Version January 25, 2023

Mathematical Methods of Physics I & II

Adrian Signer

University of Zurich: HS 2022 and FS 2023

This work is licensed under a
Creative Commons Attribution 4.0 International License.

Preface

These notes are based on the course of MMP I & II as taught at UZH in HS 2015/20 and FS 2016/21 and (slightly rearranged) for HS 2022 and FS 2023. They have many gaps (sometimes indicated by the label $\Rightarrow$ lecture) that might be filled during the lectures and/or exercise classes. The notes should be seen as a very detailed table of contents. Thus, they are not a replacement for the lectures, exercise classes, books, math.stackexchange.com or YouTube videos, but they should provide you with a guide through the jungle of MMP. It is a fairly ambitious plan we're following, so a guide might come in handy. Ideally, you read the couple of pages of what is to come before the lecture takes place to have an idea of where we're going.

As you might have discovered by now, the notes are in English. For important (possibly) new terms the German translation is given next to the English term. This is done even for blatantly obvious translations, in order to make it easy to spot the *important terms [wichtige Begriffe]* in the text.

MMP I is covered in Sections 1 – 7 and its mantra is "just calculate". This does not mean we will not prove important results or understand what we're doing. But we will only glimpse at formal structures and funny coincidences. Every now and then a miracle occurs, we will note it but then just shrug our shoulders and carry on. By the end of MMP I you should be reasonably familiar with complex analysis, be able to do contour integrals, Fourier series and Fourier expansions and solve virtually all ordinary and partial differential equations you will encounter in standard undergraduate physics lectures. So you're well equipped to "just do physics".

In MMP II, i.e. Sections 9 – 12, we are somewhat more formal and we will do a bit of justice to the first 'M' of MMP. Some of the 'things' we've done in MMP I will be put on more solid footing and some of the miracles will be understood. We will introduce metric spaces, normed spaces and finally Hilbert spaces, with a particular emphasis on infinite dimensional Hilbert spaces. In addition we will spend quite some time on group theory, a topic that shows up in many different places in physics. The hope is that MMP II provides the key to a deeper understanding of many aspects in physics. And who knows, maybe it is even a bit of fun.

Contents

1 Complex analysis **1**

1.1 Complex numbers and functions . . . 1
1.2 Analytic functions . . . 3
1.3 Contour integrals . . . 7
1.4 Cauchy's theorem and integral formula . . . 9
1.5 Taylor and Laurent series . . . 11
1.6 Residue theorem . . . 13
1.7 Analytic continuation . . . 15

2 Distributions **19**

2.1 Functionals and distributions . . . 19
2.2 The δ "function" . . . 21
2.3 Heaviside function and relations between distributions . . . 23

3 Fourier series and integral transforms **25**

3.1 Fourier series . . . 25
3.2 Fourier transform . . . 29
3.3 Laplace transform . . . 33

4 Ordinary differential equations **36**

4.1 First order ODE . . . 36
4.2 Second order ODE . . . 39
4.3 Higher-order ODE . . . 41
4.4 Frobenius method (series solution) . . . 42

5 Special ordinary differential equations **43**

5.1 Legendre differential equation . . . 43
5.2 Generalized Legendre differential equation . . . 45
5.3 Laguerre differential equation . . . 46
5.4 Hermite differential equation . . . 48
5.5 Bessel differential equation . . . 48
5.6 Airy differential equation and Laplace method . . . 50
5.7 Sturm-Liouville differential equation . . . 52

6 Partial differential equations **54**
6.1 First order PDE . . . 54
6.2 Second order PDE . . . 58

7 Special partial differential equations **63**
7.1 Laplace equation in cartesian coordinates . . . 64
7.2 Laplace equation in spherical coordinates . . . 66
7.3 Laplace equation in cylindrical coordinates . . . 69
7.4 Poisson equation and Green functions . . . 71
7.5 Heat or diffusion equation . . . 75
7.6 Schrödinger equation . . . 77

8 Intermezzo: MMP I → MMP II and 1+1=2 **81**

9 Abstract spaces **83**
9.1 Metric spaces . . . 83
9.2 Banach spaces . . . 89
9.3 Hilbert spaces . . . 93

10 Linear operators in Hilbert spaces **98**
10.1 Finite-dimensional Hilbert spaces . . . 98
10.2 Direct sum, projectors and ONB . . . 100
10.3 The adjoint operator . . . 103
10.4 The spectrum . . . 106
10.5 The spectral representation of compact operators . . . 107
10.6 Integral equations . . . 110
10.7 Convergence of operators . . . 112
10.8 The resolvent and operators with compact resolvent . . . 113
10.9 "Evil" operators and the spectral family . . . 114
10.10 Quantum mechanics . . . 117
10.11 Unitary operators and time evolution . . . 120

11 Groups **121**
11.1 Basics . . . 121
11.2 Representations . . . 121
11.3 A case study: representations of D_3 . . . 121
11.4 $SO(2)$ as Lie group . . . 121
11.5 $SO(3)$, rotations, and $SU(2)$ . . . 121
11.6 $SU(N)$ and $SU(3)$ in particular . . . 121

12 Tensors **122**
12.1 The Euclidean group 122
12.2 Tensors and coordinate transformations 123
12.3 The Poincaré and Lorentz group and Lorentz transformations 127

A Curvilinear coordinates **131**

1 Complex analysis

1.1 Complex numbers and functions

We start with a brief repetition of complex numbers and some functions, mainly to fix our notation. Complex numbers $z \in \mathbb{C}$ can be written in terms of two real numbers $x, y \in \mathbb{R}$, i.e.

$$\mathbb{C} : \{z = x + iy | x, y \in \mathbb{R}\} \tag{1.1}$$

where $i^2 = -1$. Here $\mathrm{Re}(z) = x$ and $\mathrm{Im}(z) = y$ are the real and imaginary part, respectively. Alternatively we can write $z = re^{i\varphi}$ in terms of its magnitude $|z| = r$ and phase (or argument) $\arg(z) = \varphi$, with

$$r = |z| = \sqrt{x^2 + y^2} \geq 0; \qquad \cos\varphi = \frac{x}{r}; \qquad \sin\varphi = \frac{y}{r} \tag{1.2}$$

Obviously, shifting $\varphi \to \varphi + 2\pi k$ with $k \in \mathbb{Z}$ does not change z and we usually work with $-\pi < \varphi \leq \pi$. For the complex conjugate we use the notation $z^* \equiv x - iy = re^{-i\varphi}$ (sometimes $\bar{z}$ is used). This is illustrated in Figure 1.1.

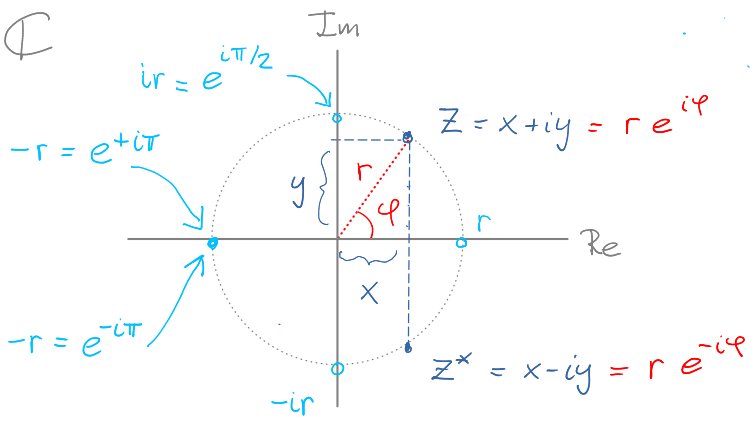

Figure 1.1: Complex plane.

Of course, we can also express x and y in terms of z and z^* as

$$x = \frac{z + z^*}{2}, \qquad y = \frac{z - z^*}{2i} \tag{1.3}$$

The exponential function with a complex argument z is related to 'ordinary' real functions through

$$e^z = e^{x+iy} = e^x \left(\cos y + i \sin y\right) \tag{1.4}$$

It can take any value apart from zero, i.e. $e^z \neq 0$, $\forall z \in \mathbb{C}$. Since $e^z = e^{z+2i\pi k}$ for $k \in \mathbb{Z}$ it actually takes any value other than 0 for (infinitely) many different values of z. Related to the exponential function are the hyperbolic functions

$$\begin{aligned} \cosh z &= \frac{1}{2}\left(e^z + e^{-z}\right) = \cos(iz) \\ \sinh z &= \frac{1}{2}\left(e^z - e^{-z}\right) = -i\sin(iz) \end{aligned} \tag{1.5}$$

So far, everything was fairly straightforward, but turning to the logarithm we will encounter an issue related to $e^z = e^{z+2i\pi k}$ for $k \in \mathbb{Z}$ mentioned above. If we want to define $\log w$ (we always use the natural logarithm, so $\log w = \ln w$) for $w \in \mathbb{C}$ we are looking for a $z \in \mathbb{C}$ such that $e^z = w$. If we find such a z, we get infinitely many other solutions $z + 2i\pi k$. Thus, for

$$w = |w|e^{i\varphi} \qquad \xrightarrow{?} \qquad \log w = \log(|w|) + i\big(\varphi + 2\pi k\big) \tag{1.6}$$

we end up with an ambiguity as to which value $k \in \mathbb{Z}$ we are supposed to take. For a proper definition of the logarithm function in the complex and to prevent it from being multivalued we have to specify the *branch* *[Zweig]*. We will follow the most common choice and pick k such that $-\pi < \varphi + 2\pi k < \pi$. As illustrated in Figure 1.2 this entails that

$$\begin{aligned} \lim_{\varepsilon \searrow 0} \log(-1 + i\varepsilon) &= +i\pi \\ \lim_{\varepsilon \searrow 0} \log(-1 - i\varepsilon) &= -i\pi \end{aligned} \tag{1.7}$$

where here (and in the remainder of this course) it is understood that $\varepsilon > 0$. In other words, we cut the complex plane from 0 along the negative real axis (dashed magenta line in Figure 1.2) and require $\log z$ to be continuous as long as we do not cross this cut.

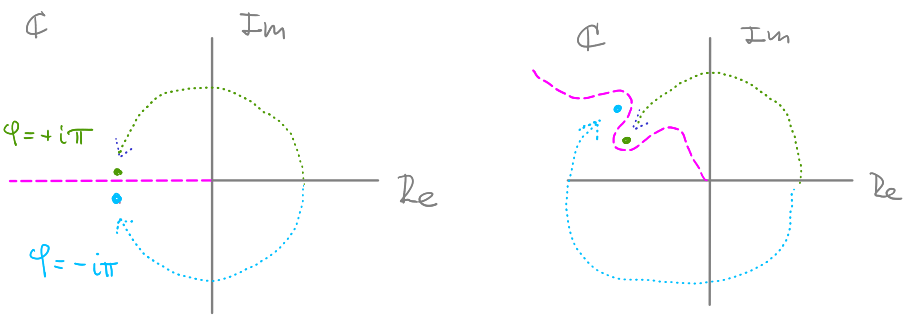

Figure 1.2: Left panel: standard choice for the branch cut (magenta) of the logarithm. For the green point $\mathrm{Im}(\log z) > 0$ whereas for the light blue point $\mathrm{Im}(\log z) < 0$. Crossing the cut, $\log z$ is not continuous. Right panel: another possible choice for the branch cut that might seem attractive after a couple of pints, but is usually disregarded the morning after. Again for the green point $\mathrm{Im}(\log z) > 0$ whereas for the light blue point $\mathrm{Im}(\log z) < 0$.

Once we have the logarithm we can generalize expressions like z^w to $z, w \in \mathbb{C}$. For $z \neq 0$ we define

$$z^w = \left(e^{\log z}\right)^w = e^{w \log z} \tag{1.8}$$

If the exponent is real, $w \in \mathbb{R}$, it is particularly convenient to use

$$z = |z|e^{i\varphi} \qquad \text{then} \qquad z^w = |z|^w\, e^{iw\varphi} \tag{1.9}$$

This leads to another case of ambiguities, namely in square (or higher) roots (even for real arguments $x > 0$, $\sqrt{x} \in \mathbb{R}$ has two solutions). For the nth root ($n \in \mathbb{N}$) of $z = |z|e^{i\varphi} \in \mathbb{C}$ this is generalized as

$$\sqrt[n]{z} = \sqrt[n]{|z|}\, e^{i\frac{\varphi + 2\pi k}{n}} \tag{1.10}$$

with $\sqrt[n]{|z|} \geq 0$. Choosing all values $k \in \mathbb{Z}$ leads to n different solutions for $\sqrt[n]{z}$. ⇒ lecture

Now that we've got to know some functions with complex arguments let us consider an arbitrary function

$$f(x,y) = u(x,y) + iv(x,y) = f\Big(\frac{z+z^*}{2}, \frac{z-z^*}{2i}\Big) \tag{1.11}$$

Here $u(x,y) \in \mathbb{R}$ and $v(x,y) \in \mathbb{R}$ are the real and imaginary part of the function f and for the last expression we simply used (1.3). As shown in Figure 1.3 we can interpret this as a map from $z = x + iy \in \mathbb{C}$ (often the function is only defined for a subset of $\mathbb{C}$) to $f = u + iv \in \mathbb{C}$.

As a simple example consider a particular (set of) function(s)

$$f_{(k)} = \underbrace{\big(x^2 + kx - y^2\big)}_{u(x,y)} + i\,\underbrace{\big(2xy - ky\big)}_{v(x,y)} = z^2 + k\,z^* \tag{1.12}$$

Here $k \in \mathbb{R}$ is to be considered as a fixed (real) parameter. Taking as an example $k = 4$, the function $f_{(k=4)}$ maps e.g. $z = 1 - i \to f_{(k=4)}(1,-1) = 4 + 2i$, as indicated in Figure 1.3.

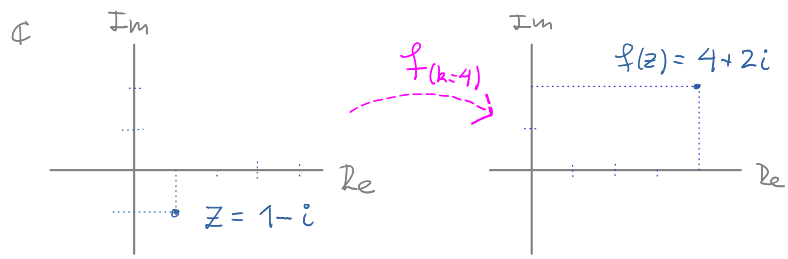

Figure 1.3: The function $f_{(k=4)}$ of (1.12) as a map of (usually a subset of) $\mathbb{C}$ to $\mathbb{C}$.

In the last step of (1.12) we have expressed the function in terms of z and z^*, in analogy with (1.11). As we will see, the function's properties will alter dramatically for the particular choice $k = 0$.

1.2 Analytic functions

We consider a function $f(x,y)$ defined on a *domain [Gebiet]* $\mathcal{G} \subset \mathbb{C}$. In Section 1, the term domain will always refer to a subset of $\mathbb{C}$ that is *connected [zusammenhängend]* and *open [offen]*. Both these terms should be familiar, so a quick recap will have to do. Open means that $\forall z_0 \in \mathcal{G}$ there exists a *neighbourhood [Umgebung]* $U_\varepsilon(z_0)$ such that $U_\varepsilon(z_0) \subset \mathcal{G}$. Loosely speaking, there is no sharp boundary. Connected means that for any two points $z_0, z_1 \in \mathcal{G}$ we can find a *path [Weg]* or *contour [Kurve]* from z_0 to z_1 that lies entirely within $\mathcal{G}$. For the moment we can think of a path or contour in an intuitive way, but we will soon be more precise. Figure 1.4 shows three areas, two of which qualify as a domain (as used in this section) and one that is not connected. In all these figures it is understood that the boundary is not part of the domain. The first domain (to the left) is not only connected, but in fact *simply connected [einfach zusammenhängend]*. Again, for the moment we take a very intuitive approach to this as to 'there are no holes'. If you're not happy with this sloppy

quick fixes (nor should you be) you will have to be a bit patient (for some concepts even wait for MMP II) or brush up on your previous Mathematics lectures.

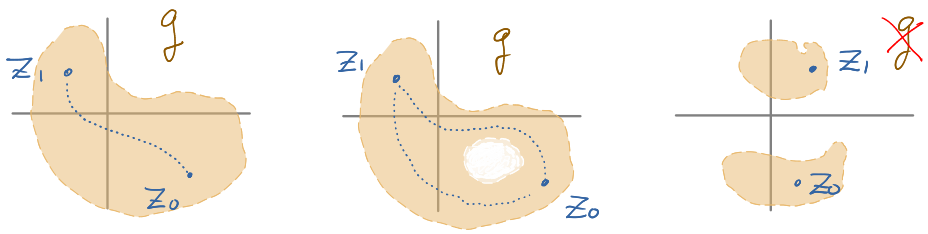

Figure 1.4: A simply connected domain $\mathcal{G}$ on the left, a (not simply connected) domain $\mathcal{G}$ in the middle and a subset of $\mathbb{C}$ that is not connected. The boundaries are drawn dashed to indicate that they are not part of the domain.

A typical example of a simply connected domain is $\mathcal{G}_1 : \{z|\ |z| < 1\}$, i.e. all points with magnitude (strictly) less than (say) 1. A domain that is not simply connected is $\mathcal{G}_1 \backslash z_0$ with $z_0 \in \mathcal{G}_1$, i.e. the previous $\mathcal{G}_1$ with a 'hole' at z_0. Finally, $\bar{\mathcal{G}}_1 : \{z|\ |z| \leq 1\}$ is not open, so not what we call a domain here.

From now on we will always tacitly assume that our functions are defined on a domain $\mathcal{G}$, i.e. $f : \mathcal{G} \to \mathbb{C}$. Note that $\mathbb{C}$ itself is a domain.

In order to prepare for a definition of dramatic importance we introduce some notation and assumptions. For the partial derivatives we will use the short-hand notation

$$u_x \equiv \frac{\partial u(x,y)}{\partial x} \qquad \text{etc.} \tag{1.13}$$

and we assume u_x, u_y, v_x, v_y exist and are continuous in $\mathcal{G}$. Then we can write

$$f_z \equiv \frac{\partial f(z,z^*)}{\partial z} = f_x\, x_z + f_y\, y_z = f_x\, \frac{1}{2} + f_y\, \frac{-i}{2} \tag{1.14}$$

where we have used (1.3). Similarly $f_{z^*} = f_x/2 + i f_y/2$.

And now (drum rolls ...) a **definition**: A (complex) function $f : \mathcal{G} \to \mathbb{C}$ is called *analytic [analytisch]* or *holomorphic [holomorph]* (at $z = z_0$) if $f_{z^*} = f_x/2 + i f_y/2 = 0$ (at $z = z_0$).

We can rephrase this as follows: $f_{z^*} = 0 \Leftrightarrow f_x + i f_y = 0 \Leftrightarrow u_x + i v_x + i u_y - v_y = 0$ which leads to the Cauchy-Riemann differential equations

$$u_x = v_y \quad \wedge \quad u_y = -v_x \qquad \Leftrightarrow \qquad f \text{ analytic} \tag{1.15}$$

Looking at our example (1.12) we get $u_x = 2x + k$ and $v_y = 2x - k$ as well as $u_y = -2y$ and $-v_x = -2y$. Thus, (1.15) is satisfied only if $k = 0$. In this case $f_{(k=0)} = z^2$, if expressed in terms of z and z^*, does only depend on z but not on z^*. So no surprise, $\partial f_{(k=0)}/\partial z^* = 0$. To summarize, $f_{(k)}$ as defined in (1.12) is not analytic (anywhere) except if $k = 0$. Then it is analytic everywhere, except at $z = \infty$. The last statement is a bit mysterious at the moment, and we will come back to this point later.

So far the concept of analyticity is simply a rather obscure property of a complex function. The deep significance of this concept comes with the following statement which in fact is an **equivalent definition** of analyticity:

A (complex) function $f : \mathcal{G} \to \mathbb{C}$ is called *analytic [analytisch]* or *holomorphic [holomorph]* (at $z = z_0$) if f is differentiable and single valued (at $z = z_0$).

Thus, for an analytic function (at $z = z_0$) the limit

$$\lim_{\delta z \to 0} \frac{f(z_0 + \delta z) - f(z_0)}{\delta z} \equiv \lim_{\delta z \to 0} \frac{\delta f(z_0)}{\delta z} \equiv \frac{df}{dz}\Big|_{z=z_0} \equiv f'(z_0) \tag{1.16}$$

exists (at $z = z_0$) and in particular, the limit is independent of the direction of approach to z_0. The latter point has a deeper significance for complex functions compared to real functions. For a real function, the point x_0 can only be approached either from below or above. If both limits exist and they are the same, the real function is differentiable at $x = x_0$. In the complex case, we can approach $z = z_0$ from infinitely many different directions, as illustrated in Figure 1.5. Only if all these infinitely many limits exist and are the same is the function differentiable (hence analytic) at $z = z_0$. Requesting this is a deep intrusion into the personality of a function. Hence, we should not be surprised if such functions have many other amazing properties.

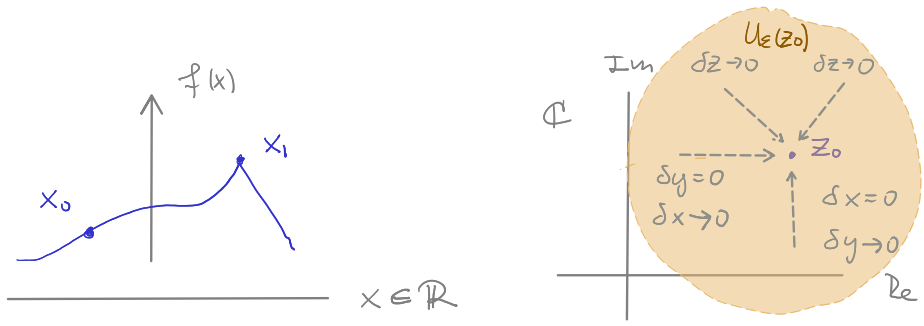

Figure 1.5: Left panel: a real function that is differentiable at x_0 but not at x_1 (the limit approaching from the left is different to the limit approaching from the right). Right panel: illustration of taking the limit $\delta z \to 0$ from various directions for a complex function defined in a neighbourhood $U_\varepsilon(z_0)$ around z_0.

We still need to prove that the definition of analyticity through the Cauchy-Riemann equation (1.15) and through differentiability are equivalent. Thus, we need to show that ⇒ lecture

(i) if f is differentiable, (1.15) is satisfied

(ii) if (1.15) is satisfied, the limit (1.16) exists and is unique.

For an analytic function the derivative as defined in (1.16) is the same as (1.14) and entails the rules for differentiation familiar from real functions. With the concept of an analytic function now available we give an outlook on some common functions and their behaviour regarding analyticity. A more careful treatment will follow in Section 1.5.

A simple starting point is a polynomial (of rank n)

$$f_{ex1}(z) = a_n\, z^n + a_{n-1}\, z^{n-1} + \ldots a_1\, z + a_0 \tag{1.17}$$

This function is analytic $\forall z$ except again for the mysterious $z = \infty$. Using the rules of differentiation familiar from real functions we know very well what the derivative of f_{ex1} is (and that it exists) and, indeed, $f_{ex1}(z)$ does not depend on z^*, so $(f_{ex1})_{z^*} = 0$.

Another example is a rational function, for example

$$f_{ex2}(z) = \frac{1}{z-2} \tag{1.18}$$

This function does not depend on z^*, so it is analytic. The only problem is $z = 2$ where the function is not defined (like $f_{ex1}(z)$ was not defined for $z = \infty$). In fact, at $z = 2$ the function f_{ex2} is not analytic, but we say it has a *pole [Pol]*. It is a single pole (will become clearer in Section 1.5). Similarly, the function $1/(z-2)^2$ has a double pole and $1/(z-2)^n$ with $n \in \mathbb{N}$ and n fold pole.

So far so good, what about

$$f_{ex3}(z) = |z| \tag{1.19}$$

This can be rewritten as $f_{ex3}(z) = \sqrt{z\,z^*}$, so is not analytic anywhere, since the presence of z^* spoils the party (what about $f_{ex3}(z)$ at $z = 0$?).

Going back to the examples we have considered before, $f(z) = e^z$ is analytic everywhere apart from $z = \infty$. This also holds for $\sin(z)$ and $\cos(z)$ or the hyperbolic functions (1.5).

To get a first encounter with a feature we have not yet come across let's consider the function

$$f_{ex4}(z) = e^{1/z} = 1 + \frac{1}{z} + \frac{1}{2!\,z^2} + \frac{1}{3!\,z^3} + \dots \tag{1.20}$$

This function is analytic everywhere apart from $z = 0$. In fact, the standard expansion of the exponential function shown in (1.20) reveals that at $z = 0$ we get poles of arbitrarily high powers. This is much 'worse' than an n-fold pole encountered before. However bad you think the situation is, it is actually worse. The function has a complete nervous breakdown. We will get to know such a behaviour as *essential singularity [wesentliche Singularität]*.

We have repeatedly mentioned the point $z = \infty$. For the moment we just define the value of a complex function $f(z)$ at infinity as $\lim_{w\to 0} f(1/w)$. Then we can see that f_{ex1} has an n-fold pole at infinity and f_{ex2} is well defined (actually 0) at infinity. Similarly, $f_{ex4} = 1$ at infinity. For $f(z) = e^z$, we get arbitrary powers of $(1/w)^n$ for $w \to 0$, so all hell breaks loose at $z = \infty$ and we get again an essential singularity.

Turning to the logarithm, we have already encountered the issue that to start with it is multivalued. If we sort this out, using the branch as indicated in the left panel of Figure 1.2 we are still left with the problem as to what is $\log(z)$ for $z < 0$, $z \in \mathbb{R}$. There $\log z$ is simply not defined, very much like f_{ex2} is not defined for $z = 2$. Of course, we could choose the cut on the right panel and make e.g. $\log(-2)$ well defined. But this would just move the problem to other points, namely those on the new cut. The real issue is that the logarithm has a *branch point [Verzweigungspunkt]* at $z = 0$. If we follow on a closed path around this point, the function changes its value. As we will see, this is related to the fact that at $z = 0$ the derivative of the logarithm does not exist. Whichever branch we choose to make $\log z$ single valued, we have to introduce a cut starting at the branch point $z = 0$ and going off to ∞. If we make the standard choice, $\log z$ is well defined and analytic for the domain $\mathcal{G} = \mathbb{C}\backslash\{\text{negative real axis}\}$.

A similar situation occurs for the square (or nth) root. Again there is a branch point at $z = 0$, where the derivative of the function does not exist. To make the function single valued (i.e. decide which solution of (1.10) we are taking) we need to introduce (a) branch cut(s). Up to this ambiguity $f(z) = \sqrt{z}$ is analytic.

1.3 Contour integrals

We consider a (complex) function $f(x, y)$ that is continuous but not necessarily analytic and want to introduce the concept of a *contour integral* *[Kurvenintegral]*, i.e. an integration of $f(z)$ along a contour γ denoted by $\int_\gamma f(z)dz$.

First let us define a contour in the complex plane. This is done by taking a real parameter $t \in \mathbb{R}$ and associate to each value of t in an interval $[a, b] \subset \mathbb{R}$ a complex number $z(t) \in \mathbb{C}$.

$$\begin{aligned} \gamma : \quad [a,b] \subset \mathbb{R} \quad &\to \quad \mathcal{G} \subset \mathbb{C} \\ t \in [a,b] \quad &\to \quad z(t) = x(t) + i\, y(t) \end{aligned} \tag{1.21}$$

In addition we require this map $t \to z(t)$ to be continuous and (at least piecewise) differentiable, see left panel of Figure 1.6. $z(a)$ and $z(b)$ are the starting and end points of the contour, both could be at infinity. If $z(a) = z(b)$ the contour is closed and usually the symbol $\oint$ is used for integration along closed contours. Every contour has a direction, thus if we swap $z(a)$ and $z(b)$ we convert $\gamma \to -\gamma$. We can combine contours, as e.g. $\gamma = \gamma_1 + \gamma_2 + \gamma_3$ in the left panel of Figure 1.6.

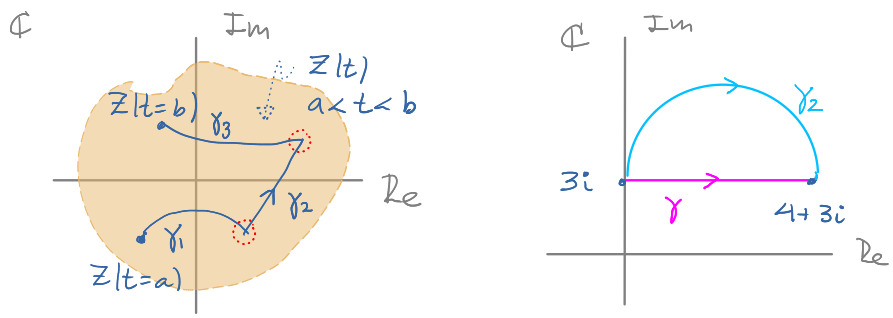

Figure 1.6: Left panel: a piecewise smooth contour γ made up of three smooth contours γ_i. At the red circled points the contour is continuous but not differentiable. Right panel: explicit examples for contours γ (magenta) and γ_2 (blue) for the integral (1.24) and (1.25).

The same contour can be parameterized in (infinitely many) different ways. To illustrate this, we give three parameterizations of the magenta contour shown in the right panel of Figure 1.6.

$$\text{parameterization 1}: \quad \gamma : [0,1] \quad t \to z(t) = 4t + 3i \tag{1.22a}$$

$$\text{parameterization 2}: \quad \gamma : [0,4] \quad t \to z(t) = t + 3i \tag{1.22b}$$

$$\text{parameterization 3}: \quad \gamma : [0,2] \quad t \to z(t) = t^2 + 3i \tag{1.22c}$$

With the contour γ defined we now define the contour integral of the function $f(z)$ (that

might also depend on z^*) as

$$\int_\gamma f(z)dz = \int_{t=a}^{t=b} f(z(t))\, z'(t)\, dt \qquad \text{with} \qquad z'(t) = \frac{dz(t)}{dt} \tag{1.23}$$

Taking separately the real and imaginary part of f in (1.23) we get ordinary real integrals w.r.t. t. The complete (complex) value of the contour integral is obtained by combining the real and imaginary parts. It is important to realize that the contour integral does not depend on the parameterization of the contour. Changing the parameterization simply amounts to a change of variables in (1.23) and, hence, does not affect the value of the integral. As an example, we can integrate the function $f_{(k)}(z)$ defined in (1.12) along the contours (1.22) using all three parameterizations and always obtain

$$\int_\gamma f_{(k)}(z)dz = -\frac{44}{3} + 48\, i + k(8 - 12\, i) \tag{1.24}$$

However, changing the contour will of course in general change the value of the integral. This is even the case, if the starting and end points are the same. Indeed, for the blue path γ_2 in Figure 1.6 we obtain ⇒ lecture

$$\int_{\gamma_2} f_{(k)}(z)dz = -\frac{44}{3} + 48\, i + k(8 - 12\, i - 4i\pi) \tag{1.25}$$

The results (1.24) and (1.25) are different, as one would expect. However, for $k = 0$ they happen to agree and for $k = 0$ the function $f_{(k)}(z)$ happens to be analytic. In fact, we could have taken any other contour starting at $z_a = 3i$ and ending at $z_b = 4 + 3i$ and would have found the same result for the contour integral in the case of the analytic function $f_{k=0}(z)$. At this point one might be tempted to think that for an analytic function a contour integral from z_a to z_b is independent of the contour. However, this is not quite correct.

One way to understand why the integral of $f_{(k=0)} = z^2$ from z_a to z_b is independent of the contour is to realise that the analytic function $f_{(k=0)}$ is itself the derivative of a function $F(z) = z^3/3$ that is analytic everywhere along the contour. This additional property ensures that the integral of $f_{(k=0)} = z^2$ along a closed contour is zero or, equivalently, that $\int_{z_a}^{z_b} z^2\, dz$ is independent of the contour chosen from z_a to z_b. Indeed, if $F'(z) = f(z)$, for a path γ from z_a to z_b we can write (1.23) as

$$\int_\gamma f(z)\, dz = \int_{t=a}^{t=b} F'(z)z'(t)dt = \int_{t=a}^{t=b} \frac{dF(z(t))}{dt} dt = F(z_b) - F(z_a) \tag{1.26}$$

indicating that the result of the integration depends only on the start and end point.

Let's consider another example, the function $f(z) = z^k$ with $k \in \mathbb{Z}$. We want to integrate this function along a closed contour $\gamma_\circ$ that is a circle of radius $r = 1$ (or any other value) around $z = 0$. We find ⇒ lecture

$$\oint_{\gamma_\circ} z^k\, dz = \begin{cases} \frac{1}{k+1} e^{it(k+1)}\Big|_0^{2\pi} = 0 & k \in \mathbb{Z}\backslash\{-1\} \\ 2i\pi & k = -1 \end{cases} \tag{1.27}$$

This is in agreement with what we stated above. Let's first consider $k \in \mathbb{Z}\backslash\{-1\}$. In this case the function $f(z) = z^k$ is the derivative of the function $F(z) = z^{k+1}/(k+1)$ which is

analytic along the (closed) contour $\gamma_\circ$. Thus the corresponding contour integral vanishes. Now consider $k = -1$. We have $F(z) = \log z$, which due to the branch cut is not analytic everywhere along the contour. Hence we should not expect the contour integral to vanish. This is also the case for e.g. $k = 1/2$ and hence $\oint_{\gamma^\circ} \sqrt{z}\, dz \neq 0$.

1.4 Cauchy's theorem and integral formula

We have seen that $f : \mathcal{G} \to \mathbb{C}$ analytic does not necessarily imply that $\oint f(z)dz$ integrates to zero. We need a bit more to ensure closed contour integrals are zero. As stated above, one additional condition that suffices is that $f(z)$ is the derivative of a function $F(z)$ that is analytic everywhere along the contour.

Another possibility is to require that $f(z)$ is not only analytic along the closed contour, but also everywhere inside the contour. In other words

$$f : \mathcal{G} \to \mathbb{C} \text{ analytic} \wedge \mathcal{G} \text{ simply connected} \quad \Rightarrow \quad \oint_\gamma f(z)dz = 0 \quad \forall \gamma \subset \mathcal{G} \tag{1.28}$$

This is often called *Cauchy's theorem [Cauchy Theorem]* or the Cauchy-Goursat theorem. Making an (unnecessary) additional assumption that u_x etc are continuous the proof can be given by relating the contour integral to an integral over the area bounded by the curve, using Stoke's theorem. ⇒ lecture

Cauchy's theorem links the behaviour of a function within a closed contour to the value of the contour integral. In what follows we will turn this around and get some information of the function from the value of contour integrals. To prepare for this, we **define** the *winding number [Umlaufzahl]* of a closed contour γ as

$$n(\gamma, z_0) \equiv \frac{1}{2\pi i} \oint_\gamma \frac{dz}{z - z_0} \tag{1.29}$$

If z_0 is outside the contour, $1/(z - z_0)$ is an analytic function everywhere inside the contour and hence $n(\gamma, z_0) = 0$. More generally, writing $dz/(z-z_0) = d\log(z-z_0)$ and using the r.h.s. of (1.6) the winding number gives the integrated change of $\arg(z - z_0)$ while following along the contour γ, see Figure 1.7. If the contour circles z_0 once in the positive sense (counter clock wise), $n = +1$ and so on, as illustrated in the right panel of Figure 1.7 . ⇒ lecture

We are now ready to state *Cauchy's integral formula [Cauchy Integralformel]*. Consider a closed contour $\gamma \subset \mathcal{G}$ and a point $z_0 \in \mathcal{G}$ that does not lie on the contour, $z_0 \notin \gamma$. Then for a function $f : \mathcal{G} \to \mathbb{C}$ analytic in $\mathcal{G}$ we have

$$n(\gamma, z_0) f(z_0) = \frac{1}{2\pi i} \oint_\gamma \frac{f(z)\, dz}{z - z_0} \tag{1.30}$$

This is truly remarkable in that the values of the function $f(z)$ on the contour define the values of $f(z_0)$ inside the contour. Thus, analyticity is such a strong constraint that once the function has committed on values on a contour, it has no freedom any longer on choose values inside the contour.

Note that (1.30) also holds for points z_0 outside the contour, it is just not very interesting. The l.h.s. is zero, since $n(\gamma, z_0) = 0$ and the r.h.s. is zero, since $f(z)/(z - z_0)$ is analytic

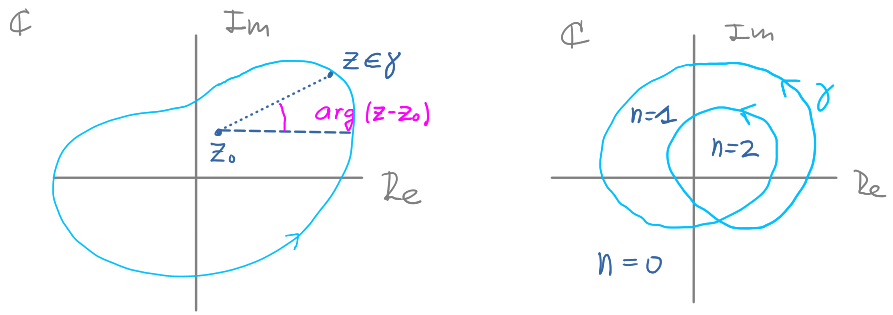

Figure 1.7: Left panel: build up of the winding number as change in $\arg(z - z_0)$. Right panel: example of a contour with winding number 0, 1 or 2.

everywhere inside the contour, and we're back to Cauchy's theorem (1.28). Also, if we change the orientation of the contour, the l.h.s. and the r.h.s. simply get a minus sign.

Often Cauchy's integral formula is written for a simple closed path leading in the positive sense around a point z_0 (that is a contour γ with $n(\gamma, z_0) = 1$). In this case, we just set $n(\gamma, z_0) = 1$ in (1.30).

To prove Cauchy's integral formula (for $n(\gamma, z_0) = 1$) we consider a contour γ_0 as depicted in Figure 1.8. The full contour $\gamma_0 = \gamma + \gamma_\rightarrow - \gamma_r + \gamma_\leftarrow$ consists of the original γ, an inner circle $-\gamma_r$ of radius r (followed through in the negative sense) and the two connections from $\gamma \to \gamma_r$ and $\gamma_r \to \gamma$. The last two are the same contour in different directions, so their contributions cancel. Since $f(z)/(z - z_0)$ is analytic everywhere inside the full contour γ_0 we have

$$\oint_\gamma \frac{f(z)\,dz}{z - z_0} = \oint_{\gamma_r} \frac{f(z)\,dz}{z - z_0} \quad \Rightarrow \text{lecture} \quad = 2\pi i\, f(z_0) \tag{1.31}$$

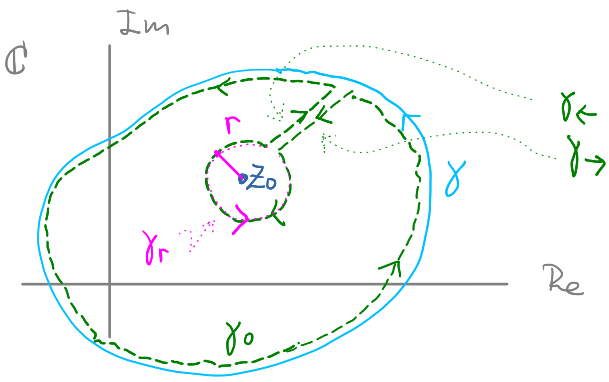

Figure 1.8: Contours used in the proof of Cauchy's integral formula.

We stated above that the values of an analytic $f(z)$ on the contour fix $f(z)$ inside the

contour. In fact, even the derivatives of $f(z)$ are determined. Indeed

$$f'(z_0) = \lim_{h\to 0} \frac{f(z_0+h)-f(z_0)}{h} = \lim_{h\to 0} \frac{1}{2\pi i} \oint_\gamma \frac{f(z)}{h}\Big(\frac{1}{z-z_0-h} - \frac{1}{z-z_0}\Big)dz$$
$$= \frac{1}{2\pi i} \oint_\gamma \frac{f(z)\,dz}{(z-z_0)^2} \tag{1.32}$$

This can be generalised to higher derivatives and for the nth derivative we find

$$f^{(n)} = \frac{n!}{2\pi i} \oint_\gamma \frac{f(z)\,dz}{(z-z_0)^{n+1}} \tag{1.33}$$

This also allows us to conclude that derivatives of analytic functions are themselves analytic functions. Note that (1.32) and (1.33) are written for a contour with $n(\gamma, z_0) = 1$.

1.5 Taylor and Laurent series

Since analytic functions can be differentiated (say at z_0 and, hence, in a neighbourhood of z_0) it is not surprising that they can be approximated in this neighbourhood by a Taylor series. To follow this up, we start by considering a function that is analytic in a simply connected domain. To be concrete, we take the open disc of radius R, i.e. $\mathcal{G}_R \equiv \{z|\,|z-z_0| < R\}$ as domain. Any function that is analytic in $\mathcal{G}_R$ can be written $\forall z \in \mathcal{G}_R$ through a *Taylor series* *[Taylorreihe]* as

$$f(z) = \sum_{n=0}^{\infty} a_n (z-z_0)^n = \sum_{n=0}^{\infty} \frac{f^{(n)}(z_0)}{n!}(z-z_0)^n \tag{1.34}$$

In fact, the statement above is sometimes taken as the definition of analyticity.

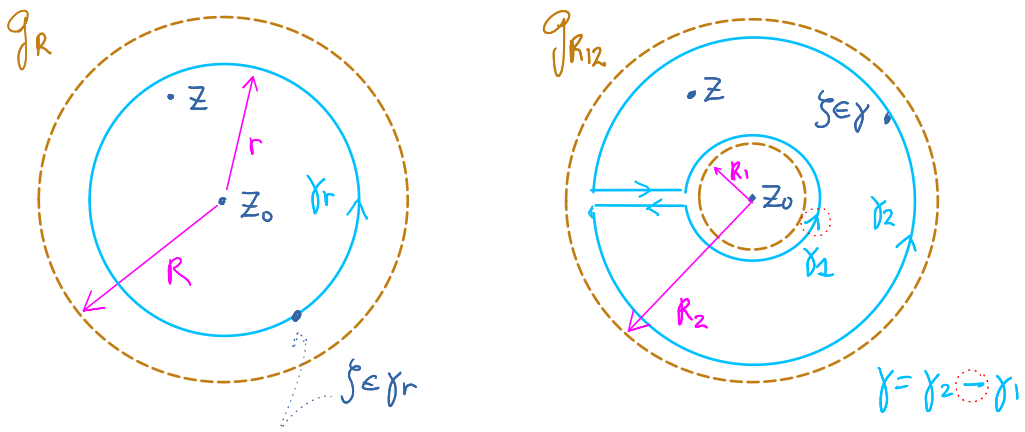

Figure 1.9: Contours used in the proof of the Taylor (left) and Laurent (right) expansion.

The idea of the proof is to pick a (say circular) contour γ_r with radius r s.t. it lies entirely $\Rightarrow$ lecture

within $\mathcal{G}_R$ and includes z and z_0, see left panel of Figure 1.9. Since $\mathcal{G}_R$ is open, this is always possible. We can then write

$$\begin{aligned} f(z) &= \frac{1}{2\pi i}\oint_{\gamma_r} \frac{f(\zeta)\,d\zeta}{\zeta - z} \\ &= \sum_{k=0}^{n-1} \frac{(z-z_0)^k}{k!}\frac{k!}{2\pi i}\oint_{\gamma_r}\frac{f(\zeta)\,d\zeta}{(\zeta-z_0)^{k+1}} + \frac{(z-z_0)^n}{2\pi i}\oint_{\gamma_r}\frac{f(\zeta)\,d\zeta}{(\zeta-z)(\zeta-z_0)^n} \end{aligned} \tag{1.35}$$

where we have used (1.30) and the algebraic identity

$$\frac{1}{\zeta - z} = \frac{1}{(\zeta - z_0)}\frac{1}{1-w} = \frac{1}{(\zeta-z_0)}\Big(1 + w + \ldots w^{n-1} + \frac{w^n}{1-w}\Big) \tag{1.36}$$

with $w \equiv (z-z_0)/(\zeta - z_0)$ (note $|z-z_0| < |\zeta - z_0|$, hence $|w| < 1$). Comparing the first term in the second line of (1.35) with (1.33) we recognize the build up of (1.34). The remainder, i.e. the second term in the second line of (1.35) can be shown to disappear for $n \to \infty$.

As a next step we consider a function that is analytic in a domain that is not simply connected. To be concrete, we take an open disc of radius R_2 and punch out a "hole" of radius R_1 in the middle, i.e. $\mathcal{G}_{R_{12}} \equiv \{z|\, R_1 < |z - z_0| < R_2\}$.

Any function that is analytic in $\mathcal{G}_{R_{12}}$ can be written $\forall z \in \mathcal{G}_{R_{12}}$ through a *Laurent series [Laurentreihe]* as

$$f(z) = \sum_{n=-\infty}^{\infty} a_n (z-z_0)^n \tag{1.37}$$

This looks very much like (1.34) with the crucial difference that the sum starts at $-\infty$ rather than 0. The terms with $n \leq -1$ are called *principal part [Hauptteil]*.

The proof is similar to the one sketched above but starts with ⇒ lecture

$$f(z) = \frac{1}{2\pi i}\oint_{\gamma_2}\frac{f(\zeta)\,d\zeta}{\zeta - z} - \frac{1}{2\pi i}\oint_{\gamma_1}\frac{f(\zeta)\,d\zeta}{\zeta - z} \tag{1.38}$$

where the contours are shown in the right panel of Figure 1.9. The second term is present since $f(z)$ is not (necessarily) analytic for $|z - z_0| < R_1$. Performing a similar expansion as for the Taylor series (1.36) for $\zeta \in \gamma_2$ and an adapted one for $\zeta \in \gamma_1$ (since now $|\zeta - z_0| < |z - z_0|$) we see that we get negative powers $n \leq 0$ (i.e. the principal part) from the contribution due to γ_1.

We can now go back to our list at the end of Section 1.2 of 'things that can go wrong'. Here we restrict ourselves to *isolated singularities [isolierte Singularitäten]*, i.e. we assume the function $f(z)$ is analytic in $\mathcal{G}_{R_{12}}$ with $R_1 \to 0$. In other words the function is analytic in $\mathcal{G}_R \backslash \{z_0\}$. What can possibly go wrong at z_0?

- The first possibility is that nothing goes wrong, that is $a_n = 0\ \forall n \leq -1$. Sometimes this is called a *removable singularity [hebbare Singularität]*, but really, there is just no singularity. In this case the Laurent series (1.37) turns into a Taylor series (1.34).

- If we're less lucky, $a_n = 0\ \forall n < -m$, but $a_{-m} \neq 0$. Thus, the "worst" term in (1.37) behaves as $1/(z-z_0)^m$. As mentioned in Section 1.2 we call this a pole, in this particular case a *pole of order [Pol der Ordnung]* m or m-fold pole.

- If we're even less lucky, $\nexists m$ s.t. $a_n = 0\ \forall n < -m$. There is no worst power, it gets ever worse. We're back to the essential singularity.

In this context we introduce the term of a *meromorphic function [meromorphe Funktion]* as a function that is analytic in a particular domain apart from a set of isolated points. A typical meromorphic function is a rational function $f(z) = P(z)/Q(z)$, where $P(z)$ and $Q(z)$ are polynomials. The roots of $Q(z)$ correspond to the position of the poles of $f(z)$.

Finally, a **definition** of utmost importance. The coefficient a_{-1} of (1.37) is called the *residue [Residuum]* of the function f at z_0.

$$\operatorname*{res}_{z=z_0} f(z) = a_{-1} \tag{1.39}$$

As we will see in the next section, this coefficient plays a very important role.

1.6 Residue theorem

Let $f(z)$ be an analytic function in $\mathcal{G} = \{z|\, 0 < |z - z_0| < R\}$. This domain is precisely $\mathcal{G}_R \backslash \{z_0\}$ considered at the end of the previous section. Thus $f(z)$ has a Laurent expansion as in (1.37). Computing the contour integral of f along a circle $\gamma_r = \{z|\ |z - z_0| = r\}$ of radius r around z_0 in the positive sense we get

$$\oint_{\gamma_r} f(z)dz = \sum_{n=-\infty}^{\infty} a_n \oint_{\gamma_r} (z - z_0)^n dz = 2\pi i\, a_{-1} = 2\pi i \operatorname*{res}_{z=z_0} f(z) \tag{1.40}$$

where we have used (1.27) and the definition (1.39) in the last steps. The result is not affected if we change the contour, as long as the contour lies within $\mathcal{G}$ and has winding number $n(\gamma_r, z_0) = 1$, i.e. the point z_0 is within the contour.

This result can be generalized to the *residue theorem [Residuensatz]*. Let $z_1, z_2 \dots z_k \in \mathcal{G}$ and $\mathcal{G}' = \mathcal{G} \backslash \{z_1, z_2 \dots z_k\}$. Furthermore, $\gamma \subset \mathcal{G}$ is a contour that avoids all z_i and f is analytic in $\mathcal{G}'$. Then

$$\oint_{\gamma} f(z)dz = 2\pi i \sum_{j=1}^{k} n(\gamma, z_j) \operatorname*{res}_{z=z_j} f(z) \tag{1.41}$$

This is a very powerful statement as the computation of a potentially very complicated integral is reduced to the calculation of residues. The calculation of the residues is typically much simpler. For a function with (at most) a single pole we have $\operatorname{res}_{z=z_j} f(z) = \lim_{z\to z_j}(z - z_j)f(z)$. If this limit is zero, the function is analytic at $z = z_j$. If the limit does not exist, the residue has to be obtained in a different way . ⇒ lecture

As a first example consider $f(z) = e^z/(z^2 - 1)$

$$\oint \frac{e^z}{z^2 - 1} dz = 2\pi i \Big(n(\gamma, +1)\frac{e}{2} + n(\gamma, -1)\frac{-1}{2e} \Big) = 2\pi i \Big(e - \frac{1}{2e} \Big) \tag{1.42}$$

where the contour γ is shown in Figure 1.10. For the first step of (1.42) we have used that the function has single poles at $z = 1$ and $z = -1$ with residues $e/2$ and $-1/(2e)$, respectively. Combining this with the appropriate winding numbers yields the final answer in (1.42). ⇒ lecture

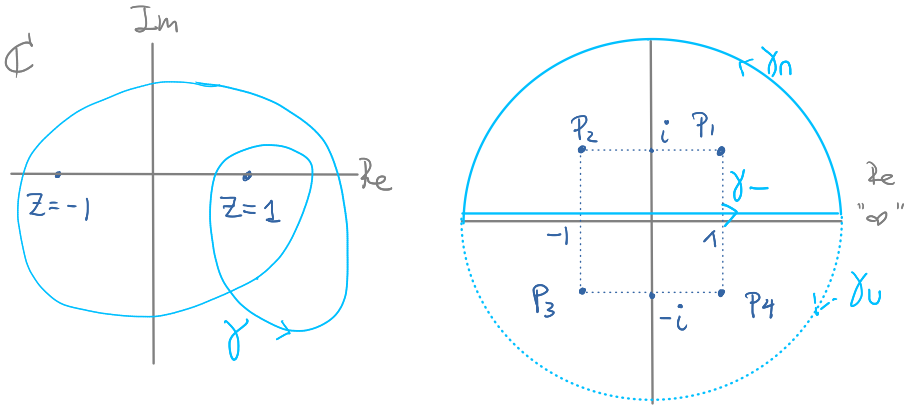

Figure 1.10: Contours used in example (1.42) (left) and example (1.43) (right).

As a second example we consider $\int_{-\infty}^{\infty} dx/(1+x^4)$. To start with, this is not a closed contour, as the integral is along the real axis denoted here by γ_-. We can still use the residue theorem by adding the integral along a half-circle in (in the upper half plane) $\gamma_\cap$. Since $f(z) \sim 1/z^4$ the contribution of the integral along a half-circle of radius R scales as $1/R^3$ and approaches zero for $R \to \infty$. Then we do end up with a contour integral and our function hat two single poles within the contour. One at $z = p_1 = (1+i)/\sqrt{2}$ and one at $z = p_2 = (-1+i)/\sqrt{2}$. Computing the residues and putting everything together we obtain ⇒ lecture

$$\int_{-\infty}^{\infty} \frac{dx}{1+x^4} = \oint_{\gamma_- + \gamma_\cap} \frac{dz}{1+z^4} = 2\pi i \Big(\frac{-1-i}{4\sqrt{2}} + \frac{1-i}{4\sqrt{2}} \Big) = \frac{\pi}{\sqrt{2}} \tag{1.43}$$

In fact, we could have closed the contour by adding a half-circle in the lower plane $\gamma_\cup$. This would pick the residues at $z = p_4 = (1-i)/\sqrt{2}$ and $z = p_3 = (-1-i)/\sqrt{2}$. Taking into account the minus sign for the negative direction of the contour, we find again the result given in (1.43). We note that this method works for all rational functions $f(z) = P(z)/Q(z)$, where the degree of the polynomial $Q(z)$ is bigger by at least 2 compared to the degree of the polynomial $P(z)$. This condition ensures that the the integral along $\gamma_\cap$ vanishes for $R \to \infty$.

As a final example, we consider the integral

$$\int_{-\infty}^{\infty} \frac{dx}{(x^2+1)(x-1)}$$

and notice with a shiver of horror that the integral is actually not defined. The integrand has a pole on the real axis at $x = 1$ and our contour is running right through it. What we can do though, is to compute the *principal value [Hauptwert]* of this integral, which is **defined** for a general function $f(x)$ as

$$\mathcal{P} \int_{-\infty}^{\infty} \frac{f(x)}{x-x_0} dx \equiv \lim_{\varepsilon \searrow 0} \Big(\int_{-\infty}^{x_0-\varepsilon} \frac{f(x)}{x-x_0} dx + \int_{x_0+\varepsilon}^{\infty} \frac{f(x)}{x-x_0} dx \Big) \tag{1.44}$$

It is by no means guaranteed that the limit exist. But if it does and hence the principal value of the integral is well defined, we can (sometimes) use the residue theorem to compute it.

For our example above we get

$$\mathcal{P}\int_{-\infty}^{\infty}\frac{dx}{(x^2+1)(x-1)} \tag{1.45a}$$
$$= \oint_{\gamma}\frac{dz}{(z^2+1)(z-1)} - \int_{\gamma_\varepsilon}\frac{dz}{(z^2+1)(z-1)} - \int_{\gamma_\cap}\frac{dz}{(z^2+1)(z-1)}$$

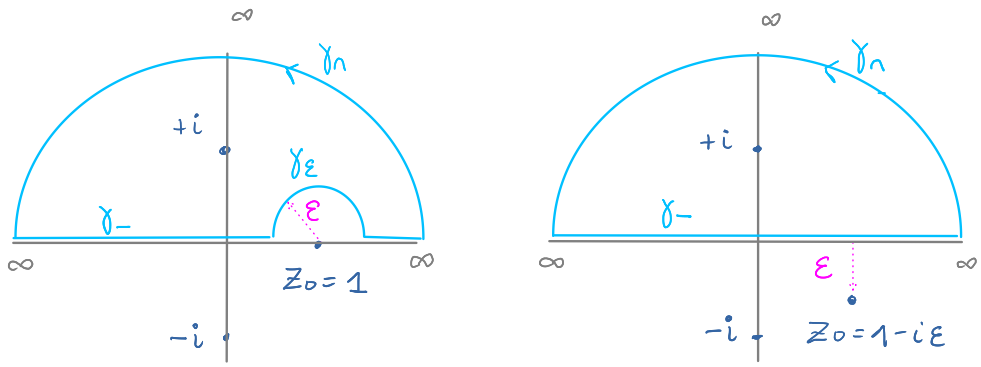

Figure 1.11: Avoiding poles on the real axes to compute the integral in (1.45a) by contour deformation (left) or shifting the pole (right).

As shown in the left panel of Figure 1.11 γ is the full closed contour (not containing the pole at $z = 1$) and γ_ε is a half-circle of (infinitesimally small) radius ϵ around the pole at $x = 1$. It is precisely the latter that ensures the avoidance of the pole and implements the principal value prescription. In the limit $R \to \infty$ the contribution of $\gamma_\cap$ vanishes. The integral along the closed contour γ can be done using the residue theorem (not taking into account the pole at $z = 1$). We are left with computing the integral along γ_ε and taking the limit $\varepsilon \searrow 0$. Doing this we find ⇒ lecture

$$\mathcal{P}\int_{-\infty}^{\infty}\frac{dx}{(x^2+1)(x-1)} = 2\pi i\frac{-1+i}{4} - i\pi\frac{-1}{2} - 0 = -\frac{\pi}{2} \tag{1.45b}$$

As expected, the final result is real. Another possibility to implement the principal value prescription is to write

$$\mathcal{P}\int_{-\infty}^{\infty}\frac{f(x)}{x-x_0}dx \equiv \lim_{\varepsilon\searrow 0}\int_{-\infty}^{\infty}\frac{f(x)}{x-x_0+i\varepsilon}dx + i\pi f(x_0)\,. \tag{1.46}$$

The first term on the r.h.s. corresponds to the integration along the real axes with the pole shifted from $z = 1$ to $z = 1 - i\varepsilon$ (where $\varepsilon \searrow 0$ is understood), as shown in the right panel of Figure 1.11. The second term on the r.h.s. is a remnant from the γ_ε contour. We will encounter such relations again in the Section 2.3 about distributions.

1.7 Analytic continuation

We have seen in (1.30) that the value of an analytic function f at a point z_0 is determined by its value on a closed path around z_0. Thus, if for two analytic functions $f_1(\zeta) = f_2(\zeta)$ on

such a contour, we know $f_1(z) = f_2(z)$ for all points inside the contour.

