STEFAN WEIGERT

# CALCULUS

# PART II

SPRING TERM 2023

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF YORK, UK

# *Contents*

# 1
# Preliminaries

WELCOME to the second part of the module *Calculus*. In the coming weeks we will study five fundamental topics of Calculus. You will have the opportunity to learn—or reinforce—important material that is going to be useful throughout your degree: *Second Order Ordinary Differential Equations*, *Fourier Series*, *Functions of More than One Variable*, *Double Integrals* and *Extrema*.

The table of contents of the notes gives you a first idea of the overall layout of the module, serving as a convenient summary of what to expect over the entire Spring term. After attending the lectures, working through the notes, engaging with the *seminars* and completing the *assignments* you should have a good grasp of these topics. I strongly recommend that you **do the exercises as you go along**, including those not set for hand-in.

## What these notes are

These notes cover the material that will be presented in **three weekly lectures** over **nine weeks**. In principle, the notes are meant to be self-contained. Each week, you should carefully work through the theory developed and the examples given.If you notice any errors, please let me know by emailing `stefan.weigert@york.ac.uk`.

The notes are based on material collected and edited over time by Eli Hawkins, Atsushi Higuchi, Christopher Hughes, Niall MacKay and Roger Colbeck.

## Additional resources

There is additional material available **online** which you should know about: a Moodle page with general and specific information about this module, recordings of this year's videos, a series of videos created for the online version of the course delivered two years ago, and a textbook.

### *Moodle*

*Moodle* is the *Virtual Learning Environment* of the Department of Mathematics which you will be familiar with by now. On *Moodle*, you will find a page dedicated to this module which you should visit frequently. It will be used to communicate further information about the module such as hand-in dates for assignments, past papers, etc. There is also a *discussion forum* which you can use to ask questions and answer those of other students, and another one for *announcements* about the course. You will need to be logged into Moodle for the link to work.

Calculus folder on Moodle

### *Textbook*

Occasionally you will find references such as "§17.9" in the notes which point to relevant chapters of the textbook *Thomas' Calculus*. You may use it for background reading and alternative perspectives on the topics we study; you also find additional exercises. This widely-used textbook is now in its 14th edition. Most editions can be used interchangeably but make sure to look at the contents if you are thinking about buying.

MD Wier, GB Thomas, FR Giordano, JR Hass: *Thomas' Calculus* (Pearson, 2010; 11th edition).

### *Videos*

This course comes with two sets of videos, both available via *Panopto*, the system used by the University of York to provide online content. Firstly, there are the recordings of the *live lectures* of this year, which you can use as a catch-up if you are unable to attend in person. Secondly, there are *videos clips* recorded for the course in the Spring Term of 2021 which had no in-person lectures. You can also access all recordings using the links on the *Moodle* page of the module.

Panopto
Spring term 2023 lectures recordings

Panopto
Spring term 2021 videos

Links to the relevant recordings from the online version of the course are provided in the margins of these notes. Again, you need to be logged in for the link to work. The clips will guide you through derivations, explain concepts, or develop new aspects of the material presented in the notes.
Clicking the link in the margin should send you to a welcome message for this module dating from two years ago. If a link in this document happens not to work, you should still be able to access the folder with the recordings through *Moodle*.

Welcome message

***

Now, with the technicalities out of the way:

LET'S BEGIN!

# 2

# Second Order Ordinary Differential Equations

*Main idea*

A second order differential equation is a relation between an independent variable $x$, an unknown function $y(x)$, say, and its first two derivatives with respect to $x$,

$$R(y'', y', y, x) = 0\,, \tag{2.1}$$

which may also include some explicitly given functions.

## 2.1 Reduction of order

Sometimes the problem of solving a second order ODE can be effectively reduced to solving a *first order* ODE.

ODE is short for ordinary differential equation characterized by the presence of a single independent variable.

*Missing dependent variable*

The more obvious case where this approach is possible occurs if the *dependent* variable does not appear directly in the equation—only its first and second derivatives appear. In this case, Eq. (2.1) takes the form

$$R(y'', y', x) = 0\,, \tag{2.2}$$

V01: Missing dependent variable

Differential equations of this type can often be solved by rewriting them as first order equations for the first derivative of the unknown function.

**Example 2.1** (Constant acceleration)**.** For $a \in \mathbb{R}$ constant, find the general solution of

$$\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} = a. \tag{2.3}$$

*Solution.* Since $x$ does not appear directly, we can treat the velocity, $v := \dot{x}$ (a dot means $\frac{\mathrm{d}}{\mathrm{d}t}$) as a new dependent variable. The equation becomes

$$\frac{\mathrm{d}v}{\mathrm{d}t} = a$$

which integrates to

$$v = at + v_0\,,$$

with a constant of integration $v_0$ given by the value of $v$ at $t = 0$. Now we remember that $\dot{x} = v$ and integrate once more, to find

$$x = \int v \, \mathrm{d}t = \tfrac{1}{2}at^2 + v_0 t + x_0,$$

where the constant of integration $x_0$ is just $x(0)$. $\triangleleft$

To check this result, compute $\dot{x} = at + v_0$ which indeed leads to $\ddot{x} = a$.

**Example 2.2.** Solve

$$\frac{\mathrm{d}^2 y}{\mathrm{d}x^2} = -\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2$$

with the initial conditions $y(0) = 0$ and $y'(0) = 1$, where the prime denotes a derivative with respect to the variable $x$, i.e., is a shorthand for $\frac{\mathrm{d}}{\mathrm{d}x}$.

*Solution.* Define $u := y'$ (Step 1). This gives the equation

$$\frac{\mathrm{d}u}{\mathrm{d}x} = -u^2,$$

which is separable. Integrating (Step 2) gives

$$x = -\int \frac{1}{u^2}\, \mathrm{d}u = x_0 + u^{-1},$$

where $x_0$ is a constant, hence $y' = u = (x - x_0)^{-1}$, so that, finally (Step 3)

$$y = \int \frac{1}{x - x_0}\, \mathrm{d}x = C + \log(x - x_0).$$

The initial conditions require

$$0 = y(0) = C + \log(-x_0) \quad \text{and} \quad 1 = y'(0) = -x_0^{-1}\,.$$

The second equation implies that $x_0 = -1$. With this, the first equation simplifies to

$$0 = C + \log 1 = C.$$

This shows that the function

$$y(x) = \log(x + 1)$$

is the solution to the initial value problem. $\triangleleft$

As a check we can compute $y' = (x + 1)^{-1}$ and $y'' = -(x + 1)^{-2}$ which satisfy $y'' = -(y')^2$, as required.

**Algorithm 1**
Second Order ODE with Missing Dependent Variable

1. Define a new dependent variable which is the derivative of the old dependent variable.
2. Solve the resulting first order equation.
3. Integrate the solution.

### *Missing independent variable*

A similar approach works–in a slightly more subtle way—for any *autonomous* ODE, i.e., whenever the *independent* variable does not appear explicitly. This situation is another special case of Eq. (2.1)

V02: Missing independent variable

$$R(y'', y', y) = 0\,, \tag{2.4}$$

With $x$ not being present, one can in fact remove every reference to it from the differential equation, simply by considering $y$ to be a new *independent* variable and define $u := \frac{\mathrm{d}y}{\mathrm{d}x}$ to be the new *dependent* variable. The crucial observation is to rewrite the second derivative as follows:

$$\frac{\mathrm{d}^2 y}{\mathrm{d}x^2} = \frac{\mathrm{d}u}{\mathrm{d}x} = \frac{\mathrm{d}y}{\mathrm{d}x}\frac{\mathrm{d}u}{\mathrm{d}y} = u\frac{\mathrm{d}u}{\mathrm{d}y}\,, \tag{2.5}$$

which eliminates the derivatives with respect to $x$.

**Example 2.3** (Constant acceleration, again)**.** Equation (2.3) is also autonomous. It can also be solved by this method. Assume $a > 0$.

*Solution.* Again use the velocity $v = \dot{x}$ as dependent variable, but now write the equation as

$$a = \ddot{x} = v\frac{\mathrm{d}v}{\mathrm{d}x}.$$

This is a separable differential equation since we have $a\,\mathrm{d}x = v\,\mathrm{d}v$. Integrating both sides leads to so

$$a(x + A) = \tfrac{1}{2}v^2 \quad \text{or} \quad \frac{\mathrm{d}x}{\mathrm{d}t} = v = \pm\sqrt{2a(x + A)}.$$

This is another separable equation which implies

$$t = \pm \int \frac{\mathrm{d}x}{\sqrt{2a(x + A)}} = B \pm \sqrt{\frac{2(x + A)}{a}}\,.$$

Finally, solving for $x$, gives

$$x = \tfrac{a}{2}(t - B)^2 - A = \tfrac{1}{2}at^2 - aBt + \left(\tfrac{aB^2}{2} - A\right)$$

This is equivalent to the previous solution; the constants have different names, but we can identify $v_0 = -aB$ and $x_0 = \frac{aB^2}{2} - A$. ◁

**Algorithm 2**
Autonomous Second Order ODE

1. Define a new independent variable as the old dependent variable.
2. Define a new dependent variable as the derivative of the old dependent variable.
3. Rewrite the old second derivative in terms of these new variables as in Eq. (2.5).
4. Solve the resulting first order equation.
5. Rewrite the solution in terms of the original variables (giving another first order ODE).
6. Solve this.

**Example 2.4.** Solve

$$0 = y'' + y\ .$$

*Solution.* Define $u := y'$ to rewrite the equation as $0 = u\frac{\mathrm{d}u}{\mathrm{d}y} + y$. Hence, we have

$$0 = y\,\mathrm{d}y + u\,\mathrm{d}u \quad \text{so that} \quad y^2 + u^2 = C_1\,,$$

with some constant $C_1$. Thus, $\frac{\mathrm{d}y}{\mathrm{d}x} = u = \pm\sqrt{C_1 - y^2}$ which implies that, for some range of $x$ and $y$, the solution is

$$\begin{aligned} x &= \pm\int \frac{\mathrm{d}y}{\sqrt{C_1 - y^2}} \\ &= \pm\int \frac{\mathrm{d}t}{\sqrt{1-t^2}} = C_2 \pm \sin^{-1} t = C_2 \pm \sin^{-1}\left(\frac{y}{\sqrt{C_1}}\right)\,, \end{aligned}$$

where we have introduced the variable $t := y/\sqrt{C_1}$. Solving for $y$, this gives

$$y = \pm\sqrt{C_1}\sin(x - C_2).$$

Using the angle addition formula for the sine, this function can also be written as

$$y = A\cos x + B\sin x \tag{2.6}$$

where $A := \mp\sqrt{C_1}\sin C_2$ and $B := \pm\sqrt{C_1}\cos C_2$. ◁

As can be readily checked, both expressions indeed solve the original differential equation. Taking the form in Eq. (2.6), for instance, we have $y' = -A\sin x + B\cos x$ and thus $y'' = -A\cos x - B\sin x$, so that finally $y'' + y = 0$.

## 2.2 *Homogeneous linear equations*

A *homogeneous linear equation* is one in which each term contains exactly one factor of the dependent variable or one of its derivatives. Even though we cannot always solve such an equation, we know that the general solution always has the same simple form.

Any second order, homogeneous linear ODE (with independent variable $x$ and dependent variable $y$) can be written in the form

$$\frac{\mathrm{d}^2 y}{\mathrm{d}x^2} + p(x)\frac{\mathrm{d}y}{\mathrm{d}x} + q(x)y = 0. \tag{2.7}$$

This relation is a rather specific case of the general relation (2.1). However, this class of ordinary differential equations has the convenient property that any *linear combination* of solutions is also a solution.

V03:Linear homogeneous ODEs

**Definition 2.1** (Linear combination)**.** A *linear combination* of functions $f_1, \ldots, f_n$ is any sum $A_1 f_1 + \cdots + A_n f_n$, where $A_1, \ldots, A_n$ are constants. A set of functions is *linearly independent* if none of the functions is a linear combination of the others.

It will sometimes be useful to give a name to the left side of Eq. (2.7), by writing

$$L[y] := \frac{\mathrm{d}^2 y}{\mathrm{d}x^2} + p(x)\frac{\mathrm{d}y}{\mathrm{d}x} + q(x)y\ . \tag{2.8}$$

The quantity $L[\cdot]$ (sometimes we'll just write $L$) is known as a *differential operator*; it is a device that takes one function as an input and outputs another function. A solution of Eq. (2.7) now is seen to be any function that is mapped to zero—or annihilated—by the operator $L$, i.e., $L[y] = 0$ holds.

The operator $L$ is also *linear* in the sense that $L[y_1 + y_2] = L[y_1] + L[y_2]$ and for any constant, $A$, $L[Ay] = A\,L[y]$. This has an immediate, nice consequence.

**Theorem 2.1** (Superposition)**.** *Any linear combination of two solutions of Eq.* (2.7) *is also a solution of Eq.* (2.7).

*Proof.* Let $y_1$ and $y_2$ be solutions of Eq. (2.7), and $A$ and $B$ be constants. If $y = Ay_1 + By_2$, then

$$L[y] = L[Ay_1] + L[By_2] = A\,L[y_1] + B\,L[y_2] = 0\ ,$$

therefore $y$ is a solution of Eq. (2.7). □

So, if $y_1$ and $y_2$ are linearly independent solutions of (2.7), then the arbitrary linear combination $Ay_1 + By_2$ is a two-parameter family of distinct solutions, and so it is the general solution of Eq. (2.7).

These results carry over to homogeneous linear ODEs of order $n$, but in that case we will need $n$ independent solutions to construct the general solution.

### *Constant coefficients*

Homogeneous linear, second order ODEs are in general impossible to solve explicitly, so we now consider just *autonomous* ones — i.e., those with constant coefficients. This situation corresponds to having constants instead of $x$-dependant functions $p(x)$ and $q(x)$ in Eq. (2.8).

**Example 2.5.** Solve the first order homogeneous linear ODE with constant coefficients given by

$$\frac{\mathrm{d}y}{\mathrm{d}x} + ay = 0\,,$$

with some constant $a$.

*Solution.* This is separable, as $\frac{\mathrm{d}y}{y} = -a\,\mathrm{d}x$. Integrating gives

$$\log y = -ax + c,$$

and hence the general solution reads

$$y = C\mathrm{e}^{-ax}\,,$$

depending on one free parameter $C = \mathrm{e}^{c}$. ◁

Before tackling second order equations with constant coefficients, let us mention an important part of solving mathematical problems which is to use an *educated guess* that we call an Ansatz.

V04: Auxiliary equation

**Definition 2.2** (Ansatz)**.** An *ansatz* is a guess for the form of the solution to a problem such as a differential equation. Typically, the proposed solution includes one or more free parameters. The idea is then to check whether the guess solves the problem for particular values of the free parameters.

The most general second order, homogeneous linear ODE with constant coefficients can be written as

$$y'' + ay' + by = 0 \tag{2.9}$$

Could we take some inspiration from Example 2.5, possibly solving Eq. (2.9) by an exponential function? In other words, we might try the exponential function $y = \mathrm{e}^{\lambda x}$ as an ansatz for solving the homogeneous linear differential equation with constant coefficients. If we do so, we obtain

$$\begin{aligned} 0 &= y'' + ay' + by \\ &= \lambda^2\mathrm{e}^{\lambda x} + a\lambda\mathrm{e}^{\lambda x} + b\mathrm{e}^{\lambda x} = (\lambda^2 + a\lambda + b)\mathrm{e}^{\lambda x}\,. \end{aligned}$$

Since $e^{\lambda x} \neq 0$, we see that the parameter $\lambda$ must satisfy a constraint in order that the ansatz be a solution,

$$0 = \lambda^2 + a\lambda + b\,. \tag{2.10}$$

This equation is called the *auxiliary equation* (or *auxiliary polynomial*) associated with Eq. (2.9). The auxiliary equation can be obtained by replacing the $n^{\text{th}}$ derivative of $y$ with the $n^{\text{th}}$ power of $\lambda$.

This recipe works equally well for higher order linear equations with constant coefficients.

If the auxiliary equation (2.10) has *different roots*, then they give *different solutions* to Eq. (2.9). Taking linear combinations of them results in the general solution of the original equation.

Let us go through a number of explicit examples for some practice.

**Example 2.6.** Find the solution of

$$y'' + y' - 2y = 0 \tag{2.11}$$

with initial conditions $y(0) = 0$ and $y'(0) = 1$.

*Solution.* Assuming a solution of the form $y = e^{\lambda x}$ we find the auxiliary equation

$$0 = \lambda^2 + \lambda - 2 = (\lambda + 2)(\lambda - 1),$$

so $e^{-2x}$ and $e^x$ are solutions of the differential equation, and the linear combination

$$y = Ae^{-2x} + Be^x$$

is the general solution, depending on two constant $A$ and $B$.

The first derivative is $y' = -2Ae^{-2x} + Be^x$, so the initial conditions require

$$\begin{aligned} 0 &= y(0) = A + B \qquad \Longrightarrow \qquad B = -A \\ 1 &= y'(0) = -2A + B = -3A \end{aligned}$$

therefore $A = -\frac{1}{3}$ and $B = \frac{1}{3}$. Hence, the solution to the initial value problem is given by

$$y(x) = -\tfrac{1}{3}e^{-2x} + \tfrac{1}{3}e^x. \qquad \triangleleft$$

To verify this result we can compute the first two derivatives $y' = \frac{2}{3}e^{-2x} + \frac{1}{3}e^x$ and $y'' = -\frac{4}{3}e^{-2x} + \frac{1}{3}e^x$, to confirm that

$$y'' + y' - 2y = \left(-\tfrac{4}{3}e^{-2x} + \tfrac{1}{3}e^x\right) + \left(\tfrac{2}{3}e^{-2x} + \tfrac{1}{3}e^x\right) + 2\left(-\tfrac{1}{3}e^{-2x} + \tfrac{1}{3}e^x\right) = 0,$$

$$y(0) = -\tfrac{1}{3} + \tfrac{1}{3} = 0,$$

and that

$$y'(0) = \tfrac{2}{3} + \tfrac{1}{3} = 1.$$

**Example 2.7.** Find the general solution of

$$y'' + y = 0,$$

again, now using an exponential function as an ansatz. [Recall that in Example 2.4 we solved this another way.]

*Solution.* Taking $y = \mathrm{e}^{\lambda x}$, the auxiliary equation becomes $0 = \lambda^2 + 1$. Its roots are imaginary, $\lambda = \pm\mathrm{i}$, which means that the complex exponentials $\mathrm{e}^{\pm \mathrm{i}x}$ are solutions of the differential equation. Their linear combination provides the desired general solution $y = C\mathrm{e}^{\mathrm{i}x} + D\mathrm{e}^{-\mathrm{i}x}$. ◁

While correct, the expression in terms of the exponentials $\mathrm{e}^{\pm\mathrm{i}x}$ is not always convenient if we are interested in *real* solutions. However, we can write

For instance, it may be clear from the problem we are solving that the solution must be real.

$$\begin{aligned} y &= C(\cos x + \mathrm{i}\sin x) + D(\cos x - \mathrm{i}\sin x) \\ &= (C+D)\cos x + \mathrm{i}(C-D)\sin x \\ &= A\cos x + B\sin x\ , \end{aligned}$$

where $A = C + D$ and $B = \mathrm{i}(C - D)$. The advantage of this formula is that $y$ is real if and only if $A$ and $B$ are real numbers.

V05: Complex roots

In general, if $\alpha + \mathrm{i}\beta$ and $\alpha - \mathrm{i}\beta$ are roots of the auxiliary equation, then these give two solutions of the differential equation:

$$\mathrm{e}^{(\alpha\pm\mathrm{i}\beta)x} = \mathrm{e}^{\alpha x}\mathrm{e}^{\pm\mathrm{i}\beta x} = \mathrm{e}^{\alpha x}(\cos\beta x \pm \mathrm{i}\sin\beta x).$$

We can combine them to obtain a pair of *real* linearly independent solutions, namely

$$\tfrac{1}{2}\left[\mathrm{e}^{(\alpha+\mathrm{i}\beta)x} + \mathrm{e}^{(\alpha-\mathrm{i}\beta)x}\right] = \mathrm{e}^{\alpha x}\cos\beta x$$

and

$$\tfrac{1}{2\mathrm{i}}\left[\mathrm{e}^{(\alpha+\mathrm{i}\beta)x} - \mathrm{e}^{(\alpha-\mathrm{i}\beta)x}\right] = \mathrm{e}^{\alpha x}\sin\beta x.$$

If the ODE has real coefficients then any complex roots of the auxiliary equation always occur with their conjugate.

**Example 2.8.** Find the general solution of

$$y'' - 2y' + 2y = 0.$$

*Solution.* Assuming $y = \mathrm{e}^{\lambda x}$ we get the auxiliary equation $0 = \lambda^2 - 2\lambda + 2$, which has roots $\lambda = 1 \pm i$, so $\mathrm{e}^x\cos x$ and $\mathrm{e}^x\sin x$ are solutions of the differential equation.

Combining these two solutions gives the general solution,

$$y = A\mathrm{e}^x\cos x + B\mathrm{e}^x\sin x = \mathrm{e}^x(A\cos x + B\sin x). \qquad \triangleleft$$

What happens if the auxiliary equation has repeated roots? In that case, exponential functions on their own will not provide the necessary number of solutions.

If we are solving a second order equation we expect there to be two unknown constants in the general solution, corresponding to two linearly independent solutions of the homogeneous equation.

**Example 2.9.** Find the general solution of

$$y'' - 2y' + y = 0. \tag{2.12}$$

*Solution.* Assuming $y = \mathrm{e}^{\lambda x}$ we obtain the auxiliary equation $0 = \lambda^2 - 2\lambda + 1 = (\lambda - 1)^2$, therefore $y = \mathrm{e}^x$ is a solution, but there is no second, independent, exponential solution.

V06: Repeated roots

By factorizing the operator, we can rewrite the differential equation as

$$0 = \left(\frac{\mathrm{d}}{\mathrm{d}x} - 1\right)^2 y\,,$$

which suggests writing $u := \left(\frac{\mathrm{d}}{\mathrm{d}x} - 1\right) y = y' - y$, so that

$$0 = u' - u\,.$$

The general solution of this first-order ODE is $u = A\mathrm{e}^x$, giving

$$y' - y = A\mathrm{e}^x\,.$$

This is another first-order linear differential equation which can be solved by spotting the *integrating factor* $\mathrm{e}^{-x}$,

$$\mathrm{e}^{-x}(y' - y) = \frac{\mathrm{d}}{\mathrm{d}x}\left(\mathrm{e}^{-x}y\right) = A\,.$$

Integrating gives $\mathrm{e}^{-x}y = B + Ax$, which means that the function

$$y = (B + Ax)\mathrm{e}^x$$

is the general solution. ◁

To check our answer, we can compute $y' = (B + A + Ax)\mathrm{e}^x$, $y'' = (B + 2A + Ax)\mathrm{e}^x$, and confirm that the terms in Eq. (2.12) add up to zero,

$$y'' - 2y' + y = (B + 2A + Ax)\mathrm{e}^x - 2(B + A + Ax)\mathrm{e}^x + (B + Ax)\mathrm{e}^x = 0.$$

Any homogeneous linear ODE with constant coefficients has solutions of the form just described. If $\lambda$ is a root of the auxiliary equation and is repeated $n$ times, then $x^k\mathrm{e}^{\lambda x}$ is a solution for $k = 0, \ldots, n-1$. Combining these together will generate the general solution.

**Example 2.10.** Find the general solution of

$$y''' + y'' - y' - y = 0.$$

**Algorithm 3**
Homogeneous Linear ODE for $y(x)$ with Constant Coefficients

1. Assume a solution of the form $e^{\lambda x}$.
2. Substitute this into the differential equation to form the auxiliary equation (this will have the $n^{\text{th}}$ derivative $\frac{d^n y}{dx^n}$ replaced by $\lambda^n$).
3. Solve the auxiliary equation.
4. If $\lambda$ is a real root of the auxiliary equation, then $e^{\lambda x}$ is a solution of the ODE.
5. If $\alpha + i\beta$ and $\alpha - i\beta$ are complex conjugate roots of the auxiliary equation, then $e^{\alpha x} \cos \beta x$ and $e^{\alpha x} \sin \beta x$ are solutions of the ODE.
6. If $\lambda$ is an $m$-times repeated root of the auxiliary equation ($m \leq n$), then multiplying the above solutions by powers of $x$ up to $x^{m-1}$ gives more solutions.
7. The general solution of the ODE is an arbitrary linear combination of the above solutions.

*Solution.* Assuming $y = e^{\lambda x}$ gives the auxiliary equation

$$\begin{aligned} 0 &= \lambda^3 + \lambda^2 - \lambda - 1 \\ &= (\lambda - 1)(\lambda + 1)^2. \end{aligned}$$

This has roots $\lambda_1 = 1$ (not repeated) and $\lambda_2 = -1$ (repeated twice), so we have a general solution of the form

$$y = Ae^x + (B + Cx)e^{-x}\,. \qquad \triangleleft$$

Again, the verification is straightforward upon calculating all three derivatives,

$$\begin{aligned} y' &= Ae^x + (C - B - Cx)e^{-x}\,, \\ y'' &= Ae^x + (B - 2C + Cx)e^{-x}, \\ y''' &= Ae^x + (3C - B - Cx)e^{-x}\,. \end{aligned}$$

Hence we find that indeed

$$\begin{aligned} y''' + y'' - y' - y = {} & (A + A - A - A)e^x \\ & + [(3C - B) + (B - 2C) - (C - B) - B]e^{-x} \\ & + (-C + C - [-C] - C)e^{-x} \\ = {} & 0. \end{aligned}$$

## 2.3 *Inhomogeneous linear equations*

Inhomogeneous linear equations are more difficult to solve than homogeneous ones, but the general solution still has a generic and fairly simple form.

Any inhomogeneous linear ODE of second order can be written in the form

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} + p(x)\frac{\mathrm{d}y}{\mathrm{d}x} + q(x)y = r(x)\ , \qquad (2.13)$$

or, in terms of the of the linear differential operator $L[\cdot]$, as

$$L[y] = r\ .$$

Now there is a useful theorem which establishes a fundamental relation between any two solutions of Eq. (2.13); the solutions of the *homogeneous* differential equation associated with it play a crucial role.

**Theorem 2.2.** *If $y_p$ is a particular solution of Eq. (2.13), then any other solution of (2.13) can be written as*

$$y = y_p + u, \qquad (2.14)$$

*where $u$ is a solution of the homogeneous equation $L[u] = 0$. Furthermore, any sum of the form (2.14) is a solution of Eq. (2.13).*

V07: 2nd-order inhomogeneous linear ODEs

*Proof.* If $y$ is some solution of Eq. (2.13) and we define $u := y - y_p$, then

$$L[u] = L[y - y_p] = L[y] - L[y_p] = r - r = 0\ .$$

Conversely, if we have $L[u] = 0$ and define $y := y_p + u$, then

$$L[y] = L[y_p + u] = L[y_p] + L[u] = r\ ,$$

so $y$ is indeed a solution of Eq. (2.13). □

If $y_p$ is a *particular* solution of a second-order inhomogeneous linear ODE, and $u_1$ and $u_2$ are two independent solutions of the corresponding homogeneous equation $L[u] = 0$, then

$$y = y_p + Au_1 + Bu_2$$

is the *general* solution. An equivalent result holds for inhomogeneous linear ODEs of higher order.

Thus, if we know how to solve the homogeneous equation, the remaining problem is to find *one particular solution* of the inhomogeneous equation. If $r(x)$ is a sum, then we can work term by term as stated by the next theorem.

**Theorem 2.3** (Linearity). *If $L[y_1] = r_1$ and $L[y_2] = r_2$, then $y := y_1 + y_2$ is a solution of $L[y] = r_1 + r_2$.*

**Algorithm 4**
Inhomogeneous Linear ODE

1. Find the general solution of the corresponding homogeneous equation.
2. Find one solution of the inhomogeneous equation (see Algorithm 5 for how to do this for equations with constant coefficients).
3. Add them together to obtain the general solution.

### *Constant coefficients*

We now consider the special case of a second-order, linear ODE with constant coefficients and some inhomogeneity $r(x)$, i.e., $p(x) = a$ and $q(x) = b$ in Eq. (2.13) are constant, but $r(x)$ might not be. The general form of the differential equation now reads

$$L[y] := y'' + ay' + by = r\,, \tag{2.15}$$

with $a, b \in \mathbb{R}$ and $r(x)$ is a given function.

There are sophisticated techniques for solving inhomogeneous linear differential equations but will limit ourselves to *the method of undetermined coefficients*. In this method, we will use an *ansatz* to find a particular solution of the differential equation. This method only works if the function $r(x)$ that makes the differential equation inhomogeneous, is a *sum of products of polynomials, exponential functions, and sines and cosines*—in other words, the type of function that is a solution of some *homogeneous* linear ODE with constant coefficients. This means we will be able to treat a case such as

$$r(x) = x^5 + x^2 \mathrm{e}^{5x} - \cos 3x + (1 + x^2)\mathrm{e}^{-2x} \sin x\,, \tag{2.16}$$

but inhomogeneities of the form $r(x) = \frac{x}{\mathrm{e}^x - 1}$ or $r(x) = \tan x$ would not be covered.

The idea of the method is that if we apply the operator $L$ to functions of this kind, then it will output similar functions albeit with possibly different coefficients. To obtain exactly the desired output $r(x)$—i.e., a solution of the inhomogeneous differential equation, we will try to inject a function resembling $r$ which, however, contains undetermined coefficients. The approach is best explained by working through examples which we will do for the remainder of this section.

**Example 2.11.** Find the general solution of

$$L[y] = y'' - 5y' + 6y = x.$$

*Solution.* Assuming $y = \mathrm{e}^{\lambda x}$ leads to the auxiliary equation is $0 = \lambda^2 -$

V08: Constant coefficients

$5\lambda + 6 = (\lambda - 2)(\lambda - 3)$, so the general solution of the homogeneous equation $L[u] = 0$ is $u(x) = A\mathrm{e}^{2x} + B\mathrm{e}^{3x}$. Now we need to hunt for a particular solution.

Since the inhomogeneous term is a polynomial of degree one, we take as an ansatz for $y_p$ the most general polynomial of that type, i.e., $y_p = \alpha x + \beta$, where $\alpha$ and $\beta$ are to be determined. We have $y_p' = \alpha$ and $y_p'' = 0$, which leads to the condition

$$x = L[\alpha x + \beta] = 0 - 5\alpha + 6(\alpha x + \beta) = 6\alpha x + (6\beta - 5\alpha).$$

The constraint is satisfied for $\alpha = \frac{1}{6}$ and $\beta = \frac{5}{36}$, so we have found a *particular solution* of the original differential equation,

$$y_p(x) = \tfrac{1}{6}x + \tfrac{5}{36}\,,$$

allowing us to write down the general solution,

$$y(x) = \tfrac{1}{6}x + \tfrac{5}{36} + A\mathrm{e}^{2x} + B\mathrm{e}^{3x}\,. \qquad \triangleleft$$

We better check for possible mistakes. To do so, we compute the derivatives $y' = \frac{1}{6} + 2A\mathrm{e}^{2x} + 3B\mathrm{e}^{3x}$ and $y'' = 4A\mathrm{e}^{2x} + 9B\mathrm{e}^{3x}$ of our proposed solution. Hence we confirm that

$$\begin{aligned} y'' - 5y' + 6y &= 4A\mathrm{e}^{2x} + 9B\mathrm{e}^{3x} - 5(\tfrac{1}{6} + 2A\mathrm{e}^{2x} + 3B\mathrm{e}^{3x}) \\ &\qquad + 6(\tfrac{1}{6}x + \tfrac{5}{36} + A\mathrm{e}^{2x} + B\mathrm{e}^{3x}) \\ &= x. \end{aligned}$$

**Example 2.12.** Find the solution of

$$y'' - 5y' + 6y = \mathrm{e}^x$$

with the initial conditions $y(0) = y'(0) = 0$.

*Solution.* Try $y_p = \alpha\mathrm{e}^x$. To determine $\alpha$, note that

$$L[\mathrm{e}^x] = \mathrm{e}^x - 5\mathrm{e}^x + 6\mathrm{e}^x = 2\mathrm{e}^x\,,$$

so $L[\frac{1}{2}\mathrm{e}^x] = \mathrm{e}^x$, and we have the particular solution,

$$y_p(x) = \tfrac{1}{2}\mathrm{e}^x.$$

The homogeneous equation is the same as in the previous example, so the general solution is

$$y(x) = \tfrac{1}{2}\mathrm{e}^x + A\mathrm{e}^{2x} + B\mathrm{e}^{3x}.$$

Its derivative equals $y'(x) = \frac{1}{2}\mathrm{e}^x + 2A\mathrm{e}^{2x} + 3B\mathrm{e}^{3x}$, so the initial conditions require

$$\begin{aligned} 0 &= y(0) = \tfrac{1}{2} + A + B, \\ 0 &= y'(0) = \tfrac{1}{2} + 2A + 3B. \end{aligned}$$

Subtracting twice the first equation from the second one gives $0 = -\frac{1}{2} + B$, so $B = \frac{1}{2}$. Plugging this value back into the first equation gives $A = -1$. Therefore the solution to the initial value problem is

$$y(x) = \tfrac{1}{2}\mathrm{e}^{x} - \mathrm{e}^{2x} + \tfrac{1}{2}\mathrm{e}^{3x}. \qquad \triangleleft$$

**Example 2.13.** Find the general solution of

$$L[y] = y'' - 5y' + 6y = \mathrm{e}^{2x}. \tag{2.17}$$

*Solution.* In this case, a multiple of $\mathrm{e}^{2x}$ does not work because it would be mapped to zero by $L$, i.e., $L[\mathrm{e}^{2x}] = 0$. Instead, let us include an additional factor of $x$, i.e., we will try $y = \alpha x \mathrm{e}^{2x}$ containing one undetermined coefficient. Using the derivatives $y' = \alpha(2x+1)\mathrm{e}^{2x}$ and $y'' = \alpha(4x+4)\mathrm{e}^{2x}$, we need to have

V09: A degenerate case

$$\mathrm{e}^{2x} = \alpha\left[(4x+4) - 5(2x+1) + 6x\right]\mathrm{e}^{2x} = -\alpha\mathrm{e}^{2x},$$

which implies $\alpha = -1$. Hence, we have found a particular solution

$$y_p(x) = -x\mathrm{e}^{2x}.$$

The homogeneous equation associated with Eq. (2.17) is the same as in Example 2.11, so the general solution is

$$y(x) = -x\mathrm{e}^{2x} + A\mathrm{e}^{2x} + B\mathrm{e}^{3x} = (A - x)\mathrm{e}^{2x} + B\mathrm{e}^{3x}. \qquad \triangleleft$$

*Remark* 2.1. There is no need to include a plain $\mathrm{e}^{2x}$ term in the ansatz $y_p$ because it would be annihilated by $L$ and hence contribute nothing to the sought-after particular solution.

