



UNIVERSITY OF YORK
DEPARTMENT OF MATHEMATICS

Quantum & Continuum Dynamics

MAT00049I

Lecture notes (spring 2024)

Stefano NEGRO
stefano.negro@york.ac.uk

April 7, 2024

Contents

Foreword	1
I Waves (Continuum Dynamics)	4
1 Waves on a string	6
1.1 The wave equation in 1 dimension	7
1.1.1 Tension forces	8
1.1.2 Applying Newton's second law	10
1.1.3 A tiny bit of dimensional analysis	10
1.1.4 Additional content: longitudinal waves	11
1.2 The solution of d'Alembert	12
1.2.1 Characteristic coordinates	13
1.2.2 General solution of the wave equation	14
1.2.3 Travelling waves	15
1.2.4 Initial value problems on the real line	16
1.3 Boundaries and interfaces	18
1.3.1 Reflection at a fixed end	18
1.3.2 Reflection at a free end	19
1.3.3 Reflection and transmission at an interface	20
1.4 The solution of Bernoulli	24
1.4.1 Separation of variables	24
1.4.2 Finite strings: standing waves and superpositions	25
1.4.3 The initial value problem	29
1.5 Additional content: damping	29
Outro: takeaway points	33
2 Energy and harmonic waves	35
2.1 Harmonic waves	35
2.1.1 More on harmonic wave properties	36
2.1.2 Complex harmonic waves	38
2.1.3 Solving PDEs with harmonic waves	39
2.2 Energy	43
2.2.1 Energy density	43
2.2.2 Energy density of example waves	44
2.2.3 Conservation equation	47
Outro: takeaway points	48

3 Out of Flatland: bodies vibrating in 3-dimensions	51
3.1 Waves on a plane	51
3.1.1 The 2-dimensional wave equation	52
3.1.2 Energy of a membrane	53
3.1.3 Plane waves	54
3.1.4 Higher dimensions	55
3.2 Rectangular membranes	56
3.3 Circular membranes	61
Outro: takeaway points	68
II Quantum Dynamics	70
4 Introduction to quantum dynamics	72
4.1 The wavefunction	75
4.1.1 The probability interpretation	76
4.1.2 Normalisation	78
4.1.3 Superposition	79
4.2 The Schrödinger equation	80
4.2.1 Conservation of the probability	82
4.2.2 Observables and expectation values	83
4.2.3 The Heisenberg uncertainty principle	85
4.2.4 The collapse of the wavefunction	87
4.3 Back to one dimension	88
4.3.1 The time-independent Schrödinger equation	88
4.3.2 A particle on a circle	90
4.3.3 A particle in a box	91
4.4 Bounded and bounded particles	93
4.4.1 Well, well, well	95
4.4.2 Throwing particles at walls	100
Outro: takeaway points	106
5 Eigenvalue problems of Sturm-Liouville type	113
5.1 Spectral problems	113
5.1.1 Sturm-Liouville problems	114
5.2 Properties of Sturm-Liouville problems	116
5.2.1 Reduction of spectral problems to Sturm-Liouville form	118
5.2.2 Quantum mechanical interpretation of Sturm-Liouville boundary conditions	120
5.2.3 Regular Sturm-Liouville problems	121
Outro: takeaway points	122
6 Power series method	125
6.1 The harmonic oscillator	125
6.1.1 Taylor series and analytic functions	126
6.1.2 A reduction of the harmonic oscillator equation	127
6.1.3 Power series solution and the energy spectrum	129
6.1.4 The wavefunctions	131
6.2 The method of Frobenius	134

Outro: takeaway points	139
7 A quantum particle in 3 dimensions	141
7.1 Separation of variables, but quantum	141
7.1.1 Commutation relations	143
7.1.2 The eigenfunctions	146
7.1.3 The Legendre equation	148
7.1.4 Spherical harmonics	150
7.2 The Hydrogen atom	153
7.2.1 Solving the radial equation	155
7.2.2 The wavefunction is finite at the origin	156
7.2.3 The energy spectrum	157
7.2.4 The wavefunction	160
Outro: takeaway points	160

Foreword

Welcome! These are the lecture notes for **Quantum & Continuum Dynamics**, a 2nd year module at the University of York in Spring 2024. The content is subdivided into chapters roughly corresponding to a week (4 hours) worth of lectures.

These notes are far from being original. I inherited this course from professors Gustav Delius and Kasia Rejzner, whose notes constitute the blueprint on which these pages were written. I have also shamelessly drew upon the excellent lecture notes of prof. David Tong¹ for Part II.

This document will be periodically revised and improved. If you see something, say something. Either via e-mail stefano.negro@york.ac.uk or through the correction form <https://forms.gle/WjFFY6gmyVEpYAc08>.

During these lectures, we are going to take what you learned in the Classical Dynamics module and generalise it in two important ways.

1. Firstly, we will move from the analysis of discrete systems – i.e. systems possessing a discrete number of *degrees of freedom*, equivalently of *generalised coordinates* – to the modelling of the behaviour of materials that can be treated as continua. We will see that the interaction of a large (infinite actually!) number of particles generates an emergent phenomenon: waves. We will derive a model of this phenomenon employing such mathematical tools as *partial differential equations* and use it to extract predictions on the behaviour of simple continuous systems.
2. We will then start a journey into the baffling behaviours that Nature exhibits at small scales. To account for the seemingly contradictory fact that light (and all matter actually) is *both* a wave *and* made of particles, we will have to ditch our classical, comforting view of the world and employ a new, slightly unsettling, quantum perspective. We will see that waves play a fundamental role in the quantum world. However, they are different than the ones we dealt with before. They are not tangible and concrete, like the vibrations of a string. They are *waves of probability* that we can use to predict the chance that a particle has, at any given time, to, say, be in some position or have a certain momentum. If this makes little sense to you, do not worry: you are in good company². These lectures will introduce you to the outlandish and wondrous quantum world and the mathematical tools we employ to describe it.

¹You can find them here: <https://www.damtp.cam.ac.uk/user/tong/qm/qm.pdf>.

²Richard Feynman famously said “*I think I can safely say that nobody really understands quantum mechanics*” (pag. 129 of [**feynman2017character**]). With this he meant that, while physicists know perfectly well how to use it as a tool, there is still little consensus on what this theory is actually saying about reality [**schlosshauer2013snapshot**].

3. After the shock of your first quantum bath, we will return to the safe embrace of classical physics by introducing the dynamics of fluids. This is the study of the motion of gases and liquids from a macroscopic perspective, from which they can be treated as continua. Fluid dynamics is a huge subject. In this introduction, we will concentrate on the founding ideas and tools that underlie the mathematical modelling fluids.

The concepts and methods presented in this module form the bedrock of a large part of Applied Mathematics. You will meet them time and again in future Applied Mathematics classes, as they constitute core modelling tools in diverse fields such as biology, ecology, physics, medicine, sociology, and economics. The examples presented in this module are less important than the principles and way of thinking they introduce.

The goal of this module is to equip you with the skills and knowledge to apply mathematics to a wide array of phenomena in the real world. In your third and fourth years, you will have the opportunity to deepen this ability through various modules.

Reading material

These notes are self-standing, in the sense that the lectures will follow them closely. The exercises and the exam will focus on their content only.

Still, it is good practice to draw information from a multitude of distinct sources, a fact that is true of news and of learning in general. Physics and mathematics are no exceptions. Reading different explanations of the same concept helps in making it “three-dimensional”, more approachable and more intuitive. This becomes particularly important when the concept is new, unfamiliar or complicated. Part II of this module most probably qualifies for one or more of these descriptors.

So, by all means, read away!

Part I: Waves

- Coulson, C. A. and Jeffrey, A. *Waves: A mathematical approach to the common types of wave motion*. Longman Group Limited (1977).

A classical text on the mathematics of wave phenomena. You will want to read through chapters 1 to 3, as they deal with most of the topics of part I of this module.

- Knobel, Roger. *An introduction to the mathematical theory of waves*. Vol. 3. American Mathematical Soc., 2000. DOI: [10.1090/stml/003](https://doi.org/10.1090/stml/003).

A book on the concept of waves, written from a mathematical angle. In chapters 1 and 2 it deals with all the topics we will cover in the first two weeks of the module. You are encouraged to read through this material and to perform the exercises it presents.

- Bettini, Alessandro. *A Course in Classical Physics 4 – Waves and Light*. Springer International Publishing, 2017. DOI: [10.1007/978-3-319-48329-0](https://doi.org/10.1007/978-3-319-48329-0).

A nice book on the physics of waves with a focus on light. It starts with a review of oscillatory phenomena in systems with finite number of degrees of freedom and

then moves to a description of waves in continuous media.

For the purpose of this module, you want to look at sections 2.4 to 2.7 (dealing with harmonic – i.e. Fourier – analysis) and Chapter 3 up to 3.5 included (dealing with waves) and excluding 3.4.

The rest of the book deals mainly with electromagnetic waves (light) and various phenomena, such as dispersion, diffraction and interference. While these topics are not contained in the syllabus of the module, you are encouraged to have a read as an in-depth study.

Part II: Quantum Dynamics

- Griffiths, David J. and Schroeter, Darrell F. *Introduction to quantum mechanics* (3rd ed.). Cambridge University Press (2018).

An excellent introductory book, it will guide you through an introduction to quantum mechanics. For this module, you really only need Chapters 1, 2 and parts of 4. However, you are encouraged to read through all of part I for in-depth study.

- Rae, Alastair I. M., Napolitano, Jim. *Quantum Mechanics, 6th edition*. CRC Press, Taylor & Francis Group (2016). DOI: [10.1201/b19619](https://doi.org/10.1201/b19619).

One of the most well-regarded introductory books on the subject. Part I contains a review of waves, electromagnetism, of the limits of classical physics and the birth of quantum mechanics. Part II deals with elementary wave mechanics and Part III with the foundations – i.e. the principles at the base of quantum mechanics and their interpretations. The remaining parts treat applications and more advanced topics.

You are encouraged to read through chapters 1, 4, 5, 6 and part of 7. The rest of the book is for you to peruse, should you be interested in the topic.

There are very many books on quantum mechanics. Some of them are more advanced than others, some more illuminating. If you wish to get more in-depth understanding of the quantum world, I suggest you refer to the following textbooks. You will find that they will accompany you a long way in your journey.

- Shankar, R. *Principles of Quantum Mechanics*. Springer New York, NY (1994). DOI: [10.1007/978-1-4757-0576-8](https://doi.org/10.1007/978-1-4757-0576-8).

A long book that takes your hand and guides you through all the basic topics of quantum mechanics, leading you gently through challenging computations.

- Weinberg, Steven. *Lectures on Quantum Mechanics*. Cambridge University Press (2015). DOI: [10.1017/CBO9781316276105](https://doi.org/10.1017/CBO9781316276105).

A book written by a giant of 20th Century theoretical physics. He is always worth listening to.

Part I

Waves (Continuum Dynamics)

Waves play a fundamental role in our universe, as they are the only means by which information can propagate through space and time. Some examples of waves that carry information are well-known, such as sound waves, light waves, radio waves, and electrical waves that travel along our neurons. Other examples are less obvious, such as the waves that describe the motion of particles in quantum mechanics and the gravitational waves that convey the effects of gravity.

Because waves travel at a finite speed, information can only propagate at the speed of the wave, a fact that has profound implications, as we see in the theory of special relativity. For instance, we can observe distant events in the past because some of the light waves that were emitted soon after the Big Bang, roughly 16 billion years ago, are only now reaching us. These waves are strongly red-shifted and detectable as microwaves.

The study of waves is not only important from a theoretical perspective but also from a practical standpoint. Waves play a vital role in our modern technological world. Advances in our understanding of how to generate and control electromagnetic waves have led to the development of essential technologies such as radio, radar, and mobile phones.

Furthermore, understanding how waves propagate can also help us prepare for and prevent potential threats. For instance, studying how pests spread in the form of invasion waves into previously uninfected areas can help us develop effective interventions to prevent their spread. Similarly, understanding how density waves form in traffic flow and lead to traffic jams can help us design interventions to improve traffic flow and reduce congestion.

By studying waves, we can gain insights into the underlying mechanisms that govern a wide range of phenomena, from the behaviour of subatomic particles to the dynamics of traffic flow. These insights find applications in various fields, including engineering, medicine, and environmental science. By deepening our understanding of waves, we can continue to make advancements in technology and address real-world challenges more effectively.

Chapter 1

Waves on a string

Given the wild diversity of wave-like phenomena of our world, we need to focus on a simple example from which to start our investigation into the nature of wave dynamics. Hence, we will start from the simplest setting possible: a vibrating string. Effectively, we are going to build and analyse a mathematical model that describes what happens when you pluck a guitar string¹.

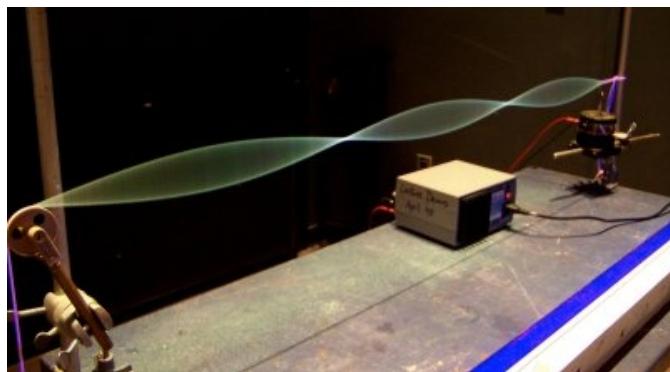


Figure 1.1: 3rd harmonic of a 1.5m length of string driven at one end and fixed at the other. See <https://sciedem.demonstrations.fas.harvard.edu/presentations/vibrating-string>.
©The President and Fellows of Harvard College

While the vibrating string is a simple model – at least compared to other real-world wave phenomena – by studying it thoroughly you will develop the intuition and the skills needed to tackle more complicated setups. Later in the course we will apply what we have learned to vibrating membranes (chapter 3) and waves on the surface of a fluid.

An additional piece of motivation – at least to those amongst you that would like to delve into applications of mathematics to theoretical physics – comes from high-energy physics. Simple as it is, the model of a vibrating string is the foundation on which physicists built (super-)string theory, a “theory of everything” positing that Nature is made of tiny strings, with different vibrational states corresponding to different elementary particles. The mathematics underlying string theory is certainly expansive – touching topics that go from differential geometry to algebraic topology – but, at its core, it is no more complicated than what we will discuss in this module: vibrating strings and quantum mechanics.

¹For a beautiful visual inspiration see the video <https://www.youtube.com/watch?v=8YGQmV3NxMI>.

1.1 The wave equation in 1 dimension

Let us consider our investigation subject: a flexible, elastic string of *length* L , *uniform linear density* ρ (mass per unit length), stretched at its ends with a *tension* T . As it is often the case, in order to isolate and describe quantitatively the specific phenomenon we wish to study – i.e. waves propagating along the string – we will make several simplifications and assumptions:

1. We suppose that the only force acting on the string is the tension T . Stated differently, we will ignore the contribution of external forces, such as gravity and drag due to air friction. As a consequence, the equilibrium configuration of the string is a **straight segment**.
2. We neglect the fact that the string is actually made up of lots of individual atoms and instead pretend that the mass is spread out continuously along the string. This is known as the *continuum approximation* and we will meet this again in the fluid dynamics part of this module.
3. We suppose that each element of the string can only move in the direction **perpendicular to the equilibrium line**. In other words, we focus on *transverse vibrations*. Additionally, we assume that the string vibrates on a plane, disregarding the existence of a third dimension.
4. We assume that the amplitude of the oscillations remains small enough, so that the magnitude of the string tension T can be taken as constant throughout the string and the angle that the string makes with its equilibrium line at each point is small.

With these assumptions in place, we can start our derivation of an equation describing the time evolution of the string displacement with respect to the equilibrium line.

We choose our coordinate system so that the string in its equilibrium configuration lies on the interval $(0, L)$ on the x -axis. The displacement from it is then described by a function $y(x, t)$ of the coordinate x and of the time t . See Fig. 1.2. At any fixed t_0 , the

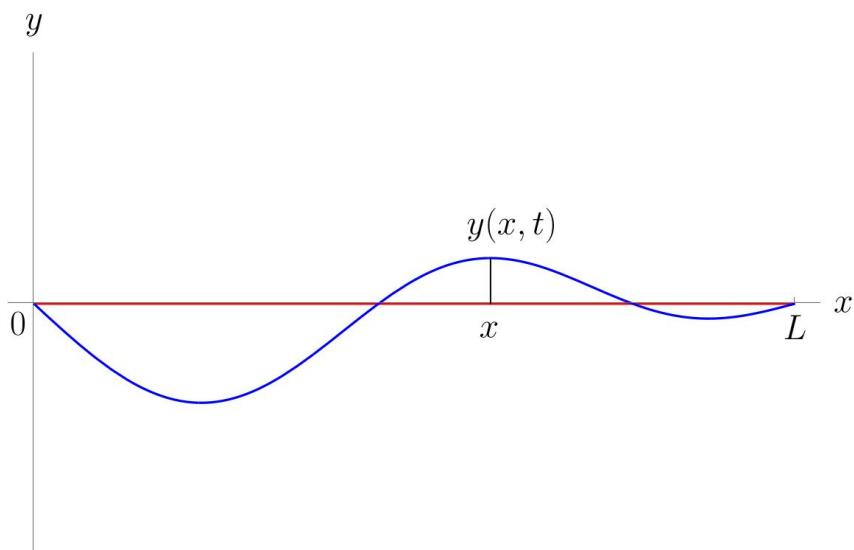


Figure 1.2: Co-ordinatisation of a string vibrating transversely. In red is the equilibrium configuration.

function $y(x, t_0)$ describes the shape of the string at the instant $t = t_0$. Let us focus on a specific moment in time. In the next few paragraphs, while studying the forces acting on the string, we will drop the explicit dependence on time, agreeing that all the quantities are taken at the same instant. We will return to the time dependence later on.

1.1.1 Tension forces

Let us isolate an infinitesimal portion of the string, between x and $x + \delta x$, where we agree that $0 < x < L$ and $\delta x \ll 1$ is an infinitesimal quantity that we will eventually send to 0. Since we assumed that the only force at play in our setting is the tension of the string, the total force acting on the portion we isolated is

$$\vec{F} = \vec{T}(x + \delta x) + \vec{T}(x). \quad (1.1)$$

We assumed the tension $\vec{T}(x)$ to have constant magnitude $|\vec{T}(x)| = T$ along the string, however, its direction varies since it always acts tangentially. Figure 1.3 depicts a schematic representation of these forces. Decomposing the total force into its horizontal and vertical

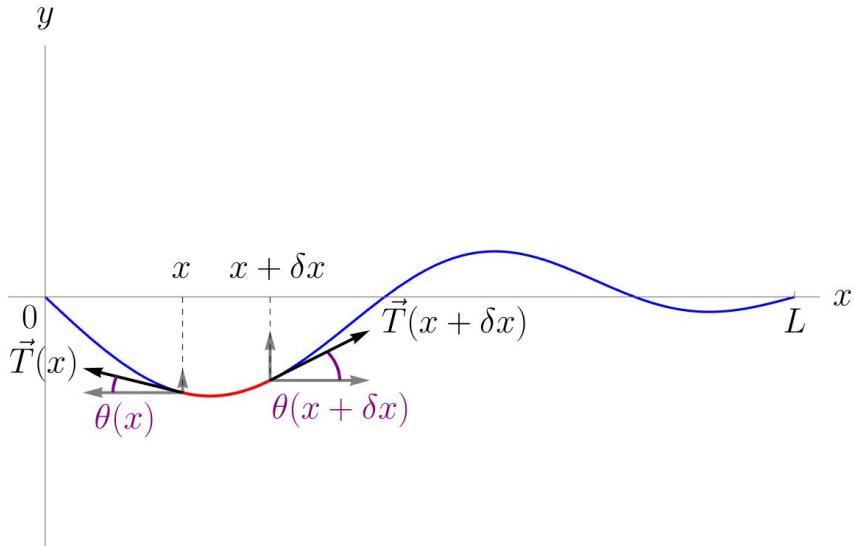


Figure 1.3: Tension forces acting on an infinitesimal portion of the string.

components, we find

$$\vec{F} = \begin{pmatrix} F_x \\ F_y \end{pmatrix} = T \begin{pmatrix} \cos \theta(x + \delta x) - \cos \theta(x) \\ \sin \theta(x + \delta x) - \sin \theta(x) \end{pmatrix}. \quad (1.2)$$

Now we are going to use the fact that $\delta x \ll 1$, so, in the hypothesis that the function $\theta(x)$ is smooth – which we implicitly assumed by asking for the oscillation amplitudes to be small – we can perform a Taylor expansion

$$\theta(x + \delta x) = \theta(x) + \delta x \frac{\partial}{\partial x} \theta(x) + \mathcal{O}(\delta x^2). \quad (1.3)$$

By further expanding the sine and cosine functions, we arrive at

$$\vec{F} = T \delta x \begin{pmatrix} -\sin \theta(x) \\ \cos \theta(x) \end{pmatrix} \frac{\partial}{\partial x} \theta(x) + \mathcal{O}(\delta x^2). \quad (1.4)$$

Finally, let us remember that we supposed the oscillations to be small. Stated differently, the angle function $\theta(x)$ is small everywhere along the string and we can approximate the trigonometric functions by their zero-order Taylor expansions

$$\sin \theta(x) = \mathcal{O}(\theta(x)) , \quad \cos \theta(x) = 1 + \mathcal{O}(\theta(x)^2) . \quad (1.5)$$

We will also need to relate $\partial\theta(x)/\partial x$ to the shape function $y(x)$. To this end, we can use the identity

$$\frac{\partial}{\partial x} y(x) = \tan \theta(x) \implies \frac{\partial^2}{\partial x^2} y(x) = \frac{\frac{\partial}{\partial x} \theta(x)}{\cos^2 \theta(x)} = \frac{\partial}{\partial x} \theta(x) + \mathcal{O}(\theta(x)^2) , \quad (1.6)$$

using which, we arrive at the following expression

$$\vec{F} = \left[T \frac{\partial^2}{\partial x^2} y(x) \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \mathcal{O}(\theta(x)) \right] \delta x + \mathcal{O}(\delta x^2) . \quad (1.7)$$

So, wrapping up this subsection, we have seen that, under the assumptions listed at the beginning of §1.1, the forces acting on an infinitesimally small portion of a string approximately sum up to zero along the horizontal direction and to a generally non-vanishing quantity proportional to the *curvature* $y''(x)$ of the shape function

$$F_x = 0 + \mathcal{O}(\theta(x)) + \mathcal{O}(\delta x^2) , \quad F_y = \left[T \frac{\partial^2}{\partial x^2} y(x) + \mathcal{O}(\theta(x)) \right] \delta x + \mathcal{O}(\delta x^2) , \quad (1.8)$$

in accordance with our hypothesis that the string only moves perpendicularly to the equilibrium line.

We arrived at this last expression by assuming two distinct quantities to be small: the angle $\theta(x)$ (equivalently, the slope $y'(x)$ of the shape function) and the length δx of the string portion. It is worth remarking that the nature of these two quantities is quite different. The length δx is infinitesimal **by construction**. Ultimately, we will be taking the limit $\delta x \rightarrow 0$, so we expect the corrections $\mathcal{O}(\delta x^2)$ to not be physically relevant, as they will vanish in the limit². On the other hand, the angle $\theta(x)$ is small **by assumption** and we will not send it to zero in our formulae. Consequently, though we expect them to be small, the corrections $\mathcal{O}(\theta(x))$ are physically relevant when comparing our theoretical model to the real world.

What about the endpoints? Another, not explicitly stated assumption we have made is that the portion of the string under scrutiny does not end on one of the string's endpoints $x = 0$ or $x = L$. In fact the dynamics at the endpoints requires its specific analysis and we will deal with this later on. For the moment we will content ourselves with deriving the dynamics of the string "far away" from its endpoints.

²This is actually not exactly true, since the string is really made of individual atoms and is not a true continuum. Thus δx can only be as small as the mean separation between atoms. Additionally, when we get to typical inter-atomic distances we will need to account for quantum effects in our description, which brings about a whole slew of problems that we really don't want to get into. Ultimately, for any conceivable macroscopic phenomenon, these corrections are so tiny to be in all respects undetectable.

1.1.2 Applying Newton's second law

Now that we have derived the force acting on the portion of our string, we can derive the equation of motion by applying Newton's second law

$$m\vec{a}(x, t) = \vec{F}(x, t), \quad (1.9)$$

where m is the mass of the string portion and $\vec{a}(x, t)$ is its acceleration. Note that we reintroduced the explicit dependence on time. The force along the x axis is 0 at our level of approximation, so we can disregard the motion along the horizontal direction. The mass of the string portion is obtained by multiplying the linear density ρ by the length δx . Thus we arrive at

$$ma_y(x, t) \equiv \rho\delta x \frac{\partial^2}{\partial t^2}y(x, t) = \left[T \frac{\partial^2}{\partial x^2}y(x) + \mathcal{O}(\theta(x)) \right] \delta x + \mathcal{O}(\delta x^2). \quad (1.10)$$

Dropping the $\mathcal{O}(\theta(x))$ terms, dividing by δx and then taking the limit $\delta x \rightarrow 0$, we finally arrive at the *1-dimensional wave equation*

$$\boxed{\frac{\partial^2}{\partial t^2}y(x, t) = c^2 \frac{\partial^2}{\partial x^2}y(x, t)}, \quad (1.11)$$

where we combined the tension T and the density ρ in a single constant, the *wave speed*

$$c = \sqrt{\frac{T}{\rho}}. \quad (1.12)$$

The reason for this nomenclature will become clear momentarily.

1.1.3 A tiny bit of dimensional analysis

It is always good practice when deriving equations to check the agreement of dimensions amongst its various terms. This type of scrutiny is known as *dimensional analysis* and – despite its simplicity – is a surprisingly powerful method, with applications going far beyond simple consistency checks³.

We will use square brackets to denote the dimension of a quantity. Thus, by writing $[y] = L$ we are saying that y has the dimensions of a length. Similarly $[m] = M$ says that m has dimensions of a mass and $[t] = T$ that t is measured in dimensions of time. So the dimensional analysis of our wave equation tells us that⁴

$$\left[\frac{\partial^2 y(x, t)}{\partial t^2} \right] = [c]^2 \left[\frac{\partial^2 y(x, t)}{\partial x^2} \right] \implies \frac{L}{T^2} = [c]^2 \frac{1}{L} \implies [c] = \frac{L}{T}. \quad (1.13)$$

Let us check that these dimensions agree with the definition (1.12). Since T is a force, we have $[T] = ML/T^2$, while ρ is a linear density, hence $[\rho] = M/L$. From (1.12), we then find

$$[c] = \sqrt{\frac{[T]}{[\rho]}} = \sqrt{\frac{ML/T^2}{M/L}} = \sqrt{\frac{L^2}{T^2}} = \frac{L}{T}. \quad (1.14)$$

This confirms the dimensional consistency of the wave equation (1.11). Additionally, we see that c has the dimensions of a velocity, which agrees with our choice of nomenclature. As for what this is the velocity of, we will get to that in §1.2.

³For a nice discussion on some applications, see the excellent notes by prof. David Tong <https://www.damtp.cam.ac.uk/user/tong/relativity/three.pdf>.

⁴It is quite evident that the dimension of a product is the product of the dimensions: $[ab] = [a][b]$.

1.1.4 Additional content: longitudinal waves

In this subsection we are going to model the behaviour of *longitudinal waves*, that is waves propagating parallel to the direction a body extends in. Think of a narrow metal rod that is being struck on one of its ends by a hammer. While the rod is a rigid body, the hit will ever so slightly compress it at its end. This compressed bit will in turn “push” on the nearby material, slightly displacing it and this phenomenon will intuitively cascade throughout the length of the rod. A similar phenomenon happens to the column of air inside of a wind instrument, say a trumpet, when the player blows in the mouthpiece⁵. We wish to model these type of phenomena.

As before, we are going to greatly simplify our setting and model the metal rod (or column of air) as an array of weights of equal mass m connected to each neighbour by a massless spring of rest length h and spring constant k . See Figure 1.4 for a schematic depiction along with a reference coordinate system that we assign to the rest configuration. We suppose that the masses

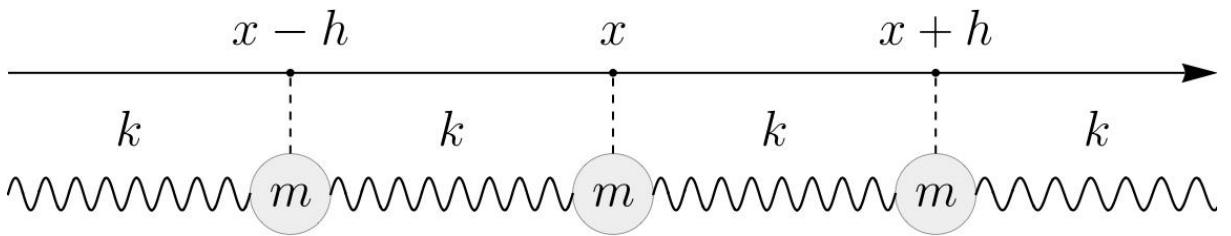


Figure 1.4: An array of masses modelling a 1-dimensional material supporting the propagation of longitudinal waves. Here is the rest configuration, along with its reference coordinatisation.

can only move in the horizontal direction. A convenient measure of the position of each mass is in terms of its displacement $u(x)$ from the rest position. See Figure 1.5.

Let us consider the forces acting on one of these little weights, say the one at rest position x . Hooke’s law tells us that each spring exerts a force equal to the product of the spring constant k and the length Δ by which the spring is deformed, compared to its rest length h . This latter quantity is easily computed in terms of the displacements

$$\Delta(x) = [x + h + u(x + h)] - [x + u(x)] - h = u(x + h) - u(x) . \quad (1.15)$$

Now, on the mass with rest position x there are two springs, acting in opposite directions. Thus the total force is

$$F(x) = k\Delta(x) - k\Delta(x - h) = k[u(x + h) - 2u(x) + u(x - h)] . \quad (1.16)$$

Now, as before, we apply Newton’s 2nd law, obtaining (we reinstate the time dependence of the displacements that we conveniently omitted above)

$$\frac{\partial^2}{\partial t^2}u(x, t) = \frac{k}{m}[u(x + h, t) - 2u(x, t) + u(x - h, t)] . \quad (1.17)$$

The last step is to perform a *continuum limit*: we want the distance between the masses h to go to zero. At the same time, however, we also want the length of the array to remain constant! For finite h , the length of the array is

$$L = Nh , \quad (1.18)$$

⁵Actually, as it is often the case, the real life physics is much more complicated than what we are going to explore here. The physical study of instruments and, more generally, of acoustics is a deeply fascinating topic. If you are interested I suggest you check the excellent books [fletcher2012physics, benade1990].

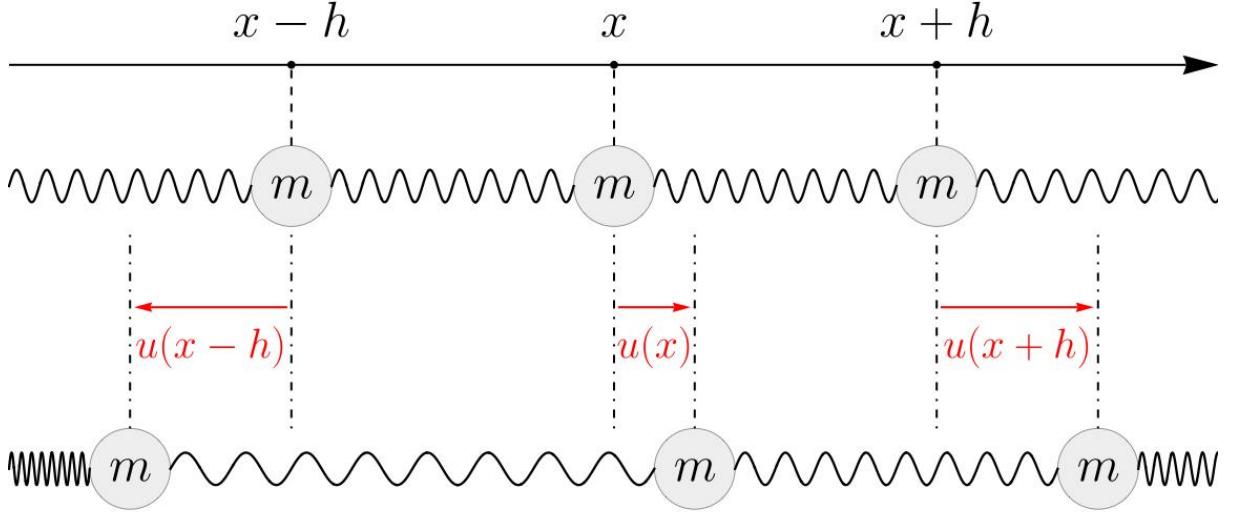


Figure 1.5: The array of masses in its rest position (upper) compared to an excited configuration (lower). We assign a function $u(x)$ to each of the masses that measures their displacement from the rest position.

where N is the total number of weights in the array. To keep L constant as $h \rightarrow 0$, we need to send $N \rightarrow \infty$. In theoretical physics, this type of procedure in which we take two limits simultaneously in order to keep certain quantities constant is commonplace and takes the name of *double scaling limit*. To properly take this limit, we must also express the “local” quantities m and k in terms of the corresponding “global” ones M and K , that is the total mass and effective spring constant of the array. These are simply⁶

$$M = Nm, \quad K = k/N. \quad (1.19)$$

Putting everything together and expressing N in terms of h and L , we arrive at the equation

$$\frac{\partial^2}{\partial t^2}u(x,t) = \frac{KL^2}{M} \frac{u(x+h,t) - 2u(x,t) + u(x-h,t)}{h^2}, \quad (1.20)$$

where we recognise the discrete version of a second derivative in the right-hand side. We can now safely take the limit $h \rightarrow 0$, obtaining

$$\frac{\partial^2}{\partial t^2}u(x,t) = c^2 \frac{\partial^2}{\partial x^2}u(x,t), \quad (1.21)$$

with $c = L\sqrt{K/M}$ being the wave speed⁷.

1.2 The solution of d'Alembert

Having obtained an equation that describes the dynamics of transverse waves on a 1-dimensional string, we will now commence an analysis of its solutions. First, we are going to study the dynamics *in the bulk*, meaning we are going to consider a very long string

⁶Remember that, for springs in series the effective spring constant is computed as the inverse of the sum of inverses: $1/K = \sum_{j=1}^N 1/k_j$.

⁷It is worth to confirm that c truly is a velocity by performing a simple dimensional analysis, remembering that K is a force per unit of length.

and focus on what happens far from its endpoints. Effectively this means we can forget about boundary effects and search for solutions to the wave equation

$$\partial_t^2 y(x, t) - c^2 \partial_x^2 y(x, t) = 0 , \quad (1.22)$$

on the real axis $x \in (-\infty, \infty)$. We will improve our model by including the effects of boundaries in §1.3. Note that in (1.22) we introduced the convenient short-hand notation $\partial_t^m = \partial^m / \partial t^m$ and $\partial_x^m = \partial^m / \partial x^m$ to lighten the expressions.

The strategy to find solutions to (1.22) is to first find the most general solution – we will employ the *method of characteristics* to this end – and subsequently select a particular solution by imposing initial conditions. So, without further ado, let us begin.

1.2.1 Characteristic coordinates

We are going to solve (1.22) with the *method of characteristics*. The key idea behind this approach is to simplify the expression of a Partial Differential Equation (PDE) by finding an appropriate coordinate transformation

$$(x, t) \longrightarrow (\xi(x, t), \eta(x, t)) . \quad (1.23)$$

We then define a new function \tilde{y} of the coordinates ξ and η , such that it has the same values as y

$$y(x, t) \equiv \tilde{y}(\xi(x, t), \eta(x, t)) , \quad \forall x, t , \quad (1.24)$$

and fix the function $\xi(x, t)$ and $\eta(x, t)$ in such a way that our PDE takes the simple *canonical* form

$$\partial_\xi \partial_\eta \tilde{y}(\xi, \eta) = \ell(\tilde{y}) , \quad (1.25)$$

with $\ell(\tilde{y})$ being a, possibly vanishing, first-order linear operator acting on \tilde{y} . A discussion of this method for a general 2-variables PDE is presented in Appendix ???. Here we jump over some derivations⁸ and present directly the *characteristic coordinates* for (1.22):

$$\xi(x, t) = x + ct , \quad \eta(x, t) = x - ct . \quad (1.26)$$

From here onwards we will make an abuse of notation and drop the tilde above the function in the transformed coordinates. This is a conventional abuse, aimed at avoiding excessive clutter in the expressions.

Let us check that the wave equation truly takes the canonical form in the characteristic coordinates. We need to express the derivatives ∂_t and ∂_x in terms of ∂_ξ and ∂_η . This is easily done using the chain rule

$$\begin{aligned} \partial_t y(\xi(x, t), \eta(x, t)) &= \partial_t \xi(x, t) \partial_\xi y(\xi, \eta) + \partial_t \eta(x, t) \partial_\eta y(\xi, \eta) = \\ &= c(\partial_\xi - \partial_\eta)y(\xi, \eta) , \end{aligned} \quad (1.27a)$$

⁸It is instructive for you to derive these coordinates by yourselves, before looking at the general method in Appendix ???. Simply substitute (1.24) into (1.22) and try to cancel all the terms but $\partial_\xi \partial_\eta \tilde{y}$. Note that you will have some freedom in the definition of ξ and η in the form of undetermined constants and a possible swapping $\xi \leftrightarrow \eta$. The choice (1.26) is the simplest and is commonplace in the literature.

$$\begin{aligned}\partial_x y(\xi(x, t), \eta(x, t)) &= \partial_x \xi(x, t) \partial_\xi y(\xi, \eta) + \partial_x \eta(x, t) \partial_\eta y(\xi, \eta) = \\ &= (\partial_\xi + \partial_\eta) y(\xi, \eta).\end{aligned}\tag{1.27b}$$

Applying ∂_t and ∂_x again to the above expressions, we find

$$\partial_t y(\xi(x, t), \eta(x, t)) = c^2 [\partial_\xi^2 y(\xi, \eta) - 2\partial_\xi \partial_\eta y(\xi, \eta) + \partial_\eta^2 y(\xi, \eta)], \tag{1.28a}$$

$$\partial_x y(\xi(x, t), \eta(x, t)) = \partial_\xi^2 y(\xi, \eta) + 2\partial_\xi \partial_\eta y(\xi, \eta) + \partial_\eta^2 y(\xi, \eta), \tag{1.28b}$$

and it is a matter of simple substitution into (1.22) to arrive at

$$-4c^2 \partial_\xi \partial_\eta y(\xi, \eta) = 0. \tag{1.29}$$

Finally, we divide by the non-zero constant $-4c^2$ and write the wave equation in its canonical form

$$\boxed{\partial_\xi \partial_\eta y(\xi, \eta) = 0, \quad \begin{cases} \xi = x + ct \\ \eta = x - ct \end{cases}}. \tag{1.30}$$

1.2.2 General solution of the wave equation

The wave equation in its canonical form is extremely simple to solve. In fact, let us interpret what it is telling us. If we denote $\partial_\eta y(\xi, \eta) = y_\eta(\xi, \eta)$, equation (1.30) implies that y_η is independent of ξ

$$y_\eta(\xi, \eta) \equiv \partial_\eta y(\xi, \eta) \equiv f(\eta). \tag{1.31}$$

The function $f(\eta)$ is arbitrary, so we can express it in terms of its primitive $F(\eta)$ as

$$f(\eta) = \partial_\eta F(\eta). \tag{1.32}$$

Thus we find the equation

$$\partial_\eta [y(\xi, \eta) - F(\eta)] = 0, \tag{1.33}$$

and it appears we can conclude that $y(\xi, \eta) = F(\eta)$. This however is not quite right. In fact, we can play the same game swapping the roles of ξ and η , concluding that

$$\partial_\xi [y(\xi, \eta) - G(\xi)] = 0, \tag{1.34}$$

with $G(\xi)$ being another arbitrary function. The way out this apparent mismatch is that we can add any function of ξ inside the square brackets in (1.33): the action of the derivative ∂_η is identically zero. Thus, we conclude – correctly this time – that the general solution of (1.30) is⁹

$$y(\xi, \eta) = F(\eta) + G(\xi), \tag{1.35}$$

with $F(\eta)$ and $G(\xi)$ being arbitrary functions of one variable.

⁹In passing from (1.33) to the expression (1.35) we are performing an indefinite integration with respect to η . Remember that this operation is defined up to an additive constant C . When integrating functions of more than one variable, C needs only to be constant with respect to the variable being integrated: it can depend arbitrarily on all the other ones.

1.2.3 Travelling waves

Let us gain some understanding of the general solution (1.35). We will do so by isolating two special cases: $G = 0$ and $F = 0$.

When $G(\xi)$ vanish identically, the solution reads

$$y(x, t) = F(x - ct) . \quad (1.36)$$

We see that this solution evolves in time by **rigidly moving to the right** – i.e. in the direction of increasing values of x – and its overall shape remains unchanged. See Figure 1.6 for a graphical depiction. We call such a solution a *right-moving wave*.

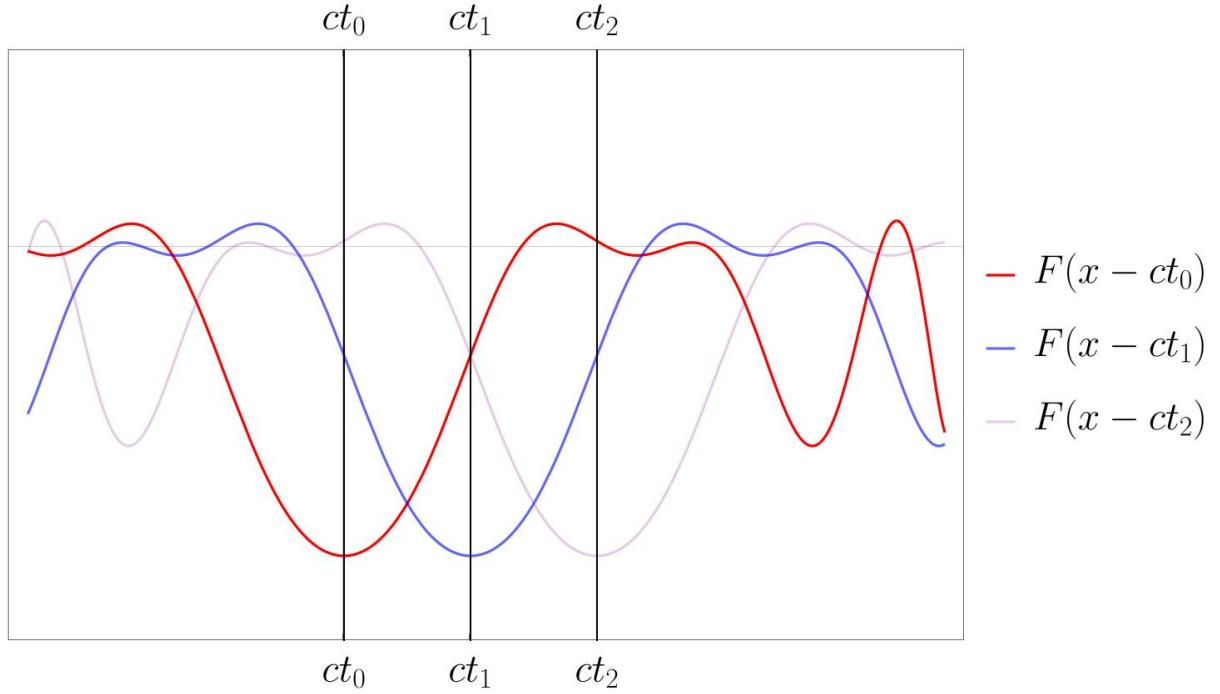


Figure 1.6: Graph of a function $F(x - ct)$ for different values of t : $t_0 = 0$ and $0 < t_1 < t_2$. As you can see, the function evolves in time by preserving its shape and rigidly moving to the right. This function is a right-moving wave.

Conversely, when $F(\eta)$ vanishes identically, the solution will take the form

$$y(x, t) = G(x + ct) , \quad (1.37)$$

and, similarly but opposed to the previous case, it will evolve in time by **moving rigidly to the left**, that is in the direction of decreasing values of x . We call this solutions *left-moving waves*.

Recapitulating, $F(x - ct)$ and $G(x + ct)$ describe waves that propagate, preserving their shape, to the right and to the left, respectively. The general solution (1.35) is the *superposition* – a fancy name for “sum” – of two such waves.

Remark The fact that the superposition of two solution is still a solution is an elementary consequence of the linearity of the wave equation. Yet, this simple fact has profound implications, as we will see later in this course.