In fact, we can make a much stronger statement, the *identity theorem [Identitätsprinzip]*. Let f_1 and f_2 be two analytic functions on a domain $\mathcal{G}$. If they happen to agree on $S \subset \mathcal{G}$ where S has an accumulation point, then they actually agree for all points in $\mathcal{G}$. The proof consists of showing that all coefficients in the Taylor expansion of the analytic function $g(z) \equiv f_1(z) - f_2(z)$ are 0, and hence so is $g(z) = 0$. ⇒ lecture

This theorem can be applied to the case illustrated in the left panel of Figure 1.12. Let f_1 be a function that is analytic in the domain $\mathcal{G}_1$ and f_2 is analytig in $\mathcal{G}_2$. Furthermore, assume there is a proper intersection $S = \mathcal{G}_1 \cap \mathcal{G}_2$, i.e. S contains an accummulation point. If $f_1(z) = f_2(z)$, $\forall z \in S$ then these functions have to agree on $\mathcal{G}_1 \cup \mathcal{G}_2$. More precisely, the function $f_2(z)$ is the unique analytic continuation of $f_1(z)$ from the domain $\mathcal{G}_1$ to the larger domain $\mathcal{G}_1 \cup \mathcal{G}_2$. Or, the other way round, $f_1(z)$ is the unique analytic continuation of $f_2(z)$ from the domain $\mathcal{G}_2$ to the larger domain $\mathcal{G}_1 \cup \mathcal{G}_2$.

Let us start with a simple example. Consider the function $f_1(z) = \sum_{n=0}^{\infty} z^n$. For the series to converge, we need $|z| < 1$. In the domain $\mathcal{G}_1 = \{z||z| < 1\}$ the function $f_1(z)$ is analytic, but $f_1(z)$ is not defined for e.g. $z = 2$. Consider now the function $f_2(z) = 1/(1-z)$. This function is analytic in $\mathcal{G}_2 = \mathbb{C}\backslash\{1\}$ and for $S = \mathcal{G}_1 \cap \mathcal{G}_2 = \mathcal{G}_1$ we have $f_1 = f_2$. Hence, f_2 is the unique analytic continuation of f_1 from $\mathcal{G}_1$ to $\mathcal{G}_2$. This might not sound very impressive. But to have a bit of fun, we could now ask the question what is $1+3+9+27+81+\ldots$. The first reaction is to say this series does not converge. But we could also say this is nothing but $f_1(3)$. And, since f_2 is the unique analytic continuation of f_1, when we say $f_1(3)$ we actually mean $f_2(3) = -1/2$. This leads us to make the bold statement $1+3+9+27+81+\ldots = -1/2$.

In case you think this is just fooling about, let us look at another example that will show up several times during this course. We define the Γ (or gamma) function as

$$\Gamma(z) \equiv \int_0^{\infty} t^{z-1} e^{-t} dt \qquad \text{for} \quad \mathcal{G}_1 = \{z|\mathrm{Re}(z) > 0\} \tag{1.47}$$

where the integral converges and $\Gamma(z)$ is analytic $\forall z \in \mathcal{G}_1$. Using integration by parts we can derive a functional relation ⇒ lecture

$$\Gamma(z+1) = z\,\Gamma(z) \tag{1.48}$$

and direct integration yields $\Gamma(1) = 1$. Hence, for $n \in \mathbb{N}$ we have $\Gamma(n) = (n-1)!$. Thus, the Γ function is a generalisation of the factorial from integer values to $z \in \mathcal{G}_1$ and if we insist this should be an analytic finction, this generalisation is unique. We can now use (1.47) to compute for example $\Gamma(1/2) = \sqrt{\pi}$ and get the wonderfully weird statements $(-1/2)! \stackrel{?}{=} \sqrt{\pi}$, or $(1/2)! \stackrel{?}{=} \Gamma(3/2) = \sqrt{\pi}/2$.

But we're not yet done. We can use (1.47) to compute $\Gamma(z)$ in $\mathcal{G}_1$. In a next step, we use (1.48) in the form $\Gamma(z-1) = \Gamma(z)/(z-1)$ to analytically continue $\Gamma(z)$ to an even larger domain, namely $\mathcal{G}_2 = \{x|\mathrm{Re}(z) > -1 \wedge z \neq 0\}$. The only point that gives us trouble in $\mathcal{G}_2$ is the point 0, due to the pole $1/(z-1)$. Hence, the Γ function is not analytic at $z = 0$ but has a single pole. For all other points in $\mathcal{G}_2$ the Γ function is analytic. This is illustrated in the right panel of Figure 1.12. No prize, for guessing our next step. We use $\Gamma(z-1) = \Gamma(z)/(z-1)$ again, and again, and again to increase the domain further and further. Apart from negative integer values (including 0) for z, we do not encounter any problems. The function $\Gamma(z)$ as defined in (1.47) and analytically continued is analytic everywhere in $\mathbb{C}$ apart from $z = -n$ with $n \in \mathbb{N}_0$, where it has a single pole.

Since derivatives of analytic functions are analytic again, we can define the Ψ function or digamma function through the relation

$$\Psi(z) \equiv \frac{\Gamma'(z)}{\Gamma(z)} = \frac{1}{\Gamma(z)}\frac{d}{dz}\Gamma(z) = \frac{d}{dz}\ln\big(\Gamma(z)\big) \tag{1.49}$$

Again, this function is analytic everywhere in $\mathbb{C}\backslash\{0, -1, -2\ldots\}$ and has single poles at negative integer values (including 0).

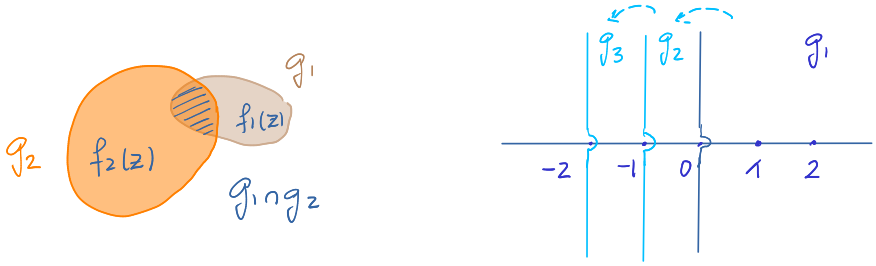

Figure 1.12: Analytic continuation in generic case (left) and for the Γ function (right).

As a final example af a function that is defined through an analytic continuation we mention the (Riemann) ζ function. For $n \in \mathbb{N}$ we define

$$\zeta(n) \equiv \sum_{m=1}^{\infty} \frac{1}{m^n} = 1 + \frac{1}{2^n} + \frac{1}{3^n} + \ldots \tag{1.50}$$

and can evaluate the series to get for example $\zeta(2) = \pi^2/6$. Proceeding in analogy to the Γ function, we define $\zeta(z)$ through (1.50) for $\mathrm{Re}(z) > 1$, to ensure the series converges. So far so good. The fun starts when we apply a functional relation similar to (1.48) (but considerably more complicated) to relate $\zeta(1-z)$ to $\zeta(z)$. This allows us to analytically continue the ζ function to $\mathbb{C}\backslash\{1\}$. The function is analytic everywhere except at $z = 1$ where it has a single pole. Again, this leads to wonderfully weird statements like $1 + 2 + 3 + \ldots \stackrel{?}{=} \zeta(-1) = -\frac{1}{12}$. ⇒ lecture

Above we have seen that we can analytically continue a function if there is an overlap S (with accumulation point) of the two domains. Next, we look at a somewhat more subtle case. Consider a function $f(z)$ analalytic in a domain $\mathcal{G}$ in the upper half plane, but bordering on the real axis, see left panel of Figure 1.13. Here S denotes the boundary of $\mathcal{G}$ that lies on the real axis. Thus S is not part of $\mathcal{G}$. Can we still analytically continue to a larger domain, in particular one that extends to the lower half plane?

The answer is given by the *reflection principle [Spiegelungsprinzip]* of Schwarz. If in addition to the analyticity of f in $\mathcal{G}$ we also know that $f(z) \in \mathbb{R}$ for $z \in S$ then we can extend the function analytically to the domain $\mathcal{G} \cup \mathcal{G}^*$, where $\mathcal{G}^*$ corresponds to the domain $\mathcal{G}$ reflected about the real axis, as depicted in the left panel of Figure 1.13. For z^* in the lower half plane, i.e. $z^* \in \mathcal{G}^*$ we have $z \in \mathcal{G}$ and the extended function is defined through $f(z^*) = f^*(z)$.

As a final point we consider an example of a *dispersion relation [Dispersionsrelation]*. Let $f(z)$ be a function that is analytic everywhere except for a branch cut on the positive real axis, starting at x_0. For any point z not on this branch cut, we can use (1.30) to relate $f(z)$ to a contour integral along γ, as depicted in the right panel of Figure 1.13. If we assume

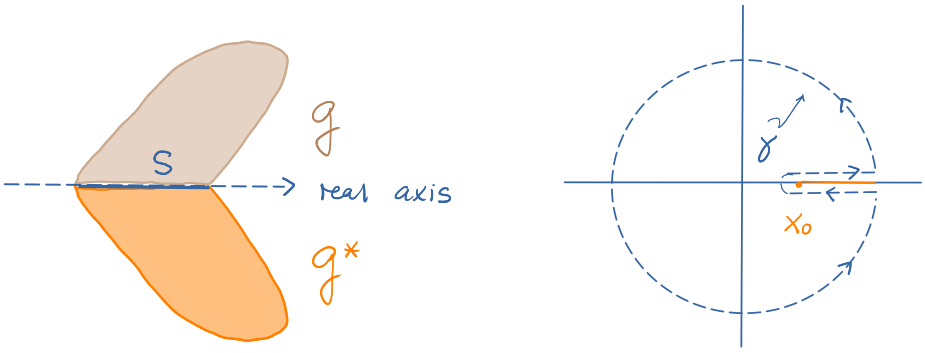

Figure 1.13: Extending the domain of an analytic function through the Schwarz reflection principle (left panel) and an application for a dispersion relation (right panel).

that $f(z) \to 0$ fast enough for $|z| \to \infty$ and using the fact that the small half circle of radius ϵ around x_0 does not contribute to the contour integral (as always, it is implied that $\epsilon > 0$ tends to zero), we find

$$f(z) = \frac{1}{2i\pi} \oint_\gamma \frac{f(\zeta)d\zeta}{\zeta - z} = \frac{1}{2i\pi} \int_{x_0}^{\infty} \frac{f(x+i\epsilon) - f^*(x-i\epsilon)}{x-z} dx = \frac{1}{\pi} \int_{x_0}^{\infty} \frac{\mathrm{Im}\big(f(x+i\epsilon)\big)}{x-z} dx \tag{1.51}$$

Thus, the function at any point z is determined by the imaginary part of f just above the branch cut. This is often useful in physics, since it is possible to obtain information on the imaginary part of certain quantities by other means. Sometimes the function $f(z)$ does not fall of sufficiently rapidly to ensure that the contribution from the outer circle of γ does not contribute. In such cases it is possible to consider so-called subtraced dispersion relations with additional factors of x in the denominator.

2 Distributions

You will have encountered the $\delta(x)$ "function" and heard statements like: this is a function that is 0 everywhere except at $x = 0$, but it integrates to 1. Taken literally, this is of course utter nonsense. Yet we seem to be getting away treating $\delta(x)$ like this and in this section we want to understand a bit better what we're actually doing with objects like the $\delta(x)$ "function".

2.1 Functionals and distributions

We are familiar with the concept of a real function $f(x)$ where to an input variable $x \in \mathbb{R}$ we associate a value $f \in \mathbb{R}$. This can be generalized to more complicated inputs (e.g. $z \in \mathbb{C}$ as in the previous section, or more than one variable) and more complicated outputs. What we will be concerned with in this section are *functionals [Funktionale]*. The input of a functional is not a (or several) variable(s) but a function. The output is a real or complex number. As for functions, we could consider very complicated inputs, several functions of several variables each, but we restrict ourselves to functionals $T[\varphi]$ of a single function of a single variable $\varphi(x)$. Similar to functions, where $f(x)$ might not be defined for any value of x, the functional $T[\varphi]$ might also be defined only for 'decent enough' functions $\varphi(x)$. Typically, we consider $\varphi(x) \in C_c^\infty$, i.e. infinitely differentiable functions with compact support (not messing about at $x \to \infty$). Thus, we finally **define** the functional as a map

$$\varphi(x) \in C_c^\infty : \quad \varphi(x) \to T[\varphi(x)] \in \mathbb{R} \quad (\text{or } \mathbb{C}) \tag{2.1}$$

restricting ourselves to real functionals. A typical example of a functional is

$$T[\varphi(x)] = \int \mathcal{T}\big(\varphi(x), \varphi'(x), \dots\big) dx \tag{2.2}$$

where $\mathcal{T}$ is a function that depends on $\varphi(x)$ and its derivatives. As a side remark we mention that in Classical Mechanics you will consider just such a functional simply with a different notation $T \to S$, $\mathcal{T} \to \mathcal{L}$, $\varphi \to q$ and $x \to t$, namely $S[q(t)] = \int \mathcal{L}(q(t), \dot{q}(t), t) dt$ where S is the action, $\mathcal{L}$ the Lagrangian and $q(t)$ the trajectory of (say) a point mass as a function of time t.

In addition, from now on we typically consider *linear [lineare]* functionals (even though we will often not explicitly state this), i.e. $\forall c_1, c_2 \in \mathbb{R}$ (or $\mathbb{C}$) and $\forall y_1, y_2 \in C_c^\infty$ we have

$$T[c_1\, y_1(x) + c_2\, y_2(x)] = c_1 T[y_1(x)] + c_2 T[y_2(x)] \tag{2.3}$$

In fact, the notion of a (linear) functional is much more general. As we will see in MMP II, a functional is defined as a (linear) map of a general vector space X to $\mathbb{R}$ (or $\mathbb{C}$). In this section we only consider the case where $X = C_c^\infty$.

We are now ready to **define** *distributions [Distributionen]* as continuous, linear functionals. Since distributions are functionals, you should always think of them as acting on *test functions [Testfunktionen]*, in our case the $\varphi \in C_c^\infty$. A continuous (linear) functional (i.e. a distribution) satisfies

$$\lim_{n\to\infty} \varphi_n \to \varphi \quad \Rightarrow \quad \lim_{n\to\infty} T[\varphi_n] \to T[\varphi] \tag{2.4}$$

In words, if we have a sequence of test functions φ_n approaching φ, the result of a continuous functional acting on φ_n will approach $T[\varphi]$. What is meant by $T[\varphi_n] \to T[\varphi]$ is clear, since $T[\varphi_n]$ are (real) numbers. What is not entirely clear at the moment is what is meant by $\varphi_n \to \varphi$. For the time being we take an intuitive approach to this and will wait for MMP II for a more precise formulation.

It is now time to get to know the mother of all linear functionals, T_τ, where $\tau(x)$ is a (integrable) function:

$$T_\tau[\varphi] \equiv \int_{-\infty}^{\infty} \tau^*(x)\,\varphi(x)\,dx \qquad \text{later written as } \langle\tau|\varphi\rangle \tag{2.5}$$

You might recognize this as the scalar product of two 'vectors' τ and φ of the vector space C_c^∞. Thus, to every function (i.e. vector) $\tau \in C_c^\infty$ we can associate a linear functional T_τ. As another side remark, in MMP II we will see that the T_τ themselves form a vector space, called the dual space. In the current context, we note that to any integrable function τ there belongs a functional T_τ, through (2.5). This functional is continuous, hence it is even a distribution. Sometimes the term *regular distributions [reguläre Distributionen]* is used for distributions that can be associated via (2.5) to functions $\tau(x)$.

There are also (some extremely important) distributions that are not regular, for example the well known δ distribution. These distributions cannot be associated with a 'normal' function $\tau(x)$. Still, as we will see (and probably have done already) we can often compute with them as if they were functions. The term *generalized function [verallgemeinerte Funktion]* is often used in this context.

An example of a non-regular distribution we have already encountered is the principal value. In the notation of this section we write this (1.44) (with say $x_0 = 0$) as

$$T_{\mathcal{P}\frac{1}{x}}[\varphi] \equiv \mathcal{P}\int_{-\infty}^{\infty} \frac{\varphi(x)}{x}dx = \lim_{\varepsilon\searrow 0}\left(\int_{-\infty}^{x_0-\varepsilon} \frac{\varphi(x)}{x}dx + \int_{x_0-\varepsilon}^{\infty} \frac{\varphi(x)}{x}dx\right) \tag{2.6}$$

Usually a simplified notation $T_{\mathcal{P}\frac{1}{x}} \to \mathcal{P}\frac{1}{x}$ is used. There is no function $\tau(x)$ that we could use in (2.5) to reproduce the result of (1.44). Yet, (1.44) is a well defined procedure that associates to each (sufficiently decent) function φ a real number, as we have seen in (1.45b). It is also fairly obvious that the 'procedure' is continuous. ⇒ lecture

While we cannot necessarily find a function $\tau(x)$ to represent a distribution, we can always represent a distribution through a limit of functions $\tau_n(x)$. Thus, distributions can be represented as

$$T[\varphi] = \lim_{n\to\infty}\int_{-\infty}^{\infty} \tau_n(x)\varphi(x)dx \tag{2.7}$$

If T is regular, we need not bother with the limit and can directly write $\tau(x)$ on the r.h.s. and, thus, associate the distribution with a function as in (2.5). If T is not regular, this cannot be done and trying to take the limit $n \to \infty$ before the integration is the ultimate sin. Distributions are always to be interpreted as a beast waiting for a test function to act upon. But we can still associate a non-regular distribution with a sequence of functions $\tau_n(x)$ and a corresponding limit $n \to \infty$ after the integration. This reaffirms the notion of a distribution as a generalized function.

We can also define derivatives of distributions. In the case of a regular distribution we **define**

$$T'_\tau[\varphi] \equiv \int_{-\infty}^{\infty} \tau'(x)\varphi(x)dx = -\int_{-\infty}^{\infty} \tau(x)\varphi'(x)dx = -T_\tau[\varphi'] \tag{2.8}$$

with an obvious extension to non-regular distributions. Since $\varphi \in C_c^\infty$ the derivative φ' exists.

Of course, we can also define linear combinations of distributions. After all, functionals form a vector space. What is generally not defined is a product of distributions. Whenever you see something like $T_1[\varphi]T_2[\varphi]$, ring all alarm bells.

So far we have played it safe in requiring $\varphi \in C_c^\infty$. In practical applications this condition is often relaxed. Of course, this requires some care as e.g. (2.8) might not be defined any longer. Also, the range of the integration in (2.5) or (2.7) is often modified to a finite range $[a, b]$, in other words we consider $\varphi \in C^\infty[a, b]$.

2.2 The δ "function"

The one distribution every physicist will come across is the δ distribution. Within the setup of Section 2.1 we **define**

$$T_{\delta(x-x_0)}[\varphi] = \delta(x - x_0)[\varphi] \equiv \varphi(x_0) \tag{2.9}$$

where as for the principal value, (2.6), we usually drop the T in the notation. This is a linear continuous functional, hence a distribution. The notation in (2.9) is probably not quite the one you're familiar with. What you will have seen is

$$\int_{-\infty}^{\infty} \delta(x - x_0)\varphi(x)\, dx = \varphi(x_0) \tag{2.10}$$

This very much looks like (2.5), i.e. we pretend we found a function $\tau \to \delta(x - x_0)$ that reproduces the distribution (2.9) and the δ distribution is regular. Sadly this is simply wrong. However, we can find a (in fact many) sequence(s) of functions τ_n such that (2.7) reproduces (2.9). Thus, $\delta(x - x_0)$ can be understood as a generalized function, or a non-regular distribution.

Let us now consider an example of such a sequence of functions, where for convenience we set $x_0 = 0$.

$$\delta(x)[\varphi] = \lim_{n\to\infty} \int_{-\infty}^{\infty} \underbrace{\sqrt{\frac{n}{\pi}} e^{-nx^2}}_{\tau_n(x)} \varphi(x)\, dx \tag{2.11}$$

This is in one-to-one correspondence with (2.7). The functions $\tau_n(x)$, plotted in the left panel of Figure 2.1, tend to become more peaked at $x_0 = 0$ for increasing n. However, $\forall n$ we have $\int_{-\infty}^{\infty} \tau_n(x)dx = 1$. For any 'decent' function φ (in particular for $\varphi \in C_c^\infty$) we can take the limit $n \to \infty$ in (2.11) and find $\varphi(0)$. This is often written as ⇒ lecture

$$\delta(x) = \lim_{n\to\infty} \sqrt{\frac{n}{\pi}} e^{-nx^2} \tag{2.12a}$$

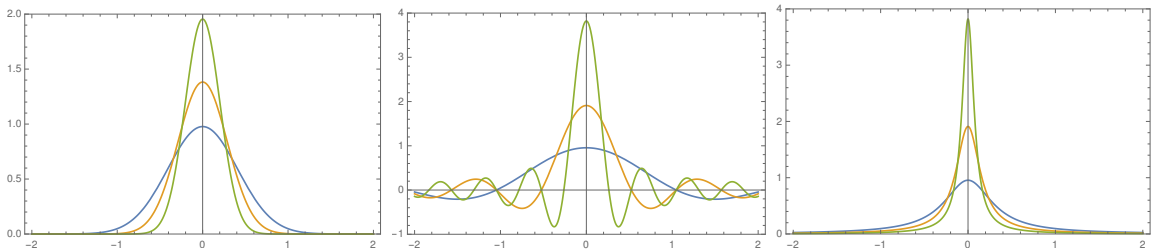

Figure 2.1: Plots for $\tau_n(x)$ as in (2.12a) (left), (2.12b) (middle), (2.12c) (right) for $n = 1/\varepsilon = 3$ (blue), $n = 1/\varepsilon = 6$ (orange), $n = 1/\varepsilon = 12$ (green).

where we have to remember that the limit cannot be taken in (2.12a), but only after the integration (recall the ultimate sin). The expression (2.12a) is a typical generalized function.

As mentioned above, (2.12a) is by no means unique. Other possibilities also plotted in Figure 2.1 are ⇒ lecture

$$\delta(x) = \lim_{n\to\infty} \frac{1}{\pi} \frac{\sin(nx)}{x} \tag{2.12b}$$

$$\delta(x) = \lim_{\varepsilon\to 0} \frac{1}{\pi} \frac{\varepsilon}{x^2 + \varepsilon^2} \tag{2.12c}$$

In the spirit of (2.12) we are now ready to write down the probably most often used representation of the δ "function"

$$\delta(x) = \lim_{n\to\infty} \underbrace{\frac{1}{2\pi} \int_{-n}^{n} e^{ikx} dk}_{\sin(nx)/(\pi x)} = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ikx} dk \tag{2.13}$$

where we have used (2.12b). And, yes, in the last step we have committed the ultimate sin.

Note that all versions of the functions that correspond to τ_n used in (2.11) integrate to 1. Hence

$$\int_{-\infty}^{\infty} \delta(x) dx = \lim_{n\to\infty} \int_{-\infty}^{\infty} \tau_n(x) dx = \lim_{n\to\infty} 1 = 1 \tag{2.14}$$

At the same time, for $x \neq 0$ we can actually take the limit $n \to \infty$ (or $\varepsilon \to 0$) and get $\lim_{n\to\infty} \tau_n(x \neq 0) = 0$, also illustrated in Figure 2.1.

As mentioned at the very end of Section 2.1 in applications we often use finite intervals $x \in [a, b]$. The generalization of (2.10) then reads

$$\int_a^b \delta(x - x_0)\varphi(x)\, dx = \begin{cases} \varphi(x_0) & \text{if } x_0 \in [a, b] \\ 0 & \text{if } x_0 \notin [a, b] \end{cases} \tag{2.15}$$

In actual calculations we often encounter expression where the argument of the δ "function" is more complicated, say itself a function $g(x)$. To deal with such expressions we use

$$\delta\big(g(x)\big) = \sum_{x_i} \frac{1}{|g'(x_i)|} \delta(x - x_i) \tag{2.16}$$

where the sum runs over all $x_i \in [a,b]$ for which $g(x_i) = 0$ and $g'(x_i) \neq 0$. This is an identity between distributions. Thus, it holds if applied on any test function. It can be proved by making a change of variables $x \to g(x)$ in the corresponding integral. ⇒ lecture

A particularly important special case of (2.16) is if $g(x) = c$ is a constant,

$$\delta(c\,x) = \frac{1}{|c|}\delta(x) \qquad \Rightarrow \qquad \delta(-x) = \delta(x) \tag{2.17}$$

The last identity can also be read off immediately from (2.12).

With this we are now well equipped to compute e.g.

$$\int_{1/2}^{7/2} \delta\big(\sin(\pi x)\big) x^2\, dx = \frac{14}{\pi} \tag{2.18}$$

or similar expressions. ⇒ lecture

2.3 Heaviside function and relations between distributions

The Heaviside or step function $\Theta(x)$ is usually defined as

$$\Theta(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x < 0 \end{cases} \tag{2.19}$$

with different conventions for $\Theta(x = 0)$. It is often useful to consider an apparently much more complicated definition of the $\Theta(x)$ function through an integral representation, namely

$$\Theta(x) = \lim_{\varepsilon \to 0} \frac{1}{2i\pi} \int_{-\infty}^{\infty} \frac{e^{ikx}}{k - i\varepsilon} dk \qquad \text{or} \qquad \Theta(x) = \lim_{\varepsilon \to 0} \frac{-1}{2i\pi} \int_{-\infty}^{\infty} \frac{e^{-ikx}}{k + i\varepsilon} dk \tag{2.20}$$

In (2.20) we have to think of k as a complex variable, like z in Section 1. (Compare (2.20) to (1.46)) We can then convert the integration in (2.20) to a closed contour integral. Focusing on the first expression in (2.20), for $x < 0$ we have to close the contour in the lower half plane to ensure we get no contribution from the half circle. In this case no pole is within the ⇒ lecture contour and according to Cauchy's theorem, (1.28), we get 0. For $x > 0$ we have to close in the upper half plane and there is the pole $k = +i\varepsilon$. Applying the residue theorem we get 1, in agreement with (2.19). Similar arguments can be made for the second expression in (2.20).

What happens if we want to differentiate $\Theta(x)$? The function is continuous (actually constant) everywhere except at $x = 0$. Thus for $x \neq 0$ we have $\Theta'(x \neq 0) = 0$. What about $x = 0$? Normally we would just say the function is not differentiable, but the reason we discuss this in the section about distributions is that we can do better and write

$$\Theta'(x - x_0) = \delta(x - x_0) \tag{2.21}$$

This is again an identity between distributions. Thus it has to hold for all (sufficiently decent) test functions. To prove the identity we can start from $\int \Theta'(x-x_0)\varphi(x)dx$ and use integration by parts or start from the integral representation (2.20) and perform the differentiation there. ⇒ lecture

Another relation between distributions we actually have already encountered is

$$\mathcal{P}\frac{1}{x - x_0} = \frac{1}{x - x_0 + i\varepsilon} + i\pi\,\delta(x - x_0) \tag{2.22}$$

This is nothing but (1.46) written without the test function $f(x)$. Note that in such expressions the limit $\varepsilon \to 0$ is understood (beware of the ultimate sin) but often not explicitly indicated. You will also often encounter (2.20) without the $\lim_{\varepsilon\to 0}$ explicitly written.

For our final example of an identity between distributions and an illustration of how to work with them we introduce the 3-dimensional δ function (leaving out the quotation marks from now on). It is defined as

$$\delta(\vec{r} - \vec{r}_0) = \delta(x - x_0)\delta(y - y_0)\delta(z - z_0) \tag{2.23}$$

where in Cartesian coordinates the vector $\vec{r}_0 = (x_0, y_0, z_0)$. Sometimes the notation $\delta^{(3)}$ is used. Using (2.16) we can write the δ function in spherical coordinates as

$$\delta(\vec{r} - \vec{r}_0) = \frac{1}{r^2 \sin\theta}\delta(r - r_0)\delta(\theta - \theta_0)\delta(\phi - \phi_0) \tag{2.24}$$

An identity that is very useful in the context of the Poisson equation $\Delta\phi(\vec{r}) = -4\pi\rho(\vec{r})$ (a more systematic treatment of partial differential equations will follow towards the end of MMP I) is

$$\Delta\left(\frac{1}{r}\right) = -4\pi\delta(\vec{r}) \tag{2.25}$$

Once more, this is a relation between distributions. The Laplace operator Δ acting on $1/r$ can trivially be shown to vanish for $r \neq 0$. To complete the proof of (2.25) we also need to show $\int \Delta(1/r)d^3\vec{r} = 1$ for an integral over a volume that contains the point $\vec{r} = 0$. ⇒ lecture

With the help of (2.25) we can now show that the solution of the Poisson equation for an arbitrary (charge distribution) $\rho(\vec{r})$ is given as ⇒ lecture

$$\phi(\vec{r}) = \int d^3\vec{r}\,' \frac{\rho(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} \qquad \Rightarrow \qquad \Delta\phi(\vec{r}) = -4\pi\rho(\vec{r}) \tag{2.26}$$

As an outlook for later sections we mention that $1/|\vec{r} - \vec{r}\,'|$ is the Green function of the Laplace equation.

3 Fourier series and integral transforms

3.1 Fourier series

We are familiar with the expansion of a vector $\vec{f}$ of an N dimensional vector space in a basis $\vec{e}_n$ as

$$\vec{f} = \sum_{n=1}^{N} c_n \, \vec{e}_n \tag{3.1}$$

For an orthonormal basis $\vec{e}_n \cdot \vec{e}_m \equiv \langle e_n | e_m \rangle = \delta_{nm}$ we have $c_n = \vec{f} \cdot \vec{e}_n \equiv \langle f | e_n \rangle$.

Now we make an attempt to generalise this to the case where the vector is a function $f(x)$. To start with, we consider a periodic function $f(x) = f(x + 2\pi)$. In practice we will simply restrict ourselves to $-\pi < x \leq \pi$ and also assume $f(x) \in \mathbb{R}$. More general cases will be considered later.

It is not surprising that such functions form a (real) vector space, as they can be multiplied by (real) numbers and added. But to generalise (3.1) we also need a scalar product. While we will consider this more carefully only in MMP II, from the Mathematics lectures you should be familiar with the scalar product

$$\vec{f} \cdot \vec{g} \qquad \rightsquigarrow \qquad \langle f | g \rangle = \int_{-\pi}^{\pi} f^*(x) g(x) dx \tag{3.2}$$

This looks very similar to (2.5) and we have left the complex conjugate sign in (3.2) even though it has no effect for a real function $f(x)$.

To generalise (3.1) to a 'vector' $\vec{f} \rightsquigarrow f(x)$ we have to realise that this vector space is infinite dimensional. It is convenient to label the basis vectors not as $n = 1 \ldots \infty$ but as $n = -\infty \ldots \infty$, or simply $n \in \mathbb{Z}$

The *Fourier series [Fourierreihe]* of a (periodic) function $f(x)$ is defined as the expansion of f in the basis

$$\vec{e}_n \;\rightsquigarrow\; e_n(x) \equiv \begin{cases} \frac{1}{\sqrt{2\pi}} & n = 0 \\ \frac{\sin(nx)}{\sqrt{\pi}} & n \in \mathbb{N} \\ \frac{\cos(nx)}{\sqrt{\pi}} & -n \in \mathbb{N} \end{cases} \tag{3.3}$$

This basis is orthonormal since ⇒ lecture

$$\langle e_n | e_m \rangle \rightsquigarrow \frac{1}{\pi} \int_{-\pi}^{\pi} \begin{cases} \sin(nx)\sin(mx)dx = \delta_{mn} \\ \cos(nx)\cos(mx)dx = \delta_{mn} \\ \sin(nx)\cos(mx)dx = 0 \end{cases} \qquad n, m = 1 \ldots \infty \tag{3.4}$$

and similar relations involving $e_0(x)$. Thus we are lead to generalise (3.1) by making the replacements $\vec{f} \rightsquigarrow f(x)$, $\vec{e}_n \rightsquigarrow e_n(x)$ and extending the sum from $n = -\infty \ldots \infty$ and write

$$f(x) = \sum_{n=-\infty}^{\infty} c_n \, e_n(x) \qquad \text{with} \qquad c_n = \int_{-\pi}^{\pi} f^*(x) \, e_n(x) dx \tag{3.5}$$

This was very brave! Actually, foolhardy is a better word. What does the = sign mean in the first equation of (3.5)? Does the series converge? In what sense? Can this be done (whatever 'this' is) for any (periodic) function $f(x)$?

While a more satisfactory answer will have to wait for MMP II, for the particular case of the Fourier series, these questions are answered as follows: We restrict ourselves to (periodic) functions $f(x)$ that satisfy the *Dirichlet conditions [Dirichlet Bedingungen]*. They state that we can split the range $-\pi < x \leq \pi$ into a finite number of intervals such that $f(x)$ is continuous and monotonic in each interval. In addition, the left-sided and right-sided limits f_- and f_+ exist at the boundaries of the intervals (and obviously also within each interval). Another way to put this is to say the function $f(x)$ has only a finite number of maxima and minima and at most a finite number of discontinuities in the interval $-\pi < x \leq \pi$. Then we have

$$f(x) \;\to\; \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(nx) + \sum_{n=1}^{\infty} b_n \sin(nx) \tag{3.6}$$

The arrow $\to$ indicates point-wise convergence to $f(x)$ if f is continuous at x. If we are unlucky enough to pick a point x where $f(x)$ is not continuous, the arrow indicates convergence to the average $(f_+(x) + f_-(x))/2$. This expression exists according to our assumptions.

The *Fourier coefficients [Fourierkoeffizient]* a_n and b_n in (3.6) are computed as

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)dx \qquad\qquad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(nx)dx \tag{3.7}$$

in accordance with the second equation in (3.5). Let us stress that the structure of (3.6) is precisely the one of (3.5). We simply set $c_0 = \sqrt{\pi/2}\,a_0$, $c_{n<0} = \sqrt{\pi}\,a_n$ and $c_{n>0} = \sqrt{\pi}\,b_n$. The expression (3.5) hints at a much more general structure and it is mainly for historic reasons that the Fourier series is often written as in (3.6).

As an example consider the function $f(x) = x$ for $-\pi < x \leq \pi$ and periodically extended for other values of x. This function is odd, $f(x) = -f(-x)$. Hence, all coefficients $a_n = 0$. For the b_n coefficients we get

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} x\,\sin(nx)dx = -\frac{2(-1)^n}{n} \tag{3.8}$$

However many terms we include in the series

$$f_N \equiv \sum_{n=1}^{N} \frac{-2(-1)^n}{n}\,\sin(nx) \tag{3.9}$$

we always have $f_N(\pi) = f_N(-\pi) = 0$. This is in agreement with (3.6), since at the point $x = \pi$ the function $f(x)$ has a discontinuity and $(f_-(\pi) + f_+(\pi))/2 = (\pi - (-\pi))/2 = 0$. For all other values of x, $-\pi < x < \pi$, the function $f_N(x)$ approaches $f(x) = x$, as illustrated in Figure 3.1. Note that the convergence is not uniform. The 'tails' at $x = \pm\pi$ do not get smaller for $N \to \infty$. ⇒ lecture

Any function $f(x)$ can be split into an even and odd function as

$$f(x) = f_{\text{even}}(x) + f_{\text{odd}}(x) \equiv \frac{1}{2}\Big(f(x) + f(-x)\Big) + \frac{1}{2}\Big(f(x) - f(-x)\Big) \tag{3.10}$$

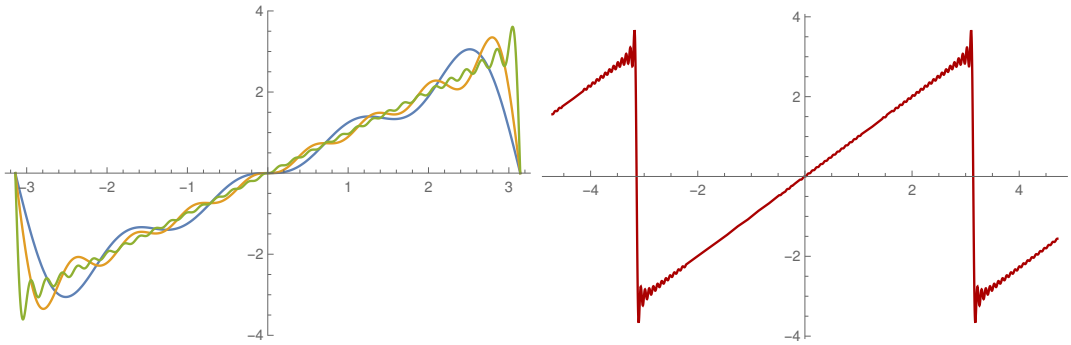

Figure 3.1: Fourier series of $f(x) = x$ taking into account 4 (blue), 8 (orange), 32 (green) and 80 (red) terms.

Only $a_n(b_n)$ will contribute to $f_{\text{even}}(f_{\text{odd}})$.

In (3.6) we have chosen the sine and cosine function as basis vectors/functions. Often it is more convenient to choose the functions e^{inx} as basis and re-write (3.6) as

$$f(x) = \sum_{n=-\infty}^{\infty} c_n\, e^{inx} = \sum_{n=-\infty}^{\infty} \sqrt{2\pi}\, c_n \frac{e^{inx}}{\sqrt{2\pi}} \tag{3.11}$$

The coefficients c_n in (3.11) are related to a_n and b_n of (3.6) as

$$c_0 = \frac{a_0}{2}; \qquad \text{and for } n > 0 \qquad c_n = \frac{a_n + ib_n}{2}; \qquad c_{-n} = \frac{a_n - ib_n}{2}; \tag{3.12}$$

The factors of $\sqrt{2\pi}$ ensure that the basis vectors form an orthonormal set

$$\int_{-\pi}^{\pi} dx \left(\frac{e^{imx}}{\sqrt{2\pi}}\right)^* \left(\frac{e^{inx}}{\sqrt{2\pi}}\right) = \delta_{mn} \tag{3.13}$$

The advantage of (3.11) is that it is typically simpler from an algebraic point of view and the distinction as in (3.3) is not required, whereas (3.6) is more obvious for disentangling the even and odd parts of a function. For a real function $f(x)$ the imaginary parts of $c_n\, e^{inx} + c_{-n}\, e^{-inx}$ cancel such that we do end up with a real expression despite the basis function being complex.

Returning to a N dimensional vector space with $\vec{f}$ as in (3.1), we know that the scalar product $\vec{f} \cdot \vec{f} = \sum_{n=1}^{N} |c_n|^2$. The generalization of this identity to $f(x)$ as in (3.6) or (3.11) is called *Parseval's identity* *[Parseval Identität]* and is simply the completeness relation

$$\langle f|f\rangle = \int_{-\pi}^{\pi} |f(x)|^2 dx = 2\pi \sum_{n=-\infty}^{\infty} |c_n|^2 = \frac{\pi a_0^2}{2} + \pi \sum_{n=1}^{\infty} (|a_n|^2 + |b_n|^2) \tag{3.14}$$

In some cases this relation can be used to compute particular sums. Applying (3.14) to the example $f(x) = x$ with the Fourier coefficients (3.8) we get

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{1}{4\pi} \int_{-\pi}^{\pi} x^2 dx = \frac{\pi^2}{6} \tag{3.15}$$

So far we have considered functions that have a period 2π. The generalization to an arbitrary period L, i.e. functions $f(x+L) = f(x)$ is obtained by minimal modifications to (3.6) or (3.11) as

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\left(\frac{2\pi n x}{L}\right) + \sum_{n=1}^{\infty} b_n \sin\left(\frac{2\pi n x}{L}\right) = \sum_{n=-\infty}^{\infty} c_n \, e^{\frac{2i\pi n x}{L}} \tag{3.16a}$$

with

$$a_n = \frac{2}{L}\int_{-L/2}^{L/2} f(x)\cos\left(\frac{2\pi n x}{L}\right)dx \qquad b_n = \frac{2}{L}\int_{-L/2}^{L/2} f(x)\sin\left(\frac{2\pi n x}{L}\right)dx \tag{3.16b}$$

and

$$c_n = \frac{1}{L}\int_{-L/2}^{L/2} f(x) e^{-\frac{2i\pi n x}{L}} dx$$

To conclude this subsection let us consider a function that is not periodic. As an example we take $f(x) = 1 + x^2$. We can approximate this function in the region $-L/2 < x < L/2$ by periodically extending $f(x)$.

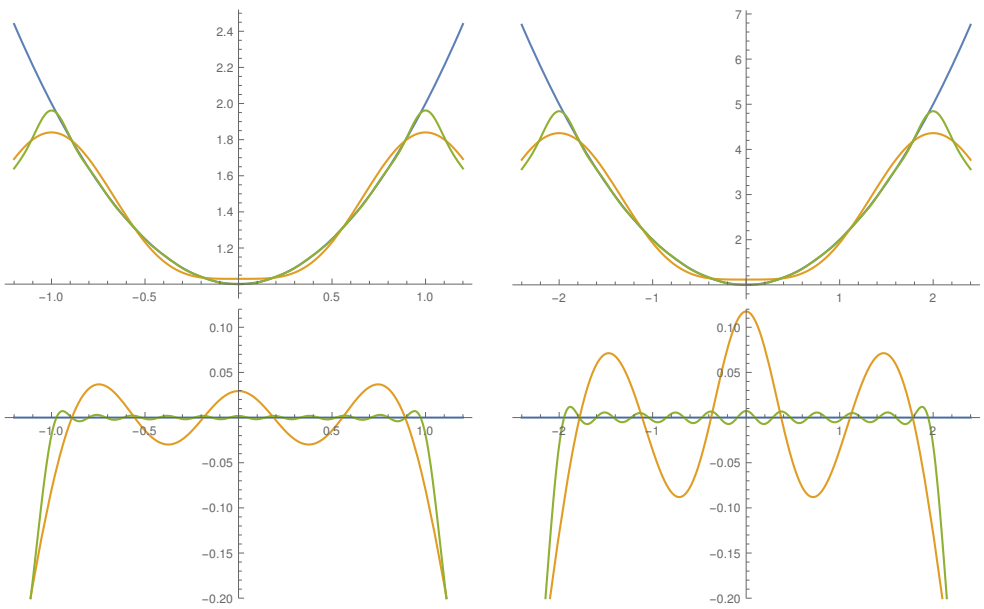

Figure 3.2: Fourier series of $f(x) = 1+x^2$ in the range $-1 < x < 1$ (left panel) and $-2 < x < 2$ (right panel) taking into account terms up to $n = 3$ (orange) and $n = 10$ (green), compared to the actual function (blue). The bottom panels show the relative difference between the Fourier series and the function.

Let us take $L = 2$ to start with. The coefficients $b_n = 0$ since $f(x)$ is even. For the other coefficients we obtain $a_0 = 2 + 2/3$ and $a_n = (-1)^n 4/(n\pi)^2$. The comparison of $f(x)$ and its Fourier series is shown in the left panel of Figure 3.2, taking into account terms up to $n = 3$ (orange) or $n = 10$ (green). As can be seen in the top-left panel, the Fourier series yields an ever better approximation of $f(x)$ in the range $-1 < x < 1$, but breaks down completely for $|x| > 1$. This is also confirmed in the bottom-left panel, where the relative difference of $f(x)$ and its Fourier series is depicted. If we want a good approximation for a larger range

of x, say $-2 < x < 2$, we just repeat the procedure for $L = 4$. Then we get $a_0 = 2 + 8/3$ and $a_n = (-1)^n 16/(n\pi)^2$ and $f(x)$ is well approximated by its Fourier series in the range $-2 < x < 2$, as shown in the right panel of Figure 3.2. However, more terms in the series are needed to obtain the same relative accuracy if L is increased. This is evident from comparing the relative differences for $L = 1$ (left) and $L = 2$ (right) in the bottom panel of Figure 3.2.

3.2 Fourier transform

We now turn to an apparently completely unrelated topic, namely the *Fourier transform [Fouriertransformation]*. Whereas the Fourier series expands a function in terms of basis functions, the Fourier transform is a special case of an integral transform. Integral transforms $\mathcal{I}$ take a function $f(x)$ and transform it into another function $\tilde{f}$ of another variable k

$$f(x) \to \mathcal{I}\big[f(x)\big] = \tilde{f}(k) \equiv \int_a^b f(x)K(x,k)dx \tag{3.17}$$

The *kernel [Kern]* $K(x,k)$ of the integral transform together with the integration boundaries a and b define the integral transform. Of course, we implicitly assume the integral in (3.17) exists. We can interpret $\mathcal{I}$ as an operator acting on suitable functions. This operator is linear, i.e. for two functions f_i and constants λ_i it satisfies

$$\mathcal{I}\big[\lambda_1 f_1(x) + \lambda_2 f_2(x)\big] = \lambda_1 \mathcal{I}\big[f_1(x)\big] + \lambda_2 \mathcal{I}\big[f_2(x)\big] \tag{3.18}$$

The special case of the Fourier transform $\mathcal{F}$ is obtained by taking $K(x,k) = e^{-ikx}/\sqrt{2\pi}$ as

$$f(x) \to \mathcal{F}\big[f(x)\big] = \hat{f}(k) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x)e^{-ikx}dx \tag{3.19}$$

where we use the short hand notation $\hat{f}$ rather than $\tilde{f}$ and we assume $\int_{-\infty}^{\infty} |f(x)|dx$ is finite, i.e. $f \in L^1(\mathbb{R})$. In the literature different conventions regarding the prefactor $1/\sqrt{2\pi}$ exist and (3.19) is also often written with variables $x \to t$ and $k \to \omega$.

If we start with a function $f(x)$ and Fourier transform it to $\hat{f}(k)$, how can we get back $f(x)$? The inverse Fourier transform in given by

$$\hat{f}(k) \to \mathcal{F}^{-1}\big[\hat{f}(k)\big] = f(x) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{f}(k)e^{+ikx}dk \tag{3.20}$$

This looks very similar to the Fourier transform itself, apart from the sign in the exponential. To convince ourselves of this we consider the inverse Fourier transform of the Fourier transform of $f(x_0)$ and get

$$\begin{aligned} \mathcal{F}^{-1}\Big[\mathcal{F}\big[f(x)\big]\Big] &= \frac{1}{2\pi} \int_{-\infty}^{\infty} dk \Big[\int_{-\infty}^{\infty} dx_0 f(x_0)e^{-ikx_0}\Big]e^{+ikx} \\ &\overset{(2.13)}{=} \int_{-\infty}^{\infty} dx_0 f(x_0)\delta(x - x_0) = f(x) \end{aligned} \tag{3.21}$$

where we have used (2.13) in the second step.

Before we consider some properties of the Fourier transform and give some examples let us comment on the connection between the Fourier series and the Fourier transform. If we define $k_n = 2\pi n/L$ and $\Delta k = 2\pi/L$ the distance between two k_n we can write (3.16a) as

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{ik_n x} \qquad \text{with} \qquad c_n = \frac{\Delta k}{2\pi} \int_{-L/2}^{L/2} f(x_0) e^{-ik_n x_0} dx_0 \tag{3.22}$$

Taking the continuum limit $L \to \infty$ and $\sum \Delta k \to \int dk$ this reads as

$$f(x) = \sum_{n=-\infty}^{\infty} \frac{\Delta k}{2\pi} \int_{-L/2}^{L/2} f(x_0) e^{-ik_n x_0} dx_0 \, e^{ik_n x} \quad \to \frac{1}{2\pi} \int_{-\infty}^{\infty} dk \int_{-\infty}^{\infty} dx_0 f(x_0) e^{-ikx_0} e^{+ikx} \tag{3.23}$$

and reproduces (3.21). Thus, somewhat naively speaking, we can view the Fourier series in the limit $L \to \infty$ as the inverse Fourier transform of the Fourier transform, i.e. reproducing the function itself.

As mentioned in connection with (3.18) the Fourier transform is linear. Furthermore, it has the properties ⇒ lecture

$$\mathcal{F}\big[f(ax)\big] = \frac{1}{|a|} \hat{f}\Big(\frac{k}{a}\Big) \tag{3.24a}$$

$$\mathcal{F}\big[e^{ik_0 x} f(x)\big] = \hat{f}(k - k_0) \tag{3.24b}$$

$$\mathcal{F}\big[f'(x)\big] = ik\hat{f}(k) \tag{3.24c}$$

$$\langle \hat{g}(k) | \hat{f}(k) \rangle = \langle g(x) | f(x) \rangle \tag{3.24d}$$

where a is a constant. The last property stating that the scalar product as defined in (3.2) is invariant under a Fourier transform has huge importance in Quantum Mechanics. Also (3.24c) is remarkable in that derivatives are transformed into multiplications.

Another structure that is simplified through a Fourier transform is a *convolution [Faltung]* of two functions $f(x)$ and $g(x)$, defined as

$$h(x) = (f * g)(x) \equiv \int_{-\infty}^{\infty} f(x_0) g(x - x_0) dx_0 \tag{3.25}$$

Convolutions show up in many different contexts. As a particularly important example we mention a measurement (or more general a signal processing) device. Let's assume the true value to be measured (or input signal) is $f(x_0)$. Due to imperfections, the measurement apparatus produces an output $h(x)$ as in (3.25). The resolution function $g(x - x_0)$ is characteristic of the device. For the perfect device it is $\delta(x - x_0)$. Indeed, (2.10) can be seen as the convolution of the test function (true value to be measured) φ with the δ 'function'. In this case the device produces precisely the desired output $h(x) = f(x)$. However, in reality a measurement (or signal processing) device has a more smeared resolution function, as shown e.g. in the left panel of Figure 2.1. If the device is accurate (but not necessarily precise) the resolution function will be centred around $x - x_0 = 0$. It could also introduce a bias (i.e. not be accurate) if it is not centred at $x - x_0 = 0$.

The important point is that if the resolution function is known, we can recover the true value $f(x)$ from the measured value $h(x)$. Under a Fourier transform (3.25) becomes ⇒ lecture

$$\hat{h}(k) = \mathcal{F}\big[h(x)\big] = \mathcal{F}\big[(f * g)(x)\big] = \sqrt{2\pi}\,\hat{f}(k)\hat{g}(k) \tag{3.26}$$

Thus we can recover

$$f(x) = \frac{1}{\sqrt{2\pi}}\,\mathcal{F}^{-1}\Big[\hat{h}(k)/\hat{g}(k)\Big] \tag{3.27}$$

In practice this is often hampered by incomplete knowledge of $g(x)$ and $h(x)$.

Now it is time for an example. We consider the function

$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ a\,e^{-\lambda x} & \text{for } x \geq 0 \end{cases} \tag{3.28}$$

where a and $\lambda > 0$ are real constants. According to (3.19) the Fourier transform is calculated as

$$\hat{f}(k) = \frac{a}{\sqrt{2\pi}} \int_0^\infty e^{-\lambda x} e^{-ikx} dx = \frac{-ia}{\sqrt{2\pi}(k - i\lambda)} \tag{3.29}$$

The obtain the inverse Fourier transform of $\hat{f}(k)$ we have to do the integral

$$\mathcal{F}^{-1}\big[\hat{f}(k)\big] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty \frac{-ia}{\sqrt{2\pi}(k - i\lambda)} e^{ikx}\, dk = \frac{a}{2i\pi} \int_{-\infty}^\infty \frac{e^{ikx}}{k - i\lambda}\, dk \tag{3.30}$$

The last integral is crying out loud for the residue theorem. The function has a pole at $k = +i\lambda$ in the upper half of the k plane. For $x > 0$ we can close the contour in the upper half plane without getting a contribution form the semi circle. According to the residue theorem we then get $\mathcal{F}^{-1}\big[\hat{f}(k)\big] = a\,e^{-\lambda x}$ for $x > 0$. For $x < 0$ we can close the contour in the lower half plane without getting a contribution form the semi circle and end up with $\mathcal{F}^{-1}\big[\hat{f}(k)\big] = 0$ for $x < 0$, in agreement with (3.28). At this point we also make the obvious remark that the Fourier transform deals with complex functions. Even if $f(x)$ happens to be real, as (3.28), $\hat{f}(k)$ is in general a complex function of k, see (3.29). While x and k are real to start with, it is often useful to think of them as complex variables, as illustrated in (3.30). Thus, everything we have learned in Section 1 comes in handy. ⇒ lecture

To illustrate another feature of the Fourier transform let us consider the example

$$f_L(x) = \frac{1}{\sqrt{L}}\Theta\Big(\frac{L}{2} - |x|\Big) \qquad \rightarrow \qquad \hat{f}_L(k) = \sqrt{\frac{2}{\pi L}}\,\frac{\sin\big(kL/2\big)}{k} \tag{3.31}$$

The function $f_L(x)$ is normalized as $\langle f(x)|f(x)\rangle = \int_{-\infty}^\infty |f_L(x)|^2 dx = 1$ and in accordance with (3.24d) we also have $\int_{-\infty}^\infty |\hat{f}_L(k)|^2 dk = 1$. However, for $L \to 0$ the function $f_L(x)$ becomes ever more peaked around $x \simeq 0$. The Fourier transform $\hat{f}_L(k)$ has just the opposite property in that it gets ever broader for $L \to 0$. This is illustrated in Figure 3.3 where $f_L(x)$ and $\hat{f}_L(k)$ are shown for the values $L \in \{1, 3, 9\}$. This is a generic property of the Fourier transform. Its most extreme case is the Fourier transform of a constant $f(x) = a$ (not localized at all) being a δ function (maximally localized),

$$\mathcal{F}\big[a\big] = \frac{a}{\sqrt{2\pi}}\,\delta(k) \tag{3.32}$$

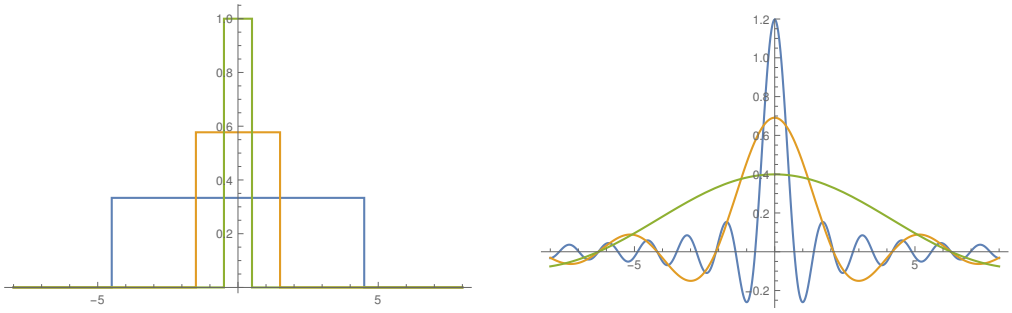

Figure 3.3: The function $f_L(x)$ (left panel) and its Fourier transform $\hat{f}_L(k)$ (right panel) as defined in (3.31) for $L = 1$ (green), $L = 3$ (orange), and $L = 9$ (blue).

So far we have considered the one dimensional Fourier transform. We can extend this to three dimensions, with a function $f(\vec{r}) = f(x, y, z) = f(r, \varphi, \theta)$, where $\vec{r} \in \mathbb{R}^3$ is a three-dimensional vector with cartesian coordinates (x, y, z) and spherical coordinates (r, φ, θ). Accordingly, the Fourier transform of $f(\vec{r})$ will also depend on three variables $\vec{k} = (k_x, k_y, k_z)$. It is defined by generalizing (3.19) to

$$f(\vec{r}) \to \mathcal{F}\big[f(\vec{r})\big] = \hat{f}(\vec{k}) \equiv \frac{1}{(2\pi)^{3/2}} \int_{-\infty}^{\infty} f(\vec{r}) e^{-i\vec{k}\cdot\vec{r}} d^3\vec{r} \tag{3.33}$$

We are often concerned with the special case of a rotationally invariant function $f(r)$, i.e. a function that does not depend individually on all components of $\vec{r}$ but only on its magnitude r. In this case (3.33) simplifies to

$$f(r) \to \mathcal{F}\big[f(r)\big] = \hat{f}(k) \equiv \frac{1}{(2\pi)^{3/2}} \int_0^{\infty} dr\, r^2 f(r) \int d\Omega_2\, e^{-i\vec{k}\cdot\vec{r}} \tag{3.34}$$

where $d\Omega_2 = \sin\theta\, d\theta\, d\phi = d\cos\theta\, d\phi$ denotes the angular integration. In such a case the Fourier transform also depends only on the magnitude k and not on the full $\vec{k}$. In three dimensions we can evaluate the angular integrals in (3.34) as ⇒ lecture

$$\hat{f}(k) = \frac{1}{(2\pi)^{3/2}} \int_0^{\infty} dr\, r^2 f(r)\, 4\pi \Big(\frac{\sin(kr)}{kr}\Big) \tag{3.35}$$

Extension from 3 to n dimensions with $\vec{r} = (r_1 \dots r_n)$ and $\vec{k} = (k_1 \dots k_n)$ leads us to consider expressions like

$$G_n(k) = \int d\Omega_{n-1}\, e^{-i\vec{k}\cdot\vec{r}} \tag{3.36}$$

As a teaser for the sections on differential equations we note that a (somewhat unconventional) way to solve these integrals is to transform them into a differential equation and solve the latter. If we act with the Laplace operator on both sides of (3.36) we get

$$\Delta G_n(k) = \Big(\frac{\partial^2}{\partial k_1^2} + \frac{\partial^2}{\partial k_2^2} \dots \frac{\partial^2}{\partial k_n^2}\Big) G_n(k) = \int d\Omega_{n-1}\, (-r^2) e^{-i\vec{k}\cdot\vec{r}} = -r^2 G_n(k) \tag{3.37}$$

Since $G_n(k)$ depends only on $k = |\vec{k}|$ we can transform the partial derivatives in (3.37) to normal derivatives w.r.t k and setting $r = 1$ (for simplicity) we find that $G_n(k)$ has to satisfy the differential equation

$$G_n'' + \frac{n-1}{k} G_n' + G_n = 0 \tag{3.38}$$

This is an ordinary differential equation that is closely related to one we will get to know and love in Section 5. The latter is called the Bessel differential equation and its solutions are, surprise, surprise, the Bessel functions. Thus, $G_n(k)$ can be expressed in terms of Bessel functions.

Normally it is more difficult to solve a differential equation than to do an integral. However, in this particular case the differential equation is well known. Another intriguing feature is that n in (3.36) indicates the number of integrations, whereas n in (3.38) is simply a parameter. So who says n needs to be an integer? Can we integrate n times with $n = 1.623$? Weird and wonderful!