**Example 2.14.** Find the general solution of

$$y'' - 5y' + 6y = \mathrm{e}^{x} + \mathrm{e}^{2x}.$$

*Solution.* The left-hand-side is the same as in the previous examples, and the terms on the right are, in fact, a sum of those from the previous examples. Thus, a particular solution is given by the sum of those particular solutions,

$$y_p(x) = \tfrac{1}{2}\mathrm{e}^{x} - x\mathrm{e}^{2x},$$

and the general solution is given by adding it to the solution of the homogeneous equation.

$$y(x) = \tfrac{1}{2}\mathrm{e}^{x} - x\mathrm{e}^{2x} + A\mathrm{e}^{2x} + B\mathrm{e}^{3x} = \tfrac{1}{2}\mathrm{e}^{x} + (A - x)\mathrm{e}^{2x} + B\mathrm{e}^{3x}. \qquad \triangleleft$$

**Example 2.15.** Find the general solution of

$$L[y] = y'' + 2y' - 3y = x\mathrm{e}^{x}.$$

*Solution.* Assuming $y = \mathrm{e}^{\lambda x}$ we obtain the auxiliary equation $0 = \lambda^2 + 2\lambda - 3 = (\lambda - 1)(\lambda + 3)$, so the general solution of the homogeneous equation is $A\mathrm{e}^{x} + B\mathrm{e}^{-3x}$.

Since $x\mathrm{e}^x$ is a polynomial of degree one times $\mathrm{e}^x$, it is natural to use the function $y = (\alpha x + \beta)\mathrm{e}^x$ as an ansatz, with derivatives $y' = (\alpha x + [\alpha + \beta])\mathrm{e}^x$ and $y'' = (\alpha x + [2\alpha + \beta])\mathrm{e}^x$. However, we find that

$$\begin{aligned} y'' + 2y' - 3y &= (\alpha x + [2\alpha + \beta])\mathrm{e}^x + 2(\alpha x + [\alpha + \beta])\mathrm{e}^x - 3(\alpha x + \beta)\mathrm{e}^x \\ &= 4\alpha \mathrm{e}^x\,, \end{aligned}$$

which cannot equal $x\mathrm{e}^x$ and hence our ansatz does not work. The reason is that $L[\mathrm{e}^x] = 0$, so the second part of our ansatz is not needed. The situation is similar to Example 2.13 where $\mathrm{e}^{2x}$ was also found to be a solution to the homogeneous equation and hence would not help in constructing a particular solution. However, if we generalize the ansatz to $y = (\alpha x^2 + \beta x)\mathrm{e}^x$, we require

$$\begin{aligned} x\mathrm{e}^{2x} &= \Big((\alpha x^2 + [4\alpha + \beta]x + [2\alpha + 2\beta]) \\ &\qquad + 2(\alpha x^2 + [2\alpha + \beta]x + \beta) - 3(\alpha x^2 + \beta x)\Big)\,\mathrm{e}^x \\ &= (8\alpha x + [2\alpha + 4\beta])\mathrm{e}^x. \end{aligned}$$

Now two parameters are present and we can satisfy the condition for $\alpha = \frac{1}{8}$ and $\beta = -\frac{1}{16}$, leading to the particular solution

$$y_p(x) = \left(\tfrac{1}{8}x^2 - \tfrac{1}{16}x\right)\mathrm{e}^x.$$

Finally, we can spell out the general solution of the ODE, namely

$$y(x) = \left(\tfrac{1}{8}x^2 - \tfrac{1}{16}x + A\right)\mathrm{e}^x + B\mathrm{e}^{-3x}. \qquad \triangleleft$$

Note that we need to multiply by $x$ precisely when the right hand side of the equation is of the same type as a solution of the homogeneous equation. Without introducing the additional factor of $x$, the ansatz for a particular solution "overlaps" with the general solution of the homogeneous equation.

For this reason it is advisable to solve the homogeneous equation before looking for a particular solution.

**Example 2.16.** Find the general solution of

$$L[y] = y'' + 2y' - 3y = 2\cos x.$$

*Solution.* Acting with the operator $L$ on the function $\cos x$ does not just produce a multiple of it; instead $L[\cos(x)]$ outputs a linear combination of the trigonometric functions $\cos x$ and $\sin x$. Hence, a reasonable guess to obtain a particular solution is to try the ansatz $y_p = \alpha\cos x + \beta \sin x$. We compute

$$\begin{aligned} y_p'' + 2y_p' - 3y_p &= -\alpha\cos x - \beta\sin x \\ &\qquad + 2(-\alpha\sin x + \beta\cos x) - 3(\alpha\cos x + \beta\sin x) \\ &= (-4\alpha + 2\beta)\cos x + (-2\alpha - 4\beta)\sin x. \end{aligned}$$

**Algorithm 5**
Inhomogeneous Linear Equation with Constant Coefficients

1. Guess a particular solution of the same type as the right hand side of the differential equation, but with undetermined coefficients.
   (a) The allowed types of functions are polynomials, exponentials, sines and cosines, and products of any of these.
   (b) Polynomials should be of the same degree as on the right hand side.
   (c) The coefficients of $x$ in exponentials or in sines and cosines should be the same as on the right hand side.
   (d) Sines and cosines are of the same type.
2. If this ansatz overlaps with the general solution of the homogeneous equation, then multiply that part of the guess by $x$. [Because of this, it is advisable to solve the homogeneous equation before trying to find a particular solution.]
3. Insert this ansatz into the differential equation and determine the values of the coefficients for which it is a solution.
4. Insert these values into the ansatz to get a particular solution.
5. Add this particular solution to the general solution of the homogeneous equation to get the general solution of the differential equation.

For this to equal $2\cos x$ we need

$$2\beta - 4\alpha = 2, \quad \text{and} \quad 2\alpha + 4\beta = 0\,,$$

giving $\alpha = -2/5$ and $\beta = 1/5$ which result in the general solution

$$y(x) = Ae^x + Be^{-3x} - \tfrac{2}{5}\cos x + \tfrac{1}{5}\sin x. \qquad \triangleleft$$

In some cases, the use of complex numbers allows us to set up an alternative approach to construct particular solutions of inhomogeneous equations. It is conceptually slightly more sophisticated but computationally simpler.

V10: Complex-valued inhomogeneity

**Example 2.17.** Find the general solution of

$$L[y] = y'' + 2y' - 3y = e^x \cos x. \tag{2.18}$$

*Solution.* Instead, let us introduce a differential equation for a complex-valued function $z(x)$,

$$L[z] = e^{(1+i)x}. \tag{2.19}$$

where the inhomogeneous term $\mathrm{e}^x \cos x$ has been replace by a suitable complex-valued function whose real part coincides with the inhomogeneity present in the original equation. Then, the *real part* $\operatorname{Re} z(x)$ of a complex-valued solution $z(x)$ of the new differential equation, is a solution of (2.18) since

$$\operatorname{Re}\left(\mathrm{e}^{(1+\mathrm{i})x}\right) = \frac{1}{2}\left(\mathrm{e}^{(1+\mathrm{i})x} + \mathrm{e}^{(1-\mathrm{i})x}\right) = \mathrm{e}^x \cos x\,.$$

We try the ansatz $z = \alpha \mathrm{e}^{(1+\mathrm{i})x}$, allowing the free parameter $\alpha$ to be a complex number, $\alpha \in \mathbb{C}$. The derivatives of $\mathrm{e}^{(1+\mathrm{i})x}$ are

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x}\mathrm{e}^{(1+\mathrm{i})x} &= (1+\mathrm{i})\mathrm{e}^{(1+\mathrm{i})x} \quad \text{and} \\ \frac{\mathrm{d}^2}{\mathrm{d}x^2}\mathrm{e}^{(1+\mathrm{i})x} &= (1+\mathrm{i})^2\mathrm{e}^{(1+\mathrm{i})x} = 2\mathrm{i}\mathrm{e}^{(1+\mathrm{i})x}\,. \end{aligned}$$

Hence, applying the operator $L$ to $\mathrm{e}^{(1+\mathrm{i})x}$ is equivalent to a multiplication of the ansatz by

$$2\mathrm{i} + 2(1+\mathrm{i}) - 3 = -1 + 4\mathrm{i}\,.$$

Consequently, for the value

$$\alpha = \frac{1}{-1+4\mathrm{i}} = \frac{1}{-1+4\mathrm{i}} \cdot \frac{-1-4\mathrm{i}}{-1-4\mathrm{i}} = \frac{-1-4\mathrm{i}}{17}\,,$$

the ansatz turns into a particular solution of Eq. (2.19), given by

$$z_p(x) = \frac{-1-4\mathrm{i}}{17}\mathrm{e}^{(1+\mathrm{i})x}\,.$$

A particular solution of Eq. (2.18) is given by the real part of $z_p(x)$,

Note that we get a particular solution to $y'' + 2y' - 3y = \mathrm{e}^x \sin x$ at the same time by taking the imaginary part of $z_p$.

$$\begin{aligned} y_p(x) &= \operatorname{Re}\left(z_p(x)\right) = \operatorname{Re}\left(\tfrac{-1-4\mathrm{i}}{17}\mathrm{e}^{(1+\mathrm{i})x}\right) \\ &= \tfrac{1}{17}\mathrm{e}^x\left(-\operatorname{Re}\mathrm{e}^{\mathrm{i}x} - 4\operatorname{Re}\mathrm{i}\mathrm{e}^{\mathrm{i}x}\right) = \tfrac{1}{17}\mathrm{e}^x\left(-\cos x + 4\sin x\right)\,. \end{aligned}$$

With the homogeneous equation being the same as in the previous two examples, the general solution of Eq. (2.18) is finally given by

$$y(x) = \tfrac{1}{17}\mathrm{e}^x\left(-\cos x + 4\sin x\right) + A\mathrm{e}^x + B\mathrm{e}^{-3x}. \qquad \triangleleft$$

It is a useful exercise to instead determine the particular solution $y_p(x)$ using the real differential equation Eq. (2.18) as it will show that the calculations are more cumbersome in that case.

## 2.4 *Coupled first-order ODEs*

We can solve a system of two coupled first-order ODEs by turning it into a single second-order ODE. In the following examples, the independent variable is $t$, the dependent variables are $x$ and $y$, and dots denote derivatives with respect to $t$, often called "time."

V11: Coupled 1st-order ODEs

**Example 2.18.** Find the general solution of the system of equations

$$\dot{x} = y \tag{2.20a}$$
$$\dot{y} = -x. \tag{2.20b}$$

*Solution.* The derivative of Eq. (2.20a) is

$$\ddot{x} = \dot{y}\,,$$

which we can use to eliminate $\dot{y}$ from Eq. (2.20b), resulting in a second-order ODE for the unknown function $x(t)$,

$$\ddot{x} = -x.$$

The general solution of this equation can be found using the methods introduced earlier and is

$$x = A\cos t + B\sin t\,,$$

and it now becomes straightforward to find $y$ using Eq. (2.20a),

$$y = \dot{x} = B\cos t - A\sin t\,. \qquad \triangleleft$$

The *general* solution to a system of two coupled first order ODEs contains two constants, one for each integration which must be performed while constructing the solutions. To fix a unique solution, we need two initial conditions: the values of $x$ and $y$ at some specific time $t$ such as $t = 0$.

**Example 2.19.** Solve the system of equations

$$\dot{x} = 5x + 6y \tag{2.21a}$$
$$\dot{y} = -3x - 4y \tag{2.21b}$$

with initial conditions

$$x(0) = 1 \quad \text{and} \quad y(0) = 0.$$

*Solution.* Differentiate Eq. (2.21a),

$$\begin{aligned}
\ddot{x} &= 5\dot{x} + 6\dot{y} \\
&= 5\dot{x} + 6(-3x - 4y) = 5\dot{x} - 18x - 24y && \text{by (2.21b)} \\
&= 5\dot{x} - 18x - 4(\dot{x} - 5x) = \dot{x} + 2x && \text{by (2.21a)}.
\end{aligned}$$

Therefore, the unknown function $x(t)$ satisfies $0 = \ddot{x} - \dot{x} - 2x$. This is a second-order ODE with constant coefficients which can be solved with the ansatz $x = \mathrm{e}^{\lambda t}$. We obtain the auxiliary equation

$$0 = \lambda^2 - \lambda - 2 = (\lambda - 2)(\lambda + 1)\,,$$

which implies that the general solution is given by

$$x(t) = A\mathrm{e}^{2t} + B\mathrm{e}^{-t}.$$

To find $y(t)$ we calculate the derivative $\dot{x} = 2A\mathrm{e}^{2t} - B\mathrm{e}^{-t}$ and use Eq. (2.21a),

$$6y = \dot{x} - 5x = (2-5)A\mathrm{e}^{2t} + (-1-5)B\mathrm{e}^{-t} = -3A\mathrm{e}^{2t} - 6B\mathrm{e}^{-t}.$$

Hence, the pair $(x(t), y(t))$ with

$$y(t) = -\tfrac{1}{2}A\mathrm{e}^{2t} - B\mathrm{e}^{-t},$$

represents the general solution of the system of equations (2.21).

Now we must identify the specific solutions singled out by the initial conditions. The conditions are

$$1 = x(0) = A + B \quad \text{and} \quad 0 = y(0) = -\tfrac{1}{2}A - B.$$

Adding these equations gives $1 = \frac{1}{2}A$, so $A = 2$ and $B = -1$. Hence, the final solution to the initial value problem is given by the functions

$$x(t) = 2\mathrm{e}^{2t} - \mathrm{e}^{-t} \quad \text{and} \quad y(t) = -\mathrm{e}^{2t} + \mathrm{e}^{-t}. \qquad \triangleleft$$

As always, it is a good idea to verify one's computations: we easily confirm that

$$\begin{aligned} \dot{x} &= 4\mathrm{e}^{2t} + \mathrm{e}^{-t} \\ 5x + 6y &= 10\mathrm{e}^{2t} - 5\mathrm{e}^{-t} - 6\mathrm{e}^{2t} + 6\mathrm{e}^{-t} = 4\mathrm{e}^{2t} + \mathrm{e}^{-t}, \end{aligned}$$

and that

$$\begin{aligned} \dot{y} &= -2\mathrm{e}^{2t} - \mathrm{e}^{-t} \\ -3x - 4y &= -6\mathrm{e}^{2t} + 3\mathrm{e}^{-t} + 4\mathrm{e}^{2t} - 4\mathrm{e}^{-t} \\ &= -2\mathrm{e}^{2t} - \mathrm{e}^{-t}. \end{aligned}$$

Here is another example of coupled equations, only marginally more involved than the previous one.

**Example 2.20.** Determine the solutions of the coupled pair of equations

$$\dot{x} = 3x - y\,, \tag{2.22a}$$
$$\dot{y} = 4x - y\,. \tag{2.22b}$$

*Solution.* First, we construct a second-order ODE for the function $x(t)$ by differentiating Eq. (2.22a) and plugging (2.22b) and Eq. (2.22a) into the result,

$$\begin{aligned} \ddot{x} &= 3\dot{x} - \dot{y} = 3\dot{x} - (4x - y) \\ &= 3\dot{x} - 4x + y = 2\dot{x} - x\,. \end{aligned}$$

**Algorithm 6**
A System of Coupled ODEs for $x(t)$ and $y(t)$

1. Differentiate the equation for $\dot{x}$, giving a formula for $\ddot{x}$ in terms of $\dot{x}$ and $\dot{y}$.

2. Use the equation for $\dot{y}$ to eliminate $\dot{y}$.

3. Use the equation for $\dot{x}$ to eliminate $y$, thus getting a second order ODE for $x(t)$.

4. Find the general solution of this ODE.

5. Compute the derivative $\dot{x}(t)$ of this solution.

6. Use the equation for $\dot{x}$ to write $y$ in terms of $x$ and $\dot{x}$, and compute $y(t)$ from the general solution for $x(t)$.

(Note: This procedure works equally well if one constructs the second-order ODE for the function $y(t)$ first.)

Next, we assume a solution of the form $x = e^{\lambda t}$ to generate the auxiliary equation

$$0 = \lambda^2 - 2\lambda + 1 = (\lambda - 1)^2 ,$$

which has the *double* root $\lambda = 1$. Hence, the solution takes the form

$$x = (A + Bt)e^t .$$

Finally, we use the fact that we know the function $x(t)$ to construct $y(t)$. Upon calculating $\dot{x}(t) = (A + B + Bt)e^t$, we can obtain $y(t)$ from Eq. (2.22a) as

$$\begin{aligned} y &= 3x - \dot{x} \\ &= 3(A + Bt)e^t - (A + B + Bt)e^t \\ &= ([2A - B] + 2Bt)e^t . \end{aligned}$$

Combining what we found, we write down the general solution to the system of equations (2.22),

$$x = (A + Bt)e^t \quad \text{and} \quad y = ([2A - B] + 2Bt)e^t . \qquad \triangleleft$$

To verify our result is a matter of working out derivatives:

$$\begin{aligned} \dot{x} &= (A + B + Bt)e^t \\ 3x - y &= (3A + 3Bt)e^t - ([2A - B] + 2Bt)e^t = (A + B + Bt)e^t , \end{aligned}$$

and

$$\begin{aligned} \dot{y} &= ([2A + B] + 2Bt)e^t \\ 4x - y &= 4(A + Bt)e^t - ([2A - B] + 2Bt)e^t \\ &= ([2A + B] + 2Bt)e^t , \end{aligned}$$

hence the equations are satisfied.

The final example proceeds along the same lines except that we encounter another auxiliary polynomial with complex-valued roots, as in Eq. (2.20).

**Example 2.21.** Determine the solutions of the coupled pair of equations

$$\dot{x} = x + y\,, \tag{2.23a}$$

$$\dot{y} = -x + y\,. \tag{2.23b}$$

*Solution.* Upon eliminating the function $y(t)$, we obtain a second-order ODE for $x(t)$,

$$0 = \ddot{x} - 2\dot{x} + 2x\,,$$

which, after assuming $x = \mathrm{e}^{\lambda t}$ gives the auxiliary equation

V12: Spiralling out

$$0 = \lambda^2 - 2\lambda + 2$$

having roots $\lambda_\pm = 1 \pm \mathrm{i}$. Hence, we find the general expression for the function $x(t)$, namely

$$x = (A\cos t + B\sin t)\mathrm{e}^t\,,$$

which depends on two free parameters, $A, B \in \mathbb{R}$. Solving Eq. (2.23a) for $y$, we obtain upon differentiating $x(t)$ that

$$y = (B\cos t - A\sin t)\mathrm{e}^t\,,$$

which completes our calculation: the general solution to the system of equations (2.23) is given by the pair

$$x(t) = (A\cos t + B\sin t)\mathrm{e}^t \quad \text{and} \quad y(t) = (B\cos t - A\sin t)\mathrm{e}^t\,. \quad \lhd$$

# 3
# Fourier Series

## A little bit of history

In 1807, Fourier made the astonishing claim that *any* function defined on some interval could be expanded as an infinite sum of sines and cosines. In fact, mathematicians had already employed series of this type. In particular, D. Bernoulli had made use of them in his work on vibrating strings (1753) and Euler was aware of them, too. However, it was thought that such an expansion could only be valid for analytic, i.e., specific well-behaved functions. The radical, and not entirely correct, element in Fourier's proposal was the assertion that the expansion is valid for arbitrary functions. Over the years that followed, many mathematicians of the 19th century—including Gauss, Cauchy, Dirichlet and Riemann—gradually put Fourier's work on a sound footing and clarified the class of functions to which it applies. In the process, they resolved many fundamental problems concerning the nature of the real numbers, the concept of a function, the theory of infinite series and integration. Arguably, much of modern analysis was developed specifically to make sense of Fourier series.

## Main idea

Suppose that the note A is played on a violin and is recorded with a microphone. The note A has a frequency of 440 Hz, meaning that the signal received by the microphone repeats itself 440 times per second. However, this signal is not a simple sine wave. It also contains *overtones* at multiples of the fundamental frequency. We can decompose the sound of the violin as a sum of pure tones (sine waves) at each of these frequencies. This is essentially what a Fourier series is.

The simplest sine wave is $\sin x$ which has period $2\pi$ (or equivalently, a frequency of $\frac{1}{2\pi}$). Any sine wave with period $2\pi$ is proportional to $\sin(x + x_0)$ and hence can be written as

$$a_1 \cos x + b_1 \sin x,$$

for some constants $a_1$ and $b_1$. Any sine wave oscillating with an integer multiple of the fundamental frequency takes the form,

$$a_n \cos nx + b_n \sin nx \,,$$

with some further constants. Adding up all these sine waves (and one with frequency 0) leads to what is known as a *Fourier series*

$$\begin{aligned} S(x) &= \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx) \qquad (3.1) \\ &= \frac{a_0}{2} + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + a_3 \cos 3x + b_3 \sin 3x + \dots \end{aligned}$$

where $a_0, a_1, a_2, \dots$ and $b_1, b_2, \dots$ are real numbers. The factor of $\frac{1}{2}$ in front of the coefficient $a_0$ is a convention whose reasoning will be justified later.

The main interest in Fourier theory is the fact that any sufficiently well behaved function defined on $[-\pi, \pi]$ can be written in the form (3.1), for suitable constants $\{a_n\}_{n=0}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$. A Fourier series is similar in spirit to a Taylor series in which a function $f(x)$ is decomposed into powers of $(x - x_0)$. A Fourier series uses sines and cosines instead but aims to achieve the same goal.

We will mainly consider functions defined on $[-\pi, \pi]$, but everything we do can easily be extended to other intervals — see Section 3.6.

Being an infinite series, a Taylor series is, effectively, defined only as a limit of *partial sums*. That is also true for a Fourier series which we should think of as the limit

$$S(x) := \lim_{N \to \infty} S_N(x)$$

with finite partial sums

$$S_N(x) := \frac{a_0}{2} + \sum_{n=1}^{N} (a_n \cos nx + b_n \sin nx) \,. \qquad (3.2)$$

If $f(x) = S(x)$, then the partial sums approximate the function $f(x)$, i.e., $f(x) \approx S_N(x)$, and the approximation gets better as $N$ increases and converges as $N \to \infty$.

We can think of a Fourier series in the form (3.1) as defined on $[-\pi, \pi]$, but we can also consider it to be defined on the whole real line. In the latter case, $S(x)$ is always $2\pi$-periodic, i.e., $S(x + 2k\pi) = S(x)$ for any integer $k$.

Most functions cannot be represented by Taylor series. Fourier series can be used to represent a much broader class of functions including discontinuous ones. The importance of Fourier series arises partly because they are valid for a wide class of functions, and partly because the trigonometric functions have many nice properties.

V11: Taylor and Fourier series

Fourier expansions frequently reduce complex problems in mathematics, physics and engineering to much simpler, tractable problems.

The theory was developed in order to describe the conduction of heat; today, Fourier series have applications ranging from signal processing to quantum theory, including JPEG image compression and MPEG video compression.

## 3.1 Fourier coefficients

Given a function $f$, how can we work out the coefficients $a_n$ and $b_n$ of its Fourier series in Eq. (3.1)? Interestingly, the decomposition of $f$ into a sum over sines and cosines is conceptually equivalent to the decomposition of a vector $\underline{r} \in \mathbb{R}$ into basis vectors,

$$\underline{r} = x\,\underline{i} + y\,\underline{j} + z\,\underline{k}\,. \tag{3.3}$$

Let us recall how we determine the coefficients $x$, $y$, and $z$ from a given vector $\underline{r}$. The essential tool to use is the *scalar product*: multiplying both sides of Eq. (3.3) with the unit vector $\underline{i}$, for example, we find

$$\begin{aligned} \underline{i}\cdot\underline{r} &= x\,\underline{i}\cdot\underline{i} + y\,\underline{i}\cdot\underline{j} + z\,\underline{i}\cdot\underline{k} \\ &= x\,1 + y\,0 + z\,0 \\ &= x\,, \end{aligned} \tag{3.4}$$

using the fact that the basic vectors $\underline{i}$, $\underline{j}$ and $\underline{k}$ form an orthonormal triad, i.e., they are *pairwise orthogonal* and have *length one*. Somewhat surprisingly, it turns out that it makes perfect sense to think of the sine and cosine functions as *pairwise orthogonal vectors* in a higher-dimensional space if we introduce an appropriate *inner product for functions*.

Given two functions defined on the interval $-\pi \leq x \leq \pi$, i.e., $f, g : [-\pi, \pi] \to \mathbb{R}$, let us define their inner product as the integral over their product,

V12: Inner product for functions

$$\langle f, g\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)g(x)\,\mathrm{d}x. \tag{3.5}$$

This is a good choice because (see Exercise Sheet 2) we can directly confirm the relations that for integers $n, m \geq 1$,

$$\frac{1}{\pi}\int_{-\pi}^{\pi} \cos nx\,\cos mx\,\mathrm{d}x = \begin{cases} 1 & n = m \\ 0 & n \neq m \end{cases}$$

$$\frac{1}{\pi}\int_{-\pi}^{\pi} \sin nx\,\sin mx\,\mathrm{d}x = \begin{cases} 1 & n = m \\ 0 & n \neq m \end{cases}$$

$$\frac{1}{\pi}\int_{-\pi}^{\pi} \cos nx\,\sin mx\,\mathrm{d}x = 0$$

stating that the set of functions $\{\cos x, \sin x, \cos 2x, \sin 2x, \cos 3x, \sin 3x, \ldots\}$ are indeed "orthogonal", just as $\underline{i}$, $\underline{j}$ and $\underline{k}$ are—except that we have infinitely many of them. Note that we can add the constant function to this set, maintaining orthogonality, so the set of functions

$$\{1/\sqrt{2}, \cos x, \sin x, \cos 2x, \sin 2x, \cos 3x, \sin 3x, \ldots\}$$

Check that $\frac{1}{\pi}\int_{-\pi}^{\pi} \cos mx \,\mathrm{d}x = 0$, $\frac{1}{\pi}\int_{-\pi}^{\pi} \sin mx \,\mathrm{d}x = 0$. The choice $f_1 = 1/\sqrt{2}$ (rather than $f_1 = 1$) ensures that $\langle f_1, f_1 \rangle = 1$.

are orthogonal. If we name them $\{f_i\}_{i=1}^{\infty}$ in this order then

$$\langle f_n, f_m \rangle = \begin{cases} 1 & n = m \\ 0 & n \neq m \end{cases}.$$

Now, let $f : [-\pi, \pi] \to \mathbb{R}$ be some function and suppose that it is equal to a Fourier series (3.1), i.e., that there exist values of the constants $a_0, a_1, \ldots$ and $b_1, b_2, \ldots$ such that $f(x) = S(x)$ for $-\pi \leq x \leq \pi$. Multiplying both sides of Eq. (3.1) with $(1/\pi)\cos mx$ and integrating over the interval $[-\pi, \pi]$, we find for $m > 0$,

$$\begin{aligned}
\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos mx \,\mathrm{d}x &= \frac{a_0}{2\pi}\int_{-\pi}^{\pi} \cos mx \,\mathrm{d}x + \sum_{n=1}^{\infty} \frac{a_n}{\pi}\int_{-\pi}^{\pi} \cos nx \, \cos mx \,\mathrm{d}x \\
&\quad + \sum_{n=1}^{\infty} \frac{b_n}{\pi}\int_{-\pi}^{\pi} \sin nx \, \cos mx \,\mathrm{d}x \\
&= 0 + a_m + 0 \\
&= a_m.
\end{aligned}$$

Alternatively, we can write this using our inner product:

This is exactly analogous to what happens in Eqs. (3.4).

$$\begin{aligned}
\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos mx \,\mathrm{d}x &= \langle f(x), \cos mx \rangle \\
&= \left\langle \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos nx + b_n \sin nx\right), \cos mx \right\rangle \\
&= \frac{a_0}{\sqrt{2}} \langle \frac{1}{\sqrt{2}}, \cos mx \rangle + \sum_{n=1}^{\infty} \left(a_n \langle \cos nx, \cos mx \rangle + b_n \langle \sin nx, \cos mx \rangle\right) \\
&= a_m \,.
\end{aligned}$$

The case $m = 0$ must be treated separately, since $\cos(0x) \equiv 1$,

$$\begin{aligned}
\frac{1}{\pi}\int_{-\pi}^{\pi} f(x) \,\mathrm{d}x &= \frac{a_0}{2\pi}\int_{-\pi}^{\pi} \mathrm{d}x + \sum_{n=1}^{\infty} \frac{a_n}{\pi}\int_{-\pi}^{\pi} \cos nx \,\mathrm{d}x + \sum_{n=1}^{\infty} \frac{b_n}{\pi}\int_{-\pi}^{\pi} \sin nx \,\mathrm{d}x \\
&= a_0 + 0 + 0 \\
&= a_0.
\end{aligned}$$

Notice that the factor $\frac{1}{2}$ in Eq. (3.1) ensures that we get out exactly $a_0$ (instead of $2a_0$) which will be useful later on.

Now for the sine-functions which only come with $m > 0$, leading to a similar calculation,

$$\begin{aligned}
\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin mx\,\mathrm{d}x &= \frac{a_0}{2\pi}\int_{-\pi}^{\pi}\sin mx\,\mathrm{d}x + \sum_{n=1}^{\infty}\frac{a_n}{\pi}\int_{-\pi}^{\pi}\cos nx\,\sin mx\,\mathrm{d}x \\
&\quad + \sum_{n=1}^{\infty}\frac{b_n}{\pi}\int_{-\pi}^{\pi}\sin nx\,\sin mx\,\mathrm{d}x \\
&= 0 + 0 + b_m \\
&= b_m .
\end{aligned}$$

Thus, the Fourier coefficients which characterize a given function $f(x)$ over the interval $[-\pi, \pi]$, are given by the expressions

$$\begin{aligned}
a_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx\,\mathrm{d}x\,, \qquad n = 0, 1, 2, \dots\,, \\
b_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\,\mathrm{d}x\,, \qquad n = 1, 2, \dots
\end{aligned} \tag{3.6}$$

Let us summarize our observations in a definition.

**Definition 3.1** (Fourier series of a function)**.** The *Fourier series* associated with a function $f : [-\pi, \pi] \to \mathbb{R}$ is given by

$$S(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty}\left(a_n \cos nx + b_n \sin nx\right)\,,$$

with Fourier coefficients given by Eqs. (3.6).

We proceed with a first example of a calculation of the Fourier series of a specific function.

V13: Step function
This function is sometimes called the "step function" on the interval $[-\pi, \pi]$. You might like to sketch it.

**Example 3.1.** Find the Fourier series of $f : [-\pi, \pi] \to \mathbb{R}$ defined as

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

*Solution.*

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,\mathrm{d}x = \frac{1}{\pi}\int_{0}^{\pi}\mathrm{d}x = 1$$

All cosine-coefficients vanish since for $n > 0$ we have

$$a_n = \frac{1}{\pi}\int_0^{\pi}\cos nx\,\mathrm{d}x = \frac{1}{\pi}\left[\frac{\sin nx}{n}\right]_{x=0}^{x=\pi} = 0\,.$$

For the sine coefficients we will also do the calculations for all $n > 0$ in one go:

$$\begin{aligned}
b_n &= \frac{1}{\pi}\int_0^{\pi}\sin nx\,\mathrm{d}x = \frac{1}{\pi}\left[\frac{-\cos nx}{n}\right]_{x=0}^{x=\pi} \\
&= \frac{1-(-1)^n}{n\pi} = \begin{cases} \frac{2}{n\pi} & n \text{ odd} \\ 0 & n \text{ even} \end{cases}
\end{aligned}$$

Note that $1-(-1)^n = 0$ for $n$ even, so all of the even-numbered terms vanish.

Putting our results together, we have obtained the Fourier series of the unit step function.

$$S(x) = \frac{1}{2} + \sum_{n=1}^{\infty} \frac{1-(-1)^n}{n\pi} \sin nx \,.$$

This expression can be simplified since half of the terms are zero: we only need to sum over the odd integers. Since any odd number can be written as $n = 2k+1$, with $k$ an integer, we can rewrite the Fourier series in another form containing only non-zero terms,

$$S(x) = \frac{1}{2} + \sum_{k=0}^{\infty} \frac{2}{(2k+1)\pi} \sin\left[(2k+1)x\right]. \qquad \triangleleft$$

*Even and odd functions*

**Definition 3.2** (Even and odd functions)**.** A function $f$ is called *even* if $f(-x) = f(x)$, and *odd* if $f(-x) = -f(x)$.

The reason for these names is that the Taylor series of an even function about the origin $x = 0$ contains only even powers of $x$, and for an odd function it contains only odd powers.