1.2.4 Initial value problems on the real line

Now that we have the general solution to the wave equation in our pocket, we can study the *initial value problem*. The question we want to answer is: “how does an initial configuration of the string evolves in time?” Mathematically this question amounts to finding the particular solution to

$$\partial_t^2 y(x, t) - c^2 \partial_x^2 y(x, t) = 0 , \quad (1.38)$$

for $-\infty < x < \infty$ and $0 \leq t < \infty$, subject to the initial conditions

$$y(x, 0) = y_0(x) , \quad \partial_t y(x, t) \Big|_{t=0} = v_0(x) . \quad (1.39)$$

Here $y_0(x)$ and $v_0(x)$ are given functions of x , the input of the problem we wish to solve. They represent, respectively, the initial shape of the string and the initial velocity of each of its points. Since the equation (1.38) is of second order in t , these conditions are sufficient to uniquely fix a particular solution.

To solve the initial problem, we substitute the solution (1.35) into the initial conditions

$$F(x) + G(x) = y_0(x) , \quad F'(x) - G'(x) = -\frac{1}{c}v_0(x) . \quad (1.40)$$

Let us start by solving the second one. To save us from writing too many integral signs, we introduce the primitive

$$V(x) \equiv \int_0^x ds v_0(s) + C , \quad (1.41)$$

where C is an arbitrary constant. Then

$$F(x) = G(x) - \frac{V(x)}{c} . \quad (1.42)$$

Now we can work on the first equation that, together with (1.42), gives us

$$F(x) = \frac{1}{2}y_0(x) - \frac{1}{2c}V(x) , \quad (1.43a)$$

$$G(x) = \frac{1}{2}y_0(x) + \frac{1}{2c}V(x) . \quad (1.43b)$$

Finally we plug these solution into the general expression (1.35)

$$y(x, t) = \frac{y_0(x - ct) + y_0(x + ct)}{2} + \frac{V(x + ct) - V(x - ct)}{2c} . \quad (1.44)$$

We notice that the arbitrary constant cancels, as we predicted

$$V(x + ct) - V(x - ct) = \int_0^{x+ct} ds v_0(s) + C - \left[\int_0^{x-ct} ds v_0(s) + C \right] = \int_{x-ct}^{x+ct} ds v_0(s) , \quad (1.45)$$

so, in conclusion, we arrive at

$$y(x, t) = \frac{y_0(x - ct) + y_0(x + ct)}{2} + \frac{1}{2c} \int_{x-ct}^{x+ct} ds v_0(s) . \quad (1.46)$$

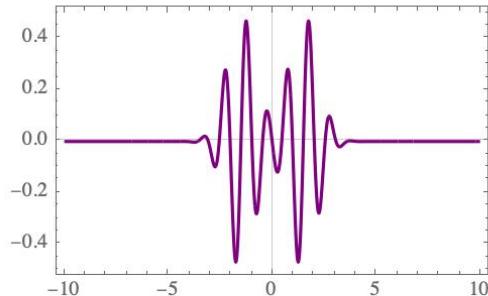
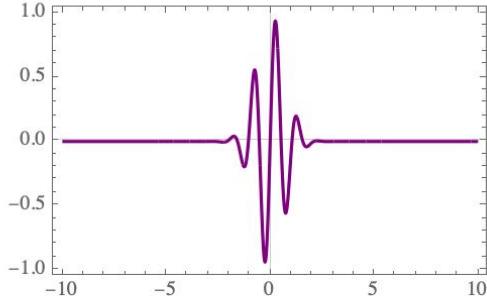
This expression is known as *d'Alembert formula*, after the french mathematician Jean-Baptiste le Rond d'Alembert, who derived it in 1747.

To give you a feel for the solution (1.46), in Figure 1.7 we plotted it for¹⁰

$$y_0(x) = e^{-x^2} \sin(6x) , \quad v_0(x) = 0 . \quad (1.47)$$

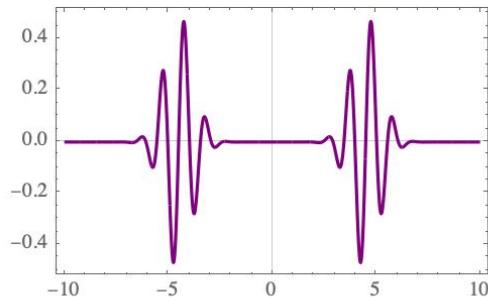
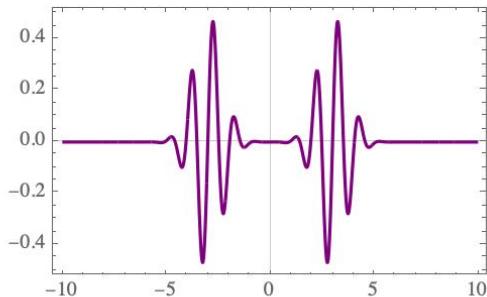
$t = 0$

$t = 1$



$t = 2$

$t = 3$



$t = 4$

$t = 5$

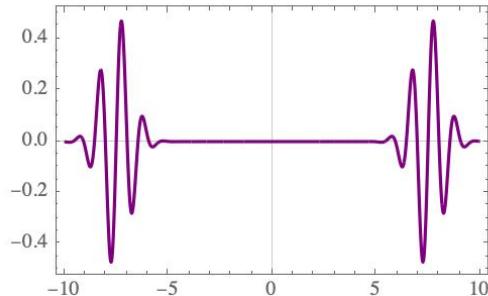
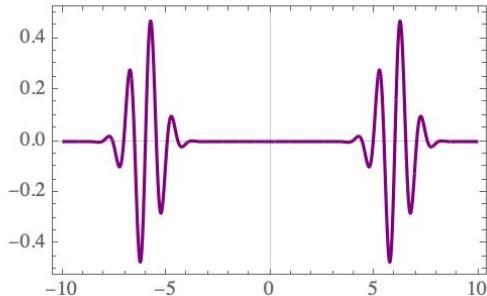


Figure 1.7: Time snapshots of the solution to the wave equation with initial conditions $y_0(x) = e^{-x^2} \sin(6x)$ and $v_0(x) = 0$. You can see an initial wave packet splitting in two copies that travel in opposite directions (notice the scale of the y axis in the various plots).

Remark In the derivation of (1.46), the only mathematical requirements we have to impose on the initial conditions is that the function $v_0(x)$ must be integrable on arbitrary subsets of the real line. Solutions of the initial value problem (1.38, 1.39) corresponding to piece-wise smooth (or even piece-wise continuous) initial functions $y_0(x)$ and $v_0(x)$ are usually referred to as *generalised solutions of the wave equation*. These are defined everywhere in the upper half of the (x, t) plane except for a finite number of lines where values of $y(x, t)$ and/or of its first derivatives are discontinuous.

¹⁰This type of functions are known as *wave packets*.

1.3 Boundaries and interfaces

Up until now we have considered our string to be effectively infinite. We saw that the wave equation predicts the existence of left- and right-moving waves that travel at constant speed – determined by the tension and the density of the string – preserving their overall shape. However, in the real world these waves are bound to meet the end of the string at some point. What will happen then? Even though we have not thoroughly discussed it – we will do so later in §2.2 – intuitively we know that waves carry some energy: after all the string moves with a certain velocity, thus it must possess a kinetic energy. In absence of dissipation, energy is conserved so the wave cannot simply disappear at the ends of the string. It must be somehow *reflected*. In this section we will work out the details of how a wave reflects at the boundaries of a string and what happens when a wave crosses an *interface*, that is a point where the physical properties of the propagation medium change abruptly.

1.3.1 Reflection at a fixed end

Let us start simply by looking at what happens when a wave reaches one endpoint. We choose a right-moving wave that arrives at the right end of the string. We are not interested in the other boundary that we suppose lie someplace far to the left. Mathematically, we are considering the wave equation on the half-line $-\infty < x < 0$. First we consider the case in which the right end, that we placed at $x = 0$, is held *fixed*, meaning

$$y(0, t) = 0, \quad \forall t \in \mathbb{R}. \quad (1.48)$$

This kind of constraint are called *Dirichlet boundary conditions*.

The solution procedure is quite straightforward. We take the general solution to the wave equation

$$y(x, t) = f(x - ct) + g(x + ct), \quad (1.49)$$

and we substitute it into the boundary condition, finding that

$$g(ct) = -f(-ct) \implies g(s) = -f(-s), \quad \forall s \in \mathbb{R}. \quad (1.50)$$

Thus our solution is

$$y(x, t) = f(x - ct) - f(-x - ct), \quad \forall x \leq 0, \forall t \in \mathbb{R}. \quad (1.51)$$

Let us spend a few lines to unpack this expression. The solution (1.51) is the sum of two waves. A right-moving one $f(x - ct)$ and a left-moving one $-f(-x - ct)$. The latter is equal to the former, flipped front-to-back and up-to-low. Some of you might be slightly confused: we said we would be looking at what happens to a single right-moving wave! Why are there two waves in the solution? Let us look at a simple example to lift the fog. Consider $f(x)$ to be a function highly localised around $x = -|x_0| < 0$ and with typical width δ . With this we mean that $f(x)$ rapidly approaches 0 outside the interval $(-|x_0| - \delta, -|x_0| + \delta)$. We suppose that x_0 is sufficiently large that $f(x) \sim 0$ for $x \lesssim 0$, so that the packet does not “feel” the boundary at the beginning of our “experiment” (i.e. for small t). With this choice, the second wave in the solution (1.51) is localised, at $t = 0$, around $x = +|x_0| > 0$. But in our setting the right half of the real line does not exist! So, effectively, we don’t see the second wave until the time $t \sim |x_0| - \delta$, when the right-moving wave starts to interact with the boundary. This simple thought experiment is depicted in Figure 1.8.

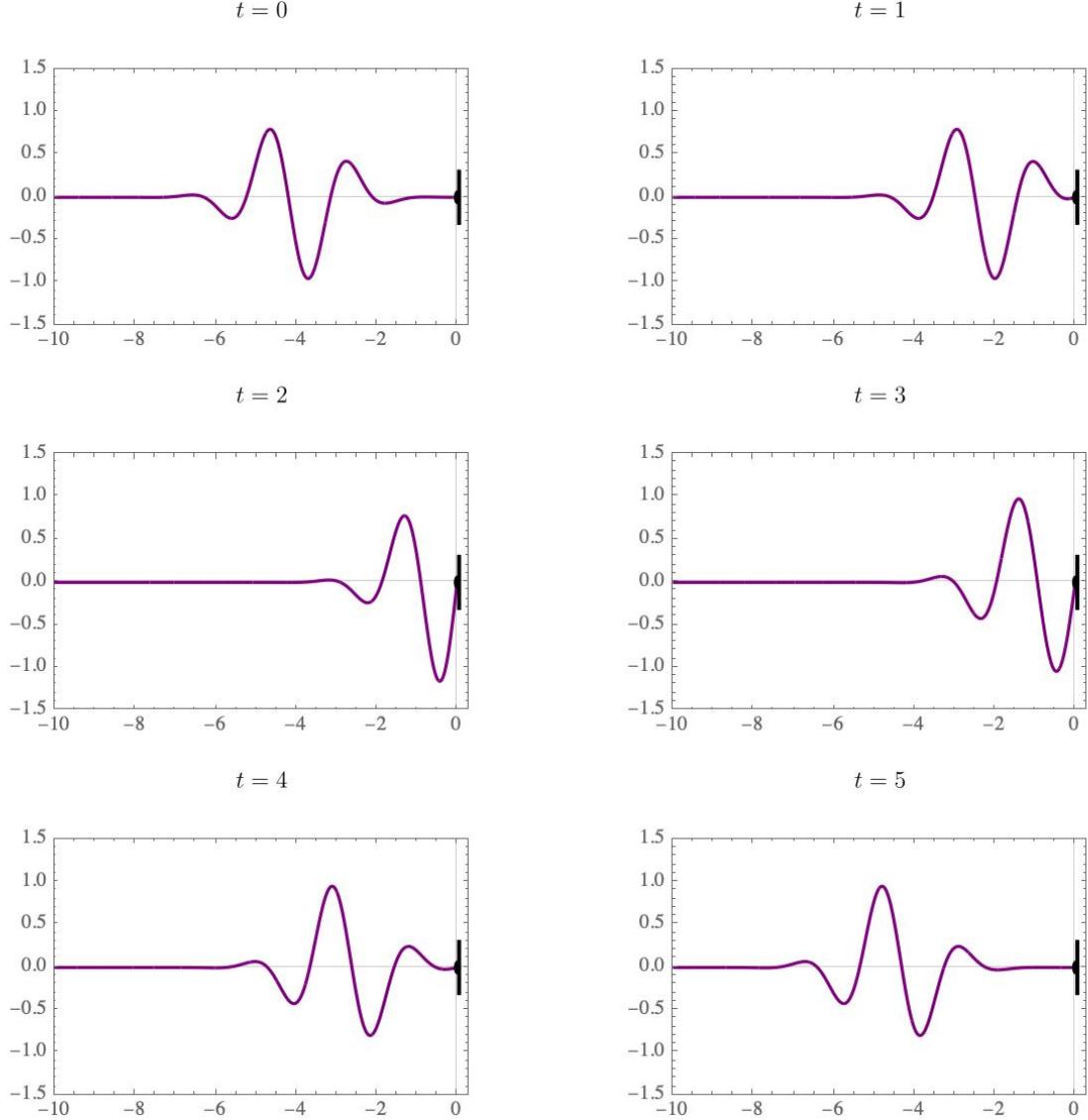


Figure 1.8: Time snapshots of the solution to the wave equation with a Dirichlet boundary condition at $x = 0$ and $f(x) = -e^{-(x+4)^2/2} \sin(3x)$. You can see the wave packet is localised on the negative real line for $t = 0$. As time passes, it moves to the right, eventually interacting with the boundary at $x = 0$. After enough time has passed, the wave is completely reflected and travels undisturbed leftwards to $-\infty$.

1.3.2 Reflection at a free end

Now let us perform the same analysis as before, but for an endpoint that is free to move. For example, you can think that the end of the string is attached to a (massless) contraption able to slide vertically along a rod without any friction. If you think a moment about what this means, you realise that the vertical component of the force at the endpoint must be zero. In turn, this means that the slope of the shape function has to be horizontal, that is

$$\partial_x y(x, t) \Big|_{x=0} = 0 , \quad \forall t \in \mathbb{R} . \quad (1.52)$$

Constraints that specify the value of the derivatives of the unknown function at certain points are known as *Neumann boundary conditions*. The specific one above is the *homogeneous Neumann boundary condition*.

geneous Neumann boundary condition.

Just as before, we substitute the general solution into (1.52), obtaining

$$f'(-ct) + g'(ct) = 0 \implies g(s) = f(-s) + C, \quad \forall s \in \mathbb{R}, \quad (1.53)$$

where C is an integration constant. The only effect it has is to rigidly move the whole string up or down the vertical direction. We set it to 0. Hence the solution reads

$$y(x, t) = f(x - ct) + f(-x - ct), \quad \forall x \leq 0, \quad \forall t \in \mathbb{R}. \quad (1.54)$$

Again, we have two waves, the incoming one and the reflected one which has the same shape as the former but is reflected front-to-back (no up-to-down reflection occurs). Figure 1.9 depicts the interaction with a free boundary of the same wave-packet as in 1.8.

1.3.3 Reflection and transmission at an interface

We have seen what happens when a wave interacts with a boundary, either fix or free to move. Now think about the following setup: we have two semi-infinite strings of different densities $\rho_1 \neq \rho_2$. We join them at their ends and suppose that the tension stays the same on the two sides of the juncture. In this subsection we wish to study the propagation of a wave as it travel from the first string, across the joint and into the second string. Pause a moment to think about what might happen. First, we immediately know that the wave speed is different for the two strings

$$c_1 = \sqrt{\frac{T}{\rho_1}}, \quad c_2 = \sqrt{\frac{T}{\rho_2}}. \quad (1.55)$$

Suppose for a moment that $\rho_1 \ll \rho_2$, that is, the second string is much heavier than the first one. When a wave travelling on the first string reaches the junction, the second, heavier string will offer a lot of resistance and will act kind of like a fixed end. Thus we expect the wave to be reflected back along the first string. However, the second string is not completely immovable and will yield, albeit slightly, to the motion of the incident wave. Hence we also expect part of the wave to be *transmitted* along the second string. Let us see how this phenomenon comes about mathematically.

We place the first string of density ρ_1 along the negative part of the real line $-\infty < x < 0$ and the second string, with density ρ_2 on the positive half-line $0 < x < \infty$. The junction is at $x = 0$ and we take $\rho_1 \neq \rho_2$ but otherwise arbitrary. Since the wave speed is different on the two sides of the junction, we start with a piece-wise solution

$$y(x, t) = \begin{cases} f_1(x - c_1 t) + g_1(x + c_1 t), & x < 0 \\ f_2(x - c_2 t) + g_2(x + c_2 t), & x > 0 \end{cases}. \quad (1.56)$$

Our study subject is a right-moving wave that is localised, at $t = 0$, on the left half-line. Mathematically, this means that we are asking¹¹

$$\begin{cases} f_1(x) \approx 0, & \forall x > 0 \\ g_1(x) \approx 0, & \forall x < 0 \end{cases}, \quad \begin{cases} f_2(x) \approx 0, & \forall x > 0 \\ g_2(x) \approx 0, & \forall x > 0 \end{cases}. \quad (1.57)$$

¹¹We are using the \approx sign out of an excess of zeal. In fact, we cannot ask a smooth function to be identically zero on an interval. However, we can ask it to be so small as to be completely negligible (think of something like $e^{-(x+10)^2}$ for $x > 0$).

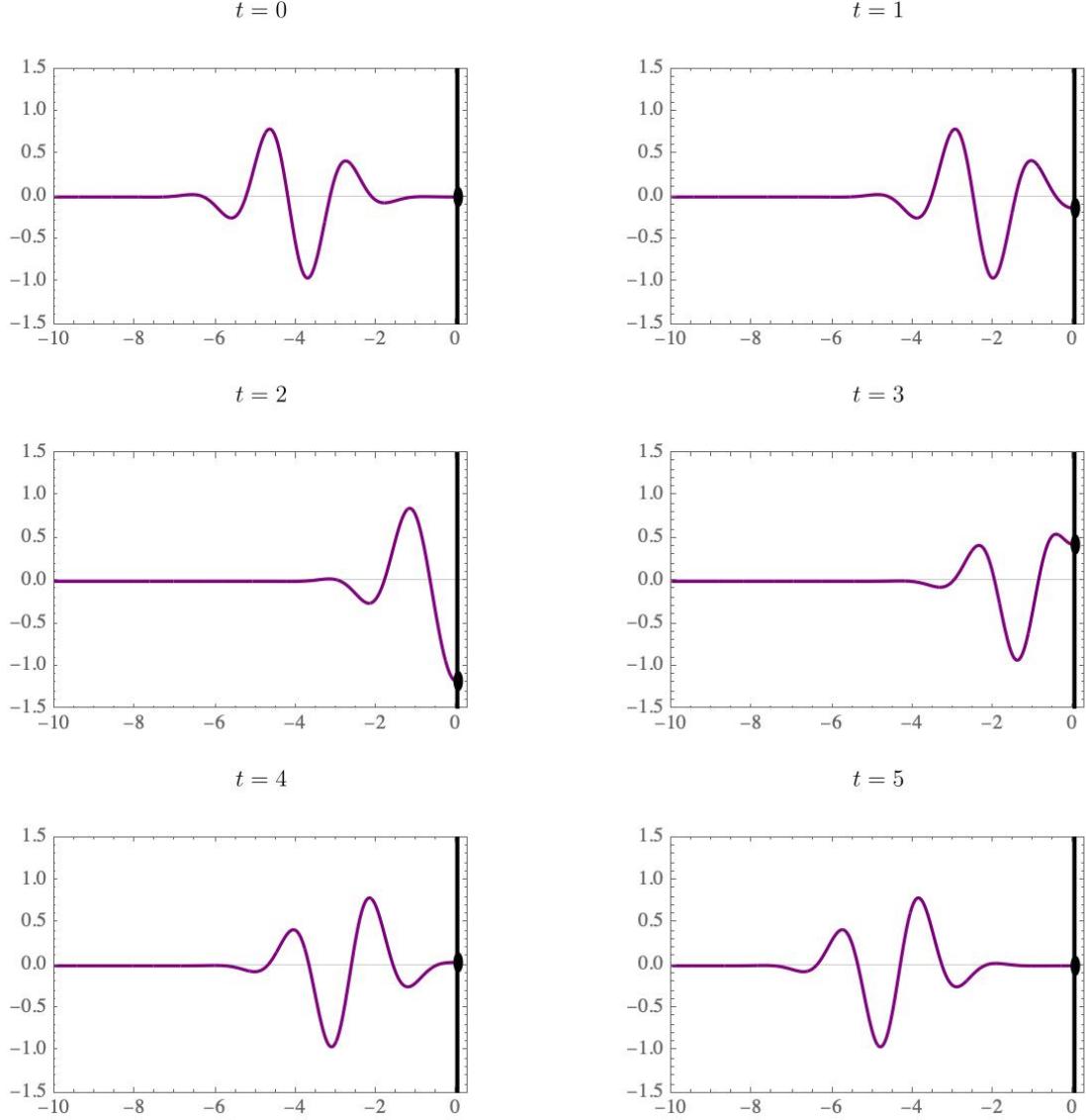


Figure 1.9: Time snapshots of the solution to the wave equation with a homogeneous Neumann boundary condition at $x = 0$ and $f(x) = -e^{-(x+4)^2/2} \sin(3x)$. You can see that, when the wave interacts with the boundary, the connection point slides up and down on the $x = 0$ vertical line. Again, the wave is reflected but this time there is no up-to-down swapping.

Notice that, since $c_2 > 0$ and we are taking the evolution to happen for $0 \leq t < \infty$, these conditions allow us to fix $g_2(x) = 0$ identically.

So, our piece-wise solution is

$$y(x, t) = \begin{cases} f_I(x - c_1 t) + f_R(x + c_1 t), & x < 0 \\ f_T(x - c_2 t), & x > 0 \end{cases}, \quad (1.58)$$

where we renamed our functions to highlight their role: f_I is the incident, right-moving wave, f_R is the reflected, left-moving wave, and f_T is the transmitted, right-moving one. We take the first to be a given function, our input, and we wish to determine the other two. To do so, we impose *continuity* of the solution and of its first derivative at the

junction

$$f_I(-c_1 t) + f_R(+c_1 t) = f_T(-c_2 t) , \quad (1.59a)$$

$$f'_I(-c_1 t) + f'_R(+c_1 t) = f'_T(-c_2 t) . \quad (1.59b)$$

The first equation imposes the condition that the strings are joined at $x = 0$, while the second asks the slope to be continuous across $x = 0$. If this was not so, then it would imply the presence of an unphysical, external force. Let us solve them, starting from the first one

$$f_T(s) = f_I\left(\frac{c_1}{c_2}s\right) + f_R\left(-\frac{c_1}{c_2}s\right) , \quad s = -c_2 t . \quad (1.60)$$

We can substitute this into the other equation

$$f'_I\left(\frac{c_1}{c_2}s\right) + f'_R\left(-\frac{c_1}{c_2}s\right) = \frac{c_1}{c_2}f'_I\left(\frac{c_1}{c_2}s\right) - \frac{c_1}{c_2}f'_R\left(-\frac{c_1}{c_2}s\right) , \quad (1.61)$$

and, shuffling the furniture, arrive at

$$f'_R(\sigma) = \frac{c_1 - c_2}{c_1 + c_2}f'_I(-\sigma) , \quad \sigma = -\frac{c_1}{c_2}s . \quad (1.62)$$

Integrating this expression¹² we can then express f_R and f_T in terms of f_I

$$f_R(\sigma) = -\frac{c_1 - c_2}{c_1 + c_2}f_I(-\sigma) , \quad (1.63a)$$

$$f_T(s) = 2\frac{c_2}{c_1 + c_2}f_I\left(\frac{c_1}{c_2}s\right) . \quad (1.63b)$$

In conclusion the solution formula reads as follows

$$y(x, t) = \begin{cases} f_I(x - c_1 t) + A_R f_I(-x - c_1 t) , & x < 0 \\ A_T f_I\left(c_1 \frac{x - c_2 t}{c_2}\right) , & x > 0 \end{cases} , \quad (1.64)$$

where we introduced the *reflection amplitude* A_R and the *transmission amplitude* A_T , given by the expressions

$$A_R = \frac{c_2 - c_1}{c_1 + c_2} , \quad A_T = 2\frac{c_2}{c_1 + c_2} . \quad (1.65)$$

Let us consider some limiting cases to check whether our solution is sensible.

- Take the case $\rho_1 = \rho_2$. Then $c_1 = c_2$ and

$$A_R = 0 , \quad A_T = 1 , \quad y(x, t) = f_I(x - c_1 t) , \quad \forall x \in \mathbb{R} .$$

This is exactly what we expect from an infinite homogeneous string.

- Suppose that $\rho_1 \ll \rho_2$, so that the right string is much heavier than the left one. Then we have

$$A_R \approx -1 , \quad A_T \approx 0 .$$

The heavy string here acts quite like a fixed point, reflecting almost completely the incoming wave.

¹²We fixed the integration constant to 0 for the same reasons we did so in §§1.3.2.

- Take the opposite case $\rho_1 \gg \rho_2$. Now

$$A_R \approx 1, \quad A_T \approx 2.$$

In this case the second string is so light that it practically does not affect the dynamics at $x = 0$. This acts almost as a free point as in the Neumann boundary condition.

As we did for the cases we studied before, we collected in Figure 1.10 the plots of the same wave packet as in 1.8, evolving along the juncture of two strings with different densities.

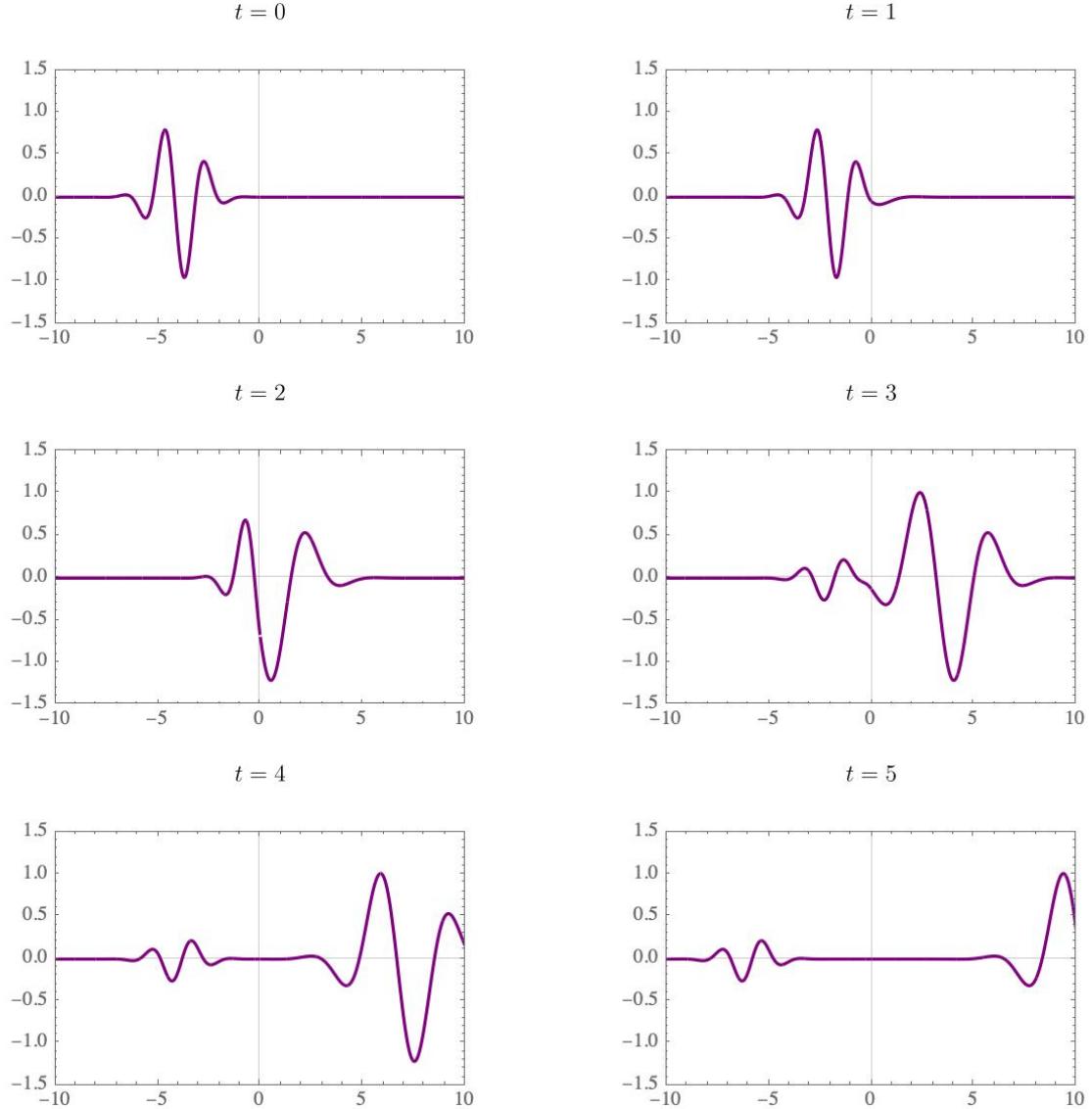


Figure 1.10: Time snapshots of the solution to the wave equation on an infinite string with a change-of-density interface at $x = 0$. The incoming wave packet is $f(x) = -e^{-(x+4)^2/2} \sin(3x)$. You can see that, when the wave reaches the interface, it deforms and, after some time, it splits into a reflected and a transmitted wave. In these plots we used a ratio $c_2/c_1 = 7/4$.

1.4 The solution of Bernoulli

Now that we have studied how waves propagate on infinitely long strings and how boundaries and interfaces affect this propagation, it is time to get somewhat more real. In this section we are going to study the dynamics of waves on a finite string. Before doing that, though, we will solve the wave equation again, using a completely different method than the d'Alembert one. We are going to follow the approach of the swiss mathematician Daniel Bernoulli. It is actually a procedure you have encountered before, under the name of *separation of variables*.

1.4.1 Separation of variables

The method of separation of variables for the solutions of PDEs starts with an *ansatz*¹³: we assume that the solution factorises into functions, each depending on only one of the variables. In our case, we start from the form

$$y(x, t) = X(x)T(t) . \quad (1.66)$$

By substituting this expression into the wave equation

$$\partial_t^2 y(x, t) = c^2 \partial_x^2 y(x, t) , \quad (1.67)$$

we obtain

$$\frac{X''(x)}{X(x)} = \frac{1}{c^2} \frac{T''(t)}{T(t)} . \quad (1.68)$$

This equation is in *separated form*: the left-hand side only depends on x , while the right-hand side only depends on t . Since the equation must hold for all values of x and t , the equation can only hold if both sides are equal to a constant, that we call $-k^2$ for future convenience

$$\frac{X''(x)}{X(x)} = \frac{1}{c^2} \frac{T''(t)}{T(t)} = -k^2 . \quad (1.69)$$

By splitting the variable dependence, we have passed from a single PDE to two ODEs, one for $X(x)$ and one for $T(t)$

$$X''(x) = -k^2 X(x) , \quad (1.70a)$$

$$T''(t) = -k^2 c^2 T(t) . \quad (1.70b)$$

Their general solution is famously

$$X(x) = A \cos(kx) + B \sin(kx) , \quad (1.71a)$$

$$T(t) = F \cos(kct) + G \sin(kct) , \quad (1.71b)$$

so that the propagating wave takes the form

$$y(x, t) = \left(A \cos(kx) + B \sin(kx) \right) \left(F \cos(kct) + G \sin(kct) \right) . \quad (1.72)$$

¹³In physics and mathematics, an *ansatz* is an educated guess made to facilitate the solution of a problem. The validity of such a guess is checked *ex post* by checking that the final results agree with the initial setting and hypotheses. The word *ansatz* comes from German and it literally translates as “initial placement of a tool at a work piece”. Grammar note: the plural of *ansatz* is *ansätze*.

Such a solution depends on 5 arbitrary constants: A, B, F, G and k .

You might be feeling a sense of unease at this point. In the previous sections, we saw that the general solution to the wave equation is the sum of two travelling waves $f(x - ct)$ and $g(x + ct)$, with f and g being arbitrary functions. Here we just derived another solution in terms of sines and cosines, with a factorisation of the dependence on x and t and only 5 arbitrary constants! How do these two facts square together? The typical $x \pm ct$ dependence is easily reconstructed in (1.72) by using some trigonometric identities

$$\begin{aligned} y(x, t) = & \frac{AF + BG}{2} \cos(k(x - ct)) + \frac{BF - AG}{2} \sin(k(x - ct)) + \\ & + \frac{AF - BG}{2} \cos(k(x + ct)) + \frac{BF + AG}{2} \sin(k(x + ct)) . \end{aligned} \quad (1.73)$$

To reconcile the apparent difference in the undetermined data (functions vs. constants), we need to appeal to an incredibly powerful concept: *the superposition principle*.

1.4.2 Finite strings: standing waves and superpositions

Let us consider a string of finite length, with both ends fixed. We choose coordinates so that, in its equilibrium position, the string lies on the interval $[0, \pi]$, with $x = 0$ and $x = \pi$ being the fixed ends. In this setup, we have two Dirichlet conditions

$$y(0, t) = 0 , \quad y(\pi, t) = 0 , \quad \forall t . \quad (1.74)$$

Using the separated form (1.66), we see that these conditions are satisfied if

$$X(0) = 0 , \quad X(\pi) = 0 . \quad (1.75)$$

Now, from the solution (1.71a) we immediately find that

$$A = 0 , \quad B \sin(k\pi) = 0 , \quad (1.76)$$

which, by discarding the trivial zero solution $A = B = 0$, implies

$$A = 0 , \quad k \in \mathbb{Z} . \quad (1.77)$$

Thus we found that the wave equation on a finite string of length π with fixed ends has the form

$$y_k^{\text{D,D}}(x, t) = \sin(kx) \left(F_k \cos(kct) + G_k \sin(kct) \right) , \quad \forall k \in \mathbb{Z} , \quad (1.78)$$

where we redefined the constants $F_k = BF$ and $G_k = BG$, and we added to y an index k and a superscript D,D to remind that this solution is valid for Dirichlet conditions at both ends.

This was fun. Let us try different boundary conditions: two Neumann ones

$$\partial_x y(x, t) \Big|_{x=0} = 0 , \quad \partial_x y(x, t) \Big|_{x=\pi} = 0 . \quad (1.79)$$

Following the same steps we arrive at the equations

$$B = 0 , \quad A \sin(k\pi) = 0 . \quad (1.80)$$

So, just as before, the constant k has to be an integer $k \in \mathbb{Z}$, but the solution will now depend on x through the cosine function

$$y_k^{\text{N,N}}(x, t) = \cos(kx) \left(F_k \cos(kct) + G_k \sin(kct) \right), \quad \forall k \in \mathbb{Z}. \quad (1.81)$$

We can also consider a string with different conditions at the two ends, like Dirichlet at $x = 0$ and Neumann at $x = \pi$. Try to work out what the solution will look like and what values are allowed for the constant k . Figure 1.11 collects plots of the evolution of waves on a finite string with the 4 possible combinations of boundary conditions.

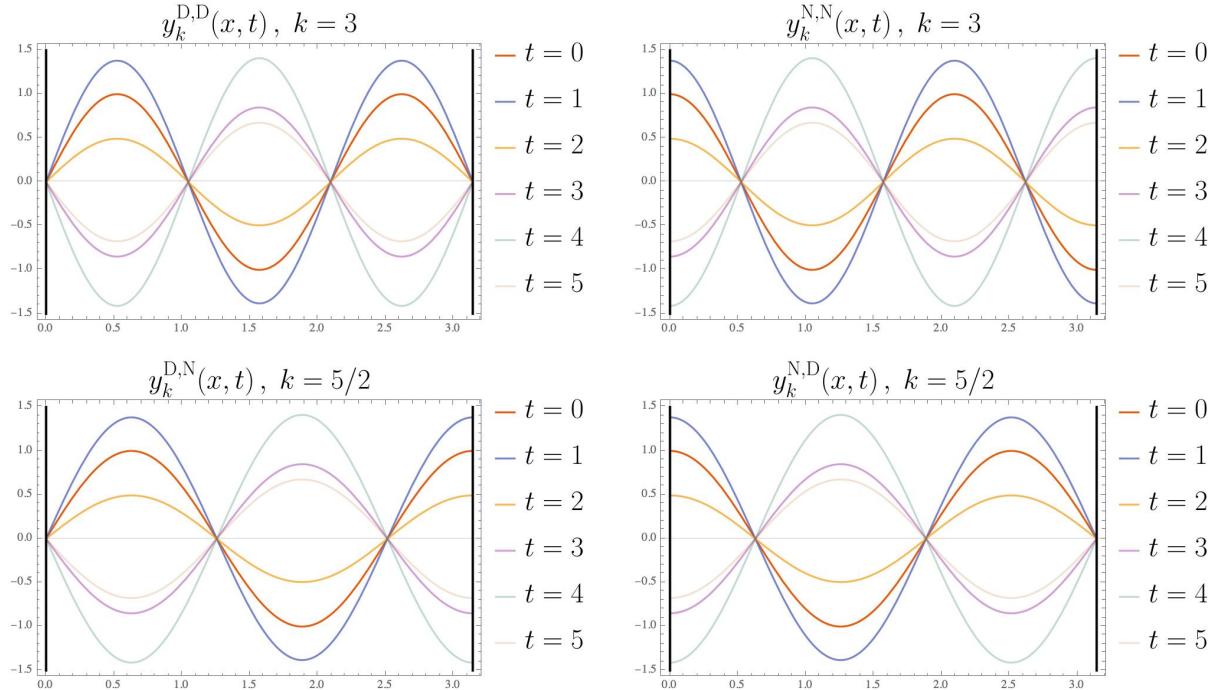


Figure 1.11: Time evolution of standing waves on a finite string with 4 different combinations of boundary conditions. Notice how the waves “do not move” horizontally. They simply vary in amplitude as time passes.

Functions of the type (1.78) and (1.81) are known as *standing waves*. They are solutions of the wave equation that do not travel along the x direction, but simply oscillate in amplitude, preserving their shape (up to re-scaling in the vertical direction). Here we are hitting again on the stark contrast between these type of solutions and the one we obtained with d’Alembert approach in the previous sections. These were travelling waves, moving leftwards or rightwards along the string with a speed c dictated by the physical properties of the string. Now we found perfectly legal solutions that do not travel at all, but wobble in their place, even though the wave speed c is still non-zero. The way out of this conflict is the already-mentioned *superposition principle*.

The superposition principle

In the words of Bernoulli himself,

The general motion of a vibrating system is given by a superposition of its proper vibrations.

These words are the first known instance of the much more general superposition principle, that states

For all linear systems, the sum of any number of solutions is still a solution.

Mathematically this means that, given a linear system^a

$$LX = 0 ,$$

with solutions $\{X_i\}_{i=1}^m$ for some $m \geq 1$, then any linear combination

$$Y = \sum_{i=1}^m \alpha_i X_i , \quad \alpha_i \in \mathbb{C} ,$$

is still a solution

$$L(Y) \equiv 0 .$$

^aHere you can take L to be a matrix and X a vector, or L a differential operator (such as $\partial_t^2 - c^2 \partial_x^2$) and X a function. You can also consider systems of linear differential equations and more complicated examples.

Our wave equation clearly qualifies as a linear system and the boundary conditions we imposed are linear as well. This means that any linear combination of, say, (1.78) with different values of $k \in \mathbb{Z}$ will still solve the wave equation. Then the most general solution for the Dirichlet-Dirichlet boundary problem is¹⁴

$$y^{D,D}(x, t) = \sum_{k=1}^{\infty} \sin(kx) \left(F_k \cos(kct) + G_k \sin(kct) \right) . \quad (1.82)$$

It turns out that a superposition of standing wave such as (1.82) can give rise to a travelling wave. Even more is true: a superposition of sine and cosine waves can generate any shape! This last fact lies at the core of *Fourier analysis*, a topic that you will meet later in your curriculum. For the moment, you can get a taste of this superposition phenomenon by looking at (1.73). There, two proper travelling waves are superimposed to generate a standing one. Figure 1.12 collects some snapshots of this solution, along with the two travelling waves that it is composed of.

¹⁴Note that there is no need to include negative integers $k < 0$ in the sum, since they produce the same solution (up to a sign). Also $k = 0$ is omitted since $y_0^{D,D}(x, t) = 0$.

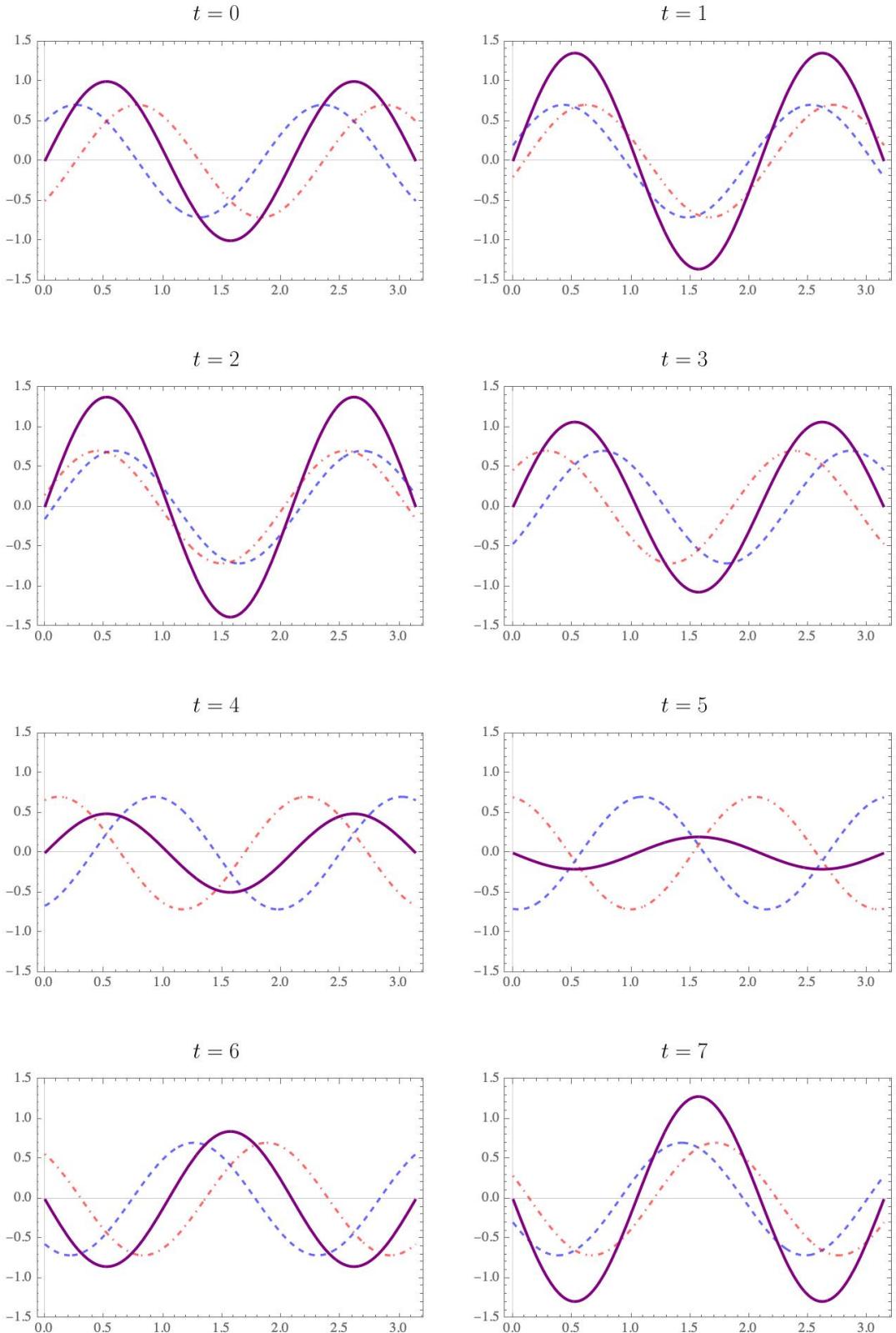


Figure 1.12: Time snapshots of the solution (1.82) (in purple, full). In blue (dashed) and red (dot-dashed) are plotted, respectively, the right-moving and left-moving waves whose superposition produces the full solution.