3.3 Laplace transform

The Laplace transform is another integral transform (3.17), defined as

$$\mathcal{L}\big[f(x)\big] = \tilde{f}(s) \equiv \int_0^\infty f(x) e^{-xs}\, dx \tag{3.39}$$

The differences to (3.19) are (apart from notation and boundary condition) that the kernel $K(x,s) = e^{-xs}$ is real for $s \in \mathbb{R}$. Hence, the Laplace transform of a real function $f(x)$, if it exists, is a real function $\tilde{f}(s)$. For $s > 0$ there is an exponential damping due to the kernel. This is often sufficient to ensure the existence of the integral in (3.39). As an example we consider $f(x) = e^{\beta x}$ and obtain for its Laplace transform

$$\mathcal{L}\big[e^{\beta x}\big] = \int_0^\infty e^{x(\beta - s)}\, dx = \frac{1}{s-\beta} \tag{3.40}$$

where the last step holds for $\beta < s$. Another example is connected to the Γ function that made a short appearance in Section **??**.

$$\mathcal{L}\big[x^n\big] = \int_0^\infty x^n e^{-xs}\, dx = \frac{1}{s^{n+1}} \int_0^\infty t^n e^{-t}\, dt \equiv \frac{\Gamma(n+1)}{s^{n+1}} \tag{3.41}$$

where we made a change of variables $xs = t$ and in the last step we used the definition of the Γ function. For (3.41) to hold, we need $s > 0$ and also $n > -1$. Otherwise the integral diverges due to the region $x \to \infty$ or $x \to 0$. In fact, $\Gamma(z)$ has a pole at $z = 0$ (and all negative integers) which means that $\Gamma(n+1)$ is not defined for $n = -1$.

Since the Laplace transform is rather similar to the Fourier transform it is not surprising that identities similar to (3.24) hold. We have ⇒ lecture

$$\mathcal{L}\big[x^n f(x)\big] = (-1)^n \frac{d^n}{ds^n} \tilde{f}(s) \tag{3.42a}$$

$$\mathcal{L}\big[e^{ax} f(x)\big] = \tilde{f}(s-a) \tag{3.42b}$$

$$\mathcal{L}\big[f'(x)\big] = s\tilde{f}(s) - f(0) \tag{3.42c}$$

and if we define a new convolution similar to (3.25) as

$$(f \star g)(x) \equiv \int_0^x f(x_0) g(x - x_0) dx_0 \tag{3.43}$$

but with different integration boundaries, we have

$$\mathcal{L}\big[(f \star g)(x)\big] = \mathcal{L}\big[f(x)\big]\mathcal{L}\big[g(x)\big] = \tilde{f}(s)\,\tilde{g}(s) \tag{3.44}$$

in analogy to (3.26). It is possible to generalize (3.42c) to the nth derivative as

$$\mathcal{L}\big[f^{(n)}(x)\big] = s^n \tilde{f}(s) - \sum_{k=0}^{n-1} s^k f^{(n-k-1)}(0) \tag{3.45}$$

A typical application of the Laplace transform is in connection with solving differential equations. Since we will discuss ordinary differential equations in the next two sections, we simply illustrate this here with an example. We want to find a function $f(x)$ that satisfies the differential equation ⇒ lecture

$$f'''(x) - 3f''(x) + 4f(x) = 1 \tag{3.46}$$

and has the boundary conditions $f(0) = 1$, $f'(0) = 0$ and $f''(0) = 0$. Applying the Laplace transform to (3.46) turns the differential equation in the variable x into an algebraic equation in the variable s

$$s^3 \tilde{f}(s) - s^2 - 3(s^2 \tilde{f}(s) - s) + 4\tilde{f}(s) = \frac{1}{s} \tag{3.47}$$

where we have used (3.45) with $n = 3$ and $n = 2$ together with the boundary conditions of $f(x)$. For the r.h.s. we applied (3.41) with $n = 0$. The solution to (3.47) is trivially found to be

$$\tilde{f}(s) = \frac{s^3 - 3s^2 + 1}{(s-2)^2 s(s+1)} = \frac{1}{4s} + \frac{1}{3(s+1)} + \frac{5}{12(s-2)} - \frac{1}{2(s-2)^2} \tag{3.48}$$

All that remains to be done now is to perform the inverse Laplace transform of the solution given in (3.48). To facilitate this we used partial fraction in the second step. In fact, staring at (3.41) and (3.42b) reveals ⇒ lecture

$$f(x) = \frac{1}{4} + \frac{1}{3}\,e^{-x} + \frac{5}{12}\,e^{2x} - \frac{1}{2}\,xe^{2x} \tag{3.49}$$

In this particular example we managed to guess the inverse Laplace transform by using a dictionary. While this might work in simple cases, we clearly need a general expression for the inverse transform. It is given as

$$\mathcal{L}^{-1}\big[\tilde{f}(s)\big] = f(x) = \frac{1}{2i\pi} \int_{\lambda - i\infty}^{\lambda + i\infty} e^{sx} \tilde{f}(s)\, ds \tag{3.50}$$

through a contour integral in the complex s plane, with a contour parallel to the imaginary axis with an offset λ such that all singularities of $\tilde{f}(s)$ are to the left of the contour. To show this we follow a similar procedure as in (3.21) and show that the inverse Laplace transform

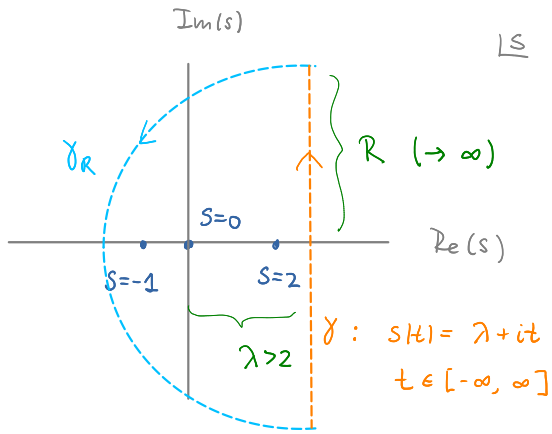

Figure 3.4: Poles of the function $\tilde{f}(s)$, (3.48), in the complex s plane and contour (orange) to be used for inverse Laplace transform (3.51).

as defined in (3.50) of the Laplace transform of $f(x)$, as defined in (3.39) reproduces $f(x)$, i.e.

$$\mathcal{L}^{-1}\Big[\mathcal{L}\big[f(x)\big]\Big] = \frac{1}{2i\pi}\int_{\lambda-i\infty}^{\lambda+i\infty} ds\, e^{sx} \int_0^\infty f(x')e^{-x's}\, dx' = f(x) \tag{3.51}$$

This is a lovely application of what we have learned in Section 1. ⇒ lecture

Applying (3.50) to (3.48) we note that $\tilde{f}(s)$ has poles at $s = 0$, $s = -1$ and $s = 2$. Thus we choose $\lambda > 2$. In addition, we note that $\tilde{f}(s) \to 0$ for $|s| \to \infty$. Thus, we can evaluate the integral in (3.50) by closing the contour with a circle of radius R in the left half plane (see Figure 3.4) and let $R \to \infty$. The integral along the (light blue) half circle γ_R vanishes for $R \to \infty$ and we can evaluate (3.50) using the residue theorem as ⇒ lecture

$$\begin{aligned} f(x) &= \operatorname*{res}_{s=0}\big(e^{sx}\tilde{f}(s)\big) + \operatorname*{res}_{s=-1}\big(e^{sx}\tilde{f}(s)\big) + \operatorname*{res}_{s=2}\big(e^{sx}\tilde{f}(s)\big) \\ &= \frac{1}{4} + \frac{1}{3}\,e^{-x} + \frac{5-6x}{12}\,e^{2x} \end{aligned} \tag{3.52}$$

in agreement with (3.49). Thus, in words (3.50) states that taking the inverse Laplace transform is nothing but summing up all residues of $e^{sx}\tilde{f}(s)$ and making sure not to forget one (the choice of λ). Isn't this quite simply gorgeous?

4 Ordinary differential equations

This section deals with *ordinary differential equations* (ODE) *[gewöhnliche Differentialgleichung]*. What is an ODE? First and foremost it is an equation, the 'E' part, and it is implicitly understood that we want to solve it. The unknown y is not a variable, but a function $y(x)$. The equation contains the function itself but also its derivatives y', $y''\ldots$ w.r.t. the argument x, hence the 'D'. The 'O' signifies that only ordinary derivatives are involved and no partial derivatives appear. In other words, we can think of the function $y(x)$ as a function of a single variable x. Often the term *dependent variable* is used for y and *independent variable* for x. If you are saddened by the lack of partial derivatives, don't worry. They will appear in Sections 6 and 7, where we consider partial differential equations (PDE).

Needless to say that Physics is littered with ODE (and PDE). In fact even the most beloved equation $F = m\,a$ is nothing but an ODE. Using x and t for the dependent and independent variables, we have $x''(t) = F(t, x, x', (x')^2 \ldots)/m$, where the force F might depend on time t, the position x, the velocity (squared) x' and $(x')^2$ an who knows what else. This is a second order ODE (the highest derivative it contains is the second derivative) and to fix $x(t)$ completely, we need two initial conditions, typically $x(t = 0)$ and $x'(t = 0)$.

4.1 First order ODE

We start by considering first order ODE, i.e. differential equations where only the function $y(x)$ and its first derivative $y'(x)$ appear. In particular, we consider a generic form of such an ODE, namely

$$\frac{dy(x)}{dx} \equiv y' = f(y, x) \tag{4.1}$$

where $f(y, x)$ is an arbitrary function of the dependent variable y and the independent variable x. In what follows we consider several important special cases for the function f.

The first special case is a *separable [separierbar]* ODE, where we can write

$$\text{separable:} \qquad y' = f(y, x) = g(x)h(y) \tag{4.2}$$

In other words, the function $f(y(x), x)$ can be separated into a product of two functions g and h that each depend only on one of the two variables. Then we can write

$$\frac{dy}{h(y)} = g(x)dx \qquad \rightarrow \qquad \int \frac{dy}{h(y)} = \int g(x)dx \tag{4.3}$$

and simply solve (4.2) through integration. Let's illustrate this with a concrete example. We want to solve ⇒ lecture

$$xy' - y = 3 \tag{4.4}$$

which corresponds to (4.2) with $h(y) = 3 + y$ and $g(x) = 1/x$. Using (4.3) with an indefinite integral we get the *general solution [allgemeine Lösung]*

$$\log(3 + y) + c = \log(x) \qquad \rightarrow \qquad y(x) = k\,x - 3 \tag{4.5}$$

where c and $k = e^{-c}$ are constants that appear due to the indefinite integration. Plugging (4.5) back into (4.4) we can verify that we have found a solution. In fact, we have found infinitely many solutions! To obtain a unique solution, i.e. fix the single constant of integration, c or k, we need a single *initial condition [Anfangsbedingung]* or *boundary condition* (recall $F = m\,a$ where we needed two initial conditions, since we had a second order ODE). Assuming we want to solve (4.4) with $y(1) = 2$, we can simply take the general solution and determine $k = 5$ to fulfill the initial condition. The we obtain the *particular solution [spezielle Lösung]* of (4.4) with $y(1) = 2$ as $y(x) = 5\,x - 3$. We can also directly include the boundaries in the integration and modify (4.3) to

$$\int_{y(1)=2}^{y} \frac{dy}{3+y} = \int_{x=1}^{x} \frac{1}{x} dx \qquad \rightarrow \qquad y(x) = 5\,x - 3 \tag{4.6}$$

We were lucky. (4.4) is a linear ODE, i.e. it does not contain terms like y^2 or $y\,y'$. If such terms appear, we can still use (4.3) but we might miss *singular solutions [singuläre Lösungen]*. Consider another example, ⇒ lecture

$$y' = \sqrt{1-y^2} \qquad \xrightarrow{(4.3)} \qquad y(x) = \sin(x-k) \tag{4.7}$$

Indeed, we find some solutions using (4.3), but we missed some others, namely $y(x) = 1$ and $y(x) = -1$. Non-linear equations like (4.7) are treacherous.

Having had our fingers burnt with non-linear ODE, let's consider a general linear first order ODE

$$\text{linear:} \qquad y' + p(x)\,y = q(x) \tag{4.8}$$

where p and q are arbitrary functions of the independent variable x but do not depend on the dependent variable y. To start with, we note that the *homogeneous [homogene]* version of (4.8), that is the version with $q = 0$, is separable and can be solved as

$$y_h(x) = c\,\exp\Big(-\int p(x)dx\Big) \tag{4.9}$$

The function $y_h(x)$ is called the *complementary function [Komplementärfunktion]*. It contains a constant c and is the general solution to (4.8) with $q = 0$. To obtain the general solution to the inhomogeneous problem, i.e. (4.8) with $q(x)$ not set to zero, we need a single *particular [partikuläre]* solution $y_p(x)$. If we have found one, then

$$y(x) = y_h(x) + y_p(x) \tag{4.10}$$

is the general solution of (4.8) as can be explicitly verified. ⇒ lecture

This leaves us with the question of how to find a particular solution. An option that should not be underestimated is guessing a solution. After all, by simply plugging in it is very easy to verify whether a certain function is a solution of the ODE or not. For cases, where guessing is not bearing fruits, we present two more systematic possibilities.

The first is called *variation of constant [Variation der Konstanten]*. We make an ansatz by promoting the constant c above to a function $k(x)$. In other words, we are looking for a solution to (4.8) in the form $y_p(x) = k(x)y_h(x)$. Plugging this ansatz into (4.8) we find a condition for $k(x)$, namely ⇒ lecture

$$k(x) = \int dx \frac{q(x)}{y_h(x)} \tag{4.11}$$

Thus, we obtain a particular solution and, hence, the general solution by 'simply' doing the integral given in (4.11).

The second possibility is to use the Laplace transform. We have already encountered the usefulness of the Laplace transform for solving ODE in Section 3.3.

Let us conclude the discussion on the linear first order ODE with an example. We want to find the general solution to

$$y' = x - y \tag{4.12}$$

This corresponds to (4.8) with $p(x) = 1$ and $q(x) = x$. We first solve the homogeneous problem $y' = -y$ and find for the complementary solution $y_h(x) = c\,e^{-x}$. To find a particular solution we illustrate both options. First, with variation of the constant we obtain from (4.11)

$$k(x) = \int dx \frac{x}{e^{-x}} = (x-1)e^x \qquad \rightarrow \qquad y_p(x) = x - 1 \tag{4.13}$$

Alternatively, using the Laplace transform on both sides of (4.12) we find ⇒ lecture

$$s\,\tilde{y}(s) - y(0) = \frac{1}{s^2} - \tilde{y}(s) \qquad \rightarrow \qquad \tilde{y}(s) = \frac{1}{s^2(s+1)} + \frac{y(0)}{s+1} \tag{4.14}$$

and after the inverse Laplace transform we get

$$y_p = x - 1 + e^{-x}\big(1 + y(0)\big) \tag{4.15}$$

Note that the two particular solutions found in (4.13) and (4.15) are different. However, combined with the complementary solution $y_h(x) = c\,e^{-x}$ as in (4.10) they both lead to the same general solution

$$y(x) = y_h + y_p = (x-1) + c\,e^{-x} \tag{4.16}$$

And as a final remark related to this example, guessing the particular solution might well have worked in this case!

To illustrate the potential use of variable transformations we consider the

$$\text{Bernoulli ODE}: \qquad y' + p(x)y = q(x)\,y^n \tag{4.17}$$

In general, (4.17) is neither separable nor linear. However, if we set $z(x) = y^{1-n}(x)$ the ODE (4.17) can be transformed to a liner ODE for $z(x)$,

$$z' + (1-n)p(x)z = (1-n)q(x) \tag{4.18}$$

This ODE can be solved with the methods presented above. Of course, as with integration, finding a useful variable transformation is a bit of black magic.

4.2 Second order ODE

It won't come as a surprise that second order ODE are generally speaking more complicated than first order ODE. Hence, in this section we will focus mainly on linear ODE and only at the end mention a few special cases of non-linear second order ODE.

Let us start with a linear, homogeneous ODE

$$y'' + p(x)\, y' + q(x) y = 0 \tag{4.19}$$

with arbitrary functions $p(x)$ and $q(x)$. What is new compared to the homogeneous version of (4.8) is that now we have two linearly independent solutions, $y_1(x)$ and $y_2(x)$ and the general solution to (4.19) is given as a linear combination

$$y(x) = c_1\, y_1(x) + c_2\, y_2(x) \tag{4.20}$$

Consequently, we need two initial conditions to fix the two constants c_1 and c_2, as already alluded to in connection with $F = m\, a$. As a solace to having to find two solutions we remark that if one is known, e.g. $y_1(x)$, the second one can be found as follows. We plug in the ansatz $y_2(x) = v(x) y_1(x)$ into (4.19) and obtain ⇒ lecture

$$v'' y_1 + 2 v' y_1' + p v' y_1 = 0 \qquad \rightarrow \qquad w' = -\Big(\frac{2 y_1'}{y_1} + p\Big) w \tag{4.21}$$

where for the second step we have set $w(x) \equiv v'(x)$. Thus we end up with a separable ODE for $w(x)$ that we can solve with the method of Section 4.1. Once we have $w(x)$ we get the second linearly independent solution to (4.19) as

$$v(x) = \int dx\, w(x) \qquad \rightarrow \qquad y_2(x) = v(x) y_1(x) \tag{4.22}$$

This leaves us with the problem to find one solution in the first place. For the very important special case of an ODE with

$$\text{constant coefficients} \qquad y'' + a_1\, y' + a_0 y = 0 \tag{4.23}$$

this can be done by educated guessing, or making a clever ansatz. In fact, if a_1 and a_0 do not depend on x and we plug $y(x) = e^{\lambda x}$ into (4.23) we obtain

$$\lambda^2 + a_1\, \lambda + a_0 = 0 \tag{4.24}$$

Derivatives have been replaced by powers of λ. It happens often that a clever ansatz transforms an ODE into an algebraic equation that is much easier to solve. (4.24) is called the *characteristic equation [charakteristische Gleichung]*. It is a quadratic equation in λ and has two solutions λ_1 and λ_2. If $\lambda_1 \neq \lambda_2$, the general solution (4.20) for constant coefficients reads

$$\lambda_1 \neq \lambda_2 : \qquad y(x) = c_1\, e^{\lambda_1 x} + c_2\, e^{\lambda_2 x} \tag{4.25}$$

If it happens that $\lambda_1 = \lambda_2$ we only get one independent solution. To get the other, we use the method described in (4.21) and (4.22) and find ⇒ lecture

$$\lambda_1 = \lambda_2 : \qquad y(x) = c_1\, e^{\lambda_1 x} + c_2\, x\, e^{\lambda_1 x} \tag{4.26}$$

The generalization to the inhomogeneous case

$$y'' + p(x)\, y' + q(x) y = f(x) \tag{4.27}$$

follows a similar pattern as in the case of first order ODE. We need to find a single particular solution $y_p(x)$ to (4.27) and then the general solution is given by a generalization of (4.10) as

$$y(x) = y_h(x) + y_p(x) = c_1\, y_1(x) + c_2\, y_2(x) + y_p(x) \tag{4.28}$$

Again we are faced with the question on how to find a particular solution y_p. As for the first order case, there are several possibilities.

Let us start with 'guessing' or making a clever ansatz. This is promising in the case of constant coefficients (4.23) if $f(x)$ has a simple or special form. For example, if $f(x)$ is a combination of exponential and polynomials with constants b_i, then the ansatz

$$f(x) = e^{qx}(b_m x^m + b_{m-1} x^{m-1} \ldots) \qquad \rightarrow \qquad y_p = e^{qx}(d_m x^m + d_{m-1} x^{m-1} \ldots d_0) \tag{4.29}$$

will transform the ODE to an algebraic equation. Before deploying heavy artillery it is a good idea to check if you get away with a clever ansatz.

If this does not work, we can again use variation of constants. However, it is now more involved since we have two constants c_1 and c_2 that need to be promoted to functions $k_1(x)$ and $k_2(x)$. If we set

$$y_p(x) = k_1(x) y_1(x) + k_2(x) y_2(x) \tag{4.30}$$

and plug this into (4.27), we can show that the functions $k_1(x)$ and $k_2(x)$ satisfy the first order ODE ⇒ lecture

$$k_1' = \frac{-f\, y_2}{y_1\, y_2' - y_1' y_2} \qquad \text{and} \qquad k_2' = \frac{f\, y_1}{y_1\, y_2' - y_1' y_2} \tag{4.31}$$

This can be solved $k_i(x) = \int k_i'(x) dx$ and we get y_p from (4.30).

Of course, we also have again the option to use the Laplace transform, as in Section 4.1. For constant coefficients, this transforms the ODE in x into an algebraic equation in s. The latter can be solved and we use the inverse Laplace transform to obtain the solution to the ODE. In this context we remark that the Laplace transform is less useful if the functions a_i actually do depend on x. From (3.42a) we see that powers of x transform to derivatives w.r.t. s. Thus taking the Laplace transform in such cases typically transforms on ODE in x into an ODE in s.

All of a sudden guessing seems very attractive, doesn't it? Now if guessing doesn't quite work, we can try to enforce it. To do so we use the *superposition principle [Superpositions-prinzip]*. It states that if we split $f(x) = f_1(x) + f_2(x)$ and we find particular solutions $y_{p1}(x)$ and $y_{p2}(x)$ for the inhomogeneous problems with $f_1(x)$ and $f_2(x)$, then $y_p \equiv y_{p1} + y_{p2}$ is a particular solution to the inhomogeneous problem with $f(x)$. This is simply due to the linearity of the ODE. Thus, the idea is to divide and conquer, i.e. split the r.h.s. of (4.27) into simple parts and guess the particular solution to each simple part individually. In particular, we can use the Fourier series expansion of $f(x)$. Then the r.h.s. is written as a combination as in (3.11) and for each term separately we can obtain a particular solution as described in (4.29).

Let us now look at a few special cases for non-linear second order ODE. A simple case is

$$\text{dependent variable missing:} \qquad y'' + f(y') = g(x) \tag{4.32}$$

We can set $z(x) = y'(x)$ and transform (4.32) into the first order ODE $z' + f(z) = g(x)$. Another lucky case is

$$y' \text{ missing:} \qquad y'' + f(y) = 0 \tag{4.33}$$

and no explicit x dependence. Multiplying (4.33) by y' and using $y'\,y'' = [(y')^2]'/2$ we can convert (4.33) into a separable first order ODE

$$y' = \Big(-2 \int f(y) dy \Big)^{1/2} \tag{4.34}$$

The last special case we consider is

$$\text{independent variable missing:} \qquad y'' + f(y') + g(y) = 0 \tag{4.35}$$

This is a generalisation of (4.33). Then we can set $y' = p$ and with $y'' = p\dot{p}$, where $\dot{p} \equiv dp/dy$ and using p as new dependent and y as new independent variable we transform (4.35) into a first order ODE for $p(y)$

$$p\,\dot{p} + f(p) + g(y) = 0 \tag{4.36}$$

All our attempts and partial successes in Sections 4.1 and 4.2 should not cover up the fact that very often we cannot obtain exact analytic solutions of ODE. In such cases we have to revert to approximations or numerical methods.

4.3 Higher-order ODE

Luckily, with very few exceptions, the ODE we encounter in Physics are at most second order. Hence, we will spend very little time with ODE beyond second order.

One example we will be looking at quickly is an nth order homogeneous, linear ODE with constant coefficients

$$y^{(n)}(x) + a_{n-1} y^{(n-1)}(x) + \ldots a_1 y'(x) + a_0 y(x) = 0 \tag{4.37}$$

where the coefficients a_i do not depend on x. Very similar to the corresponding case of the second order ODE, the ansatz $y(x) = e^{\lambda x}$ transforms (4.37) into an algebraic equation

$$\lambda^n + a_{n-1}\lambda^{n-1} + \ldots a_1 \lambda + a_0 \lambda = 0 \tag{4.38}$$

that is again called the characteristic equation. It is a polynomial equation in λ of order n and has n solutions λ_i, $1 \leq i \leq n$. If all λ_i are different, the general solution (complementary function) is given as

$$y_h(x) = \sum_{i=1}^{n} c_i\, e^{\lambda_i x} \tag{4.39}$$

If two or more λ_i are equal, we have to include $x^j e^{\lambda_i x}$ as solutions. In any case, we now have n linearly independent solutions and we need n initial conditions to uniquely determine the solution. In the inhomogeneous case we have to find once more a single particular solution to obtain the general solution. Obviously, it is typically more complicated than in Section 4.2. Variation of constants gets very tedious and the best bet is a good guess or using the Laplace transform.

4.4 Frobenius method (series solution)

As a preparation for the next section we present a brute force method to deal with ODE of the form (4.19), i.e. second order ODE with non-constant coefficients. In fact, you will encounter many such ODE in Physics and, therefore, in Section 5. some of them will be investigated in detail.

The idea is as simple as it is powerful. We make an ansatz for $y(x)$ as a power series in x

$$y(x) = x^\rho \big(k_0 + k_1 x + \dots\big) = \sum_{n=0}^{\infty} k_n \, x^{n+\rho} \tag{4.40}$$

and try to determine the overall power ρ and the constants k_i. As this is an infinite amount of work, the k_i are hopefully determined through a recursion relation. Plugging (4.40) into (4.19) we obtain ⇒ lecture

$$\sum_{n=0}^{\infty} \Big[(\rho+n)(\rho+n-1) + xp(x)(\rho+n) + x^2 q(x)\Big] k_n \, x^{n+\rho-2} = 0 \tag{4.41}$$

If the functions $p(x)$ and $q(x)$ are kind enough to allow a Taylor expansion of

$$x\,p(x) = (p_0 + xp_1 \dots) \tag{4.42a}$$

$$x^2\,q(x) = (q_0 + xq_1 + \dots) \tag{4.42b}$$

around $x = 0$ we can expand all terms in the l.h.s. of (4.41). To ensure that we get 0, each power of x on the l.h.s. has to be zero. In particular, the lowest power, i.e. the term with $n = 0$ has to vanish. This results in the *indicial equation [charakteristische Gleichung]*[1]

$$\rho(\rho-1) + p_0\,\rho + q_0 = 0 \tag{4.43}$$

This is a quadratic equation in ρ with two solutions, ρ_1 and ρ_2, giving rise to two linearly independent solutions. For both of those the hope is to determine all coefficients k_i though a recursion relation. Alternatively, once one of the two solutions is fixed, the second one can be found as described in connection with (4.21) and (4.22).

If all this seems very abstract, we will practise this method intensely in the next section.

[1] German speaking mathematicians don't seem to be very imaginative giving names to equations and call every second equation coming their way 'charakteristische Gleichung'.

5 Special ordinary differential equations

This section is devoted to a rather detailed study of a selection of second order ODE that play a particularly important role in Physics. In doing so we get to know some functions that appear in different contexts in physical problems. While we will investigate some properties of these functions, our study will be by no means exhaustive. Still, this section is rather heavy in doing calculations and getting familiar with and used to some computational techniques is another aim of this section.

5.1 Legendre differential equation

The Legendre differential equation reads

$$(1-x^2)\,y''(x) - 2x\,y'(x) + l(l+1)\,y(x) = 0 \tag{5.1a}$$

$$\text{or} \quad [(1-x^2)\,y']' + l(l+1)\,y = 0 \tag{5.1b}$$

where for the moment we consider l simply as an external parameter. It is a linear second order ODE (4.19) with non-constant $p(x) = -2x/(1-x^2)$ and $q(x) = l(l+1)/(1-x^2)$. Before we jump to solving it let us note that (5.1a) appears in connection with the Laplace operator in spherical coordinates $\vec{r} = (r, \phi, \theta)$ and the variable x corresponds to $x \sim \cos\theta$. Hence we will primarily look for solutions that are well defined in the range $-1 \leq x \leq 1$. As we will see below this entails $l \in \mathbb{N}$.

We will use the Frobenius method introduced in Section 4.4 to look for solutions. Writing $y(x)$ as in (4.40) as a first step we need to solve the indicial equation (4.43). In our current case we have $p_0 = q_0 = 0$ and hence (4.43) simplifies to $\rho(\rho-1) = 0$. The two solutions are $\rho_1 = 0$ and $\rho_2 = 1$.

Considering first the case $\rho_1 = 0$ we plug $y(x) = \sum k_n x^n$ into (5.1b) and obtain the relation

$\Rightarrow$ lecture

$$\sum_{n=0}^{\infty} \Big(n(n-1)k_n x^{n-2} - \big(n(n+1) - l(l+1)\big)k_n x^n \Big) = 0 \tag{5.2}$$

This can only hold if there is a relation between k_n and k_{n+2}. In fact, we need

$$k_{n+2} = k_n \frac{n(n+1) - l(l+1)}{(n+2)(n+1)} = k_n \frac{(n-l)(n+l+1)}{(n+2)(n+1)} \tag{5.3}$$

Thus, to obtain the first linearly independent solution (called $y_1(x)$ in (4.20)) we fix k_0 (called c_1 in (4.20)) and then obtain all further k_n, with n even through (5.3). The second linearly independent solution (called $y_2(x)$ in (4.20)) is obtained using the second solution of the indicial equation, $\rho = 1$. The recursion relation we obtain is identical with (5.3). Hence, we fix k_1 (called c_2 in (4.20)) and then obtain all further k_n, with n odd through (5.3).

In general one would expect that the recursion never stops and the solutions are given by a non-terminating series. However, if l happens to be a positive integer, i.e. $l \in \mathbb{N}$ then at some point we will have $n = l$. This means $k_{l+2} = 0$, as can be inferred from (5.3). But then (5.3) also entails that all further $k_{l+4} = 0$ and for all $m > l$ we have $k_m = 0$. In other words, our solution is a terminating series, a polynomial of rank l. No prize for guessing that these

polynomials, suitably normalized, are called the *Legendre polynomials* *[Legendre Polynome]*. They are conventionally denoted by $P_l(x)$.

As an example let's construct $P_3(x)$. For this we start with an arbitrary k_1 and get from (5.3) $k_3 = -5/3\, k_1$. All further $k_m = 0$. Our 3rd Legendre polynomial therefore reads

$$P_3 \sim k_1\, x - \frac{5}{3} k_1\, x^3 \qquad \xrightarrow{P_l(1)=1} \qquad P_3 = -\frac{3}{2}\, x + \frac{5}{2}\, x^3 \tag{5.4}$$

The normalization condition that is conventionally used is $P_l(1) = 1$ which fixes k_1 for odd l and k_0 for even l.

Taking stock we note that for each $l \in \mathbb{N}$ we have found a solution to (5.1a) that is a polynomial of degree l and contains either only even or only odd powers of x.

Before we go on and indulge ourselves with studying these Legendre polynomials we stress that there are also other solutions. After all, for each l we expect two linearly independent solutions. Taking e.g. again $l = 3$, we can start with k_0, rather than k_1, and the recursion will never stop. Thus, we get a second, non-terminating solution. In addition, there are also solutions to (5.1a) with $l \notin \mathbb{N}$. Again, in these cases the recursion (5.3) never stops and we get a non-terminating solution. This as such is not a problem. But (5.3) reveals that for $n \to \infty$ we get $k_{n+2} \simeq k_n$. It is then clear that the (non-terminating) series $y(x) = \sum k_n x^n$ cannot possibly converge for $x = 1$. Thus, if we insist on a solution in the range $-1 \leq x \leq 1$ we have to make sure the series terminates. The Legendre polynomials are the only solutions that are defined in $-1 \leq x \leq 1$ and we need $l \in \mathbb{N}$ and get one such solution $\forall l \in \mathbb{N}$. In fact, l is related to the angular momentum in quantum mechanics and what we have witnessed is nothing but its quantized nature.

Getting back to the Legendre polynomials P_l we list some properties. First, they can be obtained through a *Rodrigues* formula *[Rodrigues Formel]*

$$\text{Rodrigues:} \qquad P_l(x) = \frac{1}{2^l\, l!} \frac{d^l}{dx^l} (x^2 - 1)^l \equiv \frac{1}{2^l\, l!} D^l (x^2 - 1)^l \tag{5.5}$$

Thus, up to a normalization, P_l is obtained by acting l times with the differentiation operator $D \equiv d/dx$ on $(x^2-1)^l$. To prove (5.5) we show that $D^l(x^2-1)^l$ satisfies (5.1a) and the prefactor ensures the normalization condition $P(1) = 1$. This is basically an exercise in differentiation. ⇒ lecture

A property that will be very important is the orthogonality relation of Legendre polynomials, or concretely

$$\text{orthogonality:} \qquad \int_{-1}^{1} dx\, P_l(x) P_{l'}(x) = \frac{2}{2l+1} \delta_{ll'} \tag{5.6}$$

Indeed, we can use the functions $\sqrt{(2l+1)/2}\, P_l$ as a basis in the vector space of continuous functions $f(x)$ in $-1 \leq x \leq 1$ with the scalar product ⇒ lecture

$$\langle f | g \rangle = \int_{-1}^{1} f^*(x) g(x) dx \tag{5.7}$$

and we can expand such functions as a "Legendre" series

$$f(x) = \sum_{l=0}^{\infty} c_l\, P_l \qquad \text{with} \qquad c_l = \frac{2l+1}{2} \int_{-1}^{1} P_l(x) f(x) dx \tag{5.8}$$

This should look awfully familiar! If not, have a look at (3.16a) and (3.16b).

We mention two further properties. First, we can obtain the Legendre polynomials through a *recursion relation [Rekursionsrelation]*. If we fix $P_0 = 1$ and $P_1 = x$ we can obtain P_l for $l > 1$ through

$$\text{recursion:} \qquad (l+1)P_{l+1} = (2l+1)x\,P_l - l\,P_{l-1} \tag{5.9}$$

Finally we have a *generating function [erzeugende Funktion]*. Assuming $|t| < 1$ we have

$$\text{generating function:} \qquad \sum_{l=0}^{\infty} P_l(x)\,t^l = \frac{1}{\sqrt{1-2xt+t^2}} \tag{5.10}$$

The r.h.s. of (5.10) allows us to obtain P_l by differentiating l times w.r.t. t (dividing by $l!$) and then setting $t = 0$. In some sense, all information about all Legendre polynomials is contained in the function on the r.h.s. of (5.10). For example (5.9) can be shown using (5.10). Hence the name generating function. ⇒ lecture

5.2 Generalized Legendre differential equation

The Legendre polynomials as discussed in Section 5.1 appear in connection with the Laplace equation with azimuthal symmetry, i.e. if there is no ϕ dependence. In the general case, i.e. with ϕ dependence, we will encounter the generalized Legendre equation

$$(1-x^2)\,y''(x) - 2x\,y'(x) + \Big(l(l+1) - \frac{m^2}{1-x^2}\Big)y(x) = 0 \tag{5.11}$$

which in addition to l contains another parameter m. If you want a physical interpretation of m think of it as the z component of the angular momentum l. We will see $|m| \leq l$. All this will come back to haunt you in Section 7.

Once more, we are looking for solutions to (5.11) that exist $\forall x$ in $-1 \leq x \leq 1$. As we will see these are again terminating series solutions, i.e. polynomials. They are called *associated [assoziierte]* Legendre polynomials $P_l^{(m)}(x)$. To find these solutions we change the dependent variable form $y(x)$ to $z(x)$ with

$$y(x) = (1-x^2)^{m/2} z(x) \tag{5.12}$$

plug in this ansatz into (5.11) and multiply the full equation by $(1-x^2)^{m/2}$. This results in an ODE for $z(x)$, namely ⇒ lecture

$$(1-x^2)z'' - 2(m+1)x\,z' + \big(l(l+1) - m(m+1)\big)z = 0 \tag{5.13}$$

For $m = 0$ (5.13) is equivalent to (5.1a). Hence the solution is $z(x) = P_l^{(m=0)} = P_l(x)$. In order to obtain solutions for $m \neq 0$ we differentiate (5.13) w.r.t. x and set $v(x) = z'(x)$ to obtain

$$(1-x^2)v'' - 2(m+2)x\,v' + \big(l(l+1) - (m+1)(m+2)\big)v = 0 \tag{5.14}$$

(5.14) can be obtained from (5.13) by replacing $m \to m+1$ (and $z \to v$). Thus P_l' is the solution of (5.13) with $m = 1$. Similarly we conclude that the nth derivative $P_l^{(n)}$ is

the solution of (5.13) with $m = n$. Remembering the prefactor $(1-x^2)^{m/2}$ and adding a conventional $(-1)^m$, we find that

$$P_l^{(m)}(x) \equiv (-1)^m (1-x^2)^{m/2} \frac{d^m}{dx^m} P_l(x) \tag{5.15}$$

are the solutions to (5.1a), i.e. the associated Legendre polynomials. Since P_l is a polynomial of degree l, we only get non-trivial solutions $P_l^{(m)}(x)$ for $m \leq l$. Combining (5.15) with (5.5) the definition can be extended to negative m with $-l \leq m \leq l$ as

$$P_l^{(m)}(x) = \frac{(-1)^m}{2^l\, l!} (1-x^2)^{m/2} D^{l+m} (x^2-1)^l \tag{5.16}$$

and we have ⇒ lecture

$$P_l^{(-m)}(x) = (-1)^m \frac{(l-m)!}{(l+m)!} P_l^{(m)}(x) \tag{5.17}$$

The associated Legendre polynomials also satisfy an orthogonality relation similar to (5.6), namely

$$\int_{-1}^{1} dx\, P_l^{(m)}(x) P_{l'}^{(m)}(x) = \frac{2}{2l+1} \frac{(l+m)!}{(l-m)!} \delta_{ll'} \tag{5.18}$$

They are closely related to the *spherical harmonics* *[Kugelflächenfunktionen]*

$$Y_l^{(m)}(\phi, \theta) = \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}}\, P_l^{(m)}(\cos\theta)\, e^{im\phi} \tag{5.19}$$

These functions will play an important role in connection with the Laplace equation in spherical coordinates. They form an orthonormal basis on the 2-dimensional surface Ω_2 of the 3-dimensional unit sphere

$$\int d\Omega_2 \big(Y_l^{(m)}\big)^* Y_{l'}^{(m')} = \delta_{ll'} \delta_{mm'} \tag{5.20}$$

You should recognize the expression on the r.h.s. of (5.20) as a scalar product as e.g. in (3.2). The domain of integration $d\Omega_2$ we have already encountered in connection with the 3-dimensional Fourier transform (3.34).

5.3 Laguerre differential equation

The most popular appearance of the Laguerre ODE is related to quantum mechanics with the Coulomb potential, e.g. the hydrogen atom. But for us it is just another ODE that reads

$$x\, y'' + (1-x)\, y' + n\, y = 0 \tag{5.21}$$

where n to start with is simply a parameter, very similar to l in Section 5.1. This ODE has the form of (4.19) with non-constant $p(x) = (1-x)/x$ and $q(x) = n/x$. This leads to the characteristic equation (4.43) in the form $\rho(\rho-1) + \rho = 0$ and, hence, $\rho = 0$. Thus, we make the ansatz $y(x) = \sum_{j=0}^{\infty} k_j\, x^j$ and using the Frobenius method arrive at the recursion relation ⇒ lecture

$$k_{j+1} = k_j \frac{j-n}{(j+1)^2} \qquad \rightarrow \qquad k_j = (-1)^j \frac{1}{j!} \binom{n}{j} k_0 \tag{5.22}$$

where in the second step we expressed k_j directly in terms of k_0 with the help of the binomial coefficient. If the parameter $n \in \mathbb{N}$ we obtain again terminating solutions, since $k_{n+1} = 0$ and $k_m = 0$, $\forall m$ with $m \geq n+1$. These solutions with the normalization fixed through $k_0 = 1$ are called ... yes, you've guessed it, *Laguerre polynomials* *[Laguerre Polynome]* and deserve a displayed equation ⇒ lecture

$$L_n = \sum_{j=0}^{n} \binom{n}{j} \frac{(-x)^j}{j!} \tag{5.23}$$

Very much like the Legendre polynomials, the Laguerre polynomials have a bunch of relations.

$$\text{Rodrigues:} \qquad L_n = \frac{e^x}{n!} D^n \left(x^n e^{-x}\right) \tag{5.24a}$$

$$\text{orthogonality:} \qquad \int_0^\infty dx\, e^{-x} L_m(x) L_n(x) = \delta_{mn} \tag{5.24b}$$

$$\text{recursion:} \qquad (n+1) L_{n+1} = (2n+1-x) L_n - n L_{n-1} \tag{5.24c}$$

$$\text{generating function:} \qquad \sum_{n=0}^{\infty} L_n(x)\, t^n = \frac{e^{-\frac{xt}{1-t}}}{1-t} \tag{5.24d}$$

The proof of (5.24a) is an exercise in differentiation whereas the proof of (5.24d) is a lovely application of what we learned in Section 1. Indeed, starting from (5.24a) and using (1.33) to express the nth derivative as a integral along a simple closed contour, we can write ⇒ lecture

$$L_n(z) = \frac{e^z}{2i\pi} \oint \frac{\zeta^n e^{-\zeta}\, d\zeta}{(\zeta - z)^{n+1}} \qquad \xrightarrow{t=1-z/\zeta} \qquad \frac{1}{2i\pi} \oint \frac{e^{-\frac{zt}{1-t}}\, dt}{(1-t)(t-0)^{n+1}} \tag{5.25}$$

where we simply performed a change of variable $\zeta \to t$ in the contour integral. Using (1.33) again, the last expression is recognised (up to a factor $n!$) as the nth derivative w.r.t. t at $t = 0$ of the function given on the r.h.s. of (5.24d).

In case you wondered where the second linearly independent solution has gone, or what happens if $n \notin \mathbb{N}$: as for the Laguerre case, there are such solutions, but they are not terminating solutions, i.e. not polynomials. This is not to say they are not important. The Legendre polynomials are related to solutions with the Coulomb potential that correspond to bound states. The fact that they exist only for integer values of n will be related to bound states only existing for very particular energies (think Balmer series). The non-terminating solutions play a role for unbound systems governed by the Coulomb potential. We will not discuss them any further here, but they might come back with a vengeance once you discuss scattering off a Coulomb potential in Quantum Mechanics.

As for the Legendre ODE, there is also a generalized Laguerre ODE. It reads

$$x\, y'' + (k+1-x)\, y' + n\, y = 0 \tag{5.26}$$

and for $k = 0$ reduces to the standard Laguerre ODE, (5.21). There are polynomial solutions to (5.26) given by the *associated* *[assoziierte]* Laguerre polynomials

$$L_n^{(k)}(x) = (-1)^k D^k L_{n+k}(x) \qquad \text{with} \qquad D = \frac{d}{dx} \tag{5.27}$$

Again, there are additional non-terminating solutions.

5.4 Hermite differential equation

The Hermite ODE will make an entrance in the quantum mechanical treatment of the harmonic oscillator, i.e. the potential $V(x) \sim x^2$. The Schrödinger equation for the wave function $\psi(x)$ reads $\psi'' - x^2\psi + (2n+1)\psi = 0$ where n again is a parameter. If we write $\psi(x) = e^{-x^2/2}H(x)$ the Schrödinger equation results in a ODE for the function $H(x)$, the Hermite ODE

$$H'' - 2x\,H' + 2n\,H = 0 \tag{5.28}$$

Once more, this ODE has the form of (4.19), this time with non-constant $p(x) = -2x$ and constant $q = 2n$, hence $\rho = 0$ or $\rho = 1$. We follow standard procedure (by now we can do this in our sleep) to obtain for $n \in \mathbb{N}$ the polynomial solutions, called *Hermite polynomials* ⇒ lecture *[Hermite Polynome]*

$$H_n(x) = \sum_{k=0}^{n} c_k\, x^{k+\rho} \qquad \text{with} \qquad c_{k+2} = \frac{2(k+\rho-n)}{(k+2+\rho)(k+1+\rho)}\, c_k \tag{5.29}$$

Their relations read

Rodrigues: $$H_n = (-1)^n e^{x^2}\, D^n\big(e^{-x^2}\big) \tag{5.30a}$$

orthogonality: $$\int_{-\infty}^{\infty} dx\, e^{-x^2} H_m(x) H_n(x) = 2^n \sqrt{\pi}\, n!\, \delta_{mn} \tag{5.30b}$$

recursion: $$H_{n+1} = 2x\, H_n - 2n H_{n-1} \tag{5.30c}$$

generating function: $$\sum_{n=0}^{\infty} H_n(x)\, \frac{t^n}{n!} = e^{2xt-t^2} \tag{5.30d}$$

The nth Hermite polynomial is a polynomial of degree n with only even or odd powers of x. Once more, there are additional non-terminating solutions to (5.28).

5.5 Bessel differential equation

Before we get bored stiff with another polynomial solution to an ODE, let us consider an ODE where we are actually interested in non-terminating solutions. A very important example of such an ODE is the Bessel equation

$$x^2\, y'' + x\, y' + (x^2 - \alpha^2)\, y = 0 \tag{5.31}$$

where α is a parameter. The solutions of this ODE show up regularly in various applications. We will encounter them when we solve the Laplace equation in cylindrical coordinates and, actually, have already had a near miss with them at the end of Section 3.2. We can rewrite (5.31) in the form of (4.19) with non-constant $p(x) = 1/x$ and $q(x) = (1 - \alpha^2/x^2)$. The characteristic equation (4.43) therefore reads

$$\rho(\rho-1) + \rho - \alpha^2 = 0 \tag{5.32}$$

and has the two solutions $\rho_{1/2} = \pm\alpha$. Considering first $\rho_1 = +\alpha$ we obtain a series solution ⇒ lecture

with the recursion relation of the coefficients as

$$y = \sum_{n=0}^{\infty} k_n \, x^{n+\alpha} \qquad\qquad k_{n+2} = \frac{-k_n}{(n+2)(n+2+2\alpha)} \tag{5.33}$$

Fixing k_0 we obtain a non-terminating series of coefficients k_n with n even as ⇒ lecture

$$y(x) = \sum_{n=0}^{\infty} k_{2n} x^{2n+\alpha} \qquad \text{with} \qquad k_{2n} = k_0 \, \frac{(-1)^n \, \alpha!}{2^{2n} n! (n+\alpha)!} \tag{5.34}$$

In (5.34) we have implicitly assumed α is an integer. However, if we express $(n+\alpha)! \to \Gamma(n+\alpha+1)$ we can relax this restriction. Fixing the initial condition by $k_0 = (2^\alpha \, \alpha!)^{-1}$ and changing notation $y(x) \to J_\alpha(x)$ we obtain the *Bessel function of the first kind [Besselfunktion erster Art]* as

$$J_\alpha(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!(n+\alpha)!} \left(\frac{x}{2}\right)^{2n+\alpha} \quad \to \quad \sum_{n=0}^{\infty} \frac{(-1)^n}{\Gamma(n+1)\Gamma(n+\alpha+1)} \left(\frac{x}{2}\right)^{2n+\alpha} \tag{5.35a}$$

$$J_{-\alpha}(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{\Gamma(n+1)\Gamma(n-\alpha+1)} \left(\frac{x}{2}\right)^{2n-\alpha} \tag{5.35b}$$

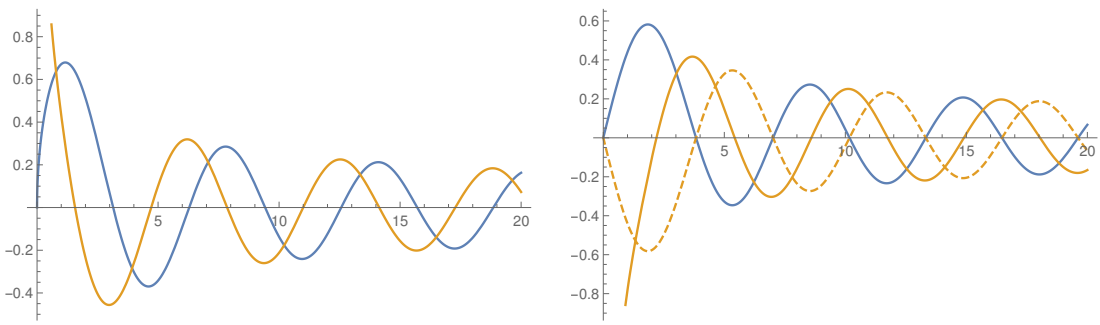

Figure 5.1: Left panel: the linearly independent Bessel functions $J_{1/2}(x)$ (blue) and $J_{-1/2}(x)$ (orange); right panel: the linearly dependent Bessel functions $J_1(x)$ (blue) and $J_{-1}(x)$ (orange dashed). To obtain a second linearly independent solution for $\alpha \in \mathbb{Z}$ a customary choice is $Y_1(x)$ (orange, solid).

In (5.35b) we have directly given the second solution that is obtained for $\rho_2 = -\alpha$. Both solutions given in (5.35) are non terminating solutions to (5.31). They are linearly independent if $\alpha \notin \mathbb{Z}$. This is evident from looking at the limit $x \to 0$ which is

$$\lim_{x\to 0} J_{\alpha>0}(x) \to 0 \qquad \text{and} \qquad \lim_{x\to 0} J_{\alpha<0}(x) \to \infty \tag{5.36}$$

As an example, in the left panel of Figure 5.1 the Bessel functions $J_{1/2}(x)$ and $J_{-1/2}(x)$ are shown, with the behaviour as in (5.36). However, we can show that ⇒ lecture

$$J_{-\alpha} = (-1)^\alpha J_\alpha \qquad\qquad \text{if} \quad \alpha \in \mathbb{Z} \tag{5.37}$$

For this we need a property of the Γ function, namely that it has a (single) pole at $\Gamma(-z_n)$ where $z_n \in \mathbb{N}$, i.e. for negative integers, including 0. Thus, for $\alpha \in \mathbb{N}$ we need a solution

different from $J_{-\alpha}$ to obtain a second linearly independent solution. Without going into detail we just mention that the functions

$$J_\alpha(x) \qquad \text{and} \qquad Y_\alpha \equiv \lim_{\nu\to\alpha} \frac{J_\nu(x)\cos(\nu\pi) - J_{-\nu}(x)}{\sin(\nu\pi)} \tag{5.38}$$

are linearly independent even for $\alpha \in \mathbb{N}$. The function Y_α is called Bessel function of the *second kind* *[zweiter Art]* or sometimes Neumann function. The limit $\nu \to \alpha$ is only delicate for $\alpha \in \mathbb{N}$.

For completeness we also mention that sometimes Bessel functions of the 3rd kind, or Hankel functions are used. They are defined as linear combinations

$$H_\alpha^{(1)}(x) \equiv J_\alpha(x) + i\, Y_\alpha(x) \tag{5.39a}$$

$$H_\alpha^{(2)}(x) \equiv J_\alpha(x) - i\, Y_\alpha(x) \tag{5.39b}$$

Let's now go back to (3.37) where we have found an ODE for $G_n(k)$ defined through a $n-1$ dimensional integral, (3.36). The ODE (3.38) is equivalent to the Bessel ODE (5.31). This can be seen by setting $G_n(k) = (2\pi)^{n/2} k^{-\alpha} J_\alpha(k)$ with $\alpha = n/2 - 1$ and plugging this into (3.38). The factor $(2\pi)^{n/2}$ is just a normalization that can be fixed by considering $k = 0$ and making sure $G_n(0) = \int d\Omega_{n-1}$ corresponds to the surface of the n dimensional unit sphere. Then we find

$$G_n(k) = \int d\Omega_{n-1}\, e^{-i\vec{k}\cdot\vec{r}} = (2\pi)^{n/2} k^{1-\frac{n}{2}} J_{\frac{n}{2}-1}(k) \tag{5.40}$$

Indeed, setting e.g. $n = 3$ in (5.40) we get $G_n(k) = 4\pi \sin(k)/k$, in agreement with the result obtained by actually doing the integration. But as already alluded to at the end of Section 3.2, we are now not restricted to the case $n \in \mathbb{N}$.

5.6 Airy differential equation and Laplace method

As a last concrete example of an ODE in this section we consider

$$y'' - \lambda x\, y = 0 \tag{5.41}$$

This ODE appears for example in connection with the Schrödinger equation in the gravitational potential. Without loss of generality we can set $\lambda = 1$, as we can always rescale $x \to \lambda^{1/3} x$.

We can again use the Frobenius method. The indicial equation reads $\rho(\rho - 1) = 0$ with solutions $\rho = 0$ and $\rho = 1$. The recursion relation is found as usual. It relates terms x^n to x^{n+3} and starting with k_0 and k_1 leads to two independent solutions as ⇒ lecture

$$y_1(x) = \sum_{n\geq 0} k_{3n} x^{3n} \qquad \text{with} \qquad k_{n+3} = \frac{\lambda}{(n+3)(n+2)} k_n \tag{5.42a}$$

$$y_2(x) = \sum_{n\geq 0} k_{3n+1} x^{3n+1} \qquad \text{with} \qquad k_{n+3} = \frac{\lambda}{(n+4)(n+3)} k_n \tag{5.42b}$$

If we fix k_0 and k_1 we have two linearly independent solutions. In the literature two particular combinations are used which are called Airy function(s). They are obtained (with $\lambda := 1$)

by setting $k_0 = 1/\Gamma(2/3)$ for $y_1(x)$ and $k_1 = 1/\Gamma(1/3)$ for $y_2(x)$ and then forming the silly looking linear combinations

$$\mathrm{Ai}(x) = 3^{-2/3}\, y_1(x) - 3^{-1/3}\, y_2(x) \tag{5.43a}$$

$$\mathrm{Bi}(x) = 3^{-1/6}\, y_1(x) + 3^{1/6}\, y_2(x) \tag{5.43b}$$

The linear combination (5.43a) has been chosen such that $\lim_{x\to\infty} \mathrm{Ai}(x) = 0$. The reason for the other funny coefficients is to do with the integral representations, (5.49) and (5.50).

We will use (5.41) to illustrate another method to solve ODE, namely the *Laplace method [Laplace Methode]*. As the name suggests, this is closely related (but not quite the same) as using the Laplace transform to solve an ODE. In fact, we write

$$y(x) = \int_\gamma \tilde{y}(s) e^{xs} ds \tag{5.44}$$

which looks very similar to the inverse Laplace transform (3.50). Apart form the trivial difference related to the missing prefactor $1/(2i\pi)$ that we will take into account at the very end, there is a more important difference in that in (5.44) the contour γ is not yet fixed. Still, we can insert (5.44) into (5.41) and obtain

$$0 = \int_\gamma (s^2 - x) e^{sx} \tilde{y}(s) ds = \int_\gamma (s^2 \tilde{y} + \tilde{y}') e^{sx} ds - e^{sx}\, \tilde{y}\Big|_\gamma \tag{5.45}$$

where we have used (3.42a) with $n = 1$. The last term corresponds to the boundary term obtained while using integration by parts. Here we use our freedom to choose the contour γ such that the boundary terms vanish. Then (5.45) corresponds to a linear, homogeneous (thus separable) first order ODE ⇒ lecture

$$\tilde{y}'(s) + s^2 \tilde{y}(s) = 0 \tag{5.46}$$

that can be solved readily as

$$\tilde{y}(s) = c\, e^{-s^3/3} \qquad \Rightarrow \qquad y(x) = c \int_\gamma e^{-s^3/3} e^{xs} ds \tag{5.47}$$

How do we choose γ such that the boundary terms vanish? We could take a closed path. But then we just get $y(x) = 0$ due to Cauchy's theorem. While this is a solution to (5.41) we could do with more interesting ones. We have to make sure the contour γ starts and ends in the domain of the complex s plane where $e^{sx}\, \tilde{y} = e^{xs} e^{-s^3/3} \to 0$. This is the case in the domain with $\mathrm{Re}(s^3) > 0$ and $|s| \to \infty$. This domain is shown in light blue in Figure 5.2. There are three directions we can take with $|s| \to \infty$ to end up in a 'safe' end point, i.e. an end point s.t. the boundary terms vanish. This means we have three paths, γ_1, γ_2, and γ_3 as shown in Figure 5.2. However, the corresponding integrals are not independent, as the sum of them vanishes, due to Cauchy. Therefore, we end up with two independent choices for the contour and, hence, two linearly independent solutions, as expected.

If we take γ_1 which is equivalent to integrating along the imaginary axis we obtain ⇒ lecture

$$\int_{\gamma_1} ds\, e^{sx - s^3/3} = i \int_{-\infty}^{\infty} dt\, e^{i(tx + t^3/3)} = i \int_{-\infty}^{\infty} dt\, \cos(tx + t^3/3) \tag{5.48}$$

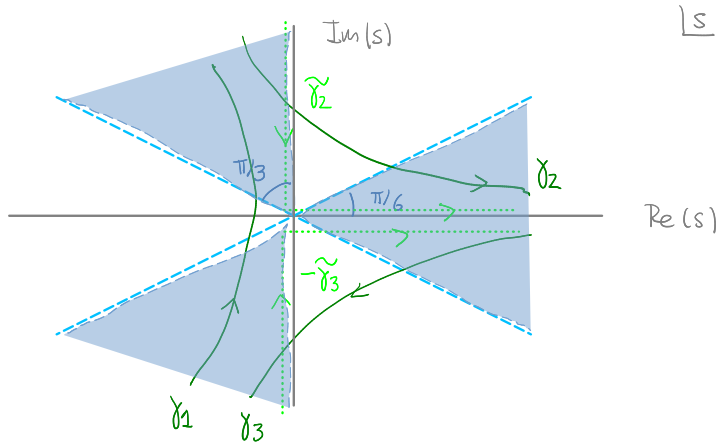

Figure 5.2: Complex s plane with contour γ_1 chosen for the Ai function and $\gamma_2 - \gamma_3$ for Bi.

Including the prefactor $1/(2i\pi)$ we have omitted in (5.44) we end up with the integral representation of the Airy function

$$\mathrm{Ai}(x) = \frac{1}{\pi}\int_0^\infty dt\,\cos(tx + t^3/3) \tag{5.49}$$

For the second linearly independent solution we can choose e.g. $\gamma_2 - \gamma_3$ and writing this as a combination of integrations along the real and imaginary axes (the bright green contours $\tilde{\gamma}_2$ and $\tilde{\gamma}_3$ depicted in Figure 5.2) we obtain ⇒ lecture

$$\mathrm{Bi}(x) = \frac{1}{\pi}\int_0^\infty dt\Big(\sin(tx + t^3/3) + \exp(tx - t^3/3)\Big) \tag{5.50}$$

Depending on the applications, the series representation or the integral representation is more useful. In fact, also the Bessel functions have integral representations and the Airy functions can be expressed in terms of the Bessel functions $J_{1/3}$ and $J_{-1/3}$.

Needless to say that we could spend the rest of our lives studying the properties and different representations of solutions to all sorts of ODE. Often the one ODE that you really need is one that has not been covered here. But at least you should now have a set of tools at hand to tackle it.