These are useful in the present context because the Fourier series of an odd function does not need any odd functions for its construction, and, likewise, the Fourier series of an even function does not need any even functions for its construction. The functions $\cos nx, n \in \mathbb{N}$, are even, since $\cos(-nx) = \cos nx$, and the functions $\sin nx$ are odd, $\sin(-nx) = -\sin nx$. Hence the Fourier series of an odd function has $a_n = 0$ for all $n$, and the Fourier series of an even function has $b_n = 0$ for all $n$. This property is a direct consequence of Eq. (3.6), using the fact that the integral from $-\pi$ to $\pi$ of an odd function vanishes. Using symmetry can save work calculating coefficients.

**Example 3.2.** Find the Fourier series of $f : [-\pi, \pi] \to \mathbb{R}$, $f(x) = x$.

*Solution.* Because $f$ is an odd function, we conclude that $a_n = 0$ for all $n$. The other coefficients can be found from

V14: Fourier series of $f(x) = x$

$$\begin{aligned} b_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} x \sin nx \,\mathrm{d}x \\ &= \frac{-1}{n\pi} \left\{ \left[x \cos nx\right]_{x=-\pi}^{x=\pi} - \int_{-\pi}^{\pi} \cos nx \,\mathrm{d}x \right\} = \frac{-1}{n\pi} \left[ x \cos nx - \frac{\sin nx}{n} \right]_{x=-\pi}^{x=\pi} \\ &= \frac{-1}{n\pi} 2\pi \cos n\pi = 2 \frac{(-1)^{n+1}}{n} \,. \end{aligned}$$

We hence find obtain

$$S(x) = 2\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \sin nx\,. \tag{3.7}$$

as the Fourier series. $\triangleleft$

*Remark* 3.1. We have not yet proven rigorously that a function is actually equal to the associated Fourier series. Interestingly, the series in (3.7) can be shown to coincide with the original function $f(x) = x$ whenever $x \in (-\pi, \pi)$ but it *fails* to reproduce the correct values at $x = \pm\pi$, i.e., the endpoints of the interval over which the function $f(x)$ had been defined. This can be seen directly by evaluating the expression $S(x)$ at the endpoints giving $S(\pm\pi) = 0$, since each term of the sum on the right-hand-side of Eq. (3.7) is equal to zero: $\sin(\pm n\pi) = 0$, $n \in \mathbb{N}_0$. The somewhat subtle reason for the mismatch will be explained in Section 3.2, and because of possible mismatches we use a different symbol ($S$) for the Fourier series compared to the function itself ($f$).

**Example 3.3.** For some real number $\alpha \notin \mathbb{Z}$, find the Fourier series of $f : [-\pi, \pi] \to \mathbb{R}$, $f(x) = \cos\alpha x$.

*Solution.* This is an even function, so $b_n = 0$. The other coefficients can be found in one go:

$$\begin{aligned}
a_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} \cos\alpha x\,\cos nx\,\mathrm{d}x = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\cos(\alpha+n)x + \cos(\alpha-n)x\right]\mathrm{d}x \\
&= \frac{1}{2\pi}\left[\frac{\sin(\alpha+n)x}{\alpha+n} + \frac{\sin(\alpha-n)x}{\alpha-n}\right]_{x=-\pi}^{x=\pi} \\
&= \frac{1}{\pi}\left(\frac{\sin(\alpha+n)\pi}{\alpha+n} + \frac{\sin(\alpha-n)\pi}{\alpha-n}\right) = \frac{(-1)^n}{\pi}\left(\frac{1}{\alpha+n} + \frac{1}{\alpha-n}\right)\sin\pi\alpha \\
&= \frac{2\alpha(-1)^n \sin\pi\alpha}{\pi(\alpha^2-n^2)}\,,
\end{aligned}$$

so that we obtain the Fourier series

$$S(x) = \frac{\sin\pi\alpha}{\pi\alpha} + \frac{2\alpha\sin\pi\alpha}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^n}{\alpha^2-n^2}\cos nx\,. \qquad \triangleleft$$

*Remark* 3.2. In this case, the expansion can be shown to actually hold throughout the entire interval *including* the endpoints, i.e., $S(x) = \cos\alpha x$ for $-\pi \le x \le \pi$. This is in contrast to the Example 3.2, and we will find an explanation through Theorem 3.1 presented in the following section.

Here is another example where the Fourier series $S(x)$ of a function $f(x)$ does *not* match its values at the endpoint of the interval, $S(\pm\pi) \neq f(\pm\pi)$.

**Algorithm 7**
Fourier series of a function $f : [-\pi, \pi] \to \mathbb{R}$

1. The Fourier series of $f$ is
$S(x) := \dfrac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx)$.
2. Check for any symmetries of the function that imply certain coefficients are zero.
3. The $a_n$ coefficients can be calculated using
$a_n := \dfrac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx \, \mathrm{d}x$.
(although the same formula works, often the case $n = 0$ needs to be calculated separately — check for any division by $n$ in the expression for $a_n$ which would be come problematic if $n = 0$).
4. The $b_n$ coefficients can be calculated using
$b_n := \dfrac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx \, \mathrm{d}x$.

**Example 3.4.** Find the Fourier series of $f : [-\pi, \pi] \to \mathbb{R}$, $f(x) = \sinh x$.

*Solution.* This is an odd function, therefore $a_n = 0$ and we can calculate $b_n$ via

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} \sinh x \, \sin nx \, \mathrm{d}x = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{e}^x \sin nx \, \mathrm{d}x - \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{e}^{-x} \sin nx \, \mathrm{d}x \,.$$

In the last integral, substituting $t = -x$ gives

$$\int_{-\pi}^{\pi} \mathrm{e}^{-x} \sin nx \, \mathrm{d}x = \int_{\pi}^{-\pi} \mathrm{e}^{t} (-\sin nt)(-1) \, \mathrm{d}t = -\int_{-\pi}^{\pi} \mathrm{e}^{t} \sin nt \, \mathrm{d}t = -\int_{-\pi}^{\pi} \mathrm{e}^{x} \sin nx \, \mathrm{d}x,$$

so the two terms are equal. We continue

$$\begin{aligned}
b_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} \mathrm{e}^x \sin nx \, \mathrm{d}x \\
&= \frac{1}{\pi} \mathrm{Im} \left( \int_{-\pi}^{\pi} \mathrm{e}^{(1+\mathrm{i}n)x} \, \mathrm{d}x \right) \qquad \text{since } \sin nx = \mathrm{Im}\left(\mathrm{e}^{\mathrm{i}nx}\right) \\
&= \frac{1}{\pi} \mathrm{Im} \left( \left[ \frac{\mathrm{e}^{(1+\mathrm{i}n)x}}{1+\mathrm{i}n} \right]_{x=-\pi}^{x=\pi} \right) = \frac{1}{\pi} \mathrm{Im} \left( \frac{1-\mathrm{i}n}{1+n^2} (-1)^n \left( \mathrm{e}^{\pi} - \mathrm{e}^{-\pi} \right) \right) \\
&= (-1)^{n+1} \frac{2n \sinh \pi}{\pi(1+n^2)} \,.
\end{aligned}$$

Thus, we have found all required coefficients and

$$S(x) = \frac{2 \sinh \pi}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^{n+1} n}{n^2 + 1} \sin nx. \qquad \triangleleft$$

Theorem 3.1 will show that the Fourier series in the previous example agrees with $f(x) = \sinh x$ only in the *open* interval $\pi < x < \pi$.

## 3.2 Convergence

### *Piecewise continuously differentiable functions*

We began the discussion by effectively assuming that a Fourier series can be found to represent a given function, $S(x) = f(x)$, for $-\pi \leq x \leq \pi$, i.e., in a *closed* interval which contains the endpoints. On that basis we determined expressions for the coefficients $a_n$ and $b_n$ defining the Fourier series associated with the function $f(x)$. However, we have not proven that the formulae for the coefficients guarantee that $S(x) = f(x)$ holds everywhere in the interval considered. In fact, we found that this seems to be true only in some cases such $f(x) = \cos \alpha x$ (cf. Example 3.3) but for $f(x) = x$ and $f(x) = \sinh x$ (see Examples 3.2 and 3.4) we noticed some subtle discrepancy between $f(x)$ and its Fourier series $S(x)$. We need to clear this up.

To understand what is really going on, we first recall that a Fourier series is an *infinite* series, which is defined as a *limit* of finite sums—and limits are known to be tricky. The worst-case scenario is that (for some values of $x$) this limit does not exist. In that case we say that the Fourier series does not converge.

To be on safe ground, will now identify a class of functions that is broad enough to include our examples but, at the same time, restrictive enough so that their Fourier series actually converge. The following three definitions will help us to single out a desirable set of functions.

Let us recall what it means for a function to be *continuous*. A convenient tool to describe continuity are the *one-sided limits* of a function at a point.

**Definition 3.3** (One-sided limits)**.** The left and right limits of a function $f(x)$ at a point $x$ are given by

$$f(x^{\pm}) := \lim_{\xi \to x^{\pm}} f(\xi).$$

If it does not matter whether we approach a point $x$ from either the left or from the right, then (and only then) do we call $f$ *continuous* at $x$,

$$f(x) = f(x^{+}) = f(x^{-}) \, .$$

Let us relax this condition just a little bit by allowing a function $f(x)$ not to be continuous at a few points, resulting in the slightly weaker property of *piecewise continuity.*

**Definition 3.4** (Piecewise continuity)**.** A function $f$ is *piecewise continuous* on an interval $[a, b]$ if $f$ is continuous everywhere except at finitely many points in $[a, b]$, and at those points $f$ has at least finite limits from the left and right.

The graph of a piecewise continuous function is shown in Fig. 3.1. It consists of three continuous parts but the values of the function are

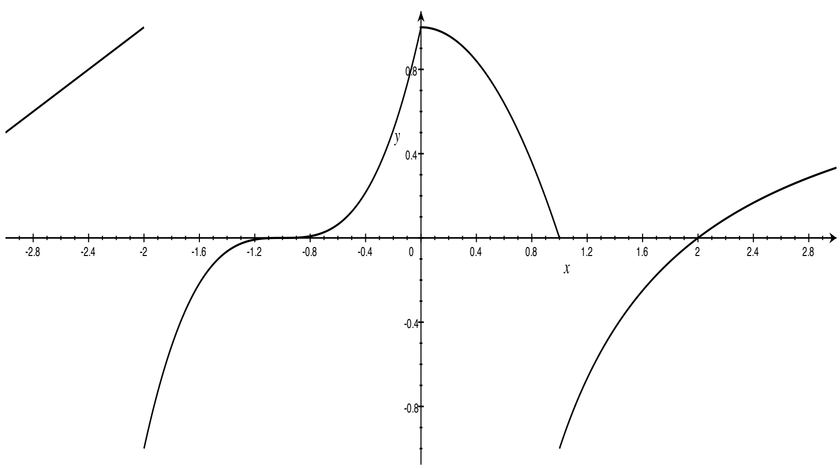



Figure 3.1: A piecewise continuous and piecewise continuously differentiable function.

seen to jump at the points $x = -2$ and $x = 1$, creating *discontinuities*. At these points the function is *not differentiable* since it is not even continuous. In addition, the slopes of the graph take different values when approaching the point $x = 0$ from the left or the right which also prevents differentiability at this point. However, since the derivative of the function remains well-defined for most values of $x$, it makes sense to introduce a slightly weaker version of differentiability.

V15: Discontinuities

**Definition 3.5** (Piecewise continuous differentiabilty)**.** A function $f$ is *piecewise continuously differentiable* if it is (i) piecewise continuous and (ii) differentiable except at finitely many points and (iii) its derivative is piecewise continuous.

The graph in Fig. 3.1 also serves as an example of a piecewise continuously differentiable function as it has the properties required by the definition. Note that the non-differentiability at the origin will manifest itself by a gap in the graph of the derivative, at $x = 0$.

### *Periodic extension of a function*

Typically, a Taylor series for a "nice" function $f(x)$—such as the one defined over $(0, 1)$ in Fig. 3.1—converges to it in the given interval but may not necessarily converge outside of that interval. In contrast, the Fourier series for a piecewise continuously differentiable function—such as the one defined over all of $[-\pi, \pi]$ in Fig. 3.1—will converge throughout but it may not equal the given function at some isolated points in $[-\pi, \pi]$.

Let $f(x)$ be a piecewise continuously differentiable function on the interval $[-\pi, \pi]$. As we shall see below in Theorem 3.1, the Fourier series of $f$ will converge, so we are safe talking about the associated function $S(x)$ defined by the Fourier series of $f(x)$.

Due to the periodicity of a Fourier series, $S(x + 2\pi) = S(x)$, it will just repeat the values of $f(x)$ from the interval $[-\pi, \pi)$ which are known from $S(x) = f(x)$ for $-\pi \leq x < \pi$. In other words, the Fourier series will converge for all $x \in \mathbb{R}$ and has well-defined values on the entire real line. If the function $f(x)$ had, from the outset, been defined differently outside of the interval $[-\pi, \pi)$, then the Fourier series would not "know" about it. We see that a Fourier series defined in the interval $[-\pi, \pi]$ naturally *extends* to all of $\mathbb{R}$, and we will apply that idea now arbitrary functions $f(x)$.

**Definition 3.6** (Periodic extension)**.** If $f : [-\pi, \pi) \to \mathbb{R}$, then its *periodic extension* $\tilde{f} : \mathbb{R} \to \mathbb{R}$ is defined by

$$\tilde{f}(x + 2\pi k) := f(x) \quad \text{for } k \in \mathbb{Z} \quad \text{and } -\pi \leq x < \pi.$$

Put differently, the periodic extension of a function $f(x)$ is the unique periodic function $\tilde{f}(x)$ such that they agree on the original interval, i.e., $\tilde{f}(x) = f(x)$ for $-\pi \leq x < \pi$. Even if $f(x)$ is continuous, its periodic extension need not be, depending on whether the value of $f(x)$ at the left endpoint of the interval, at $x = -\pi$, matches the limiting value when approaching the right endpoint from the left. If $f(x)$ is continuous, then the condition for the continuity of $\tilde{f}(x)$ can be expressed in terms of the function $f(x)$ as

V16: Periodic extension

$$f(-\pi) = f(\pi^-) \, .$$

Let us illustrate the two cases that may occur.

**Example 3.5.** Sketch the periodic extension of $f(x) = x^2$, $x \in [-\pi, \pi)$.

*Solution.* The periodic extension $\tilde{f}$ is shown in Fig. 3.2. Since we have $f(-\pi) = f(\pi^-) = \pi^2$, the extension $\tilde{f}(x)$ is a *continuous* function defined for $x \in \mathbb{R}$. ◁

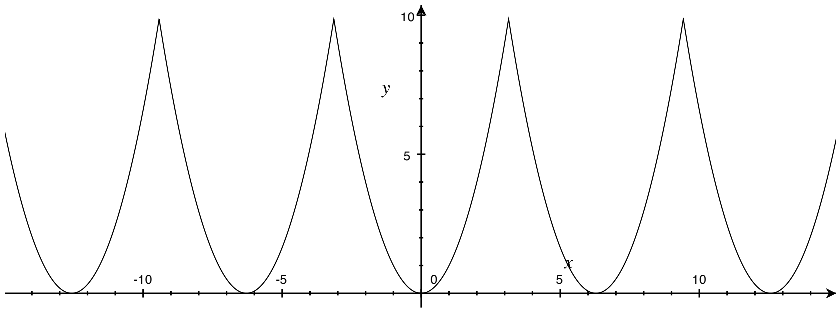



Figure 3.2: The periodic extension $\tilde{f}(x)$ of $f(x) = x^2$.

**Example 3.6.** Sketch the periodic extension of $f(x) = x$, $x \in [-\pi, \pi)$.

*Solution.* The extension $\tilde{f}$ (drawn in Fig. 3.3) is *not continuous*, because

$$f(-\pi) = -\pi \neq \pi = f(\pi^-)\,. \qquad \triangleleft$$

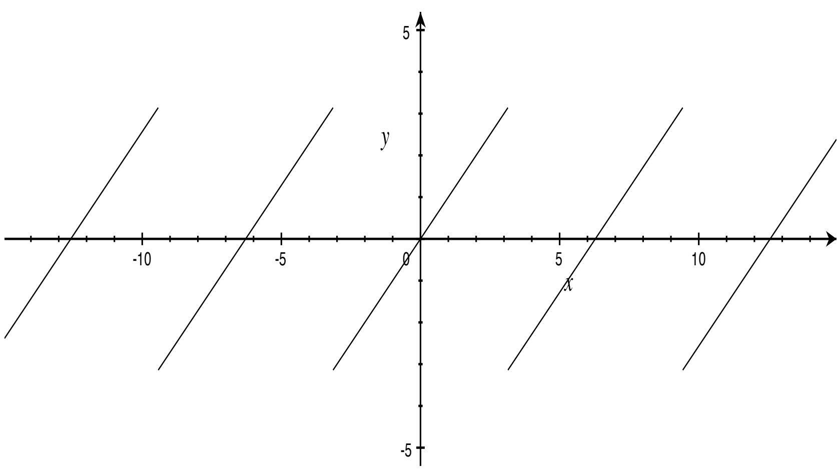



Figure 3.3: The periodic extension of $f(x) = x$.

*Main result*

The best-case scenario would be that the Fourier series of a function $f$ converges to the periodic extension, i.e., $S(x) = \tilde{f}(x)$ for all $x \in \mathbb{R}$. However, this is not the case for all piecewise continuously differentiable functions.

**Example 3.7.** The function $f : [-\pi, \pi] \to \mathbb{R}$,

$$f(x) = \begin{cases} 0 & x \neq 0 \\ 1 & x = 0 \end{cases}$$

is a piecewise continuously differentiable function. What is its Fourier series?

*Solution.* In the integrals that define the Fourier coefficients of $f(x)$, the integrands vanish everywhere except at the single point $x = 0$. However, a non-zero value at only one point contributes nothing to the integrals which means that all coefficients vanish resulting in a Fourier series of the function $f(x)$ equal to zero, i.e., $S(x) = 0$. $\triangleleft$

The example shows that the Fourier series does not "see" the difference between two functions that only differ at one point. More generally, the Fourier series of two piecewise continuously differentiable functions $f$ and $g$ will coincide if there are *at most finitely many points* where the functions differ, i.e., where $f(x) \neq g(x)$.

This observation is important in order to understand how the Fourier series of a function $f$ relates to its periodic extension $\tilde{f}$. If the periodic extension is continuous, then we can reasonably hope that the Fourier series of $f(x)$ will converge to $\tilde{f}$. But what will happen if $\tilde{f}$ has a discontinuity at some point $x_0$? The Fourier series cannot "sense" the value $\tilde{f}(x_0)$ taken by the extension at that point: we can always change the value of $f(x)$—and hence $\tilde{f}(x)$—at the discontinuity without changing its Fourier series.

However, the series will "know" the one-sided limits $\tilde{f}(x_0^+)$ and $\tilde{f}(x_0^-)$, because they are determined by the values of $f(x)$ nearby and the Fourier coefficients pick up this information. There is no reason for the Fourier series to prefer $\tilde{f}(x_0^+)$ over $\tilde{f}(x_0^-)$. The only reasonable thing for the Fourier series is to converge to the average of these two values. That is exactly what happens as the theorem about convergence of Fourier series states.

V17: Midpoint rule and convergence

**Theorem 3.1** (Fourier convergence theorem)**.** If $f : [-\pi, \pi] \to \mathbb{R}$ is a piecewise continuously differentiable function and $\tilde{f} : \mathbb{R} \to \mathbb{R}$ is its $2\pi$-periodic extension, then at $x \in \mathbb{R}$, the Fourier series of $f$ *converges* to

$$\lim_{N\to\infty} S_N(x) = S(x) = \frac{\tilde{f}(x^+) + \tilde{f}(x^-)}{2}\,.$$

In particular, if $\tilde{f}$ is continuous at $x$, then $S(x) = \tilde{f}(x)$ there.

*Proof.* Rather technical and outside the scope of this module. $\square$

The content of the theorem is often referred to as the "midpoint rule." If the one-sided limits of a function $f(x)$ coincide at some point $x$, then the Fourier series reproduces the value exactly. If a piecewise continuously differentiable function $f(x)$ has a discontinuity at a point $x$, i.e., the one-sided limits do not agree, then its graph has a vertical *gap*. The graph of $S(x)$ has an *isolated point* at the midpoint of this gap, which corresponds to a symmetric interpolation between the limits. The "averaging" also happens at the endpoints, i.e., if the periodic extension of a function defined on the interval $[-\pi, \pi)$ develops gaps at odd multiples of $\pi$.

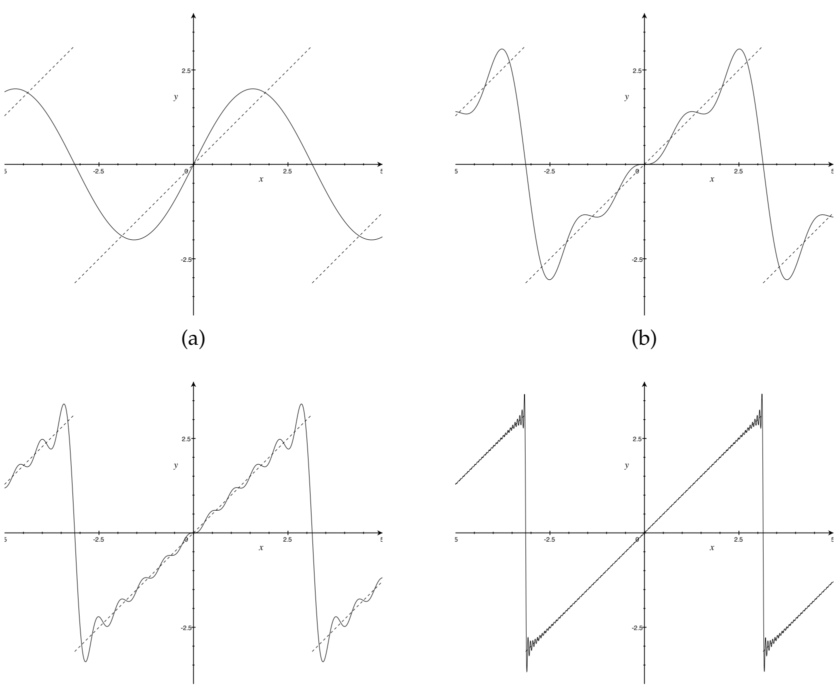
A pictorial representation of the partial sums $S_n(x)$ which approach (and define!) the series $S(x)$ in the limit $N \to \infty$ illustrates that the behaviour of $S(x)$ is entirely natural. In addition, given an explicit Fourier series, one can sometimes check very easily that it conforms with the predictions of Theorem 3.1.

**Example 3.8.** Consider $f : [-\pi, \pi] \to \mathbb{R}$, $f(x) = x$. In Example 3.2 we computed the Fourier series for $f$. If we take the first $N$ terms we can define

V18: Midpoint rule in action

$$S_N(x) := 2 \sum_{n=1}^{N} \frac{(-1)^{n+1}}{n} \sin nx \,,$$

so that $S(x) = \lim_{N\to\infty} S_N(x)$. Four of these are illustrated in Fig. 3.4. Note how the partial sums manage to hug the straight line ever better when more terms are included.



(a) (b) (c) (d)

Figure 3.4: Partial sums $S_1, S_4, S_{10}$ and $S_{100}$ for $f(x) = x$.

In the open interval $x \in (-\pi, \pi)$, the function $f(x) = x$ is *continuous*, and $S(x)$ appears to be approaching $f(x) = x$ for such values of $x$. The Fourier convergence theorem tells us that the series $S(x)$ coincides with $f(x)$ for $x \in (-\pi, \pi)$, and, more generally, $S(x)$ coincides with the periodic extension $\tilde{f}(x)$ for $x \in \mathbb{R}$ except possibly at odd multiples of $\pi$.

When $x = \pi$, the extension $\tilde{f}(x)$ is *discontinuous* jumping from $\tilde{f}(\pi^-) = \pi$ to $\tilde{f}(\pi^+) = \tilde{f}(\pi) = -\pi$. Theorem 3.1 implies that we will find

$$S(\pi) = \frac{\pi + (-\pi)}{2} = 0\,.$$

A visual inspection of the graphs is consistent with this. In this case it is straightforward to confirm this observation analytically since $\sin n\pi = 0$ for any integer $n$,

$$S_n(\pi) = 2\sum_{n=1}^{n} \frac{(-1)^{n+1}}{n} \sin n\pi = 0\,.$$

Consequently, when restricted to the interval $[-\pi, \pi]$, the Fourier series of $f(x) = x$ equals

$$S(x) = \begin{cases} x & -\pi < x < \pi \\ 0 & x = \pm\pi\,, \end{cases}$$

which means that it actually *differs* from the function $f(x)$, but only at the end points $x = \pm\pi$. Finally, extending these considerations to all of the real line, we conclude that $S(x)$ coincides with $\tilde{f}(x)$ except for $x = (2k+1)\pi$ with $k \in \mathbb{Z}$. ■

*Remark* 3.3. The graphs in Fig. 3.4 show that the Fourier series develops "overshoots" near the discontinuities of the extension $\tilde{f}(x)$, known under name of the *Gibbs phenomenon*. In the limit of $N \to \infty$, these apparent deviations do not affect the match between the Fourier series and the periodic extension $\tilde{f}(x)$ but are present for any partial sum.

Let us look at another example where the Fourier series does not agree everywhere with the function from which it stems.

**Example 3.9.** Determine the Fourier series of the function

Sketch it!

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

and compare it with the original function.

*Solution.* This function is piecewise continuously differentiable on the open interval $(-\pi, \pi)$. Let us calculate its Fourier coefficients. The expansion coefficient for the constant term is given by

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,\mathrm{d}x = \frac{1}{\pi}\int_0^{\pi} x\,\mathrm{d}x = \frac{\pi}{2}\,,$$

and for the others we obtain

$$\begin{aligned} a_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx\,\mathrm{d}x = \frac{1}{\pi}\int_0^{\pi} x\cos nx\,\mathrm{d}x \\ &= \frac{1}{\pi n}\left[x\sin nx\right]_0^{\pi} - \frac{1}{\pi n}\int_0^{\pi} \sin nx\,\mathrm{d}x = \frac{1}{\pi n}\left[x\sin nx + \frac{\cos nx}{n}\right]_0^{\pi} \\ &= \frac{\cos n\pi - \cos 0}{\pi n^2} = \frac{(-1)^n - 1}{\pi n^2}\,, \qquad n \in \mathbb{N}\,. \end{aligned}$$

Using integration by parts, the coefficients of sine functions read

$$
\begin{aligned}
b_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\,\mathrm{d}x = \frac{1}{\pi}\int_{0}^{\pi} x\sin nx\,\mathrm{d}x = \\
&= \frac{1}{\pi n}\left[-x\cos nx + \int \cos nx\,\mathrm{d}x\right]_0^{\pi} = \frac{1}{\pi}\left[-\frac{x\cos nx}{n} + \frac{\sin nx}{n^2}\right]_0^{\pi} \\
&= \frac{(-1)^{n+1}}{n}, \qquad n \in \mathbb{N}.
\end{aligned}
$$

Putting these results together, we find the Fourier series of $f(x)$ as

$$
S(x) = \frac{\pi}{4} + \sum_{n=1}^{\infty}\left(\frac{(-1)^n - 1}{\pi n^2}\cos nx + \frac{(-1)^{n+1}}{n}\sin nx\right).
$$

The Fourier series will equal the periodic extension $\tilde{f}(x)$ whenever it is continuous. At odd multiples of $\pi$, however, the extension is discontinuous. According to Theorem 3.1, at $x = \pi$ the series $S(x)$ will take the value

$$
S(\pi) = \frac{f(\pi^+) + f(\pi^-)}{2} = \frac{0+\pi}{2} = \frac{\pi}{2},
$$

and by periodicity, this is the value at any odd multiple of $\pi$, too, so that

$$
S(x) = \begin{cases} \frac{\pi}{2} & x = \pi m,\ m \text{ odd} \\ \tilde{f}(x) & \text{otherwise}. \end{cases}
$$

is the desired Fourier series. $\triangleleft$

*Evaluating infinite series*

**Example 3.10.** Let $f : [-\pi, \pi] \to \mathbb{R}$, $f(x) = x$. What does the Fourier convergence theorem tell us for $x = \frac{\pi}{2}$?

*Proof.* Using the series given in Eq. (3.7) and evaluating $S(x) = f(x)$ at $x = \pi/2$, we find the relation

$$
\frac{\pi}{2} = 2\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}\sin\left(\tfrac{n\pi}{2}\right),
$$

but

$$
\sin\left(\tfrac{n\pi}{2}\right) = \begin{cases} 0 & n \text{ even} \\ (-1)^k & n = 2k+1, \quad k \in \mathbb{Z} \end{cases}
$$

so

$$
\frac{\pi}{2} = 2\sum_{k=0}^{\infty}\frac{(-1)^{(2k+1)+1}}{2k+1}(-1)^k = 2\sum_{k=0}^{\infty}\frac{(-1)^k}{2k+1},
$$

or equivalently,

$$
\frac{\pi}{4} = \sum_{k=0}^{\infty}\frac{(-1)^k}{2k+1} = 1 - \tfrac{1}{3} + \tfrac{1}{5} - \tfrac{1}{7} + \tfrac{1}{9} - \dots.
$$

This is the same formula we would obtain by setting $x = 1$ in the Taylor series for $\tan^{-1} x$. Unfortunately, this is not an efficient way to compute $\pi$ since the sum converges very slowly. $\triangleleft$

**Example 3.11.** What does the Fourier convergence theorem for the function $f(x)$ in Example (3.9) tell us for $x = \pi$?

*Proof.* As explained above, the Fourier series has a discontinuity at $x = \pi$ which means that we have to invoke the mid-point rule which leads to $S(\pi) = \pi/2$. Therefore, equating this value with the Fourier series at $x = \pi$, we find

$$\begin{aligned}\frac{\pi}{2} &= \frac{\pi}{4} + \sum_{n=1}^{\infty}\left(\frac{(-1)^n - 1}{\pi n^2}\cos n\pi + \frac{(-1)^{n+1}}{n}\sin n\pi\right)\\ &= \frac{\pi}{4} + \sum_{n=1}^{\infty}\left(\frac{(-1)^n - 1}{\pi n^2}(-1)^n + 0\right),\end{aligned}$$

or

$$\frac{\pi}{4} = \sum_{n=1}^{\infty}\frac{1-(-1)^n}{\pi n^2} = \sum_{k=0}^{\infty}\frac{2}{\pi(2k+1)^2}.$$

Hence, we have obtained a closed-form expression for the value of the sum over the inverse squares of all odd integers,

$$\frac{\pi^2}{8} = \sum_{k=0}^{\infty}\frac{1}{(2k+1)^2} = 1 + \frac{1}{9} + \frac{1}{25} + \frac{1}{49} + \dots. \qquad \triangleleft$$

## 3.3 *Parseval's theorem*

Recall that, effectively, we compute the Fourier coefficients of a function by using an *inner product* of functions defined in Eq. (3.5). The analogy with the familiar inner product between vectors is perfect, including its geometric meaning, except for the dimension of the vector spaces involved. In this section, we will flesh out one specific aspect of the analogy.

The inner product of a vector $\underline{r} \in \mathbb{R}^3$, say, with itself is the square of its norm (or length), $\underline{r} \cdot \underline{r} = \|\underline{r}\|^2$. Given some orthonormal basis of the space $\mathbb{R}^3$, one can express it as the sum of the squares of the expansion coefficients of $\underline{r}$. If we define the squared norm of a function $f : [-\pi, \pi] \to \mathbb{R}$ by

If $\underline{r} = x\,\underline{i} + y\,\underline{j} + z\,\underline{k}$ then $\|\underline{r}\|^2 = x^2 + y^2 + z^2$, for example.

$$\|f\|^2 = \langle f, f\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f^2(x)\,\mathrm{d}x\,,$$

then structurally equivalent relation exists between the norm $\|f\|^2$. Parseval's identity makes this relation explicit.

**Theorem 3.2** (Parseval). *If $f : [-\pi, \pi] \to \mathbb{R}$ is a piecewise continuously differentiable function with Fourier coefficients $a_0$, $a_n$ and $b_n$, $n \in \mathbb{N}$, then we have the identity*

$$\frac{1}{\pi}\int_{-\pi}^{\pi} f^2(x)\,\mathrm{d}x = \frac{a_0^2}{2} + \sum_{n=1}^{\infty}(a_n^2 + b_n^2)\,.$$

V19: Parseval's identity

*Proof.* Since $f(x)$ is piecewise continuously differentiable, it is equal to its Fourier series $S(x)$ for almost all points in the interval $-\pi \le x \le \pi$. Thus, we can replace $f(x)$ by the series $S(x)$ under the integral, directly establishing the result,

$$\begin{aligned}
\frac{1}{\pi}\int_{-\pi}^{\pi} f^2(x)\,\mathrm{d}x &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\left(\frac{a_0}{2} + \sum_{n=1}^{\infty}\left[a_n \cos nx + b_n \sin nx\right]\right)\mathrm{d}x \\
&= \frac{a_0}{2}\left(\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,\mathrm{d}x\right) + \sum_{n=1}^{\infty} a_n \left(\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx\,\mathrm{d}x\right) \\
&\qquad + \sum_{n=1}^{\infty} b_n \left(\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\,\mathrm{d}x\right) \\
&= \frac{a_0^2}{2} + \sum_{n=1}^{\infty}(a_n^2 + b_n^2)\,.
\end{aligned}$$

□

Theorem 3.2 can be used to evaluate infinite series as the following example shows.

**Example 3.12.** Consider the function $f(x) = x$ as defined in Example 3.10. Apply Parseval's theorem to this case to find an infinite series for $\pi^2$.

