should be able to obtain both A_0 and A_1 from the indefinite integral.

66. You will need the formula for the sum of the first n terms of the geometric series.

67. Cf. remark 3.2 for the meaning of 'well-defined" in this context. For fixed n, we obtain the largest possible Riemann sum by letting the c_j be the points at which f takes its maximal value on $[x_{j-1}, x_j]$. Similarly for the smallest possible Riemann sum. What is the difference between those two extremes, and what happens for $n \to \infty$?

68. Careful, the FTC can not immediately by applied, because the x appears in the integrand as well. Hence pull out the x and then differentiate the product $x \cdot \int_0^{x^2} \ldots$. Note that there is a different way to solve this: find an explicit formula for F in terms of x – without an integral – and then differentiate. Try this approach as well and make sure your answers agree!

69. First, you need a function whose graph is the circle or part of the circle – use the formula $x^2 + y^2 = R^2$ for that.

70. The value you obtained is correct if it rounds to 0.309.

71. It is useful to consider the area function with $a = 0$, $A(x) = \int_0^x (1 + |t|)^2 \, dt$.

72. Taking the union of the set of p values, for which the first improper exists, with the set of p's for the second, you should obtain the whole real line with only one point missing.

73. Compute I_0 directly and use integration by parts to find a recursive formula for I_n. Can you derive an explicit formula!

74. The value you obtained is correct if it rounds to $1.571 + 1.317 = 2.888$.

75. The pyramid is quite symmetric – this allows to consider a sub-domain S of the base and later multiply the volume that is sitting on top of it by a suitable factor. Choose S so that the walls of the pyramid do not have any edges over it – i.e. the restriction of the shell to S is a plane!

76. Sketch the domain and the smallest rectangle that contains it, and argue that the integrand $f(x) \cdot f(y)$ integrates to $2I$ over the rectangle.

77. If you divide that number by $\iint_D 1\,dA$, the result will help you balance a cut-out of D on a single point.

78. E.g., letting y be the outer integral: there are three different sets of formulas for the x boundaries.

79. The value you obtained is correct if it rounds to 39.478.

80. You will need the general version of theorem 3.52 and "spherical polar coordinates" for this.
    Answer: The value you obtained is correct if it rounds to 47.124.

81. Check the initial condition for both possible choices.

82. Initially, it does not make sense, but setting the solution equal to 0 from a certain t_0 on, one obtains a function for the water level $h(t)$ that does agrees with how one would imagine a tank to drain.

83. 20 seconds.

84. $y(x) \approx x \cdot \arcsin(\ln x + 0.866)$, but you should obtain the constant exactly.

85. $y(x) = \frac{\sin^2 x}{\cos x}$.

86. Perhaps stating this more formally makes it more clear what needs to be done: Let $f = f(x,y)$ be continuous on $\mathbb{R}^2 \backslash \{(0,0)\}$. Then

$$\exists\, F = F(t) : f(x,y) = F(^y/_x) \,\forall x \neq 0 \iff f(x,y) = f(\lambda x, \lambda y)\,\forall \lambda \neq 0.$$

The direction "$\Longrightarrow$" is the easier one. The case $x = 0$ causes some extra work – one can use continuity for that – but it would be alright to skip this technicality.

87. The general solution is $y_{\text{gen}}(x) = A\,e^x \cos x + B\,e^x \sin x + 5$, and you can verify the constants you have found for the particular solution by checking whether it satisfies the boundary conditions.

88. $y(x) = A\,e^{-x} + B\,x e^{-x} + ^1/_2\,x^2 e^{-x} + ^5/_{32} \cosh(3x) - ^3/_{32} \sinh(3x)$, where the hyperbolic functions could also be written out as a sum of exponential expressions.

89. No need to provide the answer here – just check it yourself in the way described in the previous exercise. This is a very good thing to do, as it shows you what we are actually doing here.

90. $u(x,y) = (y+1) \cdot \cosh x$, but, again, this would have been easy to check – always try to check your answers yourself!

91. Use the same change of variables as for s-o PDEs, and then make a suitable choice for (a,b). It is less straightforward what to do now, but with some flexibility, you will find a solution (containing a single-variable function) that can be easily verified.