5.7 Sturm-Liouville differential equation

In this subsection we will not look at yet another ODE. Instead, we go back to Section 5.1–5.4, marvel at the similarities and make a connection to topics that will be covered in MMP 2. We found a set of solutions that satisfy some kind of orthogonality relation, (5.6), (5.18), (5.24b), and (5.30b). Could it be this is not a coincidence?

All these ODE are of a certain type, called *Sturm-Liouville* differential equation. These are of the form

$$-D\big[p(x)D\big]y(x) + q(x)y(x) = \lambda\, w(x)\, y(x) \qquad \text{with} \qquad D \equiv \frac{d}{dx} \tag{5.51}$$

where λ is a parameter, $p(x)$ and $q(x)$ are suitable functions of the independent variable x and $w(x) \geq 0$ is a weight function to be discussed in connection with (5.52). The concrete expressions are summarised as follows:

ODE	$p(x)$	$q(x)$	$w(x)$	λ	range of x
Legendre	$1-x^2$	$\frac{m^2}{1-x^2}$	1	$l(l+1)$	$-1 \leq x \leq 1$
Laguerre	$x^{k+1}\,e^{-x}$	0	$x^k\,e^{-x}$	n	$0 \leq x < \infty$
Hermite	e^{-x^2}	0	e^{-x^2}	$2n$	$-\infty < x < \infty$

For the Legendre case we took m as a fixed constant and express the parameter λ through l. This corresponds to the orthogonality relation (5.18).

The deeper reason why the solutions to such an ODE are orthogonal or, in fact can even be made into an orthonormal basis, is the following: we can view $y(x)$ as a vector in an infinite dimensional vector space. In this vector space we are able to introduce a scalar product, very much like (3.2), but generalized with the weight function $w(x)$ as

$$\langle f|g\rangle = \int f^*(x)g(x)\,w(x)dx \tag{5.52}$$

The integration domain depends on the type of functions (i.e. the vector space) we consider. Roughly speaking (more details in MMP 2) we therefore obtain a Hilbert space. Now we define

$$SL \equiv -D\big[p(x)D\big] + q(x) \tag{5.53}$$

as a linear operator in this Hilbert space. With this we can write (5.51) as an eigenvalue problem

$$SL\,y(x) = \lambda\,y(x) \tag{5.54}$$

and solving the ODE amounts to finding the eigenvalues λ and eigenvectors of the operator SL. The crucial property of SL is that it is a self-adjoint operator. In a finite-dimensional vector space, a self-adjoint operator is represented by a hermitian matrix and has an orthonormal eigenbasis. As we will learn in MMP 2 this can be extended to infinite-dimensional Hilbert spaces. In other words, for self-adjoint operators such as SL we can find eigenvectors that are mutually orthogonal. Even more, they form a basis, i.e. any vector in this Hilbert space can be expressed in the basis of eigenvectors $y(x)$. This will explain why we can expand any vector in this Hilbert space in terms of the solutions to the ODE. An example of such an expansion was given in (5.8). It also makes sure we did not miss some solutions which are in the Hilbert space. As we will see in Section 7, this is of utmost importance when we are going to study solutions of partial differential equations.

6 Partial differential equations

As promised in Section 4 we now turn to *partial differential equations* (PDE) *[partielle Differentialgleichungen]*. Thus, our dependent variable now depends on several independent variables. Our standard notation will be Ψ for the former, and x, y, z, t ... for the latter. Since we are dealing with a function of several variables, $\Psi(x, y \ldots)$ we will encounter partial derivatives in our PDE, such as $\partial\Psi/\partial x \equiv \partial_x \Psi$. If only single derivatives occur, we have a first order PDE. Second order PDE involve derivatives like $\partial^2\Psi(x, y \ldots)/(\partial x \partial y) \equiv \partial_{xy}\Psi(x, y \ldots)$. To cheer you up, we will only consider linear PDE, at most second order.

A particularly important operator that appears in numerous PDE in physics is the Laplace operator that has made a brief appearance already on several occasions, but now has its well deserved and triumphant entrance

$$\Delta = \nabla^2 \equiv \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) = \partial_{xx} + \partial_{yy} + \partial_{zz} \tag{6.1}$$

To convince you of the importance of this operator, we list some PDE you will have encountered, or will encounter in physics

Laplace equation	$\Delta\Psi(\vec{r}) = 0$	(6.2a)
Poisson equation	$\Delta\Psi(\vec{r}) = \rho(\vec{r})$	(6.2b)
Diffusion or heat equation	$\frac{\partial}{\partial t}\Psi(\vec{r}, t) = \kappa\, \Delta\Psi(\vec{r}, t)$	(6.2c)
Helmholtz	$\Delta\Psi(\vec{r}) + k^2\, \Psi(\vec{r}) = 0$	(6.2d)
Wave equation	$\Delta\Psi(\vec{r}, t) = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\Psi(\vec{r}, t)$	(6.2e)
Schrödinger time independent	$-\frac{\hbar^2}{2m}\Delta\Psi(\vec{r}) + V(\vec{r})\Psi(\vec{r}) = E\, \Psi(\vec{r})$	(6.2f)
Schrödinger time dependent	$i\hbar\frac{\partial}{\partial t}\Psi(\vec{r}, t) = -\frac{\hbar^2}{2m}\Delta\Psi(\vec{r}, t) + V(\vec{r})\Psi(\vec{r}, t)$	(6.2g)

where $\vec{r} = x$ or $\vec{r} = (x, y)$ or $\vec{r} = (x, y, z)$. Thus, you will not be surprised that we will investigate PDE with the Laplace operator in great detail, in cartesian, spherical and whatever other coordinate systems that come to our mind. However, before we do this in Section 7, we want to get familiar with some more general concepts of PDE. Therefore, in this section we take a more generic look at first and second order PDE. In particular, we will see that the issue of setting boundary conditions is much more subtle for PDE. A thorough treatment of these issues is well beyond the scope of this lecture, but at least we will get an overview and a reasonable preparation to tackle the most prominent PDE in physics.

6.1 First order PDE

Let us start with a concrete, simple example with two independent variables. The dependent variable $\Psi(x, y)$ is to satisfy the PDE

$$a\,\frac{\partial\Psi}{\partial x} + b\,\frac{\partial\Psi}{\partial y} = 0 \tag{6.3}$$

where a and b are constants. Thus, we are dealing with a linear, first order PDE with constant coefficients. How hard can it be.

The key idea is to look for new independent variables $s(x, y)$ and $t(x, y)$ such that the PDE turns into an ODE. In the case of (6.3) this is achieved by setting

$$\left.\begin{aligned} s &\equiv a\,x + b\,y \\ t &\equiv b\,x - a\,y \end{aligned}\right\} \qquad \begin{aligned} \partial_x \Psi &= a\,\partial_s \Psi + b\,\partial_t \Psi \\ \partial_y \Psi &= b\,\partial_s \Psi - a\,\partial_t \Psi \end{aligned} \tag{6.4}$$

The coefficients in the definition of s and t have cunningly be chosen to ensure the $\partial_t \Psi$ terms cancel when the r.h.s. of (6.4) is inserted into (6.3). Indeed, the latter now reads

$$(a^2 + b^2)\,\frac{\partial \Psi(s,t)}{\partial s} = 0 \tag{6.5}$$

with the solution $\Psi(s, t) = f(t)$, where $f(t)$ is any function depending solely on t. In fact, (6.5) is an ODE (only one independent variable s) and t is just a parameter, like e.g. l in (5.1a). Going back to our original independent variables using $t = b\,x - a\,y$ we find the solution to (6.3) as

$$\Psi(x, y) = f(b\,x - a\,y) \tag{6.6}$$

Since f can be any function as long as its derivative exists, we have found a multitude of solutions to (6.3). In some sense (6.6) can be seen as the general solution to (6.3) and we want to investigate how we can consistently fix boundary conditions to obtain a special solution. To do so, we need to introduce a new term.

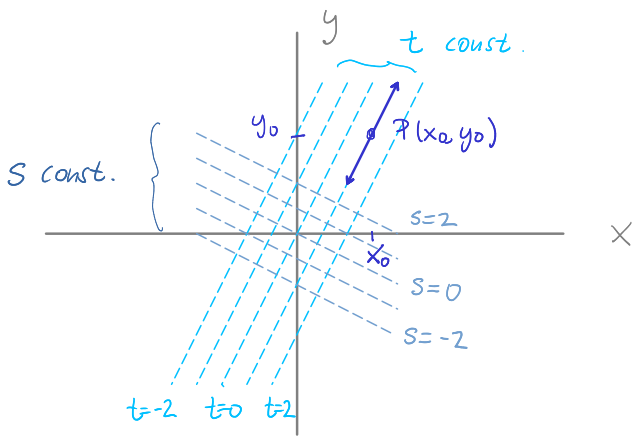

Figure 6.1: Characteristics of the PDE (6.3) for $a = 1$ and $b = 2$ (azure lines). Along these, t is constant. Lines along which s is constant are perpendicular to the characteristics and are shown in darker blue.

Contours in the x-y plane with $t = c$ constant are called *characteristics* *[Charakteristik]* of the PDE. In our case, (6.3), the equation describing the characteristics is

$$y = \frac{b}{a}x - \frac{t}{a} \qquad \text{or} \qquad \frac{dy}{dx} = \frac{b}{a} \tag{6.7}$$

i.e. straight lines with slope b/a. For each value of c we get a characteristics. For the concrete case $a = 1$, $b = 2$ they are shown in Figure 6.1 in azure, for $t \in \{2, 1, 0, -1, -2\}$. Along these characteristics t is constant, but s changes. The solution $\Psi(x, y)$ changes w.r.t. s (i.e. along the characteristics) as governed by the ODE (6.3). Thus, if we e.g. fix $\Psi(x_0, y_0)$, as indicated by the point P in Figure 6.1, then Ψ is determined everywhere along the characteristics through the point P by (6.3) . When fixing boundary conditions, we better make sure we do not impose a value of $\Psi(x, y)$ for any other point on this characteristics! This is bound to lead to an inconsistency.

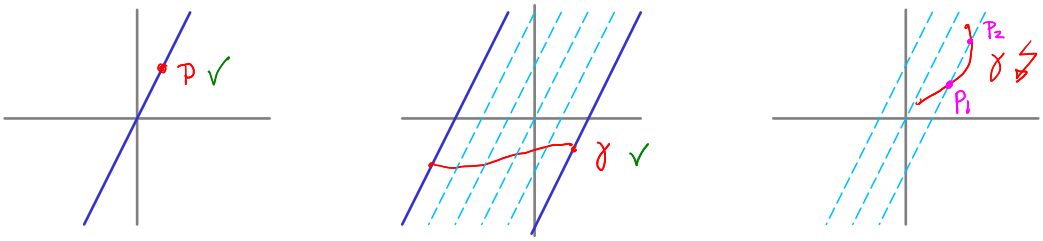

Figure 6.2: Left panel: fixing Ψ at P determines Ψ everywhere on the (blue) characteristics. Middle panel: fixing Ψ on γ determines Ψ everywhere between the two dark blue characteristics. Right panel: imposing boundary conditions on Ψ on γ will lead to inconsistencies and no solution in general.

How then can we fix boundary conditions without messing up everything? This is illustrated in Figure 6.2, still for (6.3) with $a = 1$ and $b = 2$. As just discussed, if we fix Ψ at a single point through a boundary condition (red point P, left panel), the function Ψ is determined for all points on the characteristics through the point P. We can impose more boundary conditions, as long as it is not for another point on the same characteristics. An example is the red contour γ (middle panel). This contour intersects any characteristics at most once. Thus we are free to fix Ψ on γ. If we do this, Ψ is determined for all points that lie between the two dark blue characteristics in the middle panel of Figure 6.2. A case when things go wrong is illustrated in the right panel of Figure 6.2. There, the path γ crosses at least one characteristics more than once. Thus, if we fix Ψ at P_1, this determines Ψ at P_2 and we are not allowed to impose boundary conditions on this contour if we are keen to get a solution.

In a next step we make our PDE more complicated in that we allow the coefficients a and b of (6.3) to depend on x and y, i.e.

$$a(x, y)\,\frac{\partial \Psi}{\partial x} + b(x, y)\,\frac{\partial \Psi}{\partial y} = 0 \tag{6.8}$$

To find the characteristics of this PDE, we are looking for curves $(x(t), y(t))$ in the x-y plane, such that $\Psi(x(t), y(t))$ is independent of t. You can think of such a curve as a contour in the complex $z = x(t) + iy(t)$ plane, very much like (1.21). The independence of $\Psi(x(t), y(t))$ on t means

$$0 = \frac{d}{dt}\Psi(x(t), y(t)) = \partial_x \Psi\,\frac{dx(t)}{dt} + \partial_y \Psi\,\frac{dy(t)}{dt} = \Big(-\frac{b(x, y)}{a(x, y)}\frac{dx}{dt} + \frac{dy}{dt}\Big)\partial_y \Psi \tag{6.9}$$

from which we get an ODE

$$\frac{dy}{dx} = \frac{b(x,y)}{a(x,y)} \tag{6.10}$$

to find the characteristics. This is precisely the last expression in (6.7) that apparently also holds for x- and y-dependent coefficients a and b.

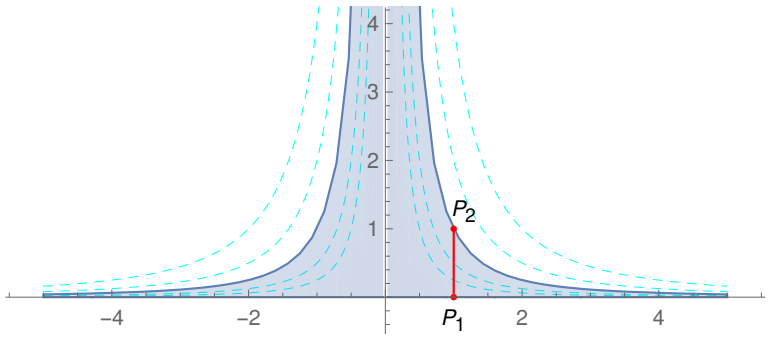

Figure 6.3: Characteristics of the PDE (6.11) for $c \in \{1/4, 1/2, 2, 4\}$ as dashed azure lines and for $c = 1$ as blue line in the region $y > 0$. The boundary condition along the red contour fixes the value of the solution within the coloured area.

This cries out for an example. Let us consider the PDE

$$x\,\frac{\partial \Psi}{\partial x} - 2y\,\frac{\partial \Psi}{\partial y} = 0 \tag{6.11}$$

We want to find a solution with the boundary condition $\Psi(x,y) = 2y + 1$ on the contour $\gamma : t \in [0,1] \to (x(t) = 1, y(t) = t)$, i.e. the straight line from P_1 to P_2 in Figure 6.3. With the help of (6.10) we find for each constant c a characteristics ⇒ lecture

$$y = \frac{c}{x^2} \tag{6.12}$$

Hence, the general solution to (6.11) is $\Psi(x,y) = f(x^2\,y)$ where f is an arbitrary function. To satisfy the boundary condition we set $f(x^2\,y)|_\gamma = f(y)|_\gamma = 2y + 1$ and we get our solution as

$$\Psi(x,y) = 2\,x^2\,y + 1 \tag{6.13}$$

within the blue shaded region of Figure 6.3. Outside this region, the solution is not fixed by the boundary condition. If we want to fix it everywhere, we have to fix the boundary condition on a contour that is intersected by every characteristics precisely once.

As we can see, fixing consistent boundary conditions to determine the solution everywhere is quite a bit more involved for first order PDE than ODE. In order to help us with this issue in more complicated cases and, in particular, second order PDE, let us look at this once more in a slightly different way.

If we have a boundary condition for Ψ in (6.8) along a contour γ parameterized as $(x(t), y(t))$ we get two equations for the two 'unknowns' $\partial_x\Psi$ and $\partial_y\Psi$. They can be written as

$$\begin{pmatrix} a & b \\ \frac{dx}{dt} & \frac{dy}{dt} \end{pmatrix} \cdot \begin{pmatrix} \partial_x\Psi \\ \partial_y\Psi \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{d\Psi}{dt} \end{pmatrix} \tag{6.14}$$

This is simply (6.8) and (6.9) (with $\frac{d\Psi}{dt} \neq 0$) rewritten in matrix form. This system of equations can be solved (for a particular point (x, y)), if the determinant of the 2×2 matrix is not zero, i.e. if

$$a \frac{dy}{dt} - b \frac{dx}{dt} \neq 0 \tag{6.15}$$

Hence, the system has a solution if the contour does *not* satisfy the equation for the characteristics, (6.10). If we want a solution everywhere, the contour must not coincide with (i.e. be parallel to) a characteristics anywhere. In other words, every characteristics has to intersect the contour precisely at one point. Note that this line of argumentation does not rely on the r.h.s. of (6.8) being zero. It also applies to non-homogeneous first order PDE.

If we go from two to three independent variables we have to fix the function on a two dimensional surface (parameterized by two parameters t_1 and t_2) such that each characteristics intersects this surface precisely once. (6.14) becomes a 3×3 system. One equation is the PDE, the other two are (6.9) for the parameters t_1 and t_2. Again, we must ensure the determinant of the corresponding 3×3 matrix is non-vanishing which means every characteristics must not lie within the surface anywhere, but simply intersect it at precisely one point.

6.2 Second order PDE

We could spend much more time with first order PDE. But the main points have been made. We want to determine the characteristics to convert the PDE to an ODE and then we have to be careful to fix consistent boundary conditions. While a systematic treatment of second order PDE is well beyond the scope of this lecture, this knowledge will guide us in our crash course. We will understand the different types of second order PDE and why the naive expectation of how to set boundary conditions often fails.

So here it is, the general linear second order PDE with (for the moment) two independent variables

$$a\, \partial_{xx}\Psi + b\, \partial_{xy}\Psi + c\, \partial_{yy}\Psi = F(x, y, \Psi, \partial_x\Psi, \partial_y\Psi) \tag{6.16}$$

where a and b are functions of x and y, as in (6.8), and in the inhomogeneous case we have a non-vanishing r.h.s. given by the function F that is supposed to be linear in Ψ, $\partial_x\Psi$, and $\partial_y\Psi$. From our experience of linear second order ODE we might expect we can set boundary conditions by specifying $\Psi(x, y)$ as well as the normal derivative $\partial\Psi(x, y)/\partial n \equiv \nabla\Psi \cdot n$ along a suitably chosen contour. Such boundary conditions are called *Cauchy boundary conditions [Cauchy Randbedingungen]*. Let us now try to impose such boundary conditions and see what happens.

Once more, we start from a contour γ determined by points $(x(t), y(t))$ as a function of a parameter t. We require

$$\text{Cauchy:} \qquad \Psi(x, y)\big|_\gamma = \alpha(t) \qquad \text{and} \qquad \frac{\partial\Psi(x, y)}{\partial n}\Big|_\gamma = \beta(t) \tag{6.17}$$

where the functions α and β are the boundary conditions, as illustrated in Figure 6.4. The two boundary conditions can be solved for $\partial_x\Psi = f_x$ and $\partial_y\Psi = f_y$, where the precise form of the functions $f_{x/y}(d\alpha/dt, \beta, dx/dt, dy/dt)$ does not matter for the moment. In a second step we express $df_{x/y}/dt$ in terms of second partial derivatives, and with the PDE (6.16) we ⇒ lecture

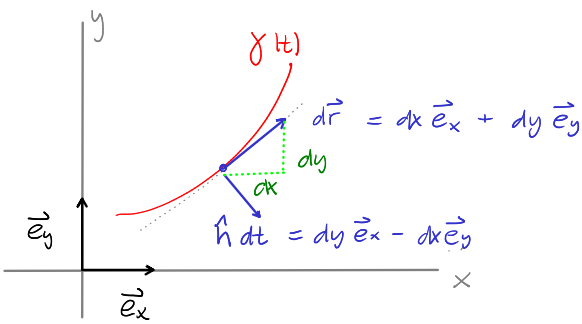

Figure 6.4: Fixing Cauchy boundary conditions (6.17) along a contour/boundary γ.

obtain a system of 3 equations with 3 unknowns as

$$\begin{pmatrix} a & b & c \\ \frac{dx}{dt} & \frac{dy}{dt} & 0 \\ 0 & \frac{dx}{dt} & \frac{dy}{dt} \end{pmatrix} \cdot \begin{pmatrix} \partial_{xx}\Psi \\ \partial_{xy}\Psi \\ \partial_{yy}\Psi \end{pmatrix} = \begin{pmatrix} F \\ \frac{df_x}{dt} \\ \frac{df_y}{dt} \end{pmatrix} \tag{6.18}$$

This is the generalization of (6.14) to a second order PDE with Cauchy boundary conditions. Again, we ask the question whether this system can be solved. Note that the precise form of the r.h.s. has no impact on the answer. But for a solution to exist, we need the determinant of the 3×3 matrix in (6.18) to be different from zero. Computing the determinant and multiplying by $(dt/dx)^2$ we find that a solution to (6.16) does *not* exist if ⇒ lecture

$$a\left(\frac{dy}{dx}\right)^2 - b\left(\frac{dy}{dx}\right) + c = 0 \tag{6.19}$$

This is an ODE for the characteristics. It is the generalization of (6.10) to the second order PDE (6.19). As it is a quadratic equation we have to distinguish three cases for the potential solutions

$$\left(\frac{dy}{dx}\right) = \frac{b \pm \sqrt{b^2 - 4\,ac}}{2a} \qquad \begin{cases} b^2 > 4\,ac & \text{hyperbolic PDE} \\ b^2 = 4\,ac & \text{parabolic PDE} \\ b^2 < 4\,ac & \text{elliptic PDE} \end{cases} \tag{6.20}$$

If a, b or c depend on x and y the nature of the PDE might depend on the region of the x-y plane.

In the case of a *hyperbolic [hyperbolische]* PDE we get two different ODE and, therefore, two different sets of (real) characteristics. To see this we write

$$\Big(a\,\partial_{xx} + b\,\partial_{xy} + c\,\partial_{yy}\Big)\Psi = a\left(\partial_x + \frac{b - \sqrt{b^2 - 4\,ac}}{2a}\partial_y\right)\left(\partial_x + \frac{b + \sqrt{b^2 - 4\,ac}}{2a}\partial_y\right)\Psi \tag{6.21}$$

to obtain a 'product' of two real first order PDE. Each has a set of characteristics and the two conditions in (6.17) potentially can be used to uniquely fix the solution.

Let us make a concrete example and look at the wave equation (6.2e) for a one-dimensional wave $\Psi(x,t)$, the 'classic' example of a hyperbolic PDE. The equation reads

$$\frac{\partial^2\,\Psi(x,t)}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2\,\Psi(x,t)}{\partial t^2} = \left(\partial_x - \frac{1}{c}\partial_t\right)\left(\partial_x + \frac{1}{c}\partial_t\right)\Psi = 0 \tag{6.22}$$

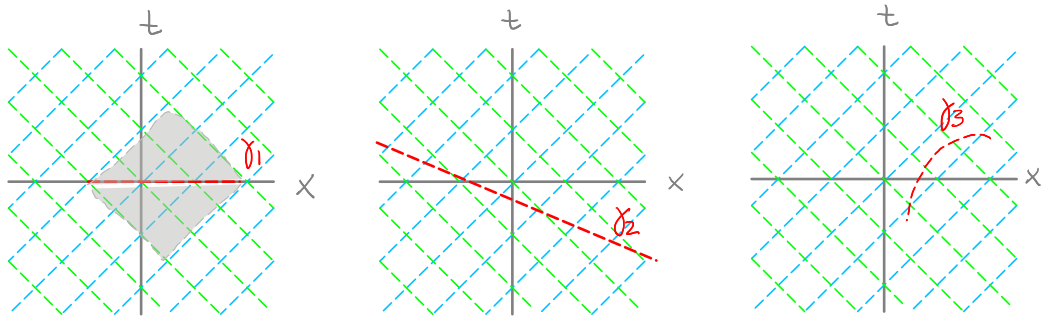

Figure 6.5: Characteristics of the wave equation corresponding to $x - ct$ constant (azure) and $x + ct$ constant (green). Fixing Cauchy conditions along γ_1 (left panel) determines a solution in the shaded region; along (non-ending) γ_2 (middle panel) everywhere; along γ_3 (right panel) results in no solution.

and corresponds to (6.16) with $a \to 1$, $b \to 0$, $c \to -1/c^2$, and $F = 0$. The equations for the characteristics follow from (6.20) as

$$\frac{dt}{dx} = \pm\frac{1}{c} \qquad \Rightarrow \qquad \frac{dx}{dt} = \pm c \qquad \Rightarrow \qquad x \pm ct = \text{constant} \tag{6.23}$$

and the general solution to (6.22) is $\Psi(x,t) = f_1(x - ct) + f_2(x + ct)$. Since we are dealing with a hyperbolic PDE, everything in (6.23) is real, we have two sets of characteristics, shown in azure and green in Figure 6.5, and the solution is a sum of two real functions f_1 and f_2.

Note the difference between Figure 6.1 and Figure 6.5. In the former, there is only one set of characteristics (azure), as it is a first order PDE. The (darker blue) lines perpendicular to those are just for illustration of the new coordinate system. These are not characteristics. However, in Figure 6.5 both sets, the azure and green lines are chracteristics of the PDE. We have two sets, since we are dealing with a hyperbolic second order PDE.

If we choose a suitable contour γ in the x-t plane such that the characteristics of each set intersect the contour at most once, we can consistently impose the two boundary conditions (6.17) on γ to fix the two functions f_1 and f_2. This determines the solution to (6.22) in the domain of the x-t plane covered by the characteristics running through γ, as illustrated by γ_1 in Figure 6.5. If we want to fix the solution *everywhere*, we have to set Cauchy boundary conditions on a contour that is intersected by *each* characteristics of both sets precisely once, such as γ_2. If we take a γ_3 such that a single characteristics of either set intersects twice, we get no solution.

To summarise, for a hyperbolic PDE, if we are "careful" in the choice of the boundary γ, we can impose Cauchy boundary conditions and find a solution.

In the case of elliptic (or parabolic) PDE this is not the case. If we were to use (6.21) we would find no (or a single) real set of characteristics. As a result, we will not have enough freedom in the general solution to be able to accommodate Cauchy boundary conditions.

Let us make this a bit more concrete first for a *parabolic [parabolische]* PDE. We take the

simple example

$$\frac{\partial^2 \Psi(x,y)}{\partial x^2} + 2\,\frac{\partial^2 \Psi(x,y)}{\partial x \partial y} + \frac{\partial^2 \Psi(x,y)}{\partial y^2} = 0 \tag{6.24}$$

Using (6.19) we get a single equation (corresponding to the azure characteristics in Figure 6.5)

$$\frac{dy}{dx} = 1 \qquad \Rightarrow \qquad x - y = \text{constant} \tag{6.25}$$

and, hence, our general solution is $\Psi(x,y) = f_1(x-y)$. Sadly, we have only a single function f_1 at our hands to satisfy the boundary conditions. Requiring both conditions in (6.17) is asking too much of f_1. However, with a bit of luck we can impose either of the two conditions, corresponding to *Dirichlet [Dirichlet]* or *Neumann [Neumann]* boundary conditions

$$\text{Dirichlet:} \qquad \Psi(x,y)\big|_\gamma = \alpha(t) \tag{6.26a}$$

$$\text{Neumann:} \qquad \frac{\partial \Psi(x,y)}{\partial n}\bigg|_\gamma = \beta(t) \tag{6.26b}$$

Of course, we still can mess it up by choosing a stupid contour for the boundary conditions, such as γ_3 in Figure 6.5.

Boundary conditions that fix the solution everywhere for hyperbolic and parabolic PDE are to be given on a open contour, or going to more than two independent variables, on an open hypersurface. We can think of a hypersurface simply as a $n-1$ dimensional surface in an n dimensional space. With 2 independent variables, it is just a (non-ending) line, like γ_2. If we had 3 independent variables, we can fix boundary conditions on an (suitably chosen) open 2-dimensional surface (e.g. a plane).

For elliptic PDE things are different. Here we do not even have a single real characteristics. In this case, a general rule of thumb is that we can impose either Dirichlet or Neumann boundary conditions on a closed hypersurface. To give a little bit of substance to this claim, let us consider the Laplace (or Poisson) PDE

$$\frac{\partial^2 \Psi(x,y)}{\partial x^2} + \frac{\partial^2 \Psi(x,y)}{\partial y^2} = 0 \qquad (\text{or } = \rho) \tag{6.27}$$

Referring to (6.16) we have $a = 1$, $b = 0$, $c = 1$ and this is our prime example for an elliptic PDE. We will study it thoroughly in Section 7 in three dimensions and, as a preparation, we present here a uniqueness theorem:

Let $\Psi(\vec{r})$ be a solution to the Laplace equation (6.2a) or Poisson equation (6.2b) with continuous partial derivatives in a region V and on its boundary $\sigma = \partial V$, satisfying Dirichlet (Neumann) boundary conditions on σ. Then Ψ is the only such solution (up to an additive constant).

To show this, let us assume that we have two functions $\Psi_1(x,y)$ and $\Psi_2(x,y)$ that satisfy (6.27). Then $\Phi = \Psi_1 - \Psi_2$ satisifies $\Delta\Phi = 0$ with zero (Dirichlet or Neumann) boundary conditions. Consider ⇒ lecture

$$0 = \int\limits_{\sigma=\partial V} \Phi \frac{\partial \Phi}{\partial n}\, d\sigma = \int\limits_{\sigma=\partial V} \Phi \vec{\nabla}\Phi\, d\sigma = \int\limits_V \vec{\nabla}\big(\Phi\vec{\nabla}\Phi\big)\, dV = \int\limits_V \big(\Phi\Delta\Phi + |\vec{\nabla}\Phi|^2\big)\, dV \tag{6.28}$$

Since $|\vec{\nabla}\Phi|^2 \geq 0$ we conclude $\vec{\nabla}\Phi = 0$ in V. Thus Ψ_1 and Ψ_2 differ at most by a constant. In the case of Dirichlet boundary conditions (6.26a), the constant must be zero.

This uniqueness theorem will turn out to be extremely useful in Section 7 in that if we have found a solution, we know it is unique. However, it is by no means guaranteed that we can find a solution. Taking Neumann boundary conditions, $\partial\Psi/\partial n = f(\vec{r})$, $\forall\, \vec{r} \in \sigma$, where $\sigma = \partial V$, there is a solvability condition. Indeed, if $\Delta\Psi = \rho$, we have

$$\int_V \rho(\vec{r})\, d^3\vec{r} = \int_V \Delta\Psi\, dV = \int_V \vec{\nabla}\cdot\vec{\nabla}\Psi\, dV = \int_{\sigma=\partial V} \frac{\partial\Psi}{\partial n}\, d\sigma = \int_\sigma f(\vec{r})\, d\sigma \tag{6.29}$$

i.e. a condition on the integral of the function $f(\vec{r})$ over the full boundary.

In the case of Dirichlet we insist on continuous boundary conditions. After all, we are typically interested in continuous solutions, in fact even solutions with continuous partial derivatives, used as an assumption in the uniqueness theorem.

To conclude our generic deliberations on second order PDE we stress again that in this section we only presented the basic concepts and ideas. A more serious analysis would require dedicated lecture courses. But a general rule of thumbs is that if we are "careful" (beware of γ_3) we can specify boundary conditions for second order PDE as follows:

PDE	boundary condition	hypersurface
hyperbolic	Cauchy	open
parabolic	Dirichlet or Neumann	open
elliptic	Dirichlet or Neumann	closed

Armed with this knowledge, in the next section we will look for explicit solutions to concrete PDE showing up in physics, (6.2). In true Hollywood style, we will be reunited with many good old friends, sadly missed since Section 5.

7 Special partial differential equations

This section is devoted to finding solutions for some of the most important PDE in physics. Roughly speaking, we will be working our way through the equations listed in (6.2).

Since we will typically be working in 3 dimensions, our PDE will get rather involved. The choice of the coordinate system has to be adapted to the physical situation. We start by working in cartesian coordinates in 3 dimensions, i.e. $\vec{r} = (x, y, z)$ and then consider spherical $\vec{r} = (r, \theta, \phi)$ and cylindrical $\vec{r} = (r, \phi, z)$ coordinate systems. In any case, our PDE will contain 3 or in the case of time dependence 4 independent variables. This looks daunting. In order to cope with this, throughout this section we will apply a very important concept, *separation of variables [Separation der Variablen]*. The idea is to look for solutions that factorize, e.g. $\Psi(\vec{r}, t) = T(t)\,\psi(\vec{r})$. Using this as an ansatz we aim to transform the PDE for $\Psi(\vec{r}, t)$ into an ODE for $T(t)$ and a PDE for $\psi(\vec{r})$. We can continue with this procedure and factorize $\psi(\vec{r}) = \psi_1(r_1)\,\psi_2(r_2)\,\psi_3(r_3)$, where $\vec{r} = (r_1, r_2, r_3)$ according to the coordinate system we have chosen. As we will see, with a bit of luck we might be able to reformulate the PDE in terms of ODE for each of the functions $\psi_i(r_i)$ and $T(t)$.

Let us illustrate this by making a connection between the Helmholtz equation (6.2d) and the wave equation (6.2e). If we are looking for solutions to (6.2e) in the form of

$$\Psi(\vec{r}, t) = T(t)\,\psi(\vec{r}) \qquad \xrightarrow{(6.2e)} \qquad \frac{1}{\psi(\vec{r})}\,\Delta\psi(\vec{r}) = \frac{1}{T(t)}\frac{1}{c^2}\frac{d^2T(t)}{dt^2} \tag{7.1}$$

The crucial argument is now made by considering the second equation in (7.1). The l.h.s. is obviously independent of t. The r.h.s. is obviously independent of r. Since the l.h.s. equals the r.h.s. both of them cannot depend on either r or t. Hence, they must be a constant. Let us denote this constant by $-k^2$. Then, we can rewrite (7.1) as

$$\Delta\psi(\vec{r}) = -k^2\,\psi(\vec{r}) \qquad \text{and} \qquad \frac{d^2T(t)}{dt^2} = -c^2k^2\,T(t) \tag{7.2}$$

The second equation is a second order ODE for which we can easily find the general solution as $T(t) = c_1\,e^{ick\,t} + c_2\,e^{-ick\,t}$ with c_1 and c_2 constants. The first equation in (7.2) is nothing but the Helmholtz equation (6.2d). It can be seen as a time-independent wave equation. In fact, the time-independent and time-dependent Schrödinger equations, (6.2f) and (6.2g), are related in precisely the same way.

So far so good. But this procedure leads to several questions. Have we found all solutions by making the factorized ansatz, or are there some we have missed? Can we always do this cheap trick?

The answer to the second question is "no". However, as we will see, separation of variables will lead to solutions in many important cases. Regarding the first question, we will argue case by case. We will look at the solutions we obtain and see whether they are complete. In doing so, we will have to refer to concepts that will only be carefully considered in MMP 2. In particular the notion of what 'complete' actually means will remain somewhat murky. In any case, the fact that some of the solutions to ODE found in Section 5 form a basis will be of crucial importance, as will the discussion at the end of Section 6.2.

7.1 Laplace equation in cartesian coordinates

The first equation we consider is the Laplace equation (6.2a) or its generalisation, the Helmholtz equation (6.2d) in cartesian coordinates. Using separation of variables we will look for solutions to

$$\Delta\Psi(x,y,z) + k^2\,\Psi(x,y,z) = 0 \tag{7.3a}$$

in the form

$$\Psi(x,y,z) = X(x)\,Y(y)\,Z(z) \tag{7.3b}$$

Before we go on to find solutions, let us come back to the questions whether they are unique. In the case of the Laplace equation we refer to the discussion at the end of Section 6.2 in connection with (6.27). There we have seen that for the Laplace equation, imposing Dirichlet or Neumann boundary conditions on a closed hypersurface will allow at most one solution (up to a constant). Thus, if we find a solution of the form (7.3b) to such an boundary condition problem, we know it will be the 'one and only'.

So let's see if we can find one. Plugging the ansatz (7.3b) into the equation (7.3a) we get ⇒ lecture

$$\frac{X''(x)}{X(x)} = -\frac{Y''(y)}{Y(y)} - \frac{Z''(z)}{Z(z)} - k^2 \equiv -l^2 \tag{7.4}$$

where the (ordinary) derivatives are w.r.t. the variable of the function. Reasoning as in connection with (7.2) we conclude that the expression in (7.4) must be a constant (i.e. independent of x, y and z) that we denote by $-l^2$. Following the same line of arguments for Y''/Y and Z''/Z we find

$$\frac{X''(x)}{X(x)} = -l^2; \qquad \frac{Y''(y)}{Y(y)} = -m^2; \qquad \frac{Z''(z)}{Z(z)} = -n^2; \qquad l^2 + m^2 + n^2 = k^2; \tag{7.5}$$

Thus we have enormously simplified the problem. Rather than dealing with a PDE with 3 independent variables, we have 3 decoupled (i.e. independent of each other) ODE. For this to work, it is essential that k^2 factorizes itself. If k^2 is of the form $k^2(x,y,z) = f(x)+g(y)+h(z)$, separation of variables will lead to three decoupled ODE. Since in our case k^2 is a constant, the ODE become particularly simple, namely a second order ODE with constant coefficients. Thus X, Y, and Z can be given in terms of exponential or trigonometric functions.

To illustrate this further, we consider a concrete example. For simplicity we work in 2 dimensions, i.e. $\Psi(x,y)$, but the concepts are precisely the same in 3 (and, actually n) dimensions. Our task is to find a solution to the Laplace equation $\Delta\Psi(x,y) = 0$ in the region V with Dirichlet boundary conditions on the hypersurface ∂V. The region V is the rectangle in the x-y plane determined by the points $(x,y) = \{(0,0),(\pi,0),(0,\pi),(\pi,\pi)\}$ and we demand that $\Psi = 0$ on the boundary except for the top line, where $\Psi(x,y=\pi) = f(x)$. To be concrete, we set $f(x) = (\pi - x)\,x^2/4$, see Figure 7.1. Note that we have chosen $f(x)$ such that the boundary conditions are continuous, i.e. $f(0) = f(\pi) = 0$.

Using separation of variables $\Psi(x,y) = X(x)\,Y(y)$ we arrive at

$$X''(x) = -l^2\,X(x) \qquad \Rightarrow \qquad X_l(x) = a_{xl}\sin(lx) + b_{xl}\cos(lx) \tag{7.6a}$$

$$Y''(y) = l^2\,Y(y) \qquad \Rightarrow \qquad Y_l(y) = a_{yl}\,e^{ly} + b_{yl}\,e^{-ly} \tag{7.6b}$$

In (7.6) we have decided to write $X_l(x)$ in terms of trigonometric functions and $Y_l(y)$ in terms of exponential functions. At this moment, this is just a choice. The important observation is

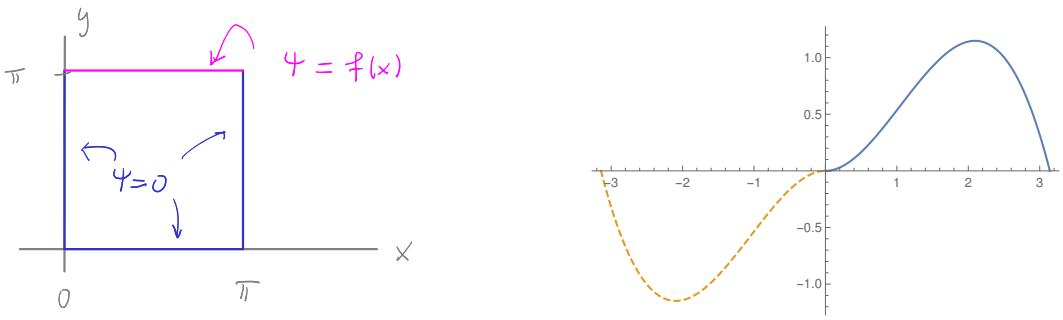

Figure 7.1: Boundary value problem for the two-dimensional Laplace equation. The function $f(x) = (\pi - x)\, x^2/4$ periodically extended to the region $-\pi < x < \pi$ is shown in the right panel.

that $\forall\, l$ we obtain four constants a_{xl}, b_{xl}, a_{yl}, and b_{yl}. As we will see, with the choice above, they will turn out to be real.

Ignoring boundary conditions for the moment, we can write the general solution as

$$\Psi(x,y) = \sum\!\!\!\!\!\!\int_l \Big(k_{l1}\sin(lx)e^{ly} + k_{l2}\cos(lx)e^{ly} + k_{l3}\sin(lx)e^{-ly} + k_{l4}\cos(lx)e^{-ly}\Big) \tag{7.7}$$

still with four constants k_i for each value of l. Since at the moment $l \in \mathbb{R}$ the sum over l is actually an integration $\int dl$. The reason we have written the funny summation-integration sign is that in the next step, we will get constraints on l and select discrete values.

We now apply the boundary conditions, one by one ⇒ lecture

$$\Psi(0,y) = 0 = \sum\!\!\!\!\!\!\int_l \Big(k_{l2}\, e^{ly} + k_{l4}\, e^{-ly}\Big) \qquad \Rightarrow \qquad k_{l2} = k_{l4} = 0 \tag{7.8a}$$

$$\Psi(x,0) = 0 = \sum\!\!\!\!\!\!\int_l \big(k_{l1} + k_{l3}\big)\sin(lx) \qquad \Rightarrow \qquad k_{l1} = -k_{l3} \tag{7.8b}$$

$$\Psi(\pi,y) = 0 = \sum\!\!\!\!\!\!\int_l 2\, k_{l1}\sin(l\pi)\sinh(ly) \qquad \Rightarrow \qquad \sin(l\pi) = 0 \tag{7.8c}$$

It is the last condition $\sin(l\pi) = 0 \Rightarrow l \in \mathbb{Z}$ that converts the integration over l into a sum, since only discrete values of l are allowed. In fact, it is sufficient to consider $l \in \mathbb{N}$ since negative values of l just add a factor (-1) to the constant k_{l1}. Insisting on the last remaining boundary condition, the non-trivial boundary condition for the top magenta boundary of Figure 7.1, we require

$$\Psi(x,\pi) = \sum_{l\in\mathbb{N}} 2\, k_{l1}\sin(lx)\sinh(l\pi) = f(x) = \frac{x^2}{4}(\pi - x) \tag{7.9}$$

Just when we thought we might never need the Fourier series again, here it is, back in its full glory. To match (7.9) we extend $f(x)$ defined in $0 \le x \le \pi$ to make it an odd function $\bar{f}(x)$

in the range $-\pi \leq x \leq \pi$, as shown in the right panel of Figure 7.1. Computing its Fourier series we find

$$\bar{f}(x) = \sum_{n=1}^{\infty} b_n \sin(nx) \qquad \text{with} \qquad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} dx\, \bar{f}(x)\sin(nx) = -\frac{1+2(-1)^n}{n^3} \tag{7.10}$$

Finally, we arrive at our unique solution ⇒ lecture

$$\Psi(x,y) = \sum_{l\in\mathbb{N}} \frac{b_l}{\sinh(l\pi)} \sin(lx)\sinh(ly) = \sum_{l=1}^{\infty} \frac{-(1+2(-1)^l)}{l^3\,\sinh(l\pi)} \sin(lx)\sinh(ly) \tag{7.11}$$

Had we chosen a different $f(x)$ for the boundary condition at the top, this simply would modify the coefficients b_l in (7.11) according to the Fourier series of the boundary condition. The other boundary conditions have been chosen in a way that was convenient to show how the multitude of coefficients in the general solutions are fixed. Needless to say that if these conditions are more complicated, the work to be done is more tedious. But, as long as we fix continuous boundary conditions, we will get a unique solution.

7.2 Laplace equation in spherical coordinates

Very often we are dealing with situations where spherical coordinates are better suited. In these cases we need to express the Laplace operator (6.1) in terms of (r,θ,ϕ) where

$$\begin{aligned} x &= r\,\sin\theta\,\cos\phi \\ y &= r\,\sin\theta\,\sin\phi \\ z &= r\,\cos\theta \end{aligned} \tag{7.12}$$

and $0 \leq r < \infty$, $0 \leq \phi < 2\pi$, and $0 \leq \theta \leq \pi$ often implemented as $-1 \leq \cos\theta \leq 1$. Expressing the partial derivatives in (6.1) through partial derivatives w.r.t. r, ϕ and θ is somewhat tedious, but straightforward (see Section A) and we get

$$\Delta = \Big(\frac{1}{r}\frac{\partial^2}{\partial r^2}\,r + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta} + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\Big) \tag{7.13}$$

The first term in (7.13) can be written in several useful ways

$$\frac{1}{r}\frac{\partial^2}{\partial r^2}\,r = \frac{1}{r^2}\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r} = \frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r} \tag{7.14}$$

To find solutions to the Laplace equation in spherical coordinates, we use separation of variables in analogy to (7.3b) as

$$\Psi(r,\theta,\phi) = \frac{U(r)}{r}P(\theta)Q(\phi) \tag{7.15}$$

where the additional factor of $1/r$ will turn out to be a convenient convention. Inserting the ansatz (7.15) into (7.13) and following the line of argumentation leading to (7.4) we find ⇒ lecture

$$\frac{Q''(\phi)}{Q(\phi)} = -m^2 \qquad \Rightarrow \qquad Q(\phi) = e^{\pm im\,\phi} \tag{7.16}$$

To start with, m^2 is just a constant, precisely as l^2 in (7.4). However, if we insist on the uniqueness of $Q(\phi)$ under $\phi \to \phi + 2\pi$ we must have $m \in \mathbb{Z}$. For the special value $m = 0$ the function $\Psi(r, \theta, \phi)$ does not depend on ϕ.

Following the same procedure for $U(r)$ and $P(\theta)$ we get ⇒ lecture

$$r^2 \frac{U''(r)}{U(r)} = l(l+1) \qquad \text{and} \qquad \frac{1}{P(\theta)\,\sin\theta}\frac{d}{d\theta}\big(\sin\theta\, P'(\theta)\big) - \frac{m^2}{\sin^2\theta} = -l(l+1) \tag{7.17}$$

where for reasons to be revealed we have decided to write the constant as $l(l+1)$. Once more, we managed to transform a PDE with three independent variables into a set of three ODE. This can also be done for (7.3a) with $k^2(r, \theta, \phi) = f(r) + g(\theta) + h(\phi)$. Thus, if our problem factorises in terms of (r, θ, ϕ) but not in (x, y, z) or has boundary conditions given in terms of (r, θ, ϕ), using spherical coordinates is the obvious choice.

In the case of cartesian coordinates, the ODE were trivial, (7.5). Here, they seem complicated. However, the second equation in (7.17) is a good old friend, the Legendre ODE. If we change variables from θ to x (an unfortunate clash of notation, this x is not a cartesian coordinate) through $\cos\theta = x$, hence $\sin^2\theta = 1 - x^2$, and $d/d\theta = -\sin\theta\, d/dx$, the ODE for $P(\theta)$ can written as ⇒ lecture

$$\frac{d}{dx}(1-x^2)P'(x) + \Big(l(l+1) - \frac{m^2}{1-x^2}\Big)P(x) = 0 \tag{7.18}$$

A glance at (5.11) saves us a lot of work.

As we have seen in Section 5.1 and Section 5.2 the only solutions that are defined for all values of θ, i.e. everywhere in $-1 \le x = \cos\theta \le 1$ are the (associated) Legendre polynomials, with $l \in \mathbb{N}$. This also explains the choice of writing the constant in (7.17) as $l(l+1)$. Combining $P(\theta)\,Q(\phi)$ we obtain precisely the spherical harmonics, introduced in (5.19).

What about the ODE for $U(r)$. This is a second order ODE (of the Euler type) that can be solved by making an ansatz $U(r) \sim r^{\text{pow}}$ and we get the general solution ⇒ lecture

$$\frac{U(r)}{r} = \frac{1}{r}\Big(k_{l1}\, r^{l+1} + k_{l2}\, r^{-l}\Big) \tag{7.19}$$

As expected, we have two linearly independent solutions. Since $l \ge 0$ one of them, r^{l+1}, leads to a solution that is regular at the origin and one, r^{-l}, yields a solution that is singular at the origin but potentially well defined for $r \to \infty$. Both of them can be relevant, depending on the nature of the physics problem. Assembling the ingredients of (7.15) we obtain the general solution to the Laplace equation in spherical coordinates as

$$\Psi(r, \theta, \phi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\Big(a_{lm}\, r^l\, Y_l^{(m)}(\theta, \phi) + b_{lm}\, r^{-l-1}\, Y_l^{(m)}(\theta, \phi)\Big) \tag{7.20}$$

It might seem strange that (7.20) contains only discrete labels (i.e. the solution is a sum) whereas (7.7) contains a continuous label. The difference is due to us insisting that (7.20) is defined $\forall\,\theta$ and $\forall\,\phi$. If we restrict the range of ϕ and θ, (7.16) and (7.17) do not any longer entail that m and l are integers.

The constants a_{lm} and b_{lm} can be used to try to satisfy boundary conditions. We have written 'try to satisfy' because it is by no means guaranteed that a solution does exist.

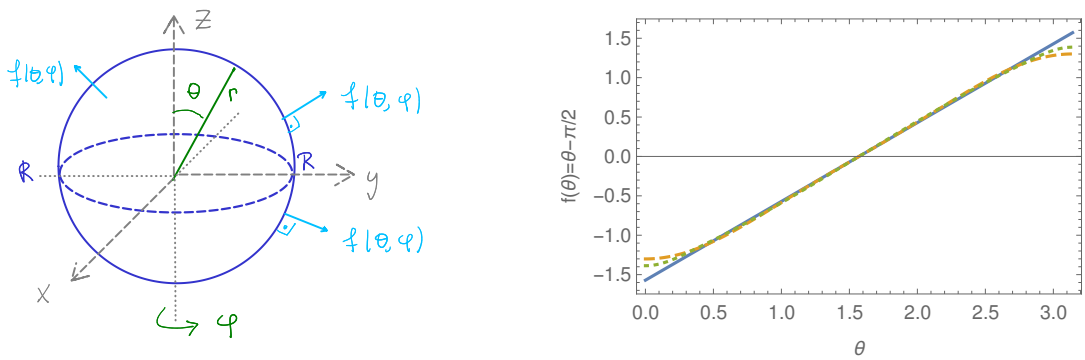

Figure 7.2: Neumann boundary value problem for Laplace equation. The volume V is a sphere of radius R, its boundary $\partial V = \sigma$ is the surface denoted by $d\Omega_2$ earlier, e.g. in (5.40). In this case, the normal derivative is simply the derivative w.r.t. r. The right panel shows the approximation of the Legendre series (7.26) to the full function $f(\theta) = \theta - \pi/2$ (blue) if using terms up to $l = 3$ (orange dashed) or $l = 11$ (green dotted).

In order to illustrate this, we consider a concrete problem. We want to find the solution to the Laplace equation (i.e. $\rho = 0$) defined on the inside of a sphere of radius R, satisfying Neumann boundary conditions on the surface of the sphere as follows (see Figure 7.2): ⇒ lecture

$$\Delta\Psi(r,\theta,\phi) = 0 \quad \text{for } r < R \qquad \text{with} \qquad \frac{\partial\Psi}{\partial n}\Big|_{r=R} = f(\theta,\phi) = \theta - \frac{\pi}{2} \tag{7.21}$$

First we need to check if this particular $f(\theta,\phi)$ satisfies the solvability condition (6.29). Luckily, this is the case, since

$$\int_\sigma f(\vec{r})\, d^2\sigma \to \int d\Omega_2 \Big(\theta - \frac{\pi}{2}\Big) = \int_0^{2\pi} d\phi \int_0^\pi \sin\theta\, d\theta \left(\theta - \frac{\pi}{2}\right) = 0 \tag{7.22}$$

Since we need a solution that is well defined at $r = 0$, we take the general solution (7.20) with $b_{lm} = 0$ and determine the coefficients a_{lm} such that the boundary conditions are satisfied. We are lucky in that there is no ϕ dependence, hence we only need $m = 0$ and taking into account (5.19) by setting $a_l = a_{l0}\sqrt{(2l+1)/(4\pi)}$ we get ⇒ lecture

$$\frac{\partial\Psi}{\partial n}\Big|_{r=R} = \sum_{l=0}^{\infty} a_l\, l\, R^{l-1} P_l(\cos\theta) \stackrel{!}{=} f(\theta) = \theta - \frac{\pi}{2} \tag{7.23}$$

In (7.9) we used the Fourier series to match our general solution to the boundary condition. Now we use the 'Legendre' series, (5.8). Thus we express

$$f(\theta) = \theta - \frac{\pi}{2} = \sum_{l=0}^{\infty} c_l\, P_l(\cos\theta) \qquad \text{with} \qquad c_l = \frac{2l+1}{2}\int_0^\pi d\theta\, \sin\theta f(\theta)\, P_l(\cos\theta) \tag{7.24}$$

We find $c_0 = 0$ as required by the solvability condition. But for $l > 0$ we get an infinite series of non-vanishing terms

$$c_1 = -\frac{3\pi}{8}, \qquad c_2 = 0, \qquad c_3 = -\frac{7\pi}{128}, \qquad c_4 = 0, \qquad c_5 = -\frac{11\pi}{512}, \qquad \ldots \tag{7.25}$$

Once we have all c_l we match them to a_l and write our solution as

$$\Psi(r,\theta) = \sum_{l>0} \frac{c_l}{l} \frac{r^l}{R^{l-1}} P_l(\cos\theta) + K = -\frac{3\pi}{8} r\cos\theta - \frac{7\pi}{128} \frac{r^3}{R^2} \frac{(5\cos^3\theta - 3\cos\theta)}{3\cdot 2} + \dots \tag{7.26}$$

where K is a constant, as expected from the uniqueness theorem with Neumann boundary conditions.

In Section 5.1 we boldly claimed that the P_l (or $Y_l^{(m)}$ in case of a ϕ dependent problem) form a basis, without actually explaining what this means. Keeping waiving with our arms, we can approximate any boundary condition (as long as it satisfies the solvability condition) on the surface of the sphere through a unique linear combination of P_l or $Y_l^{(m)}$. The uniqueness is in agreement with the notion of a basis as well as our discussion at the end of Section 6.2. In the right panel of Figure 7.2 we see that the series in (7.26) is indeed "approaching" the function $f(\theta)$. Once more we have to be patient and wait for MMP 2 to get a more solid understanding, what "approaching" actually means.

7.3 Laplace equation in cylindrical coordinates

Last but not least, we consider the Laplace equation in cylindrical coordinates (r,ϕ,z) related to cartesian coordinates as

$$\begin{aligned} x &= r\,\cos\phi \\ y &= r\,\sin\phi \\ z &= z \end{aligned} \tag{7.27}$$

where $0 \le r < \infty$ and $0 \le \phi < 2\pi$. In analogy to (7.13) we express the Laplace operator in these coordinates and get

$$\Delta = \Big(\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial\phi^2} + \frac{\partial^2}{\partial z^2}\Big) \tag{7.28}$$

where the first two terms correspond to the Laplace operator in two-dimensional polar coordinates.

Now we simply turn the crank. We set $\Psi(r,\phi,z) = R(r)Q(\phi)Z(z)$, and rewrite the PDE $\Delta\Psi(r,\phi,z) = 0$ in terms of ODE as

$$\frac{R''(r)}{R(r)} + \frac{1}{r}\frac{R'(r)}{R(r)} + \underbrace{\frac{1}{r^2}\frac{Q''(\phi)}{Q(\phi)}}_{-\nu^2} + \underbrace{\frac{Z''(z)}{Z(z)}}_{k^2} = 0 \tag{7.29}$$

where we introduced constants $-\nu^2$ and k^2. The ODE for $Q(\phi)$ and $Z(z)$ can readily be solved as

$$Q_\nu(\phi) = s_\nu\,\sin(\nu\phi) + c_\nu\,\cos(\nu\phi) \tag{7.30}$$

$$Z_k(z) = p_k\,e^{kz} + m_k\,e^{-kz} \tag{7.31}$$

where s_ν, c_ν, p_k, and m_k are constants. For the solution to be unique under $\phi \to \phi + 2\pi$ we need again $\nu \in \mathbb{Z}$. This leaves us with the ODE for $R(r)$ which reads

$$r^2\,R''(r) + r\,R'(r) + \big(r^2\,k^2 - \nu^2\big)R(r) = 0 \tag{7.32}$$

A shudder of excitement runs down our spine as we realise this is the Bessel ODE, (5.31). Thus, the solutions are the Bessel functions $J_{\pm\nu}(kr)$. However, since $\nu \in \mathbb{Z}$, these two solutions are not linearly independent, (5.37). Never mind, we are prepared and use the Bessel function of the second kind, (5.38) to write the general solution as ⇒ lecture

$$\Psi(r,\phi,z) = \sum\!\!\!\!\!\!\int_k \sum_{\nu\in\mathbb{N}} \big(j_{\nu k}\, J_\nu(kr) + y_{\nu k}\, Y_\nu(kr)\big)\big(s_\nu\, \sin(\nu\phi) + c_\nu\, \cos(\nu\phi)\big)\big(p_k\, e^{kz} + m_k\, e^{-kz}\big) \tag{7.33}$$

To state the obvious, this can be generalised to (7.3a) with $k^2(r,\phi,z) = f(r) + g(z) + h(\phi)$.

Again, we look at a concrete example to illustrate how we fix the constants in the general solution (7.33) through the boundary conditions. Since we used Neumann in the case of spherical coordinates, it is nothing but fair to use Dirichlet in the case of cylindrical coordinates. Thus, we want to solve the Laplace equation in the inside of a cylinder of radius R and height h with the boundary condition that $\Psi(r,\phi,z) = 0$ on the surface of the cylinder, except for the 'lid', where we want $\Psi(r,\phi,h) = f(r,\phi)$ for a given function f, see Figure 7.3. We will not bother with giving an explicit example of $f(r,\phi)$, as by now it should be obvious how to accommodate different choices.