*Solution.* We calculated the Fourier-series representation of the line segment in Example 3.2. The only non-zero coefficients are given by $b_n = \frac{2(-1)^{n+1}}{n}$, $n \in \mathbb{N}$, so that Parseval's identity takes the form

$$\frac{1}{\pi}\int_{-\pi}^{\pi} x^2\,\mathrm{d}x = \sum_{n=1}^{\infty}\left(\frac{2(-1)^{n+1}}{n}\right)^2 = \sum_{n=1}^{\infty}\frac{4}{n^2}\,.$$

The integral on the left is easy to calculate,

$$\frac{1}{\pi}\int_{-\pi}^{\pi} x^2\,\mathrm{d}x = \frac{1}{\pi}\left[\frac{x^3}{3}\right]_{-\pi}^{\pi} = \frac{2\pi^2}{3}\,,$$

which leads to the identity

$$\pi^2 = 6\sum_{n=1}^{\infty}\frac{1}{n^2}\,. \tag{3.8}$$

giving us the value of another infinite sum. ◁

*Intermezzo: the Riemann zeta function (non-examinable)*

The sum over the inverse squares of all positive integers can be thought of in an interesting way, namely as the value of a specific and rather important function.

**Definition 3.7** (Riemann zeta function)**.** The *Riemann zeta function* is defined as

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s}, \qquad s > 1.$$

Thus, we see that Eq. (3.8) can be stated in the form

$$\zeta(2) = \frac{\pi^2}{6}.$$

The Riemann zeta function is important in many fields of mathematics. It plays a particularly important role in number theory because Euler was able to show that it can be written as an infinite product involving *all prime numbers*,

$$\zeta(s) = \frac{1}{1-2^{-s}} \frac{1}{1-3^{-s}} \frac{1}{1-5^{-s}} \frac{1}{1-7^{-s}} \cdots = \prod_{p \text{ prime}} \frac{1}{1-p^{-s}}.$$

a truly astonishing relation. It is possible to extend the definition of the zeta function by allowing its argument to be complex numbers, $s \in \mathbb{C}$. In general, it is difficult to obtain analytic expressions for the values of $\zeta(s)$, even when $s$ is real or an arbitrary integer. The result given in (3.8) belongs to a family of cases which can all be treated using Fourier series.

**Example 3.13.** Calculate the values of the Riemann zeta function for positive even integers, i.e., $\zeta(2m)$, $m \in \mathbb{N}$.

*Solution.* A first few values are

$$\zeta(2) = \frac{\pi^2}{6}, \qquad \zeta(4) = \frac{\pi^4}{90}, \qquad \zeta(6) = \frac{\pi^6}{945}, \qquad \zeta(8) = \frac{\pi^8}{9450}, \qquad \text{etc.} \tag{3.9}$$

Recursion relations for the denominators $6, 90, 945, 9450, 93555, \dots$ are known which means that given the first $k$ values $\zeta(2), \zeta(4), \dots, \zeta(2k)$, one can determine the next one from those, that is $\zeta(2k+2)$.

The complete derivation of the values $\zeta(2k)$ uses techniques which are not relevant to us. At some point, however, the argument uses the Fourier series of the continuous function

V20: When Riemann meets Fourier

$$f_\lambda(x) = \mathrm{e}^{\lambda x}, \qquad x \in [-\pi, \pi],$$

depending on the real parameter $\lambda$. The periodic extension $\tilde{f}_\lambda(x)$ has discontinuities at all odd integers. As in Example (3.11), the crucial

Sketch $f_\lambda(x)$ and its periodic extension!

information comes from the Fourier series evaluated at a discontinuity of the extension which is only possible thanks to Theorem 3.1.

Using the definition of the Fourier coefficient of the cosine terms, we find

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} \mathrm{e}^{\lambda x}\cos nx\,\mathrm{d}x = (-1)^n \frac{2}{\pi}\frac{\lambda \sinh \pi\lambda}{\lambda^2+n^2}\,, \qquad n \in \mathbb{N}_0\,.$$

Using the fact that $\cos nx = \operatorname{Re} \mathrm{e}^{inx}$, the intermediate steps are not too complicated,

$$\begin{aligned}\frac{1}{\pi}\int_{-\pi}^{\pi} \mathrm{e}^{\lambda x}\cos nx\,\mathrm{d}x &= \frac{1}{\pi}\operatorname{Re}\int_{-\pi}^{\pi} \mathrm{e}^{(\lambda+\mathrm{i}n)x}\,\mathrm{d}x = \frac{1}{\pi}\operatorname{Re}\frac{\left[\mathrm{e}^{(\lambda+\mathrm{i}n)x}\right]_{-\pi}^{\pi}}{\lambda+\mathrm{i}n} \\ &= (-1)^n\frac{2}{\pi}\sinh\pi\lambda\operatorname{Re}\frac{1}{\lambda+\mathrm{i}n} = (-1)^n\frac{2}{\pi}\frac{\lambda\sinh\pi\lambda}{\lambda^2+n^2}\,.\end{aligned}$$

The other coefficients follow from an identical calculation if we use $\sin nx = \operatorname{Im} \mathrm{e}^{inx}$,

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} \mathrm{e}^{\lambda x}\sin nx\,\mathrm{d}x = (-1)^{n+1} \frac{2}{\pi}\frac{n \sinh \pi\lambda}{\lambda^2+n^2}\,, \qquad n \in \mathbb{N}_0\,,$$

so that altogether the Fourier series associated with the function $f_\lambda(x)$ becomes

$$S_\lambda(x) = \frac{\sinh\pi\lambda}{\pi\lambda} + \frac{2}{\pi}\sinh\pi\lambda\sum_{n=1}^{\infty}(-1)^n\left(\frac{\lambda\cos nx}{(\lambda^2+n^2)} - \frac{n\sin nx}{(\lambda^2+n^2)}\right).$$

In the next step, one considers the value of the series at the point $x=\pi$ where all sine terms vanish, leading to

$$S_\lambda(\pi) = \frac{\sinh\pi\lambda}{\pi\lambda}\left(1+\sum_{n=1}^{\infty}\frac{2\lambda^2}{n^2+\lambda^2}\right).$$

Now one invokes Theorem 3.1 providing an alternative expression for the series at $x=\pi$, namely

$$S_\lambda(\pi) = \frac{f_\lambda(\pi)+f_\lambda(-\pi)}{2} = \frac{\mathrm{e}^{\pi\lambda}+\mathrm{e}^{-\pi\lambda}}{2} = \cosh\lambda\pi\,,$$

which implies the identity

$$\lambda\pi\coth\lambda\pi = 1+\sum_{n=1}^{\infty}\frac{2\lambda^2}{(n^2+\lambda^2)}.$$

Finally, a rather technical use of Taylor series allows one to extract the relations 3.9 from this identity but we will suppress the details. The essential point is to understand that the midpoint rule provides crucial information to solve interesting problems, in particular, in the presence of discontinuities. $\triangleleft$

It is possible to extend the definition of $\zeta(s)$ by allowing $s$ to be complex-valued. One of the great unsolved problems of mathematics is the *Riemann hypothesis*. If you can prove the conjecture, you will win a Fields medal which is the mathematics equivalent of a Nobel prize.

**Conjecture** (Riemann)**.** *The (nontrivial) zeros of the Riemann zeta function $\zeta(s)$,$s \in \mathbb{C}$, all lie on the* critical line *in the complex plane, i.e., $\zeta(s) = 0$ implies* $\operatorname{Re} s = \frac{1}{2}$.

The trivial zeros of the Riemann zeta function are located at the points $s = -2, -4, -6, \ldots$, on the real axis.

## 3.4 *Half-range series*

The general Fourier series (3.1) is a bit messy, since it includes both sines and cosines. It is simpler to work with only sines or only cosines. these special Fourier series are important in the applications to partial differential equations. Since the cosine and sine functions are even and odd, respectively, they can only represent even or odd functions. Any even or odd function on the interval $[-\pi, \pi]$, however, is essentially equivalent to a function on the interval $[0, \pi]$. This observations underpins the idea of *half-range* Fourier series.

Suppose that we have a function $f : [0, \pi] \to \mathbb{R}$ and wish to represent it in terms of a Fourier series. We can extend it in two ways to a function defined on the interval $[-\pi, \pi]$. We could introduce an *even* function by

$$f_{\text{even}}(x) := \begin{cases} f(-x) & -\pi \leq x < 0 \\ f(x) & 0 \leq x \leq \pi \,, \end{cases}$$

simply by reflecting $f(x)$ about the vertical axis. Fig. 3.5 shows the effect on the functions $g(x) = 1$ and $h(x) = x$, both defined on the interval $0 \leq x \leq \pi$.

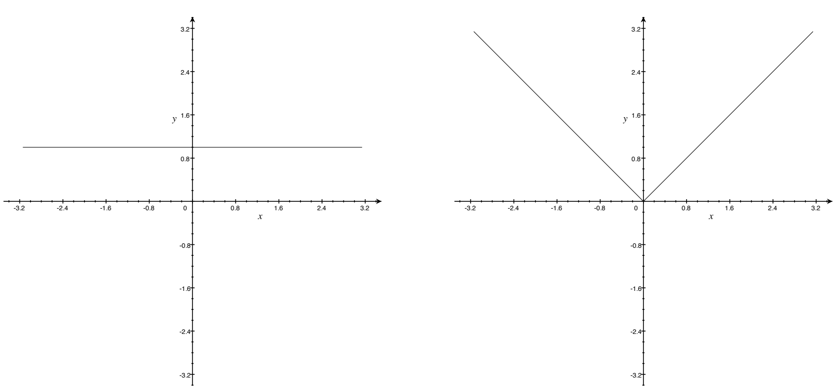


Figure 3.5: Even extensions $g_{\text{even}}$ and $h_{\text{even}}$ of $g(x) = 1$ and $h(x) = x$, $x \in [0, \pi]$.

With this definition, the *half-range cosine series* of $f$ on $[0, \pi]$ is given by the Fourier series of $f_{\text{even}}$ on the interval $[-\pi, \pi]$. Being an even function, the sine terms vanish, and the Fourier coefficients of the cosine terms (including the constant one) are readily determined to be

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f_{\text{even}}(x) \cos nx \, \mathrm{d}x = \frac{2}{\pi} \int_0^{\pi} f(x) \cos nx \, \mathrm{d}x \,, \qquad n \in \mathbb{N}_0.$$

Note that the integration is over the interval $[0, \pi]$ only due to the symmetry of the function $f_{\text{even}}$.

Given initially the function $f : [0, \pi] \to \mathbb{R}$, we could, alternatively, generate an *odd* function on the interval $[-\pi, \pi]$ by defining

$$f_{\text{odd}}(x) := \begin{cases} -f(-x) & -\pi \leq x < 0 \\ 0 & x = 0 \\ f(x) & 0 < x \leq \pi \,; \end{cases}$$

at the origin, the value of $f_{\text{odd}}$ may differ from the value of $f$. Starting with the same two functions $g(x)$ and $h(x)$ as before, Fig. 3.6 illustrates the result of anti-symmetrizing them.

V21: Odd extensions

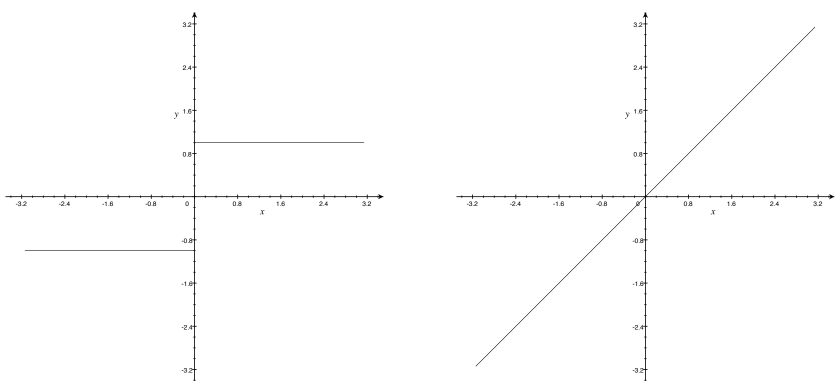


Figure 3.6: Odd extensions $g_{\text{odd}}$ and $h_{\text{odd}}$ of $g(x) = 1$ and $h(x) = x$, $x \in [0, \pi]$.

Now, the *half-range sine series* of $f$ on $[0, \pi]$ equals the Fourier series of $f_{\text{odd}}$ on $[-\pi, \pi]$. The constant and the cosine terms vanish by construction, and the coefficients of the sine terms can be obtained from an integration over the interval $[0, \pi]$,

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f_{\text{odd}}(x) \sin nx \, \mathrm{d}x = \frac{2}{\pi} \int_0^{\pi} f(x) \sin nx \, \mathrm{d}x \,.$$

Let us look at an example in which we calculate the half-range Fourier series for one and the same function.

**Example 3.14.** Find the cosine and sine series for $f(x) = 1$ on the interval $[0, \pi]$.

*Solution.* For the cosine series, we calculate

$$a_n = \frac{2}{\pi} \int_0^{\pi} 1 \cdot \cos nx \, \mathrm{d}x = \begin{cases} \frac{2}{\pi} \left[\frac{\sin nx}{n}\right]_0^{\pi} = 0 & \text{if} \quad n \neq 0 \\ = \frac{2}{\pi} \int_0^{\pi} 1 \, \mathrm{d}x = 2 & \text{if} \quad n = 0 \,, \end{cases}$$

**Algorithm 8**
Cosine and Sine Series for a function $f : [0,\pi] \to \mathbb{R}$

1. The half-range cosine and sine series of $f$ are given by
$S_c(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx$ and $S_s(x) = \sum_{n=1}^{\infty} b_n \sin nx$,
respective.

2. To calculate the coefficients, use the following.
Cosine series: $a_n = \frac{2}{\pi}\int_0^{\pi} f(x)\cos nx\,\mathrm{d}x$ and $b_n = 0$.
Sine series: $b_n := \frac{2}{\pi}\int_0^{\pi} f(x)\sin nx\,\mathrm{d}x$, and $a_n = 0$.

leading to a particularly simple cosine series,

$$S_{\mathrm{c}}(x) = 1\,.$$

For the sine series, we obtain

$$\begin{aligned} b_n &= \frac{2}{\pi}\int_0^{\pi} 1 \cdot \sin nx\,\mathrm{d}x = \frac{2}{\pi}\left[-\frac{\cos nx}{n}\right]_0^{\pi} \\ &= \frac{2}{\pi n}\left(-\cos n\pi + \cos 0\right) = \frac{2}{\pi n}\left[1-(-1)^n\right] = \begin{cases} \frac{4}{\pi n} & n \text{ odd} \\ 0 & n \text{ even.} \end{cases} \end{aligned}$$

giving the sine series,

$$S_{\mathrm{s}}(x) = \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{1}{2k+1}\sin[(2k+1)x] \qquad \triangleleft$$

Since $S_{\mathrm{c}}(x) = S_{\mathrm{s}}(x) = f(x)$ for $x \in (0,\pi)$, we have found a surprisingly convoluted way to rewrite the constant 1. Another surprising feature: from the Fourier convergence theorem, $S_{\mathrm{s}}(x) = 1$ for any $x \in (0,\pi)$, while dropping to 0 at $x = \pi$ and then taking the value $-1$ for $x \in (\pi, 2\pi)$ and so on. Looking at the form of the series, this is far from obvious.

## 3.5 *Complex exponential series*

Since the cosine and sine functions are the real and imaginary parts of the complex exponential function, one can write the Fourier series (3.1) in a slightly different but completely equivalent way,

$$S(x) = \sum_{n=-\infty}^{\infty} c_n \mathrm{e}^{\mathrm{i}nx}, \tag{3.10}$$

with complex-valued coefficients $c_n \in \mathbb{C}$, and a doubly infinite sum, i.e., $n \in \mathbb{Z}$.

The complex form of the Fourier series simplifies the computation of the expansion coefficients since it is no longer necessary to treat the sine and cosine terms separately. Analogously to the sine and cosine form, we can extend the inner product for functions we introduced earlier to complex-valued functions $f, g : [-\pi, \pi] \to \mathbb{C}$ as

$$\langle f, g\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\bar{g}(x)\,\mathrm{d}x. \tag{3.11}$$

The set of functions used in the complex Fourier series, $f_n := \mathrm{e}^{\mathrm{i}n\pi}$, $n \in \mathbb{Z}$, are orthogonal under this inner product. In fact,

$$\langle f_n, f_m\rangle = \begin{cases} 0 & m \neq n \\ 2 & m = n \end{cases},$$

as can be verified from the fact that, for any integer $k \in \mathbb{Z}$, we have

$$\frac{1}{\pi}\int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}kx}\,\mathrm{d}x = \begin{cases} \frac{1}{\pi}\left[\frac{1}{\mathrm{i}k}\mathrm{e}^{\mathrm{i}kx}\right]_{-\pi}^{\pi} = 0 & \text{if} \quad k \neq 0 \\ \frac{1}{\pi}\int_{-\pi}^{\pi}\mathrm{d}x = 2 & \text{if} \quad k = 0 \end{cases}.$$

Assuming that the Fourier series (3.10) indeed represents a function $f(x)$, i.e., $f(x) = S(x)$, we can hence compute the coefficients:

$$\begin{aligned}\frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\,\mathrm{e}^{-\mathrm{i}mx}\,\mathrm{d}x &= \frac{1}{2}\langle f, f_m\rangle = \frac{1}{2}\left\langle \sum_{n=-\infty}^{\infty} c_n f_n, f_m \right\rangle \\ &= \frac{1}{2}\sum_{n=-\infty}^{\infty} c_n \langle f_n, f_m\rangle = c_m\,.\end{aligned}$$

In other words, the coefficient $c_n$ is found from multiplying $f(x)$ with $\mathrm{e}^{-\mathrm{i}nx}$ and integrating over $x$ between $-\pi$ and $\pi$:

Note the negative sign in the exponent!

$$c_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\,\mathrm{e}^{-\mathrm{i}nx}\,\mathrm{d}x. \tag{3.12}$$

Taking the complex conjugate of this expression leads to

$$\bar{c}_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\,\mathrm{e}^{\mathrm{i}nx}\,\mathrm{d}x = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\,\mathrm{e}^{-\mathrm{i}(-n)x}\,\mathrm{d}x = c_{-n}.$$

This property reflects the fact that we use a *complex* Fourier series to represent a *real* function $f(x)$. The real and imaginary parts of $\mathrm{e}^{-\mathrm{i}nx}$ being $\cos nx$ and $-\sin nx$, we can establish the relation between the real and the complex Fourier coefficients,

V22: Real vs complex Fourier series

$$a_n = 2\,\mathrm{Re}\,c_n, \quad b_n = -2\,\mathrm{Im}\,c_n\,.$$

The mathematical equivalence of the real and the complex Fourier series ensures that Theorem 3.1 continues to hold.

It is also possible to consider Fourier series for complex valued functions. For these, it need not be the case that $\bar{c}_n = c_{-n}$, and there is also a different expression for $a_n$ and $b_n$ in terms of $c_n$.

You might like to work this out.

**Example 3.15.** Find the complex exponential series for $f : [-\pi, \pi] \to \mathbb{R}$, $f(x) = x$.

*Solution.* Starting with $n = 0$, we immediately obtain

$$c_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} x \, dx = \left[\tfrac{1}{2}x^2\right]_{-\pi}^{\pi} = 0 \,.$$

When $n \neq 0$, we integrate by parts to find

$$\begin{aligned} c_n &= \frac{1}{2\pi} \int_{-\pi}^{\pi} x \, e^{-inx} \, dx = \frac{i}{2\pi n} \left\{ \left[ x \, e^{-inx} \right]_{-\pi}^{\pi} - \int_{-\pi}^{\pi} e^{-inx} \, dx \right\} \\ &= \frac{i}{2\pi n} \left[ \pi(-1)^n - (-\pi)(-1)^n \right] = \frac{(-1)^n}{n} i \,, \end{aligned}$$

leading to the complex Fourier series of $f(x) = x$,

$$S(x) = i \sum_{n \in \mathbb{Z} \setminus \{0\}} \frac{(-1)^n}{n} e^{inx} \,.$$

Note that $n = 0$ is explicitly excluded from the sum. ◁

The complex series is equivalent to the sine series (3.7) which we calculated in Example 3.2 since we can rearrange its terms,

$$S(x) = i \sum_{n=1}^{\infty} \frac{(-1)^n}{n} \left( e^{inx} - e^{-inx} \right) = \sum_{n=1}^{\infty} \frac{(-1)^n i}{n} 2i \sin nx = \sum_{n=1}^{\infty} \frac{2(-1)^{n+1}}{n} \sin nx \,.$$

**Example 3.16.** Find the complex exponential series of $f : [-\pi, \pi] \to \mathbb{R}$, $f(x) = x^2$.

*Solution.* The Fourier coefficients are defined as

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} x^2 \, e^{-inx} \, dx \,.$$

The case $n = 0$ gives

$$c_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} x^2 \, dx = \frac{1}{2\pi} \left[ \tfrac{1}{3} x^3 \right]_{-\pi}^{\pi} = \frac{\pi^2}{3}.$$

For $n \neq 0$, let us first compute the integral

$$\begin{aligned} \int x^2 \, e^{-inx} \, dx &= \frac{i}{n} x^2 e^{-inx} - \frac{2i}{n} \int x \, e^{-inx} \, dx \\ &= \frac{i}{n} x^2 e^{-inx} + \frac{2}{n^2} x \, e^{-inx} - \frac{2}{n^2} \int e^{-inx} \, dx \\ &= \left( \frac{i}{n} x^2 + \frac{2}{n^2} x - \frac{2i}{n^3} \right) e^{-inx} + \text{Const.} \end{aligned}$$

Using the limits $\pm\pi$ in the integration, we obtain

$$c_n = \frac{(-1)^n}{2\pi} \left[ \frac{i}{n} x^2 + \frac{2}{n^2} x - \frac{2i}{n^3} \right]_{-\pi}^{\pi} = 2 \frac{(-1)^n}{n^2}$$

so that, finally,

$$S(x) = \frac{\pi^2}{3} + 2 \sum_{n \in \mathbb{Z} \setminus \{0\}} \frac{(-1)^n}{n^2} \mathrm{e}^{\mathrm{i}nx} , \qquad x \in [-\pi, \pi] ,$$

which reproduces the values $f(x) = x^2$ at all points of the *closed* interval. $\triangleleft$

## 3.6 Fourier series on other intervals

For completeness, we briefly mention that one can modify Fourier series for functions on other intervals, rather than $[-\pi, \pi]$. For instance, if we have a function defined on $[-L/2, L/2]$ (so that its periodic extension has period $L$), we can also come up with a Fourier series. In essence we are rescaling the function, replacing $x$ by $2\pi x/L$. We hence expect a Fourier series of the form

$$S(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left( a_n \cos\left(\frac{2n\pi x}{L}\right) + b_n \sin\left(\frac{2n\pi x}{L}\right) \right) .$$

The formulae for the coefficients come out with an analogous change:

$$\begin{aligned} a_n &= \frac{2}{L} \int_{-L/2}^{L/2} f(x) \cos\left(\frac{2n\pi x}{L}\right) \mathrm{d}x \\ b_n &= \frac{2}{L} \int_{-L/2}^{L/2} f(x) \sin\left(\frac{2n\pi x}{L}\right) \mathrm{d}x . \end{aligned}$$

Note that these match our previous formulae if $L = 2\pi$. You might like to confirm the orthogonality properties of the new set of functions in the series with the modified inner product $\frac{2}{L} \int_{-L/2}^{L/2} f(x) g(x) \, \mathrm{d}x$.

Similarly, for half-range series defined on intervals of arbitrary length. For the interval $[0, L/2]$, the sine series will take the form

$$S(x) = \sum_{n=1}^{\infty} b_n \sin\left(\frac{2n\pi x}{L}\right) ,$$

with a necessary adjustment of the formula for the sine series coefficients. The change comes from the need to have the correct inner products among the sine functions so that they provide a set of orthonormal sets of functions on $[0, L/2]$, i.e.,

$$\frac{4}{L} \int_0^{L/2} \sin\left(\frac{2n\pi x}{L}\right) \sin\left(\frac{2m\pi x}{L}\right) \mathrm{d}x = \begin{cases} 1 & \text{if} \quad n = m \\ 0 & \text{if} \quad n \neq m . \end{cases}$$

The formula for the coefficients $b_n$ now reads

$$b_n = \frac{4}{L} \int_0^{L/2} f(x) \sin\left(\frac{2n\pi x}{L}\right) \mathrm{d}x.$$

An analogous relation can be made for the cosine series.

# 4
# *Functions of More than One Variable*

### *Main idea*

Imagine a landscape with lovely rolling hills. Take a big knife and cut through them along a vertical line. Now look at the cross section along the cut: you can think of the *height* as defining a function $f(x)$ where the variable $x$ tells you the position along the cut. Next, imagine cutting along another vertical line perpendicular to the first. This will give another function which we can call $h(0, y)$, assuming that the second cut hits the first one at $x = 0$. The function $h(0, y)$ describes the height of the hills above the ground when you move along the second cut, i.e., by varying the value of the variable $y$.

The height of the hills varies not only along these two straight lines but over the entire two-dimensional region, the points of which can be identified by pairs of numbers $(x, y)$, Their values can be chosen independently from each other. We are able to describe the height of the hills by a function $h(x, y)$ depending on *two* independent variables. In other words, the *height* function $h(x, y)$ fully records all ups and downs of the rolling hills. When specifying a location in terms of the variables $x$ and $y$, the function $h(x, y)$ outputs the height at particular point, just as the height function $f(x)$ of the first cut did along a line. We know about $f$ if we know $h$ since $h(x, 0) \equiv f(x)$.

In this chapter we will introduce number of useful tools to describe the properties of functions depending on two (and occasionally even more) variables.

### *Geometric pictures (§14.1)*

There are two main ways to visualize a function of two variables. Recalling that the graph of a function of one variable is a *curve* in a plane suggests associating a function $f(x, y)$ with a *surface* embedded in a three-dimensional space.

**Definition 4.1** (Graph of a function)**.** If $f$ is a function of two variables, then its *graph* is the surface

$$\{(x, y, z) \in \mathbb{R}^3 \mid z = f(x, y)\}.$$

**Example 4.1.** If $f(x, y) = Ax + By + C$, $(x, y) \in \mathbb{R}^2$, with constants $A$, $B$ and $C$, then the graph of $f$ is a plane.

**Example 4.2.** If $f(x, y) = \sqrt{1 - x^2 - y^2}$ defined on $U = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$, then the graph of $f$ is a hemisphere.

Alternatively, on can represent the surfaces associated with a function $f(x, y)$ by identifying lines of identical "height" and projecting onto a plane.

V23: Graphs and level sets

**Definition 4.2** (Level sets and curves)**.** If $f$ is a function of two variables, then a *level set* for $f$ is the set

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

for some constant $c \in \mathbb{R}$. Provided $f$ is a nice enough function, its level set will be a smooth curve and thus called a *level curve*.

**Example 4.3.** If $x$ is longitude and $y$ is latitude, and $f(x, y)$ is the height of the point, then the contour lines on a map are the level curves of $f$.

**Example 4.4.** Let $f(x, y) = x^2 + y^2$. The level set $x^2 + y^2 = 0$ consists of the single point $(0, 0)$. For $c > 0$, the level curve $x^2 + y^2 = c$ is a circle of radius $\sqrt{c}$.

## 4.1 *Partial derivatives (§14.3)*

Just as in the one-dimensional case, we would like determine where a function has steep slopes and where it varies only slowly—or not at all, which would correspond to extremal points. Our brains actually know how to do all this, at least qualitatively as you can see from the following example.

After completing this module, you will be taking your well-deserved skiing holiday in Switzerland (other destinations are available). You find yourself at the top of a mountain, ready to go. Of course you want to *maximise* the fun you have which means to ski as fast as possible. What to do? Well, you simply look around to check for the direction with the steepest slope. By identifying that specific direction, what you effectively have done is to compare the values of the *partial derivatives* along the different directions you could travel. Let us formalize this intuitive idea.

**Definition 4.3** (Partial derivatives)**.** The rate of change of $f(x, y)$ as $x$ changes with $y$ fixed, is called the *partial derivative* of $f$ with respect to $x$,

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

and the partial derivative of $f$ with respect to $y$ is given by

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

Thus, to compute $\frac{\partial f}{\partial x}$ we simply apply the usual rules of differentiation to $x$ regarding $y$ as a constant. The symbol "$\partial$" is a curly "d" pronounced "partial".

*Remark* 4.1. There are several different notations for partial derivatives:

$$\frac{\partial f(x, y)}{\partial x} = \partial_x f(x, y) = f_x(x, y) = f_1(x, y). \tag{4.1}$$

The expression $f_x$ is a convenient shorthand for a derivative with respect to the variable $x$ while writing $f_1(x, y)$ prompts us to take a derivative with respect to the *first argument* of the function $f(x, y)$. Thus, upon listing the arguments of $f$ in the opposite order, we would find $f_x(y, x) = f_2(y, x)$, in contrast to Eq. (4.1).

**Example 4.5.** Find the partial derivatives of the function $f(x, y) = x^2 + xy + y$.

*Solution.*

$$\frac{\partial f}{\partial x} = 2x + y, \qquad \frac{\partial f}{\partial y} = x + 1. \qquad \triangleleft$$

*Second order partial derivatives*

The derivatives $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ are functions of $x$ and $y$, so we can consider partial derivatives of these. Applying the same derivative to each of the expressions again we find one pair of *second-order partial derivatives* of $f$:

V24: Partial derivatives

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial x^2} = f_{xx} = f_{11}(x, y),$$
$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial y^2} = f_{yy} = f_{22}(x, y).$$

There are two more second-order partial derivatives we can form,

$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial y \partial x} = f_{xy} = (f_x)_y = f_{12}(x, y)\,,$$
$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial x \partial y} = f_{yx} = (f_y)_x = f_{21}(x, y)\,,$$

known as *mixed* partial derivatives. Hence, one needs to be careful about the order of the subscripts: for example, the derivative $f_{yx}$ is obtained by differentiating the function $f$ with respect to $y$ first and then with respect to $x$ which, in principle, differs from calculating $f_{xy}$.

**Example 4.6.** Calculate the second derivatives for the function $f$ in the previous example.

*Solution.*

$$\frac{\partial^2 f}{\partial x^2} = 2\,, \qquad \frac{\partial^2 f}{\partial y \partial x} = 1\,, \qquad \frac{\partial^2 f}{\partial x \partial y} = 1\,, \qquad \frac{\partial^2 f}{\partial y^2} = 0\,. \qquad \triangleleft$$

Thus, for this specific example, we notice that the mixed partial derivatives coincide,

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}\,.$$

According to the following theorem, this property is, however, not a coincidence but holds for "sufficiently nice" functions.

**Theorem 4.1** (Clairaut). *If a function $f(x, y)$ and its partial derivatives $f_x$, $f_y$, $f_{xy}$, $f_{yx}$ are defined throughout an open region containing the point $(a, b)$* and they are all continuous at $(a, b)$, *then*

$$f_{xy}(a, b) = f_{yx}(a, b)\,.$$

Clairaut's theorem is a useful property since changing the order of differentiation can sometimes make the calculation simpler.

**Example 4.7.** Find $\frac{\partial^2 f}{\partial x \partial y}$ for

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

*Solution.* We have

$$\frac{\partial f}{\partial y} = x + \frac{e^y}{y^2+1} - \frac{2ye^y}{(y^2+1)^2} \qquad \text{so} \qquad \frac{\partial^2 f}{\partial x \partial y} = 1$$

whereas if we do the differentiation the other way around, first $x$ then $y$, we have

$$\frac{\partial f}{\partial x} = y, \qquad \text{so} \qquad \frac{\partial^2 f}{\partial y \partial x} = 1 \qquad \triangleleft$$

For a function of *three* variables $f(x,y,z)$, we define $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, and $\frac{\partial f}{\partial z}$ in a similar manner. For example, $\frac{\partial f}{\partial x}$, the partial derivative of $f$ with respect to $x$, is the derivative of $f$ with respect to $x$ with $y$ and $z$ fixed. The alternative notations extend naturally to more than two variables,

$$\frac{\partial f}{\partial z} = f_z = f_3(x,y,z)\,, \qquad \frac{\partial^2 f}{\partial x \partial z} = f_{zx} = f_{31}(x,y,z)\,, \qquad \text{etc.}$$

Clairaut's theorem also holds for functions of three (or even more) variables. It becomes ever more useful if we need to know calculate all partial derivatives of some higher order since it reduces the number of computations we need to carry out.

### *Generalizing the chain rule (§14.4)*

The chain rule from single-variable calculus states that, given two differentiable functions $f(t)$ and $g(x)$, the derivative of their composition $F(t) = f[x(t)]$ is given by $F'(t) = x'(t)f'[x(t)]$. We will need to generalize the rule to functions of more variables. When doing so, *vectors* will appear naturally. Most of the time they will be written in component form $(x,y)$ but sometimes the expressions are more readable when using basis vectors such as $\underline{i} = (1,0)$, etc.

To establish a chain rule for functions of more than one variable, we will need them to be differentiable in an appropriate sense.

**Definition 4.4** (Continuous differentiability)**.** A function $f(x,y)$ is *continuously differentiable* if $f$ and its partial derivatives $f_x$ and $f_y$ are all continuous.

This property is sufficient to introduce the derivative of a function depending on two variables, each of which depends on the independent variable $t$, say.

We will not cover the proof in this course, but you might like to read about it in Thomas' book.

**Theorem 4.2** (Chain rule)**.** *Let $f(x,y)$ be a continuously differentiable function, and $\underline{r}(t) = (x(t),y(t))$ be a pair of differentiable functions. The derivative of the composite function*

$$F(t) = f[\underline{r}(t)] = f[x(t),y(t)]$$

*with respect to the variable t is given by*

$$F'(t) = x'(t)\, f_x[x(t),y(t)] + y'(t)\, f_y[x(t),y(t)]\,.$$

There are other suggestive ways to write down the chain rule, for example

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

often, no distinction between $F$ and $f$ is being made leading to

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

Let us work through an example for the chain rule the result of which we can check directly by an alternative calculation.