1.4.3 The initial value problem

The last step in our analysis of waves on a finite string is to impose initial conditions on the Bernoulli solution (1.82) we have found. Choosing some initial shape $y_0(x)$ and initial velocity $v_0(x)$, we impose

$$y(x, 0) = y_0(x), \quad \partial_t y(x, t) \Big|_{t=0} = v_0(x), \quad \forall x \in [0, \pi]. \quad (1.83)$$

The goal now is to completely fix the solution, i.e. to determine F_k and G_k for all $k \in \mathbb{Z}$. Substitution of (1.82) into (1.83) produces

$$y_0(x) = \sum_{k=1}^{\infty} F_k \sin(kx), \quad v_0(x) = \sum_{k=1}^{\infty} kcG_k \sin(kx). \quad (1.84)$$

To extract the coefficients from the sums, we recall a nifty little identity from Fourier analysis, known as *orthogonality relation*:

$$\int_0^\pi dx \sin(kx) \sin(lx) = \frac{\pi}{2} \delta_{k,l}, \quad (1.85)$$

where $\delta_{k,l}$ is the *Kronecker delta* function, equal to 1 if $k = l$ and zero otherwise

$$\delta_{k,l} = \begin{cases} 1, & k = l \\ 0, & k \neq l \end{cases}. \quad (1.86)$$

Thanks to this formula, we can invert the equations (1.84), finding

$$F_k = \frac{2}{\pi} \int_0^\pi dx y_0(x) \sin(kx), \quad G_k = \frac{2}{\pi k c} \int_0^\pi dx v_0(x) \sin(kx). \quad (1.87)$$

The method of separation of variables that we have used in this lecture to investigate the wave equation is a powerful technique that can be used to solve a wide range of PDEs beyond the one we considered here. The idea behind this method is very simple: we assume that the solution can be expressed as a product of functions, each depending only on one of the variables (there may be more than two!). Asking that the solution to take this factorised form reduces the PDE to a set of independent ODEs, which can be solved one by one. We will use this method again later in this module, when dealing with waves on 2-dimensional surfaces and when studying quantum mechanics. Beyond this module, this method will serve you well in several instances: in fluid mechanics, advanced quantum mechanics, mathematical ecology and epidemiology and many other fields.

1.5 Additional content: damping

Most of you will have first-hand experience with vibrating bodies. After they are struck, the vibration wanes and, after enough time has elapsed, it stops completely. The model that we constructed, on the other hand, predicts that a wave will persevere in its motion forever, without

ever losing even a pinch of its energy. We can obviate this departure from reality by introducing a *damping force* in our model. We will not need to be precise as to what physical phenomenon is causing the damping. It could be the drag of the air, for example. We will model this effect as a force proportional to the transverse velocity of the string

$$F_{\text{drag}} = -\beta \delta s \partial_t y(x, t) , \quad (1.88)$$

where β is a constant measuring the strength of the damping effect. It is measured in units $[\beta] = M/(LT)$. The tension forces remain the same as before, so we can immediately write down the modified equation

$$\partial_t^2 y(x, t) + \lambda \partial_t y(x, t) = c^2 \partial_x^2 y(x, t) , \quad (1.89)$$

where $\lambda \equiv \beta/\rho$ is measured in units $[\lambda] = 1/T$.

We need to choose the type of system we wish to consider. Let us take a string with two fixed ends

$$y(0, t) = 0 , \quad y(\pi, t) = 0 , \quad \forall t . \quad (1.90)$$

To find the solution of (1.89) we follow Bernoulli and take the factorised ansatz

$$y(x, t) = X(x)T(t) , \quad X(0) = X(\pi) = 0 , \quad (1.91)$$

that brings us to the equations

$$X''(x) = -k^2 X(x) , \quad (1.92a)$$

$$T''(t) = -\lambda T'(t) - k^2 c^2 T(t) . \quad (1.92b)$$

We know how to solve both of these. Their solutions are

$$X(x) = A \cos(kx) + B \sin(kx) , \quad (1.93a)$$

$$T(t) = F_+ e^{-\omega_+(k)t} + F_- e^{-\omega_-(k)t} , \quad (1.93b)$$

where

$$\omega_{\pm}(k) = \frac{\lambda \pm \sqrt{\lambda^2 - 4k^2 c^2}}{2} . \quad (1.94)$$

The boundary conditions, just as the case we studied above, tell us that

$$A = 0 , \quad k \in \mathbb{Z} , \quad (1.95)$$

and so, remembering the superposition principle, we can write the general solution as

$$y(x, t) = \sum_{k=1}^{\infty} \sin(kx) \left[F_+(k) e^{-\omega_+(k)t} + F_-(k) e^{-\omega_-(k)t} \right] . \quad (1.96)$$

Notice that we have two kind of waves that enter the superposition (1.96), depending on the sign of $\lambda^2 - 4k^2 c^2$. All the terms with $k^2 < \lambda^2/(4c^2)$ will have purely real exponents, so the corresponding coefficients $F_{\pm}(k)$ can be taken to be real. When $k^2 > \lambda^2/(4c^2)$, however, the exponents will be complex

$$\omega_{\pm}(k) = \kappa \pm i\Omega(k) , \quad \begin{cases} \kappa = \lambda/2 \\ \Omega(k) = \sqrt{4k^2 c^2 - \lambda^2}/2 \end{cases} , \quad \forall k > \frac{\lambda}{2c} . \quad (1.97)$$

In this case, the coefficients $F_{\pm}(k)$ will need to be complex as well, in order for the solution $y(x, t)$ to be real. In particular, we take

$$F_+(k) = \frac{C(k) + iD(k)}{2} , \quad F_-(k) = \frac{C(k) - iD(k)}{2} , \quad (1.98)$$

so that the solution can be expressed as

$$y(x, t) = \sum_{k=1}^{\lfloor \frac{\lambda}{2c} \rfloor} \sin(kx) [F_+(k)e^{-\omega_+(k)t} + F_-(k)e^{-\omega_-(k)t}] + \sum_{k=\lceil \frac{\lambda}{2c} \rceil}^{\infty} \sin(kx)e^{-\kappa t} [C(k) \cos(\Omega(k)t) + D(k) \sin(\Omega(k)t)]. \quad (1.99)$$

Here, the symbols $\lfloor x \rfloor$ and $\lceil x \rceil$ stand, respectively, for the *floor* and *ceiling* – i.e. the lowest and highest closest integer – of the real number x . We see that all the components – the *harmonics* to borrow a term from a future section – of the above wave with k smaller than $\lambda/(2c)$ are completely damped, in the sense that they do not evolve in an oscillatory way: they experience an exponential decay. On the contrary, the components with $k > \lambda/(2c)$ do present an oscillatory evolution, though they also decay due to the exponential $e^{-\kappa t}$. Eventually, the entire string will settle on its rest position and all movement will stop. Technically, this only happens in the limit $t \rightarrow \infty$, but for $t \gg \kappa$ any motion is effectively imperceptible.

With the expression (1.99) available, one can impose some initial conditions $y(x, 0) = y_0(x)$, $\partial_t y(x, t)|_{t=0} = v_0(x)$ and study the propagation of an initial wave along the string. Figure 1.13 plots an example.

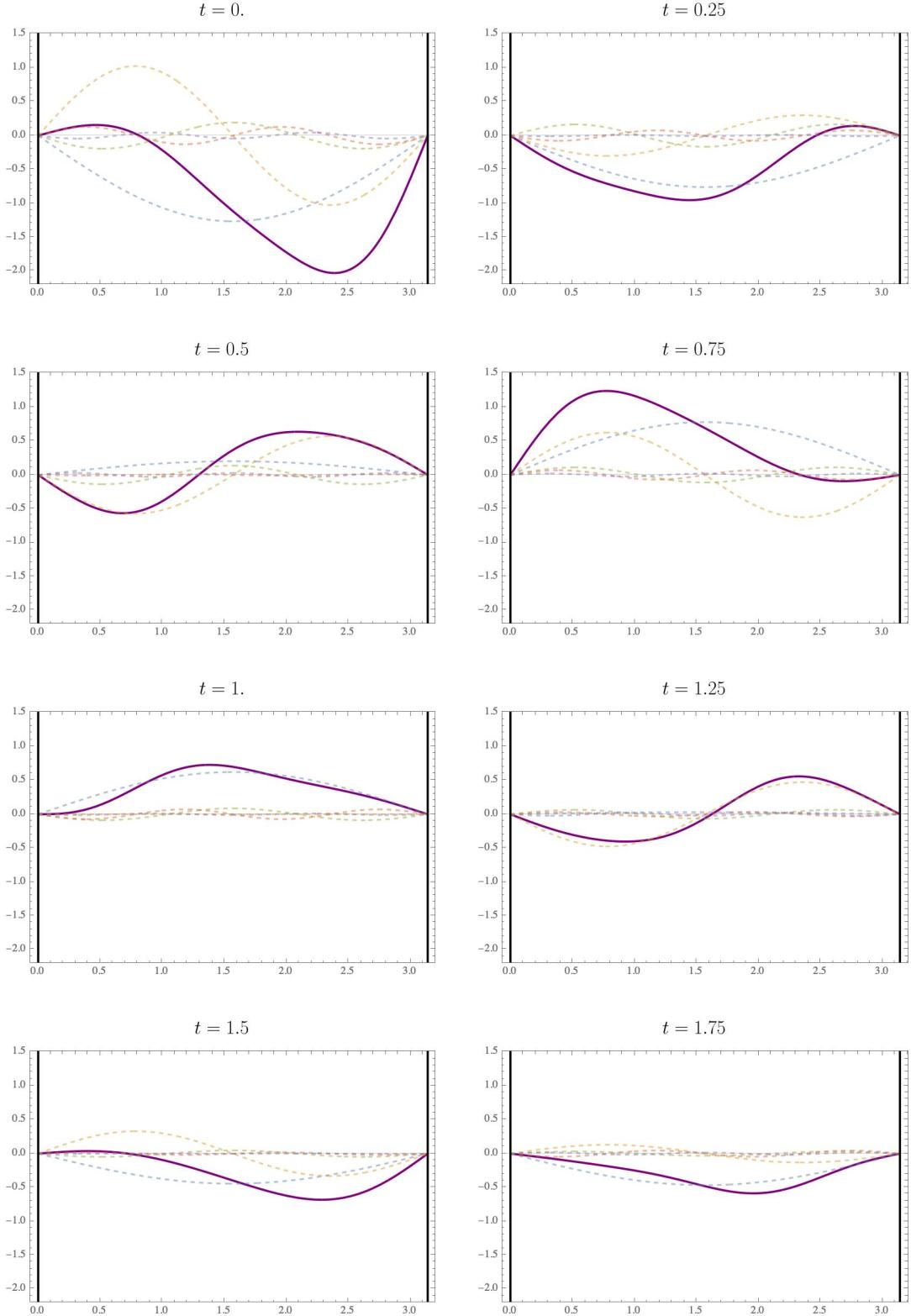


Figure 1.13: Time snapshots of the solution (1.99) (in purple, full) for initial conditions $v_0(x) = 0$ and $y_0(x) = a(x - \pi/8)^4 + b(x - \pi/8)^2 + c$, with a, b, c some constants chosen to have nice plots. We set $c = 2$ and $\lambda = 1.1$. The dashed lines are the first 5 harmonics of the solution. Notice how the higher harmonics are decaying faster in amplitude with respect to the fundamental.

Outro: takeaway points

In this chapter we have explored the physics of waves in 1 space dimension. The takeaway points are follow.

- The dynamics of waves on an idealised string is governed by the 1D wave equation

$$\frac{\partial^2}{\partial t^2}y(x,t) = c^2 \frac{\partial^2}{\partial x^2}y(x,t) .$$

The parameter c is the wave speed and is related to the physical parameters of the string, the tension T and the linear density ρ , as

$$c = \sqrt{\frac{T}{\rho}} .$$

- The 1D wave equation can be simplified by a change of variables $(x, t) \rightarrow (\xi, \eta) = (x + ct, x - ct)$

$$\frac{\partial^2}{\partial \xi \partial \eta}y(\xi, \eta) = 0 .$$

This forms makes it clear that the general solution is the sum of a right- and a left-moving waves

$$y(x, t) = F(x - ct) + G(x + ct) .$$

- With the general solution in hand, it is possible to solve the initial value problem and pin down a specific solution with initial profile $y_0(x)$ and initial speed $v_0(x)$ at time $t = 0$

$$y(x, t) = \frac{y_0(x - ct) + y_0(x + ct)}{2} + \frac{1}{2c} \int_{x-ct}^{x+ct} ds v_0(s) .$$

- We studied how waves reflect at boundaries, both in the Dirichlet (fixed) and in the Neumann (free) cases. We saw that a wave on a semi-infinite string with a boundary at $x = 0$ has a profile

$$y(x, t) = \begin{cases} f(x - ct) - f(-x - ct) , & \text{Dirichlet , } y(0, t) = 0 \\ f(x - ct) + f(-x - ct) , & \text{Neumann , } \partial_x y(x, t) \Big|_{x=0} = 0 \end{cases} .$$

- We studied how waves reflect and transmit at an interface – i.e. a junction between two strings of different densities (or tensions). If the wave speeds of the strings are c_1 and c_2 , the wave propagating on the strings joined at $x = 0$ has a profile

$$y(x, t) = \begin{cases} f_1(x - c_1 t) + A_R f_1(-x - c_1 t) , & x < 0 \\ A_T f_1 \left(c_1 \frac{x - c_2 t}{c_2} \right) , & x > 0 \end{cases} ,$$

where

$$A_R = \frac{c_2 - c_1}{c_1 + c_2} , \quad A_T = 2 \frac{c_2}{c_1 + c_2} ,$$

are the reflection and transmission coefficients, respectively.

- We learned how to solve the wave equation using the separation of variable method. This brought us to a solution of the form

$$y(x, t) = \left(A \cos(kx) + B \sin(kx) \right) \left(F \cos(kct) + G \sin(kct) \right).$$

- Studying waves on a finite string we learned the concept of standing waves

$$y_k^{D,D}(x, t) = \sin(kx) \left(F_k \cos(kct) + G_k \sin(kct) \right), \quad \forall k \in \mathbb{Z},$$

$$y_k^{N,N}(x, t) = \cos(kx) \left(F_k \cos(kct) + G_k \sin(kct) \right), \quad \forall k \in \mathbb{Z},$$

with D,D and N,N standing for Dirichlet-Dirichlet and Neumann-Neumann respectively. The two connotations indicate which type of boundary condition is imposed on the two endpoints of the string.

- We learned of the superposition principle and how a general solution can be built by summing an arbitrary (possibly infinite) number of standing waves

$$y^{D,D}(x, t) = \sum_{k=1}^{\infty} \sin(kx) \left(F_k \cos(kct) + G_k \sin(kct) \right).$$

- We saw how to pin down a specific solution with initial conditions $y_0(x)$ (profile) and $v_0(x)$ (speed) at $t = 0$, by using the orthogonality relation of sinusoidal functions. The end results are

$$F_k = \frac{2}{\pi} \int_0^{\pi} dx y_0(x) \sin(kx), \quad G_k = \frac{2}{\pi k c} \int_0^{\pi} dx v_0(x) \sin(kx).$$

Chapter 2

Energy and harmonic waves

Now that we have developed our sea legs on the example of a vibrating string, it is time to move to a more general investigation of waves. We will study in particular *harmonic waves*, also called *sinusoidal* or *monochromatic* waves. In this chapter, we will present their definition, introduce the relevant terminology, detail their importance in the analysis of PDEs and expound on their energy properties.

2.1 Harmonic waves

The wave equation $\partial_t^2 y - c^2 \partial_x^2 y = 0$ admits solutions of the form

$$y(x, t) = A \cos(k(x - ct)) + B \sin(k(x - ct)) + \\ + C \cos(k(x + ct)) + D \sin(k(x + ct)), \quad (2.1)$$

composed of 4 sinusoidal functions, two right-moving and two left-moving. Each of these is a specific instance of the general function

$$h(x, t) = a \cos(kx - \omega t + \phi), \quad (2.2)$$

sometimes also written as

$$h(x, t) = a \cos(2\pi(\hat{k}x - \nu t) + \phi). \quad (2.3)$$

This class of function depends on 4 real constants a , \hat{k} , ν and ϕ . They are called, respectively, *amplitude*, *wave number*, *frequency* and *phase*. The rescaled constants

$$k = 2\pi\hat{k}, \quad \omega = 2\pi\nu, \quad (2.4)$$

are called *angular wave number* and *angular frequency* to distinguish them from \hat{k} and ν . These rescaled quantities k, ω are used more frequently by mathematicians, though when mentioned the connotation “angular” is often dropped. Thus pay attention when perusing

the literature. Concerning dimensions and units of measure, we have

$$[a] \equiv [h], \quad \text{the unit of measure depends on what } h \text{ describes ,} \quad (2.5a)$$

$$[k] \equiv [\hat{k}] = \frac{1}{L}, \quad \text{the unit of measure is m}^{-1}, \quad (2.5b)$$

$$[\omega] \equiv [\nu] = \frac{1}{T}, \quad \text{the unit of measure is Hz = s}^{-1}, \text{ short-hand for "Hertz" ,} \quad (2.5c)$$

$$[\phi] = 1, \quad \text{the phase is a pure number, measured in radians .} \quad (2.5d)$$

Other quantities of widespread use are the *period* $P = 1/\nu = 2\pi/\omega$, measured in seconds and the *wavelength* $\lambda = 1/\hat{k} = 2\pi/k$, measured in meters.

The functions of the form (2.2) are so relevant and widespread as to deserve their own name: *harmonic waves*. Sometimes they are also referred to as *sinusoidal waves* for evident reasons, or *monochromatic waves*. The term monochromatic refers to the fact that $h(x, t)$ above oscillate in time with a unique frequency ν . As you are going to see later on in your curriculum, light is also composed of waves and each frequency is associated to a specific colour¹, hence the name.

2.1.1 More on harmonic wave properties

Let us look more closely at these quantities that we just introduced. Figure 2.1 plots a harmonic wave $h(x, t)$ and depicts the geometry of its characteristic quantities.

- Since all the constants and variables in $h(x, t)$ are assumed real, we have

$$\max_{x \text{ or } t} h(x, t) = a, \quad (2.6)$$

given that a cosine oscillate between ± 1 .

So the amplitude a determines the maximal displacement of the wave from the x axis². The points at which $h(x, t) = a$ are called *crests* or *peaks*, while those at which $h(x, t) = -a$ are called *troughs*. On the other hand, the points at which $h(x, t) = 0$ are called *nodes*.

- The harmonic wave (2.2), being a sinusoid, is periodic in x

$$h(x + n/\hat{k}, t) = h(x, t), \quad \forall n \in \mathbb{Z}. \quad (2.7)$$

This spatial period is precisely the wavelength $\lambda = 1/\hat{k}$. This quantity can be equivalently defined as the distance between two consecutive crests (or troughs).

- The wave number \hat{k} is the number of crests (or troughs) inside the interval $x \in (0, 1]$. Note that the interval is opened on one end and closed on the other.

¹Of course, frequency is a continuum as is the colour spectrum. For example, red is usually taken to lie in the range $\nu \in (400, 484) \times 10^{14}\text{Hz}$. Although you cannot really perceive visually the difference between, say, a wave with $\nu = 424 \times 10^{14}\text{Hz}$ and one with $\nu = 425 \times 10^{14}\text{Hz}$, they correspond to distinct colours.

²We will always be thinking of a wave oscillating along x and evolving in the time t , so it is natural to picture ourselves a plot of $h(x, t)$ on the (x, h) plane at a fixed instant $t = t_0$. Obviously one can swap the roles of x and t and plot the evolution of a single point $x = x_0$ in the (t, h) plane. It is however important to keep distinct the different physical roles of the variables and constants.

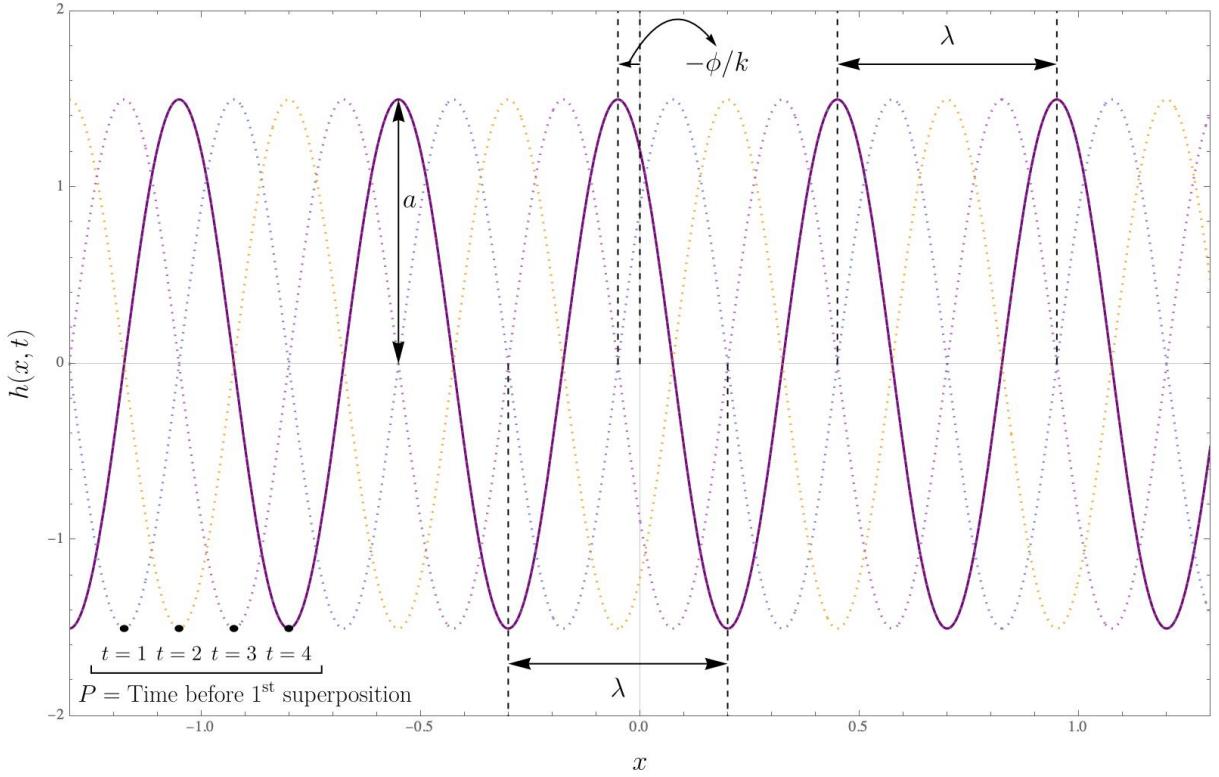


Figure 2.1: A harmonic wave (2.2) with $a = 3/2$, $\hat{k} = 2$, $\nu = 1/4$ and $\phi = \pi/5$. In full line, purple, is the plot of $h(x, t = 0)$, while the dotted lines are the function $h(x, t)$ at $t = 1, 2, 3$. For $t = 4$, the plot of $h(x, t)$ coincides with that of $h(x, 0)$. In the figure it is shown how to measure the characteristic features of the wave.

- Similarly, the wave $h(x, t)$ is periodic in t

$$h(x, t + n/\nu) = h(x, t), \quad \forall n \in \mathbb{Z}. \quad (2.8)$$

This period is, appropriately enough, the period $P = 1/\nu$. It is the time that elapses from a reference instant t_0 before the (x, h) plot of the wave superimposes with itself for the first time.

- The frequency ν is the number of times the (x, h) plot of the wave has superimposed with itself in a unit time interval $t \in (0, 1]$. Again, make sure the interval is open on one end.
- The phase ϕ measures, in units of the angular wave number, the displacement of the crest closest to the reference point $x = 0$, at a reference time $t = 0$. In fact, this displacement \tilde{x} can be found by asking $h(\tilde{x}, 0) = a$, which means³

$$\phi = -k\tilde{x}. \quad (2.9)$$

- One important additional characteristic quantity is the *wave speed* c , sometime also called the *phase speed*, which measure the speed at which the crests – equivalently

³In general, the equation should be $k\tilde{x} + \phi = 2\pi n$ for some $n \in \mathbb{Z}$. You can check that $n = 0$ gives the crest closest to $x = 0$. In any case, ϕ is defined *modulo* 2π , so it does not really matter which crest you choose to measure it.

the troughs or the nodes or any other point on the wave profile – is travelling. A moment of thought gives the formula

$$c = \lambda\nu \equiv \frac{\lambda}{P} \equiv \frac{\nu}{\hat{k}} \equiv \frac{\omega}{k}. \quad (2.10)$$

Terminology recap

- **Amplitude** a : height of crests (measured from the x -axis);
- **Frequency** ν : # of crests in a unit time interval (can be non-integer!);
- **Wave number** \hat{k} : # of crests in a unit space interval (can be non-integer!);
- **Period** $P = 1/\nu$: amount of time it takes for a crest to superimpose with the original position of a neighbouring crest;
- **Wave length** $\lambda = 1/\hat{k}$: distance between consecutive crests;
- **angular frequency** $\omega = 2\pi\nu$;
- **angular wave number** $k = 2\pi\hat{k}$;
- **wave speed** $c = \lambda\nu$: speed at which any one point of the wave moves (also called *phase speed*).

2.1.2 Complex harmonic waves

Remember Euler's identity $e^{ix} = \cos(x) + i\sin(x)$? We can use it to write our harmonic wave as follows

$$h(x, t) = \operatorname{Re} [ae^{i\phi} e^{i(kx - \omega t)}] = \operatorname{Re} [Ae^{i(kx - \omega t)}], \quad (2.11)$$

where we absorbed the phase factor $e^{i\phi}$ into a *complex amplitude*⁴ $A = ae^{i\phi}$. Now, let us introduce a new, complex function $H(x, t)$ by simply dropping the real part from the above expression of $h(x, t)$

$$H(x, t) = Ae^{i(kx - \omega t)}, \quad (2.12)$$

with $A \in \mathbb{C}$ and $k, \omega \in \mathbb{R}$. Functions of this form are called *complex harmonic waves*. As it is often the case, working in the realm of complex numbers turn out to simplify things. We can then go back to real numbers when needed by taking real and imaginary parts of our expressions.

Amongst the class of complex harmonic waves (2.12), we find solutions to the wave equation. In fact

$$\partial_t^2 H(x, t) = c^2 \partial_x^2 H(x, t) \implies -A\omega^2 = -Ac^2 k^2. \quad (2.13)$$

Therefore, all complex harmonic waves with

$$\omega = \omega(k) = \pm ck \implies \nu = \nu(\hat{k}) = \pm \hat{c}\hat{k}, \quad (2.14)$$

⁴Note that the number of free parameters did not change, now we have 1 complex quantity A and 2 real ones k, ω , which amount to 4 real parameters.

are solutions to the 1-dimensional wave equation. Equations that tie the wave number and the frequency together such as (2.14) are known as *dispersion relations*.

Insight: dispersion relation

There is a reason for this name, although it is not completely evident in the case of the wave equation considered here. As just said, a dispersion relation relates the frequency and wave number of a wave $\omega = \omega(k)$, where the function $\omega(k)$ depends on the specific form of the PDE under consideration. This means that, on a medium that support the propagation of waves according to this specific PDE, any wave will move with a speed

$$c = c(k) = \frac{\omega(k)}{k}.$$

While, in the case of equation (2.13), this velocity is a constant, in general it will depend on k (equivalently on ν).

Now, take an initial wave packet, that is a superposition of infinite harmonic waves that is localised in space. All of these waves will have different frequencies ν and, consequently, will move at different speeds! So the initially localised wave packet will lose cohesion as time passes and slowly spread out: it will **disperse**. Figure 2.2 depicts an example.

2.1.3 Solving PDEs with harmonic waves

Harmonic waves – complex or real – are not only useful when dealing with the wave equation (2.13). In fact they are incredibly useful and have a wide range of applications, due to the following fact

Any linear homogeneous PDE with constant coefficients admit solutions in the form of complex harmonic waves.

And this is true in any number of dimensions, not just for one space and one time variables! Let us work out an example in detail.

The heat equation

Consider the following PDE

$$\partial_t u(x, t) = \alpha \partial_x^2 u(x, t), \quad (2.15)$$

with $\alpha > 0$ a coefficient called *thermal diffusivity*. This is called the *heat equation* and was first proposed by the french mathematician Joseph Fourier in 1822 for the purpose of modelling how heat diffuses through a given body. This equation appear in many disparate branches of mathematics and physics, such as the study of random walks and Brownian motion, quantum mechanics, image analysis and hydrodynamics, just to name a few.

Substituting a complex harmonic wave in place of the function $u(x, t)$, we obtain the relation

$$\omega = -i\alpha k^2. \quad (2.16)$$

You should be slightly confused. We said ω and k were real parameters, but this relation requires at least one of them to be complex! Well, this is kind of a trick. We are silently expanding the class of complex harmonic functions to include those with $\omega \in \mathbb{C}$. Let us

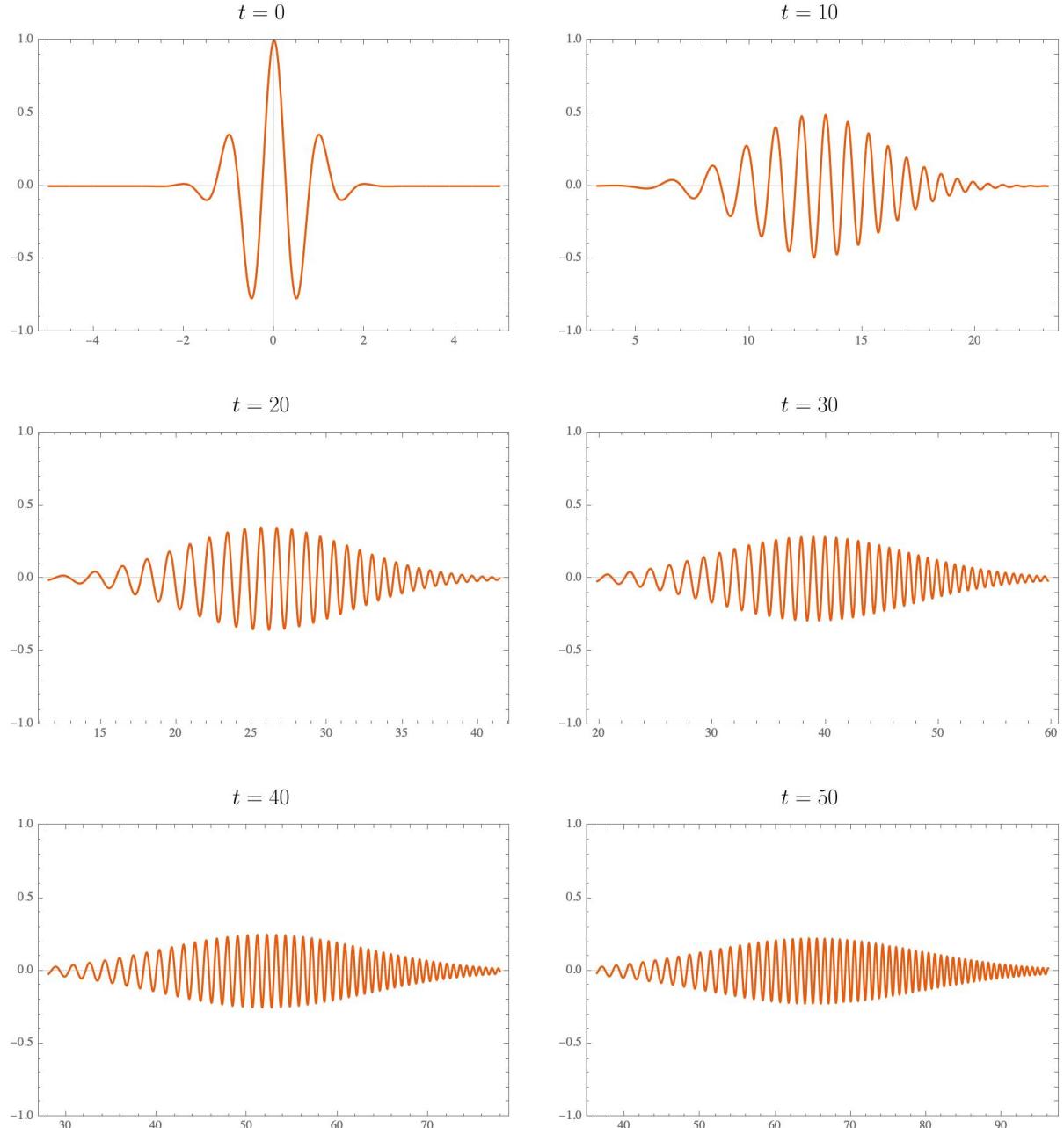


Figure 2.2: An initially localised wave packet $e^{-x^2} \cos(6x)$ propagating on a dispersive medium, with dispersion relation $\omega(k) = ck(1+k)$. You can see the packet losing cohesion as time passes and spreading over the whole real axis.

sweep this fact under the rug for the moment and proceed. Our harmonic wave solution to the heat equation thus looks as follows

$$u_k(x, t) = A_k e^{ikx - \alpha k^2 t}. \quad (2.17)$$

Now, let us impose some boundary conditions. We will place two Dirichlet-like conditions at the ends:

$$u(0, t) = u_0, \quad u(\pi, t) = u_\pi < u_0. \quad (2.18)$$

In Fourier's interpretation of the equation $u(x, t)$ is the time-dependent, point-wise tem-

perature of a rod (that we suppose infinitely thin) of length π and thermal diffusivity⁵ α . The boundary conditions above mean that we placed the rod in contact with two heat reservoirs of temperature u_0 at $x = 0$ and u_π at $x = \pi$. Asking for u_π to be smaller than u_0 means the temperature will “flow” from $x = 0$ to $x = \pi$. Clearly, (2.17) cannot satisfy the boundary conditions by itself and the superposition principle does not help in this case. But, we can use a trick: any function such as $a + bx$ turns to zero identically when acted upon by both ∂_t and ∂_x^2 . This means that we can consider a solution of the form

$$u(x, t) = u_0 - \frac{u_0 - u_\pi}{\pi}x + U(x, t), \quad (2.19)$$

where now $U(x, t)$ has to satisfy standard Dirichlet conditions

$$U(0, t) = U(\pi, t) = 0. \quad (2.20)$$

We know the solution: the x appears in a sine function (achieved with $A_{-k} = -A_k$) and $k \in \mathbb{Z}$. We can now write down our general solution in the form of an infinite series:

$$u(x, t) = u_0 - \frac{u_0 - u_\pi}{\pi}x + \sum_{k=1}^{\infty} a_k \sin(kx) e^{-\alpha k^2 t}, \quad (2.21)$$

where a_k are now real coefficients (they are related to A_k by multiplication by 2i).

The last step are initial conditions. The heat equation is of first order in time, so we only need to impose an initial profile for the function $u(x, 0)$. We will choose a simple setup

$$u(x, 0) = u_\pi. \quad (2.22)$$

In other words our rod was in thermal equilibrium with the reservoir at $x = \pi$ and, at time $t = 0$, we placed it in contact with the reservoir at $x = 0$. The initial condition reads

$$u_0 - \frac{u_0 - u_\pi}{\pi}x + \sum_{k=1}^{\infty} a_k \sin(kx) = u_\pi. \quad (2.23)$$

Remembering the relation⁶

$$\frac{2}{\pi} \int_0^\pi dx \sin(kx) \sin(lx) = \delta_{k,l}, \quad (2.24)$$

and the trivial integrals

$$\frac{2}{\pi} \int_0^\pi dx \sin(lx) = \begin{cases} 4/(\pi l), & l \in 2\mathbb{Z} + 1 \\ 0, & l \in 2\mathbb{Z} \end{cases}, \quad \frac{2}{\pi} \int_0^\pi dx x \sin(lx) = \frac{2}{l}(-1)^{l+1}, \quad (2.25)$$

we integrate against $2/\pi \sin(lx)$ the two sides of (2.23). We find a remarkably simple expression for the coefficients a_l

$$a_l = -2 \frac{u_0 - u_\pi}{\pi l}. \quad (2.26)$$

⁵For the interested reader, the thermal diffusivity for a 1-dimensional rod is $\alpha = k/(c\rho)$ where k is the *thermal conductivity* of the material the rod is made of, c its *specific heat capacity* and ρ its density.

⁶Fun terminology fact, this is called an *orthogonality relation*, since you can think of $\sin(kx)$ as a “vector” (with index k) and $2/\pi \int_0^\pi dx f(x)g(x)$ as a “scalar product”.

We are almost there, promise. We need to resum the series over k . You can check that

$$\sum_{k=1}^{\infty} \frac{\sin(kx)}{k} = \frac{\pi - x}{2}, \quad \forall x \in [0, \pi], \quad (2.27)$$

so that (2.23) is satisfied. Sadly enough, the sum for general t is not known in closed form, so we must settle for

$$u(x, t) = u_0 - \frac{u_0 - u_{\pi}}{\pi} x - 2 \frac{u_0 - u_{\pi}}{\pi} \sum_{k=1}^{\infty} \frac{\sin(kx)}{k} e^{-\alpha k^2 t}. \quad (2.28)$$

We can still extract some information though. For example, we see easily that for $t \rightarrow \infty$, the solution will relax onto a linear function

$$\lim_{t \rightarrow \infty} u(x, t) = u_0 - \frac{u_0 - u_{\pi}}{\pi} x, \quad (2.29)$$

interpolating from the temperatures u_0 and u_{π} . On the other hand, at $t = 0$, the solution has a discontinuity at $x = 0$, due to the Dirichlet conditions

$$u(x, 0) = \begin{cases} u_0, & x = 0 \\ u_{\pi}, & 0 < x \leq \pi \end{cases}, \quad (2.30)$$

taking an ‘‘L shape’’. The curves $u(x, t)$ will smoothly deform with t from $u(x, 0)$ to $u(x, \infty)$. In Figure 2.3 is plotted the evolution of this solution.

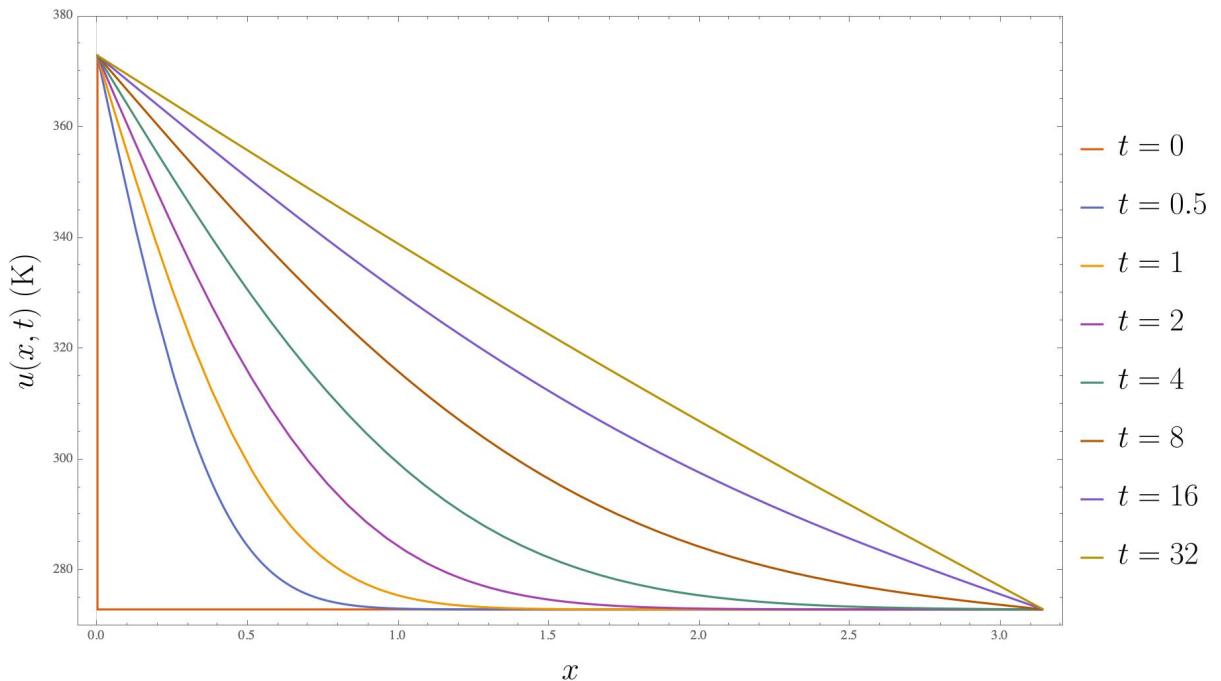


Figure 2.3: Plot of the solution (2.28) for $u_0 \simeq 373.15\text{K}$ (the boiling water point) and $u_{\pi} \simeq 273.15\text{K}$ (the freezing water point).

2.2 Energy

Let us stop our marathon of mathematical formulae and derivation to think a bit more deeply about what a wave is. Consider a travelling wave, say, just for definiteness, a d'Alembert one. Try answering the question: *what is it that is travelling along the string?* Clearly it is not matter: we actually imposed that the string bits experience no longitudinal displacement. Of course, this was an approximation and, in the real world, the string bits do experience a longitudinal displacement along the transverse one. However this displacement can be at most an oscillation about their rest point since the string overall position is supposed to be fixed.

We are touching on an important point here: the string and the wave are separate entities. The former is a medium, with fixed overall position, while the latter is, well, *something* that propagates along the medium. What is this something? The first guess is that what travels on the string is a *disturbance* of the string's position. This appears to be quite reasonable and, in fact, it is often taken as a definition:

a wave is a disturbance that propagates through a medium.

However, there is a subtlety in this statement that is often overlooked and is important to keep sharp. In the case of a transverse wave on a string, the disturbance is of mechanical type: it involves the position and velocity of the medium. We can see the position of the string, it is a quantity that we can measure. For this reason, the *measurable effect* of the disturbance conflates with the disturbance itself. This is not the case for other types of waves, such as the electromagnetic ones, where the medium is the electromagnetic field that pervades our world. This field is intangible, we cannot measure it directly: we need a *probe*, for example in the form of a moving body of known electric charge. In other words, we need to perform an experiment (i.e. a *measurement*). The electromagnetic waves exist even if no probe is there for us to see its effects. In the case of the string, the probe and the propagation medium are one and the same.

So, our waves are disturbances propagating through a medium whose effects can be measured with an appropriate probe. Since the passing of a wave excites the motion of this probe in some measurable way, we see that one important quantity that waves carry is *energy*. In this section, we will introduce the expressions of energy and *energy density* of waves on strings and calculate a few examples. Finally, we will discuss a fundamental property, not just of waves but of Nature in general: the *conservation of energy*.

2.2.1 Energy density

The total energy of a string is the sum of the *kinetic energy* K and the *potential energy* V . Let us look at these two quantities one at a time.

Consider an infinitesimal bit of the string, say between x and $x + \delta x$. Its kinetic energy is

$$\delta K(x, t) = \frac{1}{2}mv^2 = \frac{1}{2}\rho\delta x [\partial_t y(x, t)]^2. \quad (2.31)$$

To obtain the total kinetic energy of the string, we take the limit $\delta x \rightarrow 0$ while summing over all the string bits. This becomes an integral

$$K(t) = \int_0^L dx \mathcal{K}(x, t) \quad \mathcal{K}(x, t) \equiv \frac{\rho}{2} [\partial_t y(x, t)]^2, \quad (2.32)$$

where L is the total length of the string and we introduced the *kinetic energy density* $\mathcal{K}(x, t)$.

To derive the potential energy, we again start with an infinitesimal segment of the string. When not in its rest configuration, this segment will be stretched from a length of δx to a longer length δs . If T is the tension, the work that has to be performed to stretch the string of an amount $\delta s - \delta x$ is simply $T\delta s - T\delta x$. Since we are neglecting effects coming from other forces, such as gravity, this gives the string bit's potential energy. We need an expression for δs which is the *arc-length* of the curve described by $y(x, t)$ between x and $x + \delta x$. We know how this quantity is computed

$$\delta s(x, t) = \sqrt{1 + (\partial_x y(x, t))^2} \delta x \approx \delta x \left(1 + \frac{1}{2} (\partial_x y(x, t))^2 + \mathcal{O}((\partial_x y)^4) \right), \quad (2.33)$$

where we approximated the expression to the first two terms of its Taylor expansion, since we are assuming that the slope $\partial_x y(x, t)$ is a small quantity, as we did when deriving the wave equation. The potential energy of an infinitesimal bit of the string thus reads⁷

$$\delta V(x, t) = T\delta s(x, t) - T\delta x = \frac{T}{2} (\partial_x y(x, t))^2 \delta x. \quad (2.34)$$

Again, the total potential energy is given by summing all the infinitesimal portions' contributions

$$V(t) = \int_0^L dx \mathcal{V}(x, t), \quad \mathcal{V}(x, t) = \frac{T}{2} (\partial_x y(x, t))^2. \quad (2.35)$$

The quantity $\mathcal{V}(x, t)$ is the *potential energy density*.

Finally, the total energy is given by

$$E = K(t) + V(t) = \int_0^L dx \mathcal{E}(x, t), \quad \mathcal{E}(x, t) = \frac{1}{2} [\rho(\partial_t y(x, t))^2 + T(\partial_x y(x, t))^2], \quad (2.36)$$

where $\mathcal{E}(x, t)$ is the *energy density*. Notice that the missing dependence on t of E in (2.36) is not a typo. As we will see below, the energy truly is independent of time.

2.2.2 Energy density of example waves

Let us work out explicitly a few examples.