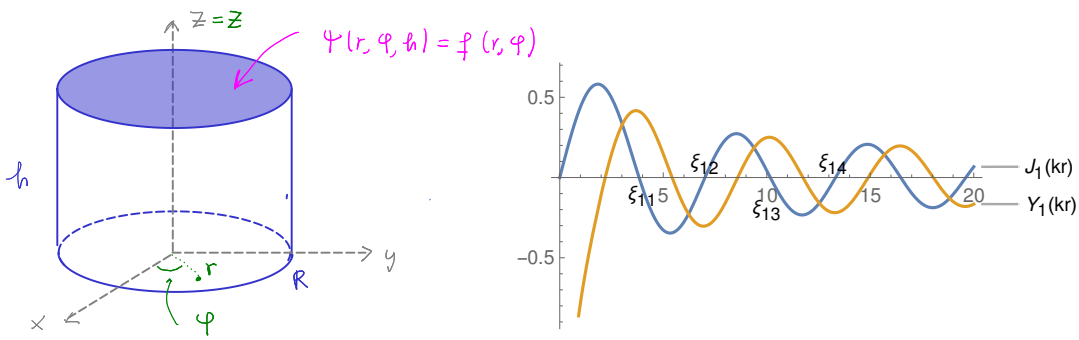

Figure 7.3: Dirichlet boundary value problem for Laplace equation in cylindrical coordinates. The volume V is a cylinder of radius R and height h. The right panel shows $Y_1(kr)$ (orange) and $J_1(kr)$ (blue) with its zeros ξ_{1n} with $J_1(\xi_{1n}) = 0$.

Let us now take into account the various boundary conditions as done in (7.8a) ⇒ lecture

$$\Psi(r,\phi,0) = 0 \qquad \Rightarrow \qquad p_k = -m_k \tag{7.34a}$$

$$\Psi(r<R) \quad \text{finite} \qquad \Rightarrow \qquad y_{\nu k} = 0,\ \forall\,\nu,\ \forall\,k \tag{7.34b}$$

$$\Psi(R,\phi,z) = 0 \qquad \Rightarrow \qquad J_\nu(kR) = 0 \tag{7.34c}$$

The boundary condition (7.34b) is due to the fact that $Y_\nu(r)$ diverges for $r \to 0$ for $\nu \geq 0$, as illustrated for $\nu = 1$ in the right panel of Figure 7.3. For each choice of $\nu \in \mathbb{N}$, (7.34c) picks discrete values from the continuum of possible k, namely the zeros of $J_\nu(kR)$. More precisely, $k = \xi_{\nu n}/R$ such that $J_\nu(\xi_{\nu n}) = 0$. Thus the integration over dk in (7.33) becomes a sum over discrete values of k, as has happened in connection with (7.8c).

Assembling what we have so far we get

$$\Psi(r,\phi,z) = \sum_{\nu=0}^{\infty}\sum_{n=1}^{\infty} J_\nu\big(\xi_{\nu n}\frac{r}{R}\big)\, \sinh\big(\xi_{\nu n}\frac{z}{R}\big)\Big(c_{\nu n}\cos(\nu\phi) + s_{\nu n}\sin(\nu\phi)\Big) \tag{7.35}$$

which satisfies (7.34). To fix the remaining constants $c_{\nu n}$ and $s_{\nu n}$ we use the boundary condition of the 'lid' of the cylinder $\Psi(r,\phi,h) = f(r,\phi)$. This is done in a way similar to the Fourier or Legendre series by computing

$$c_{\nu n} = \frac{1}{\sinh(\xi_{\nu n}\frac{h}{R})} \int_0^{2\pi} d\phi \int_0^{\infty} dr\, r\, f(r,\phi)\Big(J_\nu\big(\xi_{\nu n}\frac{r}{R}\big)\cos(\nu\phi)\Big)^* \tag{7.36}$$

and analogous for $s_{\nu n}$.

7.4 Poisson equation and Green functions

So far, all PDE we have considered were homogeneous. We know from ODE that an inhomogeneous problem can be considerably more involved, and now we want to address this question first by introducing the concept of a *Green function [Greensche Funktion]* and then apply it to the Poisson equation. In fact, we already had a brief encounter with this topic at the end of Section 2.3.

We write a generic partial differential equation as

$$\mathcal{D}_{\vec{r}}\,\Psi(\vec{r}) = \rho(\vec{r}) \tag{7.37}$$

where $\mathcal{D}_{\vec{r}}$ is any (combination of) differential operator(s) acting on the function Ψ. If $\vec{r}$ is one dimensional, (7.37) is an ODE, if $\vec{r}$ has more than one dimension it is a PDE and $\mathcal{D}_{\vec{r}}$ contains partial derivatives w.r.t. the coordinates (i.e. independent variables) collected in $\vec{r}$. The problem with (7.37) is the r.h.s., the inhomogeneity $\rho(\vec{r})$. In our list of PDE, (6.2), only the Poisson equation (6.2b) has such a term, but obviously there are many more cases where we have to deal with such terms. Not only does $\rho(\vec{r})$ make (7.37) more complicated it is also annoying in that for each new choice of $\rho(\vec{r})$ we have to find a new solution. Very much like for each inhomogeneity in an ODE we have to find a new particular solution.

Wouldn't it be nice to solve (7.37) once and for all and for any $\rho(\vec{r})$? Prepare the stage for the Green function. Let us assume we knew a function $G_{\rm b}(\vec{r}-\vec{r}_0)$ that satisfies

$$\mathcal{D}_{\vec{r}}\,G_{\rm b}(\vec{r}-\vec{r}_0) = \delta(\vec{r}-\vec{r}_0) \tag{7.38}$$

subject to some particular boundary conditions, as indicated by the subscript 'b'. Such a function is like a master key to finding a solution Ψ (with boundary conditions 'b') to (7.37) for any $\rho(\vec{r})$. Indeed, ⇒ lecture

$$\Psi(\vec{r}) = \int d\vec{r}_0\; G_{\rm b}(\vec{r}-\vec{r}_0)\rho(\vec{r}_0) \tag{7.39}$$

reduces the problem of (7.37) to 'simply' doing an integral. In some sense, the Green function is the inverse of the operator $\mathcal{D}_{\vec{r}}$, subject to particular boundary conditions and, yes, (7.39) is a convolution as defined in (3.25).

Needless to say that finding the Green function is not always easy. We will focus on the Poisson equation and, to start with, on Dirichlet boundary conditions $\lim_{r\to\infty}\Psi(\vec{r}) = 0$. The reason is that this problem is of utmost importance in electromagnetism (EM), writing

$$\Delta\,\Psi(\vec{r}) = \rho(\vec{r}) \qquad \text{in EM} \rightsquigarrow \qquad \Delta\,\phi(\vec{r}) = -4\pi\,\rho(\vec{r}) \tag{7.40}$$

where $\rho(\vec{r})$ is a given charge distribution and we are looking for the potential ϕ that is related to the electric field as $\vec{E} = -\vec{\nabla}\phi$. If the charge distribution is restricted to a finite space, as depicted in Figure 7.4, i.e. $\exists\, R$ such that $\rho(\vec{r}) = 0$ for $r = |\vec{r}| > R$ then we are looking for ϕ as in (7.40) with the boundary condition $\lim_{r\to\infty} \phi(\vec{r}) = 0$.

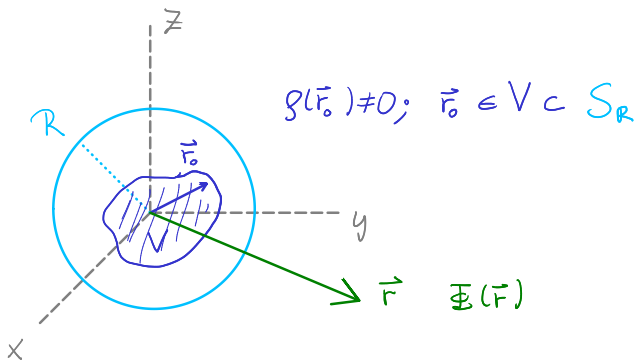

Figure 7.4: Determination of the potential $\phi(\vec{r})$ due to a charge distribution $\rho(\vec{r}_0)$.

But really, this is precisely (2.26). In Section 2.3 we have introduced (2.25) as an equation between distributions. In the current context, it is nothing but a version of (7.38) with $\mathcal{D}_{\vec{r}} = \Delta$. The proof of (2.25) given in Section 2.3 can now be made more direct since we have the Fourier transform available as a tool. Taking the Fourier transform of the equation

$$\Delta\, G_\infty(\vec{r} - \vec{r}_0) = \delta(\vec{r} - \vec{r}_0) \tag{7.41}$$

we obtain ⇒ lecture

$$G_\infty(\vec{r} - \vec{r}_0) = -\frac{1}{4\pi}\frac{1}{|\vec{r} - \vec{r}_0|} \tag{7.42}$$

as the Green function of the Laplace operator with boundary condition $\lim_{r\to\infty} G \to 0$. This can be interpreted as the potential due to a point charge positioned at $\vec{r} = \vec{r}_0$. A continuous charge distribution is made up of a sum (actually an integral) over point charges. Due to the linearity of the Laplace operator we can use the superposition principle introduced in Section 4.2. The potential induced due to ρ is simply the sum (actually the integral) over the individual potentials of the point charges. This is the content of (2.26).

What if we would like to have the Green function with different boundary conditions? Let's assume we want to solve again the Poisson equation $\Delta\Psi(\vec{r}) = \rho(\vec{r})$ for $\vec{r} \in V$, where V is some region, let's say a 'potato'. But this time we want Dirichlet boundary conditions $\Psi(\vec{r}) = f(\vec{r})$ on the surface $\sigma = \partial V$ of the 'potato'. The general form of this Green function is

$$G_V(\vec{r} - \vec{r}_0) = -\frac{1}{4\pi}\frac{1}{|\vec{r} - \vec{r}_0|} + F(\vec{r}, \vec{r}_0) \tag{7.43}$$

where the function F satisfies $\Delta\, F = 0$. The first term in (7.43) is sometimes referred to as the singular solution, whereas the second term is called the regular solution. It is clear that

G_V also satisfies (7.37) with $\mathcal{D}_{\vec{r}} = \Delta$. The task is to find F such that the new boundary condition is satisfied.

As a first example we consider V as the region with $z \geq 0$. The method we are going to use to find G_V is known as image (or mirror) charges in electrostatics. If we want to cancel out the effect of the point charge at $\vec{r}_0 = (x, y, z)$ on the plane $z = 0$, we just put a mirror charge (same magnitude, different sign) at $\vec{r}_1 = (x, y, -z)$, as illustrated in Figure 7.5, and obtain

$$G_V(\vec{r} - \vec{r}_0) = -\frac{1}{4\pi}\frac{1}{|\vec{r} - \vec{r}_0|} + \frac{1}{4\pi}\frac{1}{|\vec{r} - \vec{r}_1|} \tag{7.44}$$

For $\vec{r} \in V$ the second part $F \sim 1/|\vec{r} - \vec{r}_1|$ has no singularities, hence the name regular solution.

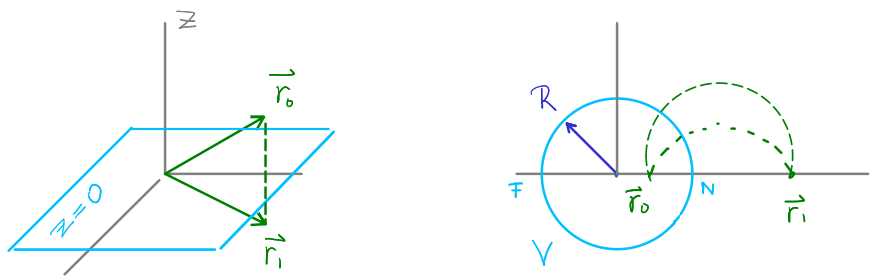

Figure 7.5: Examples of using mirror charges to determine Green functions with special boundary conditions. Left panel: V is the upper half-space $z \geq 0$ and $\partial V = \sigma$ is the x-y plane. Right panel: V is the inside of a sphere of radius R and σ is its surface.

As a second example we consider a perfectly spherical potato, i.e. we want to find G_{S_R}, the Green function with Dirichlet boundary conditions $\Psi(\vec{r}) = 0$ on the surface of a sphere of radius R. Again we put a mirror charge, this time of a different magnitude k, at a point $\vec{r}_1$, amounting to a particular choice of F, namely

$$G_{S_R}(\vec{r} - \vec{r}\,') = -\frac{1}{4\pi}\frac{1}{|\vec{r} - \vec{r}_0|} + \frac{1}{4\pi}\frac{k}{|\vec{r} - \vec{r}_1|} \tag{7.45}$$

As long as $\vec{r}_1$ is outside the sphere, the second term in (7.45) satisfies the Laplace equation $\forall\, \vec{r} \in S_R$ and it is regular. We have to find a $\vec{r}_1$ (the position of the image) outside the sphere and a constant k (the charge of the image) such that the boundary condition is satisfied. The simplest way to do this is to consider the points N and F from which we get $k/(R - r_1) = 1/(R - r_0)$ and $k/(R + r_1) = 1/(R + r_0)$, hence ⇒ lecture

$$G_{S_R}(\vec{r} - \vec{r}\,') = -\frac{1}{4\pi}\frac{1}{|\vec{r} - \vec{r}_0|} + \frac{1}{4\pi}\frac{R/r_0}{|\vec{r} - (R/r_0)^2\,\vec{r}_0|} \tag{7.46}$$

As a final topic in this section we return to the problem depicted in Figure 7.4 and consider the expansion of the Green function (7.42) in terms of spherical harmonics. In a first step we expand in $t = r_0/r < 1$ and use (5.10) to write

$$\frac{1}{|\vec{r} - \vec{r}_0|} = \frac{r^{-1}}{\sqrt{1 - 2xt + t^2}} = \sum_{l=0}^{\infty} \frac{r_0^l}{r^{l+1}}\, P_l(x) \tag{7.47}$$

where $x = \cos\gamma$ and $\gamma\angle(\vec{r}, \vec{r}_0)$ the angle between $\vec{r} = (r, \theta, \phi)$ and $\vec{r}_0 = (r_0, \theta_0, \phi_0)$ as depicted in Figure 7.6.

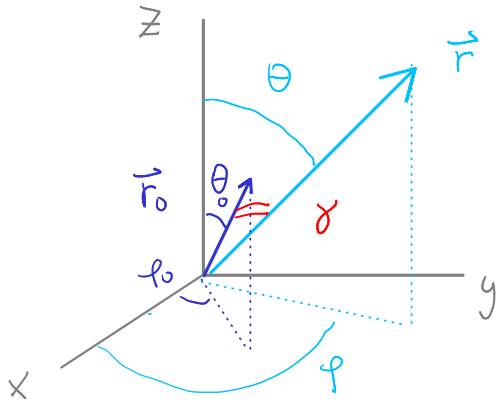

Figure 7.6: The vectors $\vec{r} = (r, \theta, \phi)$ and $\vec{r}_0 = (r_0, \theta_0, \phi_0)$ and the angle γ between them.

For the second step we need the *addition theorem [Additionstheorem]* of spherical harmonics

$$P_l(\cos\gamma) = \frac{4\pi}{2l+1} \sum_{m=-l}^{l} Y_l^{(m)}(\theta, \phi) \left(Y_l^{(m)}(\theta_0, \phi_0) \right)^* \tag{7.48a}$$

$$\cos\gamma = \cos\theta \, \cos\theta_0 + \sin\theta \, \sin\theta_0 \, \cos(\phi - \phi_0) \tag{7.48b}$$

The proof relies on the fact that the spherical harmonics form a basis on the surface of the unit sphere. ⇒ lecture

Combining (7.47) with (7.48a) we can express the Green function (7.42) as

$$G_\infty(\vec{r} - \vec{r}_0) = -\frac{1}{4\pi} \frac{1}{|\vec{r} - \vec{r}_0|} \overset{r_0 < r}{=} \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{r_0^l}{r^{l+1}} \frac{-1}{2l+1} Y_l^{(m)}(\theta, \phi) \left(Y_l^{(m)}(\theta_0, \phi_0) \right)^* \tag{7.49}$$

While it seems we have not exactly simplified the expression, the structure of G_∞ as an expansion in spherical harmonics offers the option to do a *multipole expansion [Multipolentwicklung]* in r_0/r of the potential ϕ, (2.26), i.e. for distances large compared to the extension of the charge distribution. Indeed, we can write ⇒ lecture

$$\phi(\vec{r}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{4\pi}{2l+1} \frac{b_{lm}}{r^{l+1}} Y_l^{(m)}(\theta, \phi) \tag{7.50a}$$

$$b_{lm} = \int d\vec{r}_0 \, r_0^l \, \rho(\vec{r}_0) \left(Y_l^{(m)}(\theta_0, \phi_0) \right)^* \tag{7.50b}$$

It is instructive to compare (7.50a) to the general structure of the solution of the Laplace equation in spherical coordinates, (7.20). Yes, it is exactly the solution that is well behaved for $r \to \infty$.

The leading term of (7.50a) is obtained for $l = 0$ and hence $m = 0$. We get $b_{00} = Q/\sqrt{4\pi}$ ⇒ lecture
where $Q = \int d\vec{r}_0\, \rho$ is the total charge. The next term is obtained for $l = 1$ and $m \in \{-1, 0, 1\}$.
Evaluating the coefficients b_{1m} we can write the potential as ⇒ lecture

$$\phi(\vec{r}) = \frac{Q}{r} + \frac{\vec{d}\cdot\vec{r}}{r^3} + \mathcal{O}\Big(\frac{1}{r^3}\Big) \tag{7.51}$$

where we have defined the dipole moment of the charge distribution ρ as

$$\vec{d} = \int d\vec{r}_0\, \vec{r}_0\, \rho(\vec{r}_0) \tag{7.52}$$

The first term in (7.51) approximates the charge distribution as a point charge and behaves as $1/r$. If we go a bit closer, the dipole structure will become visible, behaving as $1/r^2$. Going even closer, the higher moments (quadrupole $\sim 1/r^3$ etc.) that have been omitted in (7.51) will also become relevant. Once more, our result is composed of infinitely many terms. But on many occasions, taking only the first few gives an excellent approximation.

7.5 Heat or diffusion equation

The heat or diffusion equation (6.2c) describes the flow of heat or chemical substances in a medium. It is a parabolic PDE. We will consider its solutions with Dirichlet boundary conditions, i.e.

$$\frac{\partial}{\partial t}\Psi(\vec{r},t) - \kappa\,\Delta\Psi(\vec{r},t) = 0 \qquad \text{with} \qquad \Psi(\vec{r},t=0) = f(\vec{r}) \tag{7.53}$$

An elegant way to solve (7.53) is to use the Fourier transform

$$\hat{\Psi}(\vec{k},t) = \frac{1}{(2\pi)^{n/2}} \int d^n\vec{r}\,\Psi(\vec{r},t)\, e^{-i\,\vec{k}\cdot\vec{r}} \tag{7.54a}$$

$$\hat{f}(\vec{k}) = \frac{1}{(2\pi)^{n/2}} \int d^n\vec{r}\, f(\vec{r})\, e^{-i\,\vec{k}\cdot\vec{r}} \tag{7.54b}$$

Doing this without making an assumption on the dimension n of $\vec{r}$ we find

$$\partial_t\hat{\Psi}(\vec{k},t) = -\kappa\, k^2\, \hat{\Psi}(\vec{k},t) \qquad \text{with} \qquad \hat{\Psi}(\vec{k},0) = \hat{f}(\vec{k}) \tag{7.55}$$

This is an ODE w.r.t. t and we can replace $\partial_t \to d/dt$. The partial derivatives w.r.t. the coordinates in $\vec{r}$ have been replaced by a multiplication with a parameter k^2, a consequence of (3.24c). Even better, solving the ODE is a piece of cake and we get

$$\hat{\Psi}(\vec{k},t) = e^{-\kappa\, k^2\, t}\, \hat{f}(\vec{k}) \tag{7.56}$$

All we need to do is to take the inverse Fourier transform to obtain ⇒ lecture

$$\Psi(\vec{r},t) = \int d^n\vec{r}\,'\, f(\vec{r}\,')K_t(\vec{r}-\vec{r}\,') \quad \text{with} \quad K_t(\vec{r}-\vec{r}\,') = \frac{1}{(2\pi)^n}\int d^n\vec{k}\, e^{-\kappa\, k^2\, t}\, e^{-i\vec{k}\cdot(\vec{r}-\vec{r}\,')} \tag{7.57}$$

Again, this is a convolution, (3.25), and ⇒ lecture

$$K_t = \left(\frac{1}{4\pi\,\kappa\,t}\right)^{n/2} e^{-\frac{(\vec{r}-\vec{r}\,')^2}{4\kappa\,t}} \tag{7.58}$$

computed by completing the square in the exponent is called the kernel of the heat equation. Note that ⇒ lecture

$$\int d^n\vec{r}\;\Psi(\vec{r},t) = \int d^n\vec{r}\;f(\vec{r}) \tag{7.59}$$

corresponds to the total amount of substance (or heat) and is time independent. Thus, (7.53) describes how the substance is distributed over space as a function of time, without creating or destroying any of the substance. This is illustrated in Figure 7.7 where we take an initial condition $f(\vec{r}) = \delta(\vec{r})$, i.e. a point-like source at $t = 0$. Then $\Psi(\vec{r},t) = K_t(\vec{r})$ and the latter is plotted for increasing t in one dimension with $\kappa = 1$.

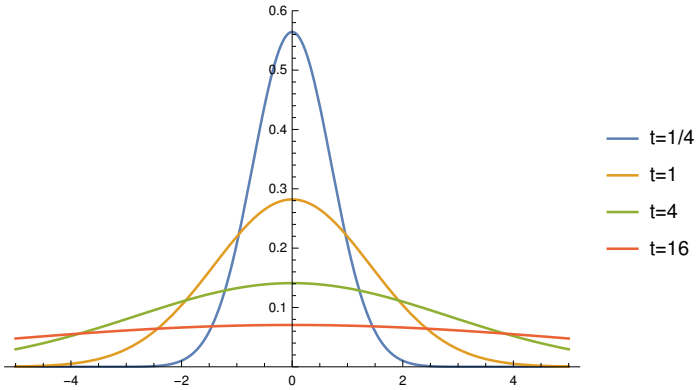

Figure 7.7: Flow of heat as a function of time, starting with a δ peak at $t = 0$

Of course, the heat equation can also be solved by using separation of variables. Restricting ourselves to the one-dimensional case and writing $\Psi(x,t) = T(t)\,X(x)$ we obtain

$$T'(t) = -\lambda^2\,T(t) \qquad \Rightarrow \qquad T_\lambda(t) = e^{-\lambda^2\,t} \tag{7.60a}$$

$$X''(x) = -\frac{\lambda^2}{\kappa}\,X(x) \qquad \Rightarrow \qquad X_\lambda(x) = c_\lambda\,\cos\left(\frac{\lambda\,x}{\sqrt{\kappa}}\right) + s_\lambda\,\sin\left(\frac{\lambda\,x}{\sqrt{\kappa}}\right) \tag{7.60b}$$

This form is particularly useful, if a solution in a confined region is needed. To illustrate this consider a rod of length (for convenience) 2π. At $t = 0$, let the substance be distributed according to a particular function $\Psi(x,0) = f(x)$ for $-\pi \le x \le \pi$. The function $f(x)$ can be expressed as a Fourier series with Fourier coefficients a_n and b_n, (3.6). The label $\lambda = n\sqrt{\kappa}$ then is discrete with $n \in \mathbb{N}$ and the time-dependent solution is given as

$$\Psi_f(x,t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} e^{-n^2\kappa\,t}\Big(a_n\cos(nx) + b_n\sin(nx)\Big) \tag{7.61}$$

How is this solution related to the one in (7.57)? The difference is that $\Psi_f(x,t)$ is defined only within $-\pi \le x \le \pi$ and its analogue to (7.59) is

$$\int_{-\pi}^{\pi} dx\;\Psi_f(x,t) = \int_{-\pi}^{\pi} dx\;f(x) = \pi\,a_0 \tag{7.62}$$

Also here, the total amount of substance is conserved. However, it will remain within the interval $-\pi \leq x \leq \pi$ at all times, whereas for Ψ as given in (7.57) the substance will wander off and spread on the full x axis. Consider an initial condition $f(x) = \Theta(d - |x|)/(2d)$ with say $d = 1/4$. In Figure 7.7 we compare the two solutions (the calculation of the Fourier coefficients is a nice little exercise). The left panel shows $\Psi(x,t)$ as in (7.57) for increasing values of t (with $\kappa = 0.2$). It is clear that the total amount of substance within $-\pi \leq x \leq \pi$ approaches 0 for $t \to \infty$, i.e. $\Psi(x,\infty) \to 0$. On the contrary, the corresponding solution $\Psi_f(x,t)$ as given in (7.61) conserves all of the substance within the interval $-\pi \leq x \leq \pi$. For small t, both solutions agree very well, but for large t they start to differ. In particular $\Psi_f(x,\infty) \to a_0/2$. ⇒ lecture

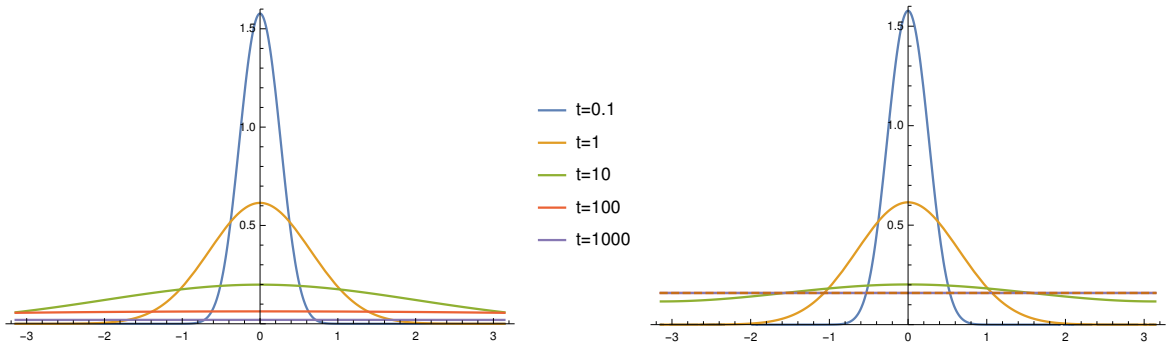

Figure 7.8: Comparison of the solutions $\Psi(x,t)$, (7.57) and $\Psi_f(x,t)$, (7.61) for an initial condition $f(x) = 2\,\Theta(1/4 - |x|)$. The solutions agree very well until they reach the boundary $|x| = \pi$.

The moral of the story is that solutions to PDE have to be adapted to the situation. This not only concerns setting the boundary conditions, but in fact even the path to finding a solution.

7.6 Schrödinger equation

The Schrödinger equation (6.2g) describes the quantum mechanical treatment of a point mass m subject to the potential $V(\vec{r})$. The dependent variable of this PDE, $\Psi(\vec{r},t)$ is the wave function and is supposed to be square integrable. In fact, typically the wave fuction is normalised as

$$\int d\vec{r}\,|\Psi(\vec{r},t)|^2 = 1 \tag{7.63}$$

such that $|\Psi(\vec{r},t)|^2 d\vec{r}$ has a direct interpretation of a probablility to find the point mass at time t in a particular (infinitesimal) volume element $d\vec{r}$.

Using separation of variables as $\Psi(\vec{r},t) = T(t)\psi(\vec{r})$ we obtain from (6.2g)

$$i\hbar\,\frac{T'(t)}{T(t)} = E \qquad \Rightarrow \qquad T(t) = e^{-i\frac{E}{\hbar}t} \tag{7.64}$$

For us E is just a constant (as e.g. $-\lambda^2$ in (7.60a)) but the letter has been chosen with its physical interpretation (energy) in mind.

Using (7.64) the time-dependent Schrödinger equation (6.2g) for $\Psi(\vec{r},t)$ is transformed into the time-independent Schrödinger equation (6.2f) for $\psi(\vec{r})$.

Ignoring constants[2] in (6.2f) we are led to consider

$$\big(-\Delta + V(\vec{r})\big)\,\psi(\vec{r}) = E\,\psi(\vec{r}) \tag{7.65}$$

We can call this a homogeneous PDE, or we can think of it along the lines presented in Section 5.7, interpreting $\psi(\vec{r})$ as eigenvector with eigenvalue E to the operator $(-\Delta + V(\vec{r}))$. Whether or not such an eigenvector exists for some V depends on the eigenvalue E. Very often, only for particular discrete values of E this is the case, as we have witnessed several times. Thus, the possible values of E (the energy) are quantised.

Let us consider some examples, starting with the one-dimensional harmonic oscillator where $V(x) \sim x^2$. The time dependent Schrödinger equation

$$i\partial_t \Psi(x,t) = (-\partial_{xx} + x^2)\Psi(x,t) \tag{7.66}$$

leads to the time-independent Schrödinger equation

$$\psi''(x) = (x^2 - E)\psi(x) \tag{7.67}$$

which is simply a second order ODE. We get two linearly independent solutions. In the region $x \to \infty$ they behave a $\psi_-(x) \sim e^{-x^2/2}$ and $\psi_+(x) \sim e^{+x^2/2}$. For the second solution the integral (7.63) does not exist. Hence we drop it from our physics driven considerations. Motivated by the behaviour of $\psi_-(x)$ we define $H(x)$ through $\psi(x) = H(x)\,e^{-x^2/2}$ and have a flashback to Section 5.4, (5.28). There we have seen that for $E = (2n+1)$ with $n \in \mathbb{N}$ we get terminating and, hence, normalisable solutions. We remark that the non-teminating solutions of (5.28) are not square integrable. This can be seen by recalling the recursion relation (5.29) which, if not terminating, for large k behaves like $c_{k+2} = c_k/k$. Thus for $r \in \mathbb{R}$ but $r \notin \mathbb{N}$ the solution of (5.28) (with $2m = k$) behaves as

$$H_r(x) = \sum_{m=0}^{\infty} c_m x^{2m+\rho} \quad \text{with} \quad c_m \sim \frac{1}{m!} \quad \text{hence for } x \to \infty \quad H_r(x) \sim e^{x^2} \tag{7.68}$$

Therefore, $\psi_r(x) = H_r(x)\,e^{-x^2/2} \sim e^{x^2/2}$ is badly divergent for $x \to \infty$.

To summarise, we have found (countable) infinitely many eigenvalues $E_n = (2n+1)$ with $n \in \mathbb{N}$ and the corresponding eigenvectors $\psi_n(x) = H_n(x)\,e^{-x^2/2}$ with H_n the Hermite polynomials. There are no other square integrable eigenvectors.

Returning to the time-dependent problem (7.66), we can ask a similar question as for the heat equation. Assume at $t = 0$ the wave function $\Psi(x,0) = f(x)$ is given. What is the wave function at a later time t? To solve this, we expand $f(x)$ in terms of Hermite polynomials (as we did in the case of the Fourier series, or Legendre series) and obtain

$$\Psi(x,0) = f(x) = \sum_n a_n H_n(x)\,e^{-x^2/2} \quad \Rightarrow \quad \Psi(x,t) = \sum_n a_n H_n(x)\,e^{-x^2/2}\,e^{-i\,E_n\,t} \tag{7.69}$$

Note that the wave function is intrinsically complex.

[2] To save ink, we set $\hbar = 2m = 1$ in what follows. This means $\hbar\omega = 2$ in case you want to compare with QM books.

We have just solved a problem in quantum mechanics by first finding all eigenvalues and eigenvectors of the time-independent Schrödinger equation and then determined the time evolution, given initial conditions. From the MMP point of view, it is simply solving a PDE with initial conditions. The fact that we can expand $f(x)$ in terms of the solutions of (7.67) is to do with the operator in (7.67) being self-adjoint (recall Section 5.7). In fact, you will learn that physical quantities in quantum mechanics are represented by self-adjoint operators. Recall this, when you solve the harmonic oscillator (and many other problems) in quantum mechanics!

The three-dimensional (or however many dimensions you fancy) harmonic oscillator is solved by separating variables further and write $\psi(\vec{r}) = X(x)\,Y(y)\,Z(z)$ and we obtain eigenvalues $E_{n_x n_y n_z} = 2(n_x + n_y + n_z) + 3$ and eigenvectors

$$\psi_{n_x n_y n_z}(\vec{r}) = H_{n_x}(x) H_{n_y}(y) H_{n_z}(z)\, e^{-r^2/2} \tag{7.70}$$

Let us now return to (7.65) and assume the potential is spherically symmetric, i.e. depends only in $r = |\vec{r}|$. This is a common case in physics and cries out loud for using spherical coordinates. Recalling (7.13) the time-independent Schrödinger equation reads

$$\Big(-\frac{1}{r}\frac{\partial^2}{\partial r^2}\, r - \frac{1}{r^2}\Big(\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\Big) + V(r) - E\Big)\psi(r,\theta,\phi) = 0 \tag{7.71}$$

In order to get some insights into the structure of this PDE and (7.65) let us embark on a little intermezzo. We mentioned above that in quantum mechanics physical observables are represented by self-adoint operators. The notion of self-adoint is still rather murky and will not play an important role for the following observations.

Starting with the momentum, we have the correspondence

$$\vec{p} = -i\vec{\nabla} = -i\vec{e}_x\,\partial_x - i\vec{e}_y\,\partial_y - i\vec{e}_z\,\partial_z \tag{7.72}$$

(with $\hbar = 1$), where $\vec{e}_i$ are the unit vectors of a cartesian coordinate system. In fact, (7.65) is nothing but the statement $E = p^2/(2m) + V$, i.e. the total energy is equal to the kinetic energy plus the potential energy. The position operator (e.g. x in $V(x) \sim x^2$ or its three-dimensional equivalent $\vec{r} = (x, y, z)$) is just the multiplication operator. Let us now consider the angular momentum operator $\vec{L} = \vec{r} \times \vec{p}$. In cartesian coordinates it is written as

$$\vec{L} = \begin{pmatrix} L_x \\ L_y \\ L_z \end{pmatrix} = \begin{pmatrix} y\,p_z - z\,p_y \\ z\,p_x - x\,p_z \\ x\,p_y - y\,p_z \end{pmatrix} = -i \begin{pmatrix} y\,\partial_z - z\,\partial_y \\ z\,\partial_x - x\,\partial_z \\ x\,\partial_y - y\,\partial_x \end{pmatrix} \tag{7.73}$$

where we have written the coefficients of $\vec{e}_i$ as a column. Very much like $\vec{p}$ also $\vec{L}$ is a differential operator acting on vectors (i.e. the wave function). We can express $\vec{L}$ also in spherical coordinates. This is discussed in some detail in Section A, see (A.15) and (A.18). The upshot is that (7.71) can be rewritten as

$$\Big(-\frac{1}{r}\frac{\partial^2}{\partial r^2}\, r - \frac{1}{r^2}\, L^2 + V(r) - E\Big)\psi(r,\theta,\phi) = 0 \tag{7.74}$$

The only dependence on θ and ϕ in (7.74) is in the spherical harmonics $Y_l^{(m)}(\theta,\phi)$ embedded in $\psi(r,\theta,\phi)$, see (7.50a). If we take a single term $\psi_{lm} = Y_l^{(m)}(\theta,\phi)\,U_l(r)/r$ in an expansion similar to (7.20) we have ⇒ lecture

$$L_z\, Y_l^{(m)}(\theta,\phi) = m\, Y_l^{(m)}(\theta,\phi) \qquad \text{and} \qquad L^2\, Y_l^{(m)}(\theta,\phi) = l(l+1)\, Y_l^{(m)}(\theta,\phi) \tag{7.75}$$

Thus $Y_l^{(m)}$ is an eigenvector to L_z and L^2 with eigenvalue m and $l(l+1)$, respectively. The second relation in (7.75) is nothing but the generalized Legendre equation. Indeed, expressing L^2 of (A.18) in terms of $x = \cos\theta$ as done in (A.19) and using $\partial^2/\partial\phi^2 = m^2$ when acting on $Y_l^{(m)}$, (5.19), we can recognize (5.11) and (7.18). The Schrödinger equation (7.74) now becomes an ODE for $U_l(r)$, as

$$U_l''(r) + \Big(E - V(r) - \frac{l(l+1)}{r^2}\Big) U_l(r) = 0 \tag{7.76}$$

This is a huge simplification that allows to solve many problems in quantum mechanics, such as the electron bound in a hydrogen atom by the Coulomb potential.

Since we are already engulfed in some aspects of quantum mechanics and as a preparation for what is to come in the second half of MMP 2 we digress a bit into commutation relations. We have defined the momentum and position operator above. In one dimension they are $p_x = -i\partial_x$ and x. A crucial property of these operators is their commutation relation. To start with, we note that for any function (or vector) $f(x)$

$$x(-i\partial_x)\, f(x) \neq (-i\partial_x)x\, f(x) = x(-i\partial_x)f(x) + (-i)f(x) \tag{7.77}$$

where we have simply used the product rule. We can write this as a relation between the two operators

$$[x, p_x] \equiv x\, p_x - p_x\, x = i \tag{7.78}$$

Here we have defined the commutator of two operators, written as square brackets. This object will play a fundamental role and appear in many different contexts. (7.78) can be generalised to three dimensions as a commutation relation between the operators $\vec{p}$ and $\vec{r}$

$$[x_j, p_k] \equiv x_j\, p_k - p_k\, x_j = i\, \delta_{jk} \tag{7.79}$$

where we have used the more systematic notation $\vec{p} = (p_x, p_y, p_z) \to (p_1, p_2, p_3)$ for the momentum and $\vec{r} = (x, y, z) \to (x_1, x_2, x_3)$ for the position operator.

The commutation relations of the angular momentum operator (7.73) can be derived from (7.79) and are

$$[L_j, L_k] = i\, \epsilon_{jkl}\, L_l \tag{7.80}$$

where we have used the totally antisymmetric ϵ_{jkl} symbol. Written out, (7.80) reads

$$[L_1, L_2] = i\, L_3 \qquad \text{and cyclic} \tag{7.81}$$

We will encounter these relations again and again. Better get used to them and be prepared to see them with $\hbar$ not set to 1.

8 Intermezzo: MMP I $\rightarrow$ MMP II and 1+1=2

As threatened in the preface, we will now change a bit the way we look at mathematical issues, from a mainly computational perspective to a more conceptual point of view. If you think this has nothing to do with physics, think again. What we are looking at in the following is oozing with physics. In order not to get lost, this intermezzo provides a very rough outline on how we will proceed.

The most basic concept at the beginning of our journey is a *set [Menge]* $\mathbb{X}$, simply a collection of *elements [Elemente]*, typically denoted by x_i. The set can have a finite number of elements, a *countable [abzählbar]* infinite number of elements or in fact an *uncountable [überabzählbar]* infinite number of elements (in which case you cannot think of the index i running from 1 to some upper limit, possibly ∞). While mathematicians may rejoice at this, from a physics point of view a set alone is not particularly interesting, nor particularly relevant. To make it more interesting, we have to ask for a bit more and insist on some additional structure of the set.

Broadly speaking, this can be done through an 'algebraic' or an 'analytic' structure. Let's start with the former which brings us to 1+1=2. However trivial you might think this is, it is actually quite amazing. It is an example of a set (say the integers $\mathbb{Z}$) with an additional structure, a binary operation, in that for each (in general ordered) pair of elements (x_i, x_j) with $x_i \in \mathbb{X}$ and $x_j \in \mathbb{X}$ we associate uniquely one element $x_k \in \mathbb{X}$. As a shortcut, this binary operation is written as $x_i + x_j = x_k$ or $x_i \cdot x_j = x_k$ or $x_i * x_j = x_k$ etc. Everyday sets (the set of flowers in your garden) most of the time do not have such a binary operation (or do you know what 'the dandelion growing near the apple tree' + 'the daisy next to the pond' is ?).

If we ask just a little more, i.e. a few additional properties of the set and the binary operation we arrive at a *group [Gruppe]*, the simplest algebraic structure we will consider. In this case the binary operation is typically written as 'multiplication', i.e. $x_i \cdot x_j = x_k$ but it should be clear by now, that this operation can be much more general than the multiplication of e.g. real numbers. In particular, we often have $x_i \cdot x_j \neq x_j \cdot x_i$. Groups play an immensely important role in physics and we will consider them in Section 11, including finite groups (i.e. groups that are finite sets) as well as infinite groups.

A richer algebraic structure you are already familiar with is a *vector space [Vektorraum oder linearer Raum]* V. A vector space also has a binary operation (here typically written as $x_i + x_j = x_k$) but in addition also multiplication with real or complex numbers is defined. Thus, for $\alpha_i \in \mathbb{K}$ where $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$ (depending on whether we consider a real or a complex vector space) we can define linear combinations $\forall x_i, x_j \in V$ and $\forall \alpha_i, \alpha_j \in \mathbb{K}$ $\exists!\ \alpha_i\, x_i + \alpha_j\, x_j \in V$. Of course, the additional axioms you are familiar with from linear algebra apply. What is not defined in a vector space is something like $x_i \cdot x_j$ or $(x_i)^2$, hence the name linear space and investigating vector spaces is called linear algebra. Vector spaces will show up (an have shown up) throughout MMP.

An even richer algebraic structure which we will encounter peripherally in connection with groups is – sit down – an *algebra [Algebra]*. For us, an algebra is simply a vector space with an additional group-like binary operation that defines $x_i \cdot x_j$. Thus, in an algebra we can multiply elements with real (or complex) numbers, as well as elements among themselves. The familiar rules of computation hold.

So far so good, now we have an idea what was alluded to by $1 + 1 = 2$. But what about

MMP I $\to$ MMP II, or $x_i \to x_j$ if expressed in terms of elements of our set $\mathbb{X}$?

We will use the arrow $\to$ for a multitude of situations (like $+$ or $\cdot$ above), but it always signifies a 'going over to' or 'approaching'. Let's start with a sequence $\{x_l\}$ of elements of $\mathbb{X}$. If we write $x_{\{l\}} \to x$, this means the elements in the sequence approach x, or get closer to x or 'converge' to x. For this to make sense we need a measure of distance. Mathematically this means we have to introduce a *metric [Metrik]*. To each set (x_i, x_j) with $x_i \in \mathbb{X}$ and $x_j \in \mathbb{X}$ we associate a real (!) number, the distance, often denoted by $d(x_i, x_j)$. Again, this is an additional structure we need to impose on the set to promote it to a *metric space [Metrischer Raum]*. Once we have this we can define *open [offen]* subsets, *closed [abgeschlossen]* subsets, convergence, Cauchy sequence, continuity ... to our hearts desire. Typically, these are local properties we are investigating and the additional structure we have introduced, the metric, can be seen as a 'analytic' structure. Remember, where did you learn about convergence, continuity and all this? Right, in analysis, ding-a-ling (not algebra). Many things you have learnt there can be taken over and generalised to metric spaces. We will start with this in Section 9.

But in fact, we want more. The true fun starts when we combine the algebraic structure of the vector space with the analytic structure of the metric, or somewhat generalised a *norm [Norm]*. A norm is a measure of length and associates to each $x_i \in V$ a real number, its length, typically written as $|x_i| \in \mathbb{R}$ or $\|x_i\| \in \mathbb{R}$. A vector space with a norm is basically a *Banach space [Banachraum]* (a mathematician will cringe at this remark, as we 'forgot' something very important). If a vector space has a norm, we can always get a metric by defining the distance between x_i and x_j as $d(x_i, x_j) = \|x_i - x_j\|$ (note that $x_i - x_j$ is defined, since we are in a vector space!).

But in fact, we want even more! Banach spaces are fine and good, but for physics Hilbert spaces are much more important. A Hilbert space is a vector space where we have an *inner product [inneres Produkt]* or *scalar product [Skalarprodukt]* (listen to the mathematician cringing again!). Remember that in a vector space we do not have multiplication of elements. But if we have a binary operation (x_i, x_j) to $\alpha \in \mathbb{K}$ (as always with a few well chosen properties) that generalises the notion of an 'angle' between x_i and x_j, we have in fact also a norm. Following our notation from MMP I we will denote the scalar product as $\langle x_i | x_j \rangle$ and then the norm can be defined as $\|x_i\| = \sqrt{\langle x_i | x_i \rangle}$. Thus, if we have a Hilbert space we have a Banach space and if we have a Banach space we have a metric space. In Section 9 we will start from the simple metric space and work ourselves to the richer Hilbert space. Much of this will be a repetition of linear algebra, but we will include the case on infinite dimensions. You should have learnt that a finite dimensional vector space can always be made (i.e. is isomorphic) to a Hilbert space. In infinite dimensions this is not true any longer.

In Section 10 we will study linear operators in Hilbert spaces. Again, quite a few things will be very similar to finite dimensional Hilbert spaces, but others will be drastically different. Hilbert spaces (also infinite dimensional ones) are everywhere in physics. The whole drama of quantum mechanics plays in Hilbert spaces. They will also pop up again in group theory in Section 11 and, hence, anything that has to do with symmetries. Don't you dare to say MMP II is not physics.

To round this off, in Section 12 we will come back to more intuitively accessible topics, when we will apply some of the concepts to the three-dimensional Euclidean space and discuss coordinate transformations and tensors in general. To make sure you won't get overly attached to intuition, we will then (time permitting) extend this to Minkowski space-time.

9 Abstract spaces

From linear algebra in n-dimensional real or complex vector spaces you will be familiar with expressions like

$$|x\rangle = \sum_{i=1}^{n} c_i \, |e_i\rangle \tag{9.1}$$

wherby a vector $|x\rangle$ is expanded in a basis $|e_i\rangle$. Similarly you will have calculated eigenvectors and eigenvalues λ_i and (hopefully) came across the spectral representation of (say a hermitean) $n \times n$ matrix/linear operator T as

$$T = \sum_{i=1}^{n} \lambda_i \, P_i \, . \tag{9.2}$$

Here P_i are projectors onto the eigenspace of the eigenvalue λ_i and we have assumed the eigenvalues are not degenerate.

The aim of the following two sections is to deepen our understanding of these matters and generalize them to $n \to \infty$, whenever possible. The latter point is rather subtle and we will have to start by reminding ourselves of a few somewhat formal concepts that will play an important role for making the plunge $n \to \infty$.

If you prefer applied topics, these sections will be horribly formal. If you prefer mathematical topics, these sections will be horribly handwaving. I will try to find a reasonable balance between these two worlds and alert you to some subtleties and mathematical concepts, without following a completely rigorous approach.

We will first consider metric spaces. Everything we learn about them will also hold for Banach spaces, i.e. spaces with a norm. Finally, everything we learn for Banach spaces will also hold for Hilbert spaces, i.e. spaces with a scalar product.

9.1 Metric spaces

Let us consider a set $\mathbb{X}$ with elements (sometimes called points) x, y, z or x_i. In this subsection we do not require this set to be a vector space. The only additional structure we insist upon is a *metric [Metrik]*, which is a binary operation $d(x_i, x_j)$ with the properties

$$d(x,y) \in \mathbb{R} \qquad \forall \; x, y \in \mathbb{X} \tag{9.3a}$$

$$d(x,y) = 0 \qquad \Leftrightarrow \qquad x = y \tag{9.3b}$$

$$d(x,y) \leq d(x,z) + d(y,z) \tag{9.3c}$$

That's all. Everything we develop in this subsection will be built on these three properties, and nothing else. We have used the notation $d(.,.)$ for distance, as indeed the metric can be thought of as a measure of distance. (9.3c) is called the triangle inequality. The symbol $\Leftrightarrow$ stands for iff (i.e. if and only if).

As a first step we can derive two further properties related to a metric ⇒ lecture

$$d(x,y) \geq 0 \tag{9.3d}$$

$$d(x,y) = d(y,x) \tag{9.3e}$$

You might think it is 'obvious' that a distance should not be negative and the distance from x to y should be the same as the distance from y to x. But it is not about being 'obvious' or not. All that matters is that (9.3d) and (9.3e) can be derived from (9.3a)–(9.3c) and, hence, from now on can be used whenever we talk about a metric.

A *metric space [Metrischer Raum]* $M = (\mathbb{X}, d)$ is a set $\mathbb{X}$ together with a metric. Physicists are often sloppy and just give $\mathbb{X}$, thinking it is 'obvious' which d to use. Now, what did we say about 'obvious' a few lines above? Exactly. (You will find that here I am preaching water and soon I will be drinking wine, as I am a physicist after all.)

There are endless examples of metric spaces. We just mention a few that will show up again and again. As a first example, let $\mathbb{X} = \mathbb{R}^n$ with n finite, i.e. $\mathbb{X} = \{x | x = (x_1, x_2, \dots x_n)\}$ with $x_i \in \mathbb{R}$.[3] Next, let's define two metrics

$$d(x, y) = \Big(\sum_{i=1}^{n} (x_i - y_i)^2 \Big)^{1/2} \qquad M_{\mathbb{R}^n} = (\mathbb{R}^n, d) \tag{9.4a}$$

$$d_m(x, y) = \max_i |x_i - y_i| \qquad M_{\max} = (\mathbb{R}^n, d_m) \tag{9.4b}$$

Even though the set $\mathbb{X}$ is the same for the two metric spaces $M_{\mathbb{R}^n}$ and $M_{\max}$, these are two different metric spaces and have nothing to do with each other. Of course, we have to show that both examples in (9.4) satisfy (9.3a)–(9.3c). The first example, d is just the canonical metric if x and y are interpreted as points in an n dimensional space. You might say: wonderful, $\mathbb{R}^n$ is actually a vector space, and we can even make a Hilbert space out of it. While this is true, we could not care less at the moment. We are happy with what we have, a metric. In passing we note that we can easily generalize $M_{\mathbb{R}^n}$ (9.4a) to a complex vector space as

$$d(x, y) = \Big(\sum_{i=1}^{n} |x_i - y_i|^2 \Big)^{1/2} \qquad M_{\mathbb{C}^n} = (\mathbb{C}^n, d) \tag{9.5}$$

and we will often work with $M_{\mathbb{K}^n} = (\mathbb{K}^n, d)$ where it is understood that $\mathbb{K}$ is either $\mathbb{R}$ or $\mathbb{C}$.

Another example is to do with the set ℓ^p which consists of the elements $x = (x_1, x_2, \dots)$ with $x_i \in \mathbb{C}$ (i.e. the non terminating sequences of complex numbers) with $\sum_i |x_i|^p < \infty$. This can be made to a metric space using

$$d_p(x, y) = \Big(\sum_{i} |x_i - y_i|^p \Big)^{1/p} \qquad M_{\ell^p} = (\ell^p, d_p) \tag{9.6}$$

In particular the case $p = 2$ will be very important and in some sense is a generalisation of $M_{\mathbb{C}^n}$ in (9.5) to $n = \infty$. We can also think of taking the limit $p \to \infty$ and arrive at the bounded sequences ℓ^∞, i.e. $x = (x_1, x_2, \dots)$ with $x_i \in \mathbb{C}$ and $|x_i| < c_x \ \forall i$, where c_x is a constant. Then we have a metric space

$$d_\infty(x, y) = \sup_i |x_i - y_i| \qquad M_\infty = (\ell^\infty, d_\infty) \tag{9.7}$$

The last example we make for the moment is $C[a, b]$ or $C^n[a, b]$, the set of (continuous or n times continuously differentiable) functions on the real interval $[a, b]$. Again we can make

[3] Do not confuse x_i here (the ith component of the n tuple x) with x_i in the first line of this subsection (a generic element of the set $\mathbb{X}$).

different metric spaces out of this set as follows:

$$d_m(x,y) = \max_{t\in[a,b]} |x(t) - y(t)| \qquad M_1 = (C^n[a,b], d_m) \tag{9.8a}$$

$$d_i(x,y) = \int_a^b |x(t) - y(t)|\, dt \qquad M_2 = (C^n[a,b], d_i) \tag{9.8b}$$

As we will see, M_1 and M_2 in (9.8) have very different properties.

As one more example, let's come back to our favourite set $\mathbb{X}_*$, the set of flowers in our garden. Can we make a metric space out of this? The answer is yes, but .... With the discrete metric that is defined as $d_d(x,x) = 0$ and $d_d(x,y) = 1\ \forall x \neq y$ we can make a metric space out of every set. However, we do not gain too much, as all statements we will be making in the remainder of this subsection will be (correct but) fairly 'empty' for this case. Still, it might be a good exercise to apply the concepts to the flower space that we will denote by $M_* = (\mathbb{X}_*, d_\mathrm{d})$.

Finally a warning: physicists sometimes call something a metric that actually is not a metric in the sense we just have defined. The most important example of this is in (special) relativity with two 4-vectors $x^\mu = (t_x, \vec{x})$ and y^μ a "metric" is defined as $d_?(x,y) = t_x t_y - \vec{x}\cdot\vec{y}$. This contradicts (9.3b) and also (9.3d). Thus, do not blindly apply what we learn here to the "metric" of special relativity. We might come back to this in Section 12.

We can now copy numerous concepts you are familiar with from analysis so we have them ready for our generic metric spaces. Let $x_0, x, y, y_0 \in \mathbb{X}$ and $M_X = (\mathbb{X}, d)$. ⇒ lecture

- open *ball [Kugel]* with centre x_0 and radius $r \in \mathbb{R}$: $B(x_0, r) = \{x | d(x, x_0) < r\}$
- closed ball with centre x_0 and radius $r \in \mathbb{R}$: $\tilde{B}(x_0, r) = \{x | d(x, x_0) \leq r\}$
- *sphere [Sphäre]* with centre x_0 and radius $r \in \mathbb{R}$: $S(x_0, r) = \{x | d(x, x_0) = r\}$
- ϵ *neighbourhood [Umgebung]* of x_0: $B(x_0, \epsilon)$

To keep going we define a *subset [Teilmenge]* $\mathbb{Y} \subset \mathbb{X}$ with the resulting metric space $M_Y = (\mathbb{Y}, d)$ a *subspace [Teilraum]* of M_X. Its complement is $M_Y^c = (\{x | x \in \mathbb{X} \wedge x \notin \mathbb{Y}\}, d)$. Then we have ⇒ lecture

- M_Y is called *open [offen]* iff $\forall y \in \mathbb{Y}\ \exists \epsilon > 0$ s.t. $B(y, \epsilon) \subset \mathbb{Y}$. An open ball is open. This requires a proof!
- M_Y is called *closed [abgeschlossen]* iff M_Y^c is open. Don't be mislead by the everyday association of the terms open and closed. Subsets are not either open or closed. They can be one of the two, but also both or neither.
- y_0 is an *accumulation point [Häufungspunkt]* (of a set, not of a sequence) of M_Y iff $\forall \epsilon > 0$ we have $B(y_0, \epsilon) \cap \{\mathbb{Y} \backslash y_0\} \neq \varnothing$ where $\varnothing$ is the empty set. In words, any neighbourhood of an accumulation point y_0 contains at least one point of the set other than y_0 itself. Note y_0 may or may not be an element of $\mathbb{Y}$.
- the *closure [abgeschlossene Hülle]* of M_Y, denoted by $\bar{M}_Y$ contains all elements $y \in \mathbb{Y}$ and all accumulation points. A closed space contains all its accumulation points. The

closure of a subset is closed. If you think this is 'obvious', here a warning that sometimes the 'obvious' is actually wrong: one might think the closure of the open ball $B(x_0, r)$ is the closed ball $\tilde{B}(x_0, r)$. But in general this is not true!

- y_0 is a *boundary point [Randpunkt]* of M_Y iff $\forall B(y_0, \epsilon)$ we have $B \cap \mathbb{Y} \neq \varnothing$ and $B \cap \mathbb{Y}^c \neq \varnothing$. In words, any neighbourhood of a boundary point y_0 of a subset $\mathbb{Y}$ contains at least one point of the subset and at least one point that does not belong to the subset. Note y_0 may or may not be an element of $\mathbb{Y}$. Sometimes the symbol $\partial\mathbb{Y}$ is used for the set of all boundary points (i.e. the boundary) of $\mathbb{Y}$. In fact we have used this notation already in MMP I, e.g. for Neumann boundary condition.

- M_Y is called *bounded [beschränkt]* if $\exists B(x_0, r)$ such that $\mathbb{Y} \subset B(x_0, r)$, i.e. if one can find any ball that completely contains $\mathbb{Y}$.

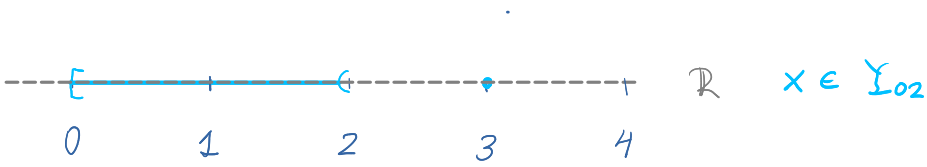

Figure 9.1: Light blue points of the set $\mathbb{Y}_{02}$.

Since we did not have a lovely picture for quite a while, let us put one in and make an example with M_X set to $M_{\mathbb{R}} \equiv M(\mathbb{R}, |\ |)$ with the canonical metric $d(x, y) = |x - y|$ as in (9.4a) with $n = 1$ and $M_{Y_{02}} = (\mathbb{Y}_{02}, |\ |)$ with $\mathbb{Y}_{02} = \{x | 0 \leq x < 2 \vee x = 3\} \subset \mathbb{R}$, as shown in Figure 9.1. Let us consider a few elements/points:

$x_0 = 0 \in M_{Y_{02}}$ is an accumulation point and a boundary point

$x_1 = 1 \in M_{Y_{02}}$ is an accumulation point but not a boundary point

$x_2 = 2 \notin M_{Y_{02}}$ is an accumulation point and a boundary point

$x_3 = 3 \in M_{Y_{02}}$ is a boundary point but not an accumulation point

$x_4 = 4 \notin M_{Y_{02}}$ is neither an accumulation nor a boundary point

The closure of $\mathbb{Y}_{02}$ is $\bar{\mathbb{Y}}_{02} = \{x | 0 \leq\ x \leq 2 \vee\ x = 3\}$ (mmmhh, this wine tastes nice) and $M_{Y_{02}}$ is bounded. Indeed, $B(25, 42152)$ is a ball with $\mathbb{Y}_{02} \subset\ B$. You might argue this choice of B is not the most efficient, but who cares. We found a B with $\mathbb{Y}_{02} \subset\ B$, so $\mathbb{Y}_{02}$ is bounded, mission accomplished. Of course, if we can find a ball with some centre y and radius r that contains a set, we can find for any other centre y_1 a ball that contains the set. Using the triangle inequality we choose the radius as $r + d(y, y_1)$.

Whether or not a certain set $\mathbb{Y}$ with a metric d is open or not depends on the space in which it is embedded. The standard example is for $\mathbb{Y}$ to take the set of rational numbers, say $0 < y < 1$ with the metric $d(y_1, y_2) = |y_1 - y_2|$, i.e. (9.4a) with $n = 1$. Then $\mathbb{Y}$ is an open subset of $\mathbb{Q}$ the set of all rational numbers. But $\mathbb{Y}$ is not an open subset of $\mathbb{R}$.