**Example 4.8.** Let $f(x,y) = \mathrm{e}^{xy}$, $\underline{r}(t) = (x(t), y(t)) = (t^2, t)$, and

$$F(t) = f[\underline{r}(t)] = f[x(t), y(t)] = \mathrm{e}^{t^2\cdot t} = \mathrm{e}^{t^3}\,.$$

Verify that the chain rule is satisfied here.

*Solution.* The partial derivatives are given by

$$f_x(x,y) = y\mathrm{e}^{xy} \qquad \text{and} \qquad f_y(x,y) = x\mathrm{e}^{xy}\,,$$

respectively, and upon substituting $x(t)$ and $y(t)$ they evaluate to

$$f_x(x(t), y(t)) = t\mathrm{e}^{t^3} \qquad \text{and} \qquad f_y(x(t), y(t)) = t^2\mathrm{e}^{t^3}\,.$$

Using

$$x'(t) = 2t \qquad \text{and} \qquad y'(t) = 1\,,$$

the chain rule states that

$$\begin{aligned} F'(t) &= x'(t)\, f_x[x(t), y(t)] + y'(t)\, f_y[x(t), y(t)] \\ &= 2t \cdot t\mathrm{e}^{t^3} + 1 \cdot t^2\mathrm{e}^{t^3} = 3t^2\mathrm{e}^{t^3}\,. \end{aligned}$$

Directly differentiating $F(t) = \mathrm{e}^{t^3}$ gives $F'(t) = 3t^2\mathrm{e}^{t^3}$ when using the chain rule for a function of a single variable, thus confirming the result found from the multi-variable chain rule. $\triangleleft$

*Remark* 4.2. From now on, all functions will be assumed to be continuously differentiable unless stated otherwise, so that the chain rule holds.

## 4.2 *The gradient (§14.5)*

The form of the derivative of a function of more than one variable shown in Theorem 4.2 suggests considering the partial derivatives as components of a vector.

**Definition 4.5** (Gradient)**.** The *gradient* of a function of two variables $f(x,y)$ is given by

$$\nabla f := \frac{\partial f}{\partial x}\,\underline{i} + \frac{\partial f}{\partial y}\,\underline{j}\,,$$

and the gradient of a three-variable function $F(x,y,z)$ reads

$$\nabla F := \frac{\partial F}{\partial x}\,\underline{i} + \frac{\partial F}{\partial y}\,\underline{j} + \frac{\partial F}{\partial z}\,\underline{k}\,.$$

The symbol $\nabla$ is called the "nabla operator" (or just "nabla"), and we will also call it grad, so $\nabla f$ is "grad $f$".

V25: Chain rule and gradient

**Example 4.9.** Find the gradient of $f(x,y) = (x^2+y)^2$.

*Solution.* Since $\frac{\partial f}{\partial x} = 4x(x^2+y)$ and $\frac{\partial f}{\partial y} = 2(x^2+y)$, we have

$$\nabla f(x,y) = 4x(x^2+y)\,\underline{i} + 2(x^2+y)\,\underline{j}\,. \qquad \triangleleft$$

Now we are in a position to write the chain rule for functions depending on $n > 1$ variables an elegant way using vector notation and the gradient. First, Theorem 4.2 extends naturally: if we compose a function $f : \mathbb{R}^n \to \mathbb{R}$ with a vector-valued function $\underline{r} : \mathbb{R} \to \mathbb{R}^n$ to $F(t) := f[\underline{r}(t)]$, then we have

$$F'(t) = \underline{r}'(t) \cdot \nabla f[\underline{r}(t)]\,. \tag{4.2}$$

The case of three variables is particularly important for applications since it involves functions which vary as from one point in space to the other. In this case we have $\underline{r}(t) = (x(t), y(t), z(t))$ and thus $F(t) = f[\underline{r}(t)] = f[x(t), y(t), z(t)]$, so that with

$$\underline{r}'(t) = (x'(t), y'(t), z'(t)) = \frac{\mathrm{d}x}{\mathrm{d}t}\,\underline{i} + \frac{\mathrm{d}y}{\mathrm{d}t}\,\underline{j} + \frac{\mathrm{d}z}{\mathrm{d}t}\,\underline{k}\,,$$

we obtain

$$\frac{\mathrm{d}F}{\mathrm{d}t} = \underline{r}'(t) \cdot \nabla f[\underline{r}(t)] = \frac{\mathrm{d}x}{\mathrm{d}t}\frac{\partial f}{\partial x} + \frac{\mathrm{d}y}{\mathrm{d}t}\frac{\partial f}{\partial y} + \frac{\mathrm{d}z}{\mathrm{d}t}\frac{\partial f}{\partial z}\,.$$

### *Directional derivatives*

The partial derivatives, $f_x$ and $f_y$, of a given function $f(x,y)$ of two variables are rates of change of $f$ in the $x$- and $y$-directions. We can find the rate of change in an arbitrary direction characterized by some unit vector $\underline{u}$ with $\|\underline{u}\| = 1$. The concept of a directional derivative is valid for any number of variables.

**Definition 4.6** (Directional derivative)**.** If $f$ is a function, $\underline{r}_0$ is a point, and $\underline{u}$ is a unit vector, then the *directional derivative* of $f$ at $\underline{r}_0$ in the direction of $\underline{u}$ is

$$D_{\underline{u}} f(\underline{r}_0) := \lim_{h\to 0} \frac{f(\underline{r}_0 + h\,\underline{u}) - f(\underline{r}_0)}{h} \tag{4.3}$$

To make the statement more explicit, consider two dimensions. Given a point $\underline{r}_0 = (a,b)$ and unit vector $\underline{u} = (u_1, u_2)$, say, we have

V26: Directional derivative

$$D_{\underline{u}} f(a,b) = \lim_{h\to 0} \frac{f(a+hu_1, b+hu_2) - f(a,b)}{h}.$$

Directional derivatives can be computed from the gradient.

**Lemma 4.1.** *If $f$ is a function, $\underline{r}_0$ is a point, and $\underline{u}$ is a unit vector, then*

$$D_{\underline{u}} f(\underline{r}_0) = \underline{u} \cdot \nabla f(\underline{r}_0). \tag{4.4}$$

*Proof.* Let $\underline{r}(h) := \underline{r}_0 + h\,\underline{u}$, i.e., a set of points on a straight line, parameterized by $h$, which passes through $\underline{r}_0$ and along $\underline{u}$. Then, the directional derivative defined in Eq. (4.3) can be written as

$$D_{\underline{u}} f(\underline{r}_0) = \lim_{h\to 0} \frac{f[\underline{r}(h)] - f[\underline{r}(0)]}{h} = \frac{\mathrm{d}}{\mathrm{d}h} f[\underline{r}(h)]\bigg|_{h=0} .$$

The derivative of $\underline{r}(h)$ is just $\underline{r}'(h) = \underline{u}$, so the chain rule gives

$$D_{\underline{u}} f(\underline{r}_0) = \underline{u} \cdot \nabla f[\underline{r}(h)]\Big|_{h=0} = \underline{u} \cdot \nabla f(\underline{r}_0). \qquad \triangleleft$$

As an illustration, consider again the two-dimensional case, with $\underline{r}_0 = (a, b)$ and $\underline{u} = (u_1, u_2)$, which gives explicitly

$$D_{\underline{u}} f(a, b) = u_1\, f_x(a, b) + u_2\, f_y(a, b)\,.$$

Note that the standard partial derivatives are the directional derivatives in the directions of the coordinate axes,

$$\begin{aligned} \frac{\partial f}{\partial x}(a, b) &= f_x(a, b) = f_1(a, b) = D_{\underline{i}} f(a, b)\,, \\ \frac{\partial f}{\partial y}(a, b) &= f_y(a, b) = f_2(a, b) = D_{\underline{j}} f(a, b)\,, \end{aligned}$$

with $\underline{i} = (1, 0)$ and $\underline{j} = (0, 1)$.

**Example 4.10.** Let $f(x, y) = \tan^{-1}(y/x)$ for $x > 0$ and $y > 0$. What is the directional derivative of this function at $(2, 1)$ in the direction of $\underline{u} = (\frac{3}{5}, \frac{4}{5})$?

*Solution.* Recall that

$$\frac{\mathrm{d}}{\mathrm{d}x}(\tan^{-1} x) = \frac{1}{1 + x^2}\,,$$

so

$$\begin{aligned} f_x &= \frac{1}{1 + y^2/x^2}\left(-\frac{y}{x^2}\right) = -\frac{y}{x^2 + y^2}\,, \qquad & f_x(2, 1) &= -\tfrac{1}{5}, \\ f_y &= \frac{1}{1 + y^2/x^2}\left(\frac{1}{x}\right) = \frac{x}{x^2 + y^2}, \qquad & f_y(2, 1) &= \tfrac{2}{5}, \end{aligned}$$

thus the gradient vector at $(2, 1)$ is

$$\nabla f(2, 1) = (-\tfrac{1}{5}, \tfrac{2}{5})\,,$$

and hence

$$D_{\underline{u}} f(2, 1) = \underline{u} \cdot \nabla f(2, 1) = \left(\tfrac{3}{5}, \tfrac{4}{5}\right) \cdot \left(-\tfrac{1}{5}, \tfrac{2}{5}\right) = \tfrac{1}{5}\,. \qquad \triangleleft$$

The directional derivative allows us to geometrically interpret the gradient and its particular role.

**Theorem 4.3** (Gradient). *The gradient $\nabla f(\underline{r}_0)$ is a vector that points in the direction of the steepest increase of the function $f$ at the point $\underline{r}_0$, and its magnitude is equal to the rate of increase.*

*Proof.* Consider any unit vector $\underline{u}$ which will enclose an angle $\alpha$ with the gradient vector $\nabla f(\underline{r}_0)$. Using $\underline{a} \cdot \underline{b} = \|\underline{a}\| \, \|\underline{b}\| \, \cos\theta$, where $\theta$ is the angle between $\underline{a}$ and $\underline{b}$, the directional derivative takes the form

$$D_{\underline{u}} f(\underline{r}_0) = \underline{u} \cdot \nabla f(\underline{r}_0) = \cos\alpha \, \|\nabla f(\underline{r}_0)\| \; ,$$

where $\alpha$ is the angle between $\nabla f(\underline{r}_0)$ and $\underline{u}$ and we have used $\|\underline{u}\| = 1$. The right-hand-side is greatest for $\alpha = 0$ which means that $\underline{u}$ points in the direction of $\nabla f(\underline{r}_0)$ and equals $D_{\underline{u}} f(\underline{r}_0) = \|\nabla f(\underline{r}_0)\|$. $\square$

Rephrasing this result in terms of the skiing example, our brain is capable to identify the direction of the (negative) gradient when searching for the steepest descent.

## 4.3 Tangents and normals (§14.6)

*Linear approximations*

For a function $f(x)$ of one variable, the best linear approximation to $f(x)$ near $x = a$ is given by the first-order Taylor polynomial of $f(x)$ about $x = a$. This is the unique linear function with the same value, $f(a)$, and first derivative, $f'(a)$, at $x = a$. Its graph is a line *tangent* to the graph of $f(x)$ at the point $(a, f(a))$.

V27: Linear approximations and gradient

An analogous statement holds functions of two (or more) variables. The *linear approximation* to $f(x,y)$ about the point $(a,b)$ is

$$f(x,y) \approx T_1(x,y) = f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b) \; . \quad (4.5)$$

When we rewrite the right-hand-side of this equation using vector notation, the gradient appears,

$$T_1(\underline{r}) = f(\underline{r}_0) + (\underline{r} - \underline{r}_0) \cdot \nabla f(\underline{r}_0) \, ,$$

where $\underline{r} = (x,y)$ and $\underline{r}_0 = (a,b)$. This approximation to the function $f(x,y)$ has the same value and first-order partial derivatives as the function $f$ at $(a,b)$. The graph $z = T_1(x,y)$ is a plane *tangent* to the graph of $f(x,y)$ at the point $(a,b,f(a,b))$.

**Example 4.11.** Let $f(x,y) = y^2 + 2\sin x$. Find the linear approximation to $f(x,y)$ about the point $\underline{r}_0 = (\frac{\pi}{6}, 1)$.

*Solution.* We have

$$f(\tfrac{\pi}{6}, 1) = 2, \qquad f_x(\tfrac{\pi}{6}, 1) = 2\cos\tfrac{\pi}{6} = \sqrt{3}, \qquad f_y(\tfrac{\pi}{6}, 1) = 2$$

so the linear approximation is

$$\begin{aligned} T_1(x, y) &= 2 + \sqrt{3}(x - \tfrac{\pi}{6}) + 2(y - 1) \\ &= \sqrt{3}\,x + 2y - \frac{\pi}{2\sqrt{3}} \end{aligned} \qquad \triangleleft$$

This is a two-dimensional plane in $\mathbb{R}^3$, intercepting the $z$-axis at $-\pi/2\sqrt{3}$ and approaching ever larger values in the first octant, i.e., for ever larger values of $x$ and $y$.

### *Level surfaces and tangent vectors*

In Definition 4.2 we had introduced the idea of *level sets* and *level curves* for functions of two variables. These concepts generalize naturally when we consider functions of three (or even more) variables.

**Definition 4.7** (Level surface)**.** If $F(x, y, z)$ is a function of three variables, then a *level set* of $F$ is a set

$$\{(x, y, z) \in \mathbb{R}^3 \mid F(x, y, z) = c\}$$

for some constant $c \in \mathbb{R}$. If this is a smooth surface, then it is usually called a *level surface*.

Given an explicit expression for a function $F(x, y, z)$, it is important to be able to describe the planes tangent to its level surfaces. To do so, we formalize the idea of a vector tangent to a surface.

**Definition 4.8** (Tangent vector)**.** A vector is *tangent* to a surface $\mathcal{S}$ at a point $\underline{r}_0 \in \mathcal{S}$ if it is tangent at $\underline{r}_0$ to some curve in $\mathcal{S}$ passing through $\underline{r}_0$.

Let $\mathcal{S}$ be the level surface of $F : \mathbb{R}^3 \to \mathbb{R}$ passing through the point $\underline{r}_0$. If $\underline{u}$ is a vector tangent to $\mathcal{S}$ at $\underline{r}_0$, then $\underline{u}$ is tangent to some curve $\mathcal{C} \subset \mathcal{S}$ at $\underline{r}_0$. Because $\mathcal{C}$ lies in the level surface, $F$ is constant along $\mathcal{C}$, which means that, according to Lemma 4.1, the directional derivative along $\underline{u}$ must vanish,

$$0 = D_{\underline{u}} F(\underline{r}_0) = \underline{u} \cdot \nabla F(\underline{r}_0).$$

This means that $\nabla F(\underline{r}_0)$ is perpendicular to $\underline{u}$.

Due to $F(x(t), y(t), z(t)) = c$ and the chain rule, which we can express using Eq. (4.2) we must have

$$F'(t) = \underline{r}'(t) \cdot \nabla F(\underline{r}(t)) = 0$$

whenever we travel on a curve located on a level surface. Thus, the vector $\underline{r}'(t)$, known as the *tangent vector* to the curve $\mathcal{C}$ at the point $\underline{r}(t)$, necessarily lies in the plane tangent to the level surface $\mathcal{S}$.

### *Normal vectors*

Normal vectors are a concept closely related to vectors tangent to a level curve or level surface. Let us begin with defining these vectors in the simplest possible case, i.e., for functions of two variables.

**Definition 4.9** (Normal vector)**.** A vector is said to be *normal* to a curve $\mathcal{C}$ at the point $(a,b) \in \mathcal{C}$ if it is perpendicular to any vector tangent to $\mathcal{C}$ at $(a,b)$.

Recall that a level set is the set of points such that $f(x,y) = c$ for some constant, $c$. If the set is a smooth curve, then it is a called a level curve. Along a level curve, the function $f(x,y)$ does not change its value. Thus, if $\underline{u}$ points in the direction of the line tangent to the level curve at the point $(a,b)$, then the derivative of $f(x,y)$ in the direction of $\underline{u}$ vanishes,

$$0 = D_{\underline{u}} f(a,b) = \underline{u} \cdot \nabla f(a,b)\,.$$

V28: Gradient and tangent vectors

The inner product only vanishes if $\nabla f(a,b)$ and $\underline{u}$ enclose a right angle which means that $\nabla f(a,b)$ is orthogonal (or normal) to the level curve at the point $(a,b)$. In other words, the normal to a level set is parallel to $\nabla f(a,b)$.

These ideas can be illustrated with a simple explicit example.

**Example 4.12.** For $f(x,y) = x^2 + y^2$, its gradient at the point $(x,y)$ is given by $\nabla f(x,y) = 2(x,y)$ corresponding to vectors all of which point away from the origin. Imagining the lines tangent to the circles depicted in Fig. 4.1, it becomes obvious that the gradient $\nabla f$ is normal to the level curve at each point.

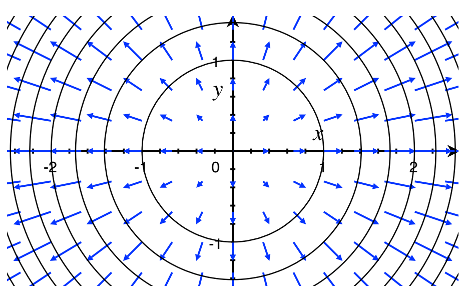



Figure 4.1: Level curves and gradient, $\nabla f$, for $f(x,y) = x^2 + y^2$.

Here is another example that involves an ellipse instead of circles.

**Example 4.13.** Find the line that intersects the curve $x^2 + 4y^2 = 8$ orthogonally at the point $(2,1)$.

*Solution.* Let $f(x,y) = x^2 + 4y^2$. We know the gradient vector $\nabla f(2,1)$ of $f(x,y)$ at the point $(2,1)$ to be normal to the level curve $f(x,y) = 8$ at that point. Hence we can determine the line we are looking by calculating $\nabla f(2,1)$.

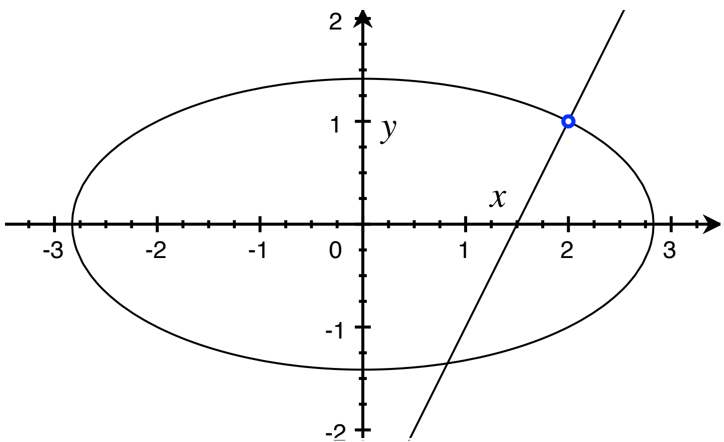



Figure 4.2: The curve $x^2 + 4y^2 = 8$ and the line perpendicular to it at $(2,1)$. Note that they are not perpendicular at the other point where they intersect.

Since $f_x(x,y) = 2x$ and $f_y(x,y) = 8y$, we have $\nabla f(2,1) = (4,8)$, thus we want the line going through $(2,1)$ that is parallel to this vector. By noting that the slope of this line must be $8/4 = 2$, we have

$$y = 2(x-2) + 1 = 2x - 3$$

as the equation of the line as shown in Fig. 4.2. Alternatively, the parametric form of the line is given by

$$\underline{r}(t) = \begin{pmatrix} 2 \\ 1 \end{pmatrix} + \begin{pmatrix} 4 \\ 8 \end{pmatrix} t = \begin{pmatrix} 2+4t \\ 1+8t \end{pmatrix}. \qquad \triangleleft$$

*Tangent lines to curves*

Consider a point $\underline{r}_0 = (a,b)$ lying on the level curve $f(x,y) = c$ and recall that $\nabla f(a,b)$ is *normal* to the level curve at $\underline{r}_0$. Suppose that the point $\underline{r} = (x,y)$ is on the line *tangent* to the curve at $\underline{r}_0$. Since $\underline{r} - \underline{r}_0$ is parallel to the tangent line, we must have that

$$\nabla f(\underline{r}_0) \cdot (\underline{r} - \underline{r}_0) = 0, \quad \text{or equivalently}$$
$$f_x(a,b)(x-a) + f_y(a,b)(y-b) = 0$$

is an equation for the line tangent to the level curve at $\underline{r}_0 = (a,b)$.

**Example 4.14.** Find the line tangent to the curve $x^2 - y^2 = 5$ at the point $(3,2)$.

*Solution.* Let $f(x,y) = x^2 - y^2$. Then $\frac{\partial f}{\partial x} = 2x$ and $\frac{\partial f}{\partial y} = -2y$ leading to $f_x(3,2) = 6$ and $f_y(3,2) = -4$. Therefore, the tangent line is given by

$$f_x(3,2)(x-3) + f_y(3,2)(y-2) = 6(x-3) - 4(y-2) = 0\,,$$

or, equivalently by $3x - 2y - 5 = 0$, which allows us to express $y$ as a function of $x$, $y = \frac{3}{2}x - \frac{5}{2}$. ◁

The graph of a function associated with $y = f(x)$ can also be described as the level curve of a function of two variables, $F(x,y) = f(x) - y$ and setting $F(x,y) = 0$. In this case, $\frac{\partial F}{\partial x} = f'(x)$ and $\frac{\partial F}{\partial y} = -1$, hence the equation for the line tangent to the graph at $(a, f(a))$ is

$$\begin{aligned} 0 &= F_x[a, f(a)](x-a) + F_y[a, f(a)][y - f(a)] \\ &= f'(a)(x-a) - [y - f(a)]\,, \end{aligned}$$

establishing the close relation to the first-order Taylor polynomial of $f(x)$ at $x = a$,

$$y = f(a) + f'(a)(x-a)\,.$$

*Tangent planes to surfaces*

In a three dimensional setting, the concept of a tangent lines generalizes to that of a tangent plane.

**Definition 4.10** (Tangent plane)**.** A plane is *tangent* to a surface $\mathcal{S}$ at a point $\underline{r}_0 \in \mathcal{S}$ if it passes through $\underline{r}_0$ and is parallel to any vector tangent to $\mathcal{S}$ at $\underline{r}_0$.

Consider the level surface of $F(x,y,z)$ through $\underline{r}_0 = (x_0, y_0, z_0)$. If the point $\underline{r} = (x,y,z)$ is located on the plane tangent to the level surface of $F$ at $\underline{r}_0$, then the vector $\underline{r} - \underline{r}_0$ is parallel to the plane and hence tangent to the surface which means that

V29: Tangent lines and planes

$$\nabla F(\underline{r}_0) \cdot (\underline{r} - \underline{r}_0) = 0$$

must hold. In terms of coordinates $x, y$ and $z$, the equation of the plane tangent to the level surface at $(x_0, y_0, z_0)$ takes the form

$$F_x(x_0,y_0,z_0)(x-x_0) + F_y(x_0,y_0,z_0)(y-y_0) + F_z(x_0,y_0,z_0)(z-z_0) = 0\,.$$

**Example 4.15.** Find the plane tangent to $x^2 + y^2 + z^2 = a^2$ at $(x_0, y_0, z_0)$. (This is a sphere of radius $a$ centred at the origin.)

*Solution.* Let $F(x,y,z) = x^2 + y^2 + z^2$. Then

$$\frac{\partial F}{\partial x} = 2x_0, \quad \frac{\partial F}{\partial y} = 2y_0, \quad \frac{\partial F}{\partial z} = 2z_0$$

at $(x_0, y_0, z_0)$. Therefore the equation for the tangent plane we are looking for is given by

$$F_x(x_0, y_0, z_0)(x - x_0) + F_y(x_0, y_0, z_0)(y - y_0) + F_z(x_0, y_0, z_0)(z - z_0)$$
$$= 2x_0(x - x_0) + 2y_0(y - y_0) + 2z_0(z - z_0) = 0\,,$$

i.e.,

$$x_0 x + y_0 y + z_0 z = x_0^2 + y_0^2 + z_0^2.$$

However, $x_0^2 + y_0^2 + z_0^2 = a^2$, because $(x_0, y_0, z_0)$ is on the sphere, hence the equation simplifies to

$$x_0 x + y_0 y + z_0 z = a^2.$$

This result can also be obtained by another more geometric argument. The vector $\underline{r}_0 = (x_0, y_0, z_0)$ is normal to the sphere at the point $\underline{r}_0$. If $\underline{r} = (x, y, z)$ is a point on the tangent plane, then $\underline{r} - \underline{r}_0$ must be tangent to the sphere, hence $0 = \underline{r}_0 \cdot (\underline{r} - \underline{r}_0) = \underline{r}_0 \cdot \underline{r} - a^2$. $\triangleleft$

*Remark* 4.3. In analogy to the two-dimensional case, the graph of the function $f(x, y)$ can also be constructed as a level surface. If we define $F(x, y, z) = f(x, y) - z$, then the graph of $f$ is the level surface when setting $F(x, y, z) = 0$.

To find the plane tangent to the graph of $f(x, y)$ at $(a, b, f(a, b))$, we compute

$$F_x(x, y, z) = f_x(x, y)\,, \quad F_y(x, y, z) = f_y(x, y)\,, \quad F_z(x, y, z) = -1\,.$$

Hence the equation of the tangent plane is

$$f_x(a, b)(x - a) + f_y(a, b)(y - b) - [z - f(a, b)] = 0\,,$$

i.e.,

$$z = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b)\,.$$

Yet again we notice the direct link between the plane and the (graph of the) linear approximation to $f(x, y)$ about the point $(a, b)$.

## 4.4 *Vector fields*

Functions map elements of one space to elements of another space, and they are given specific names depending on the dimension of their range.

**Definition 4.11** (Scalar field)**.** A function $f : \mathbb{R}^n \to \mathbb{R}$ that maps points of the $n$-dimensional space $\mathbb{R}^n$ to real numbers $\mathbb{R}$ is called a *scalar field*.

The function $F(x, y, z) = x^2 + y^2 + z^2$ defines a scalar field since it maps triples $(x, y, z) \in \mathbb{R}^3$ to real numbers $\mathbb{R}$.

V30: Scalar and vector fields

**Definition 4.12** (Vector field)**.** A function $f : \mathbb{R}^n \to \mathbb{R}^n$ that maps points of the $n$-dimensional space $\mathbb{R}^n$ to other elements of $\mathbb{R}^n$ is called a *vector field*.

For example, a vector field in two dimensions takes the form

$$\underline{f}(x, y) = u(x, y)\,\underline{i} + v(x, y)\,\underline{j}\,,$$

for some functions $u$ and $v$. Both the input of $f$ and its output are elements of the space $\mathbb{R}^2$, i.e., they can be thought of as vectors,

$$\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \qquad \text{and} \qquad \begin{pmatrix} u(x, y) \\ v(x, y) \end{pmatrix} \in \mathbb{R}^2\,.$$

**Example 4.16.** If $\underline{f}(x, y) = (x^2 + y, xy)$, then $\underline{f}(1, 2) = (3, 2)$.

The gradient operator constructs a vector field out of any scalar field.

**Example 4.17.** Give an example of a vector field obtained as a gradient of a scalar field.

*Solution.* The three partial derivatives of the scalar function $F(x, y, z) = x^2 + y^2 + z^2 \in \mathbb{R}$ are given by $F_x = 2x$, $F_y = 2y$ and $F_z = 2z$ so that we find

$$\nabla F(x, y, z) = 2 \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3\,, \qquad \text{or} \qquad \nabla F(\underline{r}) = 2\underline{r}\,. \qquad \triangleleft$$

*Remark* 4.4. Importantly, the converse is false: Not every vector field is the gradient of a scalar field. This observation raises an interesting question: which vector fields can be obtained as the gradient of a scalar field? For those that can be so obtained we will say that the vector field $\underline{f}$ *is* a gradient.

### *A necessary condition for a vector field to be a gradient*

Here is a *necessary* condition for a two-component vector field $\underline{f}$ depending on two variables $x$ and $y$ to be the gradient of a scalar field, i.e., to possess a potential function.

V31: Vector fields with potentials

**Lemma 4.2.** *If a (continuously differentiable) vector field $\underline{f} = (u, v)$ is a gradient, i.e., if $\underline{f} = \nabla\varphi$ for some scalar field $\varphi(x,y)$, then its components satisfy*

$$\frac{\partial v}{\partial x} = \frac{\partial u}{\partial y}\,.$$

*Proof.* The equation $\underline{f} = \nabla\varphi$ implies

$$u = \frac{\partial\varphi}{\partial x}, \quad v = \frac{\partial\varphi}{\partial y}.$$

Continuous differentiability of $u = \varphi_x$ and $v = \varphi_y$ implies continuity of $\varphi_{xy}$ and $\varphi_{yx}$, so the hypotheses of Clairaut's Theorem (see Theorem 4.1) are satisfied, and therefore $\varphi_{xy} = \varphi_{yx}$. This shows that

$$\frac{\partial v}{\partial x} = \frac{\partial}{\partial x}\left(\frac{\partial\varphi}{\partial y}\right) = \frac{\partial}{\partial y}\left(\frac{\partial\varphi}{\partial x}\right) = \frac{\partial u}{\partial y}. \qquad \triangleleft$$

If a vector field $\underline{f}$ emerges as the gradient of a scalar field $\varphi$, the function $\varphi(x, y)$ (or $\varphi(x, y, z)$) is called a *potential* (function) for $\underline{f}$.

**Example 4.18.** Is $\underline{f}(x, y) = (3x^2y, x^3y)$ the gradient of a scalar field?

*Solution.* Since we have

$$\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} = 3x^2y - 3x^2 \neq 0,$$

the vector field $\underline{f}$ is not the gradient of a scalar field. $\triangleleft$

**Example 4.19.** Is $\underline{f}(x, y) = (2xy + \cos x)\underline{i} + (x^2 - \sin y)\underline{j}$ a gradient?

*Solution.* Since we have

$$\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} = 2x - 2x = 0,$$

the vector field $\underline{f}$ *may* be the gradient of a scalar field.

To find out, suppose that $\nabla\varphi = \underline{f}$, i.e.,

$$\frac{\partial\varphi}{\partial x} = u = 2xy + \cos x, \tag{4.6}$$

$$\frac{\partial\varphi}{\partial y} = v = x^2 - \sin y. \tag{4.7}$$

By integrating Eq. (4.6) with respect to $x$, regarding $y$ as constant, we find

$$\varphi = x^2y + \sin x + g(y).$$

Note that $g(y)$ can be any function of $y$ because it will vanish when differentiated with respect to $x$. By substituting this expression for $\varphi$ into Eq. (4.7) we obtain

$$\frac{\partial\varphi}{\partial y} = x^2 + g'(y) = x^2 - \sin y.$$

Hence $g'(y) = -\sin y$, so $g(y) = \cos y + C$, where $C$ is a constant. Thus, the scalar field

$$\varphi = x^2y + \sin x + \cos y + C$$

has the desired gradient as can be readily verified,

$$\nabla\varphi = (2xy + \cos x)\underline{i} + (x^2 - \sin y)\underline{j} = \underline{f}(x, y). \qquad \triangleleft$$

## 4.5 *Implicit differentiation*

### *Implicitly defined functions — one variable*

An explicit definition for a function $y(x)$ is a formula for computing $y$ from $x$, via $y = y(x)$. More generally, relations of the form $f(x, y) = c$ (which cannot be solved explicitly for $y$) provide an *implicit* definition of a function $y(x)$; here $c$ is a constant. Effectively, we just have an equation that relates $x$ and $y$.

This way of defining functions implicitly can be interpreted geometrically since the formula $f(x, y) = c$ defines a level curve of $f$. The implicit definition simply declares this curve (or part of it) to be the graph of a new function.

Interestingly, it is often possible to discuss the properties of an implicitly function in detail, even without being able to write it as an explicit function such as $y = y(x)$. It is possible, for example, to determine the derivative of the implicitly defined function from the relation $f(x, y) = c$.

**Example 4.20.** The relation $x^2 + xy + y^2 = 7$ defines $y$ as a function of $x$ near $(2, 1)$. Compute $\frac{dy}{dx}$ at $(2, 1)$.

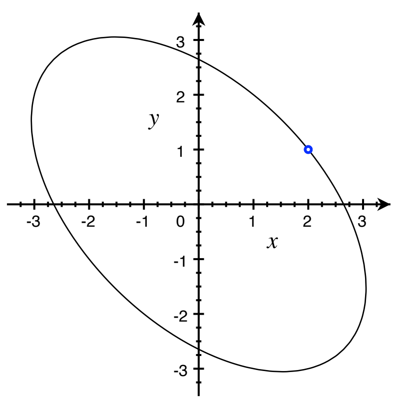



Figure 4.3: The ellipse $x^2 + xy + y^2 = 7$ and the point $(2, 1)$.

*Solution.* If $y(x)$ is the implicitly defined function, then we can think of the given relation as

$$x^2 + xy(x) + y(x)^2 = 7\,.$$

By differentiating both sides (with respect to $x$) we find

$$2x + y + x\frac{dy}{dx} + 2y\frac{dy}{dx} = 0\,.$$

Hence

$$\frac{dy}{dx} = -\frac{2x + y}{x + 2y}\,.$$

At $x = 2$, $y = 1$ we have

$$\left.\frac{\mathrm{d}y}{\mathrm{d}x}\right|_{(x,y)=(2,1)} = -\frac{5}{4}\,. \qquad \triangleleft$$

Using partial derivatives, we can make this method more systematic.