Example 2.2.1: Travelling wave First, let us consider a right-moving wave $y(x, t) = f(x - ct)$ travelling with wave speed c . We can find its energy density fairly easily by substitution into (2.36)

$$\mathcal{E}(x, t) = \frac{1}{2} [\rho c^2 (f'(x - ct))^2 + T(f'(x - ct))^2]. \quad (2.37)$$

But now remember that the wave speed c is related to the string's physical parameters ρ and T as $c = \sqrt{T/\rho}$, thus the kinetic and potential energies turn out to be equal! This phenomenon is known as *equipartition of energy*. We have

$$\mathcal{E}(x, t) = T(f'(x - ct))^2. \quad (2.38)$$

⁷Here and below we drop the correction terms $\mathcal{O}((\partial_x y)^4)$.

So the energy density travels along the string with the wave profile. This is what we meant when before we said that a wave transports energy.

Example 2.2.2: Standing wave Now consider a standing wave solution, say a Dirichlet-Dirichlet one

$$y(x, t) = \sin(kx) \left(F \cos(kct) + G \sin(kct) \right) = A \sin(kx) \cos(kct + \phi), \quad (2.39)$$

where we used phasor addition to combine the t -dependent function into a single cosine. The following relations are true

$$A^2 = F^2 + G^2, \quad \phi = -\arccos \left(\frac{F}{\sqrt{F^2 + G^2}} \right). \quad (2.40)$$

Let us compute the first derivatives

$$\partial_x y(x, t) = Ak \cos(kx) \cos(kct + \phi), \quad (2.41a)$$

$$\partial_t y(x, t) = -Ak c \sin(kx) \sin(kct + \phi). \quad (2.41b)$$

$$(2.41c)$$

Here too, the prefactors of \mathcal{K} and \mathcal{V} are the same, since $c^2\rho = T$. However the two densities are different this time

$$\mathcal{K}(x, t) = \frac{A^2 k^2 T}{2} \sin^2(kx) \sin^2(kct + \phi), \quad (2.42a)$$

$$\mathcal{V}(x, t) = \frac{A^2 k^2 T}{2} \cos^2(kx) \cos^2(kct + \phi). \quad (2.42b)$$

You can see how these densities depend on x and t in Figure 2.4. The kinetic energy density is always vanishing at the nodes, since, by definition, they are not moving and it is maximal at the crests and troughs, intuitively the points that move faster at any given time. It also vanishes for all x when the profile $y(x, t)$ reaches its maximal amplitude. On the contrary, the potential energy density vanishes at the crests and at the troughs and is maximal at the nodes. This is perhaps counterintuitive, but if you think a moment about it, they are the points doing most of the work as they pull on the rest of the string while standing still.

In this example we can integrate along the string and compute the total energy. The integrals are simple to execute (remember that k is an integer)

$$K(t) = \frac{A^2 k^2 T}{2} \sin^2(kct + \phi) \int_0^\pi dx \sin^2(kx) = \frac{A^2 k^2 T \pi}{4} \sin^2(kct + \phi), \quad (2.43a)$$

$$V(t) = \frac{A^2 k^2 T}{2} \cos^2(kct + \phi) \int_0^\infty dx \cos^2(kx) = \frac{A^2 k^2 T \pi}{4} \cos^2(kct + \phi). \quad (2.43b)$$

When summing these two terms we notice that the dependence on t cancels identically!

$$E = K(t) + V(t) = \frac{A^2 k^2 T \pi}{4}. \quad (2.44)$$

This confirms our earlier statement that the total energy is independent of time. We will see a general proof momentarily.

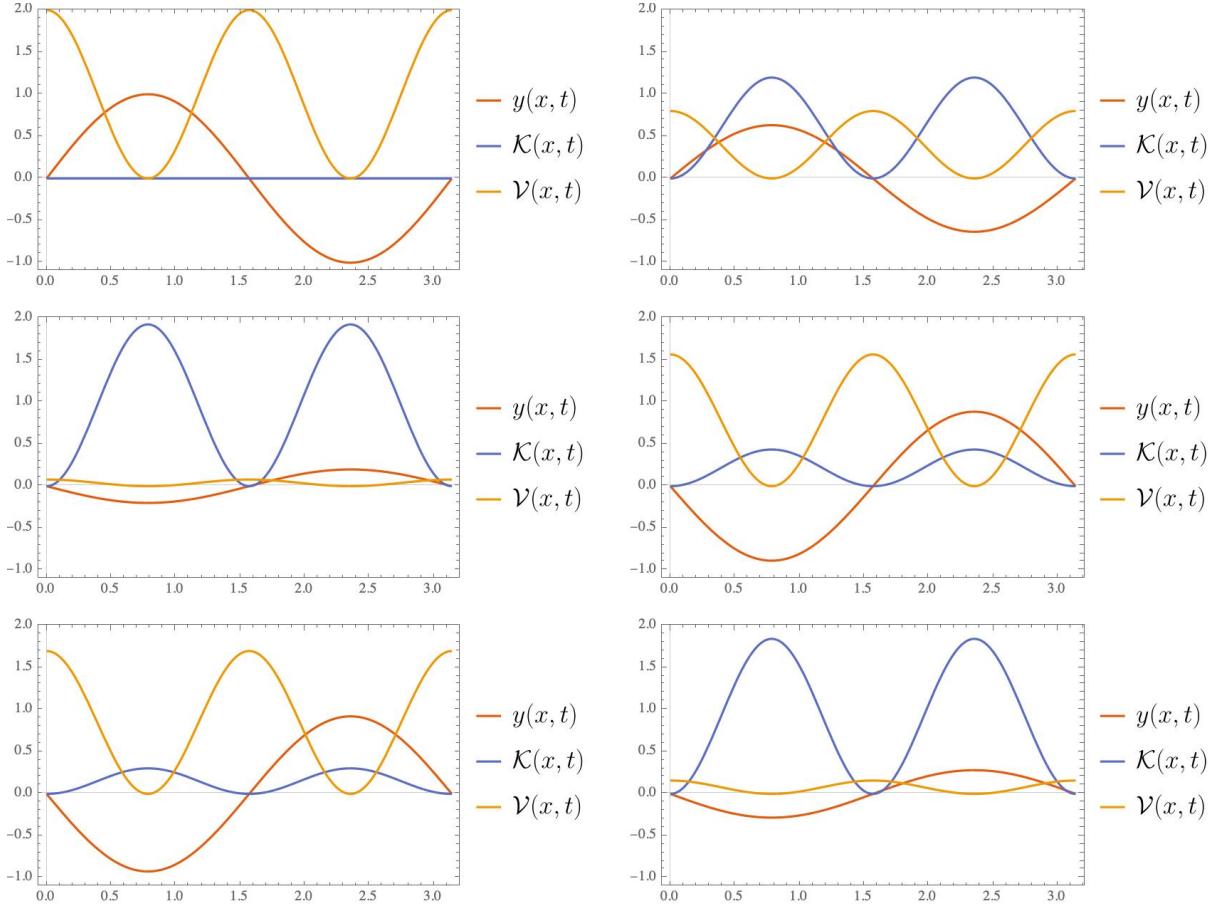


Figure 2.4: Time snapshots of (2.39), along with its kinetic and potential energy densities.

Example 2.2.3: Bi-chromatic wave As a slight increase in difficulty let us consider the sum of two standing waves

$$y(x, t) = y_k(x, t) + y_l(x, t) = A_k \sin(kx) \cos(kct + \phi_k) + A_l \sin(lx) \cos(lct + \phi_l). \quad (2.45)$$

Such a wave is sometime called *bi-chromatic* as it contains two fundamental frequencies $\omega_k = kc$ and $\omega_l = lc$. We suppose that $k \neq l$ and both need to be integers. Let us calculate the kinetic energy of $y(x, t)$

$$\begin{aligned} K(t) &= \frac{\rho}{2} \int_0^\pi dx (\partial_t y_k(x, t) + \partial_t y_l(x, t))^2 = \\ &= K_k(t) + K_l(t) + T k l \cos(kct + \phi_k) \cos(lct + \phi_l) \int_0^\pi dx \sin(kx) \sin(lx) = \\ &= K_k(t) + K_l(t), \end{aligned} \quad (2.46)$$

where $K_k(t)$ and $K_l(t)$ are the individual kinetic energies of $y_k(x, t)$ and $y_l(x, t)$, respectively. You can immediately see that the same result hold for the potential energy as well: it is the sum of the two standing waves' potential energies. Thus we have the following important result

The total energy of a sum of standing waves is equal to the sum of the individual standing waves' energies.

We write this in formulae as follows

$$E \left[\sum_k y_k(x, t) \right] = \sum_k E[y_k(x, t)] , \quad (2.47)$$

where the notation $E[f]$ stands for “the energy of the wave f ”.

Example 2.2.4: Complex harmonic wave Finally, let us look at the energy of a complex harmonic wave (2.12)

$$y(x, t) = A e^{i(kx - \omega t)} . \quad (2.48)$$

Note that both the kinetic and potential energy densities are supposed to be real quantities. However if we square $y(x, t)$ we will obtain a function returning complex numbers. The way out is to substitute the squares with absolute values squared: the energy density of a complex wave is the sum of the energy densities of its real and imaginary parts. Thus we see that

$$\mathcal{E}(x, t) = \frac{\rho}{2} \left| -i\omega A e^{i(kx - \omega t)} \right|^2 + \frac{T}{2} \left| ik A e^{i(kx - \omega t)} \right|^2 = \left(\frac{\rho}{2} \omega^2 + \frac{T}{2} k^2 \right) |A|^2 . \quad (2.49)$$

In the case of complex harmonic waves the energy density itself is a constant.

2.2.3 Conservation equation

We have seen that in some of the specific examples we considered the total energy of the string is independent of time. This is actually a general property – actually valid beyond the specific physical system considered here. Let us prove this fact.

Consider a solution to the wave equation $y(x, t)$ and its energy density

$$\mathcal{E}(x, t) = \frac{\rho}{2} (\partial_t y(x, t))^2 + \frac{T}{2} (\partial_x y(x, t))^2 . \quad (2.50)$$

Its time derivative is computed easily

$$\partial_t \mathcal{E}(x, t) = \rho \partial_t y(x, t) \partial_t^2 y(x, t) + T \partial_x y(x, t) \partial_x \partial_t y(x, t) = \quad (2.51a)$$

$$= T \left[\partial_t y(x, t) \partial_x^2 y(x, t) + \partial_x y(x, t) \partial_x \partial_t y(x, t) \right] , \quad (2.51b)$$

where we used the wave equation in the second equality. We can notice that the term in the square brackets is a derivative with respect to x . So, defining the *energy flux* \mathcal{F} as

$$\mathcal{F}(x, t) = -T \partial_x y(x, t) \partial_t y(x, t) , \quad (2.52)$$

we can write

$$\partial_t \mathcal{E}(x, t) + \partial_x \mathcal{F}(x, t) = 0 . \quad (2.53)$$

This is known as *conservation equation*, since it implies the conservation of the energy. In fact, let us integrate it along the string (we take the endpoints to be at positions x_1 and x_2 which might be infinite)

$$\frac{dE}{dt} \equiv \int_{x_1}^{x_2} dx \partial_t \mathcal{E}(x, t) = - \int_{x_1}^{x_2} dx \partial_x \mathcal{F}(x, t) = \mathcal{F}(x_1, t) - \mathcal{F}(x_2, t) . \quad (2.54)$$

But this quantity is not zero! In fact it depends on the boundary conditions. However this should not be excessively surprising. What the equation above is saying is that the energy changes in time by the same amount that flows in from the endpoints of the string! This is entirely reasonable: if somehow the “world outside” of the string injects (or absorbs) energy in it, the string energy will reflect that change. This is because the total energy of the string plus the “world outside” is a constant in time. Still, in most cases the string is studied in isolation from its environment: it is assumed to be a *closed system* that does not exchange energy with the environment. In fact the cases we studied are all of this type. Let us look at them.

- **Dirichlet-Dirichlet.** In this case, $y(x_1, t) = y(x_2, t) = 0$ for any time t . Clearly the time derivatives of y at the endpoints will vanish as well, which means that $\mathcal{F}(x_1, t) = \mathcal{F}(x_2, t) = 0$. ✓
- **Neumann-Neumann.** Now we have, $\partial_x y(x, t)|_{x=x_1} = \partial_x y(x, t)|_{x=x_2} = 0$ for any time t . This immediately implies that $\mathcal{F}(x_1, t) = \mathcal{F}(x_2, t) = 0$. ✓
- **Dirichlet-Neumann.** This case follows immediately from the previous two. ✓
- **Infinite and semi-infinite string.** In case one or both of the endpoints is at infinite distance, one should be slightly more careful. For a physicist this case follows from the above. In fact, in reality there exist no such thing as an infinite string. It is an idealisation that makes certain computations more convenient. An infinite string is really a string so long that one can study its dynamics without worrying about boundaries. However the boundaries are still there, so the conservation of energy is covered by the above considerations. ✓

Outro: takeaway points

In this chapter we have looked more closely into the concept of wave, introducing harmonic waves as general building blocks that can be used to solve Partial Differential Equations. We also studied their energy and saw it satisfies a conservation equation. The takeaway points are

- We defined the class of harmonic (or sinusoidal, or monochromatic) waves

$$h(x, t) = a \cos\left(2\pi(\hat{k}x - \nu t) + \phi\right),$$

depending of 4 real constants. These encode physical properties of the wave

- **Amplitude a :** height of crests (measured from the x -axis);
- **Frequency ν :** # of crests in a unit time interval (can be non-integer!);
- **Wave number \hat{k} :** # of crests in a unit space interval (can be non-integer!);
- **Period $P = 1/\nu$:** amount of time it takes for a crest to superimpose with the original position of a neighbouring crest;
- **Wave length $\lambda = 1/\hat{k}$:** distance between consecutive crests;
- **angular frequency $\omega = 2\pi\nu$;**
- **angular wave number $k = 2\pi\hat{k}$;**

- **wave speed** $c = \lambda\nu$: speed at which any one point of the wave moves (also called *phase speed*).
- We introduced a variant called complex harmonic waves

$$H(x, t) = Ae^{2\pi i(\hat{k}x - \nu t)} ,$$

depending on two real parameters \hat{k}, ν and a complex amplitude A .

- We mentioned the concept of dispersion relation $\nu = \nu(\hat{k})$, an equation that connects the wave number \hat{k} to the frequency ν . In the case of an idealised string, the dispersion relation is trivial

$$\nu(\hat{k}) = \pm c\hat{k} ,$$

but in general it can be more complicated. For non-linear dispersion relations, each wave number will have an associated speed of propagation (the phase speed)

$$c(\hat{k}) = \frac{\nu(\hat{k})}{\hat{k}} .$$

This means that a non-monochromatic wave (a superposition of various harmonic waves with different frequencies) will lose cohesion in time and disperse.

- We learned how to solve PDEs with constant coefficients using harmonic waves. In particular, we used the superposition principle to write the general solution of the heat equation as a sum of harmonic waves

$$u(x, t) = u_0 - \frac{u_0 - u_\pi}{\pi}x + \sum_{k=1}^{\infty} a_k \sin(kx)e^{-\alpha k^2 t} ,$$

with an additional constant u_0 and linear $(u_0 - u_\pi)x/\pi$ terms. Then we used the orthogonality relation for sinusoidal functions to find a specific solution that satisfies the chosen initial conditions.

- We learned that a wave is a disturbance that propagates through a medium and to separate the medium (e.g. the string) from the wave itself (the propagation of the disturbance).
- We saw how to write the total energy of the oscillating string as the integral of an energy density

$$E = K(t) + V(t) = \int_0^L dx \mathcal{E}(x, t) ,$$

where L is the length of the string, $K(t)$ and $V(t)$, respectively, the total kinetic and potential energies and

$$\mathcal{E}(x, t) = \frac{\rho}{2} (\partial_t y(x, t))^2 + \frac{T}{2} (\partial_x y(x, t))^2 ,$$

is the energy density.

- We computed the energy and its density for example waves. We saw that a travelling wave satisfies the equipartition of energy, for which the kinetic and potential energies are equal. We also saw that the total energy of a standing wave does not depend on time and for a wave which is the sum of standing waves, the total energy is also the sum of the components' energies.

- We saw that the energy density satisfies a conservation equation

$$\partial_t \mathcal{E}(x, t) + \partial_x \mathcal{F}(x, t) = 0 ,$$

where

$$\mathcal{F}(x, t) = -T \partial_x y(x, t) \partial_t y(x, t)$$

is the energy flux. This equation implies the conservation of energy by virtue of (the 1-dimensional version of) the divergence theorem

$$\frac{dE}{dt} = \mathcal{F}(x_2, t) - \mathcal{F}(x_1, t) ,$$

with x_1 and x_2 being the endpoint of the string. We verified that the total energy is not time-dependent for various boundary conditions on the string.

Chapter 3

Out of Flatland: bodies vibrating in 3-dimensions

Until now, we have only considered a 2-dimensional world. In this simplified setting, we studied at length the (admittedly approximate) physical laws that govern a stretched string and the waves that propagate along it. The time has come for this question: what if we allowed the string to vibrate in two transverse directions? Well, it is not that hard to see that, in the same approximations made in §1.1, the motion along the y and the z directions will decouple and we end up simply with two copies of the 1-dimensional wave equation

$$\partial_t^2 y(x, t) = c^2 \partial_x^2 y(x, t), \quad \partial_t^2 z(x, t) = c^2 \partial_x^2 z(x, t). \quad (3.1)$$

Though there are some interesting phenomena that arise when the displacement is not confined on a single direction – prime amongst them the *polarisation* – we want to do something different. Instead of adding a dimension to the oscillations, let us add it to the string itself and consider as the object of our study *membrane*, stretched in both x and y directions and allowed to vibrate in the z direction. The physical object you should keep in mind is a drum being struck¹.

3.1 Waves on a plane

Consider an infinite two-dimensional membrane of homogeneous density² ρ . In the equilibrium state, it is flat and we take coordinates so it coincides with the (x, y) plane in \mathbb{R}^3 . We assume that this membrane is stretched to a tension T^3 . To understand better what this means, consider a line on the surface. Each segment of the line will experience tension forces along the line itself, just as in the 1-dimensional case. However there will also be tension forces acting in the direction perpendicular to the line. In fact, the part of the membrane on one side of the line “pulls” on the remaining part of the membrane on the other side of the line. The combination of these tensions will produce the total force.

¹There are a number of nice slow-motion videos of drums being struck, such as <https://www.youtube.com/watch?v=STSWLX23xqc>. A cymbal being struck in slow motion is also quite impressive to watch <https://www.youtube.com/watch?v=kpoan0lb3-w>.

²In this case this is a mass per unit area, so it is measured in kg/m².

³This, on the other hand, is still a force per unit length and measured in N/m = kg/s².

3.1.1 The 2-dimensional wave equation

We make the same assumptions as in § 1.1 : that the membrane points only move in the transverse z direction, that the tension remain constant throughout and is the only force acting on the membrane, and that the latter is a continuum medium. Then, we can describe the membrane as a time-dependent surface $z = z(x, y, t)$ in \mathbb{R}^3 . The function $z(x, y, t)$ is the vertical displacement of the membrane from the equilibrium plane $z = 0$ at point (x, y) and time t . One final assumption is that the angle between the horizontal plane $z = 0$ and the plane tangent to $z(x, y, t)$ is small. In other words, we ask

$$\partial_x z(z, y, t) \ll 1 , \quad \partial_y z(x, y, t) \ll 1 , \quad \forall x, y, t . \quad (3.2)$$

This assumption implies that T can be taken constant in modulus across the whole membrane.

Let us consider an infinitesimal element of the membrane, of sides δx and δy , see figure 3.1. We can follow the same reasoning as in §1.1 . In fact, you can see that for any line

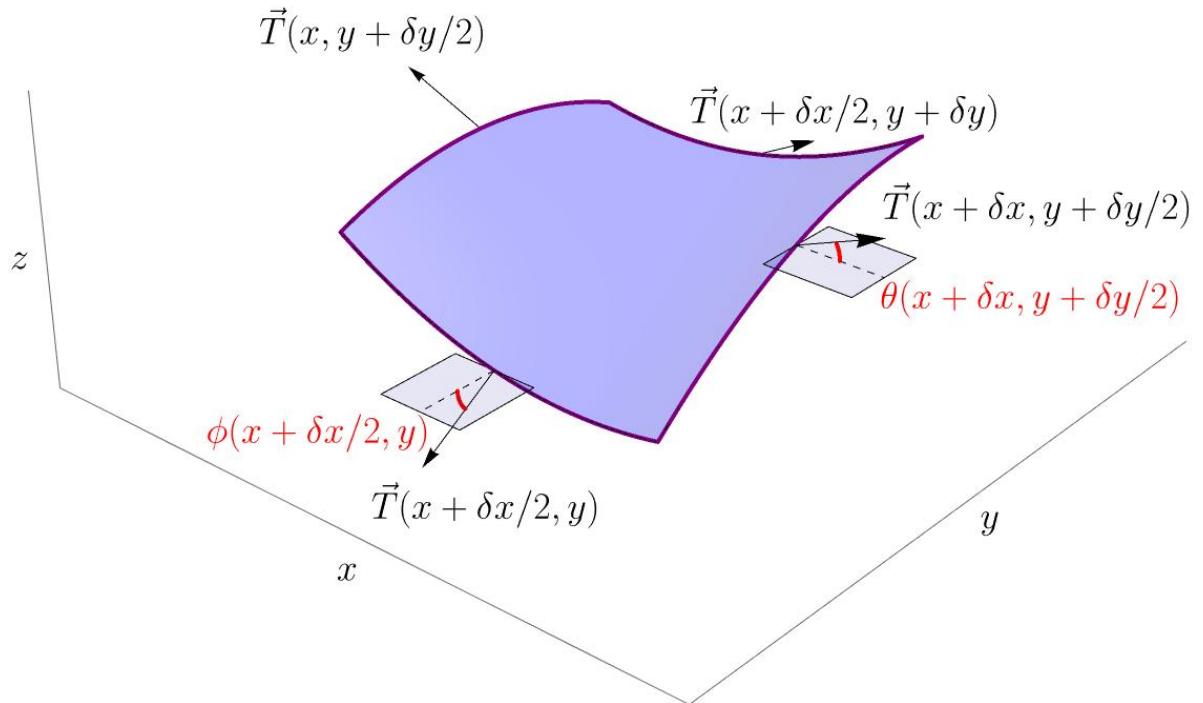


Figure 3.1: Infinitesimal portion of the membrane, of sides δx and δy and the tension forces acting on it. As we send $\delta x \rightarrow 0$ and $\delta y \rightarrow 0$, the horizontal components of the various forces compensate, while the vertical components sum to a non-vanishing quantity.

on the surface, say in the x direction, the tension forces along the line behave exactly as in the 1-dimensional case. Additionally, we have the transverse forces, in the y direction to take care of. But these also behave in the same way as in 1 dimension, only in the y direction. They simply sum to the other ones. As $\delta x \rightarrow 0$ and $\delta y \rightarrow 0$, the horizontal

components of the force compensate, while the vertical one gives

$$F_z(x, y) = T\delta x \delta y \left[\sin \theta(x + \delta x, y + \delta y/2) - \sin \theta(x, y + \delta y/2) + \sin \phi(x + \delta x/2, y + \delta y) - \sin \phi(x + \delta x/2, y) \right]. \quad (3.3)$$

Here $\theta(x, y)$ is the angle made by the vector $\vec{T}(x, y)$ along the x direction and the horizontal plane. Similarly for $\phi(x, y)$, only with vectors along the y direction. Mathematically

$$\tan(\theta(x, y, t)) = \partial_x z(x, y, t), \quad \tan(\phi(x, y, t)) = \partial_y z(x, y, t). \quad (3.4)$$

Now we use the small angle approximations and, just in the same way as in §1.1 we arrive at the expression

$$F_z(x, y, t) = T\delta x \delta y \left[\partial_x^2 z(x, y, t) + \partial_y^2 z(x, y, t) \right]. \quad (3.5)$$

Finally we apply Newton's 2nd law $F_z(x, y, t) = m\partial_t^2 z(x, y, t) = \rho\delta x \delta y \partial_t^2 z(x, y, t)$, which yields the 2-dimensional wave equation

$$\boxed{\partial_t^2 z(x, y, t) - c^2 \nabla^2 z(x, y, t) = 0}. \quad (3.6)$$

Here we introduced the wave velocity

$$c = \sqrt{\frac{T}{\rho}}, \quad (3.7)$$

and the symbol⁴ ∇^2 of the *Laplace operator*

$$\nabla^2 = \partial_x^2 + \partial_y^2. \quad (3.8)$$

Note that the Laplace operator is sometimes written as Δ ! You should pay attention to this.

3.1.2 Energy of a membrane

Before studying the solutions of the wave equation (3.6), let us make some general considerations on the energy of these. The derivation of the kinetic and potential energy densities parallels what we saw for the 1-dimensional case. We will not repeat ourselves and simply state the result

$$\mathcal{E}(x, y, t) = \underbrace{\frac{\rho}{2} \left(\partial_t z(x, y, t) \right)^2}_{\mathcal{K}(x, y, t)} + \underbrace{\frac{T}{2} \left[\left(\partial_x z(x, y, t) \right)^2 + \left(\partial_y z(x, y, t) \right)^2 \right]}_{\mathcal{V}(x, y, t)}. \quad (3.9)$$

Before moving on, let us introduce a piece of notation that will simplify our formulae and make generalisations straightforward. We denote by $\vec{\nabla} f(x, y)$ the *gradient* vector of the 2-variable function $f(x, y)$

$$\vec{\nabla} f(x, y) = \begin{pmatrix} \partial_x f(x, y) \\ \partial_y f(x, y) \end{pmatrix}. \quad (3.10)$$

⁴The symbol ∇ is called *nabla*.

With this, we can write

$$\mathcal{E}(x, y, t) = \frac{\rho}{2} \left(\partial_t z(x, y, t) \right)^2 + \frac{T}{2} \left| \vec{\nabla} z(x, y, t) \right|^2. \quad (3.11)$$

Does this quantity satisfy a conservation equation? The answer is quite easy to give. Let us compute the time derivative of the above expression

$$\partial_t \mathcal{E}(x, y, t) = \rho \partial_t z(x, y, t) \partial_t^2 z(x, y, t) + T \vec{\nabla} z(x, y, t) \cdot \vec{\nabla} \partial_t z(x, y, t), \quad (3.12)$$

where the last term is a scalar product. Using the wave equation and the identity $\rho c^2 = T$, we find

$$\begin{aligned} \partial_t \mathcal{E}(x, y, t) &= T \partial_t z(x, y, t) \nabla^2 z(x, y, t) + T \vec{\nabla} z(x, y, t) \cdot \vec{\nabla} \partial_t z(x, y, t) = \\ &= -\vec{\nabla} \cdot \left(-T \partial_t z(x, y, t) \vec{\nabla} z(x, y, t) \right). \end{aligned} \quad (3.13)$$

Thus, we define the energy flux density as the following vector

$$\vec{\mathcal{F}}(x, y, t) \equiv -T \partial_t z(x, y, t) \vec{\nabla} z(x, y, t) = -T \partial_t z(x, y, t) \begin{pmatrix} \partial_x z(x, y, t) \\ \partial_y z(x, y, t) \end{pmatrix}. \quad (3.14)$$

The conservation equation, then, relates the time derivative of the energy density to the *divergence* of the energy flux density vector

$$\partial_t \mathcal{E}(x, y, t) + \vec{\nabla} \cdot \vec{\mathcal{F}}(x, y, t) = 0. \quad (3.15)$$

In 1 dimension, the conservation equation led to the conservation of the energy. The same applies here. In fact the energy of a membrane occupying a region R of the (x, y) plane (at rest) is given by

$$E = \iint_R dA \mathcal{E}(x, y, t), \quad (3.16)$$

with $dA = dx dy$ being the area element. Its rate of change through time is then

$$\frac{dE}{dt} = \iint_R dA \partial_t \mathcal{E}(x, y, t) = - \iint_R dA \vec{\nabla} \cdot \vec{\mathcal{F}}(x, y, t) = - \oint_{\partial R} ds \hat{n} \cdot \vec{\mathcal{F}}(x, y, t). \quad (3.17)$$

In the last equality we used the *divergence theorem*⁵. There, \hat{n} the outward pointing normal to the boundary ∂R of the region R and ds the line element on it. Here, just as in 1-dimension, the change of energy in time is exactly identical to the energy that flows into the system through its boundary. If the membrane is supposed to be a closed system, then we expect the right-most side of (3.17) to vanish identically.

3.1.3 Plane waves

Let us mimic the d'Alembert solutions of §1.2. We call $z(x, y, t)$ a *2-dimensional plane wave* if it varies only in a single direction on the plane, say the one determined by the unit vector $\hat{n} = (n_x \ n_y)^T$. In other words,

$$z(x, y, t) = z(\hat{n} \cdot \vec{x}, t) \equiv z(n_x x + n_y y, t), \quad (3.18)$$

⁵This is a two-dimensional version of the divergence (or Gauss) theorem you saw in the Vector & Complex Calculus module earlier this year. See §8 of the [lecture notes](#).

where we introduced the 2-dimensional coordinate vector $\vec{x} = \begin{pmatrix} x & y \end{pmatrix}^T$. Does this class of waves satisfy the equation (3.6)? Let us see:

$$\partial_t^2 z(\xi, t) - c^2 |\hat{n}|^2 \partial_\xi^2 z(\xi, t) \equiv \partial_t^2 z(\xi, t) - c^2 \partial_\xi^2 z(\xi, t) = 0 , \quad (3.19)$$

where $\xi = \hat{n} \cdot \vec{x}$ and we used the fact that \hat{n} is a unit vector. Hence, for plane waves, the 2-dimensional wave equation reduces to the 1-dimensional one, along a specific direction dictated by the propagation direction of the wave. Then we know that travelling waves are solutions

$$z(x, y, t) = f(\hat{n} \cdot \vec{x} - ct) + g(\hat{n} \cdot \vec{x} + ct) . \quad (3.20)$$

These solutions are known as *travelling 2-dimensional plane waves*.

There exists a notion of 2-dimensional *harmonic plane wave* as well. They are the following class of functions

$$h(x, y, t) = e^{i(\vec{k} \cdot \vec{x} - \omega t)} . \quad (3.21)$$

The quantity \vec{k} is called the *wave vector*. A harmonic plane wave solves the 2-dimensional wave equation iff the following dispersion relation is satisfied

$$\omega = \omega(\vec{k}) \equiv c|\vec{k}| . \quad (3.22)$$

3.1.4 Higher dimensions

The reason for introducing vector notation in the formulae above is not only to save space. Everything that we wrote keeps on holding if we change the dimension⁶ of our problem from $d = 2$ to arbitrary $d \in \mathbb{N}$! In fact, consider a function $w : \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}$. The wave equation is simply

$$\partial_t w(\vec{x}, t) = c^2 \nabla^2 w(\vec{x}, t) , \quad \vec{x} \in \mathbb{R}^d , t \in \mathbb{R} . \quad (3.23)$$

We can even be more general and consider *complex solutions* $w : \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{C}$. All the other formulae work as well. For example, the plane wave solutions are

$$w(\vec{x}, t) = f(\hat{n} \cdot \vec{x} - ct) , \quad (3.24)$$

for any choice of $f : \mathbb{R} \rightarrow \mathbb{R}$ or $f : \mathbb{R} \rightarrow \mathbb{C}$.

As another example, the energy density is

$$\mathcal{E}(\vec{x}, t) = \frac{T}{2} \left(\frac{1}{c^2} |\partial_t w(\vec{x}, t)|^2 + |\vec{\nabla} w(\vec{x}, t)|^2 \right) , \quad (3.25)$$

and satisfies the continuity equation

$$\partial_t \mathcal{E}(\vec{x}, t) = -\vec{\nabla} \vec{\mathcal{F}}(\vec{x}, t) , \quad (3.26)$$

with energy flux density given by

$$\vec{\mathcal{F}}(\vec{x}, t) = -T \operatorname{Re} \left[\partial_t w(\vec{x}, t) \vec{\nabla} w(\vec{x}, t) \right] . \quad (3.27)$$

⁶When we say “dimension” we mean the number of independent real variables.

The conservation of energy follows from the d -dimensional divergence theorem. Let R be a region contained in d -dimensional space, with $d-1$ -dimensional boundary ∂R . Then

$$\frac{dE}{dt} = \int_R dV \partial_t \mathcal{E}(\vec{x}, t) = - \int_{\partial R} dS \hat{n} \cdot \vec{\mathcal{F}}(\vec{x}, t) . \quad (3.28)$$

All these expressions admit an even more general and more compact form, in terms of *space-time* vectors. You will be meeting them in your future module on electromagnetism and relativity.

3.2 Rectangular membranes

Now we are going to impose boundary conditions on our solutions. We will follow the same approach as in §1.4.3. First, let us consider a rectangular membrane, that is the following domain

$$D_{a,b} = \left\{ (x, y) \in \mathbb{R}^2 \mid 0 < x < a, 0 < y < b \right\} . \quad (3.29)$$

We want to find solutions to the wave equation

$$\left[\partial_t^2 - c^2 \partial_x^2 - c^2 \partial_y^2 \right] z(x, y, t) = 0 , \quad (x, y) \in D_{a,b} , \quad (3.30)$$

that satisfy Dirichlet boundary conditions on the boundary

$$z(x, y, t) = 0 , \quad \forall (x, y) \in \partial D_{a,b} . \quad (3.31)$$

These conditions are written more explicitly as

$$\begin{cases} z(0, y, t) = 0 \\ z(a, y, t) = 0 \end{cases} , \quad \begin{cases} z(x, 0, t) = 0 \\ z(x, b, t) = 0 \end{cases} . \quad (3.32)$$

We employ separation of variables, making the ansatz

$$z(x, y, t) = X(x)Y(y)T(t) . \quad (3.33)$$

Substitution in the wave equation (3.30) (dividing everything by $z(x, y, t)$) yields

$$\frac{1}{c^2} \frac{T''(t)}{T(t)} = \frac{X''(x)}{X(x)} + \frac{Y''(y)}{Y(y)} . \quad (3.34)$$

Here the left-hand side only depends on t while the right-hand side only depends on x and y . In order for it to be satisfied for any (x, y, t) , both sides must be equal to a constant. For future convenience, we choose this to be⁷ $-k^2$ for $k \in \mathbb{R}$. Hence we split the wave equation in two pieces

$$T''(t) = -c^2 k^2 T(t) , \quad (3.35a)$$

$$\frac{X''(x)}{X(x)} = -k^2 - \frac{Y''(y)}{Y(y)} , \quad (3.35b)$$

⁷The possibility of this constant to be positive is excluded *ex post* by the reality and boundary conditions.

an ODE for $T(t)$ and a 2-variable PDE for $X(x)$ and $Y(y)$. However, also in the latter the left-hand and right-hand sides depend on different variables! We can thus introduce a new constant $-\mu^2$ that both sides need to be equal to. Thus we perform a further split that brings us to three independent ODEs

$$T''(t) = -c^2 k^2 T(t) , \quad (3.36a)$$

$$X''(x) = -\mu^2 X(x) , \quad (3.36b)$$

$$Y''(y) = -\nu^2 Y(y) . \quad (3.36c)$$

Here we introduce the constant ν such that

$$k^2 = \mu^2 + \nu^2 . \quad (3.37)$$

We know perfectly well the solutions to the ODEs: they are sinusoidal functions

$$X(x) = A \cos(\mu x) + B \sin(\mu x) , \quad (3.38a)$$

$$Y(y) = C \cos(\nu y) + D \sin(\nu y) , \quad (3.38b)$$

$$T(t) = F \cos(kct) + G \sin(kct) . \quad (3.38c)$$

This solution depends on 8 arbitrary real constants A, B, C, D, F, G, μ and ν .

Now we are ready to impose the boundary conditions. Using the form (3.33), we see that they become

$$\begin{cases} X(0) = 0 \\ X(a) = 0 \end{cases} , \quad \begin{cases} Y(0) = 0 \\ Y(b) = 0 \end{cases} . \quad (3.39)$$

The upper conditions are easily solved by setting $A = C = 0$. The remaining two are asking for

$$B \sin(\mu a) = 0 , \quad D \sin(\nu b) = 0 . \quad (3.40)$$

Choosing $B = 0$ or $D = 0$ yields the trivial solution $z(x, y, t) = 0$ so we discard the possibilities. The only way to solve the above equations is to impose⁸

$$\mu = \frac{\pi}{a} m , \quad \nu = \frac{\pi}{b} n , \quad m, n \in \mathbb{N} . \quad (3.41)$$

The restriction to natural numbers \mathbb{N} instead of integers \mathbb{Z} comes from the fact that m and $-m$ give the same solution (up to redefinition of the constants F and G).

In conclusion, the wave equation on a 2-dimensional rectangular membrane with Dirichlet boundary admits the following solutions

$$z_{m,n}^D(x, y, t) = \sin\left(\frac{\pi}{a} mx\right) \sin\left(\frac{\pi}{b} ny\right) [F_{m,n} \cos(k_{m,n} ct) + G_{m,n} \sin(k_{m,n} ct)] , \quad (3.42)$$

with

$$k_{m,n} = \sqrt{\left(\frac{\pi}{a} m\right)^2 + \left(\frac{\pi}{b} n\right)^2} , \quad (3.43)$$

for any $m, n \in \mathbb{N}$ and $F_{m,n}, G_{m,n} \in \mathbb{R}$. These solutions are called *normal modes* of the membrane. Sometimes they are also called *eigenmodes*. Figure 3.2 depicts some of them.

⁸This kind of conditions – in which a parameter is forced to assume only a discrete set of values – are known as *quantisation conditions*. We will meet more of these when studying quantum mechanics.

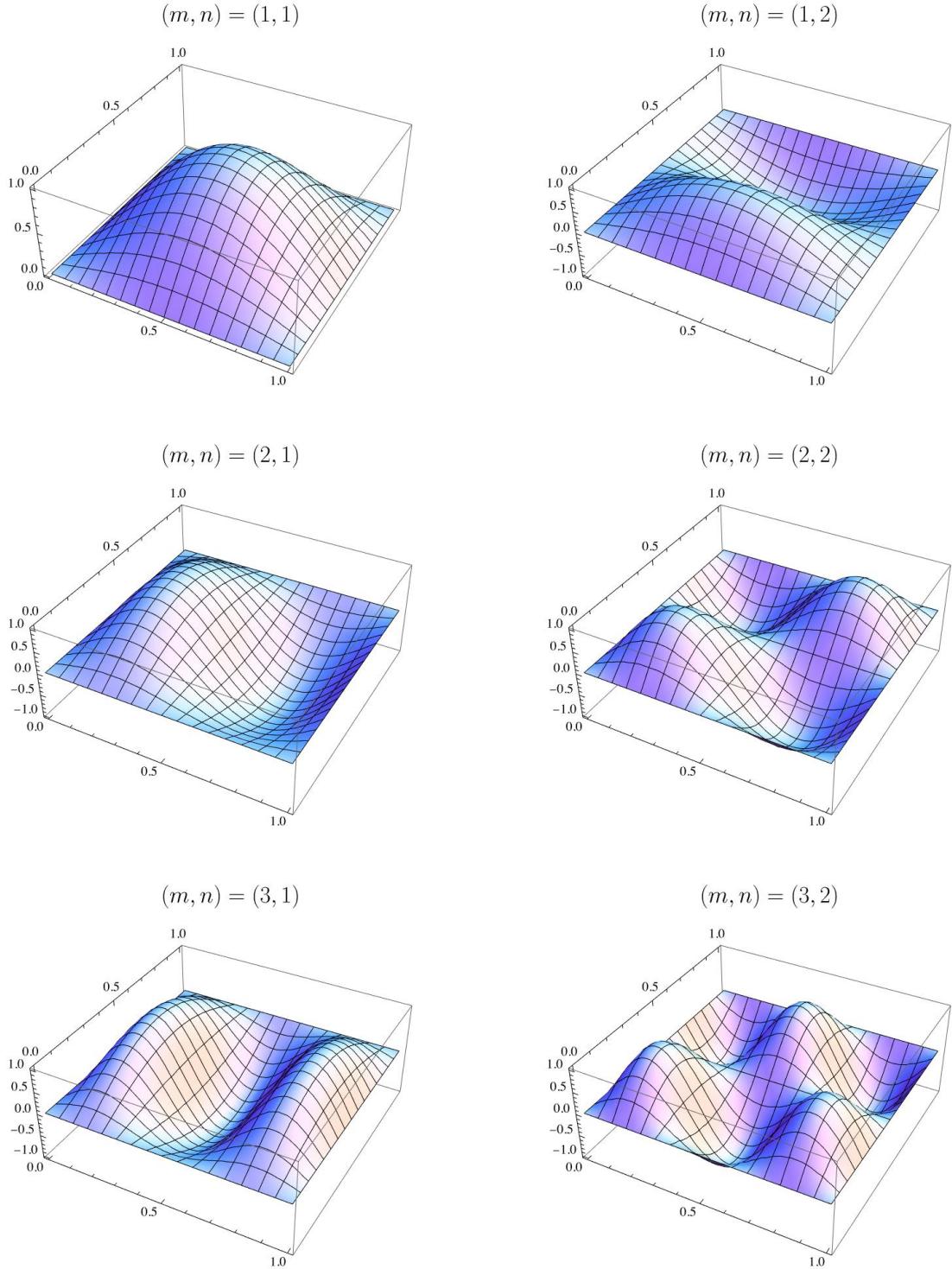


Figure 3.2: Example of normal modes of a Dirichlet membrane. Here $a = b = 1$ and $t = 0$. These solutions oscillate with time. Have a look at the video <https://www.youtube.com/watch?v=yDZsCZn31Sk> for some examples.

A generic solution will be, thanks to the superposition principle, a linear combination of normal modes

$$z^D(x, y, t) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} z_{m,n}^D(x, y, t) . \quad (3.44)$$

The free constants $F_{m,n}$ and $G_{m,n}$ will be determined by the initial conditions, exactly in the same way as for the 1-dimensional case: using the Fourier transform. Take the following generic conditions

$$z(x, y, 0) = z_0(x, y) , \quad \partial_t z(x, y, t)|_{t=0} = v_0(x, y) . \quad (3.45)$$

Substituting (3.44), we find

$$\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} F_{m,n} \sin\left(\frac{\pi}{a}mx\right) \sin\left(\frac{\pi}{b}ny\right) = z_0(x, y) , \quad (3.46a)$$

$$\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} G_{m,n} k_{n,m} c \sin\left(\frac{\pi}{a}mx\right) \sin\left(\frac{\pi}{b}ny\right) = z_0(x, y) . \quad (3.46b)$$

Do not let the double series scare you, everything works out just as you have already seen. It is sufficient to apply the orthogonality relation

$$\frac{2}{a} \int_0^a dx \sin\left(\frac{\pi}{a}mx\right) \sin\left(\frac{\pi}{a}m'x\right) = \delta_{m,m'} , \quad (3.47)$$

both for x and for y (trading a for b). In the end we compute the coefficients in terms of the initial conditions as

$$F_{m,n} = \frac{4}{ab} \int_0^a dx \int_0^b dy z_0(x, y) \sin\left(\frac{\pi}{a}mx\right) \sin\left(\frac{\pi}{b}ny\right) , \quad (3.48a)$$

$$F_{m,n} = \frac{1}{k_{m,n}} \frac{4}{ab} \int_0^a dx \int_0^b dy v_0(x, y) \sin\left(\frac{\pi}{a}mx\right) \sin\left(\frac{\pi}{b}ny\right) \quad (3.48b)$$

Free boundary

What if the membrane has a free boundary? The answer is quite easy. Almost everything we have just seen applies here as well. The main difference is that the boundary condition will now be of Neumann type, that is

$$\hat{n} \cdot \vec{\nabla} z(x, y, t) = 0 , \quad \forall (x, y) \in \partial D_{a,b} , \quad (3.49)$$

where \hat{n} is the unit normal to the boundary $\partial D_{a,b}$. These condition are explicitly written as

$$\begin{cases} \partial_x z(x, y, t)|_{x=0} = 0 \\ \partial_x z(x, y, t)|_{x=a} = 0 \end{cases} , \quad \begin{cases} \partial_y z(x, y, t)|_{y=0} = 0 \\ \partial_y z(x, y, t)|_{y=b} = 0 \end{cases} . \quad (3.50)$$

Proceeding just as we did for the Dirichlet case, we arrive at the normal modes

$$z_{m,n}^N(x, y, t) = \cos\left(\frac{\pi}{a}mx\right) \cos\left(\frac{\pi}{b}ny\right) [F_{m,n} \cos(k_{m,n}ct) + G_{m,n} \sin(k_{m,n}ct)] . \quad (3.51)$$

Figure 3.3 depicts some of them.