Another concept we can take over from analysis is convergence. Consider a sequence $x_i \in M_X$. We say the sequence x_i converges to x_0 and write $x_i \to x_0$ or $\lim x_i = x_0$ iff $\forall \epsilon > 0\ \exists N(\epsilon)$ s.t. $\forall n > N(\epsilon)$ we have $x_n \in B(x_0, \epsilon)$. The element x_0 is called the *limit* *[Grenzwert]* of the converging sequence. Looking at this, you do not even realise we're not in the Analysis 1 course. The generalisation is hidden in $B(x_0, \epsilon)$ which now is a neighbourhood in a general metric space. Now we know about convergence, we can give a different (but equivalent) definition of closed. A metric space M_Y is closed iff the limit of every convergent sequence lies itself in M_Y. ⇒ lecture

Similarly, the notion of continuity of a map is now generalised without any effort. A map $T : M_X \mapsto M_Z$ is called *continuous* *[stetig]* at x_0 iff

$$x_i \to x_0 \qquad \Rightarrow \qquad Tx_i = z_i \to Tx_0 = z_0 \tag{9.9}$$

In other words, under continuous maps convergent sequences are mapped into convergent sequences. Of course, there are many equivalent definitions of continuous, as familiar from Analysis 1. Also, the remarkable **theorem** that a map is continuous iff the inverse image of any open (closed) subspace of M_Z is open (closed) can be taken over.

What we have done so far is to regurgitate and generalize concepts we are familiar with from analysis. The next concept is also known in $M(\mathbb{R}, |\ |)$, but its generalisation to general metric (and later Hilbert) spaces is rather subtle. At the same time it will be vitally important and unfortunately has sometimes different and confusing terminology. Hence it deserves a special treatment.

It is about the notion of a *compact* *[(folgen)kompakt]* (sub)space. A (sub)space $(\mathbb{Y}, d)$ is called compact if every sequence has a convergent subsequence. What is not clear at the moment is whether the limit element is required to be in $\mathbb{Y}$ or not. We will use the notion that compact implies the limit is in $\mathbb{Y}$. In German, this is sometimes also called *[in sich kompakt]*. A less severe requirement is for a (sub)set $\mathbb{Y}$ to be *relative compact* *[relativ kompakt]* whereby each sequence is required to have a converging subsequence, but the limit of this subsequence need not be an element of $\mathbb{Y}$. To make things worse, some books (written in German) call this "kompakt". So be alert.

In $M_{\mathbb{R}^n}$, $(n < \infty)$ a subset is compact iff it is closed and bounded, according to Bolzano-Weierstrass. So we could take closed and bounded as definition for compact. However, in general this is not the case any longer. In fact, it is still true that if a subspace is compact, it is also closed and bounded. But the converse does not hold as illustrated by the following ⇒ lecture example in M_{ℓ^p}, (9.6). Consider the subspace $M_Y \subset M_{\ell^p}$ where $\mathbb{Y} = \{y_i | (0 \ldots 1, 0, \ldots)\}$, i.e. y_i has an entry 1 at the ith position and 0 everywhere else. The set $\mathbb{Y}$ is closed (it has no accumulation points, so all of them are part of the set) and bounded, take e.g. $y_i \in B(y_0, 2)$ with $y_0 = (0, 0 \ldots)$. However, the sequence of y_i has no convergent subsequence. The reason this can happen is that there are infinitely many entries and the entry 1 can wander off towards ever higher positions. If we had only a finite number of entries, this would not be possible. Related to this, we will see that the notion of compact becomes very important in infinite-dimensional spaces (even though we have not yet defined what this is) as a little bit more than bounded and closed. As a first preparation to this we note that under a continuous map the image of a compact set is again compact. The proof can be given in two lines. ⇒ lecture

Keeping to copy from Analysis 1 we recall that $M_Y \subset M_X$ is called a *dense* *[dicht]* subset of M_X if $\bar{\mathbb{Y}} = \mathbb{X}$. In words, every element $x \in M_X$ is either in M_Y or can be approximated arbitrarily well by an element in M_Y. Or, in other words, every element $x \in M_X$ is either in

M_Y or is an accumulation point of M_Y. The standard example is rational numbers being a dense subset of real numbers.

Moving on we call a metric space $M_X = (\mathbb{X}, d)$ *separable [separabel]* iff $\exists M_Y = (\mathbb{Y}, d)$ with $\bar{\mathbb{Y}} = \mathbb{X}$ and $\mathbb{Y}$ has a countable number of elements. Thus a separable space M_X is 'nearly countable' in that there exists a countable space that is dense in M_X. Our example above $M_{Y_{02}}$ is separable, since $\mathbb{Q}_{02}$ (all rational numbers between 0 and 2 and the number 3) is dense in $\mathbb{Y}_{02}$ and rational numbers are countable. Similarly $M_{\mathbb{R}^n}$, $M_{\mathbb{C}^n}$ and M_{ℓ^p} are separable. An example of a space that is not separable is M_∞, (9.7).

If M_X is compact it is separable. The idea of the proof is to show that compactness implies $\forall \epsilon > 0$ we can cover the space with a finite number of $B(x_i, \epsilon)$. Now take $\epsilon \to 1/n$ with $n \in \mathbb{N}$ and we get a countable union of finite sets that is dense and covers the space. ⇒ lecture

The for the moment last concept we are going to steal from analysis is that of a *Cauchy sequence [Cauchy Folge]*. Let x_i be a sequence in $M = (\mathbb{X}, d)$. The sequence is called a Cauchy sequence if $\forall \epsilon > 0\ \exists N(\epsilon)$ s.t. $\forall m, n > N(\epsilon)$ we have $d(x_m, x_n) < \epsilon$. In words, the elements of a Cauchy series at some point get closer and closer to each other. It is easy to see that any converging sequence is a Cauchy sequence. The reverse need not be true. However iff in $M = (\mathbb{X}, d)$ every Cauchy sequence converges to an element in M we call M *complete [vollständig]*. From Analysis 1 we know $M_\mathbb{R}$ is complete, so is $M_{\mathbb{K}^n}$ and M_{ℓ^p}.

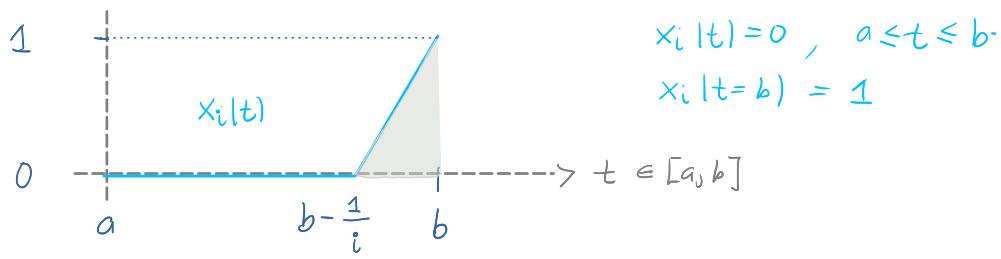

Figure 9.2: Functions $x_i(t) \in C[a, b]$ with a limit $x_0 \notin C[a, b]$.

Once more, the notion of completeness is a property of a metric space, not a set. To illustrate this, lets go back to the examples M_1 and M_2 of (9.8). In fact, M_1 is complete but M_2 is not, as the example in Figure 9.2 illustrates. For $m, n \to \infty$ we have $d(x_m, x_n) \to 0$ (the triangular area tends to zero) but in the limit $i \to \infty$ we have $x_i(t) \to x_0(t)$ with $x_0(t) = 0$ for $a \le t < b$ and $x_0(b) = 1$. The key point is that the limit x_0 is not continuous and, hence, not in $M_2 = (C[a, b], d_i)$. This is typical for incomplete spaces. They have holes and the limit of Cauchy sequences might 'fall' in one of the holes. This is precisely the situation with rational numbers, where the limit of a Cauchy series (of rational numbers) might be a real (but not rational) number. ⇒ lecture

You might think it is a bit of a disaster if a metric space is not complete. But not really. If we have a metric space $M = (\mathbb{X}, d)$ that is not complete we can simply *complete [vervollständigen]* it. This means $\exists M_c = (\mathbb{Y}, d_c)$ with M_c complete and $\bar{M} = M_c$. Or M is dense in the complete space M_c. The space M_c is in some sense unique. To be more precise with the last statement we have to introduce *isometry [Isometrie]*. Let's consider two spaces $M_1(\mathbb{X}_1, d_1)$ and $M_2(\mathbb{X}_2, d_2)$. The spaces M_1 and M_2 are called *isometric [isometrisch]* if there is a bijective map $T : M_1 \mapsto M_2$ with $d_1(x_1, y_1) = d_2(Tx_1, Ty_1)$, $\forall x_i, y_1 \in \mathbb{X}_1$. A streetwise

translation of isometry is 'they are the same' and we write $M_1 \cong M_2$. In fact you could argue that Tx_i is simply 'another name' for x_i and the structure of the x_i in M_1 is precisely copied over to Tx_i in M_2. This has copycat written all over it.

Thus, even though completeness is actually a very important property (who would want to have holes) physicists typically do not care too much. If the space is not complete, they actually simply mean its completion.

Just before me move on to Banach spaces. Isn't it remarkable how much time we managed to spend with (we could have spent much more) and how rich the consequences of the humble looking (9.3a)–(9.3c) are?

9.2 Banach spaces

The sets we have considered so far were completely general and all we were asking for was a metric. Now we make quite a radical additional demand. From now on, $\mathbb{X}$ is a *vector space [linearer Raum / Vektorraum]* and we will call its elements vectors. We will only consider real or complex vector spaces and often write $\mathbb{K}$ without specifying whether we mean $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$. Since $\mathbb{X}$ is a vector space we can form linear combinations $k_1\, x_1 + k_2\, x_2 \in \mathbb{X}$ where $k_i \in \mathbb{K}$ and $x_i \in \mathbb{X}$. Thus we have a (rather sophisticated) algebraic structure.

But we are not yet satisfied. We want more! We also want a *norm [Norm]* to be defined on the vector space which makes it a normed space. A norm associates a positive real number (also for complex vector spaces) $\|x\|$ (or $|x|$) to all $x \in \mathbb{X}$ that can be thought of as 'length'. On a formal level it has to satisfy the following properties $\forall x, y \in \mathbb{X}$ and $k \in \mathbb{K}$:

$$\|x\| \in \mathbb{R} \quad \text{with} \quad \|x\| \geq 0 \tag{9.10a}$$

$$\|x\| = 0 \qquad \Leftrightarrow \qquad x = 0 \tag{9.10b}$$

$$\|k\, x\| = |k|\, \|x\| \tag{9.10c}$$

$$\|x + y\| \leq \|x\| + \|y\| \tag{9.10d}$$

Note that (9.10b) has two 'different' 0. The first is the real number 0, the second is the 0 element of the vector space. A norm induces a metric through $d(x, y) = \|x - y\|$. It is an easy exercise to show that (9.10) entail (9.3). Furthermore, (9.10d) is directly linked to (9.3c) and is also called triangle inequality. The metric induced through a norm is invariant under translations, i.e. $d(x + z, y + z) = d(x, y)$.

Thus, every normed space is a metric space (but the converse is obviously not true at all). Hence normed spaces can be complete or not. If the normed space is not complete, we can simply replace it by its completion. Then we arrive at a *Banach space [Banachraum]*, a complete normed space. Completeness was the property that we 'forgot' in Section 8, but now we see it's not too bad. We can sort it out by completion. Still, we must not forget that the statements we are going to make about Banach spaces do not necessarily hold for non-complete normed spaces.

Let's go through some examples, familiar from Section 9.1. We start with $\mathbb{K}^n$. In Section 9.1 we treated this set just as a set, but of course it is in fact a vector space. We introduce a norm as

$$x = (x_1 \dots x_n) \in \mathbb{K}^n \qquad \text{then} \qquad \|x\| = \Big(\sum_{i=1}^{n} |x_i|^2 \Big)^{1/2} \tag{9.11}$$

It should be fairly obvious that (9.11) satisfies (9.10) and induces the metric (9.5). Also, this space is complete, thus $\mathbb{K}^n$ with the norm (9.11) is a Banach space.

Our next example is ℓ^p. This is also 'more' than a set, namely a vector space, since $\forall x, y \in \ell^p$ and $\forall k_1, k_2 \in \mathbb{C}$ we have $\sum |k_1\, x + k_2\, y|^p < \infty$. Hence the linear combination $k_1\, x + k_2\, y \in \ell^p$ and all axioms of a vector space are satisfied. We now define the norm as

$$x = (x_1, x_2 \ldots) \in \ell^p \qquad \text{then} \qquad \|x\| = \Big(\sum_{i=1}^{\infty} |x_i|^p\Big)^{1/p} \tag{9.12}$$

Again, we need to show that (9.12) satisfies (9.10). The metric that is induced by (9.12) is obviously d_p of (9.6). This metric space is also complete (proof needed!) hence ℓ^p with the norm (9.12) is a Banach space.

As for metric spaces, also for normed spaces it is not sufficient to specify just the set (or here the vector space). We can define different normed spaces with the same vector space. To illustrate this we go back to (9.8) but now treat $C[a, b]$ as a linear space. For $x, y \in C[a, b]$, i.e. $x(t)$ and $y(t)$ continuous functions in the interval $a \leq t \leq b$ we define $(k_1\, x + k_2\, y)(t) = k_1\, x(t) + k_2\, y(t)$. This is also a continuous function $C[a, b]$ and again, all axioms of a vector space are satisfied. Now we want to make a normed space out of this vector space. There is more than one possibility, very similar to (9.8a) and (9.8b):

$$x = x(t) \in C[a, b] \qquad \text{then} \qquad \|x\|_m = \max_{t \in [a,b]} |x(t)| \tag{9.13a}$$

or

$$x = x(t) \in C[a, b] \qquad \text{then} \qquad \|x\|_i = \int_a^b |x(t)|\, dt \tag{9.13b}$$

or

$$x = x(t) \in C[a, b] \qquad \text{then} \qquad \|x\|_2 = \Big(\int_a^b |x(t)|^2\, dt\Big)^{1/2} \tag{9.13c}$$

The norm (9.13a) induces the metric (9.8a) and (9.13b) induces (9.8b) and for good measure we have introduced yet another norm. But the normed spaces $(C[a, b], \|\,.\,\|_m)$ and $(C[a, b], \|\,.\,\|_i)$ and $(C[a, b], \|\,.\,\|_2)$ are quite different. The first is complete, hence a Banach space, whereas the latter two are not complete. So we should prefer $(C[a, b], \|\,.\,\|_m)$, shouldn't we. Well, mathematicians do, but we don't. We'll get to love $(C[a, b], \|\,.\,\|_2)$.

Since normed spaces are vector spaces, we can copy concepts from linear algebra. In particular we call the set of $n < \infty$ vectors $\{x_1 \ldots x_n\}$ *linearly independent [linear unabhängig]* iff $\sum_{i=1}^n k_i\, x_i = 0$ entails $k_i = 0, \forall i$. Now what about an infinite set of vectors? We call the infinite set of vectors $\{x_1 \ldots\}$ linearly independent, iff all non-empty finite subsets of $\{x_1 \ldots\}$ are linearly independent.

Another concept familiar from linear algebra for finite n is the *dimension [Dimension]* of the vector space. If $\exists n$ linearly independent vectors in $\mathbb{X}$ but $\nexists (n + 1)$ independent vectors in $\mathbb{X}$ we say $\mathbb{X}$ has dimension n or $\dim\mathbb{X} = n$. Now if $\exists n$ linearly independent vectors $\forall n \in \mathbb{N}$ we say the vector space is infinite dimensional or $\dim\mathbb{X} = \infty$. This is our first timid step into the world of infinite-dimensional vector spaces. A bit like Alice in Wonderland.

Since normed spaces (complete or not) are metric spaces we can also carry over all concepts of Section 9.1 and combine them with the above. In particular we have the notion of

convergence $x_i \to x_0$ which we now also can write as $\lim_{i\to\infty} \|x_i - x_0\| = 0$. We want to exploit this to extend the definition of a basis to infinite-dimensional vector spaces. First let's recall the definition of a basis in a n-dimensional vector space. Iff we have a set of n vectors $\{e_1 \dots e_n\}$ such that $\forall x \in \mathbb{X}$ there $\exists! \alpha_i \in \mathbb{K}$ with $x = \sum_{i=1}^n \alpha_i\, e_i$ then the set $\{e_1 \dots e_n\}$ is called a basis. The α_i are called the *coefficients [Koeffizienten]* of the vector x in this basis. This is precisely (3.1), just with a different notation. But now we want to move towards (3.5) or, more generally speaking, define a basis for infinite-dimensional vector spaces. An here it comes: Iff we have a set of (infinitely many) vectors $\{e_1 \dots e_i \dots\}$ such that

$$\forall x \in \mathbb{X} \quad \exists! \alpha_i \in \mathbb{K} \quad \text{s.t.} \quad \lim_{n\to\infty} \|x - \sum_{i=1}^{n} \alpha_i\, e_i\| = 0 \tag{9.14}$$

then the set $\{e_1 \dots e_i \dots\}$ is called a basis, the (infinitely many) α_i the coefficients of x and we write

$$x = \sum_{i=1}^{\infty} \alpha_i\, e_i \tag{9.15}$$

Now compare this with (9.1). Yep, same thing, slightly different notation, and soon it will be clear why the notation is still somewhat different. On the danger of repeating myself, (9.15) looks fairly innocuous, but is actually desperately relying on the (analytic) notion of distance for the equal sign and the (algebraic) notion of a linear combination for the r.h.s. Only the combination of these two worlds make (9.1) possible. The fact that such equations are not just academic ivory-tower chat, but crucially important in everyday physics should be clear from the Fourier series (3.16a), the 'Legendre' series (7.24), and our example of a Schrödinger equation (7.69), to mention but a few examples.

Since we are dealing with a vector space we can introduce *linear operators [lineare Operatoren]*. They will stay with us for the rest of our lives. Their importance cannot be overstated. Let $\mathbb{X}$ and $\mathbb{Y}$ be two normed or better Banach spaces (where a norm is implied, even if not explicitly indicated) and $T : \mathbb{X} \mapsto \mathbb{Y}$ a map. If

$$T \quad \text{linear} \quad \Leftrightarrow \quad T(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 T x_1 + \alpha_2 T x_2 \quad \forall \alpha_i \in \mathbb{K}, \quad \forall x_i \in \mathcal{D}_T \subset X \tag{9.16}$$

Here $\mathcal{D}_T$ is the *domain [Definitionsbereich]* of T, i.e. the set of $x \in \mathbb{X}$ for which Tx is defined. Note that $\mathcal{D}_T$ is an essential part of the definition of T, but – you've guessed it – physicists often do not bother specifying it. In any case, $\mathcal{D}_T$ is itself always a vector space and if $\mathcal{D}_T = \mathbb{X}$ we say the operator T is defined everywhere. The case $\mathbb{Y} = \mathbb{X}$ is very common and we will often use the term operator instead of linear map. The word 'linear' is always implied when we say operator.

The *range [Wertebereich]* $\mathcal{R}_T \subset \mathbb{Y}$ is the set of $y \in \mathbb{Y}$ s.t. $\exists x \in \mathcal{D}_T$ with $Tx = y$. If $\mathcal{R}_T = \mathbb{Y}$ the map/operator T is called *onto* or *surjective [surjektiv]*. Finally we define the *kernel* or *null space [Kern / Nullraum]* $\mathcal{N}_T \subset \mathbb{X}$ of the operator T as $\{x | Tx = 0\}$, i.e. those points that are mapped to the zero vector in $\mathbb{Y}$. There is a clash of terminology, as this kernel has nothing to do with the kernel of integral transforms in e.g. (3.17), (3.39) or (7.58). Thus we prefer the terminology null space. It is a nice exercise to show that $\mathcal{R}_T$ and $\mathcal{N}_T$ are also vector spaces. ⇒ lecture

While we're at it we also recall two further definitions. T is called *injective [injektiv]* if $Tx_1 = Tx_2 \;\Rightarrow\; x_1 = x_2$. Thus, for injective maps different elements in $\mathcal{D}_T \subset \mathbb{X}$ are mapped

to different elements in $\mathcal{R}_T \subset \mathbb{Y}$. This gives us a chance to define an inverse of T. Finally, T is called *one-to-one* or *bijective [bijektiv]* if it is injective and surjective.

A map like (9.16) is also sometimes called a vector space *homomorphism [Homomorphismus]*. A homomorphism is a map that respects the structure of a space. This means, if in $\mathbb{X}$ we have say three vectors satisfying $x_1 + x_2 = x_3$ then under T they are mapped into y_1, y_2, and y_3 in $\mathbb{Y}$ such that the relation still holds, $y_1 + y_2 = y_3$. A bijective homomorphism is called an *isomorphism [Isomorphismus]*. If there exists an isomprphism between two vector spaces $\mathbb{X}$ and $\mathbb{Y}$ we call the spaces *isomorphic [isomorph]* and write $\mathbb{X} \cong \mathbb{Y}$, the same symbol we used for isometry. An isomorphism is the extension of isometry to vector spaces. It is a one-to-one map perserving the structure. Really, the two spaces are just 'the same', only giving different names to its elements.

An operator is called *bounded [beschränkt]* iff $\exists c \in \mathbb{R}$ s.t. $\|Tx\| \leq c\,\|x\| \quad \forall x \in \mathcal{D}_T \subset X$. If a certain c satisfies the constraint above, so does any $c' > c$. The smallest c for which the constraint is satisfied is called the *norm [Norm]* of the operator T, written as

$$\|T\| \equiv \sup_{x \in \mathcal{D}_T, x \neq 0} \frac{\|Tx\|}{\|x\|} = \sup_{x \in \mathcal{D}_T, \|x\|=1} \|Tx\| \tag{9.17}$$

where in the last step we used the linearity of T. By definition we then have

$$\|Tx\| \leq \|T\|\,\|x\| \tag{9.18}$$

At the risk of stating the obvious, the terms 'bounded' and 'norm' that have been introduced for operators just now must not be confused with the same words used for (sub)sets or vectors. But it should be fairly obvious, why the words have been recycled. In fact, the norm of an operator as defined in (9.17) satisfies the axioms of a norm, (9.10). ⇒ lecture

In a finite dimensional normed space all (linear) operators are bounded. But this is not true any longer in an infinite dimensional space. We will get to know several important examples at a later point.

However, there is a very important **theorem**. Let $T : \mathcal{D}_T \subset \mathbb{X} \mapsto \mathbb{Y}$ be a (linear) operator and $\mathbb{X}$ and $\mathbb{Y}$ normed spaces then

(i) T is continuous $\quad \Leftrightarrow \quad$ T is bounded

(ii) T is continuous at a single point $\quad \Leftrightarrow \quad$ T is continuous everywhere

After the (remarkably short) proof we can thus think of 'nice' operators and 'not so nice' operators. The former are the bounded = continuous ones. In fact, we will see later that 'nice' is sometimes not 'nice enough' and we will get to know compact operators that are 'even nicer' than bounded operators. All this is not to say that unbounded operators are not important. Unfortunately, there are very important unbounded operators, and we will have to do battle with them for QM's sake. ⇒ lecture

After this deluge of definitions it is time to look at some examples. The first one is the obvious one $T : \mathbb{K}^m \mapsto \mathbb{K}^n$ with m, n finite. These are the linear operators from the m-dimensional vector space $\mathbb{K}^m$ to the n-dimensional vector space $\mathbb{K}^n$ and once a basis has been chosen, they can be represented by a $n \times m$ matrix. Following from the remarks above, these operators are bounded, hence continuous.

As another example let $T : C[a,b] \mapsto \mathbb{K}$. Do you recognize this? Compare to (2.1). Yep, these are the linear functionals defined for continuous functions. If we choose instead C_c^∞ as in (2.1) and insist T is continuous (i.e. bounded) then $T : C_c^\infty \mapsto \mathbb{K}$ are the distributions and $x \in \mathbb{X} = C_c^\infty$ are the test functions. Compare (9.16) with (2.3) and (9.9) with (2.4). Indeed, it is the same and we even use the same letter, T.

We can generalise the concept of functionals from C_c^∞ to any vector space $\mathbb{X}$. To do so, we start by considering $B(\mathbb{X},\mathbb{Y})$, the space of all bounded linear operators of $\mathbb{X}$ (with some norm) to $\mathbb{Y}$ (with some norm). If $\mathbb{X} = \mathbb{K}^m$ and $\mathbb{Y} = \mathbb{K}^n$ we are back to our first example and $B(\mathbb{X},\mathbb{Y})$ is nothing but the space of all $n \times m$ matrices (once bases are chosen).

The space $B(\mathbb{X},\mathbb{Y})$ is itself a normed space. The elements of this space are maps. For $T \in B(\mathbb{X},\mathbb{Y})$ we have defined $\|T\|$ in (9.18) and $\alpha_1 T_1 + \alpha_2 T_2 \in B(\mathbb{X},\mathbb{Y})$ is defined as

$$(\alpha_1 T_1 + \alpha_2 T_2)x \equiv \alpha_1 T_1 x + \alpha_2 T_2 x \qquad \forall \alpha_i \in \mathbb{K}, \qquad \forall x \in \mathbb{X} \tag{9.19}$$

and all axioms of a vector space are satisfied. Taking the special case $\mathbb{Y} = \mathbb{K}$ with the canonical metric we arrive at $B(\mathbb{X},\mathbb{K})$, the space of the linear, bounded functionals on the vector space $\mathbb{X}$.

A final example we only touch upon is the Fourier series. We can consider (3.6) as a map of a function $f(x)$ (satisfying the Dirichlet conditions) to a set of numbers a_n, b_n, not dissimilar to ℓ^2. It is clear that this map is linear. What is not so clear at the moment is what precisely are the vector spaces and the norm. Looking at Parseval's identity, (3.14), is a bit of a give-away. But it also shows we actually really would like to have a scalar product which sadly is still missing.

Before we rectify the situation regarding the scalar product, let us make acquaintance with the *dual space [dualer Raum]*. With a bit of luck you have come across this already in linear algebra. In fact, this space is nothing but $B(\mathbb{X},\mathbb{K})$, mentioned above. What is remarkable is that this space is a Banach space (i.e. complete) independent of whether $\mathbb{X}$ is complete or not. Since this is a particularly important space, it has got a special notation. If $M_X = (\mathbb{X}, \|\,.\,\|)$ is a normed (real or complex) space, its dual space, the space of all continuous, linear (real or complex) functionals on M_X is denoted by $\tilde{M}_X$, and $\tilde{M}_X$ is a Banach space.

9.3 Hilbert spaces

Let $\mathbb{X}$ be a vector space with $x, y, z \in \mathbb{X}$ and $k_i \in \mathbb{K}$. We define a *scalar product* or *inner product [Skalarprodukt oder inneres Produkt]* as a binary operation whereby to every pair of vectors in $\mathbb{X}$ a number in $\mathbb{K}$ is associated. This number is written sometimes as $x \cdot y$ or (x, y) but we will us the Dirac notation and write $\langle x|y\rangle$. To be a scalar product the following requirements must be met $\forall x, y, z \in \mathbb{X}$ and $\forall k_i \in \mathbb{K}$:

$$\langle x|k_1 y + k_2 z\rangle = k_1\langle x|y\rangle + k_2\langle x|z\rangle \tag{9.20a}$$

$$\langle x|y\rangle = \langle y|x\rangle^* \tag{9.20b}$$

$$\langle x|x\rangle \geq 0 \qquad \text{and} \qquad \langle x|x\rangle = 0 \;\Leftrightarrow\; x = 0 \tag{9.20c}$$

We have chosen the scalar product to be linear in the second argument (sometimes it is chosen to be linear in the first argument) which means it is anti linear in the first argument, i.e. $\langle k_1 x|y\rangle = k_1^*\langle x|y\rangle$. Of course, for $\mathbb{K} = \mathbb{R}$ the complex conjugate has no effect.

In MMP I we have often spoken about a scalar product without actually defining what it is, see e.g. (2.5), (3.2), (5.7). Now, here it is, the definition, in its full glory. And we will check it was justified to call the above examples scalar product.

Before we do this, we remark that if we have a scalar product, we have in fact a norm and, hence, a metric. Indeed

$$\|x\| = \sqrt{\langle x|x\rangle} \qquad \text{or} \qquad d(x,y) = \|x-y\| = \sqrt{\langle x-y|x-y\rangle} \tag{9.21}$$

Thus, also for vector spaces with a scalar product we have automatically all concepts of a metric space, in particular the notion of completeness. A vector space with a scalar product that is complete under the norm/metric induced by this scalar product is called a *Hilbert space [Hilbertraum]*. If we just have a scalar product and are not sure yet whether or not the space is complete, we have a *pre-Hilbert space [Prähilbertraum]*. As for metric and Banach spaces, we can always find a (up to isomorphism unique) completion of a pre-Hilbert space. Thus, even if we have only a pre-Hilbert space $\mathbb{Y}$, we know we can find a Hilbert space $\mathbb{X}$ such that $\bar{\mathbb{Y}} = \mathbb{X}$ and if there was another such Hilbert space $\mathbb{X}_2$ then $\mathbb{X} \cong \mathbb{X}_2$.

We have just written this lightheartedly and without proof, but it will turn out to be quite important. Completeness is a crucial property. But we do not always give it the attention it deserves, simply because we always imply that we actually meant the completion.

Every Hilbert space is a Banach space. The converse does not hold. The norm induced through a scalar product satisfies the parallelogram equation ⇒ lecture

$$\|x+y\|^2 + \|x-y\|^2 = 2\big(\|x\|^2 + \|y\|^2\big) \tag{9.22}$$

A Banach space is a Hilbert space iff its norm satisfies (9.22).

With the scalar product we have the concept of angles for generic (pre-)Hilbert spaces. In particular we say x and y are *orthogonal [orthogonal]*

$$x \perp y \qquad \Leftrightarrow \qquad \langle x|y\rangle = 0 \tag{9.23}$$

As a final general comment before we turn to examples we mention the

$$\text{Schwarz inequality} \qquad |\langle x|y\rangle| \leq \|x\|\,\|y\| \tag{9.24}$$

and equality holds iff x and y are linearly dependent. The proof can be given in two lines. ⇒ lecture

Our first example for a Hilbert space is $\mathbb{C}^n$, or $\mathbb{R}^n$. For $x = (x_1 \dots x_n)$ and $y = (y_1 \dots y_n)$ the scalar product is the well known

$$\langle x|y\rangle = \sum_{i=1}^{n} x_i^*\, y_i = \sum_{i,j=1}^{n} x_i^*\, A_{ij}\, y_j \tag{9.25}$$

where A_{ij} is the $n\times n$ unit matrix. In fact, we could take any positive $A_{ij} = A_{ji}^*$ and still have a scalar product. We will see shortly that this is reminiscent of (5.52), where we had the normal scalar product generalized with an arbitrary but positive weight function. Another recurring remark related to (9.25) is that in special relativity people take $A_{ij} = \mathrm{diag}(1,-1,-1,-1)$ (and denote it by $g_{\mu\nu}$) to define the Minkowski scalar product. This is not a positive matrix and, hence, the Minkowski thingymabob might be whatever it is, but certainly not a scalar

product in the sense we have defined it here. It violates e.g. (9.20c). This is related to the remark in Section 9.1 about the metric in special relativity.

As a second example we consider $\mathbb{X} = \ell^2$, i.e. $x = (x_1, x_2 \ldots)$ with the scalar product

$$\langle x|y\rangle = \sum_{i=1}^{\infty} x_i^* \, y_i \tag{9.26}$$

The sum in (9.26) converges since $x, y \in \ell^2$. The norm that is induced by (9.26) is given in (9.12) with $p = 2$. Note that ℓ^p with $p \neq 2$ is not a Hilbert space. The norm (9.12) does not satisfy the parallelogram equation (9.22) unless $p = 2$. Thus, e.g. ℓ^3 is a case of a Banach space that is not a Hilbert space. But ℓ^2 is a Hilbert space, one that will be showing up regularly in physical applications.

The next space we consider – again – is $C[a, b]$. We know it is a normed space. But can we define a scalar product? With the norm (9.13a) $C[a, b]$ is even complete, but sadly, this norm does not satisfy (9.22). This is why we are not too keen on $(C[a, b], \| \, . \, \|_m)$. However, (9.13c) does satisfy (9.22) and, hence, we can get a scalar product out of it, namely

$$\langle x|y\rangle = \int_a^b x^*(t) y(t) \, dt \tag{9.27}$$

In view of (2.5), (3.2), (5.7) you should not be terribly surprised by (9.27) and this also explains why we prefer $(C[a, b], \| \, . \, \|_2)$ over the other normed spaces in (9.13). In fact, very much like we could generalise (9.25) by taking a non-trivial (positive!) matrix A_{ij}, we can also generalise (9.27) by adding a non-trivial (positive!) weight function $w(t)$ to the definition of the scalar product. Now why would anyone want to do this? Hint, hint, (5.52). For the time being, we will however work with (9.27).

Sadly though, this vector space is not complete, as can be seen from the example depicted in Figure 9.2. We have mentioned several times that this is not such a big deal, as we can simply find a completion. Doing this brings us to the mother of all Hilbert spaces, $L^2[a, b]$.

So, what is $L^2[a, b]$. First, $[a, b]$ is the same as for $C[a, b]$, a real interval. We will now explicitly allow for $a = -\infty$ and/or $b = \infty$. In doing so, it is not clear at all any longer, that (9.27) or the corresponding norm

$$\|x\| = \int_a^b |x(t)|^2 dt \tag{9.28}$$

exists. By definition and as indicated by the superscript 2 in L^2, elements of this space are square integrable, i.e. (9.28) exists for all $x \in L^2[a, b]$. Finally the L. This stands for Lebesgue or Lebesgue integrable. This is a generalisation of the Riemann integration and allows to integrate a larger class of (rather obscure) functions. This does not have to concern us greatly and we will just call $L^2[a, b]$ the space of square integrable functions in the (possibly infinite) interval $[a, b]$. Actually, elements of $L^2[a, b]$ are not really functions with a finite norm (9.28), but rather equivalence classes of functions with a finite norm (9.28). What this means is that two functions $x_1(t)$ and $x_2(t)$ are identified to be in the same class if $x_1(t) = x_2(t)$ nearly everywhere, i.e. everywhere up to a set of measure zero. ⇒ lecture

From the example in Figure 9.2 it is clear that to have any chance to get a complete space we have to allow for non-continuous functions. But if we have a $x_1(t)$ and and $x_2(t)$ that

differs from $x_1(t)$ only at a single point $a \leq t_0 \leq b$, it is clear that all integrals (9.27) and (9.28) will not be affected if we replace $x_1(t)$ by $x_2(t)$. Thus x_1 and x_2 are 'the same', or more accurately belong to the same equivalence class. Now, why stop at a difference at one point. As long as the measure of the set on which x_1 and x_2 differ is zero, the argument can be repeated. All these infinitely many functions belong to the same class, each is a representative of its class. It should be obvious that two continuous functions cannot belong to the same class, unless they are identical.

Thus, we define $L^2[a,b]$ as the space (of equivalence classes) of functions that are square (Lebesgue) integrable in the interval $[a,b]$. Which of the representative functions out of the class we take to actually compute the norm or a scalar product does not matter. We always get the same answer, by construction. Typically, if the class contains a continuous representative, why not take this one. It helps to draw lovely pictures.

Coming back to our example in Figure 9.2, how does this all help? The sequence considered there is still a Cauchy sequence, so it better converges. For all i the function x_i is continuous, hence the only continuous representative of the class. The limit is the function $x(t) = 0$ for $a \leq t < b$ and $x(b) = 1$. But this function is 'the same' as the function $x(t) = 0$ for $a \leq t \leq b$. Thus, the elements of $L^2[a,b]$ that are represented (e.g.) by the continuous functions $x_i(t)$ converges to the element of $L^2[a,b]$ that is represented by the continuous function $x(t) = 0$, i.e. the null element of $L^2[a,b]$. Thus, the limit is in $L^2[a,b]$. Obviously, this is far from a proof that $L^2[a,b]$ is complete, but it should illustrate the basic idea how taking classes of functions as elements, we can solve the problem of having holes in the vector space.

With this in mind, it should also be clear that strictly speaking, for a $x \in L^2[a,b]$ it does not make sense to ask what is e.g. $x(t_0)$, or if $x(t)$ is differentiable. The value of x at a particular point t_0 is arbitrary. Still, such statements are often made and they typically imply that there is a (single!) continuous representative of the element $x \in L^2[a,b]$ and for this $x(t_0)$ takes a certain value.

This brings us to the question if for all $x \in L^2[a,b]$ we have a continuous representative. The precise answer to this question is given as follows: The space $C[a,b] \cap L^2[a,b]$, i.e. the space of square integrable continuous functions is dense in $L^2[a,b]$. Thus every element of $L^2[a,b]$ can be approximated arbitrarily well by a continuous function. Approximated refers to the norm (9.28), thus $\forall x \in L^2[a,b]$ and $\forall \epsilon > 0$ we can find a $y \in C[a,b]$ such that $\|x - y\| < \epsilon$.

There are other spaces of 'reasonable' functions that are dense in $L[a,b]$, such as $C_c^n[a,b]$ or even $C_c^\infty[a,b]$. As explained earlier these are the n times (or infinitely often) continuously differentiable functions. The subscript c indicates the functions have a compact support. This is needed since $a = -\infty$ and/or $b = \infty$ is allowed. Such functions are automatically square integrable.

This might sound very surprising. How can it be that any wild element in $L^2[a,b]$ can be approximated arbitrarily well by a super smooth function in $C_c^\infty[a,b]$. To get an idea how this is possible let us look at the functions

$$x(t) = \begin{cases} 0 & |t| \geq 1 \\ 1 & |t| < 1 \end{cases} \qquad \text{and} \qquad x_c(t) = \begin{cases} 0 & |t| \geq 1 \\ \exp\left(c - \frac{c}{1-t^2}\right) & |t| < 1 \end{cases} \tag{9.29}$$

Clearly $x(t) \in L^2[\mathbb{R}]$ but $x(t) \notin C[\mathbb{R}]$. However, for $c > 0$ we have $x_c(t) \in C_c^\infty[\mathbb{R}]$. Indeed,

the function $x_c(t)$ satisfies

$$\lim_{|t|\to 1} \frac{d^n x_c(t)}{dt^n} = 0 \qquad \forall n \in \mathbb{N}_0 \quad \text{if} \quad c > 0 \tag{9.30}$$

where the limit can be taken from both sides, $|t| > 1$ and $|t| < 1$. The function $x(t)$ (grey) and its approximations $x_c(t)$ (dashed) for ever smaller values of $c > 0$ are depicted in Figure 9.3. The smooth limit of $x_c(t = 1)$ for $|t| < 1$ can best be seen for $c = 1$ (blue). For $c \to 0$ the $x_c \in C_c^\infty[\mathbb{R}]$ approximate $x(t)$ arbitrarily well, in that $\|x - x_c\| \to 0$ for $c \to 0$. More complicated functions in $L^2[\mathbb{R}]$ can be treated similarly, piece wise for different ranges in t.

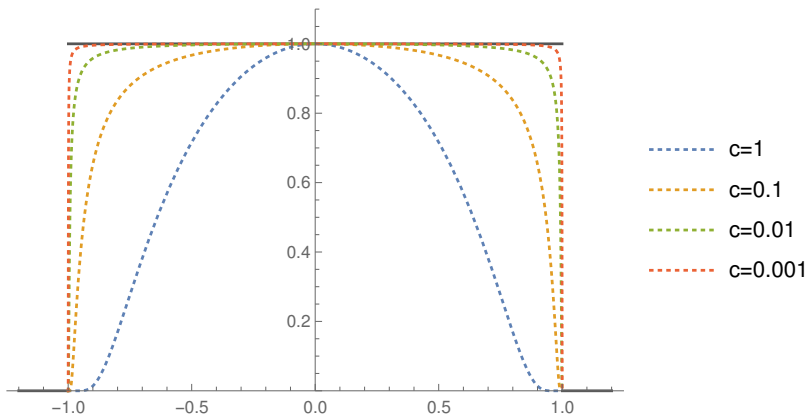

Figure 9.3: Function $x(t) \in L^2[\mathbb{R}]$ with limiting approximations $x_c(t) \in C_c^\infty[\mathbb{R}]$.

What we have achieved now is to have a Hilbert space (essentially of) of square integrable functions. This has far reaching consequences of physics. As the most prominent example we mention the wave function, as the solution to a Schrödinger equaton. This is nothing but a vector in $L^2[a, b]$ and the existence of the norm (9.28) allows for a probabilistic interpretation of the absolute value squared of the wave function, (7.63). The differential operators appearing in the Schrödinger equation, e.g. in (6.2g) are linear operators in the Hilbert space $L^2[a, b]$. Thus, we have to get to know and love linear operators.

Before we do this in Section 10 let us briefly comment on the notation $|x\rangle \in \mathbb{X}$ for a vector. This is called a ket. Acting with an operator on it we write $T|x\rangle = |y\rangle \in \mathbb{X}$, and get again a ket (i.e. another vector). We can also 'act' with $\langle y|$ on $|x\rangle$ and get $\langle y|x\rangle \in \mathbb{K}$, i.e. a number. Thus $\langle y|$ is a linear functional, linear due to (9.20a). In other words, $\langle y| \in \tilde{\mathbb{X}}$ is an element of the dual space of $\mathbb{X}$. If $\mathbb{X} = H$ is a Hilbert space, through the scalar product, $\forall\, y \in H$ we can define a $\langle y| \in \tilde{H}$. We call $\langle y|$ a bra, such that $\langle y|x\rangle$ is a bra-ket, i.e. a bracket. Yes, this happens when physicists make jokes.

10 Linear operators in Hilbert spaces

In this section we study operators on a real or complex Hilbert space H, where we include the possibility of infinite dimensions. When we say operator, we always mean a linear operator.

To start with, we have a recap of finite dimensional Hilbert spaces and recall the spectral representation for self-adjoint and unitary operators. The spectral representation allows e.g. to trivially work out the inverse of an operator and, hence, solve systems of equations. We also would like to do this in infinite-dimensional Hilbert spaces. PDE such as (7.37) for example or ODE as alluded to in Section 5.7 can be seen as an equation of a (differential) operator acting on a vector in an infinite-dimensional Hilbert space. Solving these differential equations is inverting the (differential) operator. Thus, the rest of this section will then be devoted to see how many of these nice properties in finite-dimensional Hilbert spaces we can salvage in infinite-dimensional Hilbert spaces. At the very end of this section we will apply the concepts we get familiar with to give a very basic outlook on the mathematical structure of quantum mechanics.

10.1 Finite-dimensional Hilbert spaces

Let $\dim H = n < \infty$ with $|y\rangle \in H$ and $\langle x| \in \tilde{H}$. The scalar product $\langle x|y\rangle$ is a linear and continuous functional and in fact every linear, continuous functional on H can be expressed in this way (Riesz representation theorem) and we have $H \cong \tilde{H}$.

We choose a orthonormal basis (ONB) $\{|\varphi_i\rangle\}$ with $1 \leq i \leq n$, satisfying $\langle\varphi_i|\varphi_j\rangle = \delta_{ij}$. The completeness relation then reads

$$\mathbb{I} = \sum_{i=1}^{n} |\varphi_i\rangle\langle\varphi_i| \qquad \Rightarrow \qquad |x\rangle = \mathbb{I}|x\rangle = \sum_{i=1}^{n} |\varphi_i\rangle\langle\varphi_i|x\rangle = \sum_{i=1}^{n} c_i|\varphi_i\rangle \tag{10.1}$$

where $\mathbb{I}$ is the unit operator. This is just the decomposition of a vector in a basis, very much like (3.1), just with a slightly different notation. The vector $|x\rangle$ can now be identified with $(c_1 \dots c_n) \in \mathbb{K}^n$ and every n dimensional Hilbert space is isomorphic to $\mathbb{K}^n$, i.e. $H \cong \mathbb{K}^n$ with $|x\rangle \cong (c_1 \dots c_n)$.

A (linear) operator T acting on $|x\rangle$

$$T|x\rangle = T\left(\sum_{j=1}^{n} c_j|\varphi_j\rangle\right) = \sum_{j=1}^{n} c_j T|\varphi_j\rangle = \sum_{i,j=1}^{n} T_{ij}c_j|\varphi_i\rangle \tag{10.2}$$

where we used that $T|\varphi_j\rangle \in H$, hence it can be expanded as in (10.1) and be written as $T|\varphi_j\rangle = \sum_{i=1}^{n} T_{ij}|\varphi_i\rangle$. Thus, once a basis has been chosen there is a one-to-one relation between the operator T and the $n \times n$ matrix T_{ij}. Or in other words, the space $B(H, H)$ can be represented as the space of $n \times n$ matrices. The operators are automatically bounded (hence continuous) with $\|T\| \leq \max |T_{ij}|$ and the operator T is defined everywhere, $\mathcal{D}_T = H$. The product of two operators ST is represented by the matrix $(ST)_{ij} = \sum_{k=1}^{n} S_{ik}T_{kj}$.

If we find a $|x\rangle \neq 0$ with $T|x\rangle = \lambda|x\rangle$ for $\lambda \in \mathbb{K}$ we call $|x\rangle$ an *eigenvector* (EV) *[Eigenvektor]* and λ and *eigenvalue* (EVal) *[Eigenwert]* of T. As is familiar from linear algebra, this means $\det(T - \lambda\mathbb{I}) = 0$. The solution of the corresponding equation has at most n different solutions λ_i with $i < n$. It is possible that some EVal are *degenerate [entartet]*, i.e. $\lambda_i = \lambda_j$ for $i \neq j$.

The inverse operator to T if it exists (!!) is defined through

$$T|x\rangle = |y\rangle \qquad \Rightarrow \qquad T^{-1}|y\rangle = |x\rangle \tag{10.3}$$

For T^{-1} to exist we need $\det T \neq 0$. If $\exists T^{-1}$ and $\exists S^{-1}$ we have $(S\,T)^{-1} = T^{-1}S^{-1}$. The inverse operator is the solution to the system of equations given by the l.h.s. of (10.3). We can extend this to an inhomogeneous operator equation (or system of equations) as

$$(T - k\mathbb{I})|x\rangle = |y\rangle \qquad \Rightarrow \qquad R_T(k)|y\rangle \equiv (T-k)^{-1}|y\rangle = |x\rangle \tag{10.4}$$

where $R_T(k)$ (do not confuse with the range $\mathcal{R}_T$) is the *resolvent [Resolvente]* for $k \in \mathbb{K}$. Note we have started to drop the unit operator in the notation, i.e. $k\mathbb{I} = k$. Iff k is not an EVal, the resolvent exists, is linear, and is everywhere defined, $\mathcal{D}_{R_T(k)} = H$.

The *adjoint [adjungiert]* operator $T^\dagger$ is defined through

$$\langle y|Tx\rangle = \langle T^\dagger y|x\rangle \equiv \langle y|T|x\rangle \qquad \forall x, y \in H \tag{10.5}$$

where we have introduced a common notation in the last step. Since T is linear, so is $T^\dagger$ and $(T^\dagger)^{-1} = (T^{-1})^\dagger$ and $(S\,T)^\dagger = T^\dagger S^\dagger$. If T is represented by the matrix T_{ij} the operator $T^\dagger$ is represented by the matrix T^*_{ji}. An operator is called *symmetric [symmetrisch]* if $T = T^\dagger$. Since we have anyway $\mathcal{D}_T = \mathcal{D}_{T^\dagger} = H$ in $\mathbb{K}^n$ the equality $T = T^\dagger$ is actually complete, including the domain and we say T is *self-adjoint [selbstadjungiert]*. The corresponding matrix is *hermitean [hermitesch]* $T_{ij} = T^*_{ji}$. Self-adjoint operators rule the world!

An operator U is called *unitary [unitär]* if $U^\dagger = U^{-1}$ and *normal [normal]* if $T^\dagger\,T = T\,T^\dagger$. Note that $S = T^\dagger\,T$ is self-adjoint. Working in real Hilbert spaces the term *orthogonal [orthogonal]* is used instead of unitary. The reason is that under an unitary/orthogonal map the scalar product is invariant

$$U \text{ unitary:} \qquad \langle Uy|Ux\rangle = \langle U^\dagger Uy|x\rangle = \langle y|x\rangle \tag{10.6}$$

and orthogonal vectors remain orthogonal. Unitary operators rule the world too!

For a self-adjoint operator $T = T^\dagger$ we can find a ONB $\{|\chi_i\rangle\}$ such that $T|\chi_i\rangle = \lambda_i|\chi_i\rangle$ and the EVal $\lambda_i \in \mathbb{R}$, even if we work in a complex Hilbert space. In this "eigenbasis" T is represented by the diagonal matrix with the EVal λ_i as entries. Of course, it is possible that some EVal are degenerate. Let us assume we have k different EVal and the EVal λ_i is i_d fold degenerate. We have $\sum_{i=1}^k i_d = n$. We can then define the operators

$$P_{\lambda_i} = \sum_{i_j=1}^{i_d} |\chi_{i_j}\rangle\langle\chi_{i_j}| \qquad \text{with} \qquad P^2 = P \qquad \text{and} \qquad P^\dagger = P \tag{10.7}$$

and the sum is over all i_d linearly independent (actually orthogonal) EV $|\chi_{i_j}\rangle$ with EVal λ_i. If λ_i is not degenerate there is only one term in the sum. The operator P in (10.7) is a *projector [Projektor]* onto the eigenspace of the EVal λ_i. We can thus write the *spectral representation [Spektraldarstellung]* of T as

$$T = \sum_i^k \lambda_i P_{\lambda_i} \tag{10.8}$$

where the sum is over all distinct EVal λ_i. Note that we have $P_i\, P_j = 0$, EV to different EVal are orthogonal. We can split the Hilbert space into a direct sum

$$H = H_1 \oplus H_2 \oplus \ldots \oplus H_k \tag{10.9}$$

where H_i is the i_d dimensional eigenspace of λ_i. This means that any vector in H can be uniquely decomposed as $|x\rangle = \sum_{i=1}^{k} c_i |x_i\rangle$ and $|x_i\rangle = P_i|x\rangle \in H_i$.

The spectral representation (10.8) is immensely useful since this allows to invert T on the blink of an eye, or actually do whatever we want with T in an instant. Any function f of the operator T can be computed (if it exists) as

$$f(T) = \sum_{i}^{k} f(\lambda_i) P_{\lambda_i}. \tag{10.10}$$

Thus if f is defined for the EVal, it is defined for the full operator. We can compute e^T, $\sin T$, maybe even $\log T$ or $\sqrt{T}$.

Lets assume we have $S = S^\dagger$ and $T = T^\dagger$. As we have seen it is a very good idea to choose a basis of EV of T, i.e. a basis where T is diagonal. Can we choose a basis s.t. T and S are diagonal? The answer is yes, iff $[T, S] = 0$, i.e. iff the two operators commute. The commutator has been defined in (7.79). The connection between commutation of operators and the possibility to choose a simultaneous eigenbasis has a deep meaning in quantum mechanics, related to the uncertainty principle. ⇒ lecture

So, a self-adjoint operator $T = T^\dagger$ can be tamed by finding all EVal and EV and work in the corresponding basis, or put differently, use the spectral representation. In fact, we can do the same for a unitary operator $U^\dagger = U^{-1}$. Whereas for a self-adjoint operator all EVal are real, for a unitary operator all EVal satisfy $|\lambda_i| = 1$. The rest is pretty much the same and we can also obtain a spectral representation, very much like (10.8). If $T = T^\dagger$ then e^{iT} is unitary. We will need this badly later.

Now, what if the operator is neither unitary nor self-adjoint. Well, we might be out of luck. It is still possible (but not guaranteed) that we can find n EV with their EVal (i.e. T is diagonable) but even if we do, these EV will not be orthogonal and will not form a ONB. Luckily, in what follows we will mostly deal with self-adjoint or unitary operators.

10.2 Direct sum, projectors and ONB

Yes, all the words in the title of this subsection we have just heard in Section 10.1. But now we mean business and will not any longer assume H is finite dimensional.

Let Y be a linear subspace of H, i.e. $|y_1\rangle, |y_2\rangle \in Y \;\Rightarrow\; k_1|y_1\rangle + k_2|y_2\rangle \in Y$. If Y is closed then $Y \subset H$ is itself a Hilbert space, with the same scalar product as H. The orthogonal complement of Y is defined as $Y^\perp = \{z \in H | z \perp Y\}$, i.e. it contains all $|z\rangle \in H$ such that $\langle z|y\rangle = 0,\ \forall |y\rangle \in Y$. We have $Y \cap Y^\perp = \{0\}$ and ⇒ lecture

$$H = Y \oplus Y^\perp \tag{10.11}$$

To prove (10.11) we have to show that $\forall x \in H$ there $\exists! y \in Y$ and $\exists! z \in Y^\perp$ s.t. $x = y + z$. This proof relies on the fact that for any $x \in H$ we can find a unique $y \in Y$ such that $\|x - y\|$ is minimal. For this to hold it is essential that Y is closed. As illustrated in Figure 10.1 if Y ⇒ lecture

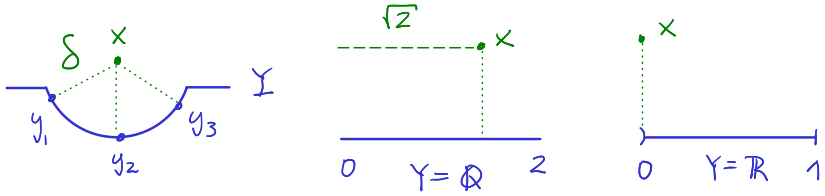

Figure 10.1: Left panel: infinitely many $y \in Y$ have minimal distance δ from $x \in H = \mathbb{R}^2$. Middle and right panel: no point $y \in Y = [0,2] \cap \mathbb{Q}$ or $y \in Y = \{y | y \in \mathbb{R}, 0 < y \leq 1\}$ has minimal distance δ from $x \in H = \mathbb{R}^2$. In all cases Y is not a Hilbert space.

is not a Hilbert space itself we might find infinitely many or no $y \in Y$ for a given $x \in H$ s.t. $\|x - y\|$ is minimal.

With this we can define the projection of H onto Y, $P_Y : H \mapsto Y$ as $P_Y|x\rangle = |y\rangle$. This operator P_Y is linear, bounded $\|P_Y\| = 1$, satisfies $P_Y^2 = P_Y$ (called idempotent) and as we will see later, self-adjoint. Since we have $P_Y Y = Y$ and $P_Y Y^\perp = \{0\}$ we see that P_Y is not injective, hence P_Y^{-1} does not exist. (10.11) is a first step towards generalising (10.9) to arbitrary Hilbert spaces. In fact, we can repeat this step and arrive at

$$H = H_1 \oplus H_2 \oplus \ldots \oplus H_k = \bigoplus_{i=1}^{k} H_i \qquad \text{with} \qquad \sum_{i=1}^{k} \dim H_i = \dim H \tag{10.12}$$

Very much like P_Y above was the projection onto the subspace Y, related to (10.12) we have the projectors P_i projecting onto H_i. Also $P_j|x\rangle \in H_i^\perp$ for $i \neq j$. These projectors satisfy $P_i^2 = P_i$, $P_i P_j = 0$ for $i \neq j$ and

$$\mathbb{I} = \sum_{i=1}^{k} P_i \tag{10.13}$$

This should be compared to (10.8). For $T = \mathbb{I}$ we have $\lambda_i = 1$ and the only difference is that in (10.8) the P_{λ_i} all project onto finite-dimensional Hilbert spaces, whereas in (10.13) some projectors P_i might project onto infinite dimensional Hilbert spaces.

The last relation of writing the identity as a sum of projectors is a first step towards constructing a ONB. Let $\{\varphi_i\}$ be an orthonormal system of vectors, with $i \in I$ where I is for the time being an arbitrary index set. Thus we have $\langle \varphi_j | \varphi_i \rangle = \delta_{ij}$. In fact, whenever we have a set of linearly independent vectors, we can use the Gram-Schmidt procedure familiar from linear algebra to construct an orthonormal system out of it. Next we define the *span [lineare Hülle]* of $\{\varphi_i\}$ as ⇒ lecture

$$\operatorname{span}\big(\{\varphi_i\}\big) = \{x | x = \sum_{i \in I} k_i |\varphi_i\rangle\} \tag{10.14}$$

i.e. the set of all linear combinations that can be constructed out of the vectors $\{\varphi_i\}$. We call $\{\varphi_i\}$ a complete orthonormal system if $\overline{\operatorname{span}}(\{\varphi_i\}) = H$, i.e. if the span of it is dense in H. This also implies that for $\forall\, y \in H$ the statement $y \in \operatorname{span}(\{\varphi_i\})^\perp$ entails $|y\rangle = |0\rangle$. So

far, I was an arbitrary index set. Iff I is countable then so is $\mathrm{span}(\{\varphi_i\})$. Thus, we have a countable set that is dense in H, in other words H is separable.

In what follows we will restrict ourselves to separable Hilbert spaces and we assume infinite dimensions (the finite-dimensional case is trivially included). These are precisely the Hilbert spaces for which there exists a countable orthonormal system $\{\varphi_i\}$ now with $i \in \{1, 2, \ldots\}$. In this case for each $|x\rangle \in H$ we can write

$$|x\rangle = \sum_{i=1}^{\infty} x_i |\varphi_i\rangle \tag{10.15}$$

precisely in the sense given in (9.15) in that the partial sums $|x^N\rangle \equiv \sum_{i=1}^{N} x_i|\varphi_i\rangle$ converge $\lim_{N\to\infty} |x^N\rangle = |x\rangle$ or equivalently $\lim_{N\to\infty} \|x^N - x\| \to 0$. Thus, we call such a system $\{\varphi_i\}$ a basis, more precisely a ONB of H. Furthermore

$$\langle\varphi_j|x\rangle = \langle\varphi_j| \sum_{i=1}^{\infty} x_i|\varphi_i\rangle = \sum_{i=1}^{\infty} x_i \langle\varphi_j|\varphi_i\rangle = x_j \tag{10.16}$$

and, combining with (10.15) we find

$$|x\rangle = \sum_{i=1}^{\infty} |\varphi_i\rangle\langle\varphi_i|x\rangle \qquad \text{or} \qquad \mathbb{I} = \sum_{i=1}^{\infty} P_{\varphi_i} \equiv \sum_{i=1}^{\infty} |\varphi_i\rangle\langle\varphi_i| \tag{10.17}$$

The last equation is called the decomposition of the identity and is a generalisation of (10.13) in that each projector P_{φ_i} projects onto the one-dimensional Hilbert space $\mathrm{span}(\varphi_i)$. It looks very much like the first equation in (10.1) with an innocuous looking modification $n \to \infty$. But this is not innocuous at all. Now we have an infinite sum of operators $|\varphi_i\rangle\langle\varphi_i|$. What the heck is that supposed to be? Does this sum converge? In what sense? To what? While we will be looking at infinite sums of operators and limits of operators more generally only in Section 10.7, for the particular case of (10.17) all these questions have been answered. This was the whole point of the prelude to (10.17). But it is important to realise that these questions must be answered. Simply replacing $n \to \infty$ in (10.1) might qualify as artwork, but has certainly nothing to do with mathematics.