V32: Implicit differentiation - one variable

**Theorem 4.4** (Implicit differentiation (1))**.** *If $y(x)$ is defined implicitly by the equation $f(x,y) = c$ then*

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{-\frac{\partial f}{\partial x}}{\frac{\partial f}{\partial y}} = -\frac{f_x(x,y)}{f_y(x,y)}\,. \tag{4.8}$$

*Proof.* Since $c$ is a constant,

$$0 = \frac{\mathrm{d}}{\mathrm{d}x} f[x, y(x)]\,.$$

Recall the chain rule (Theorem 4.2) for a function of two variables both depending on $t$,

$$\frac{\mathrm{d}}{\mathrm{d}t} f[x(t), y(t)] = x'(t)\, f_x[x(t), y(t)] + y'(t)\, f_y[x(t), y(t)]\,.$$

In particular, if $t = x$, then $\frac{\mathrm{d}x}{\mathrm{d}x} = 1$ and hence we want

$$0 = \frac{\mathrm{d}}{\mathrm{d}x} f[x, y(x)] = f_x[x, y(x)] + \frac{\mathrm{d}y}{\mathrm{d}x} f_y[x, y(x)].$$

Solving for $\frac{\mathrm{d}y}{\mathrm{d}x}$ gives Eq. (4.8). $\square$

**Example 4.21.** Again, let $x^2 + xy + y^2 = 7$ and compute $\frac{\mathrm{d}y}{\mathrm{d}x}$ at $(2, 1)$ using Theorem 4.4.

*Solution.* Denoting $f(x,y) = x^2 + xy + y^2$, we have $\frac{\partial f}{\partial x} = 2x + y$ and $\frac{\partial f}{\partial y} = x + 2y$, so

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{2x + y}{x + 2y}\,,$$

and when $x = 2$ and $y = 1$, this equals $-\frac{5}{4}$ as we have already seen. $\triangleleft$

**Example 4.22.** Let $y$ be defined implicitly by $y^2 - x^2 - \sin(xy) = 1$. Find $\frac{\mathrm{d}y}{\mathrm{d}x}$ in terms of $x$ and $y$.

*Solution.* Let $f(x,y) = y^2 - x^2 - \sin(xy)$, then

$$f_x(x,y) = -2x - y\cos(xy)$$

and

$$f_y(x,y) = 2y - x\cos(xy)\ ,$$

so

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{f_x(x,y)}{f_y(x,y)} = \frac{2x + y\cos(xy)}{2y - x\cos(xy)}. \qquad \triangleleft$$

*Remark* 4.5. Although $\frac{dy}{dx}$ is a function of $x$, this method does not (usually) give an explicit expression for the derivative. To compute $\frac{dy}{dx}$ for a given value of $x$, we must first compute $y$ (perhaps only approximately) and then insert this value into the expression.

### *Implicitly defined functions — two variables*

We have seen that the relation $F(x,y,z) = c$ defines a level surface of the three-variable function $F(x,y,z)$. We can implicitly define a function of two variables by declaring (part of) this surface to be its graph.

To illustrate this idea, consider the relation

$$x^2 + y^2 + z^2 = a^2 \,, \tag{4.9}$$

where $a > 0$ is a constant. Geometrically, Eq. (4.9) defines a sphere of radius $a$ centred at the origin. We can think of the relation as an implicit definition of a function $z(x,y)$ with a graph which is part of the sphere—just as we did in the lower-dimensional case of the ellipse in Example 4.20.

The example of the sphere allows us to solve for $z$, resulting in explicit expressions of two functions

$$z(x,y) = \pm\sqrt{a^2 - x^2 - y^2} \,. \tag{4.10}$$

If we choose the root with the positive sign, then $z$ is a function of $x$ and $y$ whose graph is the *upper* hemisphere; if we pick the root with the negative sign, then $z$ is a *different* function whose graph is the *lower* hemisphere.

It is often important to specify which variable is kept fixed when takes a derivative of an implicitly defined function. For example, when we regard $z$ as a function of $(x,y)$ and differentiate with respect to $x$, we can write $(\partial z/\partial x)_y$ to indicate that $y$ is fixed. Similarly, if we regard $y$ as a function of $(x,z)$ and differentiate it with respect to $z$, then we write $(\partial y/\partial z)_x$.

In the remainder of this section, it is important to distinguish between *partial derivatives* such as $F_x$ etc. and *subscripts* of round bracket such as $(\ldots)_x$ etc. which denote variables kept fixed. Unfortunately, this slightly confusing notation is standard practice.

To compute $(\partial z/\partial x)_y$ implicitly, we differentiate Eq. (4.9) with respect to $x$, regard $z$ as a function of $(x,y)$ and keep $y$ constant. We find

$$2x + 2z\left(\frac{\partial z}{\partial x}\right)_y = 0$$

or, when solving for the derivative,

$$\left(\frac{\partial z}{\partial x}\right)_y = -\frac{x}{z} \,.$$

Using the explicit expressions for $z(x,y)$ in Eq. (4.10), we obtain from the positive root that

$$\left(\frac{\partial z}{\partial x}\right)_y = -\frac{x}{\sqrt{a^2 - x^2 - y^2}} = -\frac{x}{z},$$

and from the negative root that

$$\left(\frac{\partial z}{\partial x}\right)_y = \frac{x}{\sqrt{a^2 - x^2 - y^2}} = -\frac{x}{z},$$

respectively. Both results these agree with what we found by implicit differentiation.

**Theorem 4.5** (Implicit differentiation (2)). *If $F(x,y,z) = c$ defines $z$ implicitly as a function of $(x,y)$ we have*

$$\left(\frac{\partial z}{\partial x}\right)_y = \frac{-\frac{\partial F}{\partial x}}{\frac{\partial F}{\partial z}} = -\frac{F_x}{F_z} \tag{4.11a}$$

*and*

$$\left(\frac{\partial z}{\partial y}\right)_x = -\frac{F_y}{F_z}. \tag{4.11b}$$

*Proof.* The function $z(x,y)$ is defined by

$$F(x,y,z(x,y)) = c\,.$$

To find $(\partial z/\partial x)_y$, we differentiate this equation *regarding $y$ as a constant*. Then the situation is exactly the same as for an implicitly defined function of one variable. By the chain rule, we find

V33: Implicit differentiation - two variables

$$\frac{\partial F}{\partial x} + \left(\frac{\partial z}{\partial x}\right)_y \frac{\partial F}{\partial z} = 0\,.$$

Solving for $\left(\frac{\partial z}{\partial x}\right)_y$ gives Eq. (4.11a). The proof of Eq. (4.11b) is the same. □

Similarly,

$$\left(\frac{\partial x}{\partial z}\right)_y = -\frac{F_z}{F_x}$$

and so forth.

**Example 4.23.** Suppose that $z$ depends upon $x$ and $y$, satisfies the relation

$$0 = z^5 + xz^3 + yz^2 + y^2 z - 1, \tag{4.12}$$

and equals 1 when $(x,y) = (-2,-2)$. Compute the partial derivatives of $z$ with respect to $x$ and $y$ when $(x,y) = (-2,-2)$.

Because this is a quintic equation, it cannot be solved explicitly for $z$.

*Solution.* Let

$$F(x,y,z) = z^5 + xz^3 + yz^2 + y^2 z - 1,$$

so that Eq. (4.12) is just $F(x,y,z) = 0$. The partial derivatives of this function are

$$\begin{aligned} F_x(x,y,z) &= z^3 \\ F_y(x,y,z) &= z^2 + 2yz \\ F_z(x,y,z) &= 5z^4 + 3xz^2 + 2yz + y^2\,, \end{aligned}$$

and at the point $(-2,-2,1)$ they take the values

$$\begin{aligned} F_x(-2,-2,1) &= 1\,, \\ F_y(-2,-2,1) &= -3\,, \\ F_z(-2,-2,1) &= -1\,, \end{aligned}$$

so that using Theorem 4.5 we obtain the derivatives at the specified point,

$$\left(\frac{\partial z}{\partial x}\right)_y\Bigg|_{(x,y)=(-2,-2)} = -\frac{F_x(-2,-2,1)}{F_z(-2,-2,1)} = 1$$

and

$$\left(\frac{\partial z}{\partial y}\right)_x\Bigg|_{(x,y)=(-2,-2)} = -\frac{F_y(-2,-2,1)}{F_z(-2,-2,1)} = -3. \qquad \triangleleft$$

## 4.6 *Polar coordinates*

Polar coordinates $r \in [0,\infty)$ and $\theta \in [0,2\pi)$ for the plane $\mathbb{R}^2$ can be thought of as an important example for the implicit definition of functions when introduced by the relations

$$x = r\cos\theta\,, \tag{4.13}$$
$$y = r\sin\theta\,. \tag{4.14}$$

The Cartesian coordinates take values $x, y \in \mathbb{R}$.

**Example 4.24.** Determine the four partial derivatives $(\partial r/\partial x)_y$, $(\partial r/\partial y)_x$, $(\partial\theta/\partial y)_x$ and $(\partial\theta/\partial y)_x$.

V34: Implicit differentiation - polar coordinates

*Solution.* The full notation makes the following computations too cluttered, so just keep in mind that $x$ and $y$ are the independent variables. We now need to read the equation $x = r\cos\theta$ as

$$x = r(x,y)\cos\theta(x,y)\,,$$

so that differentiating it with respect to $x$ gives

$$1 = \frac{\partial r}{\partial x}\cos\theta - r\sin\theta\frac{\partial\theta}{\partial x}\,, \tag{4.15}$$

and the derivative of Eq. (4.13) with respect to $y$ is obtained as

$$0 = \frac{\partial r}{\partial y}\cos\theta - r\sin\theta\frac{\partial\theta}{\partial y}. \tag{4.16}$$

Repeating this procedure for Eq. (4.14) gives us another pair of equations containing the derivatives we wish to determine,

$$0 = \frac{\partial r}{\partial x}\sin\theta + r\cos\theta\frac{\partial\theta}{\partial x} \qquad \text{and} \qquad 1 = \frac{\partial r}{\partial y}\sin\theta + r\cos\theta\frac{\partial\theta}{\partial y}\,. \tag{4.17}$$

To isolate the partial derivatives is a matter of linear algebra—we need to solve four simultaneous linear equations which can be done by Gaussian elimination or, more elegantly, by using matrix notation. Let

$$\mathbb{A} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

and

$$\mathbb{B} = \begin{pmatrix} \frac{\partial r}{\partial x} & \frac{\partial r}{\partial y} \\ \frac{\partial\theta}{\partial x} & \frac{\partial\theta}{\partial y} \end{pmatrix}.$$

The four equations we just derived can be combined into this relation

$$\begin{pmatrix} \frac{\partial x}{\partial x} & \frac{\partial x}{\partial y} \\ \frac{\partial y}{\partial x} & \frac{\partial y}{\partial y} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} \begin{pmatrix} \frac{\partial r}{\partial x} & \frac{\partial r}{\partial y} \\ \frac{\partial\theta}{\partial x} & \frac{\partial\theta}{\partial y} \end{pmatrix},$$

or $\mathbb{I} = \mathbb{A}\mathbb{B}$, Thus, we find $\mathbb{B} = \mathbb{A}^{-1}$ from inverting the matrix $\mathbb{A}$.

Using the determinant of $\mathbb{A}$,

$$\det\mathbb{A} = \cos\theta\, r\cos\theta - \sin\theta(-r\sin\theta) = r\,,$$

and using Cramer's rule for the inverse of a $2\times 2$ matrix, we find

$$\mathbb{B} = \mathbb{A}^{-1} = \frac{1}{r}\begin{pmatrix} r\cos\theta & r\sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\frac{\sin\theta}{r} & \frac{\cos\theta}{r} \end{pmatrix},$$

so that we can read off the answer to our question,

$$\left(\frac{\partial r}{\partial x}\right)_y = \cos\theta \qquad \left(\frac{\partial r}{\partial y}\right)_x = \sin\theta$$
$$\left(\frac{\partial\theta}{\partial x}\right)_y = -\frac{\sin\theta}{r} \qquad \left(\frac{\partial\theta}{\partial y}\right)_x = \frac{\cos\theta}{r}\,. \qquad \triangleleft$$

*Remark* 4.6. The result can, of course, also be found by solving (4.13) and (4.14) for $r$ and $\theta$ and differentiating the resulting *explicit* functions $r(x,y)$ and $\theta(x,y)$, i.e.,

$$r = \sqrt{x^2+y^2} \qquad \text{and} \qquad \theta = \tan^{-1}\frac{y}{x}\,,$$

assuming $x > 0$. We can now directly work out the partial derivatives,

$$\left(\frac{\partial r}{\partial x}\right)_y = r_x = \frac{x}{\sqrt{x^2+y^2}} = \cos\theta\,,$$
$$\left(\frac{\partial r}{\partial y}\right)_x = r_y = \frac{y}{\sqrt{x^2+y^2}} = \sin\theta\,,$$
$$\left(\frac{\partial \theta}{\partial x}\right)_y = \theta_x = -\frac{y}{x^2+y^2} = -\frac{\sin\theta}{r}\,,$$
$$\left(\frac{\partial \theta}{\partial y}\right)_x = \theta_y = \frac{x}{x^2+y^2} = \frac{\cos\theta}{r}\,.$$

which agree with the results obtained before by implicit differentiation.

### *The Laplacian in polar coordinates*

The description of many physical systems involves second-order partial derivatives of a function $f(x,y)$, the Laplacian being an important example.

**Definition 4.13** (Laplacian of a function)**.** Given a scalar field $f(x,y)$ in two dimensions, its *Laplacian* is defined as

$$\Delta f \equiv \nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}\,.$$

Often, the function $f(x,y)$ is known to possess symmetries that can be exploited by expressing it in terms of another set of coordinates such as $(r,\theta)$, say, so that $F(r,\theta) = f(r\cos\theta, r\sin\theta)$. One then needs to construct the Laplacian in terms of the new set of coordinates. We will illustrate the necessary transformations in the special case of replacing Cartesian by polar coordinates. Implicit differentiation is at the core of the procedure.

V35: Laplacian operator

**Theorem 4.6** (Laplacian in polar coordinates)**.** *Given a function $F(r,\theta) = f(r\cos\theta, r\sin\theta)$ depending on polar coordinates $(r,\theta)$, then the its Laplacian can be expressed as*

$$\Delta f \equiv \nabla^2 f = \frac{\partial^2 F}{\partial r^2} + \frac{1}{r}\frac{\partial F}{\partial r} + \frac{1}{r^2}\frac{\partial^2 F}{\partial \theta^2}\,. \tag{4.18}$$

*Proof.* We use the subscript notation for partial derivatives. By the chain rule, $f_x = r_x F_r + \theta_x F_\theta$ and

$$\begin{aligned} f_{xx} &= r_{xx}F_r + r_x\partial_x F_r + \theta_{xx}F_\theta + \theta_x \partial_x F_\theta \\ &= r_{xx}F_r + (r_x)^2 F_{rr} + r_x\theta_x F_{r\theta} + \theta_{xx}F_\theta + r_x\theta_x F_{r\theta} + (\theta_x)^2 F_{\theta\theta} \\ &= r_{xx}F_r + (r_x)^2 F_{rr} + 2r_x\theta_x F_{r\theta} + \theta_{xx}F_\theta + (\theta_x)^2 F_{\theta\theta}\,. \end{aligned}$$

Likewise,

$$f_{yy} = r_{yy}F_r + (r_y)^2 F_{rr} + 2r_y\theta_y F_{r\theta} + \theta_{yy}F_\theta + (\theta_y)^2 F_{\theta\theta}\,.$$

We found $r_x = \cos\theta$, $r_y = \sin\theta$, $\theta_x = -(1/r)\sin\theta$ and $\theta_y = (1/r)\cos\theta$ in Example 4.24. We can use these to compute

$$\begin{aligned}
r_{xx} &= \partial_x \cos\theta = -\theta_x \sin\theta = \frac{\sin^2\theta}{r} \\
r_{yy} &= \partial_y \sin\theta = \theta_y \cos\theta = \frac{\cos^2\theta}{r} \\
\theta_{xx} &= -\partial_x \frac{\sin\theta}{r} = -\frac{r\theta_x\cos\theta - r_x\sin\theta}{r^2} = \frac{2\sin\theta\cos\theta}{r^2} \\
\theta_{yy} &= \partial_y \frac{\cos\theta}{r} = \frac{-r\theta_y\sin\theta - r_y\cos\theta}{r^2} = -\frac{2\sin\theta\cos\theta}{r^2}
\end{aligned}$$

Bringing everything together we find

$$\begin{aligned}
\nabla^2 f &= f_{xx} + f_{yy} \\
&= ((r_x)^2 + (r_y)^2)F_{rr} + (r_{xx} + r_{yy})F_r + 2(r_x\theta_x + r_y\theta_y)F_{r\theta} + (\theta_{xx} + \theta_{yy})F_\theta + ((\theta_x)^2 + (\theta_y)^2)F_{\theta\theta} \\
&= F_{rr} + \frac{1}{r}F_r + \frac{1}{r^2}F_{\theta\theta}\,,
\end{aligned}$$

which is equivalent to the expression claimed. □

# 5
# Double integrals

*Main idea*

The definite integral $I_1 = \int_a^b f(x)\,\mathrm{d}x$ of a function $f(x)$ of a single variable $x \in [a,b]$ has an appealing geometrical interpretation: it is the *signed area* between the graph of $f$ and the interval $[a,b] \subset \mathbb{R}$ of the $x$-axis. If the function takes both positive and negative values, the value of the integral $I_1$ corresponds to taking the area of the region above $[a,b]$ and below the graph of $f$ and then subtracting area below $[a,b]$ but above the graph of $f$.

For a scalar function $f(x,y)$ of two variables, the idea of a definite integral over the two-variable function $f$ generalizes naturally. If $R \subset \mathbb{R}^2$ is a region of the $(x,y)$-plane, then the (double) integral of $f : R \to \mathbb{R}$ over $R$ is a number denoted

$$I_2 = \iint_R f(x,y)\,\mathrm{d}A\,,$$

where, intuitively, $\mathrm{d}A$ can be thought of as a small area just as $\mathrm{d}x$ is a small interval. The geometrical interpretation is similar to the one-dimensional case: it equals the *signed volume* between the graph of $f$ and the region $R$ in the $xy$-plane. As before, to obtain the value of the value of the integral $I_2$ we subtract volumes below the region $R$ of the $(x,y)$-plane but above $f(x,y)$ from the volumes above the region $R$ and below the graph of $f(x,y)$.

## 5.1 Integration in two dimensions (§15.2)

The rigorous definition of double integrals (see Sec. 5.2) is built upon the case of integrals over functions $f(x)$ of a single variable which we are familiar with. It turns out that one can often evaluate a double integral by using the familiar integration techniques twice. This approach is called *iterated integration*. It is simplest if the domain of integration, $R$, is a rectangle.

To establish the basics, let us briefly recall the definition of integration for functions of a single variable which is based on a limiting procedure involving Riemann sums. Suppose that we want to determine the area under the graph of a function $f(x)$—the answer will be the integral $I_1 = \int_a^b f(x)\,\mathrm{d}x$. First, we break up the domain $[a,b]$ into short intervals. If $f(x)$ does not change much over an interval, then it is close to being constant. Hence, the integral over that interval can be approximated by the *width* of the interval times a value of $f$ in the interval, defining the *height* of a rectangle above the interval we consider. Adding these approximations together gives

$$\int_a^b f(x)\,\mathrm{d}x \approx \sum_{i=1}^{n} f(x_i)\,\Delta x_i\,, \tag{5.1}$$

where $N$ is the number of small intervals, $x_i$ is a point in the $i^{\text{th}}$ interval, and $\Delta x_i$ is the width of the $i^{\text{th}}$ interval. This approximation is called a *Riemann sum*. The point is that if the intervals are made smaller and smaller, then this approximation should get better and better and converge to the actual value of the integral, i.e., they define the left-hand-side of Eq. (5.1).

The same approach works in two dimensions. We wish to define the integral $I_2 = \iint f\,\mathrm{d}A$ over some region $R$ in the $(x,y)$-plane. Imagine covering the region $R$ by small rectangles with sides parallel to the vertical and horizontal axes. The integral of $f(x,y)$ over the small rectangular region is approximately given by the area $\Delta A_{ij}$ of the rectangle times a value of $f$ at a point $(x_i, y_j)$ inside in that rectangle, i.e., it equals $f(x_i,y_j)\Delta A_{ij}$. Summing over all these contributions results in

$$\iint\limits_R f\,\mathrm{d}A \approx \sum_{\substack{\text{little rectangles}\\ \text{covering region } R}} f(x_i,y_j)\Delta A_{ij}\,.$$

The limit of an arbitrarily fine rectangular grid covering the region $R$ will provide a rigorous definition of the double integral written on the left-hand-side of the equation. We discuss the details of this procedure for rectangular regions first, subsequently moving on to more general shapes of the region $R$.

## 5.2 *Double integrals for rectangular regions (§15.1)*

V36: Rectangular regions & weak Fubini

The simplest case to make explicit the idea described above occurs when the region $R$ of interest is a rectangle, $R = [a,b] \times [c,d]$. Let us cut up this rectangle into small rectangles using vertical and horizontal lines. The double integral over the function $f(x,y)$ above any of these small rectangles is approximately equal to the area of the rectangle times a value of $f$ in that rectangle. More explicitly, if the $(i,j)^{\text{th}}$ little rectangle has width $\Delta x_i$ and height $\Delta y_j$, then its area is $\Delta A_{ij} = \Delta x_i \, \Delta y_j$. If $R$ is cut by vertical and horizontal lines into an $M \times N$ grid of rectangles, then

$$\iint\limits_R f \,\mathrm{d}A \approx \sum_{i=1}^{M} \sum_{j=1}^{n} f(x_i, y_j)\, \Delta y_j \, \Delta x_i .$$

The inner sum (the one over $j$) is just a Riemann sum for the integral $\int_c^d f(x_i, y)\,\mathrm{d}y = F(x_i)$, so that after letting $N \to \infty$ we find

$$\iint\limits_R f \,\mathrm{d}A \approx \sum_{i=1}^{M} \left( \int_c^d f(x_i, y)\,\mathrm{d}y \right) \Delta x_i \approx \int_a^b \int_c^d f(x,y)\,\mathrm{d}y\,\mathrm{d}x \,.$$

Finally, as $M \to \infty$, this approximation gets ever better and thereby effectively leading to a definition of the left-hand-side in terms of *iterated integrals*. We could repeat this procedure but sum over $x_i$ first, followed by the sum over $y_j$ and we would obtain the same answer. The precise statement is:

**Theorem 5.1** (Fubini's theorem [weak form]). *For a rectangular region*

$$R := \{(x,y) \in \mathbb{R}^2 \mid a \leq x \leq b,\ c \leq y \leq d\}$$

*and a continuous function $f : R \to \mathbb{R}$, we have*

$$\iint\limits_R f \,\mathrm{d}A = \int_a^b \left( \int_c^d f(x,y)\,\mathrm{d}y \right) \mathrm{d}x = \int_c^d \left( \int_a^b f(x,y)\,\mathrm{d}x \right) \mathrm{d}y \,.$$

Note that in this notation, iterated integration always proceeds from the inside outward, as indicated by the brackets.

*Remark* 5.1. The integral

$$\int_c^d f(x,y)\,\mathrm{d}y$$

is the signed area between the graph of $f$ and the line segment $\{x\} \times [c,d]$. This is a slice of the space under the graph. Fubini's theorem can be understood to state that, when calculating the volume under the graph $f(x,y)$, there is no difference whether the volume is cut into slices along horizontal (parallel to the $x$-axis) or vertical lines (parallel to the $y$-axis).

**Example 5.1.** Find the volume under the plane $x + y + z = 2$ and over the square region $R = [0,1] \times [0,1]$

*Solution.* The equation for the plane can be written as $z = 2 - x - y$, so that the volume can be found from two subsequent integrations,

$$V = \iint\limits_R (2 - x - y)\,\mathrm{d}A = \int_0^1 \left( \int_0^1 (2 - x - y)\,\mathrm{d}x \right) \mathrm{d}y \,.$$

The inner integration gives

$$\int_0^1 (2-x-y)\,\mathrm{d}x = \left[2x - \frac{x^2}{2} - xy\right]_{x=0}^{x=1} = 2 - \tfrac{1}{2} - y = \tfrac{3}{2} - y\,,$$

where we treat the variable $y$ as a constant. We end up with a function of $y$ alone which needs to be integrated in order to determine the volume,

$$V = \int_0^1 \left(\frac{3}{2} - y\right)\,\mathrm{d}y = \left[\frac{3y}{2} - \frac{y^2}{2}\right]_{y=0}^{y=1} = 1\,. \qquad \triangleleft$$

*Remark* 5.2. One can confirm that this answer agrees with our ordinary concept of volume by visualising the three-dimensional solid defined by the plane and the region $R$. Cutting the solid by the plane $z = 1$ and combining the part above this plane to the part below it produces a unit cube with sides of length 1.

**Example 5.2.** Find

$$\iint_R x\cos(x+y)\,\mathrm{d}A,$$

where $R = [0, 2\pi] \times [0, \pi]$.

*Solution.* In spite of a more complicated dependence on $x$ and $y$, the integral can be calculated in the same way,

$$\iint_R x\cos(x+y)\,\mathrm{d}A = \int_0^{\pi} \left(\int_0^{2\pi} x\cos(x+y)\,\mathrm{d}x\right)\,\mathrm{d}y\,.$$

The $x$-integration by parts gives

$$\begin{aligned}\int_0^{2\pi} x\cos(x+y)\,\mathrm{d}x &= \Big[x\sin(x+y)\Big]_{x=0}^{x=2\pi} - \int_0^{2\pi} \sin(x+y)\,\mathrm{d}x \\ &= 2\pi\sin(2\pi+y) + \Big[\cos(x+y)\Big]_{x=0}^{x=2\pi} \\ &= 2\pi\sin y\,,\end{aligned}$$

so that we find

$$\iint_R x\cos(x+y)\,\mathrm{d}x\,\mathrm{d}y = \int_0^{\pi} 2\pi\sin y\,\mathrm{d}y = 2\pi\Big[-\cos y\Big]_{y=0}^{y=\pi} = 4\pi\,.$$

Alternatively, we can perform the integration over $y$ first. In this case we have

$$\iint_R x\cos(x+y)\,\mathrm{d}x\,\mathrm{d}y = \int_0^{2\pi} \left(\int_0^{\pi} x\cos(x+y)\,\mathrm{d}y\right)\,\mathrm{d}x\,,$$

and thus

$$\int_0^{\pi} x\cos(x+y)\,\mathrm{d}y = \Big[x\sin(x+y)\Big]_{y=0}^{y=\pi} = x\sin(x+\pi) - x\sin x = -2x\sin x\,,$$

where we have used $\sin(x + \pi) = -\sin x$. Finally,

$$\begin{aligned}\iint\limits_R x\cos(x+y)\,\mathrm{d}x\,\mathrm{d}y &= -\int_0^{2\pi} 2x\sin x\,\mathrm{d}x \\ &= \Big[2x\cos x\Big]_{x=0}^{x=2\pi} + \int_0^{2\pi} 2\cos x\,\mathrm{d}x = 4\pi\,,\end{aligned}$$

in line with Fubini's theorem. $\lhd$

## 5.3 *General properties of double integrals*

It should not come as a surprise that many important properties of integrals of functions of a single variable are shared by those of functions over two (or even more) variables. So far, we have only considered integrals over rectangular regions. Still, in the general case, where the regions may take arbitrary shapes as explained next, in Sec. 5.4, the following list of features will continue to hold: double integrals are *linear*, *additive* and will *dominate* each other if the integrands do. Their validity is a consequence of defining (double) integrals in terms of (double) Riemann sums.

V37: General properties of double integrals

Addictive?!

*Linearity:* To calculate the integral over a sum of scalar multiples of two functions $f(x,y)$ and $g(x,y)$, one can equally well work out the integrals of the functions separately and add the results with the appropriate factors,

$$\iint\limits_R (af+bg)\,\mathrm{d}A = a\iint\limits_R f\,\mathrm{d}A + b\iint\limits_R g\,\mathrm{d}A\,,\qquad a,b\in\mathbb{R}\,.$$

We recognise this property in the single-variable case:

$$\int (au(x)+bv(x))\,\mathrm{d}x = a\int u(x)\,\mathrm{d}x + b\int v(x)\,\mathrm{d}x\,.$$

*Additivity:* Given two non-overlapping regions $R, S \subset \mathbb{R}^2$, we can evaluate the integrals separately,

$$\iint\limits_{R\cup S} f\,\mathrm{d}A = \iint\limits_R f\,\mathrm{d}A + \iint\limits_S f\,\mathrm{d}A\,.$$

Again, the one-dimensional analogue is well-known: given two intervals $J = [a,b]$ and $K = [c,d]$, we have

$$\int_{J\cup K} u(x)\,\mathrm{d}x = \int_J u(x)\,\mathrm{d}x + \int_K u(x)\,\mathrm{d}x\,.$$

*Domination:* If $f(x,y) \le g(x,y)$ for all $(x,y)\in R$, then the integral over $g(x,y)$ cannot be smaller than the integral over $f(x,y)$,

$$\iint\limits_R f\,\mathrm{d}A \le \iint\limits_R g\,\mathrm{d}A\,,$$

which makes sense in the one-dimensional case as well.

Finally, we mention the interpretation of a particular double integral: if we integrate the function $f(x,y) = 1$ over the region $R$ we obtain the *area* $\mathcal{A}$ of $R$,

$$\iint\limits_R \mathrm{d}A = \mathcal{A}(R)\,,$$

just as the integral over the function $f(x) = 1$ returns the *length* of the interval $[a,b]$, $\int_a^b \mathrm{d}x = (b-a)$.

## 5.4 *Double integrals for more general regions (§15.2)*

The restriction to rectangular regions is quite strong but a slight modification of the iterated integrals allows one to consider more general regions $R$. They are characterized by the property that one can describe its boundaries in a specific way: one of the variables varies between a minimal and a maximal value, say $x \in [a,b]$, while the other variable varies over a range that depends on the given value of $x$, i.e., $\varphi_1(x) \le y \le \varphi_2(x)$. The variables $x$ and $y$ may also swap their roles

V38: Fubini's theorem (strong)

It is useful to define the allowed regions to be of *Type 1* or of *Type 2* by if they are given by

$$R_1 = \Big\{(x,y) \in \mathbb{R}^2 \mid a \le x \le b,\ \varphi_1(x) \le y \le \varphi_2(x)\Big\} \qquad (5.2)$$

with continuous functions $\varphi_1(x)$ and $\varphi_2(x)$, or by

$$R_2 = \Big\{(x,y) \in \mathbb{R}^2 \mid c \le y \le d,\ \psi_1(y) \le x \le \psi_2(y)\Big\} \qquad (5.3)$$

with are continuous functions $\psi_1(y)$ and $\psi_2(y)$, respectively. Now we can spell out two generalizations of Fubini's theorem.

**Theorem 5.2** (Fubini's theorem [strong form])**.** *Let $f : R \to \mathbb{R}$ be a continuous function defined on a region $R \subset \mathbb{R}^2$.*

1. *If the region is of* Type 1*, then the double integral of $f(x,y)$ over $R$ can be written as an iterated integral of the form*

$$\iint\limits_R f\,\mathrm{d}A = \int_a^b \left(\int_{\varphi_1(x)}^{\varphi_2(x)} f(x,y)\,\mathrm{d}y\right)\mathrm{d}x\,. \qquad (5.4a)$$

2. *If the region is of* Type 2*, then the double integral of $f(x,y)$ over $R$ can be written as an iterated integral of the form*

$$\iint\limits_R f\,\mathrm{d}A = \int_c^d \left(\int_{\psi_1(y)}^{\psi_2(y)} f(x,y)\,\mathrm{d}x\right)\mathrm{d}y\,. \qquad (5.4b)$$

Thus, in order to calculate a double integral over a region different from a rectangle, one typically begins by sketching the domain of integration $R$. The boundary of $R$ will consist of several curves, which should be described by suitable functions of $x$ or $y$.

We have a region of *Type 1* if, according to Eq. (5.2), $a$ is the *minimum* value of $x$ for any $(x, y) \in R$, while $b$ is the *maximum* value of $x$ for any $(x, y) \in R$. If we intersect a region of this type with a *vertical* line, it must be a single line segment, otherwise the integral cannot be expressed as a single iterated integral in this way. The vertical line segment passing through some point $x$ has endpoints $(x, \varphi_1(x))$ and $(x, \varphi_2(x))$. If these properties hold, the double integral of $f$ over $R$ can be evaluated according to Eq. (5.4a).

Alternatively, one can check for the region $R$ to be of *Type 2* (this time using *horizontal* lines instead of vertical lines), and proceed in a similar way to set up the integral in Eq. (5.4b). If the region $R$ turns out to be neither *Type 1* nor 2, one may need to split up $R$ into simpler regions, using additivity of the double integrals.

Let us work through a number of examples to illustrate the evaluation of double integrals using the stronger version of Fubini's theorem.

**Example 5.3.** Compute

V39: Strong Fubini in action

$$I = \iint_R x^3 y^3 \,\mathrm{d}A\,,$$

over the region $R = \{(x, y) \in \mathbb{R}^2 \mid 0 \le x \le 1,\ x^2 \le y \le 1\}$ depicted in Fig. 5.1.

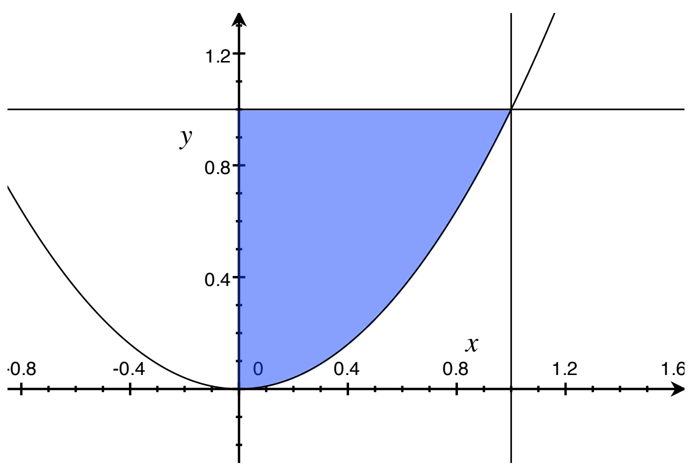



Figure 5.1: The domain of integration $R$ of Example 5.3 is both of *Type 1* and of *Type 2*.