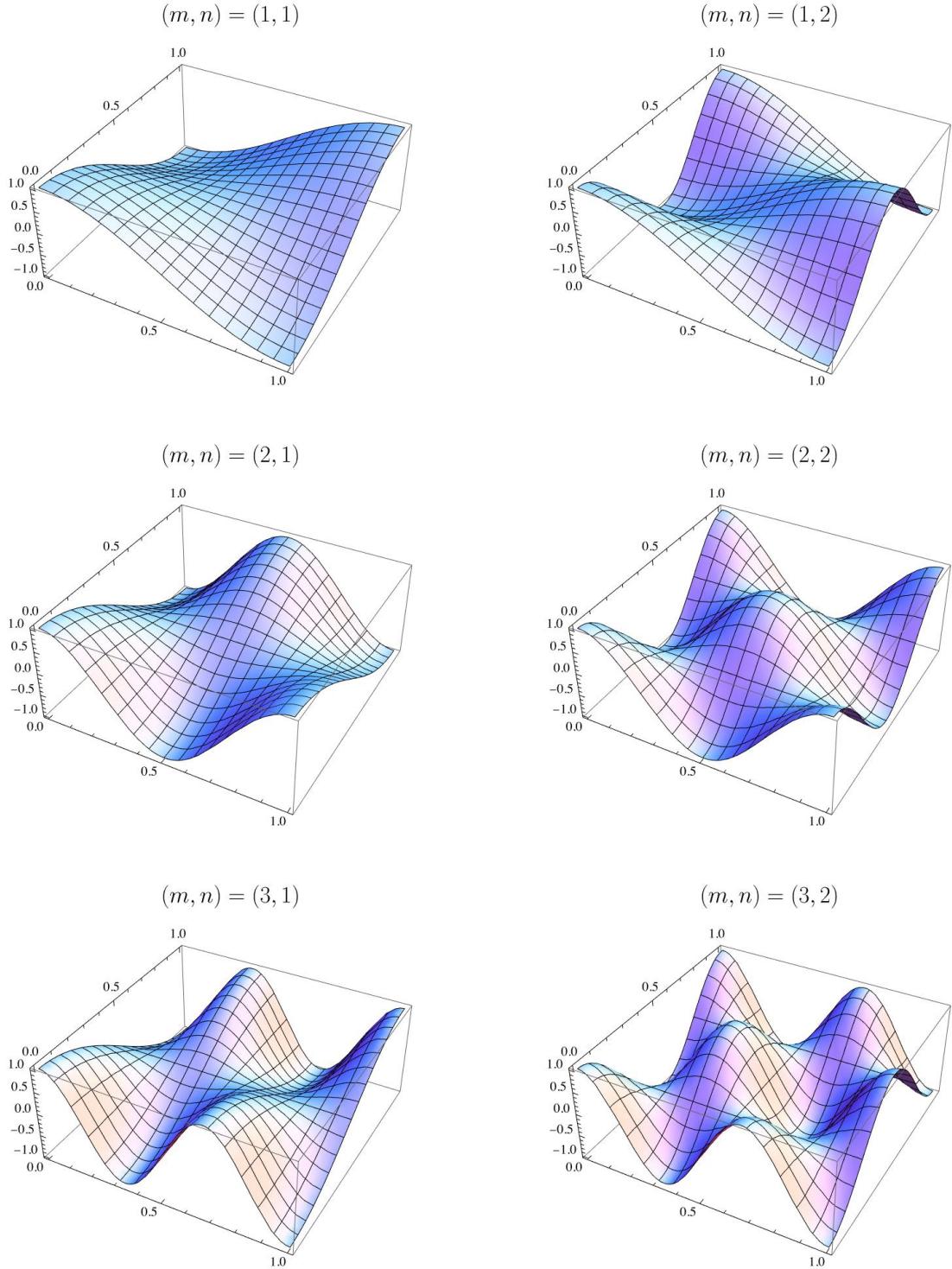


Figure 3.3: Example of normal modes of a Neumann membrane (with free boundary). Here $a = b = 1$ and $t = 0$. These solutions oscillate with time.

Chladni plates

Chladni plates take their name from the 18th Century German physicist and musician Ernst Chladni. They are a very simple and clever instrument providing a visualisation of the effects of vibrations on mechanical surfaces. They are mainly known for the beautiful patterns they produce. See e.g. <https://www.youtube.com/watch?v=OLNFrXgMJ6E> or

<https://www.youtube.com/watch?v=CsjV1gjBMbQ>. The idea is to take a thin metal plate, usually a square or rectangular one, secure it so that it can vibrate (meaning usually that its edges are free), sprinkle its surface with sand and finally make it vibrate – historically this was done by scratching its sides with a violin bow. What will happen is that the sand, at first uniformly distributed, will be pushed from the parts of the surface that vibrate the most, to those that vibrate the least. Ultimately the sand grains will accumulate along the *nodal lines*, that is the lines on the surface that experience no motion.

Mathematically, when a square plate is “played” one of the fundamental frequencies $\omega_{m,n} = k_{m,n}c = \omega_0$ is excited⁹, meaning that in the superposition (3.50), only the normal modes for which $k_{m,n}c = \omega_0$ will appear in the sum. Clearly there are two of these, related by an exchange of m and n . So the Chladni patterns are given by the solutions to the equation

$$\cos\left(\frac{\pi}{a}mx\right)\cos\left(\frac{\pi}{a}ny\right) + \cos\left(\frac{\pi}{a}nx\right)\cos\left(\frac{\pi}{a}my\right) = 0, \quad n = \sqrt{\left(\frac{a^2\omega_0^2}{\pi^2c^2} - m^2\right)}. \quad (3.52)$$

Obviously ω_0 should be chosen appropriately, so that $n \in \mathbb{Z}$. For rectangular membrane, the argument of the symmetry $(m, n) \leftrightarrow (n, m)$ does not hold anymore and more care must be made to identify all the modes with equal frequency. Figure 3.4 collects some patterns for the square plate with free boundary.

3.3 Circular membranes

Now we are going to consider a circular membrane of radius a and with boundary fixed by a Dirichlet condition. Mathematically, this means considering a domain

$$D_a = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < a^2\}, \quad (3.53)$$

and solving the wave equation on it. The nature of the boundary condition

$$z(x, y, t) = 0, \quad (x, y) \in \partial D_a, \quad (3.54)$$

makes it very difficult to solve in cartesian coordinates

$$z(x, \sqrt{a^2 - x^2}, t) = 0, \quad z(x, -\sqrt{a^2 - x^2}, t) = 0. \quad (3.55)$$

The best way to proceed is to use coordinates which are naturally adapted to the geometry of the system: polar coordinates (r, θ)

$$x = r \cos \theta, \quad y = r \sin \theta. \quad (3.56)$$

Hence, let us introduce

$$Z(r, \theta, t) \equiv z(x(r, \theta), y(r, \theta), t), \quad (3.57)$$

the solution to the wave equation in D_a , expressed in polar coordinates. It is a well-known fact¹⁰ that the Laplacian operator in polar coordinates is

$$\nabla^2 = \partial_r^2 + \frac{1}{r}\partial_r + \frac{1}{r^2}\partial_\theta^2. \quad (3.58)$$

⁹Nowadays, Chladni plates are usually excited through speakers. In this case, the speaker must emit an appropriate frequency, called *resonant*, in order for the patterns to emerge.

¹⁰You have seen how to write $\vec{\nabla}$ in polar coordinates in the Mathematics for the Sciences II module last year. See §6.5 in the [lecture notes](#).

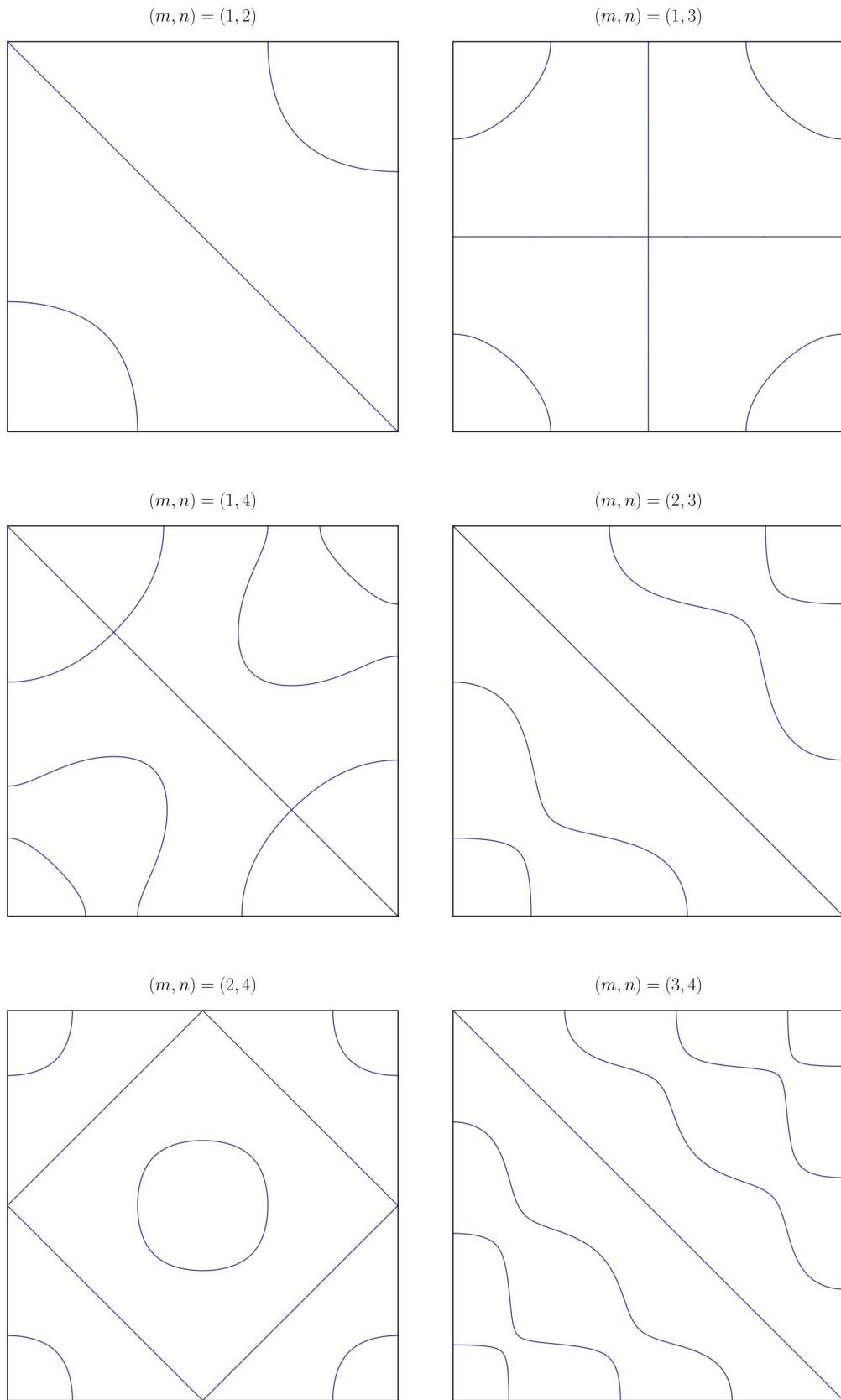


Figure 3.4: Plot of nodal lines – or Chladni patterns – in a square plate for different frequencies $\omega_{m,n} = ck_{m,n}$, with $k_{m,n}$ given by (3.43) (for $b = a = 1$).

If you wish, you can prove the above identity by acting with it on (3.57).

So, we have to deal with the following PDE

$$\frac{1}{c^2} \partial_t^2 Z(r, \theta, t) = \partial_r^2 Z(r, \theta, t) + \frac{1}{r} \partial_r Z(r, \theta, t) + \frac{1}{r^2} \partial_\theta^2 Z(r, \theta, t) . \quad (3.59)$$

We apply separation of coordinates here as well, writing

$$Z(r, \theta, t) = R(r)\Theta(\theta)T(t) . \quad (3.60)$$

The first step is the same as before and yields a first split

$$T''(t) = -k^2 c^2 T(t) , \quad (3.61a)$$

$$\frac{R''(r)}{R(r)} + \frac{R'(r)}{rR(r)} + \frac{\Theta''(\theta)}{r^2\Theta(\theta)} = -k^2 . \quad (3.61b)$$

The second equation can be further split by rearranging its terms as follows

$$r^2 \left(\frac{R''(r)}{R(r)} + \frac{R'(r)}{rR(r)} + k^2 \right) = -\frac{\Theta''(\theta)}{\Theta(\theta)} . \quad (3.62)$$

We equate both sides to a positive constant n^2 , since if we chose a negative constant we would not be able to find a solution. Hence we have completely separated our PDE in three independent ODEs

$$R''(r) + \frac{1}{r} R'(r) + \left(k^2 - \frac{n^2}{r^2} \right) R(r) = 0 , \quad (3.63a)$$

$$\Theta''(\theta) = -n^2 \Theta(\theta) , \quad (3.63b)$$

$$T''(t) = -k^2 c^2 T(t) . \quad (3.63c)$$

The solutions to the last two ODEs are well known

$$\Theta(\theta) = A \cos(n\theta) + B \sin(n\theta) , \quad (3.64a)$$

$$T(t) = F \cos(kct) + G \sin(kct) . \quad (3.64b)$$

There is one subtlety though. The variable θ is a polar angle. This means that θ and $\theta + 2\pi$ correspond to the same point on the plane. This has an important consequence: functions of θ need to be **periodic**

$$\Theta(\theta + 2\pi) = \Theta(\theta) \implies n \in \mathbb{Z} . \quad (3.65)$$

The ODE (3.63a) for $R(r)$ might have an unfamiliar form. Let us define $\rho = kr$, so that the equation now takes the form

$$\tilde{R}''(\rho) + \frac{1}{\rho} \tilde{R}'(\rho) + \left(1 - \frac{\lambda^2}{\rho^2} \right) \tilde{R}(\rho) = 0 , \quad (3.66)$$

where $\tilde{R}(\rho) = R(r)$ and $\lambda \in \mathbb{C}$ was substituted to $n \in \mathbb{Z}$ for generality. This alteration probably did not help much in the way of making the equation more familiar to you. However, (3.66) is a very well-known expression, known as the *Bessel differential equation*.

Its solutions are not elementary functions¹¹. Later in this module, we will learn a method to analyse ODEs using power series expansions, but, for the time being, we will skip a detailed analysis and simply consider the solutions as given to us by the Gods. We are interested in solutions that are non-singular¹² at $\rho = 0$. They are called *Bessel functions of the first kind*. This is a mouthful, so we will refer to them as *Bessel J functions*. They are denoted by the symbol $J_\lambda(\rho)$, with the index λ being called the *order* of the function. Bessel J functions with integer order $\lambda = n \in \mathbb{Z}$ enjoy important properties that those with non-integer order do not: they are *entire* – i.e. they have no singularity or branch cut for $x \in \mathbb{C}$ – and are even/odd under sign flip of the order

$$J_{-n}(\rho) = (-1)^n J_n(\rho), \quad \forall n \in \mathbb{Z}. \quad (3.67)$$

The Bessel J functions have an infinite amount of zeroes on the positive real line. In a sense, you can think of these functions as deformed sinusoids. We will denote by $\rho_{n,m}$ the position of the m -th positive zero of $J_n(\rho)$:

$$J_n(\rho_{n,m}) = 0, \quad 0 < \rho_{n,m-1} < \rho_{n,m}, \quad \forall m \in \mathbb{N}, \quad \forall n \in \mathbb{Z}. \quad (3.68)$$

The zeroes of Bessel J functions are *transcendental*, i.e. they are not the root of a non-zero polynomial of finite degree with rational coefficients. They are computed numerically and you can find them tabulated in several places. See e.g. [olver1960]. If you are familiar with Mathematica[©], you can obtain $\rho_{n,m}$ with the command `BesselJZero[n,m]`. These functions have several other properties¹³, but these are the only ones that matter to us. Figure 3.5 plots $J_n(\rho)$ for the first few integer values of the order n .

So, the ODE we are interested in (3.63a) admits the following solution regular at $r = 0$

$$R(r) = C J_n(kr), \quad (3.69)$$

with C an arbitrary constant. Thus the solution to the wave equation on a circular membrane takes the following form (we absorbed the constants A, B, C into the other ones)

$$\begin{aligned} Z(r, \theta, t) = J_n(kr) &\left[\cos(n\theta) \left(F \cos(kct) + G \sin(kct) \right) + \right. \\ &\left. + \sin(n\theta) \left(\tilde{F} \cos(kct) + \tilde{G} \sin(kct) \right) \right], \quad n \in \mathbb{N} \cup \{0\}. \end{aligned} \quad (3.70)$$

We still need to impose the Dirichlet boundary condition. Note however that we already have a quantisation condition for n . This came from requiring periodicity in θ . If you think about $Z(r, \theta, t)$ as a function on the (r, θ) plane, then asking periodicity in θ works just like a boundary condition, as it imposes some requirements at the boundary of the θ domain $(0, 2\pi)$. We also eliminated one of the constants by asking for regularity at $r = 0$. So, we are left with the condition

$$Z(a, \theta, t) = 0 \implies J_n(ka) = 0, \quad (3.71)$$

¹¹Elementary functions are defined as sums, products, roots and compositions of finitely many polynomial, rational, trigonometric, hyperbolic, and exponential functions, including possibly their inverse functions.

¹²The fact that the coefficients of (3.66) are singular at $\rho = 0$ is a signal that solutions may potentially diverge as $\rho \rightarrow 0$. For the Bessel equation, it turns out that there exist a solution which is regular at $\rho = 0$.

¹³If you are interested, have a look at the Digital Library of Mathematical Functions, Chapter 10 <https://dlmf.nist.gov/10>.

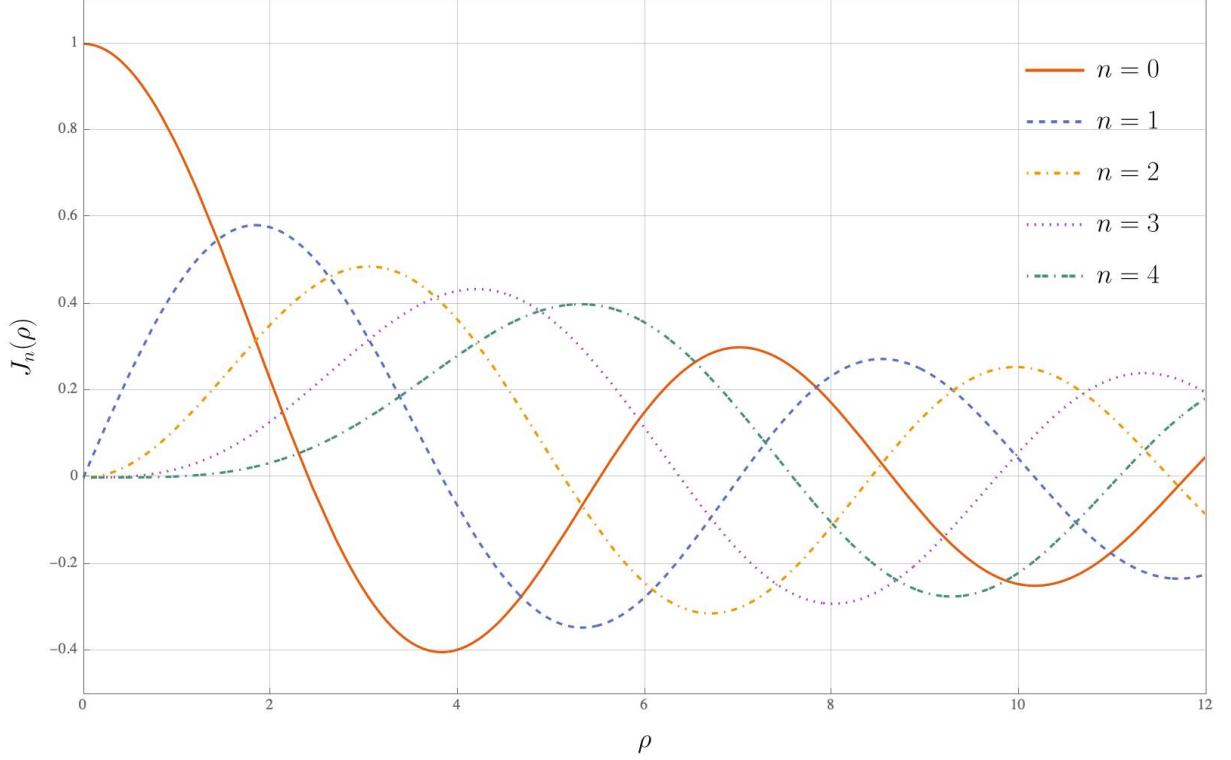


Figure 3.5: Plot of Bessel J functions for the first few integer orders. You can see that they look somewhat like sinusoidal functions.

that we solve by quantising k

$$k = \frac{\rho_{n,m}}{a} . \quad (3.72)$$

In conclusion, the normal modes of a circular membrane with fixed boundary have the following form

$$\boxed{Z_{n,m}^D(r, \theta, t) = J_n\left(\frac{\rho_{n,m}}{a}r\right) \times \left[\cos(n\theta) \left(F_{n,m} \cos\left(\frac{\rho_{n,m}}{a}ct\right) + G_{n,m} \sin\left(\frac{\rho_{n,m}}{a}ct\right) \right) + \sin(n\theta) \left(\tilde{F}_{n,m} \cos\left(\frac{\rho_{n,m}}{a}ct\right) + \tilde{G}_{n,m} \sin\left(\frac{\rho_{n,m}}{a}ct\right) \right) \right]} . \quad (3.73)$$

Figures 3.6 and 3.7 plot some of these normal modes for your visual enjoyment.

A general wave on a circular membrane with fixed boundary is given by a superposition of its normal modes

$$Z^D(r, \theta, t) = \sum_{n=0}^{\infty} \sum_{m=1}^{\infty} Z_{n,m}^D(r, \theta, t) . \quad (3.74)$$

To determine the various constants $F_{n,m}$, $G_{n,m}$, $\tilde{F}_{n,m}$ and $\tilde{G}_{n,m}$ we need to impose initial conditions

$$Z^D(r, \theta, 0) = Z_0(r, \theta) , \quad \partial_t Z^D(r, \theta, t) \Big|_{t=0} = v_0(r, \theta) . \quad (3.75)$$

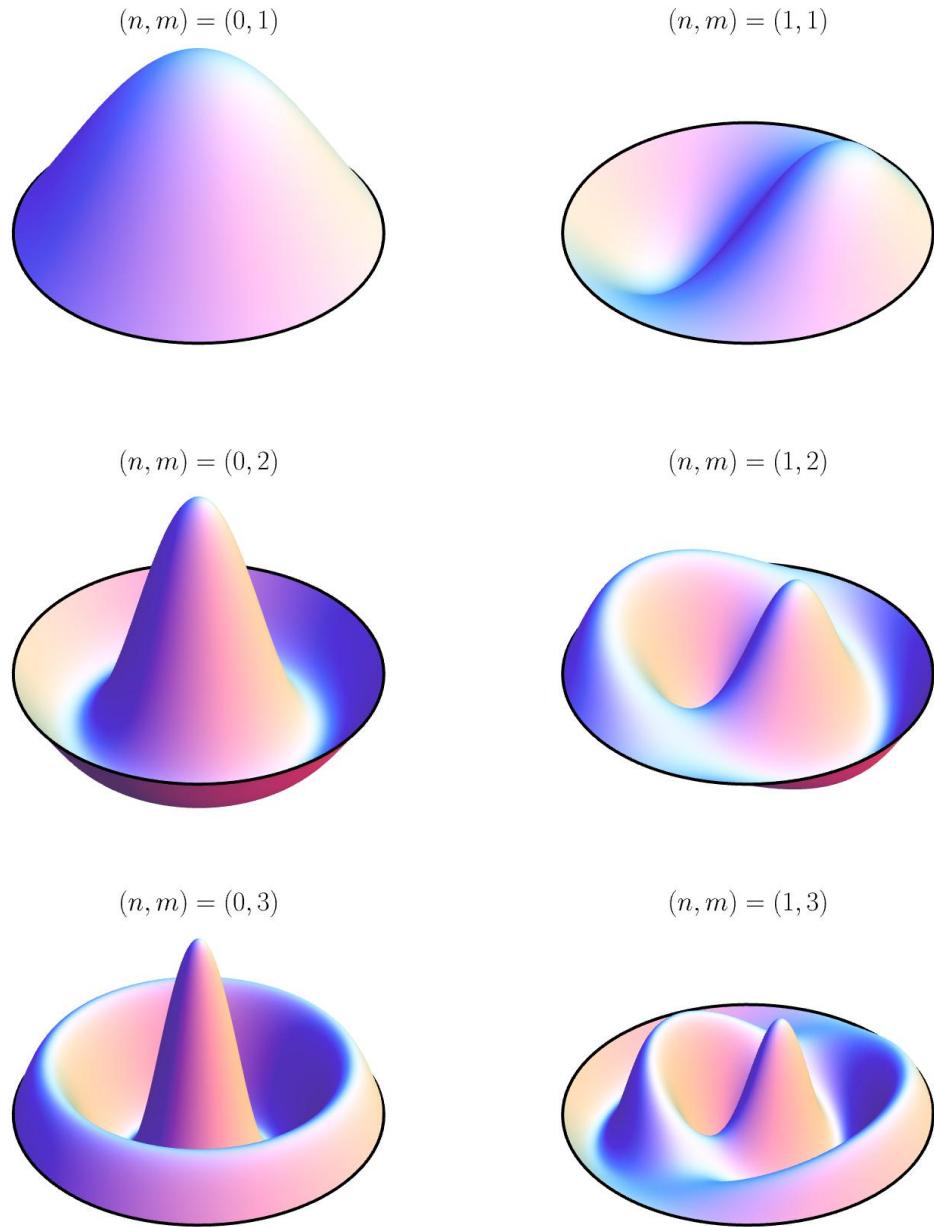


Figure 3.6: Plot of some normal modes of a circular membrane with fixed boundary.

There seems to be a problem however. We can invert the sum over n using the Fourier

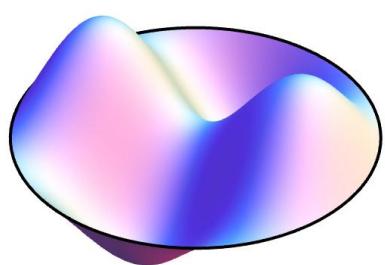
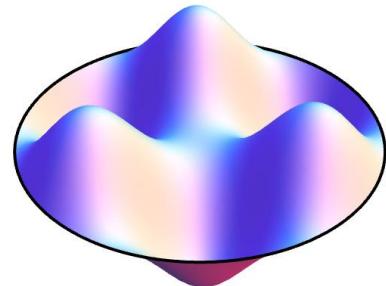
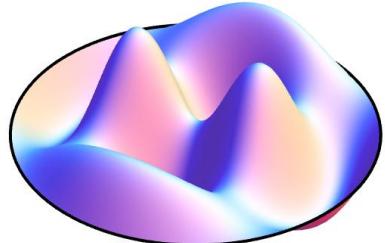
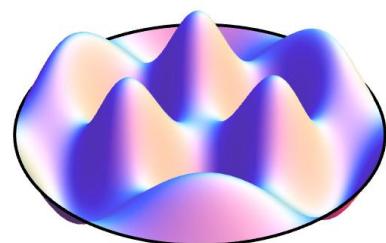
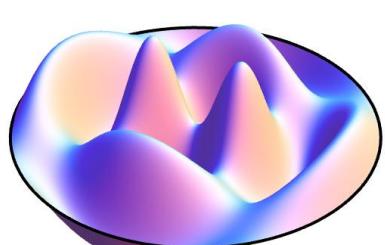
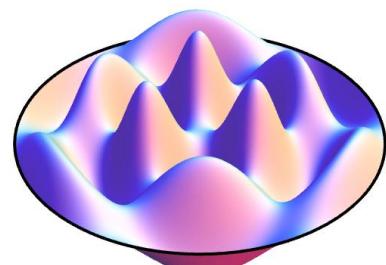
$(n, m) = (2, 1)$  $(n, m) = (3, 1)$  $(n, m) = (2, 2)$  $(n, m) = (3, 2)$  $(n, m) = (2, 3)$  $(n, m) = (3, 3)$ 

Figure 3.7: Plot of some normal modes of a circular membrane with fixed boundary, cont'd.

transform formulae (the cases $n = 0$ and/or $n' = 0$ are trivial)

$$\frac{1}{\pi} \int_0^{2\pi} d\theta \sin(n\theta) \sin(n'\theta) = \delta_{n,n'} , \quad 67$$
$$\frac{1}{\pi} \int_0^{2\pi} d\theta \cos(n\theta) \cos(n'\theta) = \delta_{n,n'} , \quad \forall n, n' \in \mathbb{N} , \quad (3.76)$$

but what about the sum over m ? Well, as it turns out, Bessel J functions enjoy their own orthogonality relation

$$\frac{2}{a^2} \int_0^a dr J_n\left(\rho_{n,m} \frac{r}{a}\right) J_n\left(\rho_{n,m'} \frac{r}{a}\right) = \left(J_{n+1}(\rho_{n,m})\right)^2 \delta_{m,m'} . \quad (3.77)$$

Due to this similarity between trigonometric and Bessel functions, series such as (3.74) are called *Fourier-Bessel series*.

Outro: takeaway points

In this chapter, we extended what we learned previously to vibrating bodies in 2 dimensions. We looked at two examples of vibrating membranes: rectangular and circular. The takeaway points are the following.

- We derived – under the same assumptions as for the string – the 2-dimensional wave equation

$$\partial_t^2 z(x, y, t) - c^2 \nabla^2 z(x, y, t) = 0 ,$$

where $z(x, y, t)$ is the vertical displacement of the membrane from the plane $(x, y, z = 0)$ in \mathbb{R}^3 , at time t , c is the wave velocity, related to the density ρ of the membrane and its tension T by

$$c = \sqrt{\frac{T}{\rho}} ,$$

and ∇^2 is the Laplace operator

$$\nabla^2 = \partial_x^2 + \partial_y^2 .$$

- We studied the energy density of the vibrating membrane

$$\mathcal{E}(x, y, t) = \frac{\rho}{2} (\partial_t z(x, y, t))^2 + \frac{T}{2} \left| \vec{\nabla} z(x, y, t) \right|^2 ,$$

where $\vec{\nabla}$ is the nabla operator, acting on scalar functions as a gradient

$$\vec{\nabla} f(x, y, t) = \begin{pmatrix} \partial_x f(x, y, t) \\ \partial_y f(x, y, t) \end{pmatrix} .$$

Further, we saw that the energy density satisfies a conservation equation analogous to the 1-dimensional case

$$\partial_t \mathcal{E}(x, y, t) + \vec{\nabla} \cdot \vec{\mathcal{F}}(x, y, t) = 0 ,$$

where $\vec{\mathcal{F}}$ is energy flux density vector

$$\vec{\mathcal{F}}(x, y, t) = -T \partial_t z(x, y, t) \vec{\nabla} z(x, y, t) .$$

Again, thanks to the divergence theorem, the variation in time of the total energy in a region R of the membrane is equal to (minus) the flux of energy across the boundary of the region ∂R

$$\frac{dE}{dt} = - \oint_{\partial R} ds \hat{n} \cdot \vec{\mathcal{F}}(x, y, t) .$$

- We introduced the concept of plane wave: a wave that only varies along a single direction $\hat{n} = (\ n_x \ n_y)^T$

$$z(x, y, t) = z(\hat{n} \cdot \vec{x}, t) .$$

We saw that for such a wave the 2-dimensional wave equation reduces to the 1-dimensional one. Hence a plane wave satisfying the wave equation must be the sum of two travelling waves

$$z(x, y, t) = f(\hat{n} \cdot x - ct) + f(\hat{n} \cdot x + ct) .$$

- We extended the definition of harmonic plane wave to the 2-dimensional case

$$h(x, y, t) = e^{i(\vec{k} \cdot \vec{x} - \omega t)} ,$$

with \vec{k} a 2-dimensional wave vector.

- We saw that, writing all the equations in vector formalism, we can immediately generalise our discussion of 2-dimensional waves to arbitrary dimensions.
- We studied the vibrations of a rectangular membrane, using the separation of variables approach. In the case of fixed boundary (Dirichlet condition) we found the general solution

$$z^D(x, y, t) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} z_{m,n}^D(x, y, t) ,$$

as a superposition of harmonics

$$z_{m,n}^D(x, y, t) = \sin\left(\frac{\pi}{a}mx\right) \sin\left(\frac{\pi}{b}ny\right) \left[F_{m,n} \cos(k_{m,n}ct) + G_{m,n} \sin(k_{m,n}ct) \right] .$$

Here a and b are the x and y lengths of the membrane, respectively, while

$$k_{m,n} = \sqrt{\left(\frac{\pi}{a}m\right)^2 + \left(\frac{\pi}{b}n\right)^2} , \quad \forall m, n \in \mathbb{N} .$$

- We repeated the above analysis for the case of a circular membrane of radius a with fixed (Dirichlet) boundary. In this situation, it is convenient to switch coordinates to polar ones $x = r \cos \theta$ and $y = r \sin \theta$. We applied again the separation of variables, arriving at a general solution of the form

$$Z^D(r, \theta, t) = \sum_{n=0}^{\infty} \sum_{m=1}^{\infty} Z_{n,m}^D(r, \theta, t) ,$$

where $Z_{n,m}^D$ are the normal modes

$$\begin{aligned} Z_{n,m}^D(r, \theta, t) &= J_n\left(\frac{\rho_{n,m}}{a}r\right) \times \\ &\times \left[\cos(n\theta) \left(F_{n,m} \cos\left(\frac{\rho_{n,m}}{a}ct\right) + G_{n,m} \sin\left(\frac{\rho_{n,m}}{a}ct\right) \right) + \right. \\ &\left. + \sin(n\theta) \left(\tilde{F}_{n,m} \cos\left(\frac{\rho_{n,m}}{a}ct\right) + \tilde{G}_{n,m} \sin\left(\frac{\rho_{n,m}}{a}ct\right) \right) \right] . \end{aligned}$$

Here $J_n(x)$ is the Bessel J function of order $n \geq 0$, while $\rho_{n,m}$ is the m -th positive zero of $J_n(x)$. Finally, we saw that the Bessel J functions, similarly to sinusoidal functions, obey an orthogonality relation that allows to fix the constants $F_{n,m}$, $G_{n,m}$, $\tilde{F}_{n,m}$ and $\tilde{G}_{n,m}$ in terms of initial conditions.

Part II

Quantum Dynamics

The advent of quantum mechanics at the beginning of 20th Century was groundbreaking, so much so that many consider it to be the single greatest achievement in the history of human civilisation. This discovery was also bewildering. The picture that quantum mechanics painted was a preposterous departure from the classical, comforting view of a “clockwork” Nature, governed by deterministic laws that we can ultimately understand and use to exactly predict the future. In this perspective, the uncertainties that plague experiments and predictions are a simple consequence of our ignorance. Any classical system that appears random does so because it is somehow difficult, but never impossible, for us to know what’s going on inside.

Quantum mechanics teaches us that this is not the way the world works. The randomness is real – it exists irrespective of us – the unpredictability is inherent. There exists a limit to how much information we can gain on any physical system. We cannot consider anymore particles as entities with, say, definite position and momentum, since the more we try to pin down their precise position, the more their momentum will be uncertain. In a sense, this is because quantum particles are fragile objects: the very act of looking – that is to say, measure – changes them, disturbing many of the other properties that they possess.

All this seems to be preposterous, and surely that was the impression that physicists had at the beginning of last Century as they attempted to make sense of a series of puzzling experimental results. But their hand was forced, because while the framework of classical physics could provide no cure to these puzzles, quantum mechanics resolved them all. As weird and outlandish as it appears, it undoubtedly is a correct description of the world we inhabit.

Among the many successes of quantum mechanics are answers to old and very basic questions. Why is matter stable? Why does the sun shine and why is it yellow? Why are all solids either conductors or insulators? But quantum mechanics also opens up vistas that we didn’t previously know existed, from novel states of matter where the constituent particles can be coaxed to perform seemingly impossible tasks¹⁴, to the subatomic world of fluctuating ethereal fields.

The goal of this second part is to begin to get to grips with the quantum world and understand some of the strange behaviour that it exhibits. Clearly, classical intuition will not be of much use as we step into the unknown. Fortunately the mathematics that you learned – along with some that you will be learning in the future – is an astonishingly appropriate language to describe the quantum world. We will embrace this mathematical description and leverage it to build a new intuition for how the universe really works.

¹⁴Think of superconductors, superfluids or Bose-Einstein condensates.

Chapter 4

Introduction to quantum dynamics

A condensed history of the birth of quantum mechanics

During the 17th Century, a debate raged between two of the most eminent natural philosophers¹ of the time: Isaac Newton and Christiaan Huygens. The debate revolved around the nature of light: *is it made of particles or is it a wave?* Following his view of reality as interactions of material points through forces, Newton elaborated his *corpuscular theory of light* which he made public in his treatise Opticks (1704). He argued that the geometric nature of reflection and refraction of light could only be explained if light were made of particles, since waves do not tend to travel in straight lines. He further supported his theory with experiments that showed no diffraction or interference, which are characteristic phenomena of waves. Huygens, on the other hand, claimed that light is a wave, first communicating this in 1678 to the Paris Académie des Sciences and later publishing his theory in his Traité de la Lumière (1690). He posited light to be a longitudinal wave, whose wavelength was so small that typical phenomena such as diffraction could not be appreciated experimentally.

Ultimately the corpuscular theory of Newton was widely accepted by the scientific community of the 18th Century as describing the true nature of light. At least until 1801, when Thomas Young reported the results of a series of interference experiments he had performed. They could not be explained by Newton's theory and so, Huygens' wave interpretation was revived and revised, mainly by Augustin-Jean Fresnel.

So Newton was wrong. Right? Well, not exactly. Sure, by the end of the 19th Century, Henry Clerk Maxwell had fully developed his theory of electromagnetism and the nature of light as a propagating electromagnetic wave was mostly cemented. However a number of strange phenomena were troubling the minds of physicists. Amongst these was the problem of the *black body radiation*. A black body is an idealised, opaque and non-reflective object². In thermal equilibrium, the black body spontaneously emits radiation in the form of electromagnetic waves. Classical theory implied that the energy of the radiation emitted by a black body is equally partitioned between the allowed frequencies³.

¹This is how adepts of the sciences used to be called at the time.

²Experimentally, it was realised as an insulated cavity with a punched-in hole so small as to be irrelevant on the thermal equilibrium of the body (i.e. very little energy escapes the cavity).

³These are the resonant frequencies of waves in a three-dimensional region that depend on the geometry of the region. Remember what we saw for waves on a string and on a membrane.

But since there is an infinite number of frequencies, the power radiated from the black body has to be infinite! This obviously makes no sense and is famously known as the *ultraviolet catastrophe*. It turned out to be the harbinger of the quantum revolution. In 1897, Max Planck set out to find a solution to the catastrophe. After three years, in 1900, he proposed an empirical law able to explain the experiments and resolved the ultraviolet catastrophe. The only physical explanation he could give⁴ to explain this formula was that the energy of light had to be *quantised*. This meant that light with frequency ω is emitted in “packets” of energy $E = \hbar\omega$, where

$$\hbar \approx 1.05 \times 10^{-34} \text{ J} \cdot \text{s} . ,$$

is known as the *Planck's constant*. It is sometimes custom to refer to $h = 2\pi\hbar$ as Planck constant and to \hbar as the *reduced Planck constant*. We find it more practical to always use \hbar . These quanta of energy are so small compared to human scales that light's energy appears continuous in our experience.

So, are these packets particles? Is this quantisation an inherent feature of light or just a feature of the black-body radiation or, worse, an ad-hoc patch to explain a mysterious phenomenon? In 1905, Albert Einstein explained the *photoelectric effect*, another baffling phenomenon wherein a material absorbs light and emits electrons. The basic idea is clear: the incident light's energy is transferred to one the electrons on the surface of the material. When one electron absorbs as much energy as that which binds it to the atom, it breaks free and escapes from the material. However, there was a problem. The emission of electrons seemed not to be influenced by the *intensity of the light*, as appears to be reasonable, but rather to its *frequency*. Experiment showed that low frequency, high intensity beams of light failed to produce photo-electrons, while high frequency, low intensity ones managed to. This brought Einstein to propose that a beam of light is not a wave propagating through space, but a swarm of discrete energy packets. So quantisation truly is an inherent feature of light and, what's more, these packets of energy act as if they were particles. They were later given the name of photons by Gilbert Lewis in 1926.

But wait, light is a wave as well! That was quite clear to physicists at the beginning of last Century. It possesses a wavelength, it experiences diffraction and interference. So what gives? The (only possible) interpretation is that light is a wave and was made of particles at the same time. This is the *wave-corpuscular nature* of light.

And the surprises were far from finished. In 1913 Niels Bohr took inspiration from Planck and Einstein results and proposed an explanation for why atoms do not collapse. You see, if you model an atom as a dense nucleus surrounded by orbiting electrons⁵ – kind of like a miniature solar system – classical physics tells you that these electrons are bound to collapse on the nucleus. That is because an accelerating electron radiates power according to Larmor formula. Bohr proposed that electrons can revolve on stable orbits – the *stationary orbits* – without radiating energy, but the radii of these orbits form a discrete sequence. The electron cannot occupy the space between stationary orbits – and thus cannot continuously radiate energy – but is allowed to jump from one orbit to another, emitting electromagnetic radiation in quanta of energy. Just like light does.

⁴Planck himself referred to his accepting this explanation as an *act of desperation*.

⁵This was the naive picture that emerged from the experiments of Ernest Rutherford in the early 20th Century.

But electrons are still particles, right? Enter the stage, Louis de Broglie. In his PhD thesis, in 1924, he proposed that all particles, matter and light, are associated with waves, having frequency ν and wavelength λ related to the energy E and momentum p of the particle through the Planck constant

$$E = h\nu , \quad p = h/\lambda .$$

With this interpretation of particles, Bohr's quantisation of the electron's orbits in an atom follow suit. Note that the de Broglie waves exist for any object, though as their average size grows their wavelength becomes increasingly small (since the mass and thus p becomes large) and for macroscopic object is essentially immaterial.

Example

Consider an average oxygen atom at room temperature (that is, about 300 K). The kinetic theory of gases tells us that the typical velocity of particles of mass m at a temperature T is given by the formula

$$v = \sqrt{\frac{3k_B T}{m}} ,$$

with $k_B = 1.38 \times 10^{-23} \text{ JK}^{-1}$ being the *Boltzmann constant*. From this information, we can compute the typical de Broglie wavelength as

$$\lambda = \frac{h}{p} = \frac{h}{mv} = \frac{h}{\sqrt{3k_B T m}} .$$

Since the mass of an oxygen atom is $m = 2.66 \times 10^{-26} \text{ kg}$, we can easily compute

$$\lambda_{\text{oxygen r.t.}} = \frac{6.63}{\sqrt{3 \cdot 1.38 \cdot 30 \cdot 2.66}} \times 10^{-10} \text{ m} = \frac{6.63}{18.2} \times 10^{-10} \text{ m} = 0.36 \times 10^{-10} \text{ m} .$$

It is on this backdrop that Werner Heisenberg, in 1925, and Erwin Schrödinger, in 1926, formulated two distinct versions of quantum mechanics. In Heisenberg's framework, physical observables are represented by matrices and a system by a state vector. The act of measuring an observable in a system correspond to the action of a matrix on the state vector which might well not be an eigenvector. Additionally, different observables might correspond to non-commuting matrices, implying that the act of measuring a quantity might influence other physical observables. This, in a nutshell, is the Heisenberg uncertainty principle. Schrödinger, on the other hand, formulated a different version of quantum mechanics, based on wave dynamics. He postulated, based on the empirical results that were available, a wave equation together with a framework to extract physical observables from its solutions. However, the interpretation of these solutions remained nebulous. It was Max Born, a few months after, that understood correctly what these solutions were: Schrödinger waves are probability amplitudes, i.e. they must be squared to obtain a probability density that can then be used to weight physical observables. We are going to see how this works in these lectures. Also in 1926 Paul Dirac proved the equivalence of Heisenberg's and Schrödinger's approaches: two different languages describing the same reality.

Finally, in 1927, the results of the celebrated double slit experiment of Clinton Davisson and Lester Germer were announced, confirming the existence of the de Broglie waves

and, in all effects, affirming the incontrovertible truth of quantum mechanics. In this experiment, electrons are fired from a beam gun towards an absorbing screen with two slits punched out. Behind this, another screen sits, registering the impact of the electrons. What was observed is that the recording screen develops interference patterns (see Figure 4.1) that confirms the wave-like nature of particles.

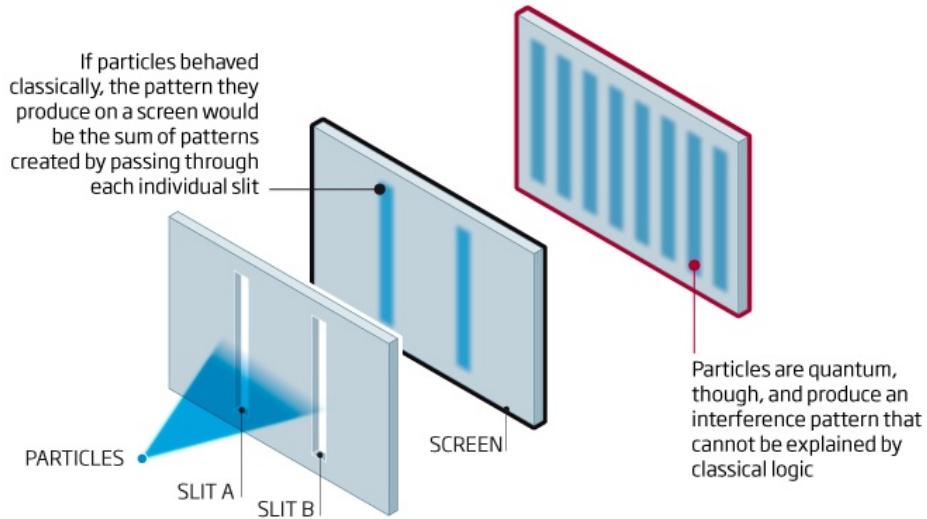


Figure 4.1: Schematic representation of a double slit experiment.