With the help of (10.15) we find

$$\langle x|x\rangle = \|x\|^2 = \sum_{i,j=1}^{\infty} x_j^* \langle\varphi_j|\varphi_i\rangle x_i = \sum_{i=1}^{\infty} |x_i|^2 \tag{10.18}$$

which is nothing but Parseval's identity that we have encountered already in connection with the Fourier series, (3.14). From (10.18) we can also deduce that the partial sums $|x^N\rangle$ form a Cauchy sequence since $\|x^N - x^M\| = \sum_{i=N+1}^{M} |x_i|^2$ must tend to zero for $M > N \to \infty$. Since we are in a Hilbert space, i.e. a complete space, this Cauchy sequence converges and, in fact, it converges to x in (10.15). ⇒ lecture

So let us look at some examples of infinite-dimensional separable Hilbert spaces.

As a first example we take $L^2[-1, 1]$. We start with the set $\{\eta_i\}$ where $|\eta_i\rangle \sim \eta_i(t) = t^i$ with $-1 \leq t \leq 1$ and $i \in \mathbb{N}_0$. These vectors are linearly independent and $\mathrm{span}(\{\eta_i\})$ is the set of polynomials in $[-1, 1]$. The set of polynomials can be shown to be dense in $C[-1, 1]$ and we argued in Section 9.3 that $C[-1, 1]$ is dense in $L^2[-1, 1]$. Hence $\{\eta_i\}$ is complete in

that $\overline{\text{span}}(\{\eta_i\}) = L^2[-1,1]$. However, the $\{\eta_i\}$ are not orthonormal. Thus, we use the Gram-Schmidt procedure to transform $\{\eta_i\}$ to a ONB. What do we get? The Legendre polynomials (up to normalisation). Thus $\forall\, |x\rangle \sim x(t) \in L^2[-1,1]$ there $\exists!\{x_i\}$ s.t. $|x\rangle = \sum_{i=1}^{\infty} x_i|\varphi_i\rangle$ as in (10.15), where $|\varphi_i\rangle \sim \varphi_i(t) = P_i(t)$ (with a most unfortunate clash in notation, as here $P_i(t)$ are the Legendre polynomials of Section 5.1 and not projectors). In fact, if you compare with (5.6) you'll notice that we have to take $\sqrt{(2l+1)/2}\, P_l(t)$ as a basis and you also notice that I lied there when stating just before (5.7) that this is a basis in the vector space of continuous functions. Sorry. It is a basis in the Hilbert space $L^2[-1,1]$ and the continuous functions just are dense in $L^2[-1,1]$. But the two relations in (5.8) are just one concrete (and very useful) example of (10.15) and (10.16). And now we even understand what the = sign in the first equation of (5.8) actually means. ⇒ lecture

Buoyed by this success we move on to $L^2[-\infty,\infty] = L^2[\mathbb{R}]$. Here we begin with the set $\{\eta_i\}$, where $|\eta_0\rangle \sim \eta_0(t) = e^{-t^2/2}$, and we define recursively $\eta_i(t) = t\,\eta_{i-1}(t)$. Thanks to the exponential damping we ensure the convergence in (9.28), i.e. these $|\eta_i\rangle \in L^2[\mathbb{R}]$. Again, they are linearly independent but not an orthonormal system. Using Gram-Schmidt again, we get ... drum rolls ... the Hermite polynomials of Section 5.4, up to a normalisation, (5.30b). Look back at (7.69). Is it not blatantly obvious now that this is just yet another version of (10.15) and that we can do this for any $\Psi(x,0) \in L^2[\mathbb{R}]$ (see also the one-dimensional version of (7.63), yes, the one-dimensional wave function is just a vector in the Hilbert space $L^2[\mathbb{R}]$).

We're on a roll. Next, take $L^2[0,\infty]$. Start with $|\eta_0\rangle \sim \eta_0(t) = e^{-t/2}$ and define again recursively $\eta_i(t) = t\,\eta_{i-1}(t)$. The usual story, these $\{\eta_i\}$ are linearly independent, using Gram-Schmidt we get an ONB. The resulting basis vectors are $\varphi_i(t) = e^{-t/2}\,L_i(t)$ with $L_i(t)$ the Laguerre polynomials of Section 5.3. You should not be surprised any longer by (5.24b).

Let us now turn to a somewhat different example, ℓ^2. Here we can directly write down an ONB $\{\varphi_i\}$ with $|\varphi_i\rangle = (0\ldots1,0\ldots)$ where the 1 is at the ith place. With the scalar product (9.26) it is clear that they are orthonormal and it is also fairly obvious that the span of these vectors is dense in ℓ^2. ⇒ lecture

For any separable infinite-dimensional Hilbert space H we can use (10.15) to associate with each $x \in H$ a sequence $(x_1, x_2 \ldots)$ and using (10.18) we see this sequence is an element of ℓ^2. Thus we have a map of H to ℓ^2. This map is bijective and a homomorphism (preserves the structure) i.e. it is an isomorphism. Thus, we arrive at the statement that any separable infinite-dimensional Hilbert space $H \cong \ell^2$. ⇒ lecture

10.3 The adjoint operator

Even though we had stated at some point that the domain $\mathcal{D}_T$ is an essential part of the definition of an operator, we were often rather cavalier about it. However, in this subsection we need to be a bit more precise about this. We want to extend the definition of adjoint operators to (separable) infinite-dimensional Hilbert spaces and generalise the concepts of symmetric and self-adjoint.

Let us start with an operator $T : H_1 \mapsto H_2$ defined everywhere on a Hilbert space H_1, i.e. $\mathcal{D}_T = H_1$. In addition, let T be bounded. Then, $\forall\, y \in H_2$ we get a linear and continuous functional $f_y(x)$ on H_1 through $f_y(x) = \langle y|Tx\rangle$. Note $\langle y| \in \tilde{H}_2$ and $\langle\, y|T \in \tilde{H}_1$.

As a second ingredient we need the **Riesz representation theorem** which states that every linear and continuous functional f on H_1 has the form $\langle\tilde{y}|x\rangle$ and the vector $\tilde{y} \in H_1$ is

unique. We have already mentioned this theorem for finite-dimensional Hilbert spaces at the beginning of Section 10.1. It can also be proved for infinite-dimensional Hilbert spaces, with the additional assumption $\mathcal{D}_f = H_1$. Furthermore, we have $\|f\| = \|\tilde{y}\|$, i.e. the norm of the operator (functional) f is equal to the norm of the vector $|\tilde{y}\rangle \in H_1$ or $\langle\tilde{y}| \in \tilde{H}_1$. ⇒ lecture

Before we use Riesz to define the adjoint operator, let us pause and look back to Section 2 where we considered distributions. They were defined as linear and continuous functionals, i.e. elements of $B(H, \mathbb{K})$. In (2.5) we defined regular distributions and now recognize those as precisely the distributions covered by Riesz. The function $\tau(x)$ in (2.5) plays the role of $\tilde{y}$. What is remarkable is that according to Riesz, in a Hilbert space there are no other (non-regular) distributions, as every linear continuous functional can be written as a scalar product, i.e. as in (2.5). In relation to (2.7) we have stated that any distribution can be considered as a limiting case of $\tau_n(x) \to \tau(x)$. But if all the $\tau_n(x)$ are in a Hilbert space, so is $\tau(x)$, due to the completeness! It all fits together nicely. But let's try to trick Riesz. What about the δ distribution? This is a non-regular distribution that we have used repeatedly in MMP I. Can we not just use (2.9) and have our example of a non-regular distribution in a Hilbert space, getting the better of Riesz? No! The expression $\varphi(x_0)$ used in (2.9) referred to a test function $\varphi \in C_c^\infty[a, b]$, or at the very least to a continuous function. But $C[a, b]$ is not complete, Riesz does not apply. Riesz-us 1:0. And using the definition (2.9) for $\varphi(x) \in L^2[a, b]$? This makes no sense, as $\varphi(x_0)$ is not defined (in fact is arbitrary) for $\varphi(x) \in L^2[a, b]$ (recall elements in $L^2[a, b]$ are equivalence classes, not functions!). Riesz-us 2:0. Final whistle. Well, at least he didn't get a hattrick.

Now we are ready to define the to $T : H_1 \mapsto H_2$ adjoint operator $T^\dagger : \tilde{H}_2 \mapsto \tilde{H}_1$ as

$$\underbrace{\langle y|}_{\in\tilde{H}_2}\underbrace{Tx\rangle}_{\in H_2} \overset{\exists!\tilde{y}}{=} \underbrace{\langle T^\dagger y|}_{\langle\tilde{y}|\in\tilde{H}_1}\underbrace{x\rangle}_{\in H_1} \equiv \langle y|T|x\rangle \tag{10.19}$$

for all $x \in H_1$ and for all $y \in H_2$. Let's do this in words. First $T|x\rangle$ is defined $\forall x \in H_1$ since T is by assumption defined everywhere. Then, for any $y \in H_2$ and, hence any $\langle y| \in \tilde{H}_2$, we get a linear bounded functional on H_1 as $f_y(x) = \langle y|Tx\rangle$. According to Riesz, this functional has the form $\langle\tilde{y}|x\rangle$. Now we define $T^\dagger$ through $\langle T^\dagger y| = \langle\tilde{y}|$.

If we now focus on $T : H \mapsto H$ linear, bounded, and defined everywhere, the situation is as in the finite-dimensional case. The operator T is called symmetric iff $\langle y|Tx\rangle = \langle Ty|x\rangle$, $\forall\, x, y \in \mathcal{D}_T = H$. Then T is also self-adjoint since $T = T^\dagger$, including $\mathcal{D}_T = \mathcal{D}_{T^\dagger} = H$. Thus, for linear, bounded and everywhere defined operators self-adjoint and symmetric is the same thing.

Let us look at an example. We take $H = \ell^2$ with $x = (x_1, x_2 \ldots)$ and use the basis $|\varphi_i\rangle = (0 \ldots 1, 0 \ldots)$ introduced at the end of Section 10.2. We define the *raising operator [Aufsteigeoperator]* T_+ and the *lowering operator [Absteigeoperator]* T_- as follows:

$$T_+|x\rangle = (0, x_1, x_2 \ldots) \quad \text{or} \quad T_+|\varphi_i\rangle = |\varphi_{i+1}\rangle \tag{10.20a}$$

$$T_-|x\rangle = (x_2, x_3 \ldots) \quad \text{or} \quad T_-|\varphi_i\rangle = |\varphi_{i-1}\rangle \text{ for } i > 1 \text{ and } T_-|\varphi_1\rangle = 0 \tag{10.20b}$$

The operators in (10.20) are also often called *ladder operators [Leiteroperatoren]* and *creation [Erzeugungs]* and *annihilation [Vernichtungs]* operators. It should now be obvious that

$$\langle y|T_+x\rangle = 0 + y_2^* x_1 + y_3^* x_2 \ldots = \langle T_- y|x\rangle \tag{10.21}$$

and, hence, $T_+^\dagger = T_-$. As a side remark, why don't you replace T by L and check out (A.17). If you act with $L_\pm$ on $Y_l^{(m)}$ the names raising and lowering operators will make even more sense.

To close our deliberations on linear, everywhere defined operators T we mention the theorem of **Hellinger-Toeplitz** stating that if such an operator T is also symmetric, it has to be bounded. In fact, to ensure that a linear, everywhere defined operator T is bounded only a little bit more has to be asked. Insisting T is symmetric is one option. Or we can ask for T to be closed. We have not defined closed[4], but suffice to say that (nearly) all operators we encounter in 'everyday' physics are closed. Thus, as a physicist you may be forgiven to think a linear, everywhere defined operator is bounded, even if from a mathematical point of view, this is not true. ⇒ lecture

But now it is time to face the devil and consider an operator that is not necessarily bounded. To make this manifest, we will use another letter, $A : H \supseteq \mathcal{D}_A \mapsto H$, for such an operator. Still, we would like to define the adjoint of A, according to (10.19), i.e.

$$\forall\, x \in \mathcal{D}_A : \qquad \langle y|Ax\rangle = \langle A^\dagger y|x\rangle \tag{10.22}$$

Whatever $A^\dagger$ is, $A^\dagger y$ has to be uniquely defined $\forall\, y \in \mathcal{D}_{A^\dagger}$. But this means, $\mathcal{D}_A$ has to be dense in H. Otherwise we can find a $z \in \mathcal{D}_A^\perp$ with $z \neq 0$. Since for this z we have $\langle z|x\rangle = 0$, (10.22) also entails $\langle y|Ax\rangle = \langle A^\dagger y + z|x\rangle$ and (10.22) cannot be used to uniquely define $A^\dagger$. Thus from now on we consider $A : H \supseteq \mathcal{D}_A \mapsto H$ with $\overline{\mathcal{D}}_A = H$, i.e. operators that are *densely defined [dicht definiert]*. For such operators we can and will use (10.22) as the (unique!) definition of the adjoint operator. ⇒ lecture

Such an operator A is called symmetric, if $\forall\, x, y \in \mathcal{D}_A$ we have $\langle y|Ax\rangle = \langle Ay|x\rangle$. Symmetric means $A \subseteq A^\dagger$, i.e. $A^\dagger|x\rangle = A|x\rangle$ for all $x \in \mathcal{D}_A$, but $\mathcal{D}_{A^\dagger} \supseteq \mathcal{D}_A$. Such an operator A is called self-adjoint if $A = A^\dagger$, i.e. $A^\dagger|x\rangle = A|x\rangle$ for all $x \in \mathcal{D}_A = \mathcal{D}_{A^\dagger}$. If you do not care about the domain of an operator, you do not care about the difference between self-adjoint and symmetric. But if you do, a self-adjoint operator is symmetric, but a symmetric operator is not necessarily self-adjoint.

If A is a densely defined linear operator its adjoint operator $A^\dagger$ is also a linear operator. In addition if A^{-1} exists and is also densely defined, then $(A^\dagger)^{-1} = (A^{-1})^\dagger$, properties we have already mentioned after (10.5) for finite-dimensional Hilbert spaces. ⇒ lecture

It is time for a few examples.

In our first example $H = \ell^2$ and $A : \mathcal{D}_A \mapsto H$ is defined as $A|x\rangle = (x_1, 2x_2, 3x_3 \ldots)$ with $x = (x_1, x_2 \ldots)$. Of course $\mathcal{D}_A = \{x|Ax \in \ell^2\}$ and A is not bounded. The questions is what is $A^\dagger$? Think, think. The answer is $A^\dagger$ cannot be defined. Why? Because $\mathcal{D}_A$ is not dense in ℓ^2. Indeed, take e.g. $z = (1, 1/2, 1/3 \ldots) \in \ell^2$, but $Az = (1, 1 \ldots) \notin \ell^2$. Thus $z \notin \mathcal{D}_A$ and it should be obvious that $\exists B(z, \epsilon)$ with $B(z, \epsilon) \cap \mathcal{D}_A = \varnothing$.

The next examples are much closer to applications[5]. For $H = L^2[0, 1]$ with $|x\rangle \sim x(t)$ we define $A = i\, d/dt$ with $\mathcal{D}_A$ all elements in $L^2[0, 1]$ for which the derivative is also in $L^2[0, 1]$. We know $\overline{\mathcal{D}}_A = H$ since $C^n[0, 1]$ is dense in $L^2[0, 1]$. We also define $A_s = i\, d/dt$ with $\mathcal{D}_{A_s} = \mathcal{D}_A \cap \{x|x(0) = x(1)\}$. Thus the domain of A_s has the additional requirement

[4] Roughly speaking, an operator T is called closed, if for a a convergent sequence $x_n \to x$ and knowing that $Tx_n \to y$ we can conclude $x \in \mathcal{D}_T$ and $Tx = y$.

[5] To make connection with applications in quantum mechanics change notation in your mind: $|x\rangle \sim x(t) \to |\psi\rangle \sim \psi(x)$.

that the functions have symmetric boundary conditions. Similarly we define $A_a = i\,d/dt$ with anti-symmetric boundary conditions $\mathcal{D}_{A_a} = \mathcal{D}_A \cap \{x|x(0) = -x(1)\}$. Finally we define $A_{00} = i\,d/dt$ with $\mathcal{D}_{A_{00}} = \mathcal{D}_A \cap \{x|x(0) = 0, x(1) = 0\}$. All domains are dense in H, thus we can define the adjoint for all four operators A, A_s, A_a, and A_{00}. These operators are closely related to the momentum operator (7.72). In physics text you will often read that $A_{00} = i\,d/dt$ (or $i\,d/dx$ to make the connection with the momentum operator p_x obvious) is a self-adjoint operator with the "proof" given as ⇒ lecture

$$\langle y|A_{00}x\rangle = \int_0^1 dt\, y^*(t)\, i\frac{d}{dt}x(t) = iy^*(t)x(t)\Big|_0^1 + \int_0^1 dt\, \left(i\frac{d}{dt}y(t)\right)^* x(t) = \langle A_{00}y|x\rangle \tag{10.23}$$

where we have used integration by parts and then the boundary conditions. Not a word about the domains! In fact, it is a nice exercise to convince yourself that A_s and A_a are self-adjoint, A_{00} is symmetric but not self-adjoint, and A is not even symmetric. Neither of these operators are bounded. Indeed, if we take $x_n(t) = \sqrt{2}\sin(n\pi t)$ with $n \in \mathbb{N}$, then $A|x\rangle \sim i\,n\sqrt{2}\,\pi\cos(n\pi t)$. Thus $\|x_n\| = 1$ but $\|Ax_n\| = \pi n\|x_n\|$ with $n \to \infty$. Thus, the momentum operator in quantum mechanics is not bounded, just in case you needed convincing that we have to learn how to deal with unbounded operators.

Let us stay in $H = L^2[0,1]$ and look at the 'multiplication' operator $V|x\rangle = V(t)\,x(t)$, where V often corresponds to a potential in the Schrödinger equation. If $V(t)$ is bounded as a function in $t \in [0,1]$ the operator V is bounded as an operator in $H = L^2[0,1]$, since $\|Vx\| \le c\|x\|$ with $c = \max_{t\in[0,1]} |V(t)|$. Then $\mathcal{D}_V = H$. For a real $V(t) = V^*(t)$ we have

$$\langle y|Vx\rangle = \int_0^1 dt\, y^*(t)\, V(t)x(t) = \int_0^1 dt\, \left(V^*(t)y(t)\right)^* x(t) = \langle Vy|x\rangle \tag{10.24}$$

Hence, a real bounded potential $V(t)$ corresponds to a self-adjoint operator. Since this operator is defined everywhere, self-adjoint and symmetric are the same. A special case is the *position operator [Ortsoperator]* $V(t) = t$.

Moving to $L^2[\mathbb{R}]$, the position operator $V(t) = t$ with $\mathcal{D}_V = \{x|x(t), tx(t) \in L^2[\mathbb{R}]\}$ is densely defined, but not bounded any longer. Still, it remains self-adjoint. The differentiation operator $A = i\,d/dt$ with $\mathcal{D}_A = \{x|x(t), x'(t) \in L^2[\mathbb{R}]\}$ is densely defined, unbounded, but self-adjoint. In physics applications we work very often on the dense subset $C_c[\mathbb{R}]$ whose elements $x(t) = 0$ for $|t| \to \infty$. This means in the "proof" given in (10.23) there are no boundary terms from the integration by parts.

10.4 The spectrum

In Section 10.1 we have defined EV and EVal of an operator and the definition carries over to infinite-dimensional Hilbert spaces. However, in infinite-dimensional spaces we still have to generalise the concept of an EVal to the spectrum of an operator, as something new can happen.

To start with, we consider the resolvent $R_T(k) = (T - k)^{-1}$. If for a particular $k \in \mathbb{K}$ the resolvent $R_T(k)$ exists (then it is automatically linear) is defined everywhere i.e. $\mathcal{D}_{R_T(k)} = H$ and is bounded[6] then we call k a *regular point [regulärer Punkt]*. In other words k is a regular point of T if $R_T(k)$ is quite simply perfect. The set of all regular points is called

[6] You will recall that linear and defined everywhere "nearly but not quite" implies bounded.

the *resolvent set* *[Resolventenmenge]* and is denoted by $\rho(T) \subset \mathbb{K}$. Finally, the *spectrum* *[Spektrum]* $\sigma(T) \subset \mathbb{K}$ of the operator T is just the complement of the resolvent set, i.e. $\sigma(T) = \mathbb{K} \backslash \rho(T)$.

Now, what's the fuss. If λ is an EVal of T then $R_T(\lambda)$ does not exist. Thus $\lambda \notin \rho(T)$ and $\lambda \in \sigma(T)$. For finite-dimensional Hilbert spaces that's pretty much all there is. The spectrum is the set of al EVal, and any $k \in \mathbb{K}$ that is not an EVal is in $\rho(T)$.

However, in infinite-dimensional Hilbert spaces, things are a bit more delicate. It is still true that all EVal λ_i of T are in the spectrum. This part of the spectrum we call the *discrete* or *point* spectrum *[diskretes oder Punktspektrum]*, denoted by $\sigma_d(T) = \{\lambda_i\}$. But something else can happen. It is possible that there is a λ that is not an EVal and in fact $R_T(\lambda)$ exists, but either $R_T(\lambda)$ is not bounded or $\mathcal{D}_{R_T(k)} \neq H$. Then we say λ is in the *continuous* *[kontinuierliches]* spectrum, $\lambda \in \sigma_c(T)$ (a slightly misleading name as we will see and sometimes the spectrum is further decomposed, but this need not concern us here). This is a new feature in infinite-dimensional Hilbert spaces and can be seen as a generalisation of the concept of an eigenvalue. The full spectrum $\sigma(T) = \sigma_d(T) \cup \sigma_c(T)$. For us, all EVal are in $\sigma_d(T)$ and all other 'funny' values are in $\sigma_c(T)$. We remark without proof that the spectrum of a linear bounded operator is closed in the complex plane, i.e. as subset of the metric space $M_{\mathbb{C}}$, (9.5).

We have already mentioned that for $H = \mathbb{K}^n$ we have $\sigma_d(T) = \{\lambda_i\}$ where λ_i are the EVal and $\sigma_c(T) = \varnothing$.

Now let us make a more subtle example, with $H = \ell^2$. Again, we look at the ladder operator T_+, defined through $T_+(x_1, x_2 \ldots) = (0, x_1, x_2 \ldots)$. We see immediately that T_+ has no EVal, since $(0, x_1, x_2 \ldots) = \lambda(x_1, x_2 \ldots)$ implies $\lambda x_1 = 0$. Hence, either we have $\lambda = 0$ which entails $|x\rangle = 0$ or $x_1 = 0$ which implies $x_2 = 0$ which implies ... again $|x\rangle = 0$. As we have agreed, this is not an EV. Thus, there are no EV. And here is the "but"... $0 \in \sigma(T_+)$. To see this lets look at the resolvent. First $(T_+ - \lambda)|x\rangle = (-\lambda\, x_1, x_1 - \lambda\, x_2, x_2 - \lambda\, x_3 \ldots)$ thus $(T_+ - \lambda)^{-1}(-\lambda\, x_1, x_1 - \lambda\, x_2, x_2 - \lambda\, x_3 \ldots) = |x\rangle$. This can be rewritten as

$$(T_+ - \lambda)^{-1}|y\rangle = \Big(-\frac{y_1}{\lambda}, -\frac{y_2}{\lambda} - \frac{y_1}{\lambda^2}, -\frac{y_3}{\lambda} - \frac{y_2}{\lambda^2} - \frac{y_1}{\lambda^3} \ldots \Big) \tag{10.25}$$

For $\lambda = 0$ this does not look good. In fact $\nexists T_+^{-1} \leftrightarrow \nexists R_{T_+}(0)$. We could restrict the domain to $\mathcal{D}_{T_+^{-1}} = \{x | x = (0, x_2 \ldots)\}$ i.e. to the elements in ℓ^2 where the first entry is 0. On this domain T_+^{-1} exists, but is not everywhere defined. Really $\lambda = 0$ is a funny point. It is not an EVal, but the resolvent is not perfect either. We have $0 \in \sigma_c(T_+) = \sigma(T_+)$. Now you understand why 'continuous spectrum' is a rather strange name. The reason for this name will become clearer when we look at unbounded operators.

You might think this is an 'academic' example that will never show up in the real physicists world. Well, well ...

10.5 The spectral representation of compact operators

We would like to generalise the spectral representation (10.8) to a self-adjoint operator T in a (separable) infinite-dimensional Hilbert space. However, as we have seen in Section 10.4, in such Hilbert spaces the notion of the spectrum is more subtle, and since the EVal λ_i as part of the spectrum are an essential component of the spectral representation, it is entirely unclear if and in what form such a generalisation exists.

In this section we will restrict ourselves to a special class of operators for which a spectral representation very similar to (10.8) can be obtained. These are *compact [kompakt oder vollstetig]* operators.

Let M be a bounded subset of H. Under a continuous (thus bounded) operator this set is mapped into another bounded subset M'. For a compact operator we ask a bit more. An everywhere defined operator T is called compact if it maps any bounded subset M onto a relative compact subset M', i.e. for any bounded M we have $T : M \mapsto M'$ with $\overline{M'}$ compact. As an alternative definition, an operator T is compact iff for any bounded sequence $\{x_n\}$ the sequence $\{T\,x_n\}$ has a convergent subsequence. It can easily be shown that any compact operator is continuous, but the converse is not true. A counter example is given by the unit operator $T = \mathbb{I}$ on ℓ^2 which is certainly continuous but not compact, as it maps the unit ball (a bounded set) to the unit ball (not a relative compact set). Once more, this is connected to the fact that bounded and closed in infinite-dimensional spaces need not imply compact. ⇒ lecture

If we have an operator $T = T_1\,T_2 \ldots T_n$ with all T_i bounded and at least one T_j compact, then T is compact. This is because a continuous operator maps a (relative) compact set to a (relative) compact set. Thus if we act with T on a bounded set, all continuous T_i will keep the set bounded, the compact T_j makes it relative compact and the following T_i keep it relative compact.

To get familiar with the concept of compact operator we look at a few examples. The raising operator T_+ on ℓ^2 is continuous. But following the same argument as for the unit operator we conclude it is not compact. On the other hand, the operator $T|x\rangle = (k_1\,x_1, k_2\,x_2 \ldots)$ with $k_i \to 0$ is compact. The last example can be generalised to ⇒ lecture

$$\text{Hilbert-Schmidt operators} \qquad \sum_{i=1}^{\infty} \|T\,\varphi_i\|^2 < \infty \tag{10.26}$$

which are compact. In (10.26) $\{\varphi_i\}$ is a ONB.

In connection with the Fourier and Laplace transform we considered integral transforms, (3.17). They can be interpreted as a (linear) operator on a (function) vector space

$$T|x\rangle = |y\rangle \sim y(t) = \int_a^b K(t,s)\,x(s)\,ds \tag{10.27}$$

with $K(t,s)$ the kernel (not null space!) that defines T. If we consider the pre-Hilbert space $C[a,b]$ then for a kernel continuous on $[a,b]\times[a,b]$ the operator T is compact. For the Hilbert space $L^2[a,b]$ if we require the kernel K to be square integrable on $[a,b]\times[a,b]$ we ensure T is compact.

More generally, if T is continuous and $\dim \mathcal{R}_T < \infty$ then T is compact. In particular in a finite-dimensional Hilbert space a continuous operator is compact (recall the counter examples above for ℓ^2). This is again related to closed and bounded being equivalent to compact for finite-dimensional spaces.

Finally if T is compact, so is $T^\dagger$.

Now that we have defined compact operators, we will consider a compact self-adjoint operator. Since $T = T^\dagger$ we can write the norm (9.17) also as $\|T\| = \sup_{\|x\|=1} |\langle x|T|x\rangle|$. Thus, there is a sequence $\{x_n\}$ with $\|x_n\| = 1$ and $\langle x_n|T|x_n\rangle \to \lambda_1 = \pm\|T\|$. We can now show that for this sequence $\|Tx_n - \lambda_1 x_n\| \to 0$. If the sequence was convergent we could conclude ⇒ lecture

that T has an EV (the limit of x_n) with EVal λ_1. This is where compactness of T comes into play. Since the sequence $\{x_n\}$ is bounded, and T is compact, the sequence $\{T\,x_n\}$ has a convergent subsequence $\{T\,x'_n\}$ for which we still have $\|T\,x'_n - \lambda_1\,x'_n\| \to 0$. Since $\{T\,x'_n\}$ and $T\,x'_n - \lambda_1\,x'_n$ converge, so does the difference, i.e. $x'_n \to y_1$. This $|y_1\rangle$ is an EV of T with EVal $\lambda_1 = \pm\|T\|$.

As a side remark we mention that for compact (and self-adjoint) operators the supremum in the definition of the norm can be replaced by the maximum.

This was a lot of work to find an EVal λ_1 and EV $|y_1\rangle$ for a compact, self-adjoint operator T. We define $H_1 = \text{span}(y_1)$ and $H = H_1 \oplus H_1^\perp$. Next we consider the operator T on $H_1^\perp$ and call it $T_1 \subset T$, i.e. $\mathcal{D}_{T_1} = H_1^\perp$. This operator is still compact and self-adjoint and $\|T_1\| \leq \|T\|$. Repeating the procedure we will find an EVal λ_2 with $|\lambda_2| \leq |\lambda_1|$ and its EV $|y_2\rangle$. Guess what. Let's do this again. Define $H_2 = \text{span}(y_2)$ and $H = H_1 \oplus H_2 \oplus H_2^\perp$ and define $T_2 \subset T$ with $\mathcal{D}_{T_2} = H_2^\perp$. We repeat this game either until we have found n EV in an n dimensional Hilbert space, or an infinite sequence of EV $|y_i\rangle$ with EVal λ_i which are ordered as $\|T\| = |\lambda_1| \geq |\lambda_2| \geq |\lambda_3| \ldots$.

It is possible to obtain further conditions. While it is perfectly possible that some EVal are degenerate, e.g. $\lambda_i = \lambda_{i+1}$ it can be shown that each $|\lambda_i| > 0$ is degenerate at most a finite number of times. Thus, the sequence $\{\lambda_i\}$ can only have 0 as an accumulation point and only the EVal 0 can be arbitrarily degenerate. The non-zero spectrum of T is discrete, i.e. consists only of the point spectrum, $\sigma(T)\backslash\{0\} = \sigma_d(T) = \{\lambda_i\}$. ⇒ lecture

Thus, for a compact self-adjoint operator T we can generalise (10.8) and (10.9) to

$$T = \sum_{i=1}^{\dim H} \lambda_i |\varphi_i\rangle\langle\varphi_i| = \big(0\,P_0 + \big)\sum_j \lambda_j P_j \tag{10.28a}$$

$$H = H_0 \oplus H_1 \oplus H_2 \ldots = H_0 \oplus \bigoplus_j H_j \tag{10.28b}$$

and, hence, use (10.10) to 'do with the operator whatever we want'. In (10.28) $\{\varphi_i\}$ are the mutually orthogonal EV to the EVal λ_i and the sum j is over all distinct non-zero EVal, with P_j the projectors onto the corresponding eigenspaces H_j. Each H_j is finite dimensional except possibly for H_0, the eigenspace of the EVal $\lambda = 0$. In (10.28a) we have written the term involving P_0, the projector onto H_0, in brackets, because $\lambda = 0$ might or might not be an EVal. If it is not, $H_0 = \{0\}$. If it is, the term P_0 has to be kept in (10.28a) even though it is multiplied by zero. The reason is that this term is needed e.g. if we want to compute $\cos(T)$ according to (10.10).

As an example we go back to $H = \ell^2$ with $T|x\rangle = (x_1, x_2/2, x_3/3\ldots)$. In this case we have $\lambda_i = 1/i$ and $|\varphi_i\rangle = (0\ldots1\ldots)$ with the 1 at the ith position. No EVal is degenerate, every H_j is one-dimensional. The spectrum $\sigma(T) = \{\lambda_i = 1/i,\, i \in \mathbb{N}\} \cup \{0\} = \sigma_d(T) \cup \sigma_c(T)$. Note that $\lambda = 0$ is not an EVal, but still part of the spectrum, since for $\lambda = 0$ the resolvent $(T-\lambda)^{-1}|x\rangle = T^{-1}|x\rangle = (x_1, 2\,x_2, 3\,x_3\ldots)$ is not bounded. Another way to see that 0 has to be part of the spectrum is to note that it is an accumulation point of the set of EVal. Since the spectrum is closed it has to contain all accumulation points. To look at such operators in $H = \ell^2$ more generally, we recall that $T|x\rangle = (k_1\,x_1, k_2\,x_2\ldots)$ is compact for $k_i \to 0$. The discrete spectrum is $\sigma_d(T) = \{k_i\}$. It is possible that some k_i are equal, i.e. this EVal can be degenerate. To ensure $k_i \to 0$, however, it is clear that no value of $k_i \neq 0$ can be repeated infinitely many times, only $k_i = 0$ can appear as many times as it likes: never, a finite number

of times, or infinitely often. This is in agreement with what we have stated in connection with (10.28).

10.6 Integral equations

In this section we want to put to use what we have learnt in Section 10.5 and show how to solve integral equations. What is an integral equation? For a given $y(t) \in C[a,b]$ and $K(t,s) \in C[a,b] \times C[a,b]$, we are looking for an $x(t)$ satisfying

$$\text{Fredholm 1st kind:} \qquad y(t) = \int_a^b K(t,s)x(s)ds \tag{10.29}$$

This is a Fredholm integral equation of the first kind. Often, the kernel has the form $K(t,s) = K(t-s)$ in which case the r.h.s. is just a convolution (3.25).

What has all this to do with Section 10.5? According to (10.27) we can write this equations in terms of a compact operator T as $|y\rangle = T|x\rangle$ and, hence, solve it as $|x\rangle = T^{-1}|y\rangle$. Even better, if we consider a kernel $K(t,s) = K^*(s,t)$ the operator T is self-adjoint. To invert it, we 'simply' find all EVal, the corresponding EV, write the spectral representation (10.28a) and we're done. In fact, we're done for any $y(t)$. Of course, the difficulty is finding all EV and EVal, but if 0 is not an EVal, we know that T can be inverted and a solution exists.

Integral equations often come in a slightly different form, namely in the form of a Volterra and Fredholm integral equations of the second kind. For $\bar{y}(t) \in C[a,b]$ and $\alpha \in \mathbb{K}$ they read

$$\text{Volterra 2nd kind:} \qquad x(t) = \bar{y}(t) + \alpha \int_a^t K(t,s)x(s)ds \tag{10.30a}$$

$$\text{Fredholm 2nd kind:} \qquad x(t) = \bar{y}(t) + \alpha \int_a^b K(t,s)x(s)ds \tag{10.30b}$$

Note the difference in the integration boundaries between in (10.30a) and (10.30b).

Again, we can write (10.30) in operator form as

$$(1-\alpha T)|x\rangle = |\bar{y}\rangle \quad \Leftrightarrow \quad (T-k)|x\rangle = |y\rangle \quad \text{with } k = \alpha^{-1} \text{ and } |y\rangle = -\alpha^{-1}|\bar{y}\rangle \tag{10.31}$$

Thus, the solution to these equations on a formal level is simply obtained by finding the resolvent $R_T(k) = (T-k)^{-1}$, (10.4), and $|x\rangle = (T-k)^{-1}|y\rangle = -k\,(T-k)^{-1)}|\bar{y}\rangle$.

For which values of k (or α) does the resolvent exist? How do we find it? These are again precisely the questions we have just considered. The answer is given in the remainder of this subsection.

For the Volterra equation (10.30a) the resolvent $R_T(\alpha)$ exists for all $\alpha \neq 0$ and then there is always a unique solution. ⇒ lecture

For the Fredholm equation (10.30b) the answer is more complicated. If $k = 1/\alpha$ is not an EVal of T, the resolvent $R_T(k)$ exists and we can find a solution as indicated below (10.31). If $k = \lambda_k$ is an EVal of T, then a solution exists only if $y \in H_k^\perp$, i.e. if y is in the orthogonal complement of the eigenspace H_k to the EVal λ_k. To understand this, we note $T - k = \sum_i (\lambda_i - k)|\varphi_i\rangle\langle\varphi_i|$. We want to build $(T-k)^{-1}|y\rangle = \sum_i (\lambda_i - k)^{-1}|\varphi_i\rangle\langle\varphi_i|y\rangle$. If $k \notin \sigma(T)$, no problem, if $k = \lambda_k$ we need the numerator to vanish, i.e. $\langle\varphi_k|y\rangle = 0$ for all EV $|\varphi_k\rangle$ to the EVal λ_k. Hence, $y \in H_k^\perp$ is necessary to ensure $y \in \mathcal{D}_{(T-k)^{-1}}$.

To summarise, the solution to (10.30b) is

$$k \notin \sigma(T): \qquad |x\rangle = \sum_i \frac{|\varphi_i\rangle\langle\varphi_i|}{\lambda_i - k} \tag{10.32a}$$

$$k = \lambda_k \in \sigma(T): \qquad |x\rangle = \sum_{i\neq k} \frac{|\varphi_i\rangle\langle\varphi_i|}{\lambda_i - k} + |y_k\rangle \tag{10.32b}$$

where $|y_k\rangle$ is an arbitrary vector in $H_k^\perp$. The resolvent is actually nothing but the Green operator. Or, to better link with Section 7.4, if we write the resolvent also as integral operator, then the corresponding kernel is the Green function.

The solution in (10.32) required that we have found all EVal with their EV. In most cases this is easier said than done, or to put it more bluntly, simply not possible. In many, really many physics applications we are in the situation that we cannot solve the system, but in (10.31) the term $|\alpha|\|T\| < 1$. In this case, we can write the inverse $(1-\alpha\,T)^{-1}$ through the *Neumann series [Neumann Reihe]* as

$$(1-\alpha\,T)^{-1} = \sum_{n=0}^{\infty} \alpha^n\,T^n \qquad \text{or explicitly} \qquad \sum_i \frac{|\varphi_i\rangle\langle\varphi_i|}{1-\alpha\lambda_i} = \sum_{n=0}^{\infty} \alpha^n \sum_i \lambda_i^n |\varphi_i\rangle\langle\varphi_i| \tag{10.33}$$

However simple the first equation above looks, to start with it is actually not clear at all what it means. We have an equality between operators, but the r.h.s. consists of an infinite sum of operators. Thus, it has to be connected to some sort of convergence of partial sums of operators. While we will be looking at this more generally only in the next Section 10.7, for a compact operator it is clear what we mean, as explicitly written in the second relation of (10.33). Since we have a spectral representation with a pure discrete spectrum, (10.28a), we can use (10.10) and simply expand the expression $(1-\alpha\,\lambda_i)^{-1}$ for $|\alpha\,\lambda_i| < 1$.

This allows us to obtain ever better approximate solutions $|x_{(n)}\rangle$ to (10.30) through iteration. The method is called *perturbation theory [Störungstheorie]* and is particularly useful, if $|\alpha|\|T\| \ll 1$:

$$\text{to solve} \qquad |x\rangle = |\bar{y}\rangle + \alpha\,T|x\rangle \tag{10.34a}$$

$$\text{step 0} \qquad |x_{(0)}\rangle = |\bar{y}\rangle \tag{10.34b}$$

$$\text{1st order} \qquad |x_{(1)}\rangle = |\bar{y} + \alpha\,T|x_{(0)}\rangle = |\bar{y}\rangle + \alpha\,T|\bar{y}\rangle \tag{10.34c}$$

$$\text{2nd order} \qquad |x_{(2)}\rangle = |\bar{y}\rangle + \alpha\,T|x_{(1)}\rangle = |\bar{y}\rangle + \alpha\,T|\bar{y}\rangle + \alpha^2\,T^2|\bar{y}\rangle \tag{10.34d}$$

$$\ldots$$

Of course, this is nothing but (10.33), including terms up to a certain power in α^n. But we do not need to know λ_i or $|\varphi_i\rangle$. What is not clear at all in general is whether the procedure in (10.34) converges. For a compact operator with $|\alpha\,\lambda_i| < 1$ it does. However, you will use this procedure also many times without worrying whether or not the series converges. Even worse, we use it in cases where we know the series does not converge. But this is a story for another day.

You should have realised that we were cheating throughout this subsection. We pretended to be in a Hilbert space, but we said $y(t) \in C[a,b]$. Implicitly, we have been working in $L^2[a,b]$ and just have chosen a $y \in L^2[a,b]$ that has a (unique) continuous representative. Again, this is something you will routinely do in quantum mechanics.

In case you wondered why we have not yet bothered at all with the Volterra integral equation of the first kind, it is because it can be transformed through differentiation w.r.t. t into the form (10.30a), as long as $K(t,t) \neq 0$ and y is differentiable. ⇒ lecture

10.7 Convergence of operators

We have written several times relations between operators where the r.h.s. was an infinite sum, such as the Neumann series, (10.33). So far, such relations were restricted to self-adjoint compact operators where we have a convenient spectral representation. For generic operators a careful discussion is required on what the equal sign actually means. It is clear that it has to be related to a convergence of partial sums. Thus, we are led to look at the concept of a converging sequence of operators.

Let us consider a sequence of operators T_n. We say this sequence *converges in the norm [Konvergenz in der Norm]* to the operator T, if $\|T - T_n\| \to 0$, or $\lim_{n\to\infty} \|T - T_n\| = 0$. We have expressed the convergence of operators through the convergence of real numbers, the norm. The infinite sums we have used so far, such as (10.17), (10.28a), or (10.33) corresopnd to convergence in the norm of the parial sums $T_n = \sum_{i=1}^{n}$ on the r.h.s. Concretely, for (10.28a) we have

$$\|T - T_n\| = \|\sum_{i=n+1}^{\infty} \lambda_i P_i\| \;\leq\; \max_{i\geq n+1} |\lambda_i| \, \|\sum_{i=n+1}^{\infty} P_i\| \;\leq\; \max_{i\geq n+1} |\lambda_i| \;\to\; 0 \tag{10.35}$$

and, hence, $\lim_{n\to\infty} \|T - T_n\| = 0$. In (10.35) we have used that P_i and, therefore, $\sum P_i$ are projectors with norm 1. If T_n converges to T in the norm, sometimes the symbol $T_n \Rightarrow T$ is used.

Another way of convergence is the *strong [stark]* convergence, often denoted as $T_n \to T$. This means

$$\|Tx - T_n x\| \to 0 \qquad \forall x \in \mathcal{D}_T \cap \mathcal{D}_{T_n} \tag{10.36}$$

Here we have expressed the convergence of operators through the convergence of vectors.

Finally, there is a third possibility to define convergence, called *weak [schwach]* convergence, namely

$$|\langle y|Tx\rangle - \langle y|T_n x\rangle| \to 0 \qquad \forall x, y \in \mathcal{D}_T \cap \mathcal{D}_{T_n} \tag{10.37}$$

In this case, the convergence of operators is expresses through the convergence of complex numbers, the scalar product. This form of convergence is often denoted by $T_n \rightharpoonup T$.

Convergence in the norm implies strong convergence, if $T \Rightarrow T$ then $T_n \to T$. This simply follows from $\|Tx - T_n x\| = \|(T - T_n)x\| \leq \|T - T_n\| \, \|x\| \to 0$, using (9.18) but the converse is not true. A counter example is $T_n|x\rangle = (0 \ldots, 0, x_{n+1} \ldots)$ where the first n entries of $x \in \ell^2$ are set to 0. We have $T_n \to T = 0$, but $T_n \not\Rightarrow T = 0$.

Strong convergence implies weak convergence, if $T \to T$ then $T_n \rightharpoonup T$. Using (9.24) we have $|\langle y|Tx\rangle - \langle y|T_n x\rangle| = |\langle y|(T - T_n)x\rangle| \leq \|y\| \|Tx - T_n x\| \to 0$. Again, the converse is not true as illustrated by the following counter example. Take $T_n = (T_+)^n$ with T_+ the raising operator introduced earlier, s.t. $T_n|x\rangle = (0, \ldots, 0, x_1, x_2 \ldots)$ for $x \in \ell^2$. Then we have $T_n \rightharpoonup T = 0$, but $T_n \not\to T = 0$. ⇒ lecture

To summarise, when writing equations between operators and one side of the equation consists of an infinite sum, we better make sure to explain what presicely is meant by the equal sign. So far, it was always meant to be the strongest possible equal sign, i.e. convergence in the norm.

10.8 The resolvent and operators with compact resolvent

The resolvent of an operator has been mentioned several times already and now we want to spend some quality time with it. Let T be any linear operator, not necessarily bounded. If k is regular, i.e. if $k \in \rho(T)$ the resolvent $R_T(k) = (T-k)^{-1}$ exists, is linear, bounded and defined everywhere. This is just the definition of the resolvent set $\rho \subset \mathbb{K}$.

Thus, we have $(T-k)R_T(k) = \mathbb{I}$ defined on the full Hilbert space. Also, $R_T(T-k) = \mathbb{I}$, but only for $x \in \mathcal{D}_T$. This means on $\mathcal{D}_T$ we have $[R_T(k), T] = 0$ and $R_T(k)T = 1 + kR_T(k)$.

Next, we consider different regular points $k_i \in \rho(T)$ and introduce the short-hand notation $R_i = R_T(k_i)$. We can show the ⇒ lecture

$$\text{first resolvent identity} \qquad R_1 - R_0 = (k_1 - k_0) R_1 \, R_0 \tag{10.38}$$

simply through multiplication by $(T-k_1)$ from the left and $(T-k_0)$ from the right. This also implies $[R_T(k_1), R_T(k_2)] = 0$ and using the esommutation of the resolvent (for different regular points k_i) commutes we can rewrite (10.38) as

$$R_1 = R_0 + (k_1 - k_0) R_0 \, R_1 = R_0 + (k_1 - k_0) R_0 \left(R_0 + (k_1 - k_0) R_0 \, R_1 \right) = \dots \tag{10.39}$$

which through iteration leads to

$$R_T(k) = \sum_{n=0}^{\infty} (k - k_0)^n \left(R_T(k_0) \right)^{n+1} \qquad \text{for} \quad k, k_0 \in \rho(T) \tag{10.40}$$

This is quite remarkable. If we consider R_T as a function of k, (10.40) states that the 'function' $R_T(k)$ is analytic, as it can be expanded in a Taylor series about k_0. The term function has been put in quotation marks, since this is an operator valued function, i.e. $\forall k \in \rho(T)$ we get an operator $R_T(k)$. Since $\sigma(T)$ is closed, the complement $\rho(T)$ is open, as we are used to from our considerations of analytic functions in Section 1.

But wait a moment. What is meant by the equal sign? We have an infinite series of operators on the r.h.s. of (10.40) so we better are clear about the convergence we are talking about. If $|k - k_0| \, \|R_0\| < 1$ the series on the r.h.s. of (10.40) converges in the strong sense. Hence, if we want to compute $R_T(k)|x\rangle$ we can express it in terms of $R_T(k_0)$ acting (multiple times) on $|x\rangle$. Roughly speaking, if we know the resolvent at one regular point, we know it in a neighbourhood of this point.

Another relation between resolvents similar to (10.34) is also of utmost importance. It is related to the situation where we have an operator T for which we cannot find the resolvent. Sadly, this is a very common situation. The cases where we can find all EV and EVal of a operator are few and far between. However, assume we are not completely out of luck in that we can write $T = T_0 + V$ and we do know the resolvent for T_0 and the 'perturbation' V is in some sense 'small'. The question now is can we somehow get information on $R_T = R_T(k)$ from our knowledge on $R_{T_0} = R_{T_0}(k)$. Again this is to do with perturbation theory as in (10.34) and has countless applications. First, we can show the ⇒ lecture

second resolvent identity $$R_T = R_{T_0} - R_{T_0}\, V\, R_T = R_{T_0} - R_T\, V\, R_{T_0} \tag{10.41}$$

This is a relation similar to (10.38), but relates the resolvent of two different operators. As it stands, (10.41) is not directly useful, since the unknown R_T appears on both sides of the equation. However, not bothering about convergence, we can use the same procedure as in (10.34) or in (10.38) to iteratively express R_T in terms of R_{T_0} and obtain

$$R_T \stackrel{?}{=} R_{T_0} \sum_{n=0}^{\infty} (-V\, R_{T_0})^n \tag{10.42}$$

This is fantastic as it allows us to approximate the unknown and complicated R_T through R_{T_0} and the perturbation V. What is not so fantastic is that the equal sign is very shady. We stated that V has to be 'small'. Then one would hope the successive terms in (10.42) become ever less important, as they have higher powers of V. But who tells us that (10.42) converges and in what sense? There is no simple and general answer to this question. It has to be investigated case by case. Physicists, as we know them, typically just don't bother, simply use (10.42) at their hearts desire and hope for the best. Funny enough, doing so they very often get excellent agreement with nature.

So far, we have considered a generic resolvent. Now we want to look at the special case of an operator T whose resolvent $R_0 = (T - k_0)^{-1}$ is compact. If $R_0 = R_T(k_0)$ is compact for any $k_0 \in \rho(T)$ then it is compact $\forall k \in \rho(T)$ as can be seen from (10.38). Since R_0 is compact, ⇒ lecture we can use the results of Section 10.5 to conclude that its spectrum consists of isolated EVal supplemented by 0, if 0 is not an EVal, i.e. $\sigma(R_0) = \sigma_d(R_0) \cup \{0\}$. What does this imply for T and its spectrum $\sigma(T)$?

To investigate this it is useful to consider the operator $Q \equiv -k\, R_0(R_0 - k)^{-1}$, which for $k \in \rho(R_0)$ is bounded, everywhere defined and even compact (since R_0 is). Multiplying the definition of Q by $(R_0 - k)$ and simple manipulations allow us to conclude that for $k \neq 0$ ⇒ lecture

$$Q = R_0 + \frac{1}{k} Q R_0 \qquad \Rightarrow \qquad Q \equiv -k\, R_0(R_0 - k)^{-1} = R_T\big(k_0 + \frac{1}{k}\big) \tag{10.43}$$

where for the last step we have used the first resolvent identity (10.38). Since Q is bounded and everywhere defined, we conclude (still for $k \in \rho(R_0)$ and $k \neq 0$) that $k_0 + 1/k \in \rho(T)$. Hence, the EVal of T can only be $k_0 + 1/\lambda_i$ with $\lambda_i \in \sigma(R_0)$ and possibly $\lambda = \infty$.

This is a very important result that explains several 'mysteries' we have encountered. A self-adjoint operator T not necessarily bounded (but certainly densely defined) with a compact resolvent has a discrete set of EVal λ_i, each degenerate at most a finite number of times, and ∞ the only possible accumulation point. The spectrum $\sigma(T) = \sigma_d(T) = \{\lambda_i\}$.

The (differential) operator showing up in the Sturm-Liouville problems (5.51) is such an operator. Thus, we can now answer the question asked at the end of the first paragraph of Section 5.7. No, this is indeed not a coincidence!

10.9 "Evil" operators and the spectral family

The operators we have worked with so far were all nice in some way. Either they were compact, or had a compact resolvent. Now we will face the harsh reality of real live and consider operators that are neither compact nor do they necessarily have a compact resolvent. In fact, they might not even be bounded.

Another step we take in this subsection is to change to a QM-inspired notation. For the remainder of Section 10 we will denote vectors by $|\psi\rangle \sim \psi(x)$ and tacitly think of the wave function. A crash course of QM with typical applications will be given in Section 10.10.

The first Hilbert space we consider is $L^2[0,1]$ and the (in this case bounded) position operator $X|\psi\rangle = x|\psi\rangle \sim x\,\psi(x)$. We want to find an EV and the corresponding EVal. At first sight it seems we have found the EVal already, namely x. But $0 \leq x \leq 1$ is not a constant, unless $\psi(x) = 0$ for all x but one, say λ. Does such a $\psi(x)$ exist? Let's go back to Section 2.2 and Figure 2.1. Indeed, in the limit $n \to \infty$ these functions τ_n do the trick. But in this limit these functions are not functions any longer, rather they are the $\delta(x-\lambda)$ distribution. While naively we could say $|\psi_\lambda\rangle \sim \delta(x-\lambda)$ with $0 \leq \lambda \leq 1$ satisfies $X|\psi_\lambda\rangle = \lambda|\psi_\lambda\rangle$ for a constant λ, sadly $\delta(x-\lambda) \notin L^2[0,1]$. Thus, our $|\psi_\lambda\rangle$ is not in the Hilbert space and, therefore, certainly not an EV. In fact, the position operator has no EV and, hence, no EVal. Since we still have $X|\psi_\lambda\rangle = \lambda|\psi_\lambda\rangle$ physicists often just pretend this is good enough and call $|\psi_\lambda\rangle$ a *generalised [uneigentlicher]* EV with λ its generalised EVal.

Even though there are no EVal, $\sigma_d(X) = \varnothing$, the position operator has a rich spectrum, $\sigma(X) = \sigma_c(X) = [0,1]$. To see this, let us try to find the resolvent $R_X(k)$. Naively it is $R_X(k)|\psi\rangle \sim \psi(x)/(x-k)$. Clearly, $\forall\, k \in [0,1]$ this operator is not defined everywhere. In some sense this failure is a left-over of the fact that all values $\lambda \in [0,1]$ are generalised EVal. And now the name continuous spectrum makes more sense.

In such a situation it is futile to hope we could ever write something like (10.28a). What are we supposed to sum over? If anything, we would have to integrate over all λ between 0 and 1. But what is the integrand? This is where the *spectral family [Spektralschar]* enters the stage.

To prepare the stage we return for a moment to a compact, self-adjoint operator T and rewrite its spectral representation (10.28a) through the spectral family

$$E(\lambda) = E_\lambda \equiv \sum_{\lambda_i \leq \lambda} P_i \tag{10.44}$$

This is an operator valued function, i.e. for each $\lambda \in \mathbb{R}$ (all EVal are real) we get a self-adjoint operator $E(\lambda) = E^\dagger(\lambda)$. Let us assume $m_1 = \min\{\lambda_i\}$ and $m_2 = \max\{\lambda_i\}$ are the smallest and largest EVal of the compact, self-adjoint operator T. Then $E(\lambda)$ as defined in (10.44) satisfies

$$\begin{aligned} &E(\lambda) = 0 \quad \text{for } \lambda < m_1 && E(\lambda) = \mathbb{I} \quad \text{for } \lambda > m_2 \\ &E^2(\lambda) = E(\lambda) && E(\mu)E(\lambda) = E(\lambda)E(\mu) = E(\min\{\mu,\lambda\}) \end{aligned} \tag{10.45}$$

The equal signs in (10.45) are to be understood in the strong sense and the identity $E(\lambda) = \mathbb{I}$ for $\lambda > m_2$ is nothing but the decomposition of the identity as already encountered e.g. in (10.17).

We can express the projection operators $P_i = E(\lambda_i) - E(\lambda_{i-1}) \equiv \Delta E_{\lambda_i}$ through the spectral family. Hence, we can also express T through the spectral family

$$T = \sum_{\lambda_i} \lambda_i P_i = \sum_{\lambda_i} \lambda_i\, \Delta E_{\lambda_i} \tag{10.46}$$

This form of T is ready to be generalised to 'continuous' or generalised EVal.

Let us generalise (10.44) and define a spectral family (or a decomposition of the identity) as a family of orthogonal projectors $E(\lambda)$, $\lambda \in \mathbb{R}$ with the following properties:

$$E(\lambda)E(\mu) = E(\min\{\mu, \lambda\}) \tag{10.47a}$$

$$E(-\infty) = 0 \qquad\qquad E(\infty) = \mathbb{I} \tag{10.47b}$$

$$E(\lambda) = E(\lambda + \epsilon) \tag{10.47c}$$

The equal signs are meant in the strong sense, e.g. the second identity of (10.47b) means $\lim_{\lambda\to\infty} E(\lambda)|\psi\rangle = |\psi\rangle$ for all $|\psi\rangle \in H$. In (10.47c) the limit $\epsilon \to 0$ with $\epsilon > 0$ is understood, in agreement with out notation on Section 1 and Section 2, i.e. $\lim_{\epsilon\to 0^+} E(\lambda+\epsilon)|\psi\rangle = E(\lambda)|\psi\rangle$. While (10.44) was obtained having in mind an operator with a discrete spectrum, (10.47) makes no reference to the nature of the spectrum at all. To illustrate this generalisation, in Figure 10.2 the spectral family related to (10.44) (assuming the sum runs over 4 EVal) is contrasted with a more general spectral family that corresponds to an operator with 3 EVal and a non-trivial continuous spectrum. The spectral family has a discontinuity at each EVal. According to (10.47c) the limit from above is continuous, but the limit from below is not continuous if λ is an EVal.

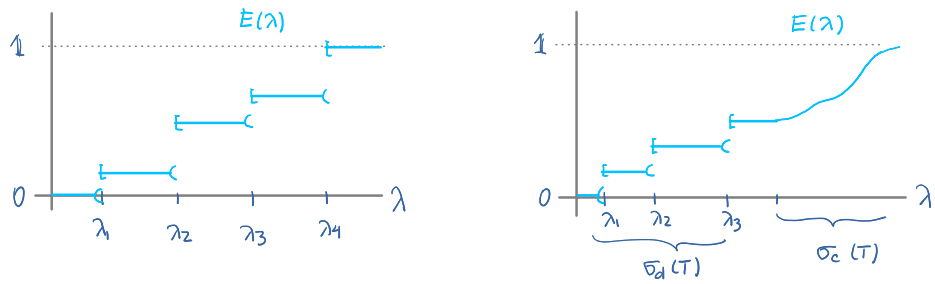

Figure 10.2: Left panel: spectral family $E(\lambda)$ of an operator with a purely discrete spectrum $\sigma(T) = \sigma_d(T) = \{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$. Right panel: spectral family for an operator with a mixed spectrum $\sigma(T) = \sigma_d(T) \cup \sigma_c(T)$.