*Solution.* Letting $a = 0$, $b = 1$, $\varphi_1(x) = x^2$ and $\varphi_2(x) = 1$, we see that the region $R$ is of *Type 1* since it conforms with the description given in Eq. (5.2). Thus, using Eq. (5.4a), we find

$$I = \iint_R x^3 y^3 \,\mathrm{d}A = \int_0^1 \left( \int_{x^2}^1 x^3 y^3 \,\mathrm{d}y \right) \mathrm{d}x\,.$$

Performing the integrations is straightforward. Starting with $y$ and treating the $x$-dependent limits as constants,

$$\int_{x^2}^{1} x^3 y^3 \, \mathrm{d}y = \left[ \frac{x^3 y^4}{4} \right]_{y=x^2}^{y=1} = \frac{x^3}{4} - \frac{x^{11}}{4} \, .$$

Finally, we integrate over $x$ to find the value of the double integral,

$$I = \int_0^1 \left( \frac{x^3}{4} - \frac{x^{11}}{4} \right) \mathrm{d}x = \left[ \frac{x^4}{16} - \frac{x^{12}}{48} \right]_{x=0}^{x=1} = \frac{1}{16} - \frac{1}{48} = \frac{1}{24} \, .$$

Interestingly, we can perform the integrations in the opposite order as well. This possibility arise since the region $R$ is also of *Type 2*,

$$R = \{ (x, y) \in \mathbb{R}^2 \mid 0 \leq y \leq 1, \; 0 \leq x \leq \sqrt{y} \} \, ,$$

i.e., we have $c = 0$, $d = 1$, $\psi_1(y) = 0$ and $\psi_2(y) = \sqrt{y}$. Thus, according to Eq. (5.4b), we find

$$I = \int_0^1 \left( \int_0^{\sqrt{y}} x^3 y^3 \, \mathrm{d}x \right) \mathrm{d}y \, .$$

Integrating over $x$ first leaves us with a function of $y$,

$$\int_0^{\sqrt{y}} x^3 y^3 \, \mathrm{d}x = \left[ \frac{x^4 y^3}{4} \right]_{x=0}^{x=\sqrt{y}} = \frac{y^5}{4} \, ,$$

and we obtain indeed the same result for $I$ as before,

$$I = \int_0^1 \frac{y^5}{4} \, \mathrm{d}y = \left[ \frac{y^6}{24} \right]_{y=0}^{y=1} = \frac{1}{24} \, . \qquad \triangleleft$$

**Example 5.4.** Compute the volume of the tetrahedron with vertices $(0,0,0)$, $(1,0,0)$, $(0,1,0)$, $(0,0,1)$.

*Solution.* The equation of the plane through the three points not at the origin is given by $x + y + z = 1$, or equivalently, $z = 1 - x - y$.

V40: Volume of tetrahedron

The body of the tetrahedron sits between this plane and a triangular region $R$ in the $(x, y)$-plane, with vertices $(0,0)$, $(1,0)$, and $(0,1)$. This region is of *Type 2* since we can describe it as

$$R = \{ (x, y) \in \mathbb{R}^2 \mid 0 \leq y \leq 1, \; 0 \leq x \leq 1 - y \} \, .$$

Hence the volume can be found by evaluating a suitable iterated integral,

$$\begin{aligned} V &= \iint_R (1 - x - y) \, \mathrm{d}A = \int_0^1 \int_0^{1-y} (1 - x - y) \, \mathrm{d}x \, \mathrm{d}y \\ &= \int_0^1 \left[ x - \frac{x^2}{2} - xy \right]_{x=0}^{x=1-y} \mathrm{d}y = \int_0^1 \left[ 1 - y - \frac{(1-y)^2}{2} - (1-y)y \right] \mathrm{d}y \\ &= \int_0^1 \left[ \frac{y^2}{2} - y + \frac{1}{2} \right] \mathrm{d}y = \left[ \frac{y^3}{6} - \frac{y^2}{2} + \frac{y}{2} \right]_{y=0}^{y=2} = \frac{1}{6} \, . \qquad \triangleleft \end{aligned}$$

*Remark* 5.3. The tetrahedron is a "triangular cone". Recall that the volume of a cone with base area $A$ and height $h$ is $\frac{1}{3}hA$. This formula agrees with the above result because $A = \frac{1}{2}$ and $h = 1$.

*Remark* 5.4. Note that the triangular region is also of *Type 1*.

Fubini's theorem provides two options to calculate a double integral. Sometimes there are computational advantages to performing the integral in one way rather than in the other. It may even happen that it is impossible to calculate a double integral without reversing the order of integration, as the following example shows.

**Example 5.5.** Evaluate

$$I = \int_0^1 \left( \int_x^1 \mathrm{e}^{-y^2} \,\mathrm{d}y \right) \mathrm{d}x.$$

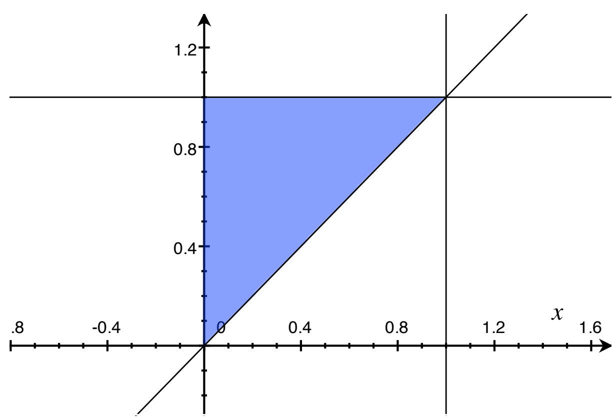



Figure 5.2: The domain of integration in Example 5.5.

*Solution.* The indefinite integral $J = \int \mathrm{e}^{-y^2}\,\mathrm{d}y$ *cannot* be written in terms of elementary functions. Thus, it seems that we are out of luck when attempting to calculate the integral $I$. However, the value of $I$ equals the value of the double integral

$$I = \iint\limits_R \mathrm{e}^{-y^2}\,\mathrm{d}A$$

over the region

$$R = \{(x, y) \in \mathbb{R}^2 \mid 0 \le x \le y \le 1\},$$

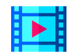
depicted in Fig. 5.2. Upon changing the order of integration—i.e., treating the region $R$ as *Type 2* and not as *Type 1*, we can actually evaluate the integral explicitly,

$$I = \int_0^1 \int_0^y \mathrm{e}^{-y^2}\,\mathrm{d}x\,\mathrm{d}y = \int_0^1 y\mathrm{e}^{-y^2}\,\mathrm{d}y = \left[-\tfrac{1}{2}\mathrm{e}^{-y^2}\right]_{y=0}^{y=1} = \tfrac{1}{2}\left(1 - \mathrm{e}^{-1}\right).$$

$\triangleleft$

*Remark* 5.5. Sometimes a region needs to be split up into two or more pieces to write a double integral as a sum of iterated integrals. This approach relies on the *additivity* of integration as discussed in Sec. 5.3.

**Example 5.6.** Let $f$ be a continuous function of two variables. Let $R$ be the region in the first quadrant bounded by the $x$-axis, the line $y = 4x/3$ and a circle of radius 5 centred at the origin (see Fig. 5.3). Express $\iint_R f \, \mathrm{d}A$ as an iterated integral or a sum of iterated integrals.



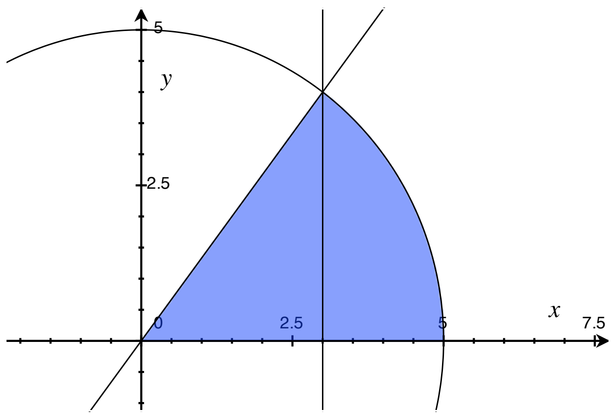



Figure 5.3: The region $R$ of Example 5.6.

V41: Additivity of double integrals

*Solution.* We could choose to integrate first over $y$ and then over $x$, i.e., consider the region to be of *Type 1*. Then the boundaries of the region $R$ are given by $y = 0$, $y = 4x/3$ and $y = \sqrt{25 - x^2}$. Consider how a vertical line at $x$ intersects with $R$. For $0 \leq x \leq 3$ the vertical line enters at $y = 0$ and leaves at $y = 4x/3$, but for $3 \leq x \leq 5$ it enters the region at $y = 0$ and leaves at $y = \sqrt{25 - x^2}$. The region can thus be expressed in the appropriate form as

$$R = \{(x, y) \in \mathbb{R}^2 \mid 0 \leq x \leq 5, \; 0 \leq y \leq \varphi_2(x)\}$$

where

$$\varphi_2(x) = \begin{cases} 4x/3 & 0 \leq x \leq 3 \\ \sqrt{25 - x^2} & 3 < x \leq 5. \end{cases}$$

This piece-wise formula for one of the limits of integration is awkward. Rather than writing a single iterated integral, it is better to write a sum of iterated integrals in which the limits of integration have simple expressions. This is equivalent to splitting $R$ into two pieces leading to

$$\iint\limits_R f \, \mathrm{d}A = \int_0^3 \int_0^{4x/3} f(x, y) \, \mathrm{d}y \, \mathrm{d}x + \int_3^5 \int_0^{\sqrt{25-x^2}} f(x, y) \, \mathrm{d}y \, \mathrm{d}x \, .$$

Alternatively we can integrate first over $x$ and then over $y$, i.e., thinking of the region $R$ as being of *Type 2*. The line $y = 4x/3$ becomes

$x = 3y/4$ and the curve $y = \sqrt{25 - x^2}$ becomes $x = \sqrt{25 - y^2}$. Hence the double integral can now be written as a single iterated integral,

$$\iint_R f \,\mathrm{d}A = \int_0^4 \int_{3y/4}^{\sqrt{25-y^2}} f(x,y) \,\mathrm{d}x \,\mathrm{d}y\,,$$

without the need split up the region. $\triangleleft$

Some double integrals have a simple interpretation with important applications in fields such as statistics or physics. The averages of the $x$- and $y$-coordinates of the points in $R$, for example, represent mean values of random variables or characterize the centre of mass of a body.

**Definition 5.1** (Centroid)**.** The *centroid* of a region $R \subset \mathbb{R}^2$ is the point $(\bar{x}, \bar{y})$ with coordinates

$$\bar{x} = \frac{1}{\mathcal{A}(R)} \iint_R x \,\mathrm{d}A\,, \qquad \text{and} \qquad \bar{y} = \frac{1}{\mathcal{A}(R)} \iint_R y \,\mathrm{d}A,$$

where $\mathcal{A}(R)$ is the area of the region.

**Example 5.7.** Find the centroid of the region bounded by one arch of a sine curve, i.e., for the points inside the region

$$R = \{(x,y) \in \mathbb{R}^2 \mid 0 \le x \le \pi,\ 0 \le y \le \sin x\}.$$

shown in Figure 5.4.

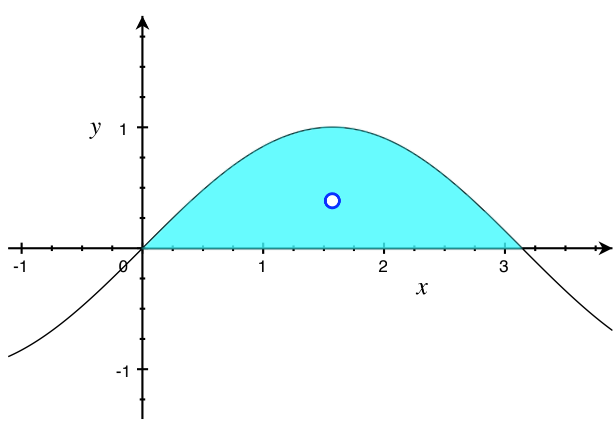



Figure 5.4: The region under an arch of $y = \sin x$ and the location of its centroid in Example 5.7.

*Solution.* The area of $R$ is the area under the curve $y = \sin x$, $0 \le x \le \pi$, which we can calculate as

$$\mathcal{A}(R) = \int_0^\pi \sin x \,\mathrm{d}x = \Big[ -\cos x \Big]_{x=0}^{x=\pi} = 2\,.$$

By symmetry, the $x$-coordinate of the centroid is $\bar{x} = \pi/2$. The $y$-coordinate of the centroid is given by two iterated integrals,

$$\bar{y} = \frac{1}{\mathcal{A}(R)} \iint\limits_R y \, \mathrm{d}A = \frac{1}{2} \int_0^\pi \left( \int_0^{\sin x} y \, \mathrm{d}y \right) \mathrm{d}x \,.$$

Using

$$\int_0^{\sin x} y \, \mathrm{d}y = \left[ \frac{y^2}{2} \right]_{y=0}^{y=\sin x} = \frac{\sin^2 x}{2} \,,$$

we find

$$\bar{y} = \frac{1}{4} \int_0^\pi \sin^2 x \, \mathrm{d}x = \frac{1}{8} \int_0^\pi (1 - \cos 2x) \, \mathrm{d}x = \frac{1}{8} \left[ x - \frac{\sin 2x}{2} \right]_{x=0}^{x=\pi} = \frac{\pi}{8} \,.$$

Thus, the centroid of $R$ is $(\pi/2, \pi/8)$. $\triangleleft$

## 5.5 *Change of variables in double integrals*

### *Polar coordinates (§15.4)*

Some integrals are easier to evaluate if, instead of the Cartesian coordinates, one uses polar coordinates $(r, \theta)$, where $x = r\cos\theta$ and $y = r\sin\theta$. To see how to write double integrals under such a change of variable we can go back to the definition of the double integral in terms of Riemann sums.

Instead of cutting $R$ by *vertical* lines (defined by constant $x$) and *horizontal* lines (defined by constant $y$), cut it up using by *radial* lines (defined by constant $\theta$) and circles (defined by constant $r$). Then we obtain

V42: Double integrals: polar coordinates

$$\iint\limits_R f \, \mathrm{d}A \approx \sum_{\substack{\text{little segments} \\ \text{covering region } R}} f(r_i \cos\theta_j, r_i \sin\theta_j) \Delta A_{ij} \tag{5.5}$$

where $\Delta A_{ij}$ is the area of the $(i, j)^{\text{th}}$ segment, and the point with polar coordinates $(r_i, \theta_j)$ lies in that segment. What form does the expression $\Delta A_{ij}$ take?

Let $\Delta r_i$ and $\Delta\theta_j$ be the coordinate widths of the $(i, j)^{\text{th}}$ segment. If $\Delta\theta_j$ is very small, then the segment is approximately rectangular, with side lengths $\Delta r_i$ and $r_i \, \Delta\theta_j$, therefore

$$\Delta A_{ij} = r_i \, \Delta r_i \, \Delta\theta_j$$

which we will need to use in the sum of Eq. (5.5). If the limit of the sum exist for arbitrarily small intervals $\Delta r_i$ and $\Delta\theta_j$, then it defines the double integral over the region $R$. Describing the region $R$ in terms of polar coordinates, i.e.,

$$S = \{(r, \theta) \in \mathbb{R}^2 \mid r \geq 0, \; 0 \leq \theta \leq 2\pi, \; (r\cos\theta, r\sin\theta) \in R\} \,,$$

we have found the desired alternative expression for the double integral as

$$\iint\limits_R f \, \mathrm{d}A = \iint\limits_S f(r\cos\theta, r\sin\theta)\, r \, \mathrm{d}r \, \mathrm{d}\theta \, .$$

The central insight is that, upon introducing polar coordinates, the area element $\mathrm{d}A = \mathrm{d}x\,\mathrm{d}y$ will take the form $\mathrm{d}A = r\,\mathrm{d}r\,\mathrm{d}\theta$.

One can see the need for an extra factor of $r$ by thinking about dimensions: $\mathrm{d}A$ is an area (dimensions length$^2$), $\theta$ is dimensionless, so $\mathrm{d}r\,\mathrm{d}\theta$ has dimensions length which cannot represent an area.

**Example 5.8.** Find the volume of the three-dimensional region $B$ given by

$$B = \{(x,y,z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 \le 1,\ y \ge 0,\ z \ge 0\}.$$

*Solution.* The projection of $B$ to the $(x, y)$-plane is the region

$$R := \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 \le 1,\ y \ge 0\} \, .$$

That is, all points inside the region, $(x, y, z) \in B$, lie above the upper half of the unit circle in the $(x, y)$-plane.

If $(x, y) \in B$, then for what values of $z$ is $(x, y, z) \in B$? Only non-negative values are allowed for $z$, i.e., $z \ge 0$. Therefore, the relation $x^2 + y^2 + z^2 \le 1$ implies that $z \le \sqrt{1 - x^2 - y^2}$, where we had to take the positive root. Consequently, $B$ is the solid above the region $R$ in the $(x, y)$-plane and under the graph of $\sqrt{1 - x^2 - y^2}$, and its volume can be calculated as

$$V = \iint\limits_R \sqrt{1 - x^2 - y^2} \, \mathrm{d}A \, .$$

In polar coordinates, the double integral is straightforward to set up and to calculate. Expressing the domain of integration in the new set of variables, we find $0 \le r \le 1$ and $0 \le \theta \le \pi$, so that, finally,

$$V = \int_0^\pi \int_0^1 \sqrt{1 - r^2}\, r \, \mathrm{d}r \, \mathrm{d}\theta = \int_0^\pi \left[ -\tfrac{1}{3}(1 - r^2)^{3/2} \right]_{r=0}^{r=1} \mathrm{d}\theta = \tfrac{1}{3}\pi \, .$$

The result is expected since a sphere has volume $\frac{4\pi}{3}$ and the region $B$ is one quarter of a unit sphere. ◁

**Example 5.9.** For $R = \{(x,y) \in \mathbb{R}^2 \mid 4 \le x^2 + y^2 \le 9,\ x \le 0\}$ find $\iint_R 3(x + y)\, \mathrm{d}A$.

*Proof.* Let us describe the region $R$ in polar coordinates. The condition $4 \le x^2 + y^2 \le 9$ is equivalent to $2 \le r \le 3$ and the condition $x \le 0$ means $\pi/2 \le \theta \le 3\pi/2$. In polar coordinates $x + y = r\cos\theta + r\sin\theta$, hence

A quick sketch of the region may make the conversion to polar coordinates easier.

$$\begin{aligned}
\iint\limits_R 3(x+y)\, \mathrm{d}A &= \int_{\pi/2}^{3\pi/2} \int_2^3 3(r\cos\theta + r\sin\theta) r \, \mathrm{d}r \, \mathrm{d}\theta \\
&= \left( \int_{\pi/2}^{3\pi/2} (\cos\theta + \sin\theta)\, \mathrm{d}\theta \right) \left( \int_2^3 3r^2 \, \mathrm{d}r \right) \\
&= [\sin\theta - \cos\theta]_{\pi/2}^{3\pi/2} \left[ r^3 \right]_2^3 \\
&= -2(27 - 8) = -38
\end{aligned}$$ ◁

Here is a classic—and quite beautiful—application of polar coordinates in a *double* integral to calculate the value of a definite *single* integral which we would not know how to evaluate directly.

**Example 5.10.** Compute the integral of the function $y(x) = \mathrm{e}^{-x^2}$, depicted in Fig. 5.5,

$$I = \int_{-\infty}^{\infty} \mathrm{e}^{-x^2}\,\mathrm{d}x.$$

V43: A double-integral detour

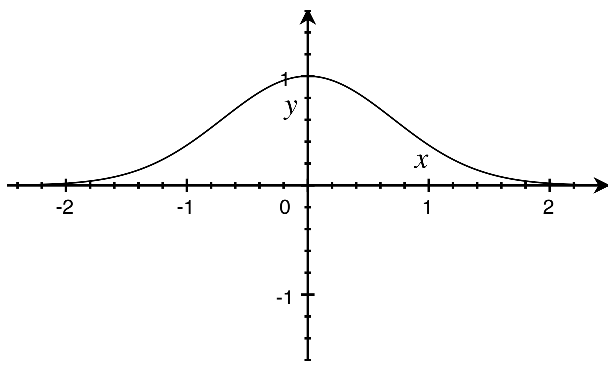



Figure 5.5: Although we cannot find an elementary function whose derivative is $\mathrm{e}^{-x^2}$, the area under the curve can be calculated to be $\sqrt{\pi}$.

*Solution.* There is no elementary function whose derivative is $\mathrm{e}^{-x^2}$, so we cannot evaluate this integral by the usual method of finding an indefinite integral.

Instead, let us start out with the product of the integral with itself which can be written as a double integral,

$$\begin{aligned} I^2 &= \left(\int_{-\infty}^{\infty} \mathrm{e}^{-x^2}\,\mathrm{d}x\right)\left(\int_{-\infty}^{\infty} \mathrm{e}^{-y^2}\,\mathrm{d}y\right) = \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} \mathrm{e}^{-x^2-y^2}\,\mathrm{d}y\right)\mathrm{d}x \\ &= \iint\limits_{\mathbb{R}^2} \mathrm{e}^{-\left(x^2+y^2\right)}\,\mathrm{d}A\,. \end{aligned}$$

The form of the exponent suggests introducing polar coordinates leading to two simple iterated integrals,

$$I^2 = \iint\limits_{\mathbb{R}^2} \mathrm{e}^{-r^2} r\,\mathrm{d}\theta\,\mathrm{d}r = \int_0^{\infty}\left(\int_0^{2\pi}\mathrm{d}\theta\right)\mathrm{e}^{-r^2} r\,\mathrm{d}r\,.$$

Calculating the integrals,

$$I^2 = 2\pi\left[-\tfrac{1}{2}\mathrm{e}^{-r^2}\right]_{r=0}^{r=\infty} = 2\pi\lim_{r\to\infty}\left(\tfrac{1}{2}-\tfrac{1}{2}\mathrm{e}^{-r^2}\right) = \pi\,,$$

we finally obtain the value of the desired integral, namely,

$$I = \int_{-\infty}^{\infty} \mathrm{e}^{-x^2}\,\mathrm{d}x = \sqrt{\pi}\,. \qquad \triangleleft$$

*General change of variables (§15.8)*

One way to introduce polar coordinates consists of establishing their relation to Cartesian coordinates, by writing $x = x(r, \theta)$ and $y = y(r, \theta)$. In general, to change old variables from $(x, y)$ to a new pair $(u, v)$, say, means to introduce a pair of equations $x = x(u, v)$ and $y = y(u, v)$. It must be possible to invert these relations everywhere—except possibly at some singular isolated point(s).

To write down a double integral in terms of the new variables, we will again go through its definition in terms of a suitable (double) Riemann sum. Once more, the domain of integration $R$ will be split up into small areas, using the lines of *constant* values of the new variables $u$ and $v$, respectively, in analogy to coordinate lines of constant $x$ and $y$, or constant $r$ and $\theta$.

What is the area enclosed by making small changes $\Delta u$ and $\Delta v$ of the new coordinates? For small $\Delta u$ and $\Delta v$ the area is approximately a parallelogram with sides of length $\Delta u$ and $\Delta v$ and we can use a linear approximation for the coordinate transformation by comparing the values of the old variables at points that differ by $\Delta u$ and $\Delta v$ of the new variables, i.e., we consider

V44: Double integrals: general coordinates

$$x(u_i + \Delta u, v_j + \Delta v) - x(u_i, v_j) \equiv \Delta x = \frac{\partial x}{\partial u}\Delta u + \frac{\partial x}{\partial v}\Delta v\,,$$

and a similar expression for $\Delta y$.

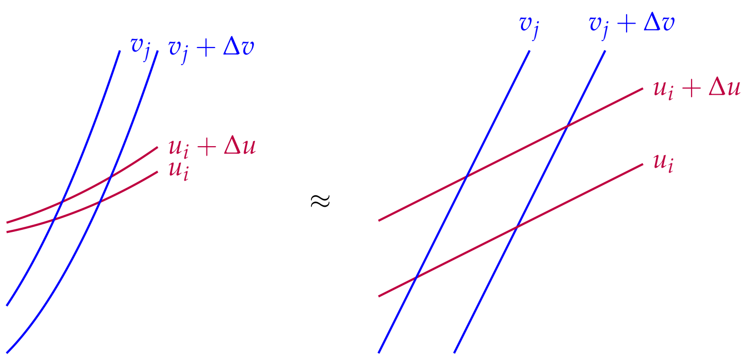



Figure 5.6: For small $\Delta u$ and $\Delta v$, the area contained by $u = u_i$, $u = u_i + \Delta u$, $v = v_j$ and $v = v_j + \Delta v$ is well approximated by a parallelogram. Note: this figure shows the lines drawn in the original $(x, y)$ coordinates.

Combining these equation into a single matrix equation, we find

$$\begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} \begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix}.$$

The matrix of partial derivatives contains all the relevant information about how changes $\Delta u$ and $\Delta v$ in the new variables are related to changes $\Delta x$ and $\Delta y$ in the old variables. As you will know from Algebra, the area of the parallelogram is the (absolute value of the) determinant of the transformation matrix. In other words, the area between the lines shown in Figure 5.6 is $\Delta A = |J|\, \Delta u\, \Delta v$, where the *Jacobian* $J$ of the coordinate transformation is defined by

On one of the exercises you are invited to confirm this for a specific example.

$$J = \det \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} .$$

If we consider a Riemann sum for the double integral in this way and use an ever finer grid, then the sum will turn into an iterated integral in terms of the variables $u$ and $v$, in perfect analogy to the specific cases of Cartesian and polar coordinates.

**Theorem 5.3** (Double integral)**.** *Let $x(u,v)$ and $y(u,v)$ be continuously differentiable functions, $f(x,y)$ a continuous function, and $R \subset \mathbb{R}^2$ a two-dimensional region in the $(x,y)$-plane. Given some region $S$ in the $(u,v)$-plane that maps 1-1 onto $R$, the following two double integrals are equal,*

$$\iint_R f(x,y)\, \mathrm{d}x\, \mathrm{d}y = \iint_S f[x(u,v), y(u,v)]\; |J(u,v)|\; \mathrm{d}u\, \mathrm{d}v\,.$$

*Remark* 5.6. The main result here can be summarized as the relation between the infinitesimal area elements, i.e., $\mathrm{d}A = \mathrm{d}x\, \mathrm{d}y = |J|\, \mathrm{d}u\, \mathrm{d}v$.

**Example 5.11.** Calculate the Jacobian of the transformation from polar coordinates to Cartesian coordinates.

*Solution.* Evaluating the partial derivatives in the Jacobian $J$, we find a positive determinant,

$$\det \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix} = \det \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} = r\cos^2\theta + r\sin^2\theta = r.$$

This accounts for the factor $r$ in the relation $\mathrm{d}A = r\, \mathrm{d}r\, \mathrm{d}\theta$. $\triangleleft$

**Example 5.12.** Calculate the integral

$$I = \iint_R y^3(2x-y)\mathrm{e}^{(2x-y)^2}\, \mathrm{d}A\,,$$

where $R$ is the region bounded by the lines $y = 2x$, $y = 2x - 4$, $y = 0$ and $y = 2$, illustrated in Fig. 5.7.

*Solution.* The integrand will simplify upon introducing new variables by defining

$$u = y \qquad \text{and} \qquad v = 2x - y\,.$$

Solving these two linear equations for $x$ and $y$, the old variables are given in terms of the new ones by

$$x = \frac{u+v}{2} \qquad \text{and} \qquad y = u\,.$$

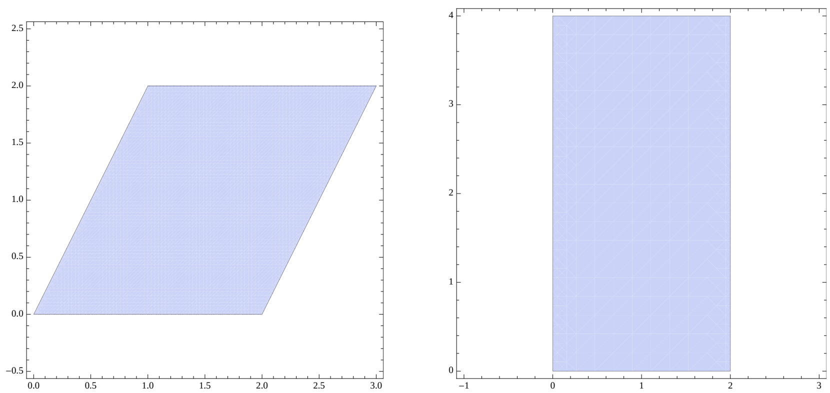



Figure 5.7: The domain of integration $R$ for Example 5.12, shown in the $(x, y)$-plane (left) and transformed to the $(u, v)$-plane (right).

Let us look at the boundaries of the region $R$ expressed in the variables $(u, v)$:

- the line $y = 2x$ corresponds to the line $u = 2(u + v)/2$, which is $v = 0$;
- the line $y = 2x - 4$ corresponds to the line $u = 2(u + v)/2 - 4$, so $v = 4$;
- the line $y = 0$ corresponds to the line $u = 0$;
- the line $y = 2$ corresponds to the line $u = 2$.

We have a rectangle in the $(u, v)$-plane. Since the Jacobian equals

$$J(u, v) = \det \begin{pmatrix} 1/2 & 1/2 \\ 1 & 0 \end{pmatrix} = -\frac{1}{2},$$

we calculate the integral $I$ in the following way,

$$\begin{aligned} I &= \int_0^2 \int_0^4 u^3 v e^{v^2} \left|\tfrac{-1}{2}\right| \mathrm{d}v\, \mathrm{d}u = \frac{1}{2} \left( \int_0^2 u^3\, \mathrm{d}u \right) \times \left( \int_0^4 v e^{v^2}\, \mathrm{d}v \right) \\ &= \frac{1}{2} \left[ \frac{1}{4} u^4 \right]_{u=0}^{u=2} \left[ \frac{1}{2} e^{v^2} \right]_{v=0}^{v=4} = e^{16} - 1. \end{aligned}$$

◁

**Example 5.13.** Evaluate the integral

$$I = \int_1^2 \int_{1/x}^{2/x} xy\, \mathrm{d}y\, \mathrm{d}x \tag{5.6}$$

in two ways: (i) by calculating the iterated integrals using the variables $(x, y)$; (ii) by introducing new variables $(u, v)$ via the relations $x = u$ and $y = \frac{v}{u}$.

*Solution.* (i) Doing the $y$-integral first, we directly obtain the integral as

$$I = \int_1^2 \left[ \tfrac{1}{2} x y^2 \right]_{y=1/x}^{y=2/x} \mathrm{d}x = \int_1^2 \left( \frac{2}{x} - \frac{1}{2x} \right) \mathrm{d}x = \left[ \tfrac{3}{2} \log x \right]_{x=1}^{x=2} = \tfrac{3}{2} \log(2)\,.$$

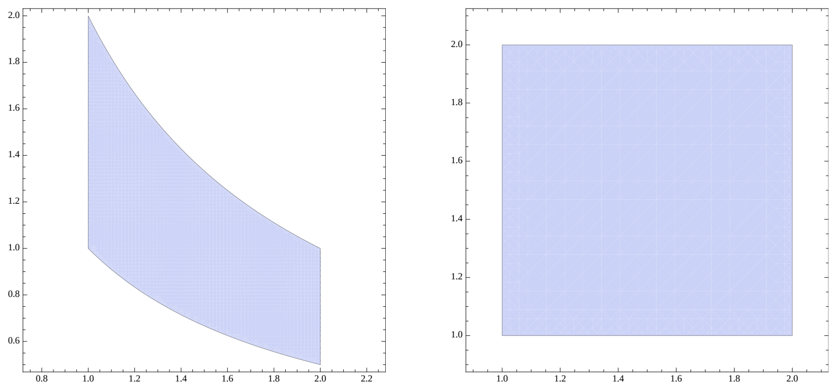



Figure 5.8: The domain of the integral (5.6) in the $(x, y)$-plane (left), and transformed to the $(u, v)$-plane (right).

V45: Hyperbolic boundaries

(ii) Now let us change variables by setting $x = u$ and $y = v/u$ and express the boundaries of the region $R$ in terms of the new variables:

- the curve $y = 1/x$ corresponds to $\frac{v}{u} = \frac{1}{u}$, so $v = 1$;
- similarly the curve $y = 2/x$ corresponds to $\frac{v}{u} = \frac{2}{u}$, so $v = 2$;
- the line $x = 1$ corresponds to $u = 1$;
- the line $x = 2$ corresponds to $u = 2$.

The resulting region is shown in Fig. 5.8. In this case, the Jacobian is not just a constant

$$J(u, v) = \det \begin{pmatrix} 1 & 0 \\ \frac{-v}{u^2} & \frac{1}{u} \end{pmatrix} = \frac{1}{u} .$$

Taking the modulus (this would be essential if the $u$-integration contained regions where $u < 0$) reproduces the result we obtained in (i),

$$\begin{aligned} I &= \int_1^2 \int_1^2 u \frac{v}{u} \left| \frac{1}{u} \right| \, \mathrm{d}u \, \mathrm{d}v = \left( \int_1^2 v \, \mathrm{d}v \right) \left( \int_1^2 \frac{1}{u} \, \mathrm{d}u \right) \\ &= \left[ \tfrac{1}{2} v^2 \right]_{v=1}^{v=2} \left[ \log u \right]_{u=1}^{u=2} = \tfrac{3}{2} \log 2 \, . \end{aligned}$$
$\triangleleft$

**Algorithm 9**
Change of Variables in a Double Integral

1. Sketch the original region of integration.
2. Choose new coordinates that either simplify the integrand or allow for a simpler description of the domain of integration.
3. Express the region of integration in terms of the new variables.
4. Calculate (the modulus of) the Jacobian and set up the transformed double integral.
5. Integrate the double integral in terms of the new variables.