4.1 The wavefunction

Quantum mechanics tells us that light and matter can display *both* wave-like and particle-like properties. The waves in question are unlike any you might have experience of. They are not propagating excitation in a medium or in a pervasive field – as is the case of electromagnetic waves. They are a mathematical construct that we use to describe the dynamics of quantum objects. Importantly, the amplitude of these waves has no physical significance⁶ but, as we will see, its modulus square does: it gives a probability density. For this reason quantum waves are sometimes referred to as probability waves.

The quantum description of reality constitutes a significant departure from the classical one. This is manifest from the very beginning as we can see by comparing how we describe the state of a system within the two frameworks. The state is the information that tells us all we need to know about the system at a fixed time, with the idea that the laws of physics will then dictate how the state evolves at all later times. Throughout these lectures we will deal mainly with the dynamics of a single particle. For the present discussion, we'll think about a particle moving in \mathbb{R}^3 .

In the classical world, the state of the particle at a given time t_0 is determined by its position $\vec{x}(t_0)$ and its velocity $\vec{v}(t_0) = \dot{\vec{x}}(t_0)$. By specifying both bits of information, then the equation of motion $\vec{F} = m\ddot{\vec{x}}$ determines $\vec{x}(t)$ and $\vec{v}(t)$ for all time. Importantly, it's not enough to just know only, say, the position of the particle at $t = t_0$. You need both

⁶This is actually a fact that led many a physicist to reject quantum mechanics as correct description of reality.

$\vec{x}(t_0)$ and $\dot{\vec{x}}(t_0)$. Mathematically, this is because the equation of motion is a second order differential equation and so you need two initial conditions to get a unique solution.

In the quantum world, the state of a particle is determined by its *wavefunction*. This is a complex valued function

$$\begin{array}{ccc} \mathbb{R}^3 \times \mathbb{R} & \longrightarrow & \mathbb{C} \\ \psi(\vec{x}, t) : & \Psi & \Psi \\ (\vec{x}, t) & \longmapsto & \psi(\vec{x}, t) \end{array} .$$

As we will see, if you know the wavefunction at some time, say t_0 , then you have all the information that you need to determine the state at any time. This description is not a small amendment to classical mechanics. We have replaced the three position coordinates $\vec{x} \in \mathbb{R}^3$ with an infinite amount of information: a (complex) functions worth of information. Moreover, we haven't specified anything about the particle's velocity. That information must also be contained, in some manner, in the wavefunction $\psi(\vec{x}, t)$.

Disclaimer

Classically, light is described by waves propagating in the electromagnetic field. There exist a serious danger to conflate these with the probability waves that describe photons quantum mechanically. It is important to keep the two notion distinct. In the quantum world, light is not a propagating continuous electromagnetic wave, but rather a photon: a propagating electromagnetic wave packet. Unlike a wave, which is spread out, it is highly localised and can behave like a particle. Associated to this particle is an intangible probability wave that weights the odds of finding the photon in a certain space region at a certain time. Once a particle's observable, say the position, is measured by whatever device, this quantity will remain determined, losing its random nature. This is why we can observe the tracks of particles as they move through a bubble chamber. See Figure 4.2

4.1.1 The probability interpretation

The wavefunction – more precisely, its mod-square – has a very simple interpretation. It tells us the probability that we will find a particle at a given position. In formulae, the *probability density* P for a particle to sit at point \vec{x} is

$$P(\vec{x}, t) = |\psi(\vec{x}, t)|^2 . \quad (4.1)$$

This is known as *Born rule*, after Max Born who first understood that quantum mechanics is a theory of probability rather than certainty.

Actual probabilities are obtained by integrating P over a volume. The probability that the particle sits, at a time t , in some infinitesimal volume dV centred around point \vec{x} is $P(\vec{x}, t)dV = |\psi(\vec{x}, t)|^2 dV$. The probability of finding a particle in a region $R \in \mathbb{R}^3$ is

$$P_R(t) = \int_R dV P(\vec{x}, t) = \int_R dV |\psi(\vec{x}, t)|^2 . \quad (4.2)$$

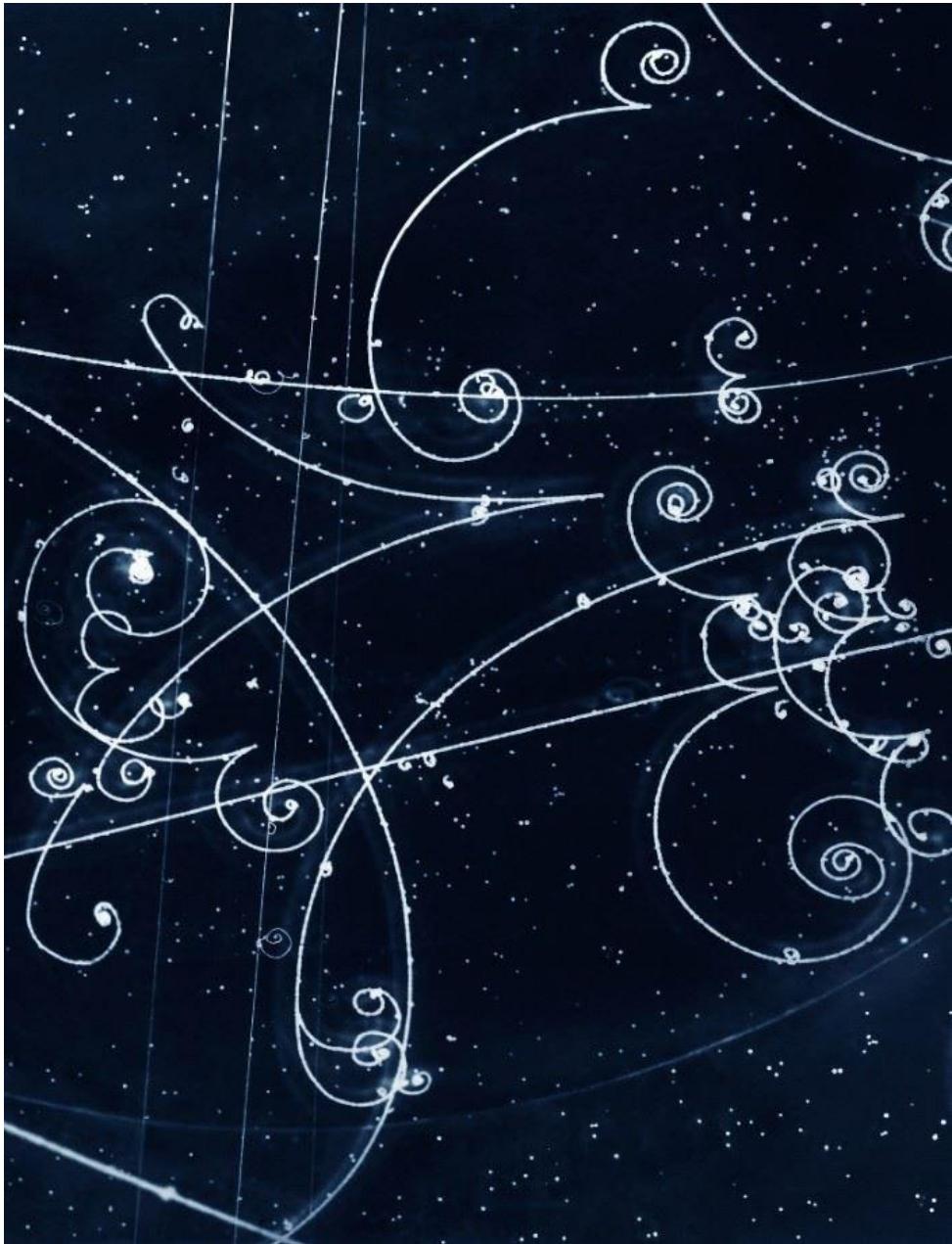


Figure 4.2: Trajectories of electrons and positrons. These tracks show part of a shower of particles, moving from right to left through a hydrogen-neon bubble chamber. The shower was triggered by a collision of a neutrino with a neon nucleus. The tracks spiralling inwards anticlockwise are produced by electrons. Anti-electrons (or positrons) are spiralling inwards clockwise. The paired tracks appearing from a point source are electron-positron pairs formed from high-energy photons in a process known as pair-creation. This photograph was taken in the Big European Bubble Chamber (BEBC) at CERN.

In one dimensions, we can ask for the probability to find a particle in an interval $[a, b]$ and we compute it as

$$P_{[a,b]} = \int_a^b dx P(x, t) = \int_a^b dx |\psi(x, t)|^2 . \quad (4.3)$$

In all other realms of science, probability arises because of our ignorance. If you throw a classical dice, and know its initial state with complete precision, then there is no doubt about what will happen. Once the dice leaves your hand, the probability that you will roll a six is either 0 or 1. However, in most circumstances we don't have a good knowledge of the initial state of the dice and, besides, differential equations are hard to solve. So we just give up, admit that we don't know what will happen, and say that the probability of rolling a six is $1/6$. Crucially, however, the introduction of probability is entirely due to our lack of knowledge; it is not an inherent property of the dice.

This is not the case in quantum mechanics. The state $\psi(\vec{x}, t)$ contains all the information about a particle. There is no additional information available that we might somehow access to and reduce our uncertainty. The probabilistic interpretation is not due to any failing on our part: it emerges from the inherent randomness of the quantum world.

4.1.2 Normalisation

Since we endowed the wavefunction with a probabilistic interpretation, we cannot take $\psi(\vec{x}, t)$ to be any complex function. The particle has to be *somewhere* in \mathbb{R}^3 , which translates into the normalisation condition

$$\int_{\mathbb{R}^3} dV P(\vec{x}, t) = \int_{\mathbb{R}^3} dV |\psi(\vec{x}, t)|^2 = 1 . \quad (4.4)$$

A wavefunction satisfying (4.4) is said to be *normalised*. Now, if we stumble upon a non-normalised function, we have a very simple fix. Suppose $\Psi(\vec{x}, t)$ is such that

$$\int_{\mathbb{R}^3} dV |\Psi(\vec{x}, t)|^2 = N < \infty , \quad (4.5)$$

then we can define a normalised wavefunction simply as

$$\psi(\vec{x}, t) \doteq \frac{1}{\sqrt{N}} \Psi(\vec{x}, t) . \quad (4.6)$$

A function $\Psi(\vec{x}, t)$ satisfying (4.5) with finite N is said to be *normalisable*. It often turns out to be more convenient to work with non-normalised but normalisable functions and remember to include the normalisation factor only at the end when computing probabilities.

Now, it is clear that a function is normalisable only if $\Psi(\vec{x}, t) \rightarrow 0$ sufficiently fast⁷ as $|\vec{x}| \rightarrow \infty$. Clearly then, we have little use for non-normalisable wavefunctions, i.e. such that

$$\int_{\mathbb{R}^3} dV |\Psi(\vec{x}, t)|^2 = \infty . \quad (4.7)$$

These have no possible probabilistic interpretation. While this seems to be an inconsequential statement, it turns out that non-normalisable functions pop up when dealing with free particles in non-compact spaces such as \mathbb{R}^3 . Special care must be taken in these situations.

⁷The mathematical statement is that normalisable wavefunctions should belong to the space $L^2(\mathbb{R}^3)$ of square-integrable functions on the 3-dimensional Euclidean space.

There is one very important fact that some of you might have already noticed. The *phase* of the wave function is totally immaterial as pertains to the probability density. In other words, all representatives of the 1-parameter family of wavefunctions

$$\psi_\alpha(\vec{x}, t) \doteq e^{i\alpha} \psi(\vec{x}, t) , \quad \alpha \in \mathbb{R} , \quad (4.8)$$

describe the same physical state. In fact $|\psi_\alpha(\vec{x}, t)|^2 = |\psi(\vec{x}, t)|^2$ and it will turn out that no other physical observable depends on α . Importantly, this statement holds true only if α is constant. Should we multiply a wavefunction by a spatially-varying phase $\exp(i\alpha(\vec{x}))$, then the probability density would not change, but other observables will.

Insight: states are rays

Taking the normalisation requirement together with the above statement on the constant phase, it is sometime useful to think of a physical state as described by a family of normalisable, complex wavefunctions, subject to an equivalence relation

$$\psi(\vec{x}, t) \simeq \lambda \psi(\vec{x}, t) , \quad \lambda \in \mathbb{C} , \quad \lambda \neq 0 . \quad (4.9)$$

This means that a state correspond to *rays* in the complex space, rather than points. Stated more mathematically, a quantum state is an element of the *projective $L^2(\mathbb{R}^3)$ space of square-integrable functions*. The connotation “projective” refers to the equivalence relation above.

4.1.3 Superposition

As we will see momentarily, the dynamics of the wavefunction is driven by the *Schrödinger equation* which is a linear wave equation. We saw in Part I that solutions to such equations obey the superposition principle. Then, if $\psi_1(\vec{x}, t)$ and $\psi_2(\vec{x}, t)$ are solutions to the Schrödinger equation, so is

$$\psi_3(\vec{x}, t) = \alpha \psi_1(\vec{x}, t) + \beta \psi_2(\vec{x}, t) , \quad \forall \alpha, \beta \in \mathbb{C} . \quad (4.10)$$

Additionally, if $\psi_1(\vec{x}, t)$ and $\psi_2(\vec{x}, t)$ are possible states of the system (meaning they are both normalisable), then so is $\psi_3(\vec{x}, t)$. In fact, let

$$\int_{\mathbb{R}^3} dV |\psi_i(\vec{x}, t)|^2 = N_i < \infty , \quad i = 1, 2 . \quad (4.11)$$

Then

$$\int_{\mathbb{R}^3} dV |\psi_3(\vec{x}, t)|^2 = \int_{\mathbb{R}^3} dV |\alpha \psi_1(\vec{x}, t) + \beta \psi_2(\vec{x}, t)|^2 \leq \int_{\mathbb{R}^3} dV (|\alpha \psi_1(\vec{x}, t)| + |\beta \psi_2(\vec{x}, t)|)^2 ,$$

where we used the triangle inequality $|z_1 + z_2| \leq |z_1| + |z_2|$. We continue

$$\begin{aligned} \int_{\mathbb{R}^3} dV |\psi_3(\vec{x}, t)|^2 &\leq \int_{\mathbb{R}^3} dV (|\alpha \psi_1(\vec{x}, t)|^2 + |\beta \psi_2(\vec{x}, t)|^2 + 2 |\alpha \psi_1(\vec{x}, t)| |\beta \psi_2(\vec{x}, t)|) \leq \\ &\leq \int_{\mathbb{R}^3} dV (2 |\alpha \psi_1(\vec{x}, t)|^2 + 2 |\beta \psi_2(\vec{x}, t)|^2) , \end{aligned}$$

where now we used the inequality $|z_1 - z_2|^2 \geq 0$, that tells us $2|z_1||z_2| \leq |z_1|^2 + |z_2|^2$. Hence we finally arrive at

$$\int_{\mathbb{R}^3} dV |\psi_3(\vec{x}, t)|^2 \leq 2|\alpha|^2 N_1 + 2|\beta|^2 N_2 < \infty, \quad (4.12)$$

showing that ψ_3 is normalisable – and so represents a physical state – if ψ_1 and ψ_2 are.

Let us look back at the double-slit experiment, to see the superposition principle at work. Let ψ_1 be the wavefunction representative of the physical state in which just one of the slits, say the left one, is open. Similarly, let ψ_2 correspond to the state in which only the right slit is open. The superposition principle tells us that when both slits are open the state of the system is represented by the wavefunction $\psi_3 = \psi_1 + \psi_2$. It is important to stress that we are adding the wavefunctions – i.e. the probability amplitudes – rather than the probability densities. Consequently, the probability density for the system with both slits open is

$$|\psi_1(\vec{x}, t) + \psi_2(\vec{x}, t)|^2 = |\psi_1(\vec{x}, t)|^2 + |\psi_2(\vec{x}, t)|^2 + 2\text{Re}\left(\overline{\psi_2(\vec{x}, t)}\psi_1(\vec{x}, t)\right), \quad (4.13)$$

$$\implies P_3(\vec{x}, t) \neq P_1(\vec{x}, t) + P_2(\vec{x}, t),$$

where a bar over a quantity \bullet stands for complex conjugation. In the above relation, the cross term $2\text{Re}\left(\overline{\psi_2(\vec{x}, t)}\psi_1(\vec{x}, t)\right)$ is what causes the interference pattern.

4.2 The Schrödinger equation

Now that we saw how, in quantum mechanics, the state of a system is encoded in the wavefunction, we need to understand the dynamics: according to what rules do the state evolve in time? The answer comes from above, so to speak, in the form of the *Schrödinger equation*

$$\boxed{i\hbar \frac{\partial}{\partial t} \psi(\vec{x}, t) = -\frac{\hbar^2}{2m} \nabla^2 \psi(\vec{x}, t) + V(\vec{x}, t) \psi(\vec{x}, t)}, \quad (4.14)$$

where $V(\vec{x}, t)$ is the *potential energy* that we allow, in full generality, to possibly depend on time, and

$$\hbar \approx 1.05 \times 10^{-34} \text{ J} \cdot \text{s}. \quad (4.15)$$

is the *Planck constant*. Unfortunately, the name of Planck constant is often assigned to the quantity $h = 2\pi\hbar$. We believe the use of h to be unpractical. First of all, using \hbar saves us to write a lot of factors of π , as already evident in the Schrödinger equation (4.14). Moreover, the symbol \hbar is used exclusively for the Planck constant, unlike h that might refer to a height or a magnetic field density. Hereafter, we will exclusively make use of \hbar . Note that \hbar has the dimensions of [energy] \times [time], the same as the classical *angular momentum* $\vec{L} = m\dot{\vec{x}} \times \vec{x}$ and as the classical action $S = \int dt m\dot{\vec{x}}/2 - V$.

While there exist no rigorous derivation of the Schrödinger equation starting from first principles⁸, we can certainly motivate its expression. In fact, the idea is that the wavefunction associated to a particle is a wave. A de Broglie wave in particular. This means

⁸In fact, Schrödinger himself postulated his equation building on intuitions that the experimental results provided.

that it possesses a frequency ω and a wave number \vec{k} , related to the total energy E and to the momentum \vec{p} as follows (we drop momentarily the explicit dependence on \vec{x} and t)

$$E = \hbar\omega, \quad \vec{p} = \hbar\vec{k}. \quad (4.16)$$

Now, suppose we are dealing with a particle of mass m and potential energy $V(\vec{x}, t)$. Then its total energy is

$$E = \frac{|\vec{p}|^2}{2m} + V, \quad (4.17)$$

which, in terms of frequency and wave number reads

$$\hbar\omega = \frac{\hbar^2}{2m} |\vec{k}|^2 + V. \quad (4.18)$$

This is already pretty similar to (4.14). However, what we want is a differential equation – i.e. a wave equation. The trick is to consider the simplest possible wave that we know, the complex plane wave

$$\psi_{\text{p.w.}} = A e^{i(\vec{k}\cdot\vec{x}-\omega t)}, \quad (4.19)$$

and use it to write ω and \vec{k} . The how is easy: we take derivatives

$$\omega = \frac{i}{\psi_{\text{p.w.}}} \frac{\partial}{\partial t} \psi_{\text{p.w.}}, \quad |\vec{k}|^2 = -\frac{1}{\psi_{\text{p.w.}}} \nabla^2 \psi_{\text{p.w.}}. \quad (4.20)$$

Substituting this expression into (4.18), we arrive at

$$i\hbar \frac{\partial}{\partial t} \psi_{\text{p.w.}} = -\frac{\hbar^2}{2m} \nabla^2 \psi_{\text{p.w.}} + V \psi_{\text{p.w.}}. \quad (4.21)$$

The last step is to take this expression, valid for plane waves, and generalise it, assuming it remain valid in the same form for any wavefunction.

A few comments on the Schrödinger equation are called for.

- Unlike the wave equation we studied in Part I, the Schrödinger equation (4.14) is of first order in time. This is the reason why $\psi(\vec{x}, t_0)$ encodes the complete information on the dynamics of a quantum system: we only need one initial condition to solve (4.14).

- The operator

$$\hat{H}(\vec{x}, t) \doteq -\frac{\hbar^2}{2m} \nabla^2 + V(\vec{x}, t),$$

is known as the *Hamiltonian*. Different choices of Hamiltonian describe different laws of physics. In particular, the equation (4.14) is only valid for non-relativistic particles, that is objects whose velocity can be considered small compared to the speed of light. A more generally applicable form of the Schrödinger equation is

$$i\hbar \frac{\partial}{\partial t} \psi(\vec{x}, t) = \hat{H}(\vec{x}, t) \psi(\vec{x}, t). \quad (4.22)$$

You might be familiar with the concept of Hamiltonian from the classical dynamics module. It is no coincidence that the same name is used to refer to \hat{H} . In fact some of you might glimpse a similarity between quantum and Hamiltonian mechanics, in

the fact that in both settings it is a Hamiltonian that generates time translations. This idea is at the core of Heisenberg's approach to quantum mechanics. We will not have the time to delve into it in this module. Just note that, in a similar spirit, the momentum \vec{p} is what generates space translations and this is reflected in the – admittedly puzzling at the moment – correspondence $\vec{p} \rightarrow -i\hbar\vec{\nabla}$.

- Not all classical theories fit in the framework of Hamiltonian mechanics. Roughly speaking, only those that satisfy the conservation of energy can be formulated in this way. Importantly, the same is true of quantum mechanics. In particular, there is no Schrödinger equation for theories with friction. Of course, this is of little importance, since friction forces are not fundamental: they result from interactions between countless, say 10^{23} , atoms. There are no such friction forces in the atomic and subatomic worlds.
- As one proceed to more advanced physical theories, such as relativistic quantum field theory or even string theory, the Schrödinger equation (4.22) remains the correct description for the evolution of the quantum state. All that changes is the Hamiltonian \hat{H} , which is replaced by something ever more complicated, as too are the wavefunctions on which the Hamiltonian acts.

4.2.1 Conservation of the probability

In §4.1 we insisted on the probabilistic interpretation of the wavefunction which, in particular, forces us to consider normalised wavefunctions. We must make sure that the Schrödinger equation does not affect the normalisation of our wavefunction. Let us look then at how the probability density changes in time

$$\frac{\partial}{\partial t}P(\vec{x},t) = \frac{\partial}{\partial t}|\psi(\vec{x},t)|^2 = \overline{\psi(\vec{x},t)}\frac{\partial}{\partial t}\psi(\vec{x},t) + \psi(\vec{x},t)\frac{\partial}{\partial t}\overline{\psi(\vec{x},t)}. \quad (4.23)$$

From Schrödinger equation we find

$$\begin{aligned} \frac{\partial}{\partial t}\psi(\vec{x},t) &= -\frac{i}{\hbar}\left(-\frac{\hbar^2}{2m}\nabla^2\psi(\vec{x},t) + V(\vec{x},t)\psi(\vec{x},t)\right), \\ \frac{\partial}{\partial t}\overline{\psi(\vec{x},t)} &= \frac{i}{\hbar}\left(-\frac{\hbar^2}{2m}\nabla^2\overline{\psi(\vec{x},t)} + V(\vec{x},t)\overline{\psi(\vec{x},t)}\right). \end{aligned} \quad (4.24)$$

Combining these, we see that the term with the potential cancels out and we find

$$\begin{aligned} \frac{\partial}{\partial t}P(\vec{x},t) &= \frac{i\hbar}{2m}\left(\overline{\psi(\vec{x},t)}\nabla^2\psi(\vec{x},t) - \psi(\vec{x},t)\nabla^2\overline{\psi(\vec{x},t)}\right) = \\ &= \frac{i\hbar}{2m}\vec{\nabla} \cdot \left(\overline{\psi(\vec{x},t)}\vec{\nabla}\psi(\vec{x},t) - \psi(\vec{x},t)\vec{\nabla}\overline{\psi(\vec{x},t)}\right). \end{aligned} \quad (4.25)$$

We write this relation as

$$\frac{\partial}{\partial t}P(\vec{x},t) + \vec{\nabla} \cdot \vec{J}(\vec{x},t) = 0, \quad (4.26)$$

where we introduced the *probability current*

$$\vec{J}(\vec{x},t) = -\frac{i\hbar}{2m}\left(\overline{\psi(\vec{x},t)}\vec{\nabla}\psi(\vec{x},t) - \psi(\vec{x},t)\vec{\nabla}\overline{\psi(\vec{x},t)}\right). \quad (4.27)$$

Thus we got ourselves a *continuity equation*, just like the one we derived in Part I for the energy density of waves. Here we proceed in the same fashion by computing the probability $P_R(t)$ of finding the particle in some region $R \subset \mathbb{R}^3$

$$P_R(t) = \int_R dV P(\vec{x}, t) . \quad (4.28)$$

Thanks to the continuity equation, we find that

$$\frac{d}{dt} P_R(t) = - \int_R dV \vec{\nabla} \vec{J}(\vec{x}, t) = - \int_{\partial R} d\vec{S} \cdot \vec{J}(\vec{x}, t) , \quad (4.29)$$

where we used the divergence theorem. We see that that the probability that the particle lies in R can change, but only if there is a flow of probability through the surface ∂R that bounds R . If we know for sure that $\vec{J} = 0$ everywhere on the surface ∂R – or if the region R has no boundary – then the probability that the particle is in the region R is time-independent. A continuity equation does not just implies that some quantity is conserved: it tells us that it is conserved locally. In the present context, this quantity is probability. The probability density cannot vanish in one region of space only to reappear in some far flung region of the universe: the evolution of the probability density is local. If it changes in some region of space then it is because it has moved into a neighbouring region. The current \vec{J} details how it does this.

If we consider $R = \mathbb{R}^3$, then intuitively we should get $\partial/\partial t P_{\mathbb{R}^3} = 0$. This is because the probability to find a particle anywhere in the Universe is always 1. In fact

$$\frac{\partial}{\partial t} P_{\mathbb{R}^3} = - \int_{S_\infty^2} d\vec{S} \cdot \vec{J}(\vec{x}, t) , \quad (4.30)$$

where S_∞^2 is the 2-sphere with infinite radius bounding asymptotically \mathbb{R}^3 . Any reasonable wavefunction decays at infinity: $\psi \rightarrow 0$ as $|\vec{x}| \rightarrow \infty$. This implies that the probability current also vanishes at infinity, which in turns means that

$$\frac{\partial}{\partial t} P_{\mathbb{R}^3} = 0 . \quad (4.31)$$

4.2.2 Observables and expectation values

Note: in this subsection we drop the explicit dependence on time.

In the classical world, the state of a particle is described by its position \vec{x} and velocity $\dot{\vec{x}}$. Equivalently, but more practically, we can choose to describe the state by the position \vec{x} and the momentum $\vec{p} = m\dot{\vec{x}}$, as in Hamiltonian mechanics. Given such a state, we can compute any function of \vec{x} and \vec{p} . These are known as *observables* and constitute the physical properties of the system. Examples are the position and momentum themselves, the energy $E(\vec{x}, \vec{p}) = |\vec{p}|^2/(2m) + V(\vec{x})$ and the angular momentum $\vec{L}(\vec{x}, \vec{p}) = \vec{x} \times \vec{p}$. Once you know the state (\vec{x}, \vec{p}) it is then simple to compute any observable $f(\vec{x}, \vec{p})$.

As we saw, in the quantum world, things work differently. The state of a particle is encoded in a wavefunction $\psi(\vec{x})$ which gives us a probability density $P(\vec{x}) = |\psi(\vec{x})|^2$. How should we think of observables in this context? Let us first look at the position \vec{x} .

We know that $P(\vec{x})dV$ gives us the probability that the particle under scrutiny happens to be in an infinitesimal neighbourhood of \vec{x} . We do not have certainty, so we cannot really speak of a value of the position. In the probabilistic setting we find ourselves in, the correct thing to ask is: what is the **mean value** of the position? This quantity, that we denote as $\langle \vec{x} \rangle$ is the sum – actually, integral – over all the values of \vec{x} , weighted by the probability density:

$$\langle \vec{x} \rangle \doteq \int_{\mathbb{R}^3} dV \vec{x} |\psi(\vec{x})|^2 . \quad (4.32)$$

So in this case, we can still think of \vec{x} as an observable and we saw that we can measure its *expectation value* – the quantum mechanics name for mean value – $\langle \vec{x} \rangle$ as above.

Now, let us look at the momentum. Following what we saw for \vec{x} , we might think of determining its expectation value as

$$\langle \vec{p} \rangle \stackrel{?}{=} \int_{\mathbb{R}^3} dV \vec{p} |\psi(\vec{x})|^2 . \quad (4.33)$$

But how do we interpret this expression? Does \vec{p} depend on \vec{x} as a function? We must determine this \vec{x} dependence in order to perform the integral. Let us remember the trick we used to justify the Schrödinger equation (remember the de Broglie relation $\vec{p} = \hbar \vec{k}$)

$$\psi_{\text{p.w.}}(\vec{x}) = A e^{i(\frac{1}{\hbar} \vec{p} \cdot \vec{x} - \omega t)} \implies \vec{p} \psi_{\text{p.w.}}(\vec{x}) = -i\hbar \vec{\nabla} \psi_{\text{p.w.}}(\vec{x}) . \quad (4.34)$$

As we did then, we suppose this relation hold for generic wavefunctions

$$\hat{\vec{p}} \psi(\vec{x}) \doteq -i\hbar \vec{\nabla} \psi(\vec{x}) , \quad (4.35)$$

where we changed notation to $\hat{\vec{p}}$ to stress the fact that we are not thinking of the momentum as a normal vector anymore, but rather as an *operator*. Now we can return to the computation of the expectation value $\langle \vec{p} \rangle$ and write it as

$$\langle \hat{\vec{p}} \rangle \doteq \int_{\mathbb{R}^3} dV \overline{\psi(\vec{x})} \hat{\vec{p}} \psi(\vec{x}) \equiv -i\hbar \int_{\mathbb{R}^3} dV \overline{\psi(\vec{x})} \vec{\nabla} \psi(\vec{x}) . \quad (4.36)$$

In the case of the momentum, then, we cannot think of \vec{p} as an observable in the classical sense anymore. We need to promote the momentum vector to a *momentum operator* and consider the latter as our quantum observable. This is a very important point that underpins the whole building of quantum mechanics

In quantum mechanics, observables are operators acting on wavefunctions.

The expectation value of an observable can be computed as in the momentum case

$$\langle \hat{O} \rangle \doteq \int_{\mathbb{R}^3} dV \overline{\psi(\vec{x})} \hat{O} \cdot \psi(\vec{x}) . \quad (4.37)$$

How do we find the operators associated to a given classical observable? This what is usually referred to with the term *quantisation* and is an ill-defined procedure⁹. The

⁹There is a simple way to understand why. Quantum mechanics is more general than classical theory. In particular, the latter is contained in the former and emerges when the Planck constant \hbar is very small compared to all other physical quantities. Classical mechanics contains less information than quantum mechanics. Thus the quantisation procedure require us to supply additional information to our theory.

standard way is to promote classical observables to operators by following the recipe we established for the momentum operator. In other words

$$O(\vec{x}, \vec{p}) \xrightarrow{\text{quant.}} \hat{O} \doteq O(\hat{\vec{x}}, \hat{\vec{p}}) . \quad (4.38)$$

So, for example, the energy $E(\vec{x}, \vec{p})$ becomes the Hamiltonian operator we saw appearing in the Schrödinger equation

$$E(\vec{x}, \vec{p}) \xrightarrow{\text{quant.}} \hat{H} = \frac{|\hat{\vec{p}}|^2}{2m} + V(\hat{x}) , \quad \hat{H}\psi(\vec{x}) = \left[-\frac{\hbar^2}{2m}\nabla^2 + V(\vec{x}) \right] \psi(\vec{x}) . \quad (4.39)$$

We are now facing a problem: operators do not necessarily commute! For example

$$\hat{\vec{p}} \cdot \hat{\vec{x}} \psi(\vec{x}) = -i\hbar \vec{\nabla} \cdot (\vec{x}\psi(\vec{x})) = -i\hbar [\psi(\vec{x}) \vec{\nabla} \cdot \vec{x} + \vec{x} \cdot \vec{\nabla} \psi(\vec{x})] \neq \hat{\vec{x}} \hat{\vec{p}} \psi(\vec{x}) . \quad (4.40)$$

The physical meaning of this is that, in the quantum world, the order in which we perform successive measurements matters. This is an issue if we want to quantise observables containing products of positions and momenta, such as the operator $x_1 p_1$. We can choose two different orderings, meaning we have at least 2 quantum operators we can associate to $x_1 p_1$! How do we choose? There are various approaches: place all momenta on the right, choose a symmetric ordering, etc... We will not delve in this matter here: you are going to meet this fact again in your future courses on quantum mechanics.

4.2.3 The Heisenberg uncertainty principle

We can do more than asking for the expectation value of an observable. For example we might wonder how much the probability distribution of the observable is spread around its mean value. This quantity is the *variance – uncertainty* in quantum mechanics parlance – and is defined as the average of the square of the deviation from the mean value

$$(\Delta O)^2 \doteq \left\langle \left(\hat{O} - \langle \hat{O} \rangle \right)^2 \right\rangle = \left\langle \hat{O}^2 - 2\hat{O}\langle \hat{O} \rangle + \langle \hat{O} \rangle^2 \right\rangle = \langle \hat{O}^2 \rangle - \langle \hat{O} \rangle^2 , \quad (4.41)$$

where we used the fact that $\langle \hat{O} \rangle$ is a number and thus $\langle \langle \hat{O} \rangle \hat{A} \rangle = \langle \hat{O} \rangle \langle \hat{A} \rangle$.

Now, let us reduce to 1 dimension for simplicity: the generalisation to higher dimensions is straightforward. One of the most famous facts of quantum mechanics is encapsulated into the Heisenberg uncertainty principle

Heisenberg Uncertainty Principle

There is a limit to the precision with which pairs of physical properties, e.g. position and momentum, can be simultaneously known. In other words, the more accurately one property is measured, the less accurately the other property can be known. For position and momentum, this principle is expressed by the formula

$$\boxed{\Delta x \Delta p \geq \frac{\hbar}{2}} . \quad (4.42)$$

Let us prove this.

Proof

First, let us assume for simplicity that $\langle \hat{x} \rangle = \langle \hat{p} \rangle = 0$. If this was not the case, we can always redefine $\hat{x} \rightarrow \hat{x} - \langle \hat{x} \rangle$ and $\hat{p} \rightarrow \hat{p} - \langle \hat{p} \rangle$ to ensure the vanishing of the mean values. Now, consider the 1-parameter family of states

$$\Psi_s(x) = (\hat{p} - is\hat{x})\psi(x) , \quad (4.43)$$

for $s \in \mathbb{R}$ and some reference state $\psi(x)$ on which we want to compute the uncertainties Δx and Δp . $\Psi_s(x)$ are bona fide states, meaning they are positive definite

$$\int_{\mathbb{R}} dx |\Psi_s(x)|^2 \geq 0 . \quad (4.44)$$

Let us develop this inequality

$$\begin{aligned} 0 &\leq \int_{\mathbb{R}} dx \overline{(\hat{p} - is\hat{x})\psi(x)}(\hat{p} - is\hat{x})\psi(x) = \\ &= \int_{\mathbb{R}} dx \overline{(-i\hbar\psi'(x) - is\hat{x}\psi(x))}(-i\hbar\psi'(x) - is\hat{x}\psi(x)) = \\ &= \int_{\mathbb{R}} dx (\hbar^2 \overline{\psi'(x)}\psi'(x) + \hbar s x \overline{\psi(x)}\psi'(x) + \hbar s x \overline{\psi'(x)}\psi(x) + s^2 x^2 |\psi(x)|^2) = \quad (4.45) \\ &= \int_{\mathbb{R}} dx (-\hbar^2 \overline{\psi(x)}\psi''(x) - \hbar s |\psi(x)|^2 + s^2 x^2 |\psi(x)|^2) = \\ &= \langle \hat{p}^2 \rangle + s^2 \langle \hat{x}^2 \rangle - \hbar s \equiv s^2(\Delta x)^2 - s\hbar + (\Delta p)^2 . \end{aligned}$$

We have integrated by parts the first and third term in going from line 3 to line 4, supposed that $\langle 1 \rangle \equiv \int_{\mathbb{R}} dx |\psi(x)|^2 = 1$ and remembered that $\langle \hat{x} \rangle = \langle \hat{p} \rangle = 0$. Now we have an inequality that must hold for any $s \in \mathbb{R}$

$$s^2(\Delta x)^2 - s\hbar + (\Delta p)^2 \geq 0 , \quad \forall s \in \mathbb{R} . \quad (4.46)$$

This can only be true if the left-hand side has one or zero roots, which means that the discriminant should be non-positive

$$\hbar^2 - 4(\Delta x)^2(\Delta p)^2 \leq 0 \implies \Delta x \Delta p \geq \frac{\hbar}{2} . \quad \blacksquare \quad (4.47)$$

If you spend a moment to think about what it means to measure the position of a particle, the Heisenberg principle start to sound less outlandish. To determine where a particle stands one can simply look at it. But to resolve something on a distance scale Δx , one needs light of wavelength $\lambda \leq \Delta x$ and light is made of photons with momentum given by the de Broglie formula $p = 2\pi\hbar/\lambda$. Hence to measure the position of a particle to a small uncertainty Δx one has to bombard it with high momentum photons that will transfer some momentum of order $\Delta p \sim 2\pi\hbar/\Delta x$. The act of measuring is thus inherently intrusive: we cannot know both the position and momentum to better than $\Delta x \Delta p \sim 2\pi\hbar$. The Heisenberg uncertainty relation (4.42) provides the more accurate bound.

Example: the gaussian wave-packet

Still unconvinced? I don't blame you. Let us work out an example to cement the understanding.

Consider the following normalised gaussian state

$$\psi(x) = \left(\frac{a}{\pi}\right)^{1/4} e^{-ax^2/2}. \quad (4.48)$$

You can verify that $\int_{\mathbb{R}} dx |\psi(x)|^2 = 1$. First, let us compute the mean values of position and momentum

$$\begin{aligned} \langle \hat{x} \rangle &= \sqrt{\frac{a}{\pi}} \int_{\mathbb{R}} dx x e^{-ax^2} = 0, \\ \langle \hat{p} \rangle &= -i\hbar \sqrt{\frac{a}{\pi}} \int_{\mathbb{R}} dx e^{-ax^2/2} \frac{d}{dx} e^{-ax^2/2} = 0. \end{aligned} \quad (4.49)$$

They both vanish due to the integrands being odd. What about the uncertainties? Again, we can easily compute them

$$\begin{aligned} \langle \hat{x}^2 \rangle &= \sqrt{\frac{a}{\pi}} \int_{\mathbb{R}} dx x^2 e^{-ax^2} = -\sqrt{\frac{a}{\pi}} \frac{\partial}{\partial a} \int_{\mathbb{R}} dx e^{-ax^2} = -\sqrt{\frac{a}{\pi}} \frac{\partial}{\partial a} \sqrt{\frac{a}{\pi}} = \frac{1}{2a}, \\ \langle \hat{p}^2 \rangle &= -\hbar^2 \sqrt{\frac{a}{\pi}} \int_{\mathbb{R}} dx e^{-ax^2/2} \frac{d^2}{dx^2} e^{-ax^2/2} = \hbar^2 \sqrt{\frac{a^3}{\pi}} \int_{\mathbb{R}} dx (1 - ax^2) e^{-ax^2} = a \frac{\hbar^2}{2}. \end{aligned}$$

We then see immediately that

$$\Delta x \Delta p = \sqrt{\langle \hat{x}^2 \rangle - \langle \hat{x} \rangle^2} \sqrt{\langle \hat{p}^2 \rangle - \langle \hat{p} \rangle^2} = \frac{\hbar}{2}. \quad (4.50)$$

This simple exercise tells us something interesting: the gaussian wave-packet saturates the Heisenberg bound. This is not the case for any old wavefunction (you are encouraged to try some examples yourselves). The class of Gaussian wavefunctions, parameterised by a , does the best job possible of balancing the competing requirements of localising both position and momentum.

4.2.4 The collapse of the wavefunction

The Schrödinger equation (4.22) governs the evolution of the wavefunction in all but one important instance. The latter is when we perform a measurement on the system. For example, suppose that, at some time t_0 , we have a particle in a superposition of two different positions \vec{X}_1 and \vec{X}_2 and that these are well-separated $|\vec{X}_1 - \vec{X}_2| \gg 1$. The wavefunction might look something like

$$\psi(\vec{x}, t_0) = \frac{1}{\sqrt{N}} \left(e^{-a|\vec{x} - \vec{X}_1|^2} + e^{-a|\vec{x} - \vec{X}_2|^2} \right). \quad (4.51)$$

If we leave the system be, it will evolve according to the Schrödinger equation. However, let us place a measuring apparatus near \vec{X}_1 to see if it captures a signal of the particle. Supposing it does, then we would know for sure that the particle sits near \vec{X}_1 and cannot be anywhere else. This means that the probability of the particle being in the neighbourhood of \vec{X}_2 must vanish and so must the support of the wavefunction! So, at a time t_1 immediately after the positive measurement, the wave function will be

$$\psi(\vec{x}, t_0) = \frac{1}{\sqrt{N'}} e^{-a|\vec{x} - \vec{X}_1|^2}. \quad (4.52)$$

This phenomenon is known as *collapse of the wavefunction* and it may sound quite disconcerting. We see that we can indeed gain knowledge on our system by performing measurements. However this will inevitably modify the wavefunction. In other words, the mere act of measuring perturbs the system, sometimes rather violently. In all effect, after the measure, the wavefunction has lost the information it contained previously on the neighbourhood of \vec{X}_2 . So, by performing a measurement, we gained a piece of information and we lost some other piece.

Another disconcerting fact about the collapse of the wave function is that it does not happen locally: the probability density simply vanish from one region of space and reemerge in the region where we detected the particle. This should be enough to make anyone who knows anything about special relativity break out in a cold sweat, as it seem to imply faster-than-light transmission of information. It turns out that this is not the case, but, sadly, we will not have the time to get to that in this module¹⁰.

4.3 Back to one dimension

While there is much to say about the postulates of quantum mechanics and their interpretation, we are going to set them on a side and dirty our hands on a few practical examples of the Schrödinger equation and its solutions. This will hopefully help you build some intuition for how to think about the wavefunction and the various pieces of information that it encodes. As always, to expose the core features, we will focus on the simplest settings: a single quantum particle living in one spatial dimension. Thus the wavefunction $\psi(x, t)$ will only depend on two real variables x and t and the Schrödinger equation (4.14) simplifies to

$$i\hbar \frac{\partial}{\partial t} \psi(x, t) = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \psi(x, t) + V(x) \psi(x, t). \quad (4.53)$$

We will solve this equation for various choices of $V(x)$ which, henceforth, will be assumed time-independent.

4.3.1 The time-independent Schrödinger equation

Since we assumed the potential $V(x)$ to be *static* – i.e. time-independent – we can deal with (4.53) with the separation of variables approach. Let

$$\psi(x, t) = u(x)T(t). \quad (4.54)$$

¹⁰Curious? You might want to peruse §5 of D. Tong's lecture notes on [Topics in Quantum Mechanics](#). Look for Bell's inequalities in particular.

Plugging this ansatz in (4.53) – and dividing both sides by $\psi(x, t)$ – we find

$$\frac{i\hbar}{T(t)} \frac{d}{dt} T(t) = \frac{1}{u(x)} \left(-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} u(x) + V(x)u(x) \right). \quad (4.55)$$

As usual, since each side depends on a different variable, we equate both to a constant that we denote by E , obtaining two ODEs. The equation for $T(t)$ is

$$\frac{d}{dt} T(t) = \frac{E}{i\hbar} T(t) \implies \frac{dT}{T} = -i\frac{E}{\hbar} dt. \quad (4.56)$$

We can directly integrate this equation, obtaining

$$T(t) = A e^{-i\frac{E}{\hbar}t}, \quad (4.57)$$

where $A = T(0)$ is an integration constant. Hence the wavefunction takes the following form

$$\boxed{\psi(x, t) = e^{-i\frac{E}{\hbar}t} u(x)}, \quad (4.58)$$

where we absorbed the constant A into the still-to-be-determined function $u(x)$, which is a solution of the *time-independent Schrödinger equation*

$$\boxed{-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} u(x) + V(x)u(x) = Eu(x)}. \quad (4.59)$$

There exist some abuse of nomenclature in the literature, since both (4.53) and (4.59) are called Schrödinger equation and both $\psi(x, t)$ and $u(x)$ are *the wavefunction*. Hopefully the context will make clear what we are dealing with.