With the generalisation of the spectral family from (10.44) to (10.47) we can now also generalise the spectral representation of an operator T from (10.46) to

$$T = \int \lambda \, dE_\lambda \qquad \text{and} \qquad f(T) = \int f(\lambda) \, dE_\lambda \tag{10.48}$$

This is a continuous sum over operators $dE_{\lambda_i} = \lim_{\epsilon\to 0^+}(E(\lambda_i) - E(\lambda_i - \epsilon))$ with $\epsilon \to 0^+$ indicating $\epsilon > 0$. The dE_λ can be seen as the infinitesimal limits of the ΔE_i in (10.46).

Every self-adjoint operator can be expressed uniquely through a spectral family (10.47). If the operator is bounded, the domain of integration over λ is bounded as well. For an unbounded operator the integration domain starts at $\lambda = -\infty$ or runs to $\lambda = +\infty$ (or both).

Taking the spectral family in the right panel of Figure 10.2 as an example let us see what happens if we perform the integration in (10.48). For $\lambda < \lambda_1$, $E(\lambda)$ is constant and, hence, $dE_\lambda = 0$. Thus there is no contribution to the integral from this region. At $\lambda = \lambda_1$ there is a discontinuity in $E(\lambda)$. This is reminiscent of the $\Theta(\lambda - \lambda_1)$ function of Section 2.3 and in

analogy to (2.21) we get a contribution to dE_λ akin to a delta distribution $\delta(\lambda - \lambda_1)P_1$. But now we are not dealing with functions but with operators. Thus, dE_λ is an operator valued distribution. If we increase λ beyond λ_1 again $dE_\lambda = 0$ until we reach λ_2. At each EVal λ_i, there is a $\delta(\lambda - \lambda_i)P_i$ contribution to the integral in (10.48). Once we are above the largest EVal, $\lambda > \lambda_3$, $E(\lambda)$ remains constant until we reach the continuous spectrum at which point it increases continuously. It has to increase according to (10.47a).

To put (10.48) into context we return to the position operator X in $L^2[0,1]$. The spectral family $E(\lambda)$ in this case is defined as $E(\lambda)|\psi\rangle = |\psi\rangle$ for $x \leq \lambda$ and $E(\lambda)|\psi\rangle = 0$ for $x > \lambda$, or $E(\lambda)|\psi\rangle = \Theta(\lambda - x)|\psi\rangle$. It should be obvious that this definition satisfies (10.47). In accordance with our observation above that the position operator has no EVal, its spectral family has no discontinuities. In a figure similar to Figure 10.2, $E(\lambda)$ would be a straight line from $E(\lambda) = 0$ at $\lambda = 0$ to $E(\lambda) = \mathbb{I}$ at $\lambda = 1$.

In quantum mechanics you will often work with a potential $V(x)$ that is an operator expressed in terms of the position operator X. This is to be interpreted exactly as in (10.48), with $T \to X$ and $f(T) \to V(x)$.

10.10 Quantum mechanics

In this section it is assumed that some encounter with QM has taken place in other lectures. For us, this serves as a playground to apply some of the concepts we have learned in Section 10. The reason is that in QM a physical system is described through a normalised vector $|\psi\rangle$ in a Hilbert space H, i.e. $\|\psi\| = 1$. In this context, the vector is often called a *state [Zustand]*, but mathematically it is just a vector. Physical quantities are represented through self-adjoint operators O which in this context are often called *observables [Observable]*. If the system is in in the state $|\psi\rangle$ and we perform a measurement of O we will get as a result a value in $\sigma(O)$. If the spectrum is purely discrete it is one of the EVal of O. In general, we do not know which EVal we get, but we can give the probability $w(\lambda_i)$ of measuring λ_i as

$$w(\lambda_i) = \langle\psi|P_i|\psi\rangle \geq 0 \qquad \in \mathbb{R} \tag{10.49}$$

where P_i is the projector onto the eigenspace of the EVal λ_i. Using the decomposition of the identity you should easily see that the probability to measure any EVal is equal to 1. If many equivalent systems are prepared in the state $|\psi\rangle$ and we average the results of measuring O in all these systems, we get the *expectation value [Erwartungswert]* of O in the state $|\psi\rangle$, ⇒ lecture

$$\langle O\rangle_\psi \equiv \langle\psi|O|\psi\rangle = \sum_{\lambda_i} \lambda_i \langle\psi|P_i|\psi\rangle \in \mathbb{R} \tag{10.50}$$

where we used the spectral decomposition of O, (10.46). As a simple-minded analogue consider rolling a dice. A roll corresponds to a measurement. Each time you do this, you get an 'EVal' of the dice, i.e. $\lambda_n = n$ with $n \in \{1,2,3,4,5,6\}$. You cannot say for sure which EVal you will get. But assuming the dice is fair, the probability to measure e.g. $\lambda = 4$ is $w(\lambda = 4) = 1/6$ and the expectation value is 3.5.[7] ⇒ lecture

[7] As a side remark we mention that in QM, after a measurement the system is in an eigenstate of O. Thus, if you roll a QM-dice and get say $\lambda = 2$, then this dice is in an eigenstate to the EVal $\lambda = 2$. If you roll the same dice again, you are guaranteed to get $\lambda = 2$. To get the expectation value, you must take many different but equivalent dice and roll each of them once, rather than rolling the same dice again and again.

Which Hilbert space we have to take depends on the observable and the system we are interested in describing. We will make two examples, one with a finite-dimensional Hilbert space $\mathbb{C}^2$ and one with an infinite-dimensional Hilbert space $L^2[-\pi,\pi]$.

As a first example we consider a spin 1/2 particle and are interested only in describing the spin (i.e. we do not care where the particle is, what its momentum is etc.). Such a system is described through a state $|\psi\rangle \in \mathbb{C}^2$, i.e $|\psi\rangle = (a,b)$ with $|a|^2 + |b|^2 = 1$. Just in passing we note that such a spin-vector is also often called *spinor [Spinor]*. Next we need the operators that describe the spin. We use the letter S rather than O. The spin operator is a *vector operator* which for us at this moment simply means that it has three components, a x, y, and z component, denoted by S_x, S_y and S_z. It is defined in terms of the Pauli matrices σ_x, σ_y, and σ_z as

$$S_x = \frac{1}{2}\sigma_x = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad S_y = \frac{1}{2}\sigma_y = \frac{1}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad S_z = \frac{1}{2}\sigma_z = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{10.51}$$

Of course, it is purely a convention how to assign the three matrices to the three coordinates x, y, and z, and we have set $\hbar = 1$. The Pauli matrices are hermitean ensuring the spin operator is self-adjoint. Each of the three operators in (10.51) has two EVal, $\lambda_1 = 1/2$ and $\lambda_2 = -1/2$ and corresponding EV $|\varphi_i\rangle$. Focusing on S_z, $|\varphi_{z1}\rangle = (1,0)$ and $|\varphi_{z2}\rangle = (0,1)$ and its spectral representation is given as

$$S_z = \frac{1}{2}\,|\varphi_{z1}\rangle\langle\varphi_{z1}| + \frac{-1}{2}\,|\varphi_{z2}\rangle\langle\varphi_{z2}| = \frac{1}{2}\,P_1 + \frac{-1}{2}\,P_2 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \frac{-1}{2}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \tag{10.52}$$

For $|\psi\rangle = (a,b)$ the probability to obtain $\pm 1/2$ in a measurement of S_z is given as $w(1/2) = \langle\psi|P_1|\psi\rangle = |a|^2$ and $w(-1/2) = \langle\psi|P_2|\psi\rangle = |b|^2$. For $|\psi\rangle = |\varphi_{z1}\rangle$ we are certain to obtain $+1/2$ when measuring S_z, as $w(1/2) = \langle\varphi_{z1}|P_1|\varphi_{z1}\rangle = 1$ and $w(-1/2) = \langle\varphi_{z1}|P_2|\varphi_{z1}\rangle = 0$. Sometimes this is referred to as $|\varphi_{z1}\rangle$ having a sharp value for S_z. We can also prepare a state that yields for sure $+1/2$ when measuring S_x (i.e. with a sharp value for S_x), namely the EV of S_x to the EVal 1/2, $|\varphi_{x1}\rangle = (1,1)/\sqrt{2}$. Can we prepare a state with a sharp value for S_x and S_z? No! Since S_x and S_z do not commute there is no common basis of EV, as discussed in Section 10.1 after (10.10). Thus, there is always a minimal uncertainty in measuring S_x times S_z. Finally, we mention that

$$S^2 = S_x^2 + S_y^2 + S_z^2 = \frac{3}{4}\,\mathbb{I} = s(s+1)\,\mathbb{I} \tag{10.53}$$

with $s = 1/2$. We will come back to this relation in Section 11.

In our second example we consider a particle trapped in a (infinite) one-dimensional potential well within the range $-\pi \le x \le \pi$ and are interested in its momentum. The Hilbert space is $L^2[-\pi,\pi]$ and the momentum operator with the conventions of Section 7.6 is $P = -i\,d/dx$. To ensure that P is self-adjoint we take as its domain $\mathcal{D}_P$ the differentiable functions with symmetric boundary conditions, i.e. $\psi(-\pi) = \psi(\pi)$ (see our discussion around (10.23) regarding the operator A_s). We can immediately find normalised EV as

$$|\varphi_n\rangle = \frac{1}{\sqrt{2\pi}}\,e^{i\,nx} \qquad\qquad P|\varphi_n\rangle = n|\varphi_n\rangle \qquad\qquad n \in \mathbb{Z} \tag{10.54}$$

where the last condition $n \in \mathbb{Z}$ is due to the symmetric boundary conditions. The fact that boundary conditions force an initially continuous parameter to become discrete is familiar to us from finding solutions to PDE. This puts the 'quantum' into QM.

The EV found in (10.54) form a ONB, since $\langle\varphi_m|\varphi_n\rangle = \delta_{mn}$ and any $|\psi\rangle \in L^2[-\pi,\pi]$ can be expanded as

$$\psi(x) \sim |\psi\rangle = \sum_n |\varphi_n\rangle\langle\varphi_n|\psi\rangle = \sum_n c_n|\varphi_n\rangle \quad \text{with} \quad c_n = \frac{1}{\sqrt{2\pi}}\int_{-\pi}^{\pi} e^{-inx}\psi(x)dx \tag{10.55}$$

This is nothing but the Fourier series of the function $\psi(x)$, (3.16), here to be interpreted as the expansion of the vector $|\psi\rangle \in L^2[-\pi,\pi]$ in the basis $|\varphi_n\rangle$ that is the eigenbasis of the self-adjoint operator $P = P^\dagger$. From (10.54) it is clear the P is not bounded, but it has a compact resolvent and a discrete spectrum $\sigma(P) = \sigma_d(P) = \mathbb{Z}$. In analogy to (10.52) we can write its spectral representation as[8]

$$P = \sum_n n\, P_n = \sum_n n\, |\varphi_n\rangle\langle\varphi_n| = \int_{-\infty}^{\infty} \lambda\, dE_\lambda \tag{10.56}$$

In the last step we have used the (for this case unnecessary) generalisation to write P in terms of its spectral family $E(\lambda)$ that has the form (10.44) with $i \to n$ and $\lambda_i = n$.

If the system is in the state $|\psi\rangle \sim \psi(x)$ and a measurement of the momentum is made, the probability to measure k is $w(k) = \langle\psi|P_k|\psi\rangle = |\langle\varphi_k|\psi\rangle|^2$. Even though here this involves doing an integral, whereas in the case of S_z we had to do matrix multiplication to get the probability, it is crucial to realise that this is indeed precisely the same formal procedure for S_z and P.

Since we have now many EVal we can also ask the question what is the probability to measure anything in the range say $k_1 \le n \le k_2$. Of course, this is just the sum of probabilities $w(n)$ given above, but using the spectral family it can be written as ⇒ lecture

$$\sum_{n=k_1}^{k_2} w(n) = \langle\psi| \sum_{n\in[k_1,k_2]} P_n\, |\psi\rangle = \langle\psi| \int_{k_1}^{k_2} dE_\lambda\, |\psi\rangle \tag{10.57}$$

Similarly, the expectation value of a momentum measurement is

$$\langle P\rangle_\psi = \langle\psi| \sum_n nP_n\, |\psi\rangle = \langle\psi| \int_{-\infty}^{\infty} \lambda\, dE_\lambda\, |\psi\rangle \tag{10.58}$$

in analogy to (10.50).

As a twist to this last example we are now interested in the position of the particle. Thus, we consider the position operator X in $L^2[-\pi,\pi]$. This operator is bounded with $\|X\| = \pi$. As already mentioned in our discussion at the very end of Section 10.9 there are no EV for this operator and, hence, no spectral representation as sum over projectors onto eigenspaces can be written. But with the spectral family for X we get as close a possible to such a representation. Indeed $\Delta E_\lambda = E(\lambda) - E(\lambda - \Delta)$ for small Δ is nearly a projector onto an eigenspace. Since $\Delta E_\lambda|\psi\rangle$ results in a state that is 0 unless $\lambda - \Delta \le x \le \lambda$, we have $X\,\Delta E_\lambda|\psi\rangle \simeq \lambda|\psi\rangle$. Only in the limit $\Delta \to 0$ this would become an EV if the result was still an element of the Hilbert space. This is consistent with our observation that $\delta(\lambda - x)$ is a generalised (but not proper) EV of X with generalised EVal λ.

[8] Another unfortunate clash of notation between the momentum operator P and the projectors P_n onto the eigenspace of P with EVal n.

Still, we can use the spectral representation $X = \int_{-\pi}^{\pi} \lambda dE_\lambda$ to answer the questions we have answered for S_z and P. If the system is in a state $|\psi\rangle \sim \psi(x)$ what is the probability to measure some value $-\pi \leq x_1 \leq x \leq x_2 \leq \pi$ when making a position measurement? Following (10.57) it is

$$w([x_1, x_2]) = \langle\psi| \int_{x_1}^{x_2} dE_\lambda \, |\psi\rangle = \int_{x_1}^{x_2} |\psi(x)|^2 dx \tag{10.59}$$

What is the probability to measure a fixed value x_0? It is zero, since the integration in (10.59) would be over a domain of measure 0. This is in contrast to the momentum operator, where the spectral family had 'delta' peaks or, equivalently, there were true EVal. Following (10.58) the expectation value of a position measurement is

$$\langle X\rangle_\psi = \langle\psi| \int_{-\pi}^{\pi} \lambda \, dE_\lambda \, |\psi\rangle = \int_{-\pi}^{\pi} x|\psi(x)|^2 dx \tag{10.60}$$

While we will never obtain a true eigenstate of the position operator, we can in principle make it arbitrarily sharp (see again Figure 2.1). However, if the position becomes very sharp, there is a minimal uncertainty in the momentum, since X and P do not commute, (7.78). Finally, if we are given a potential operator $V(X)$ we get

$$\langle V(X)\rangle_\psi = \langle\psi| \int_{-\pi}^{\pi} V(\lambda) \, dE_\lambda \, |\psi\rangle = \int_{-\pi}^{\pi} V(x)|\psi(x)|^2 dx \tag{10.61}$$

where we simply used (10.48).

10.11 Unitary operators and time evolution

We have spent a lot of time to understand self-adjoint operators. Another class of operators that play an important role are unitary operators. We will encounter them in Section 11 often related to symmetry considerations. A prominent application within QM is the time evolution of states. In Section 10.10 we described states at a particular time and a typical question is if the system is in the state $|\psi, t_0\rangle \sim \psi(x, t_0)$ at time t_0 in what state is it at a later time t. The formal answer is that $|\psi, t\rangle = U(t, t_0)|\psi, t_0\rangle$ where $U(t, t_0)$ is a unitary operator. Thus, time evolution is nothing but a unitary transformation.

Much that has been said about spectral representations on self-adjoint operators $T = T^\dagger$ can be translated to unitary operators $U^{-1} = U^\dagger$. Whereas for $T = T^\dagger$ the spectrum was real, $\lambda \in \mathbb{R}$, for unitary U the spectrum is in the complex plane with $|\lambda| = 1$.

In fact, we typically consider unitary operators that can be written as e^{iT} where $T = T^\dagger$. For such operators, we simply use (10.48) and can apply everything we have learned. For our time evolution operator in QM we have

$$U(t, t_0) = e^{-i(t-t_0)H} \tag{10.62}$$

where H is the (time-independent) self-adjoint Hamilton (or energy) operator of the system.[9] Thus, the operator H generates time evolution. We will soon realise that this concept is much more general and (10.62) is but one example of an operator H generating a (symmetry) transformation U.

[9] And not the Hilbert space. Why do they both have to start with the letter H?

11 Groups

For this chapter there is currently no script. Please follow one of the numerous books and/or texts on the internet.

11.1 Basics

- definitions: (non/abelian) group, subgroup, invariant or normal subgroups
- finite groups: cyclic C_n, symmetric S_n, dihedral D_n
- infinite groups: $O(n)$, $SO(n)$, $U(n)$, $SU(n)$ [Euclid and Poincare]
- direct product

11.2 Representations

- definitions: representation, irreducible representation
- character(tables), Schur's lemma
- relevance of representations in physics

11.3 A case study: representations of D_3

- build character table of (finite) group $S_3 = D_3$
- Clebsch-Gordan decomposition of representations

11.4 $SO(2)$ as Lie group

- definition: Lie group, generators
- irreducible representations of $SO(2)$

11.5 $SO(3)$, rotations, and $SU(2)$

- irreducible representations of $SO(3)$ and $SU(2)$
- generators, Lie algebra
- (addition of) angular momentum
- extension to $SU(3)$

11.6 $SU(N)$ and $SU(3)$ in particular

- basics of $SU(N)$
- $SU(3)$, its Lie algebra and Gell Mann matrices
- the simplest irreducible representations: 1, 3, 3^*, 8 ...

12 Tensors

12.1 The Euclidean group

We will work in the 3-dimensional Euclidean space $\mathbb{R}^3$ but most statements can be generalised to $\mathbb{R}^n$. For the time being we will denote the elements of $\mathbb{R}^3$ by $\vec{x}$, $\vec{y}$ etc. We use the canonical scalar product and its induced metric.

The Euclidean group E_3 consists of all 'motions' in $\mathbb{R}^3$ that leave $\|\vec{x}-\vec{y}\|$ invariant $\forall \vec{x}, \vec{y} \in \mathbb{R}^3$. This consists of transformations

$$\vec{x} \mapsto \vec{x}\,' = R\vec{x} + \vec{a} \qquad \text{with} \qquad \vec{a} \in \mathbb{R}^3 \qquad \text{and} \qquad R \in O(3)\,. \tag{12.1}$$

Thus, an element of E_3 can be written as $(\vec{a}, R) \in E_3$. Note that $\det(R) = -1$ is allowed, i.e. we do not insist on $R \in SO(3)$. In particular, we allow for reflections (12.14). Multiplication in E_3 is defined as subsequent transformation, i.e.

$$(\vec{a}\,', R') \cdot (\vec{a}, R) = (R'\vec{a} + \vec{a}\,', R'\,R) \tag{12.2}$$

the inverse is $(\vec{a}, R)^{-1} = (-R^{-1}\vec{a}, R^{-1})$ and the one element of E_3 is $e = (\vec{0}, 1)$.

The group E_3 has rotations $(\vec{0}, R)$ and translations $(\vec{a}, 1)$ as subgroups. The latter are actually an invariant (and abelian) subgroup. Another invariant subgroup is formed by the *proper* transformations *[eigentliche Bewegungen]*

$$E_3^+ = \{(\vec{a}, R) | \vec{a} \in \mathbb{R}^3, R \in SO(3)\} \tag{12.3}$$

The name 'proper' stems from the fact that if we move a physical object around in our everyday world, we perform such a transformation. Reflections are not included in E_3^+. They are part of the Lie group $E_3^- = \{(\vec{a}, R) |\, \vec{a} \in \mathbb{R}^3, R \in O(3), \det(R) = -1\}$. In fact, E_3^- is a left (and right) coset w.r.t. E_3^+ and, therefore, any element of E_3 is either in E_3^+ or in E_3^-.

A general transformation in E_3^+ is described in terms of 6 parameters, three rotation angles (that we write as a vector) $\vec{\theta}$ and three parameters of the translation $\vec{a}$. The three generators J_i related to rotations we have encountered already when discussing $SO(3)$. They are intimately connected to angular momentum. The three generators P_i related to translations are intimately connected to momentum.

$$\text{rotations:} \qquad R(\vec{\theta}) = e^{-i\vec{\theta}\cdot\vec{J}} \tag{12.4a}$$

$$\text{translation:} \qquad T(\vec{a}) = e^{-i\vec{a}\cdot\vec{P}} \tag{12.4b}$$

The Lie algebra reads

$$[P_i, P_j] = 0 \tag{12.5a}$$

$$[J_j, J_k] = i\,\epsilon_{jkl} J_l \tag{12.5b}$$

$$[P_j, J_k] = i\,\epsilon_{jkl} P_l \tag{12.5c}$$

where summation over repeated indices is understood. Here, (12.5a) simply expresses the abelian nature of translations, (12.5b) is familiar from $SO(3)$ and (12.5c) is a geometric property that can be understood by considering swapping the order e.g. in P_1 and J_3 . $\Rightarrow$ lecture

As a cross check for (12.5) we can verify that the infinite dimensional representation of the generators as differential operators

$$P_k = -i\,\partial_k \qquad\qquad J_k = \epsilon_{klm} x_l P_m = -i\,\epsilon_{klm} x_l \partial_m \tag{12.6}$$

satisfy the Lie algebra. The relation

$$\epsilon_{ijk}\epsilon_{lmk} = \delta_{il}\delta_{jm} - \delta_{im}\delta_{jl} \tag{12.7}$$

is extremely useful for this and similar calculations.

There are two Casimir operators, $P^2 \equiv P_i P^i$ and $\vec{J}\cdot\vec{P} \equiv J_i P^i$, since

$$[P^2, P_l] = [P^2, J_l] = 0 \qquad \text{and} \qquad [\vec{J}\cdot\vec{P}, P_l] = [\vec{J}\cdot\vec{P}, J_l] = 0 \tag{12.8}$$

We close this subsection with a remark on signs. (12.1) can be seen as an active transformation, whereby an object (or point) is moved from $\vec{x}$ to $\vec{x}\,'$. The matrix R then corresponds to the matrix used in Section 11. Alternatively, we can consider passive transformations. The object (or point) is left alone, but the coordinate system is rotated or translated. This will be our point of view in the follwing subsection. Then (12.1) and (12.4) correspond to shifting the coordinate system by $-\vec{a}$ and rotating it by $-\vec{\theta}$.

12.2 Tensors and coordinate transformations

Consider a point Q in $\mathbb{R}^3$. Once a (canonical, orthonormal) basis has been chosen, with basis vectors $\vec{e}_i$, $i \in \{1,2,3\}$ the position of Q is described by a vector $\vec{x}$, with components x^i such that

$$\vec{x} = \sum_{i=1}^{3} x^i\,\vec{e}_i = x^i\,\vec{e}_i \tag{12.9}$$

where from now on the sum over the repeated indices is always understood. Upper indices in x^i indicate the row.

If we make a coordinate transformation, more precisely a rotation of the coordinate system about the origin (also called a passive rotation since the point Q remains fixed) the 'old' basis vectors $\vec{e}_i$ are transformed into 'new' basis vectors $\vec{e}_i\,'$. The position vector of the same point Q can be written as

$$\vec{x} = x^i\,\vec{e}_i = (x')^i\,\vec{e}_i\,' \qquad \text{with} \qquad (x')^j = R^j{}_i x^i \tag{12.10}$$

The 'old' and 'new' coordinates of the point are related through matrix multiplication of the 3×3 matrix $R^j{}_i$ that corresponds to the defining representation of $SO(3)$. For a rotation about the axes $\hat{n}$ through an angle $-\theta$ it can be expressed as in (12.4a), where $\vec{\theta} = \theta\,\hat{n}$. Note that this matrix is the inverse of the matrix describing an active rotation, i.e. a rotation of the point Q. We can also consider a translation of the coordinate system and (12.4a) corresponds to a translation by $-\vec{a}$.

A 'true' vector as in (12.9) is sometimes also called a *contravariant [kontravariant]* vector and corresponds to a single column with $1 \le i \le 3$ rows. We can also define a co-vector or *covariant [kovariant]* vector in the dual space $\tilde{\mathbb{R}}^3$. This corresponds to a transposed vector

$\vec{y}^{\,T}$, a single row with $1 \leq i \leq 3$ columns. For such a co-vector we use lower indices y_i to label the columns. According to Riesz any (linear) functional in $\mathbb{R}^3$ can be expressed as a scalar product

$$\vec{y}^{\,T}\,\vec{x} = \vec{y}\cdot\vec{x} = y_i\,x^i = \langle y|x\rangle \tag{12.11}$$

which correspond to the usual matrix multiplication, rows times columns. The summation over repeated indices (one lower, one upper) is called *contraction [Kontraktion / Verjüngung]*. To understand its significance let us consider the transformation property of y_j which is ⇒ lecture

$$(\vec{y}\,')^T = \big(R\cdot\vec{y}\big)^T = \vec{y}^{\,T}\cdot R^T \qquad \text{hence} \qquad y'_j = R_j{}^k y_k \tag{12.12}$$

Let us now look at the transformation property of the scalar product (12.11). Under a coordinate transformation we have

$$y_i\,x^i \mapsto (y')_j\,(x')^j = y_k\,R_j{}^k R^j{}_i\,x^i = y_k\,\delta^k_i\,x^i = y_i\,x^i \tag{12.13}$$

In the second to last step we used the orthogonality of R, i.e. $R^\dagger\cdot R = R^T\cdot R = \mathbb{I}$, expressed in component notation. Thus, the scalar product is invariant under coordinate transformations, as we knew all along.

The transformation properties (12.10) and (12.12) as well as the invariance of the scalar product (12.13) also holds if we replace the rotation R by a *reflection [Spiegelung]*, represented by

$$\vec{x} \mapsto \vec{x}\,' = S\,\vec{x} = -\vec{x} \qquad \text{or} \qquad (x')^j = S^j{}_i x^i = -x^j \tag{12.14}$$

where $S = -\mathbb{I}$ with $\det(S) = -1$. Thus, in all equations above we can think of $R \in O(3)$. Whereas 'proper' rotations with $R \in SO(3)$ belong to E_3^+, reflections are part of E_3^-.

We are now ready to classify (classical) physical quantities according to their transformation properties under $(\vec{0}, R) \subset E_3$. To this end we use

$$O(3) = SO(3)\otimes Z_2 = SO(3)\otimes\{1,-1\} \tag{12.15}$$

which is another way to express the split of orthogonal matrices into 'proper' rotations with $\det(R) = 1$ and 'improper' rotations with $\det(R) = -1$. Now all we need to do is to specify the (preferably) irreducible representation Γ_j of $SO(3)$ and the irreducible representation of Z_2 under which the physical quantity transforms. The former we have worked out in Section 11 and for classical quantities we have $\forall j \in \mathbb{N}$ a $(2j+1)$ dimensional irreducible representation Γ_j. In quantum mechanics we also have to allow for half-integer j. For the abelian group Z_2 we have two one-dimensional irreducible representations Γ_+ and Γ_-. ⇒ lecture

A *scalar [Skalar]* is a single quantity s that does not transform (or transforms trivially or is invariant) under $(\vec{0}, R)$, i.e

$$\text{scalar:} \qquad s \overset{R}{\mapsto} s \qquad s \overset{S}{\mapsto} s \qquad \Gamma_0\otimes\Gamma_+ \tag{12.16}$$

So it is not entirely surprising that the scalar product is a scalar. We do explicitly state the transformation under a reflection, even though it is included in the transformation property under a general $R \in O(3)$.

A *pseudoscalar [Pseudoskalar]* is a single quantity $\tilde{s}$ that transform under $(\vec{0}, R)$ as

$$\text{pseudoscalar:} \qquad \tilde{s} \overset{R}{\mapsto} \det(R)\tilde{s} \qquad \tilde{s} \overset{S}{\mapsto} -\tilde{s} \qquad \Gamma_0 \otimes \Gamma_- \tag{12.17}$$

A (contravariant) *vector [Vektor]* is an entity x^i made up of three entries that transform under $(\vec{0}, R)$ as

$$\text{vector:} \qquad x^i \overset{R}{\mapsto} R^i{}_j x^j \qquad x^i \overset{S}{\mapsto} -x^i \qquad \Gamma_1 \otimes \Gamma_+ \tag{12.18a}$$

$$\text{co-vector:} \qquad x_i \overset{R}{\mapsto} R_i{}^j x_j \qquad x_i \overset{S}{\mapsto} -x_i \qquad \Gamma_1 \otimes \Gamma_+ \tag{12.18b}$$

For every contravariant vector as an element of $\mathbb{R}^3$ we have a covariant vector or co-vector as an element of the dual space $\tilde{\mathbb{R}}^3$. Note that this definition of 'vector' is a physicist's definition. So far, a vector was an element of a vector space. This is the mathematician's definition of a vector. In the latter sense, a scalar is a vector in the vector space $\mathbb{R}$. When it comes to transformation properties and the classification of physical observables, we (i.e. the physicists) use the term 'vector' is a more specialised way, namely as defined in (12.18). And we say s is not a vector, ignoring the (justified) complaints of mathematicians. It is just a matter of conventions for the language used that are adapted to the situation.

A (contravariant) *axial vector [Axialvektor]*, sometimes also called pseudovector, is an entity x^i made up of three entries that transform under $(\vec{0}, R)$ as

$$\text{axial vector:} \qquad x^i \overset{R}{\mapsto} \det(R)\, R^i{}_j x^j \qquad x^i \overset{S}{\mapsto} x^i \qquad \Gamma_1 \otimes \Gamma_- \tag{12.19}$$

We now generalise this and introduce the notion of a *tensor [Tensor]*. A scalar can also be called a tensor of rank 0, as it consists of an entity with $3^0 = 1$ entries and, hence, has 0 (contravariant+covariant) labels. A vector is a tensor of rank 1, as it consists of an entity with $3^1 = 3$ entries and, hence, has 1 (contravariant+covariant) label. Now guess what a tensor of rank 2 is. Well, done! It is indeed an entity M^{ik} with $3^2 = 9$ entries and, hence, has 2 (contravariant+covariant) labels. It can be represented as a contravariant, a covariant or a mixed rank 2 tensor as

$$\text{tensor rank 2:} \qquad M^{ik} \overset{R}{\mapsto} R^i{}_j R^k{}_l M^{jl} \qquad M^{ik} \overset{S}{\mapsto} M^{ik} \qquad (\Gamma_1 \otimes \Gamma_1) \otimes \Gamma_+ \tag{12.20a}$$

$$\text{co-tensor rank 2:} \qquad M_{ik} \overset{R}{\mapsto} R_i{}^j R_k{}^l M_{jl} \qquad M_{ik} \overset{S}{\mapsto} M_{ik} \qquad (\Gamma_1 \otimes \Gamma_1) \otimes \Gamma_+ \tag{12.20b}$$

$$\text{mixed tensor rank 2:} \qquad M^i{}_k \overset{R}{\mapsto} R^i{}_j R_k{}^l M^j{}_l \qquad M^i{}_k \overset{S}{\mapsto} M^i{}_k \qquad (\Gamma_1 \otimes \Gamma_1) \otimes \Gamma_+ \tag{12.20c}$$

It should be obvious that the three equations above (and the two equations in (12.18)) contain precisely the same information. So there is no need to give them all. We can obtain one from the other simply by lowering or raising indices. Second, the representations in (12.20) are not irreducible. Using $\Gamma_1 \otimes \Gamma_1 = \Gamma_0 \oplus \Gamma_1 \oplus \Gamma_2$ we can split a tensor of rank 2 into quantities that transform under irreducible representations, i.e. quantities that consist of $2j+1$ entries. We can also define 'pseudo tensors' in analogy to pseudoscalars and axial vectors. As a final remark regarding (12.20) we note that the transformation of M can be written in matrix form as

$$M^{ik} \mapsto R^i{}_j M^{jl} (R^T)_l{}^k \qquad \text{hence} \qquad M \mapsto R \cdot M \cdot R^{-1} \tag{12.21}$$

where we have used $R^T = R^{-1}$. This also means that relations like $\vec{y} = M\,\vec{x}$ are not affected by coordinate transformations, i.e. if they hold with one choice of a basis, they will hold with any choice of basis. We will come back to this simple observation and appreciate its fundamental importance shortly.

Before we do so, let us generalise further. A tensor of rank n is an entity $M^{i_1 i_2 \ldots i_n}$ with 3^n entries and, hence, has n (contravariant+covariant) labels. As a fully covariant tensor it transforms as

$$\text{tensor rank } n: \qquad M^{i_1 \ldots i_n} \overset{R}{\mapsto} R^{i_1}{}_{j_1} \ldots R^{i_n}{}_{j_n} M^{j_1 \ldots j_n} \qquad \Big(\bigotimes_n \Gamma_1\Big) \otimes \Gamma_+ \tag{12.22}$$

Again, this representation is not irreducible and can be decomposed into irreducible representations leading to Γ_j with $j \leq n$. The corresponding quantities transform under $\Gamma_j(R) \otimes \Gamma_+$.

If we have a (mixed) tensor of rank n whereby one upper and one lower index coincide, summing over this index produces a tensor of rank $n-2$. This is a generalisation of the contraction mentioned in connection with (12.13). We note that

$$\delta^i{}_k \;\mapsto\; R^i{}_j R_k{}^l\, \delta^j{}_l = \delta^i{}_k \tag{12.23}$$

$$\epsilon^{klm} \;\mapsto\; R^k{}_i R^l{}_j R^m{}_n\, \epsilon^{ijn} = \det(R)\, \epsilon^{klm} \tag{12.24}$$

are an invariant tensor of rank 2 and an invariant (pesudo)tensor of rank 3. As a consequence, if we have two vectors $\vec{x}$ and $\vec{y}$ then

$$\big(\vec{x} \times \vec{y}\big)^i = \epsilon^i{}_{jk} x_j y_k \tag{12.25}$$

is an axial vector. ⇒ lecture

So far we have used the representations Γ_j with integer j. In particular, for tensors of rank 2 we reduced the product representation into irreducible representations of dimension 1, 3 and 5. Clearly, this procedure can be generalised to higher rank tensors and (12.22) will lead to Γ_j with $j \leq n$.

As a preparation for quantum mechanics we also mention a *spinor [Spinor]*, a quantity that transforms under $\Gamma_{1/2}$. More precisely, a spin 1/2 spinor ψ^i is an entity with $2 \cdot 1/2 + 1 = 2$ entries transforming as

$$\text{spinor:} \qquad \psi^i \;\overset{R}{\mapsto}\; \big(\Gamma_{1/2}(R)\big)^i{}_j\, \psi^j \tag{12.26}$$

The last remark in this subsection is the most important one. We have made a catalogue of transformation properties and can sort the physical quantities like mass, momentum, torque, electric field, you-name-it-here-is-it into these boxes. Why are we so obsessed with this? The reason is that physical laws take the same form in all (inertial) frames. In our special case of rotations, this means if a physical law LHS=RHS holds in a particular basis, then it holds in all bases that are obtained by a rotation (actually a translation is ok as well). Thus, in a new frame we have LHS'=RHS'. Do it again and we get LHS"+RHS". For this to be possible, the quantities that make up the LHS and RHS have to have well-defined transformation properties. They have to transform under a representation of $O(3)$. They have to be tensors! And obviously, the nature of the LHS tensor has to be precisely the same as the RHS tensor.

Let's try this. Take whatever physics law that comes to your mind. How about $\vec{F} = m\vec{a}$. The LHS is a vector, the RHS is a scalar m times a vector $\vec{a}$, hence a vector. If we make a

coordinate transformation we get $\vec{F}\,' = m\vec{a}\,'$. How about

$$\vec{L} = \vec{r} \times \vec{p} = I \cdot \vec{\omega} \tag{12.27}$$

Now $\vec{r}$ and $\vec{p}$ are vectors. Due to the vector product $\vec{L}$ is an axial vector, (12.25). I is a rank 2 tensor, $\vec{\omega}$ an axial vector. Two of the three indices in $I \cdot \vec{\omega}$ are contracted, leading to one free index, an axial vector. All three sides of (12.27) have the same tensor structure, an axial vector, as they must. Now take all equations you have ever written in physics (yes, all of them!) and check that the tensor structure on the LHS is the same as the tensor structure on the RHS. If you ever encounter an equation 'vector'='axial vector' or 'scalar' = 'third component of vector' you know the equation is wrong.

Fundamental laws of nature are written in terms of tensors.

We can extend these considerations to fields i.e. space dependent quantities $\phi(\vec{x})$. Under a coordinate transformation we have $\vec{x} \mapsto \vec{x}\,'$ and $\phi \mapsto \phi'$. As for the quantities considered above, also fields can have various transformation properties. The simplest field is a ⇒ lecture

$$\text{scalar field:} \qquad \phi(\vec{x}) = \phi'(\vec{x}\,') \qquad \text{or} \qquad \phi(\vec{x}) \mapsto \phi'(\vec{x}) = \phi(R^{-1}\,\vec{x}) \tag{12.28}$$

This is a field that does not transform or transforms under the trivial representation under R. Thus, the 'new' field ϕ' at the point Q, represented by the 'new' vector $\vec{x}\,'$ is the same as the 'old' field ϕ at *the same* point Q represented by the 'old' vector $\vec{x}$. A typical example of such a field is the temperature as a function of position.

A familiar example is the electric field $\vec{E}(\vec{x})$. The notation is a bit of a give away. This is a vector field. At each point Q we attach a vector $\vec{E}$ that transforms as in (12.18). Hence, for a

$$\text{vector field:} \quad \vec{E}(\vec{x}) \mapsto \vec{E}\,'(\vec{x}) = R \cdot \vec{E}(R^{-1}\vec{x}) \quad \text{or} \quad (E')^i(\vec{x}) = R^i{}_j E^j(R^{-1}\vec{x}) \tag{12.29}$$

Again, we can go on and define axial vector fields (an example is the magnetic field $\vec{B}$) or in general tensor fields.

Let us now consider the action of a translation or rotation on a scalar field. Starting with translations, where we just have to replace R by T in (12.28) and (12.29) we use (12.4b) with (12.6) and get ⇒ lecture

$$\phi'(\vec{x}) = e^{-i\,\vec{a}\cdot\vec{P}}\phi(\vec{x}) = e^{-\vec{a}\cdot\vec{\nabla}}\phi(\vec{x}) = \big(1 - \vec{a}\cdot\vec{\nabla} + \dots\big)\phi(\vec{x}) = \phi(\vec{x} - \vec{a}) = \phi(T^{-1}\,\vec{x}) \tag{12.30}$$

in agreement with (12.28). The same game can be played with rotations using (12.4a) and, in fact, it is sufficient to consider infinitesimal rotations (or translations).

12.3 The Poincaré and Lorentz group and Lorentz transformations

In this subsection we want to extend the transformations under the Euclidean group (which are closely related to Galilean transformations) to (homogeneous) Lorentz and (inhomogeneous) Poincaré transformations.

Let us start with the homogeneous transformations, i.e. (12.1) with $\vec{a} = 0$. This transformation mixes the three coordinates in $\vec{x} \sim x^i$ but does not affect the time coordinates. Laws of classical (in the sense of non-relativistic) physics are invariant under such transformations.

When generalising this concept to Lorentz transformations (LT), we replace Euclidean space $\mathbb{R}^3$ by Minkowski space. Vectors in Minkowski space are written as x^μ with $\mu = 0$ labelling the 'time' component and $\mu \in \{1, 2, 3\}$ the three space components and they are called 4-vectors. The generalisation of the 'scalar product' (12.13) reads

$$y_i\, x^i = \delta_{ij} y^j x^i \;\rightarrow\; y_\mu x^\mu = g_{\mu\nu} y^\nu x^\mu \tag{12.31}$$

where the 'metric' $g_{\mu\nu} = \mathrm{diag}(1, -1, -1, -1)$. Note that the terms 'scalar product' and 'metric' above are rather dangerous. In fact, (12.31) is *not* a scalar product in the sense of (9.20). In particular (12.31) does *not* satisfy (9.20c). However, as we shall see, (12.31) is a scalar product in the sense of (12.13), i.e. it is invariant under Lorentz transformations. Similarly, the term 'metric' for $g_{\mu\nu}$ takes getting used to. While δ_{ij} induces a proper metric $d(x, y) = \delta_{ij} y^j x^i$ for $x, y \in \mathbb{R}^3$ in the sense of (9.3), $g_{\mu\nu}$ does *not* do this. In particular, (9.3b) does *not* hold. Thus, the space of 4-vectors is not a Hilbert space and whenever we use the physicist's terms scalar product and metric we have to be aware of their shortcomings. Using $g_{\mu\nu}$ we can convert upper (contravariant) indices to lower (covariant) indices,

$$x_i = \delta_{ij} x^j \;\rightarrow\; x_\mu = g_{\mu\nu} x^\nu \tag{12.32}$$

but since $g_{\mu\nu}$ is not simply the unit matrix, upper and lower spacial indices differ by a sign.

Poincaré transformations (a^μ, Λ) or LT $(0, \Lambda)$ are generalisations of E_3 and in analogy with (12.1) we write

$$x^\mu \mapsto (x')^\mu = \Lambda^\mu{}_\nu\, x^\nu + a^\mu \tag{12.33a}$$

or

$$x_\mu \mapsto (x')_\mu = \Lambda_\mu{}^\nu\, x_\nu + a_\mu \tag{12.33b}$$

where the 4×4 matrices Λ are such that the 'scalar product' (12.31) is invariant under LT. This entails ⇒ lecture

$$y_\mu x^\mu \mapsto (y')_\mu (x')^\mu = y_\nu \Lambda_\mu{}^\nu \Lambda^\mu{}_\rho x^\rho \stackrel{!}{=} y_\mu x^\mu \qquad \text{or} \qquad \Lambda^T g\, \Lambda = g \tag{12.34}$$

This is the relativistic analogue of $R^T R = \mathbb{I}$ and also results in $\det(\Lambda)^2 = 1$. In addition $\Lambda_{00} \geq 1$ or $\Lambda_{00} \leq -1$. The proper orthochronous LT are those with $\det(\Lambda) = +1$ and $\Lambda_{00} \geq 1$ and they form a subgroup of all LT, denoted by $SO(3, 1)$. These are the LT continuously connected to the one element and consist of rotations and boosts, in analogy to the proper rotations of E_3^+. The corresponding 4×4 matrices Λ have the form

$$\begin{pmatrix} \cosh\psi & -\sinh\psi & & \\ -\sinh\psi & \cosh\psi & & \\ & & 1 & \\ & & & 1 \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} 1 & & & \\ & \cos\theta & \sin\theta & \\ & -\sin\theta & \cos\theta & \\ & & & 1 \end{pmatrix} \tag{12.35}$$

for boosts (along the x axis) and rotations (about the z axis), where $\psi = \operatorname{artanh}\beta$ is the boost parameter, i.e. $\cosh\psi = \gamma$ and $\sinh\psi = \beta\gamma$.

The homogeneous transformations of E_3^+ required three generators, corresponding to three independent rotations. For LT we need 6 parameters/generators, as in addition to the three rotations we have three independent boosts. The latter mix the 0 coordinate (time) with the spacial coordinates. Also, the generators and parameters of LT will have Lorentz indices μ, rather than only space indices i. We write them as $J_{\mu\nu} = -J_{\nu\mu}$ and $\omega_{\mu\nu} = -\omega_{\nu\mu}$. As a ⇒ lecture

preparation for this generalisation we rewrite (12.5b) in terms of $J_{kl} \equiv \epsilon_{klm} J_m$ (i.e. $J_{12} = J_3$ etc.) and obtain

$$[J_{kl}, J_{mn}] = i\big(\delta_{km} J_{ln} - \delta_{kn} J_{lm} - \delta_{lm} J_{kn} + \delta_{ln} J_{km}\big) \tag{12.36}$$

This allows us to write the commutation relations of rotations in terms of Lorentz indices. Furthermore, the generators of the boosts are written as $K_m \equiv J_{m0}$. With the help of (12.36) we can now generalise (12.5) to

$$[P_\mu, P_\nu] = 0 \tag{12.37a}$$
$$[J_{\mu\nu}, J_{\rho\sigma}] = i\big(g_{\mu\rho}\, J_{\nu\sigma} - g_{\mu\sigma}\, J_{\nu\rho} - g_{\nu\rho}\, J_{\mu\sigma} + g_{\nu\sigma}\, J_{\mu\rho}\big) \tag{12.37b}$$
$$[P_\mu, J_{\nu\rho}] = i\,\big(g_{\mu\rho}\, P_\nu - g_{\mu\nu}\, P_\rho\big) \tag{12.37c}$$

where for later reference we also included the commutation relations involving the generators P_μ of 4-translations.

Thus, our transformations are written as a generalisation of (12.4)

$$\text{Lorentz transformation:} \qquad \Lambda(\omega^{\mu\nu}) = e^{\frac{i}{2}\,\omega^{\mu\nu} J_{\mu\nu}} \tag{12.38a}$$
$$\text{translation:} \qquad T(a^\mu) = e^{i a^\mu P_\mu} \tag{12.38b}$$

The factor 1/2 in (12.38a) is a convention that takes into account that each term appears twice in the sum over the antisymmetric $\omega_{\mu\nu}$ and $J_{\mu\nu}$. It is remarkable that we can write (12.37b) in terms of

$$M_j \equiv \frac{1}{2}(J_j + i\,K_j) \qquad\qquad N_j \equiv \frac{1}{2}(J_j - i\,K_j) \tag{12.39}$$

as

$$[M_j, M_k] = i\epsilon_{jkl} M_l \tag{12.40a}$$
$$[N_j, N_k] = i\epsilon_{jkl} N_l \tag{12.40b}$$
$$[M_j, N_k] = 0 \tag{12.40c}$$

These are two (commuting) copies of the $SU(2)$ Lie algebra. Thus, irreducible representations of $SO(3,1)$ can be obtained from irreducible representations of $SU(2)_M \times SU(2)_N$.

Combining proper, orthochronous LT with

$$\begin{pmatrix} 1 & & & \\ & -1, & & \\ & & -1 & \\ & & & -1 \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} -1 & & & \\ & 1, & & \\ & & 1 & \\ & & & 1 \end{pmatrix} \tag{12.41}$$

we obtain all possible LT. In (12.41) the first transformation corresponds to the reflection S (12.14), often called parity operation, whereas the second corresponds to time reversal $t \mapsto t' = -t$.

As for the Euclidean space, we can now make a catalogue of possible transformation properties of physical quantities under LT. We start by writing the first few in terms of covariant 4-tensors:

$$\text{scalar:} \qquad s \overset{\Lambda}{\mapsto} s' = s \tag{12.42a}$$

$$\text{(contravariant) 4-vector:} \qquad x^\mu \overset{\Lambda}{\mapsto} (x')^\mu = \Lambda^\mu{}_\rho x^\rho \tag{12.42b}$$

$$\text{(contravariant) tensor rank 2:} \qquad F^{\mu\nu} \overset{\Lambda}{\mapsto} (F')^{\mu\nu} = \Lambda^\mu{}_\rho \Lambda^\nu{}_\sigma F^{\rho\sigma} \tag{12.42c}$$

With just a little imagination this can be generalised to a tensor of arbitrary rank. Writing physical laws in terms of 4-tensors ensures that the laws take the same form in all inertial frames, hence, satisfy Einstein's first postulate of special relativity.

A Curvilinear coordinates

In this section we provide a brief discussion of *curvilinear coordinates [krummlinige Koordinaten]* in general and spherical coordinates in particular. For efficiency reasons, we will sometimes use the generic notation of cartesian coordinates as $\vec{r} = (x_1, x_2, x_3)$ for $\vec{r} = (x, y, z)$. We will investigate the transformation to another system of coordinates $x_i \to u_i(x_i)$, whereby the new coordinate lines might be curved. Our prime example is spherical coordinates $u_1 = r$, $u_2 = \theta$, and $u_3 = \phi$. In this case, the relation between the 'new' and 'old' coordinates is (7.12)

$$\vec{r} = \sum_{i=1}^{3} x_i \, \vec{e}_i = \sum_{i=1}^{3} x_i(u_i) \, \vec{e}_i \tag{A.1a}$$

$$\vec{r} = x \, \vec{e}_x + y \, \vec{e}_y + z \, \vec{e}_z = r \, \sin\theta \, \cos\phi \, \vec{e}_x + r \, \sin\theta \, \sin\phi \, \vec{e}_y + r \, \cos\theta \, \vec{e}_z \tag{A.1b}$$

where (A.1b) corresponds to (A.1a) expressed in the standard notation for spherical coordinates. The three basis vectors in cartesian coordinates are generically denoted by $\vec{e}_i$, and specifically by $\vec{e}_x$, $\vec{e}_y$, $\vec{e}_z$. The basis vectors in the new coordinates, $\vec{e}_{u_i}$, are obtained by (infinitesimally) varying one new coordinate, say $u_i \to u_i + du_i$, while keeping the others fixed. After normalisation (the factor h_i below), this provides the new basis vectors

$$\vec{e}_{u_i} = \frac{\partial \vec{r}}{\partial u_i} \frac{1}{h_i} \qquad \text{with} \qquad h_i = \left| \frac{\partial \vec{r}}{\partial u_i} \right| \tag{A.2}$$

In general, the new basis vectors will depend on the position, i.e. the coordinate system changes from point to point.

Let us illustrate this for the case of spherical coordinates. In this case we have ⇒ lecture

$$\vec{e}_r = \sin\theta \, \cos\phi \, \vec{e}_x + \sin\theta \, \sin\phi \, \vec{e}_y + \cos\theta \, \vec{e}_z \tag{A.3a}$$

$$\vec{e}_\theta = \cos\theta \, \cos\phi \, \vec{e}_x + \cos\theta \, \sin\phi \, \vec{e}_y - \sin\theta \, \vec{e}_z \tag{A.3b}$$

$$\vec{e}_\phi = -\sin\phi \, \vec{e}_x + \cos\phi \, \vec{e}_y \tag{A.3c}$$

where we have used

$$h_r = \left| \frac{\partial \vec{r}}{\partial r} \right| = 1 \qquad h_\theta = \left| \frac{\partial \vec{r}}{\partial \theta} \right| = r \qquad h_\phi = \left| \frac{\partial \vec{r}}{\partial \phi} \right| = r \, \sin\theta \tag{A.4}$$

At each point in space, (A.3) defines locally a new orthonormal coordinate system,

$$\vec{e}_r \times \vec{e}_\theta = \vec{e}_\phi \qquad \vec{e}_\theta \times \vec{e}_\phi = \vec{e}_r \qquad \vec{e}_\phi \times \vec{e}_r = \vec{e}_\theta \tag{A.5}$$

While moving about, the coordinate system changes, as illustrated in Figure A.1. However, locally the coordinate transformation $x_i \leftrightarrow u_i$ is invertible and we can write

$$x_i = \sum_{j=1}^{3} R_{ij} \, u_j \tag{A.6}$$

where R is a (position dependent) orthogonal 3×3 matrix.

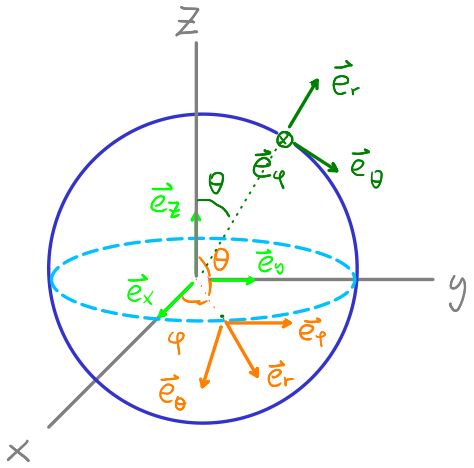

Figure A.1: Position independent system of cartesian basis vectors $\vec{e}_x$, $\vec{e}_y$ and $\vec{e}_z$ (light green) and position dependent system of spherical basis vectors $\vec{e}_r$, $\vec{e}_\theta$ and $\vec{e}_\phi$ (dark green and orange). At each point they are related through an orthogonal transfomation R.

Any vector and its total differential can now be written in either coordinate system as

$$\vec{r} = \sum_{i=1}^{3} x_i \, \vec{e}_i = \sum_{i=1}^{3} u_i \, \vec{e}_{u_i} \tag{A.7}$$

$$d\vec{r} = \sum_{i=1}^{3} dx_i \, \vec{e}_i = \sum_{i=1}^{3} h_i \, du_i \, \vec{e}_{u_i} \tag{A.8}$$

and the change of a (scalar) function $\psi(\vec{r})$ along $d\vec{r}$ is

$$\vec{r} \to \vec{r} + d\vec{r} \qquad \text{then} \qquad \psi(\vec{r}) \to \psi(\vec{r} + d\vec{r}) = \psi(\vec{r}) + d\psi(\vec{r}) + \dots \tag{A.9}$$

Identifying $d\psi$ as

$$d\psi(\vec{r}) = \vec{\nabla}\psi \cdot d\vec{r} \stackrel{!}{=} \frac{\partial \psi(u_i)}{\partial u_i} \, du_i \tag{A.10}$$

allows to extract the gradient expressed in the new coordinates as ⇒ lecture

$$\vec{\nabla} = \sum_{i=1}^{3} \vec{e}_{u_i} \, \frac{1}{h_i} \frac{\partial}{\partial u_i} \qquad \to \qquad \vec{\nabla} = \vec{e}_r \, \frac{\partial}{\partial r} + \vec{e}_\theta \, \frac{1}{r} \frac{\partial}{\partial \theta} + \vec{e}_\phi \, \frac{1}{r \, \sin\theta} \frac{\partial}{\partial \phi} \tag{A.11}$$

With (A.11) we can directly express the momentum operator $\vec{p} = -i\vec{\nabla}$ in spherical (or any other) coordinates. Also the Laplace operator in spherical coordinates, (7.13), can readily be computed. What is very important to realise in such derivations is that the new basis vectors depend on the coordinates, see (A.3). While for cartesian coordinates we can freely

interchange $\partial_x \vec{e}_x = \vec{e}_x \partial_x$ when acting on a function $\psi(\vec{r})$, such relations no longer hold for the new coordinate system and we have

$$\partial_\theta \vec{e}_\theta = -\vec{e}_r + \vec{e}_\theta \partial_\theta \qquad \partial_\phi \vec{e}_\theta = \cos\theta \, \vec{e}_\phi + \vec{e}_\theta \partial_\phi \tag{A.12}$$

$$\partial_\theta \vec{e}_\phi = \vec{e}_\phi \partial_\theta \qquad \partial_\phi \vec{e}_\phi = -\sin\theta \, \vec{e}_r - \cos\theta \, \vec{e}_\theta + \vec{e}_\phi \partial_\phi \tag{A.13}$$

$$\partial_\theta \vec{e}_r = \vec{e}_\theta + \vec{e}_r \partial_\theta \qquad \partial_\phi \vec{e}_r = \sin\theta \, \vec{e}_\phi + \vec{e}_r \partial_\phi \tag{A.14}$$

For the angular momentum operator we obtain ⇒ lecture

$$\vec{L} = \vec{r} \times \vec{p} = -ir \; \vec{e}_r \times \vec{\nabla} = -i\left(\vec{e}_\phi \frac{\partial}{\partial\theta} - \vec{e}_\theta \frac{1}{\sin\theta} \frac{\partial}{\partial\phi} \right) \tag{A.15}$$

and its individual components are

$$L_x = \vec{e}_x \cdot \vec{L} = i\left(\sin\phi \frac{\partial}{\partial\theta} + \cot\theta \, \cos\phi \frac{\partial}{\partial\phi} \right) \tag{A.16a}$$

$$L_y = \vec{e}_y \cdot \vec{L} = -i\left(\cos\phi \frac{\partial}{\partial\theta} - \cot\theta \, \sin\phi \frac{\partial}{\partial\phi} \right) \tag{A.16b}$$

$$L_z = \vec{e}_z \cdot \vec{L} = -i \frac{\partial}{\partial\phi} \tag{A.16c}$$

Often, the combinations

$$L_+ = L_x + i \, L_y = e^{i\,\phi} \left(\frac{\partial}{\partial\theta} + i \, \cot\theta \, \frac{\partial}{\partial\phi} \right) \tag{A.17a}$$

$$L_- = L_x - i \, L_y = e^{-i\,\phi} \left(-\frac{\partial}{\partial\theta} + i \, \cot\theta \, \frac{\partial}{\partial\phi} \right) \tag{A.17b}$$

are more convenient.

We can now use either $L^2 = \vec{L} \cdot \vec{L}$ with (A.15) or $L^2 = L_x^2 + L_y^2 + L_z^2$ with (A.16a) or ⇒ lecture
$L^2 = L_z^2 + (L_+ L_- + L_- L_+)/2$ with (A.17) to obtain

$$L^2 = -\frac{1}{\sin\theta} \frac{\partial}{\partial\theta} \sin\theta \frac{\partial}{\partial\theta} - \frac{1}{\sin^2\theta} \frac{\partial^2}{\partial\phi^2} \tag{A.18}$$

This should be compared with (7.17). Expressed in terms of $x = \cos\theta$ (beware of the clash of notation) we have

$$L^2 = -(1-x^2) \frac{\partial^2}{\partial x^2} + 2x \frac{\partial}{\partial x} - \frac{1}{1-x^2} \frac{\partial^2}{\partial\phi^2} = -\frac{\partial}{\partial x}(1-x^2) \frac{\partial}{\partial x} - \frac{1}{1-x^2} \frac{\partial^2}{\partial\phi^2} \tag{A.19}$$

which allows to make contact with the generalized Legendre equation (5.11) or (7.18).

Note $L^2 \neq L_z^2 + L_+ L_-$ since $L_+ L_- \neq L_- L_+$. With (A.18) we can rewrite (7.13) as

$$\Delta = \frac{1}{r} \frac{\partial^2}{\partial r^2} \, r - \frac{1}{r^2} L^2 \tag{A.20}$$

in a compact way.