# 6
# Extrema

## 6.1 Stationary points of functions (§14.7)

Extremal points of functions are particularly important in order to understand their overall behaviour. Given a function of a single variable $x$, say, the location of its local maxima and minima (plus its behaviour for $x \to \pm\infty$) we get a qualitative understanding of its structure. For similar reasons, it is useful to develop tools to identify and characterize the stationary points for functions of two variables. We will also investigate the properties of functions on subsets of their original domain of definition, the restrictions to subsets being implemented by *constraints* imposed on the initially independent variables. An example would be to calculate the highest point reached during a circular walk through a known landscape, which will typically *not* coincide with a mountain top.

### *Stationary points of functions of a single variable*

Let us begin by considering a particularly simple collection of functions of a single variable, namely all *quadratic* functions in $x$,

$$f(x) = a_0 + bx + \frac{c}{2}x^2\,, \qquad a_0, b, c \in \mathbb{R}\,, c \neq 0\,. \tag{6.1}$$

The parameter $c$ must not vanish; if it did, the function $f(x)$ would only be *linear* in $x$. As is well-known, the condition $f'(x) = 0$ identifies the value(s) of $x$ where $f(x)$ has a horizontal tangent, $x_0 = -b/c$. Defining $a = a_0 + bx_0/2$, we can write the function $f(x)$ as

$$f(x) = a + \frac{c}{2}(x - x_0)^2\,, \qquad a, c \in \mathbb{R}\,, c \neq 0\,. \tag{6.2}$$

The values of $x_0$ and $a$ simply correspond to shifts of the function (horizontally or vertically) and hence we can consider the case $a = 0$ and $x_0 = 0$ and remember that all other cases are just shifts of this. Thus we consider the simpler set of quadratic functions,

$$f(x) = \frac{c}{2}x^2, \qquad 0 \neq c \in \mathbb{R}\,, \tag{6.3}$$

with vanishing derivative at the origin and taking the value zero there.

The parabola (6.3) must either have a *minimum* or *maximum* at the point $x = 0$. The *sign* of the number $c$ tells us which of these two cases occurs. Thus, we are able to extract the relevant information by calculating the second derivative of the function $f(x)$ since it is proportional to the number $c$.

This approach to discuss points of a function with horizontal tangent can be generalized to arbitrary (smooth) functions $f : \mathbb{R} \to \mathbb{R}$, by recalling their Taylor approximation about a point $x_0$,

V46: Stationary points: single variable

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \tfrac{1}{2} f''(x_0)(x - x_0)^2 + R_2(x),$$

where the remainder $R_2(x)$ goes to 0 faster than $(x - x_0)^2$ as $x \to x_0$. If $x_0$ is a stationary point, $f'(x_0) = 0$, and we obtain

$$f(x) = f(x_0) + \tfrac{1}{2} f''(x_0)(x - x_0)^2 + R_2(x)\,.$$

If $f''(x_0) \neq 0$, then for $x$ close enough to $x_0$, the quadratic term will dominate the remainder term. This quadratic approximation of $f(x)$ near $x_0$ is of the form given in Eq. (6.2), with $c = f''(x_0)$, and $a = f(x_0)$. Again, by selecting specific values of these two parameters, we could bring (the quadratic part of) $f(x)$ into the form shown in Eq. (6.3). Now we are able to extract the behaviour of the function $f(x)$ near the extremal point $x_0$:

1. If $f''(x_0) > 0$, then for all values of $x$ in a sufficiently small interval about $x_0$, we have $f(x) > f(x_0)$, meaning that $f$ has a local *minimum* at $x_0$;

2. If $f''(x_0) < 0$, then for all values of $x$ in a sufficiently small interval about $x_0$, we have $f(x) < f(x_0)$, meaning that $f$ has a local *maximum* at $x_0$.

This result is known as the "second-derivative test".

### *Stationary points of functions of two variables*

To extend the discussion above to functions of two variables, we first need to generalize the idea of (local) maxima and minima in a suitable way. We will define the stationary points of a function $f(x, y)$ and then aim to determine their nature, proceeding in way similar to the single-variable case. We will first discuss the most general *quadratic* function of two variables and then generalize what we found to arbitrary (smooth) functions. The resulting second-derivative test can be justified by considering the *Taylor expansion* of a two-variable function near its stationary points.

**Definition 6.1** (Local minimum)**.** A function $f : \mathbb{R}^2 \to \mathbb{R}$ is said to have a *(local) minimum* at the point $(x_0, y_0)$ if $f(x_0, y_0) \leq f(x, y)$ for all $(x, y)$ in some disc centred at $(x_0, y_0)$.

The definition of a local *maximum* only differs in the sense of the inequality which holds for the values of the function $f(x, y)$ inside the disc around the point $(x_0, y_0)$.

**Definition 6.2** (Local maximum)**.** A function $f : \mathbb{R}^2 \to \mathbb{R}$ is said to have a *(local) maximum* at the point $(x_0, y_0)$ if $f(x_0, y_0) \geq f(x, y)$ for all $(x, y)$ in some disc centred at $(x_0, y_0)$.

V47: Stationary points: two variables

Let us now formalize the observation about the slopes of tangents to a (smooth) function at an extremum, i.e., either a (local) *minimum* or *maximum*. If $\frac{\partial f}{\partial x} \neq 0$ at $(x_0, y_0)$, then $f(x, y)$ increases or decreases at $(x_0, y_0)$ in the $x$-direction, and if $\frac{\partial f}{\partial y} \neq 0$ then $f(x, y)$ changes at $(x_0, y_0)$ in the $y$-direction. In contrast, the tangents to a (smooth) function are *horizontal* at a "mountain top" or at the lowest point of a depression. Assuming that the partial derivatives of $f(x, y)$ with respect to $x$ and $y$ exist, they must *vanish* at an extremum $(x_0, y_0)$,

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0 \,.$$

We can combine this pair of equations into a single one using the nabla operator, $\nabla f(x_0, y_0) = 0$, leading to the following definition.

The "0" in this equation is short for the zero vector.

**Definition 6.3** (Stationary points)**.** If $\nabla f(x_0, y_0) = 0$, then $(x_0, y_0)$ is called a *stationary point*.

**Example 6.1.** Find the stationary points of $f(x, y) = x^2 + y^2$.

*Solution.* $\frac{\partial f}{\partial x} = 2x$, $\frac{\partial f}{\partial y} = 2y$. So $\nabla f(0, 0) = 0$ and $(0, 0)$ is the only stationary point. Since $f(x, y) \geq 0$ and $f(0, 0) = 0$, the function $f$ has a minimum at $(0, 0)$. $\triangleleft$

Interestingly, there is a *third type* of stationary point that is neither a minimum nor a maximum.

**Definition 6.4** (Saddle point)**.** A stationary point where $f$ does not have an extremum is called a *saddle point*.

Saddle points cannot occur for functions of a single variable. There is, however, some similarity with points of inflection that have a horizontal tangent line but do not represent an extremum.

**Example 6.2.** Find the stationary points of $f(x,y) = x^2 - y^2$.

*Solution.* $\nabla f(0,0) = (0,0)$, so $(0,0)$ is a stationary point. But $f$ does not have an extremum at $(0,0)$, so it is a saddle point. $\triangleleft$

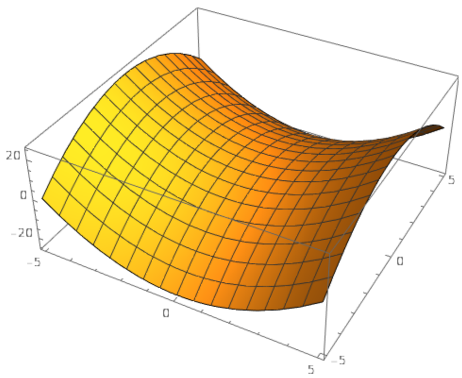



Figure 6.1: Sketch of a saddle point.

*Remark* 6.1. A function $f(x,y)$ can have a local maximum or minimum at $(x_0, y_0)$ without being differentiable at that point. In this case, the gradient $\nabla f(x_0, y_0)$ of the function does not exist at that point, and the extremum cannot be detected in this way as a stationary point.

The same is true for functions of one variable, e.g., $f(x) = |x|$ has a minimum at $x = 0$ but is not differentiable there.

## 6.2 *Classifying stationary points (§§14.7, 14.9)*

We will now classify the stationary points of a function $f(x,y)$ of two variables. The first step is to discuss the properties of the most general expression *quadratic* in the variables $x$ and $y$, just as we did in the case of a single variable.

### *The quadratic case*

V48: Stationary points: quadratic functions

The most general quadratic function of two variables $x$ and $y$ is given by

$$f(x,y) = a_0 + Kx + Ly + \frac{A}{2}x^2 + Bxy + \frac{C}{2}y^2, \tag{6.4}$$

where $a, A, B, C, K, L \in \mathbb{R}$ are constants; we exclude the case that $A = B = C = 0$ since the resulting expression would be linear in $x$ and $y$ only. Assume that there is a point $(x_0, y_0)$ where $f(x,y)$ has horizontal tangents characterized by $f_x = f_y = 0$. Then we have an alternative form for $f(x,y)$, namely

$$f(x,y) = a + \tfrac{1}{2}A(x-x_0)^2 + B(x-x_0)(y-y_0) + \tfrac{1}{2}C(y-y_0)^2, \tag{6.5}$$

in analogy to Eq. (6.2).

The values of the constants $a$, $x_0$ and $y_0$ do not affect the nature of the stationary point, they merely shift its location and height, so it will be sufficient to consider the case where $a = x_0 = y_0 = 0$.

Therefore, we consider the simpler set of quadratic polynomials given by

$$f(x,y) = \tfrac{1}{2}Ax^2 + Bxy + \tfrac{1}{2}Cy^2, \tag{6.6}$$

Notation alert: we already used the symbol $\Delta$ to denote the *Laplacian* as well as in $\Delta x$ etc. to denote *lengths* of small intervals.

which has a stationary point at $(0,0)$. Defining the quantity

$$\Delta = AC - B^2\,,$$

we can "complete the square" in order to write $f(x,y)$ as a sum over quadratic terms,

$$2Af(x,y) = A^2x^2 + 2ABxy + ACy^2 \tag{6.7}$$

$$= (Ax+By)^2 + (AC-B^2)y^2 = (Ax+By)^2 + y^2\Delta. \tag{6.8}$$

This form allows an elementary discussion of the behaviour of $f(x,y)$ at the origin where the stationary point is located. Three cases need to be considered.

$\underline{\Delta > 0:}$ The parameter $A$ must be non-zero. Hence, for points $(x,y) \neq (0,0)$, we have from (6.8) that $2Af(x,y) > 0$. There are now two possibilities:

- If $A > 0$, then $f(x,y) > 0 = f(0,0)$ holds for all $(x,y) \neq (0,0)$. Therefore $f$ has a minimum at $(0,0)$ (cf. Definition 6.1).
- If $A < 0$, then $f(x,y) < 0 = f(0,0)$ holds for all $(x,y) \neq (0,0)$. Therefore $f$ has a maximum at $(0,0)$ (cf. Definition 6.2).

  In both cases, the graph of $f$ is a *paraboloid* (see Fig. 6.2).

$\underline{\Delta < 0:}$ Now we distinguish two cases, corresponding to the parameter $A$ being equal to zero or not.

- If $A = 0$, Eq. (6.6) tells us that

$$f(x,y) = Bxy + \tfrac{1}{2}Cy^2 = \left(Bx + \tfrac{1}{2}Cy\right)y\,.$$

- If $A \neq 0$, then Eq. (6.7) can be written in the form

$$f(x,y) = \tfrac{1}{2A}(Ax + [B + \sqrt{-\Delta}]y)(Ax + [B - \sqrt{-\Delta}]y)\,,$$

  as can be verified by multiplying out the product.

  In both cases, $f(x,y)$ vanishes along *two straight lines* through the origin. These lines divide the $(x,y)$-plane into four quadrants where $f(x,y)$ takes alternating signs. Therefore, the origin $(0,0)$ is neither a minimum nor a maximum but a *saddle point* of $f$. The graph of $f$ is called a *hyperbolic paraboloid* (the graph in Fig. 6.1 is an example).

$\underline{\Delta = 0:}$ Finally, suppose that $\Delta = 0$.

- If $A \neq 0$, then the expression in Eq. (6.8) reduces to

$$f(x,y) = \tfrac{1}{2A}(Ax+By)^2\,,$$

  whatever the value of $B$.

- If $A = 0$, then $B$ must vanish as well and Eq. (6.6) reduces to

$$f(x,y) = \tfrac{1}{2}Cy^2\,, \qquad C \neq 0\,.$$

  In both cases $f$ vanishes along *one straight line* of stationary points. If $A > 0$ or both $A = 0$ and $C > 0$, then $f(x,y) \geq 0$, and all stationary points are minima; for $A < 0$ or both $A = 0$ and $C < 0$, we have $f(x,y) \leq 0$ leading to a line of stationary points that are maxima. The graph of $f$ is a called a *parabolic cylinder*.

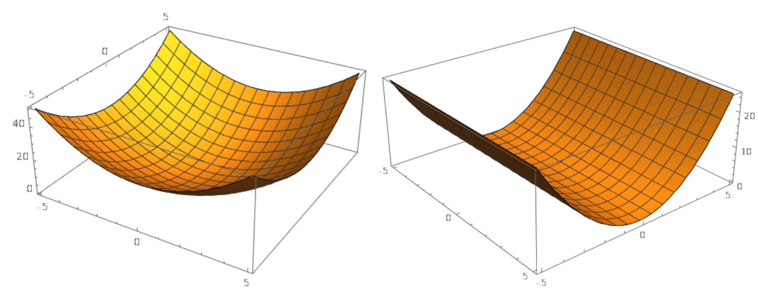


Figure 6.2: A parabaloid (left) and cylindrical parabaloid (right).

These conclusions also apply to the general quadratic polynomial (6.5).

### *The general case – second-derivative test*

In analogy to the single variable-case, the results found in for quadratic polynomials can be extended to more general functions. We first summarize the findings in terms of the *second-derivative test* and relegate their justification, which is based on the Taylor expansions for functions of two variables near a stationary point, to the next section.

To formulate the theorem which classifies the stationary points of functions $f(x,y)$, we introduce a specific combination of its second-order derivatives.

**Definition 6.5** (Hessian (matrix))**.** The *Hessian* of a continuously twice-differentiable function $f(x,y)$ at the point $(x,y)$ is

$$H(x,y) := \begin{pmatrix} f_{xx}(x,y) & f_{xy}(x,y) \\ f_{xy}(x,y) & f_{yy}(x,y) \end{pmatrix}.$$

The *discriminant* $\Delta$ of a stationary point at $(x_0,y_0)$ of $f(x,y)$ is the determinant of the Hessian, i.e., $\Delta(x_0,y_0) = \det H(x_0,y_0)$.

This quantity allows one to conveniently characterize the type of a stationary point as we have seen already in the case of quadratic functions. The discussion was based on the value of $\Delta = AC - B^2$, which is the determinant of the Hessian for the quadratic function given in Eq. (6.5)

$$H(x_0, y_0) = \begin{pmatrix} A & B \\ B & C \end{pmatrix}.$$

V49: Hessian and 2nd-derivative test

Now we can spell out the second-derivative test.

**Theorem 6.1** (Second-derivative Test). *Let $f(x, y)$ be a continuously twice-differentiable function, and assume that is has a stationary point $(x_0, y_0)$ with discriminant $\Delta = \Delta(x_0, y_0)$. If*

- $\Delta > 0$ *and* $f_{xx}(x_0, y_0) > 0$, *then* $(x_0, y_0)$ *is a local minimum;*
- $\Delta > 0$ *and* $f_{xx}(x_0, y_0) < 0$, *then* $(x_0, y_0)$ *is a local maximum;*
- $\Delta < 0$, *then* $(x_0, y_0)$ *is a saddle point.*

If discriminant vanishes at the stationary point, $\Delta = 0$, then we cannot conclude anything from the second derivatives. For example, the functions $f(x, y) = x^2 + y^4$ and $g(x, y) = x^2 - y^4$ both have a unique stationary point at $(0, 0)$ and exactly the same second derivatives there, but $f$ has a minimum there while $g$ has a saddle point.

*Remark* 6.2. The generalization of Theorem 6.1 to more than two variables is not straightforward.

**Example 6.3.** Find all stationary points of the function $f(x, y) = \frac{1}{3}(x + y)^3 - x^2 - y^2$ and determine whether $f$ has a local maximum, local minimum or neither at each stationary point.

*Solution.* If $(x, y)$ is a stationary point then

$$0 = \frac{\partial f}{\partial x} = (x + y)^2 - 2x\,, \tag{6.9a}$$

$$0 = \frac{\partial f}{\partial y} = (x + y)^2 - 2y\,. \tag{6.9b}$$

By subtracting Eq. (6.9b) from Eq. (6.9a) we find $y = x$. Substituting this relation back into Eq. (6.9a), we find $4x^2 - 2x = 2x(2x - 1) = 0$. Therefore, we must have $x = 0$ or $x = \frac{1}{2}$, resulting in the stationary points of $f(x,y)$ located at $(x, y) = (0, 0)$ and $(x, y) = (\frac{1}{2}, \frac{1}{2})$, respectively. The second derivatives are given by

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial^2 f}{\partial y^2} = 2(x + y) - 2 \qquad \text{and} \quad \frac{\partial^2 f}{\partial x \partial y} = 2(x + y)\,,$$

which produces the discriminants

$$\Delta(0, 0) = (-2)^2 - 0^2 = 4 > 0$$

and

$$\Delta(\tfrac{1}{2}, \tfrac{1}{2}) = 0^2 - 2^2 = -4 < 0\,.$$

Since $\Delta(0,0) > 0$ and $f_{xx}(0,0) = -2 < 0$, according to Theorem 6.1 the function $f$ has a local *maximum* at $(0,0)$. The second inequality implies that $(\frac{1}{2}, \frac{1}{2})$ is a *saddle point* of $f$. $\lhd$

### *Justifying the second-derivative test*

To investigate the character of a stationary point, we will make use of the fact that a function $f(x,y)$ near some point $(x_0, y_0)$ can be written in the form

$$\begin{aligned} f(x,y) =& f(x_0,y_0) + f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0) + \\ &+ \tfrac{1}{2} f_{xx}(x_0,y_0)(x-x_0)^2 + \tfrac{1}{2} f_{yy}(x_0,y_0)(y-y_0)^2 + \\ &+ f_{xy}(x_0,y_0)(x-x_0)(y-y_0) + R_2(x,y)\,. \end{aligned} \tag{6.10}$$

It is assumed that $f(x,y)$ has continuous second partial derivatives ensuring $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$ (via Theorem 4.1). The term $R_2(x,y)$ is known as the remainder term and goes to zero faster than $(x-x_0)^2 + (y-y_0)^2$ when we approach the point $(x_0, y_0)$. For the quadratic polynomial given in Eq. (6.4) the remainder vanishes exactly.

*Remark* 6.3. Note that what matters here are the *combined* power of $(x-x_0)$ and $(y-y_0)$. For example, there is no $(x-x_0)^2(y-y_0)$ term because it will vanish faster than any quadratic term upon approaching the expansion point.

If $(x_0, y_0)$ is a *stationary* point of $f(x,y)$, then the terms linear in $(x-x_0)$ and $(y-y_0)$ equal zero (6.10) leaving us with

$$\begin{aligned} f(x,y) =& f(x_0,y_0) + f_{xy}(x_0,y_0)(x-x_0)(y-y_0) + \\ &+ \tfrac{1}{2} f_{xx}(x_0,y_0)(x-x_0)^2 + \tfrac{1}{2} f_{yy}(x_0,y_0)(y-y_0)^2 + R_2(x,y)\,. \end{aligned} \tag{6.11}$$

Now we can carry over the reasoning of the quadratic function considered previously. Whenever $\Delta > 0$, the quadratic part of (6.11) is nonzero for $(x,y) \neq (x_0,y_0)$, so the remainder, which can be made arbitrarily small, cannot change the character of the stationary point at $(x_0, y_0)$, be it a maximum or a minimum. Similarly, if $\Delta < 0$, then $f$ is still both greater and smaller than $f(x_0, y_0)$ in any neighbourhood of $(x_0, y_0)$ which therefore remains a saddle point. Altogether, the remainder term is (usually) irrelevant to the question of whether the stationary point is a maximum, minimum, or saddle point.

There is an alternative way to think of this, noting that $f(x,y) - f(x_0,y_0) = \frac{1}{2}(x-x_0 \;\; y-y_0) H(x_0,y_0) \begin{pmatrix} x-x_0 \\ y-y_0 \end{pmatrix} + R_2(x,y)$. The condition that both $\Delta > 0$ and $A > 0$ is equivalent to both eigenvalues of $H(x_0,y_0)$ being positive, which ensures positivity of $f(x,y) - f(x_0,y_0)$ for $(x,y)$ close to $(x_0,y_0)$. Similarly, $\Delta > 0$ and $A < 0$ is equivalent to both eigenvalues of $H(x_0,y_0)$ being negative. The condition $\Delta < 0$ means the eigenvalues of $H(x_0,y_0)$ have different signs, so some small changes lead to positive $f(x,y) - f(x_0,y_0)$ and others negative.

## 6.3 *Extrema with a constraint (§14.8)*

Instead of extrema of a function $f(x,y)$ over the plane $\mathbb{R}^2$, we now consider extrema of $f$ along a smooth curve $\mathcal{C}$ in $\mathbb{R}^2$. To do so, we first express $\mathcal{C}$ as a level curve of another (continuously differentiable) function, i.e., for some fixed $c \in \mathbb{R}$,

$$\mathcal{C} = \{(x, y) \in \mathbb{R}^2 \mid g(x, y) = c\}\,.$$

For $f$ to have an extremal value along $\mathcal{C}$ at $(x_0, y_0) \in \mathcal{C}$ it is not necessary for $\nabla f(x_0, y_0)$ to vanish. However, the directional derivative of $f$ *in the direction tangent to* $\mathcal{C}$ at $(x_0, y_0)$ must vanish motivating the definition of a *constrained* stationary point.

V50: Constrained stationary points

**Definition 6.6** (Constrained stationary point)**.** A point $(x_0, y_0) \in \mathcal{C}$ *is called a constrained stationary point* of $f$ if

$$D_{\underline{u}} f(x_0, y_0) = \underline{u} \cdot \nabla f(x_0, y_0) = 0$$

holds for all vectors $\underline{u}$ tangent to $\mathcal{C}$ at $(x_0, y_0)$.

Assume that the gradient of $g(x, y)$ never vanishes along the curve $\mathcal{C}$, i.e., $\nabla g(x, y) \neq 0$ on $\mathcal{C}$. Since the gradient vector $\nabla g$ is normal to the level curve of $g$,

$$\nabla g(x_0, y_0) \perp \underline{u}\,,$$

and the vector $\underline{u}$ is normal to the gradient of $f$, we see that the gradients of $f$ and $g$ are parallel,

$$\nabla f(x_0, y_0) = \lambda \nabla g(x_0, y_0)\,, \qquad \text{for some } \lambda \in \mathbb{R}\,.$$

Thus, for $\nabla g(x, y) \neq 0$ everywhere on the curve, $(x_0, y_0)$ is a stationary point of $f$ subject to the constraint $g(x, y) = c$ if and only if $\nabla f(x_0, y_0) = \lambda \nabla g(x_0, y_0)$ for some $\lambda$. Consequently, the three unknown quantities $x$, $y$ and $\lambda$ must satisfy three simultaneous conditions in order that we have a constrained stationary point,

$$\frac{\partial f}{\partial x} = \lambda \frac{\partial g}{\partial x}\,, \qquad \frac{\partial f}{\partial y} = \lambda \frac{\partial g}{\partial y}\,, \qquad g(x, y) = c\,. \tag{6.12}$$

The parameter $\lambda$ is called a *Lagrange multiplier*. If an extreme value occurs, it must occur at one of the constrained stationary points found by this method, provided that $\nabla g(x, y) \neq 0$ on $\mathcal{C}$.

*Remark* 6.4. If we define the three-variable function,

$$F(x, y; \lambda) = f(x, y) - \lambda(g(x, y) - c)\,,$$

then the equations in (6.12) can be written as just $\nabla F = (0, 0, 0)$.

**Example 6.4.** What are the maximum and minimum values of $f(x, y) = 7x + 5y$ subject to the constraint $g(x, y) = x^2 + xy + y^2 = 13$?

*Solution.* The constraining curve is an ellipse shown in Fig. 6.3, along with some level curves of $f(x, y)$. It is not difficult to check that $\nabla g(x, y) \neq (0, 0)$ everywhere on the constraint curve.

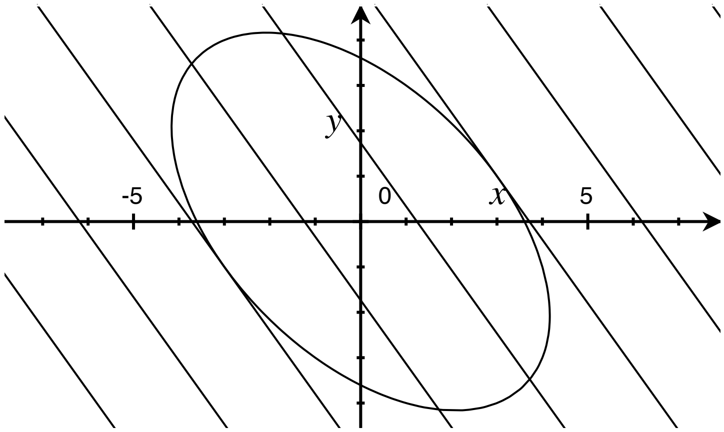



Figure 6.3: The constraint curve $x^2 + xy + y^2 = 13$ and some of the level curves of $f(x, y) = 7x + 5y$.

The equations characterizing constrained stationary points in this problem read

$$7 = \lambda(2x + y)\,, \qquad 5 = \lambda(x + 2y)\,, \qquad x^2 + xy + y^2 = 13\,,$$

implying immediately that $\lambda \neq 0$. Hence the first two equations can be written as a pair of coupled linear equations for the variables $x$ and $y$,

$$2x + y = \frac{7}{\lambda}\,, \qquad x + 2y = \frac{5}{\lambda}\,,$$

with solutions $x = 3/\lambda$ and $y = 1/\lambda$. Using these values in the third equation, we find $\lambda = \pm 1$, so that the stationary points are located at $(x_\pm, y_\pm) = \pm(3, 1)$. The values of $f(x, y)$ at these constrained extrema are $f(3, 1) = 26$ (a maximum) and $f(-3, -1) = -26$ (a minimum), respectively. $\lhd$

The method of Lagrange multipliers is not limited to the case of two variables and just one constraint.

**Example 6.5.** Find the maximum and the minimum of the function $f(x, y, z) = x + yz$ on the sphere of radius $r = 3$, centred at the origin.

*Solution.* The sphere is defined by $g(x, y, z) = x^2 + y^2 + z^2 = 9$. Using a Lagrange multiplier $\lambda$ and the four-variable function

V51: Lagrange multiplier in action

$$F(x, y, z; \lambda) = x + yz + (g(x, y, z) - 9)\ ,$$

we need to solve $\nabla F = 0$, where the four-component nabla operator is given by $\nabla = (\partial_x, \partial_y, \partial_z, \partial_\lambda)$. Explicitly, the first three equations read

$$1 = 2\lambda x\,, \qquad z = 2\lambda y\,, \qquad y = 2\lambda z\,, \tag{6.13}$$

and the last one reproduces the constraint, $x^2 + y^2 + z^2 = 9$. Solving the first equation for $\lambda$ and eliminating it from the other two in (6.13) leads to the relations $y = xz$ and $z = xy$. Eliminating $z$ from this pair leaves us with $y = x^2y$ which implies that $y = 0$ or $x^2 = 1$.

If $y = 0$, then $z = xy$ implies that $z$ must be equal to zero as well. Then the quadratic constraint reduces to $x^2 = 9$, with solutions $x = \pm 3$. Therefore, two constrained stationary points exist at $(\pm 3, 0, 0)$.

If $x^2 = 1$, then the quadratic constraint becomes

$$9 = x^2 + y^2 + z^2 = 1 + y^2 + (xy)^2 = 1 + 2y^2 ,$$

so $y^2 = 4$. In particular, if $x = 1$, then $y = \pm 2$ and $z = y$, so $(1, \pm 2, \pm 2)$ are two constrained stationary points. If $x = -1$, then again $y = \pm 2$ but $z = \mp 2$, resulting in another two constrained stationary points located at $(-1, \pm 2, \mp 2)$.

The values of $f$ at these six points are $f(\pm 3, 0, 0) = \pm 3$, $f(1, \pm 2, \pm 2) = 5$ and $f(-1, \pm 2, \mp 2) = -5$. Consequently, the largest value that the function $f(x, y, z)$ takes on the sphere is 5, and the smallest value is $-5$; each of these values is attained at *two* points. $\triangleleft$

## 6.4 *Extrema in closed regions*

The method of Lagrange multipliers can also be used to determine extrema of functions if they are defined on a subset of the $(x, y)$-plane, for example. We will illustrate the approach using a couple of examples.

**Example 6.6.** Find the minimum and maximum of the function $f(x, y) = \mathrm{e}^{-x^2-y^2}$ on the elliptical region $\mathcal{E}$ given by $x^2 + 2y^2 \leq 1$.

*Solution.* The extrema of $f$ must occur at stationary points in the *interior* of $\mathcal{E}$ where $x^2 + 2y^2 < 1$, or at constrained stationary points on the *boundary* of $\mathcal{E}$ where $x^2 + 2y^2 = 1$.

Let us first determine the location of the stationary points of $f(x, y)$ using the condition $\nabla f = 0$, or

$$\nabla f(x, y) = (-2x\mathrm{e}^{-x^2-y^2}, -2y\mathrm{e}^{-x^2-y^2}) = (0, 0) .$$

The only solutions are $x = y = 0$, identifying a single stationary point $(0, 0)$ which is located in the interior of the region $\mathcal{E}$.

To find any constrained stationary points, we use a Lagrange multiplier as we did in the previous two examples. Let $g(x, y) := x^2 + 2y^2$. We need to solve $\nabla f = \lambda \nabla g$ and $g(x, y) = 1$. The first two equations take the form

$$-2x\mathrm{e}^{-x^2-y^2} = 2x\lambda \qquad \text{and} \qquad -2y\mathrm{e}^{-x^2-y^2} = 4y\lambda .$$

The first equation has two solutions: we must either have $x = 0$ or $\lambda = -\mathrm{e}^{-x^2-y^2}$; similarly, the solutions of the second equation are $y = 0$ or $\lambda = -\frac{1}{2}\mathrm{e}^{-x^2-y^2}$. If $x = 0$, then the constraint $g = 1$ implies that

$y = \pm\frac{1}{\sqrt{2}}$. The other option, $\lambda = -e^{-x^2-y^2}$, forces $y = 0$ resulting in $x = \pm 1$, again using the constraint.

Altogether, there are thus four constrained stationary points which occur on the boundary of the region $\mathcal{E}$, namely $(0, \pm\frac{1}{\sqrt{2}})$ and $(\pm 1, 0)$. The values of $f$ at these points are

$$f(0, \pm\tfrac{1}{\sqrt{2}}) = e^{-1/2} \qquad \text{and} \qquad f(\pm 1, 0) = e^{-1}.$$

Now we are in a position to answer the original question: the value of $f$ at the origin is larger than its values at the stationary points on the boundary of $\mathcal{E}$, making $(0,0)$ the *maximum* of $f$ over $\mathcal{E}$. The smallest value of $f$ occurs (twice) on the boundary of the region $\mathcal{E}$, at the points $(\pm 1, 0)$, being equal to $e^{-1}$, corresponding to the *minima* of $f$ on $\mathcal{E}$. $\triangleleft$

**Example 6.7.** Find the maximum value of $f(x,y) = 5x - y$ in the region $R = \{(x,y) \in \mathbb{R}^2 \mid x^2 \leq y \leq 25,\ 0 \leq x \leq 5\}$, and the location where it occurs.

*Solution.* The function is linear, so has no unconstrained extrema. The maximum value hence occurs on the boundary of the region. There are three boundary curves and we deal with each separately:

- On $y = 25$, $0 \leq x \leq 5$ the function takes values $f(x,5) = 5x - 25$ so takes a maximum of 0 at $(5, 25)$.
- On $x = 0$, $0 \leq y \leq 25$ we have $f(0,y) = -y$ which is maximized at $y = 0$ with a value of 0.
- On $y = x^2$, $0 \leq x \leq 5$ we can use Lagrange multipliers to find the constrained optima. Taking $g(x,y) = x^2 - y$ we want $g = 0$ and $\nabla f = \lambda \nabla g$, which gives $5 = 2\lambda x$ and $-1 = -\lambda$. We hence have a stationary point when $\lambda = 1$ and $x = 5/2$. This optimum takes the value $25/2 - 25/4 = 25/4$ and must be a maximum on with this constraint because we have already checked the endpoints of this curve when checking the two other boundary curves.

Comparing the optima, we see that the maximum of $f(x,y)$ in $R$ is $25/4$ which occurs at $(5/2, 25/4)$. $\triangleleft$