Let us rewrite (4.59) as

$$\hat{H}(x)u(x) = Eu(x), \quad (4.60)$$

where $\hat{H}(x) = -\hbar^2/(2m)d^2/dx^2 + V(x)$ is the Hamiltonian operator introduced above. In this form, the Schrödinger equations resembles very much the eigenvalue equations you met when working with matrices. This is a very good analogy to keep in mind, so much so that equations such as (4.60) are usually called *eigenvalue problems*, even if the objects involved are not matrices and vectors – not in the sense you are familiar with at least. Just like it happens in the case of matrices, in most cases the equation (4.60) only admits solutions for very specific values of E . The notation was also chosen with hindsight, since these “eigenvalues” will have the interpretation of the possible energies of the system.

The states associated to separable solutions of the form (4.58) are sometimes referred to as *stationary states* and sometimes as *energy eigenstates*. In the same fashion as what we saw when studying waves, these solutions play a special role in that they simplify greatly the analysis of (4.53) and general solutions can be expressed as linear combinations of them, thanks to the superposition principle¹¹.

¹¹This, together with a completeness theorem guaranteeing that any “reasonable” function can be reconstructed with a Fourier-type series.

4.3.2 A particle on a circle

Perhaps surprisingly, our first case study is not the particle on an infinite line, as it turns out that it requires some specific, additional care. Instead we first focus on a particle on a compact space: either a circle S^1 or an interval. Here we look at the former.

Since there is no potential, the equation (4.59) simplifies to

$$-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} u(x) = Eu(x), \quad (4.61)$$

which we can immediately solve

$$u(x) = Ae^{ikx}, \quad (4.62)$$

where we introduced the wave number k , related to the energy by

$$E = \frac{\hbar^2}{2m} k^2. \quad (4.63)$$

Note that, since there is no potential energy, classically $E_{\text{cl}} = p^2/(2m)$, agreeing with the de Broglie relation $p_{\text{cl}} = \hbar k$.

Now, since our particle lives on a circle, we must impose periodicity conditions on our wavefunction. Let R be the radius of S^1 . Then we must ask

$$u(x + 2\pi R) = u(x) \implies k = \frac{n}{R}, \quad n \in \mathbb{Z}. \quad (4.64)$$

This is an example of a *quantisation condition*. We saw other similar cases when studying standing waves, however here the physical implications are more startling: physical quantities that are classically continuous become discrete in the quantum world. For a particle on a circle, both the momentum and the energy can only take the discrete values

$$p_n = \frac{\hbar}{R}n, \quad E_n = \frac{\hbar^2}{2mR^2}n^2, \quad n \in \mathbb{Z}, \quad (4.65)$$

The collection of all possible energies is known as the *spectrum of the Hamiltonian*. The state $u_0(x)$ associated to $n = 0$ is called *ground state*, while the other ones are named *excited states*.

Insight: going classical

Quantum mechanics is a theory that completely replaces classical physics. As such, the former must contain the latter: we'd better be able to recover classical physics expressions and expectations from quantum formulae. Practically we can achieve this by taking the $\hbar \rightarrow 0$ limit. This is known as *classical limit*. Of course, since \hbar is a universal physical constant, $\hbar \rightarrow 0$ does not make any sense! In truth, what we are doing when taking the classical limit is we are assuming that \hbar is very small compared to the system's scales. In the present example, these are the radius R and the mass m . So when, say, $mR^2 \gg \hbar^2$, we have that the relative gap between energy levels $(E_{n+1} - E_n)/E_n = (2n - 1)n^2 \hbar^2/(2mR^2)$ becomes very small and E_n will appear continuous. This is also true if n is very large. Stated differently, for very massive or energetic particles or very big circles, we are not going to perceive the discrete nature of energy levels.

As the circle gets larger, the gap between the allowed momentum and energy allowed values becomes smaller. For R large enough, it will be experimentally impossible to resolve these gaps and it will appear as if momentum and energy can take arbitrary real values. As an extreme example, we are unsure whether our Universe is infinite or else. You might think of deciding between the two by measuring the gap between allowed momenta (or energies) of free particles. However, if the Universe is finite it must have a radius $R \gtrsim 46 \times 10^9$ light years (about 4.4×10^{26} m). Needless to say, we have no way to measure a momentum gap of the order of $\hbar/R \lesssim 0.24 \times 10^{-60}$ kg m/s.

Conversely, we might use these gaps to detect the existence of additional spatial dimensions beyond the three we can experience. In fact, one of the reasons we still have not perceived these might be because they are wrapped to form a small circle. In order to “feel” their presence, you’d need to excite a probe particle enough so that they jump from one state to the next. This means provide them with an energy $E \sim 3\hbar^2/(2mR^2)$. The LHC accelerator still has not seen these extra dimensions and it can reach values $Em \sim 10^{-32}$ J kg, which put an upper limit to their radius at about¹² $R \lesssim 10^{-18}$.

The last step in our analysis is to ensure the wavefunction is correctly normalised. We easily compute

$$\int_0^{2\pi R} dx |u(x)|^2 = 2\pi R |A|^2 = 1 , \quad (4.66)$$

from which we can fix $A = 1/\sqrt{2\pi R}$. The correctly normalised eigenfunctions then are

$$u_n(x) = \frac{e^{ikx}}{\sqrt{2\pi R}} , \quad n \in \mathbb{Z} . \quad (4.67)$$

You can see that if we try to “uncompactify” the circle to recover the real line by sending $R \rightarrow \infty$, the eigenfunctions vanish identically! This is a signal that on the real line, eigenfunctions of free particles are not normalisable.

4.3.3 A particle in a box

Let us consider a particle confined in an interval $x \in (0, L)$. We can achieve this with an infinite well potential

$$V(x) = \begin{cases} 0 , & 0 < x < L \\ \infty , & \text{otherwise} \end{cases} . \quad (4.68)$$

We are really still dealing with a free particle, even though we introduced a potential. In fact the Schrödinger equation splits in two

$$V(x) = \begin{cases} -\frac{\hbar^2}{2m} u''(x) = Eu(x) , & 0 < x < L \\ \lim_{V \rightarrow \infty} \left[-\frac{\hbar^2}{2m} u''(x) + (V - E)u(x) \right] = 0 , & \text{otherwise} \end{cases} . \quad (4.69)$$

Since we are only looking for states with finite energy, outside the interval $(0, L)$ we must ask the wavefunction to vanish identically

$$u(x) = \begin{cases} u(x) , & 0 < x < L \\ 0 , & \text{otherwise} \end{cases} . \quad (4.70)$$

¹²To be precise, one will need to account for the fact that at these energy, relativistic quantum mechanics – or quantum field theory – is needed. However the conclusions remain qualitatively the same.

We already solved the free equation. In this case, we consider solutions of the form

$$u(x) = Ae^{ikx} + Be^{-ikx}, \quad k = \frac{\sqrt{2mE}}{\hbar} > 0, \quad x \in (0, L). \quad (4.71)$$

Now we simply impose that at the boundary the wavefunction vanishes. This should be familiar to you, as it is the same thing we did when studying waves on a Dirichlet-Dirichlet string. The final solutions are

$$u_n(x) = \sqrt{\frac{2}{L}} \sin\left(\frac{\pi}{L}nx\right), \quad n \in \mathbb{N}, \quad (4.72)$$

Where we fixed the constant A by requiring unit normalisation of the eigenfunctions.

There is one subtle difference between the present case and the previous one. For a particle in a box, n should be a positive integer, since $u_{-n}(x) = -u_n(x) \simeq u_n(x)$ (since signs do not matter for a state). While the expression for the energy

$$E_n = \frac{\hbar^2\pi^2}{2mL^2}n^2, \quad n \in \mathbb{N}, \quad (4.73)$$

is the same as for a particle on a circle of radius $R = L/\pi$, the momentum is not a well-defined quantity! In fact our eigenfunctions are a superposition of two states with defined momenta, e^{ikx} and e^{-ikx} .

There is an important feature of (4.73). The lowest energy level – corresponding to the ground state – is non-zero¹³. This contrasts with the case of a particle on a circle and with our classical experience.

This non-zero ground-state energy is sometimes referred to as *zero-point energy*. We can understand its existence in a simple way. For $E = 0$ and zero potential, the Schrödinger equation tells us that $u''(x) = 0$. Pairing this with the boundary conditions $u(0) = u(L) = 0$, we see that the graph of $u(x)$ should be a curve going from 0 to 0 with an everywhere vanishing curvature. The only function that achieves this is the trivial one $u(x) = 0$!

Contrast this with the particle on a circle, where the boundary condition is $u(0) = u(2\pi R)$. In this setting, we can take any constant function $u(x) = A$ to satisfy both the $E = 0$ Schrödinger equation and the boundary conditions.

So, any time we have a Dirichlet condition on some boundary, any non-vanishing wavefunction must change value from zero at a boundary point, to a non-zero value in the bulk, back to zero at another boundary point, which necessarily means that the energy must be non-zero. There are practical consequences. For example, a physical requirement for the wavefunction of an electron in a hydrogen atom is that it has to vanish at spatial infinity. Thus we immediately conclude that the electron must have a zero-point energy.

The discreteness of energy levels in the infinite well has an important application, similar in spirit to the “extra dimension” story that we told above, but significantly more practical. Consider particles moving, as particles do, in three spatial dimensions. Suppose that you trap them in a well in one dimension, but still allow them to wander in the other two. Then, provided that you can restrict their energies to be small enough, the particles will

¹³Choosing $n = 0$ gives us $u_0(x) = 0$, meaning there is no particle in the box!

act, to all intents and purposes, as if they're really two-dimensional objects. Note that this is not an approximate statement, as it would be in classical mechanics: if the particle does not have enough energy to jump to the $n = 2$ state then it really should be thought of as a 2d particle. This may not seem like a big deal at this stage but, as you might learn later in your curriculum, interesting things can happen in low dimensions that are impossible in our 3d world¹⁴. These are not mere mathematical curiosities: they can be constructed in the lab using the method above.

Before moving on, let us spend a few moments to look at the probability densities of the stationary states (4.72)

$$|u_n(x)|^2 = \frac{2}{L} \left[\sin\left(\frac{\pi}{L}nx\right) \right]^2, \quad n \in \mathbb{N}. \quad (4.74)$$

The first few look as in Figure 4.3. You can notice that, as n – and thus the energy –

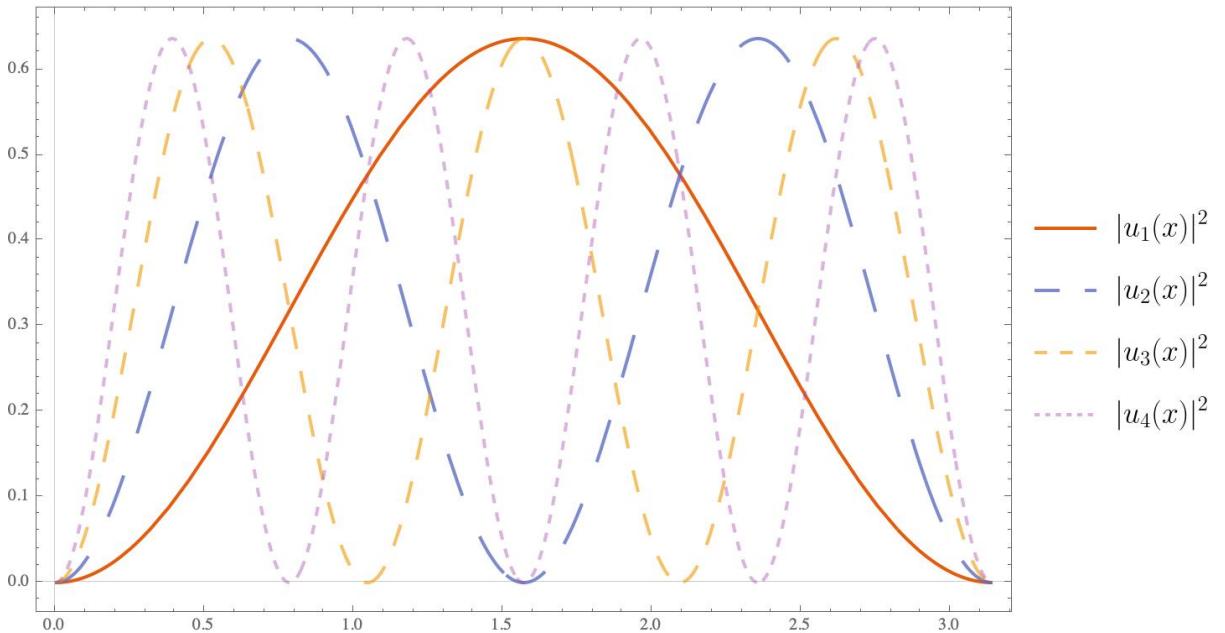


Figure 4.3: Probability densities for a particle in a box of length $L = \pi$.

increases, the particle is more and more likely to be found at n separate points, where $|u_n(x)|^2$ exhibit its maxima. Now, let us look at the classical limit, which is achieved when the combination m/\hbar^2 is very large. If we care to look at states with fixed energy in a box with fixed length L , then we should tune n to also be come very large in concert, so that the product $\hbar^2 n^2 / m$ remains roughly constant. Then we see that the oscillations of the probability density become more and more dense in the interval $(0, L)$, until we will effectively reach a uniform probability density.

4.4 Bounded and bounced particles

We have studied an example of a potential that grows as $x \rightarrow \pm\infty$. A rather extreme one, truth to be told, which is simply infinite outside an interval. What we saw is that

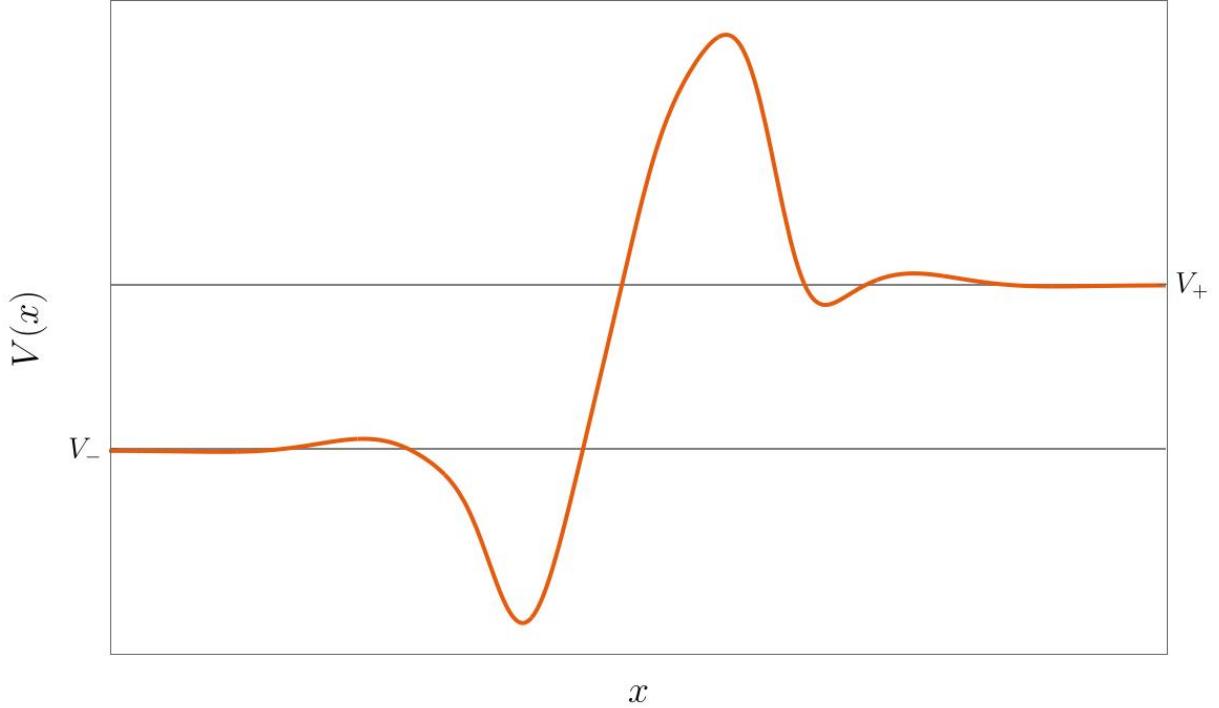
¹⁴If you want an example, in 2d – but not in 1d or 3d — it's possible for an electron to split into N objects each carrying fractional electric charge $1/N$. This is the renowned *quantum Hall effect* and can be realised experimentally!

in such a potential, a particle can only have a discrete set of energies. Later on we will deal with another canonical example of a growing potential, the harmonic oscillator $V(x) = \mu/2x^2$. Spoiling the surprise, this system also only admits a discrete set of energies. More generally when subject to any growing potential, a quantum particle will only be allowed to sit in a discrete set of states, each with its corresponding energy.

In this section, however, we wish to study a different class of potentials, those that asymptote to a constant value

$$\lim_{x \rightarrow \pm\infty} V(x) = V_{\pm} < \infty. \quad (4.75)$$

An example of such a potential is shown in the following Figure.



In the matters we are going to discuss there are a number of subtleties that relate to “how fast” the potential decays at infinity. As these risk to muddle the physical intuition that we wish to build, we are going to be quite cavalier and very simply ignore them all, claiming that our potential just need to “decay sufficiently fast”. For example, an exponential decay (or a faster one) will suit us just fine.

Let us try to infer some general features of a particle subject to this class of potentials. Using the property (4.75), we can write

$$-\frac{\hbar^2}{2m}u''(x) \approx (E - V_{\pm})u(x), \quad x \rightarrow \pm\infty. \quad (4.76)$$

In words, the particle is approximately free at very large distances. There are two qualitatively different situations

- $E - V_{\pm} > 0$.

In this case, the wavefunctions are characterised by $k_{\pm} \in \mathbb{R}$ and behave asymptotically as complex exponentials

$$u(x) \approx ae^{ik_{\pm}x}, \quad E - V_{\pm} = \frac{\hbar^2}{2m}k_{\pm}^2, \quad x \rightarrow \pm\infty. \quad (4.77)$$

Such wavefunctions are associated to *scattering states*.

- $E - V_{\pm} < 0$.

In this case, the wavefunctions are characterised by η_{\pm} and take asymptotically the form of real exponentials

$$u(x) \approx Ae^{\eta_{\pm}x} + Be^{-\eta_{\pm}x}, \quad E - V_{\pm} = -\frac{\hbar^2}{2m}\eta_{\pm}^2, \quad x \rightarrow \pm\infty. \quad (4.78)$$

Notice that neither $\psi_{\text{sc.}}(x; k) = ae^{ikx}$, nor $\psi_{\text{b.s.}}(x, \eta) = Ae^{\eta x} + Be^{-\eta x}$ are wavefunctions, since they are both non-normalisable! For the bound state, we can solve this issue by requiring the solution of the full Schrödinger equation to be such that

$$u(x) \approx \begin{cases} Be^{-|\eta_{+}|x}, & x \rightarrow +\infty \\ Ae^{|\eta_{-}|x}, & x \rightarrow -\infty \end{cases}. \quad (4.79)$$

This type of wavefunction is called *bound state*, because it is “trapped” in the wells of the potential – such as the left valley in the Figure above. As we will see shortly, these states only occur for very specific values of η_{\pm} .

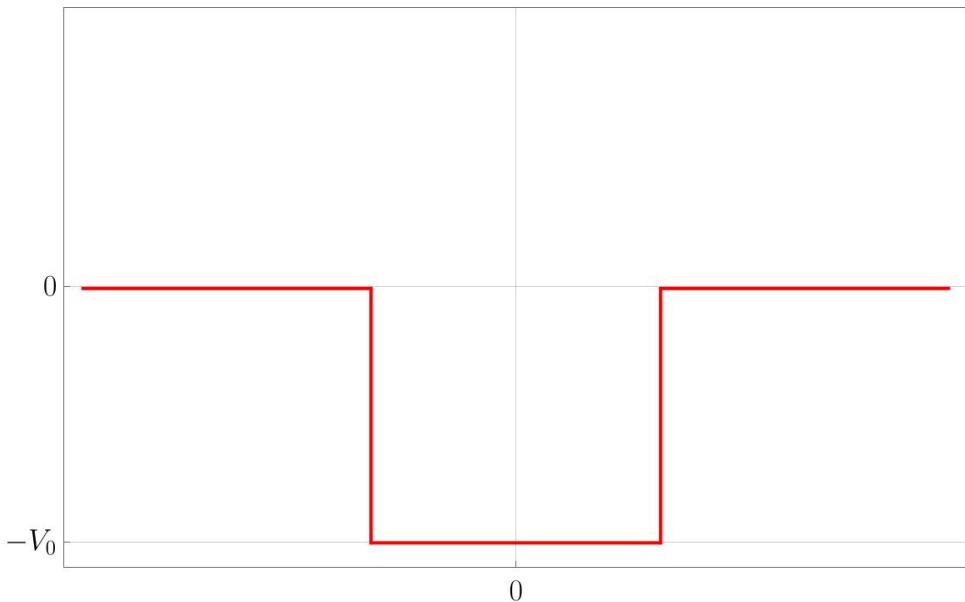
On the other hand, wavefunctions that behave asymptotically as scattering states cannot be made normalisable in the same way as bound states. We need a more careful approach to resolve the issue. We will deal with this later. First, let us build some intuition on bound states by looking at a simple example.

4.4.1 Well, well, well

We will be considering a finite and symmetric¹⁵ potential well

$$V(x) = \begin{cases} -V_0, & -L/2 < x < L/2 \\ 0, & \text{otherwise} \end{cases}, \quad V_0 > 0. \quad (4.80)$$

It looks as in the following Figure.



¹⁵The connotation “symmetric” here refers to the fact that the potential asymptotes to the same constant at $x \rightarrow +\infty$ and $x \rightarrow -\infty$.

We have set $V_+ = V_- = 0$ to simplify what follows¹⁶. You might be worried about the discontinuities at $x = \pm L/2$. In contrast to the case of the particle in a box, where the potential was infinite outside the interval, here we cannot simply set $u(\pm L/2) = 0$. So, what influence does the discontinuities have on the wavefunction? Clearly, the second derivative $u''(x)$ will be discontinuous: this is imposed on us by the Schrödinger equation

$$-\frac{\hbar^2}{2m}u''(x) = \begin{cases} (E + V_0)u(x), & -L/2 < x < L/2 \\ Eu(x), & \text{otherwise} \end{cases}. \quad (4.81)$$

What about the first derivative? Let us integrate (4.81) in a small interval around $x = -L/2$

$$-\frac{\hbar^2}{2m} \int_{-L/2-\epsilon}^{-L/2+\epsilon} dx u''(x) = \int_{-L/2-\epsilon}^{-L/2+\epsilon} dx (E - V(x))u(x). \quad (4.82)$$

In the limit $\epsilon \rightarrow 0$, the left-hand side gives us the discontinuity of the first derivative, while it is easy to see that the right-hand side is finite, since it is the integral of a piecewise constant function. So we expect the wavefunction to be of class C^1 : continuous with continuous first derivative, but discontinuous second derivative.

The task at hand here is quite simple. We are interested in bound state solutions, i.e. those for which $E < V_\pm = 0$. So we know they must have the asymptotic form (4.79). At the same time, we must also have $E > -V_0$ because, otherwise, the wavefunction would vanish identically¹⁷. Hence, inside the well, the solution is a sum of complex exponentials. What we will do is: we separately solve the Schrödinger equation (4.81) inside and outside the well and then we patch the solutions along the junctures at $x = \pm L/2$, requiring continuity of the wavefunction and its first derivative.

Before jumping in the fray, though, we add one idea that will make our life easier: that of *parity*. Notice that the potential (4.80) is an even function

$$V(-x) = V(x). \quad (4.83)$$

This means that, if $u(x)$ is a solution with energy E , then also $u(-x)$ is a solution with the same energy. Now, assuming that no two distinct states possess the same energy¹⁸, we must necessarily have

$$u(x) = \alpha u(-x), \quad \forall x \in \mathbb{R}, \quad (4.84)$$

for some $\alpha \in \mathbb{C}$. But notice that we can play the same game twice

$$u(x) = \alpha u(-x) = \alpha^2 u(-(-x)) \equiv \alpha^2 u(x), \quad (4.85)$$

which tells us that $\alpha = \pm 1$. Thus, our wavefunctions must be either even or odd functions. This results in a considerable simplification of our computations, so let us take advantage of this fact and consider each case separately.

¹⁶A shift of the energy is all it takes to set these constants from some arbitrary value to 0, in the symmetric case.

¹⁷The only possible solutions with $E < -V_0$ have the form $ae^{\eta x} + be^{-\eta x}$ both inside and outside the well (with different η s), which are not normalisable.

¹⁸This argument can be generalised also in the cases where this condition is not satisfied. We will not need to consider these situations.

Even case

We have all the ingredients we need to cook up our even solution. Outside the well, the wavefunction has the form

$$u(x) = \begin{cases} Ae^{\eta x}, & x < -L/2 \\ Ae^{-\eta x}, & x > L/2 \end{cases}, \quad \eta > 0, \quad (4.86)$$

since it is even and must be normalisable. In the well the potential is still constant, but now $E - V(x) = E + V_0 > 0$, so our solutions are complex exponentials. The parity forces

$$u(x) = B \cos(kx), \quad |x| < L/2, \quad k > 0. \quad (4.87)$$

The relations of η and k to E and V_0 are¹⁹

$$E = -\frac{\hbar^2}{2m}\eta^2 = \frac{\hbar^2}{2m}k^2 - V_0. \quad (4.88)$$

Now we glue (4.86) and (4.87) along the seams. Since we imposed the parity, we only need to look at $x = L/2$: the other conditions will be automatically satisfied. We have

$$\begin{aligned} \text{Continuity of } u(x) : \quad B \cos\left(k\frac{L}{2}\right) &= Ae^{-\eta L/2}, \\ \text{Continuity of } u'(x) : \quad -kB \sin\left(k\frac{L}{2}\right) &= -\eta Ae^{-\eta L/2}. \end{aligned} \quad (4.89)$$

We solve the first, fixing

$$B = \frac{A}{\cos\left(k\frac{L}{2}\right)} e^{-\eta L/2}, \quad (4.90)$$

and take the ratio of the two equation, arriving at the transcendental equation

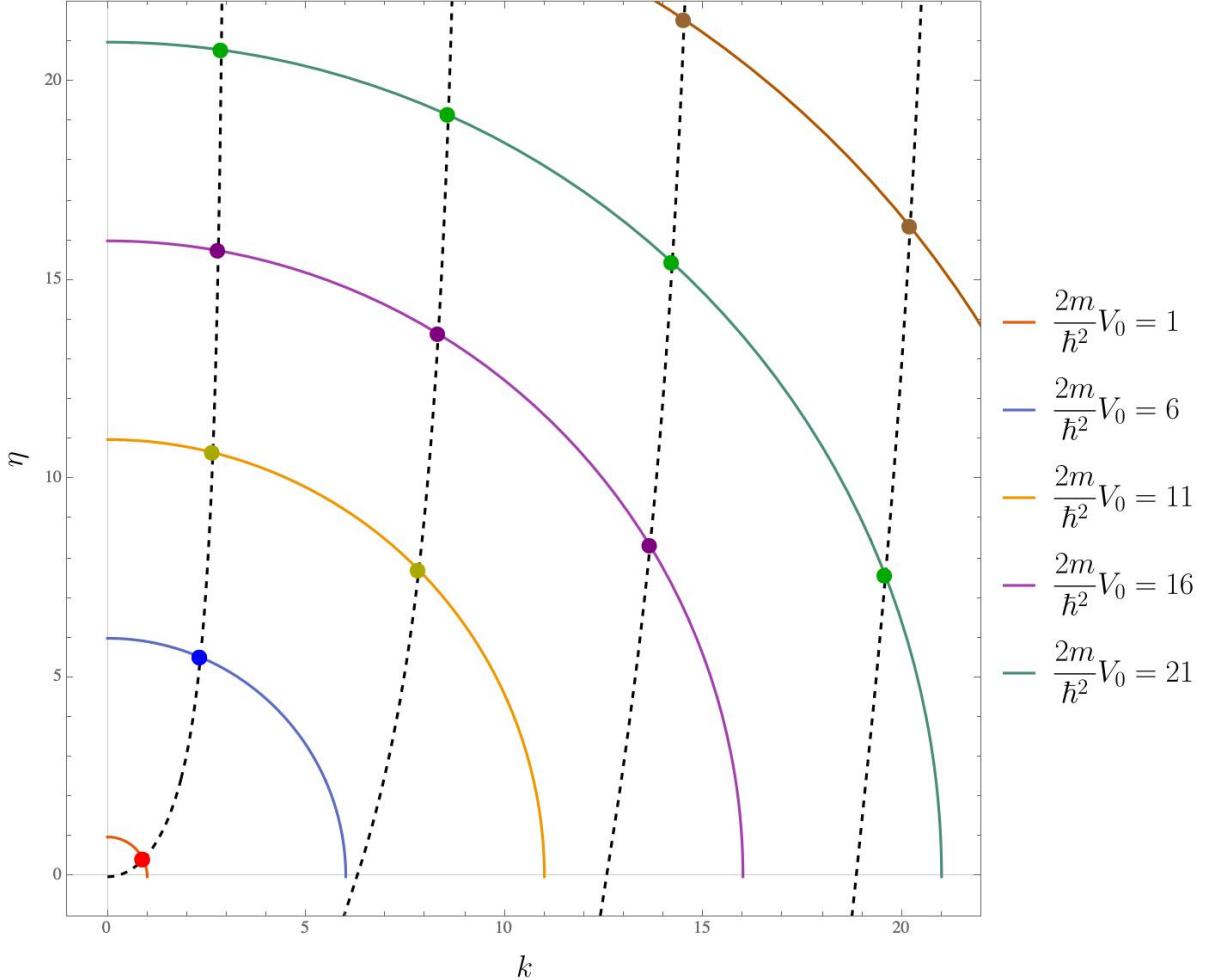
$$k \tan\left(k\frac{L}{2}\right) = \eta. \quad (4.91)$$

This equation is transcendental because, remember, η and k are not independent. They are related by (4.88), which we can write as

$$\eta^2 + k^2 = \frac{2m}{\hbar^2}V_0. \quad (4.92)$$

A solution in closed form of (4.91) constrained by (4.92) does not exist, but we can investigate these equations graphically, plotting them on the (k, η) plane, first quadrant.

¹⁹Note that $\eta > 0$ and $k > 0$ imply, respectively, $E < 0$ and $E > -V_0$.



The dashed line is $k \tan(KL/2)$ (we set $L = 1$) while the circles are (4.92) for various values of the combination $2mV_0/\hbar^2$. The solutions we are searching for are the values (k, η) where the two type of lines cross, highlighted in the Figure above by points. As expected, we only find a discrete set of values of η – equivalently of E – but, more importantly, there is only a finite amount of them. Other things that we notice is that there is always at least one solution, no matter the value of mV_0/\hbar^2 . The number of subsequent solutions, however, depends on the system's parameters. In particular, the number of solutions grows with V_0 . It also grows with L , although it is not immediately evident. You can show it yourselves²⁰.

Let us look at the limit of the infinitely deep well $V_0 \rightarrow \infty$. In this case, in order to satisfy (4.92), we can take $\eta \rightarrow \infty$ in concert. At the same time, the transcendental equation is satisfied for $kL \rightarrow (2n - 1)\pi$, with $n \in \mathbb{N}$. Now, clearly the energy $E \propto -\eta^2$ diverges to $-\infty$, but we always have the freedom to choose the reference from which we measure the energies of the system. In this limit, we choose this reference to be the floor of the potential $-V_0$, meaning we re-define the energy $E' = E + V_0$. Then from (4.88) we find

$$E' = \frac{\hbar^2}{2m}k^2 = \frac{\hbar^2\pi^2}{2mL^2}(2n - 1)^2. \quad (4.93)$$

But this is exactly the (odd part of) the energy spectrum for the particle in a box we

²⁰A smart way, if you want a suggestion, is to scale both k and η by $1/L$.

considered earlier! Well, maybe this should not be excessively surprising, since the $V_0 \rightarrow \infty$ produces exactly the potential we used then. It is a nice check that we did not commit any error, though.

There is one last thing to notice before moving on. All the bound states have energies contained in the interval $-V_0 < E < 0$, which is precisely what we would expect for a classical particle trapped in a well. However, there is a very non-classical feature in our solution: the probability of the particle being outside the well is non-zero (though is exponentially suppressed)!

Odd case

Technically, this case works out very much like the even parity one. We will be quick. The odd parity solution has the form

$$u(x) = \begin{cases} Ae^{\eta x}, & x < -L/2 \\ B \sin\left(k\frac{L}{2}\right), & -L/2 < x < L/2 \\ -Ae^{-\eta x}, & x > L/2 \end{cases} . \quad (4.94)$$

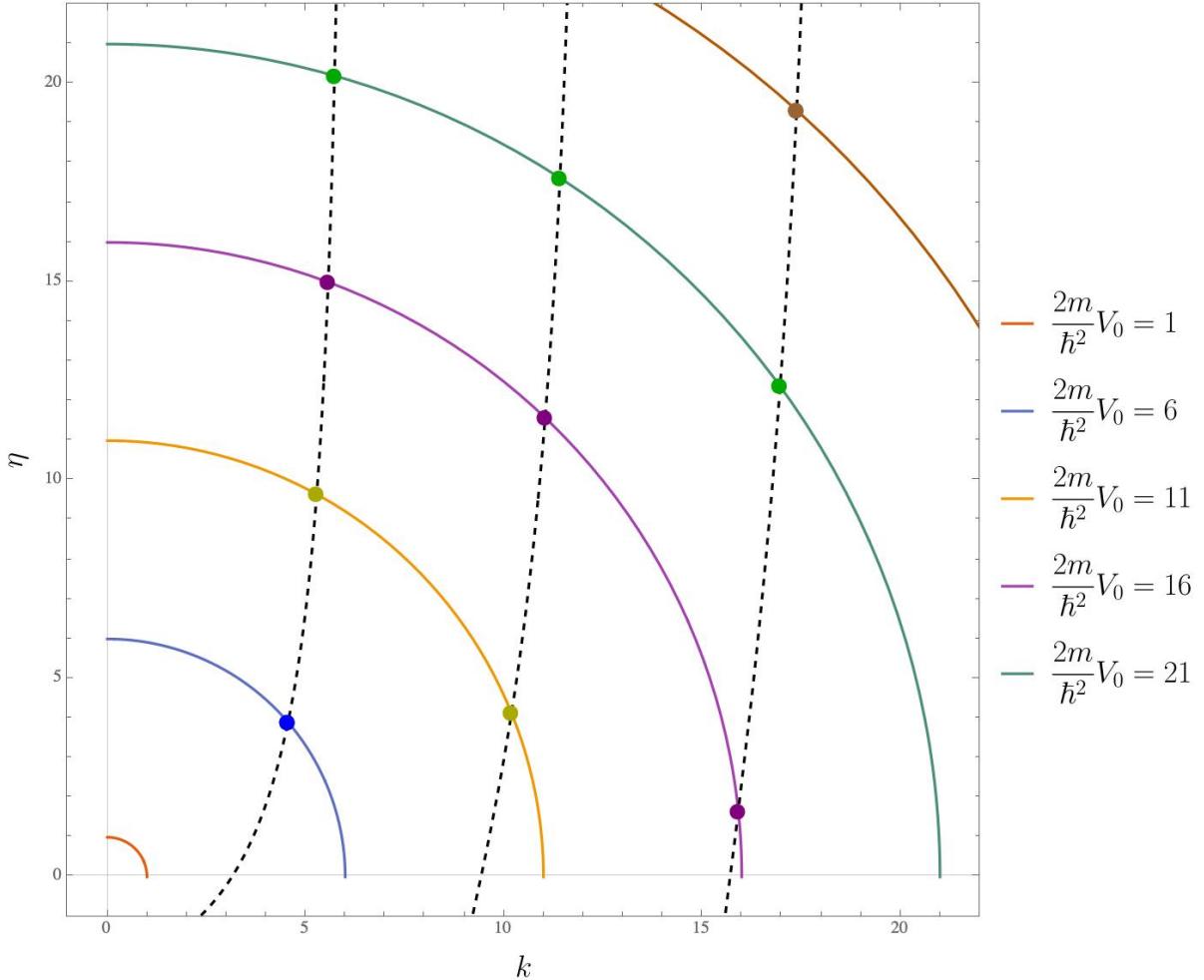
The gluing of the solutions gives

$$\begin{aligned} \text{Continuity of } u(x) : \quad B \sin\left(k\frac{L}{2}\right) &= Ae^{-\eta L/2}, \\ \text{Continuity of } u'(x) : \quad kB \cos\left(k\frac{L}{2}\right) &= -\eta A e^{-\eta L/2}, \end{aligned} \quad (4.95)$$

from which we arrive at the equation

$$\frac{k}{\tan\left(k\frac{L}{2}\right)} = -\eta . \quad (4.96)$$

Let us use the graphical method again. The plot now looks as follows.



We see that we are not guaranteed anymore to have at least one solution. In fact, the first dashed line emerges from the $k = 0$ axis into the first quadrant at $kL = \pi$. The circle (4.92) intersect this line only if

$$\frac{2mV_0}{\hbar^2} > \frac{\pi^2}{L^2} \implies \frac{L^2 m V_0}{\hbar^2} > \frac{\pi^2}{2}. \quad (4.97)$$

Thus an odd solution only exist if the well is deep or wide enough.

By similar arguments as the one used for the even parity case, we can recover in the $V_0 \rightarrow \infty$ case the missing even part of the particle in a box spectrum.

4.4.2 Throwing particles at walls

Let us now turn to study scattering states. The idea behind scattering experiments is simple: you take a particle, hurl it against some object and record how it bounces off. Ideally, if you repeat this experiment a large amount of times, you can use the data about the bouncing off to derive information on the structure of the object in question. These kind of experiments are how the structure of the atom was explored at the beginning of last Century and how LHC probes the microscopic world of fundamental particles²¹.

²¹It is actually also the way in which your eyes are able to see.

We just saw that potentials of the type (4.75), should they have some type of well, admit a finite number of negative energy bound states, trapped in the potential. Here we are asking a different question: what happens if we stand far from the potential and throw in a quantum particle. Will it bounce back, or will it pass through the potential? Intuitively, this being a quantum particle associated to a wave, we expect it to do both²². How do we describe this setup mathematically? One natural approach would be to consider a wave-packet, localised asymptotically far in the negative space direction, and send it moving towards the region where the potential is not constant. This approach has the advantage that a wave-packet gels nicely with our classical expectation of a particle. However, there is one heavy price to pay: the description of a wave-packet solution is fairly complicated even for an everywhere constant potential and becomes much more so should the potential be non-trivial.

We will take a different approach, daringly considering “wavefunctions” of the form

$$u_k(x) = A_k e^{ikx}, \quad k \in \mathbb{R}, A_k \in \mathbb{C}. \quad (4.98)$$

These have a definite momentum $p = \hbar k$ but are not admissible states, since they are not normalisable. Are we allowed to consider them? How should we think about them? Well, one possible answer involves appealing to the linearity of Schrödinger equation and, hence, to the superposition principle for solutions. We can then consider functions (4.98) as mathematical tools without direct physical significance and agree that the true physical solutions are superpositions that turn out to be normalisable.

There is an alternative, complementary interpretation that we can give to justify our approach. Rather than associating the functions (4.98) to a single particle, we take them to be describing a continuous beam of particles. Now the absolute value squared $P(x, t) = |\psi(x, t)|^2 = |u_k(x)e^{-iEt/\hbar}|^2 = |A|^2$ takes the role of the *average density of particles* in the beam. We can reinforce this interpretation by computing the current (4.27)

$$J(x, t) = -\frac{i\hbar}{2m} \left(\bar{\psi}(x, t) \frac{\partial}{\partial x} \psi(x, t) - \psi(x, t) \frac{\partial}{\partial x} \bar{\psi}(x, t) \right) = |A|^2 \frac{p}{m}, \quad (4.99)$$

which is the average density $|A|^2$ of particles times their velocity p/m , i.e. the average flux of particles.

We will focus on two particularly simple cases in which the potential is piece-wise constant. They will nonetheless be quite revealing on the general properties of quantum particles.

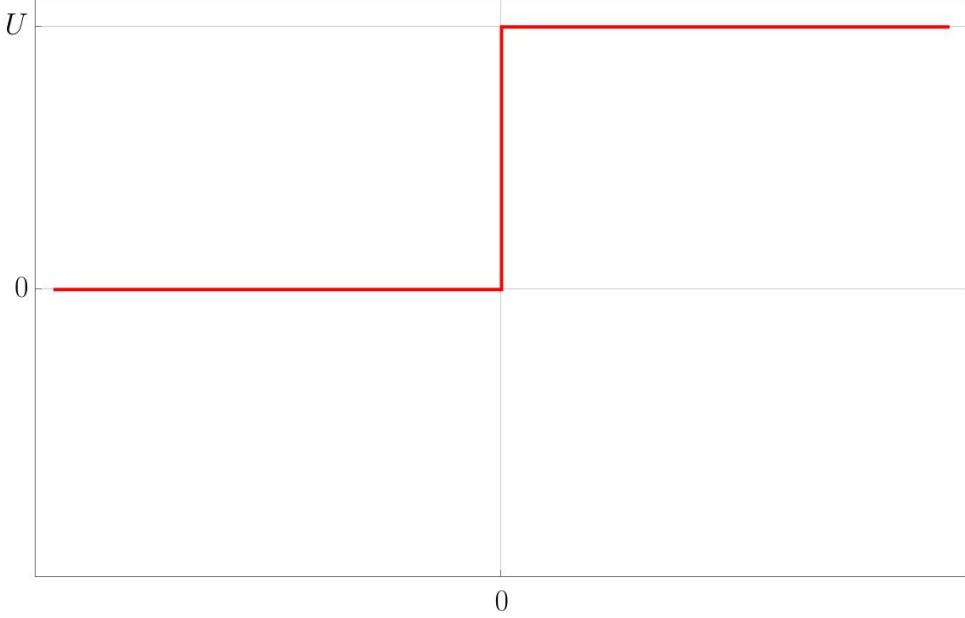
The step potential

Our first example is the step potential

$$V(x) = \begin{cases} 0, & x < 0 \\ U, & x > 0 \end{cases}, \quad U > 0. \quad (4.100)$$

It looks as in the following Figure.

²²As we will see, the results are very similar to what a wave experiences when passing through interfaces between media of different densities.



We will throw a particle – or, more physically, a beam of particles – at it from the left and study what happens. The expectation is that, if $E < U$ the particles will bounce back, while if $E > U$ they will fly by without noticing the step. Presumably.

This is similar to problems we saw when studying waves. on the half-line $x < 0$ the potential is 0, which means that the solution there is

$$u(x) = Ae^{ikx} + Be^{-ikx}, \quad k = \frac{\sqrt{2mE}}{\hbar} > 0, \quad (4.101)$$

where we choose to identify Ae^{ikx} with the right-moving wave by choosing $k > 0$. Notice that the energies of incoming and outgoing beams have to be the same, courtesy of the Schrödinger equation. However we expect the densities $|A|^2$ and $|B|^2$ to be potentially different, with $B \leq A$, as not all the particles might be reflected.

Now, in the $x > 0$ region, the potential is non-zero but constant, meaning we still have exponential solutions. We will not distinguish the case $E > U$ from $E < U$ from the outset, but simply consider a solution

$$u(x) = Ce^{ik'x} + De^{-ik'x}, \quad k' = \frac{\sqrt{2m(E-U)}}{\hbar}, \quad (4.102)$$

where k' is real for $E > U$ – which is fine – and imaginary for $E < U$ – which sound pretty strange. We will learn how to think of an imaginary momentum as we proceed.

Pondering a moment on (4.102), we realise that we were excessively general. We can do better and set $D = 0$. In fact, for $E < U$ the term $De^{-ik'x}$ makes $u(x)$ into a non-normalisable function – since $-ik' = \eta'$ with $\eta' = \sqrt{2m(U-E)}/\hbar > 0$ – and we do not want that. Conversely, if $E > U$, the term $De^{-ik'x}$ represents a left-moving wave. But we are not sending anything in from the right! Our setup is a beam of particles coming from the left and hitting the step. Left-moving particles should only exist for $x < 0$. So, in conclusion, our solution looks as follows

$$u(x) = \begin{cases} Ae^{ikx} + Be^{-ikx}, & x < 0 \\ Ce^{ik'x}, & x > 0 \end{cases}. \quad (4.103)$$

And now we patch. We already know how to proceed and we arrive at the equations

$$\begin{aligned} \text{Continuity of } u(x) : & \quad A + B = C , \\ \text{Continuity of } u'(x) : & \quad ik(A - B) = ik'C . \end{aligned} \quad (4.104)$$

Solutions are easily found

$$B = \frac{k - k'}{k + k'}A , \quad C = 2\frac{k}{k + k'}A . \quad (4.105)$$

We chose to express B and C in terms of A , since the latter determines the density of the incoming beam of particles and is thus a parameter of our setup. Compare the above expressions to those for the reflection and transmission amplitudes for waves incident on an interface between two strings (1.65). Let us recast this result in terms of fluxes. We have three beams here, an incoming – the one we are shooting from the left – a reflected and a transmitted one

$$\begin{aligned} J_{\text{inc.}} &= |A|^2 \frac{\hbar k}{m} , \\ J_{\text{refl.}} &= |B|^2 \frac{\hbar k}{m} = |A|^2 \frac{\hbar k}{m} \left(\frac{k - k'}{k + k'} \right)^2 , \\ J_{\text{tran.}} &= |C|^2 \frac{\hbar k'}{m} = |A|^2 \frac{\hbar k'}{m} \frac{4k^2}{(k + k')^2} . \end{aligned} \quad (4.106)$$

Let us spend a moment to interpret these results, starting with the case $E > U$, so that k' is real and, classically, we'd expect the particle to simple fly over the step without even noticing it. We will find it more convenient to think about the ratios of the fluxes

$$\begin{aligned} R &\doteq \frac{J_{\text{refl.}}}{J_{\text{inc.}}} = \left(\frac{k - k'}{k + k'} \right)^2 , \\ T &\doteq \frac{J_{\text{tran.}}}{J_{\text{inc.}}} = \frac{4kk'}{(k + k')^2} , \end{aligned} \quad (4.107)$$

called, respectively, *reflection and transmission coefficients*. They tells us what fraction of the incident beam is reflected and what fraction is transmitted. From the single particle perspective, these quantities give the probabilities that the particle is reflected or transmitted. As a quick sanity check, we can verify that

$$R + T = 1 , \quad (4.108)$$

which confirms that nothing of the incoming beam is lost or trapped by the potential. From the other perspective, the particle is either reflected or transmitted. There are no further alternatives. Note that the result $R + T = 1$ holds for any potential. This in particular means that if you shoot a particle towards a well, like the one we studied previously, there is no way for it to remain trapped in the potential. There is no such thing as a quantum golf game.

A fact you should remember is that k' really is a function of k or, better, both k and k' are functions of the energy E . We write them as in (4.107) for aesthetic purposes. Let us look at limiting cases

- $E \rightarrow U$: in this limit, $k' \rightarrow 0$ and $(R, T) \rightarrow (1, 0)$. So, when the particle has barely enough energy to make it over the well, it is simply reflected back with almost 100% chance.
- $E \rightarrow \infty$: now $k' \approx k$ and $(R, T) \rightarrow (0, 1)$. So as the energy grows more and more, there is less and less probability that the particle is reflected.

These limiting cases square rather well with our expectations

What about the case in which $E < U$? Well, now the wavefunction in the $x > 0$ region is

$$u(x) = Ce^{-\eta' x}, \quad \eta' = \frac{2m(U - E)}{\hbar} > 0, \quad (4.109)$$

with C still taking the expression (4.105). The meaning of an imaginary momentum can be derived from the above expression: it tells us that the wavefunction decays exponentially inside the barrier. In contrast to classical mechanics, there is some non-negligible probability to find the particle a distance $x \approx 1/\eta'$ inside the barrier but, beyond this point, the probability drops off quickly.

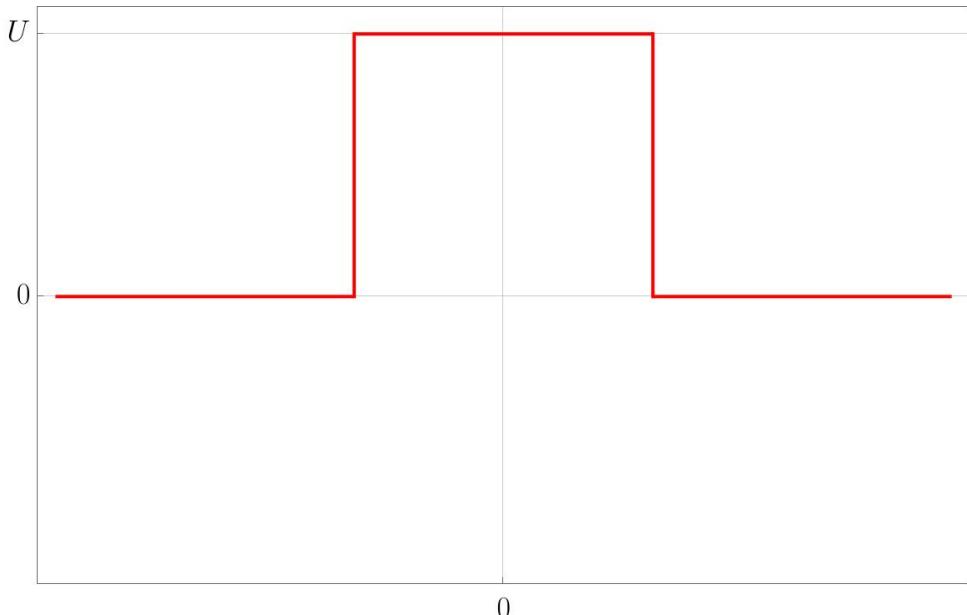
The difference between this and the previous case is that an exponentially decaying wavefunction has a vanishing flux²³ $J_{\text{tran.}} = 0$: it does not transport anything anywhere. This means that $T = 0$ and²⁴ $R = 1$.

Tunnelling

The above example did not yield excessively surprising results – aside from, possibly, that for $E > U$ there is always a chance that the incident particle is reflected. Now we come to the juicy stuff. Let us consider a “bump” potential

$$V(x) = \begin{cases} U, & -L/2 < x < L/2 \\ 0, & \text{otherwise} \end{cases}, \quad U > 0. \quad (4.110)$$

This is a mirror of the finite well we studied earlier and it looks as follows.



²³In fact, the rightmost side in (4.99) is not valid anymore, since $\bar{u}(x) = u(x)$.

²⁴It is easy to see that $R = 1$ since $|B| = 1$ as B is the ratio of two complex conjugate numbers.

We are now experts in dealing with piece-wise potentials, so we can breeze through the calculations. We are interested in the situation $0 < E < U$, in which a classical particle would simply bounce back. Our solution will then have the form

$$u(x) = \begin{cases} e^{ikx} + Be^{-ikx}, & x < -L/2 \\ Ce^{-\eta x} + De^{\eta x}, & -L/2 < x < L/2 \\ Ee^{ikx}, & x > L/2 \end{cases} \quad \begin{cases} k = \frac{\sqrt{2mE}}{\hbar} > 0 \\ \eta = \frac{\sqrt{2m(U-E)}}{\hbar} > 0 \end{cases}. \quad (4.111)$$

We choose to set the density of the incoming beam to unity ($A = 1$) in anticipation of the fact that a non-unity density will simply drop out of the quantities we wish to compute²⁵.

We have two sets of gluing conditions

$$\text{Continuity of } u(x) \text{ at } x = -\frac{L}{2} : e^{-ik\frac{L}{2}} + Be^{ik\frac{L}{2}} = Ce^{\eta\frac{L}{2}} + De^{-\eta\frac{L}{2}}, \quad (4.112a)$$

$$\text{Continuity of } u'(x) \text{ at } x = -\frac{L}{2} : ik \left(e^{-ik\frac{L}{2}} - Be^{ik\frac{L}{2}} \right) = \eta \left(De^{-\eta\frac{L}{2}} - Ce^{\eta\frac{L}{2}} \right), \quad (4.112b)$$

$$\text{Continuity of } u(x) \text{ at } x = \frac{L}{2} : Ee^{ik\frac{L}{2}} = Ce^{-\eta\frac{L}{2}} + De^{\eta\frac{L}{2}}, \quad (4.112c)$$

$$\text{Continuity of } u'(x) \text{ at } x = \frac{L}{2} : ikEe^{ik\frac{L}{2}} = \eta \left(De^{\eta\frac{L}{2}} - Ce^{-\eta\frac{L}{2}} \right). \quad (4.112d)$$

Solving these equations is only a step up in algebraic complexity from what we saw before. Ultimately, what we are interested in is the transmission coefficient $T = |E|^2 / |A|^2 = |E|^2$, so we really only need to compute E from the above equations. We can be smart about it and consider the following sum

$$k(4.112a) - i(4.112b) + (k \cosh(\eta L) + i\eta \sinh(\eta L)) (4.112c) - \frac{k \sinh(\eta L) + i\eta \cosh(\eta L)}{\eta} (4.112d),$$

from which we derive

$$E = \frac{2k\eta e^{-ikL}}{2k\eta \cosh(\eta L) - i(k^2 - \eta^2) \sinh(\eta L)}. \quad (4.113)$$

Then the transmission probability reads

$$T = |E|^2 = \frac{4k^2\eta^2}{4k^2\eta^2 \cosh^2(\eta L) + (k^2 - \eta^2)^2 \sinh^2(kL)} = \frac{1}{1 + \frac{(k^2 + \eta^2)^2}{4k^2\eta^2} \sinh^2(L\eta)}. \quad (4.114)$$

So in the quantum world there is a non-zero chance that the particle makes it through the potential wall! This effect is called *quantum tunnelling*.

Let us get some feels for this expression by looking at a limiting case: that of a particle with very low energy. To be careful, we are asking $U - E$ to be very large, compared to some feature of the system which has dimensions of an energy. Some dimensional analysis tells us that this quantity can only be $\hbar^2/(ML^2)$. So the regime we are studying is

$$U - E \gg \frac{\hbar^2}{mL^2} \implies \eta L \gg 1, \quad (4.115)$$

²⁵You saw that happening above when computing R and T .

and can be reached in various ways, either taking U much larger than E or L being very large or \hbar being very small (classical limit). They are all the same regime. We see that the transmission coefficient vanishes exponentially there, since we can take $\sinh(L\eta) \sim e^{L\eta}$ when $L\eta \gg 1$. In formulae

$$T \sim \frac{16k^2\eta^2}{(\eta^2 + k^2)^2} e^{-\frac{2L}{\hbar}\sqrt{2m(U-E)}}, \quad (4.116)$$

where both k and η above must be considered as functions of E and U . This is a key feature of the quantum tunnelling: the exponential suppression of the transmission probability with the length of the barrier or, equivalently, with its height as compared to the particle energy or, just as well, in the classical limit.

A very similar effect is at play in radioactive decay. In, admittedly rather simplified models, an alpha particle can be thought of as trapped inside the nucleus by a finite, but large potential energy barrier. A classical particle would be consigned to rattle around in the nucleus forever; a quantum particle can, with some exponentially suppressed probability, tunnel through the barrier and out the other side. The small probability manifests itself in the long lifetime of many unstable nuclei.

Outro: takeaway points

In this chapter, we introduced the concepts at the heart of quantum mechanics. This constitutes a new way to describe Nature that becomes necessary when investigating the behaviour of small objects such as atoms, electrons and the like. At these scales, classical physics is unable to account for the surprising experimental results that begun piling up at the turn of the 19th Century. After years of struggle, theoretical physicist came up with a consistent theory: quantum mechanics/dynamics. Its wild departure from a classical view and its often counter-intuitive consequences notwithstanding, its fantastical agreement with experimental results assures us that quantum mechanics/dynamics is the appropriate description of reality.

In this chapter we learned that:

- In the quantum world, any particle is associated to a *de Broglie* wave. The frequency ω and wave-number \vec{k} of this wave are related to the energy E and momentum \vec{p} of the particle

$$E = \hbar\omega, \quad \vec{p} = \hbar\vec{k},$$

where

$$\hbar \approx 1.05 \times 10^{-34} \text{ J} \cdot \text{s},$$

is the *reduced Planck's constant*.

- The state of a system (we focused on the case of a single particle) is determined by a complex-valued *wavefunction*

$$\begin{aligned} \mathbb{R}^3 \times \mathbb{R} &\longrightarrow \mathbb{C} \\ \psi(\vec{x}, t) : \quad \Psi &\quad \Psi \\ (\vec{x}, t) &\longmapsto \psi(\vec{x}, t) \end{aligned}.$$

This quantity contains all the available information on the system. So, in the quantum world, each particle is associated to a wave.

- The wavefunction is interpreted as a *probability amplitude* and its modulus squared gives the *probability density* $P(\vec{x}, t)$ for the particle to be at a point \vec{x} at a given time t

$$P(\vec{x}, t) = |\psi(\vec{x}, t)|^2 .$$

This is known as *Born rule*.

The probability to find a particle in a region $R \subseteq \mathbb{R}^3$ is given by the integral

$$P_R(t) = \int_R dV P(\vec{x}, t) .$$

- Proper wavefunctions need to be normalised

$$\int_{\mathbb{R}^3} dV P(\vec{x}, t) = \int_{\mathbb{R}^3} dV |\psi(\vec{x}, t)|^2 = 1 .$$

With some abuse of nomenclature, we refer to as wavefunctions any complex function $\Psi(\vec{x}, t)$ whose absolute value squared integrates to a finite constant over the whole space

$$\int_{\mathbb{R}^3} dV |\Psi(\vec{x}, t)|^2 = N \implies \int_{\mathbb{R}^3} dV |\psi(\vec{x}, t)|^2 = 1 , \quad \psi(\vec{x}, t) = \frac{1}{\sqrt{N}} \Psi(\vec{x}, t) .$$

These wavefunctions are called *normalisable*.

Since the phase of a wavefunction has no influence on the corresponding probability density, a normalisable wavefunction is determined *up to a multiplicative complex constant*. The wavefunctions $\psi(\vec{x}, t)$ and $\lambda\psi(\vec{x}, t)$ describe the same state.

- Quantum mechanical objects enjoy the *superposition principle*. This means that if $\psi_1(\vec{x}, t)$ and $\psi_2(\vec{x}, t)$ are two allowed wavefunctions for the system, then so is any arbitrary complex linear combination

$$\psi_3(\vec{x}, t) = \alpha\psi_1(\vec{x}, t) + \beta\psi_2(\vec{x}, t) , \quad \alpha, \beta \in \mathbb{C} .$$

In quantum systems it is the probability amplitudes – i.e. the wavefunctions – that add together, not the probability densities. This means that $P_3(\vec{x}, t) \neq P_1(\vec{x}, t) + P_2(\vec{x}, t)$, a fact that explains the interference patterns observed in the double-slit experiment.

- The dynamics of the wavefunction is encoded in the *Schrödinger equation*

$$i\hbar \frac{\partial}{\partial t} \psi(\vec{x}, t) = -\frac{\hbar^2}{2m} \nabla^2 \psi(\vec{x}, t) + V(\vec{x}, t) \psi(\vec{x}, t) ,$$

where $V(\vec{x}, t)$ is the *potential energy* – a specific feature of the system under scrutiny. This equation can be written in a more general form as

$$i\hbar \frac{\partial}{\partial t} \psi(\vec{x}, t) = \hat{H} \psi(\vec{x}, t) ,$$

where \hat{H} is the *Hamiltonian operator* of the system. In the case of a particle in a potential $V(\vec{x}, t)$ it takes the form

$$\hat{H} = -\frac{\hbar^2}{2m} \nabla^2 + V(\vec{x}, t) ,$$

but in more complicated systems it can be of a different form.

- The probability density satisfies a *continuity equation*

$$\frac{\partial}{\partial t} P(\vec{x}, t) + \vec{\nabla} \cdot \vec{J}(\vec{x}, t) = 0 ,$$

where $\vec{J}(\vec{x}, t)$ is the *probability current*

$$\vec{J}(\vec{x}, t) \doteq -\frac{i\hbar}{2m} \left(\bar{\psi}(\vec{x}, t) \vec{\nabla} \psi(\vec{x}, t) - \psi(\vec{x}, t) \vec{\nabla} \bar{\psi}(\vec{x}, t) \right) .$$

As a consequence of the continuity equation, the chance in time of the total probability to find the particle in a region $R \subseteq \mathbb{R}$ is equal to the flux of the probability current across the boundary ∂R of said region

$$\frac{d}{dt} P_R(t) = - \int_R dV \vec{\nabla} \cdot \vec{J}(\vec{x}, t) = - \int_{\partial R} d\vec{S} \cdot \vec{J}(\vec{x}, t) .$$

- Observables are the quantities pertaining to a system that we can measure. In classical mechanics they are all functions of \vec{x} and \vec{p} . In quantum mechanics, observables are *operators*. What we can measure is an observable's *expectation value*, defined as

$$\langle \hat{O} \rangle \doteq \int_{\mathbb{R}^3} dV \bar{\psi}(\vec{x}, t) \hat{O} \cdot \psi(\vec{x}, t) ,$$

where the operator \hat{O} acts on $\psi(\vec{x}, t)$.

The task of associating an operator \hat{O} to a classical observable $O(\vec{x}, \vec{p})$ is called *quantisation*. From the basic definitions

$$\hat{x}\psi(\vec{x}, t) = \vec{x}\psi(\vec{x}, t) , \quad \hat{p}\psi(\vec{x}, t) = -i\hbar \vec{\nabla} \psi(\vec{x}, t) ,$$

one constructs an operator \hat{O} as

$$\hat{O} \doteq O(\hat{x}, \hat{p}) .$$

Doing so for the energy $E(\vec{x}, \vec{p})$ yields the Hamiltonian operator \hat{H} we considered above.

It is important to remark that operators, in general, *do not commute*. This means that the order in which they appear matters and it can lead to *ordering ambiguities*.

- A consequence of the lack of commutativity of quantum observables is the *Heisenberg uncertainty principle* which states that there is an intrinsic limit to the precision with which a pair of physical properties can be known. In its 1-dimensional version it can be written as

$$\Delta x \Delta p \geq \frac{\hbar}{2} .$$

Here ΔO is the (square-root of the) variance of the operator \hat{O}

$$(\Delta O)^2 = \langle \hat{O}^2 \rangle - \langle \hat{O} \rangle^2 .$$

A relation similar to the one above holds for any pair of operators that do not commute.

We restricted to 1 dimension hereafter

- In the case of a *static potential* $V(x, t) = V(x)$, the wavefunction can be taken in factorised form

$$\psi(x, t) = u(x)e^{-i\frac{E}{\hbar}t},$$

and $u(x)$ satisfied the *time-independent Schrödinger equation* – sometimes known as *stationary Schrödinger equation*

$$\hat{H}u(x) = Eu(x).$$

For the case of a single particle in a potential $V(x)$ it takes the form

$$-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}u(x) + V(x)u(x) = Eu(x).$$

The solutions $u(x)$ are the *energy eigenstates* of the Hamiltonian operator \hat{H} .

- We studied a free particle $V(x) = 0$ on a circle S^1 . The solutions to the stationary Schrödinger equation are

$$u_k(x) = Ae^{ikx}, \quad E = \frac{\hbar^2}{2m}k^2.$$

The fact that we ask the particle to be on a circle of radius R implies the boundary condition

$$u(x + 2\pi R) = u(x) \implies k = \frac{n}{R}, \quad n \in \mathbb{Z}.$$

Hence we have a *quantisation condition* for the energy

$$E_n = \frac{\hbar^2}{2mR^2}n^2, \quad n \in \mathbb{Z}.$$

The constant A in the solution is fixed to $A = 1/\sqrt{2\pi R}$ by requiring unit normalisation.

- Next we looked at a particle in a finite interval. We realised this situation with the potential

$$V(x) = \begin{cases} 0, & 0 < x < L \\ \infty, & \text{otherwise} \end{cases}.$$

We can treat this system as a free particle whose wavefunction is forced to vanish at $x = 0$ and $x = L$. The solutions are

$$u_n(x) = \sqrt{\frac{2}{L}} \sin\left(\frac{\pi}{L}nx\right), \quad n \in \mathbb{N}.$$

Here as well, the energy is quantised

$$E_n = \frac{\hbar^2\pi^2}{2mL^2}n^2, \quad n \in \mathbb{N}.$$

- We then looked at a class of non-trivial potentials that asymptote at constant values V_{\pm} at $x \rightarrow \pm\infty$. At large distances the particle is approximately free and the wavefunction can exhibit two different behaviours

- *Scattering state*: $E - V_{\pm} > 0$
the wavefunction is oscillatory

$$u(x) \approx ae^{ik_{\pm}x}, \quad E - V_{\pm} = \frac{\hbar^2}{2m}k_{\pm}^2, \quad x \rightarrow \pm\infty.$$

- *Bound state*: $E - V_{\pm} < 0$
the wavefunction is real exponential

$$u(x) \approx Ae^{\eta_{\pm}x} + Be^{-\eta_{\pm}x}, \quad E - V_{\pm} = -\frac{\hbar^2}{2m}\eta_{\pm}^2, \quad x \rightarrow \pm\infty.$$

In order for the wavefunction of a bound state to be normalisable it has to behave as

$$u(x) \approx \begin{cases} Be^{-|\eta_{+}|x}, & x \rightarrow +\infty \\ Ae^{|\eta_{-}|x}, & x \rightarrow -\infty \end{cases}.$$

- We considered the example of a finite, symmetric potential well

$$V(x) = \begin{cases} -V_0, & -L/2 < x < L/2 \\ 0, & \text{otherwise} \end{cases}, \quad V_0 > 0.$$

We saw that bound state solutions – i.e. with $-V_0 < E < 0$ break into two families determined by their parity $u(x) = u(-x)$ and $u(x) = -u(-x)$.

Imposing continuity of the solution and its first derivative at the discontinuous points of the potential, we arrived at the following expressions

- even case:

$$u(x) = \begin{cases} Ae^{\eta x}, & x < -L/2 \\ \frac{A}{\cos(kL/2)}e^{-\eta L/2} \cos(kx), & -L/2 < x < L/2 \\ Ae^{-\eta x}, & x > L/2 \end{cases}, \quad E = -\frac{\hbar^2}{2m}\eta^2 = \frac{\hbar^2}{2m}k^2 - V_0.$$

The energy is quantised via the transcendental equation

$$k \tan\left(k\frac{L}{2}\right) = \eta, \quad \eta^2 + k^2 = \frac{2m}{\hbar^2}V_0.$$

We studied this pair of equations graphically, showing that there is a finite amount of allowed bound state energies and that there is always at least one solution, no matter how shallow or tiny is the well.

- odd case:

$$u(x) = \begin{cases} Ae^{\eta x}, & x < -L/2 \\ \frac{A}{\sin(kL/2)}e^{-\eta L/2} \sin(kx), & -L/2 < x < L/2 \\ Ae^{-\eta x}, & x > L/2 \end{cases}, \quad E = -\frac{\hbar^2}{2m}\eta^2 = \frac{\hbar^2}{2m}k^2 - V_0.$$

The energy is quantised via the transcendental equation

$$\frac{k}{\tan\left(k\frac{L}{2}\right)} = -\eta, \quad \eta^2 + k^2 = \frac{2m}{\hbar^2}V_0.$$

We studied this pair of equations graphically, showing that there is a finite amount of allowed bound state energies and that there is at least one solution only if

$$\frac{L^2 m V_0}{\hbar^2} > \frac{\pi^2}{2} .$$

- We then moved on to study *scattering problems*, that is, situations in which the energy of the system E is larger than the asymptotic values of a non-trivial potential $V(x)$. In this situation, the solution will behave asymptotically as a harmonic wave

$$u(x) \underset{x \rightarrow \pm\infty}{\approx} u_k(x) , \quad u_k(x) = A_k e^{ikx} , \quad k \in \mathbb{R} , \quad A_k \in \mathbb{C} .$$

Although the harmonic waves are not normalisable in \mathbb{R} , we swept this problem under the rug, considering them as mathematical tools without direct physical significance.

We also suggested a possible interpretation of harmonic waves as describing a continuous beam of particles, so that $|\psi(x, t)|^2 dx$ is now interpreted as an *average density of particles* in the interval $(x, x + dx)$.

- We studied the *step potential*

$$V(x) = \begin{cases} 0 , & x < 0 \\ U , & x > 0 \end{cases} , \quad U > 0 .$$

In a setup in which we shoot a particle (or a beam) from the far left towards the step, the solutions will take the form

$$u(x) = \begin{cases} Ae^{ikx} + Be^{-ikx} , & x < 0 \\ Ce^{ik'x} , & x > 0 \end{cases} , \quad E = \frac{\hbar^2}{2m} k^2 = \frac{\hbar^2}{2m} k'^2 + U .$$

The two terms in the $x < 0$ branch are an incoming, right-moving wave (with constant A) and a reflected, left-moving wave (with constant B). The term in the $x > 0$ is a transmitted, right-moving wave.

Patching the two branches of the solution together – i.e. requiring continuity of the function and of its first derivative – yields the relations

$$B = \frac{k - k'}{k + k'} A , \quad C = 2 \frac{k}{k + k'} A .$$

We defined the *reflection and transmission coefficients*

$$R \doteq \left| \frac{B}{A} \right|^2 = \left(\frac{k - k'}{k + k'} \right)^2 , \quad T \doteq \left| \frac{C}{A} \frac{k'}{k} \right|^2 = \frac{4kk'}{(k + k')^2} .$$

They tell us the fraction of the incident beam that is reflected and transmitted at $x = 0$, respectively. Correctly $R + T \equiv 1$.

- Finally we considered a “bump” potential

$$V(x) = \begin{cases} U , & -L/2 < x < L/2 \\ 0 , & \text{otherwise} \end{cases} , \quad U > 0 .$$

In the same setup in which we shoot a particle (or a beam) from the far left, and supposing $0 < E < U$, the solution takes the form

$$u(x) = \begin{cases} e^{ikx} + Be^{-ikx}, & x < -L/2 \\ Ce^{-\eta x} + De^{\eta x}, & -L/2 < x < L/2 \\ Ee^{ikx}, & x > L/2 \end{cases}, \quad E = \frac{\hbar^2}{2m}k^2 = U - \frac{\hbar^2}{2m}\eta^2,$$

where we fixed $A = 1$ from the start. Imposing continuity of the solution and of its first derivative allows to fix the constants B, C, D, E . Our interest was in the transmission coefficient, which in this case is simply $T = |Ek'/(Ak)|^2 = |E|^2$, since $A = 1$ and $k' = k$. We found

$$T = \left(1 + \frac{(k^2 + \eta^2)^2}{4k^2\eta^2} \sinh^2(L\eta)\right)^{-1},$$

which tells us that there is a non-vanishing probability that the particle makes it through the wall, even if $0 < E < U$ would imply, classically, that the particle bounces back.

Chapter 5

Eigenvalue problems of Sturm-Liouville type

In the previous chapter we have introduced some – actually a lot of – facts about quantum mechanics and studied some exemplary systems. One important thing we did was to split the solutions to the Schrödinger equation into an oscillatory, time-dependent exponential and a function $u(x)$ that only depends on the spatial variable: the stationary wavefunction. This has to satisfy the time-independend Schrödinger equation

$$-\frac{\hbar^2}{2m}u''(x) + V(x)u(x) = Eu(x) . \quad (5.1)$$

In the present chapter we are going to have a more general and formal look at this equation and, more generally, at a special class of equations that (5.1) belongs to: the *Sturm-Liouville problems*.

5.1 Spectral problems

The time-independent Schrödinger equation (5.1) is a special case of a second-order linear ODE, whose general form is

$$u''(x) + p(x)u'(x) + q(x)u(x) = \lambda\rho(x)u(x) , \quad (5.2)$$

where $u : \mathbb{R} \rightarrow \mathbb{C}$ is required to be at least of class C^2 (i.e. twice continuously differentiable), $p(x)$, $q(x)$ and $\rho(x)$ are complex-valued functions and $\lambda \in \mathbb{C}$ is the *spectral parameter*. In order to pin down a solution of a second-order linear ODE, we need to provide two conditions. As we saw in the previous chapters, spatial conditions take the form of boundary conditions. We will assume these to be linear in the function $u(x)$ and its derivative $u'(x)$. For example, Dirichlet and Neumann boundary conditions. Thus we have the problem of solving the following system

$$\begin{cases} u''(x) + p(x)u'(x) + q(x)u(x) = \lambda\rho(x)u(x) , & \forall x \in (x_L, x_R) , \\ \mathcal{B}_L(u(x), u'(x)) = 0 , \\ \mathcal{B}_R(u(x), u'(x)) = 0 , \end{cases} \quad (5.3)$$

where \mathcal{B}_L and \mathcal{B}_R are operators that enforce the two boundary conditions (here L and R stand for “left” and “right”). The problem of determining the values of λ such that a

non-trivial solution to (5.4) exists is known as *eigenvalue problem*. Another very common designation is *spectral problem*¹. Eigenvalues – or spectral – problems actually refer to more general situations, in which the ODE can be of any order. Sometime problems such as (5.4) are called *second order eigenvalue problems*. Other naming conventions we should get the habit of using are

- Eigenvalue.
The values of λ such that a non-trivial solution to (5.4) exists
- Eigenfunction.
The solution $u(x)$ associated to some eigenvalue λ .
- Spectrum.
The collection of all the eigenvalues of the problem (5.4)

Of course, there is a reason why we are using the same terminology that you encountered when studying linear algebra with matrices and vectors. The fundamental idea is that the problem (5.4) is linear. This means that if $u(x)$ and $v(x)$ are two solutions to the ODE, then so is $\alpha u(x) + \beta v(x)$, $\forall \alpha, \beta \in \mathbb{C}$. The functions behave very much like vectors, so much so that you can think of them as elements of an “infinite-dimensional vector space”². Similarly, the ODE can be thought as a *linear operator* \mathcal{L} that sends $u(x)$ to $u''(x) + q(x)u'(x) + p(x)u(x)$. This is very much like an infinitely extended – and infinitely dense – matrix. Now this is all very hand-waving but there exists a very precise way in which functions and linear operator can be treated as vectors and matrices. This is all thanks to the linearity property they satisfy.

5.1.1 Sturm-Liouville problems

Amongst all the possible spectral problems (5.4), there exists a special class called *Sturm-Liouville problems*. They take the following form

$$\begin{cases} (L \cdot u)(x) \doteq -\frac{1}{\rho(x)} \frac{d}{dx} \left(P(x)u'(x) \right) + \frac{Q(x)}{\rho(x)}u(x) = \lambda u(x) , & \forall x \in (a, b) \subset \mathbb{R} \\ \mathcal{B}_a(u, u') = 0 , \\ \mathcal{B}'_b(u, u') = 0 , \end{cases} , \quad (5.4)$$

where we used the notation $(L \cdot u)(x)$ as a short-hand to stand for the left-hand side of the ODE³. $P(x)$, $Q(x)$ and $\rho(x)$ are real-valued functions, $u(x)$ is complex-valued and the boundary points a and b can also be infinite⁴. The boundary conditions determined by \mathcal{B}_a and \mathcal{B}'_b need to satisfy an additional requirement: if $u(x)$ and $v(x)$ are any pair of class C^1 (continuously once-differentiable) functions obeying the conditions, then

$$\left[\bar{v}'(x)P(x)u(x) - \bar{v}(x)P(x)u'(x) \right]_{x=a}^{x=b} = 0 . \quad (5.5)$$

¹This is because we are tasked to find the *spectrum* of the ODE, that is the values of λ for which a solution exists.

²Think of it like this. Take a vector \vec{v} with components $v(n)$. Now suppose that this vector has an infinite number of components and, moreover, these components form a dense set. Then you can put them in one-to-one correspondence with points on the real line, simply by setting $n = x$.

³This is to highlight that we are thinking of the left-hand side of the ODE as the action of a linear operator L on the function $u(x)$.

⁴In which case, the interval is taken to be (half-)open: $(-\infty, b]$, $[a, \infty)$ or $(-\infty, \infty)$.

Why this weird-looking condition you might ask? Well, the answer is contained in the following chain of identities

$$\begin{aligned}
\int_a^b dx \rho(x) \bar{v}(x) (L \cdot u)(x) &= \int_a^b dx \bar{v}(x) \left[-\frac{d}{dx} (P(x)u'(x)) + Q(x)u(x) \right] = \\
&= \left[-\bar{v}(x)P(x)u'(x) \right]_{x=a}^{x=b} + \int_a^b dx \left[P(x)\bar{v}'(x)u'(x) + Q(x)\bar{v}(x)u(x) \right] = \\
&= \left[\bar{v}'(x)P(x)u(x) - \bar{v}(x)P(x)u'(x) \right]_{x=a}^{x=b} + \int_a^b dx \left[-\frac{d}{dx} (P(x)\bar{v}'(x)) + Q(x)\bar{v}(x) \right] u(x) = \\
&= \left[\bar{v}'(x)P(x)u(x) - \bar{v}(x)P(x)u'(x) \right]_{x=a}^{x=b} + \int_a^b dx \overline{\left[-\frac{d}{dx} (P(x)v'(x)) + Q(x)v(x) \right]} u(x) = \\
&= \left[\bar{v}'(x)P(x)u(x) - \bar{v}(x)P(x)u'(x) \right]_{x=a}^{x=b} + \int_a^b dx \rho(x) \overline{(L \cdot v)}(x) u(x) ,
\end{aligned}$$

where we integrated by parts in various places and used the fact that $P(x)$, $Q(x)$ and $\rho(x)$ are real. So, if the condition (5.5) is satisfied, one obtains the identity

$$\int_a^b dx \rho(x) \bar{v}(x) (L \cdot u)(x) = \int_a^b dx \rho(x) \overline{(L \cdot v)}(x) u(x) . \quad (5.6)$$

In the spirit of thinking of functions as vectors and operators as matrices, let us introduce the following notation

$$\langle v, u \rangle \doteq \int_a^b dx \rho(x) \bar{v}(x) u(x) . \quad (5.7)$$

Then we rewrite (5.6) as

$$\langle v, L \cdot u \rangle = \langle L \cdot v, u \rangle . \quad (5.8)$$

It would be almost criminal not to notice the analogy with vectors in \mathbb{C}^N and matrices acting on them. Before hurrying, though, we should make sure that the operation $\langle \bullet, \bullet \rangle$ we just introduced is a proper inner product. In fact it is not difficult to do so. You can verify yourselves the defining properties

- $\langle v, \alpha u_1 + \beta u_2 \rangle = \alpha \langle v, u_1 \rangle + \beta \langle v, u_2 \rangle$,
- $\overline{\langle v, u \rangle} = \langle u, v \rangle$,
- $\langle u, u \rangle \geq 0$,
- $\langle u, u \rangle = 0 \iff u \equiv 0$.

So functions on the interval $[a, b]$ keep on seeming more and more similar to vectors. Not only they form a vector space (this is again something you can easily show to be true by proving the various axioms for the addition and scalar multiplication operations), they are elements of an *normed* vector space! And linear operators such as L really are linear

maps, so they can truly be thought as generalisations of matrices. Then the identity (5.8) is telling us that the linear map L is *hermitian*⁵. Recall that for usual matrices, the hermiticity condition implied that their eigenvalues must be real and that the eigenvectors for distinct eigenvalues are orthogonal. As we are going to see next, these two properties are true for hermitian operators as well.

Before moving on let us tie this discussion to our quantum mechanical investigation. First, the Sturm-Liouville condition (5.5) is rather obscure and it is not immediately clear if Dirichlet and Neumann boundary conditions satisfy it. Let us verify that.

- D-D

Suppose the boundary conditions are $\mathcal{B}_a(u, u') = u(a)$ and $\mathcal{B}'_b(u, u') = u(b)$, so that we have Dirichlet-Dirichlet conditions on the interval $u(a) = u(b) = 0$. Then we immediately see that

$$\left[\bar{v}'(x)P(x)u(x) - \bar{v}(x)P(x)u'(x) \right]_{x=a}^{x=b} \equiv 0 . \quad \checkmark$$

- N-N

Now we take $\mathcal{B}_a(u, u') = u'(a)$ and $\mathcal{B}'_b(u, u') = u'(b)$, which give Neumann-Neumann boundary $u'(a) = u'(b) = 0$. Again the Sturm-Liouville condition is trivially solved

$$\left[\bar{v}'(x)P(x)u(x) - \bar{v}(x)P(x)u'(x) \right]_{x=a}^{x=b} \equiv 0 . \quad \checkmark$$

You can easily convince yourselves that also Dirichlet-Neumann condition is acceptable according to the standards of Sturm and Liouville.

Finally, the time-independent Schrödinger equation

$$-\frac{\hbar^2}{2m}u''(x) + V(x)u(x) = Eu(x) , \quad (5.9)$$

when supplied with appropriate boundary conditions, is clearly a Sturm-Liouville problem. It is sufficient to set $P(x) = -\hbar^2/(2m)$, $Q(x) = V(x)$, $\rho(x) = 1$ and $\lambda = E$.

5.2 Properties of Sturm-Liouville problems

So, the problems we deal with in quantum mechanics are of Sturm-Liouville type. This fact turns out to have profoundly important consequences, since the Sturm-Liouville problems enjoy some very nice properties. We actually already mentioned them: they are the same properties enjoyed by hermitian matrices: reality of the spectrum and orthogonality of the eigenvectors. What this means, physically, is that the eigenvalues arising from the Schrödinger equation – with appropriate boundary conditions – will always be real. This, admittedly, is a pretty important physical requirement since we want to identify these with the energies of the allowed states of the system. The orthogonality of the spectrum has less transparent consequences though. First of all, what does it mean for two functions

⁵Remember that for usual matrices the hermiticity condition is $\bar{L}^T = L$.

to be orthogonal? Well, taking the definition (5.7) as an inner product, we say that two function $u(x)$ and $v(x)$ are orthogonal iff

$$\langle v, u \rangle \equiv \int_a^b dx \rho(x) \bar{v}(x) u(x) = 0 . \quad (5.10)$$

Now, remember when we were talking about initial conditions for finite strings? We mentioned that sinusoidal functions satisfy an “orthogonality condition”

$$\frac{2}{\pi} \int_0^\pi dx \sin(mx) \sin(nx) = \delta_{m,n}$$

This is exactly what we are talking about. If we set $a = 0$, $b = \pi$, $\rho(x) = 1$ and define the set of functions $u_n(x) = \sqrt{2/\pi} \sin(nx)$, then these are orthogonal and normalised to 1 (in jargon, they are *orthonormal*) $\langle u_m, u_n \rangle = \delta_{m,n}$. This orthogonality condition is not an exclusive feature of sines and cosines. There exist other functions that enjoy this important property⁶. What matters here is that we can work with eigenfunctions of some Sturm-Liouville problem just as we did with sinusoids when we were studying waves.

Let us prove the reality of the spectrum and the orthogonality of eigenfunctions.

Theorem 1. Reality of the spectrum for Sturm-Liouville problems

The eigenvalues of Sturm-Liouville problems are real.

Proof 1.1. Reality of the spectrum for Sturm-Liouville problems

Take the following operator action

$$(L \cdot u)(x) = \lambda u(x) ,$$

with $\lambda \in \mathbb{C}$ and $u(x) \not\equiv 0$. Then we can compute

$$\lambda \langle u, u \rangle = \langle u, (L \cdot u) \rangle = \langle (L \cdot u), u \rangle = \langle \lambda u, u \rangle = \bar{\lambda} \langle u, u \rangle .$$

Here we used, in order, sesquilinearity^a of the inner product $\langle \bullet, \bullet \rangle$, hermiticity of the operator L , and, hence, the condition (5.5), action of L on u and finally, again, sesquilinearity of the inner product. Hence we have $(\lambda - \bar{\lambda}) \langle u, u \rangle = 0$. Since $\langle u, u \rangle = 0 \Leftrightarrow u \equiv 0$, which violates our hypothesis, we conclude that

$$\lambda = \bar{\lambda} \implies \lambda \in \mathbb{R} .$$

■

^aRemember – or not, if you never heard this term – that a bi-linear form $\langle \bullet, \bullet \rangle$ is called sesquilinear if it satisfies the conditions $\langle u, \alpha v_1 + \beta v_2 \rangle = \alpha \langle u, v_1 \rangle + \beta \langle u, v_2 \rangle$, and $\langle \alpha u_1 + \beta u_2, v \rangle = \bar{\alpha} \langle u_1, v \rangle + \bar{\beta} \langle u_2, v \rangle$.

⁶In fact, you might remember we mentioned that Bessel J functions also enjoy an orthogonality relation.

Theorem 2. Orthogonality of eigenfunctions in Sturm-Liouville problems

Let $u_1(x)$ and $u_2(x)$ be eigenfunctions of a Sturm-Liouville-type operator L , associated to different eigenvalues λ_1 and λ_2 , respectively. Then

$$\langle u_1, u_2 \rangle \equiv \int_a^b dx \rho(x) \overline{u_1}(x) u_2(x) = 0 .$$

We say that u_1, u_2 are *orthogonal* with respect to the inner product with *weight* $\rho(x)$.

Proof 2.1. Orthogonality of eigenfunctions in Sturm-Liouville problems

We follow the same chain of identities as for in the previous theorem, but for the inner product $\langle u_1, u_2 \rangle$

$$\lambda_2 \langle u_1, u_2 \rangle = \langle u_1, (L \cdot u_2) \rangle = \langle (L \cdot u_1), u_2 \rangle = \overline{\lambda_1} \langle u_1, u_2 \rangle .$$

Thus we have that $(\lambda_2 - \overline{\lambda_1}) \langle u_1, u_2 \rangle = 0$. Since, by hypothesis and thanks to the previous theorem, $\overline{\lambda_1} = \lambda_1 \neq \lambda_2$ it must necessarily be true that

$$\langle u_1, u_2 \rangle = 0$$

■

5.2.1 Reduction of spectral problems to Sturm-Liouville form

We introduced the Sturm-Liouville problems as a special class of second order spectral problems. What really makes them special is the requirement (5.5) on the boundary conditions. In fact, any second-order linear ODE can be recast in the Sturm-Liouville form. Let us see how.

Consider the general second-order linear ODE

$$u''(x) + p(x)u'(x) + q(x)u(x) = \lambda w(x)u(x) , \quad (5.11)$$

and let us multiply both sides by the arbitrary function $R(x)$

$$R(x)u''(x) + R(x)p(x)u'(x) + R(x)q(x)u(x) = \lambda R(x)w(x)u(x) . \quad (5.12)$$

We know our objective is to recast the above into a Sturm-Liouville form

$$-\frac{d}{dx} \left(P(x)u(x) \right) + Q(x)u(x) \equiv -P(x)u''(x) - P'(x)u'(x) + Q(x)u(x) = \lambda \rho(x)u(x) , \quad (5.13)$$

so we ask for $R(x)$ to satisfy the identities

$$R(x) = -P(x) , \quad R(x)p(x) = -P'(x) , \quad (5.14)$$

for some function $P(x)$. Notice that

$$\frac{P'(x)}{P(x)} = p(x) \implies P(x) = \exp \left[\int dx p(x) \right] . \quad (5.15)$$

This is everything we need. We have converted the generic second-order linear ODE (5.11) into the Sturm-Liouville form (5.13), with coefficient functions defined as

$$\begin{aligned} P(x) &= \exp \left[\int dx p(x) \right], \\ Q(x) &= -q(x) \exp \left[\int dx p(x) \right], \\ \rho(x) &= -w(x) \exp \left[\int dx p(x) \right]. \end{aligned} \quad (5.16)$$

Of course, this does not mean that any second-order spectral problem is a Sturm-Liouville problem, one needs to ensure that the boundary conditions satisfy (5.5).

Let us make an example.

Example

Consider the following second-order ODE

$$x^2 u''(x) - 2xu'(x) + u(x) = -\lambda x^4 u(x).$$

First thing: we divide everything by x^2 , so that the second derivative appears with coefficient 1.

$$u''(x) - \frac{2}{x} u'(x) + \frac{1}{x^2} u(x) = -\lambda x^2 u(x).$$

Thus our coefficient functions are

$$p(x) = -\frac{2}{x}, \quad q(x) = \frac{1}{x^2}, \quad w(x) = -x^2.$$

Let us apply the formulae we obtained above. We need the primitive of $p(x)$ which is easy to obtain

$$\int dx p(x) = -2 \int \frac{dx}{x} = -2 \log(x).$$

Now we simply substitute

$$\begin{aligned} P(x) &= e^{-2 \log(x)} = \frac{1}{x^2}, \\ Q(x) &= -\frac{1}{x^2} e^{-2 \log(x)} = -\frac{1}{x^4}, \\ \rho(x) &= x^2 e^{-2 \log(x)} = 1. \end{aligned}$$

Hence the Sturm-Liouville form of the equation we started with is

$$-\frac{d}{dx} \left(\frac{1}{x^2} u'(x) \right) - \frac{u(x)}{x^4} = \lambda u(x).$$

You might verify this is correct by expanding the above expression.

5.2.2 Quantum mechanical interpretation of Sturm-Liouville boundary conditions

The requirement (5.5) on the boundary conditions for a Sturm-Liouville problem possesses a surprisingly beautiful physical interpretation. Let us hark back to the solution $\psi(x, t)$ to the full Schrödinger equation (the time-dependent one). We saw that we can write $\psi(x, t)$ in factorised form

$$\psi(x, t) = u(x)e^{-i\frac{E}{\hbar}t}, \quad (5.17)$$

with the function $u(x)$ satisfying the Sturm-Liouville-type ODE

$$-\left(P(x)u'(x)\right) + Q(x)u(x) = \lambda\rho(x)u(x), \quad \begin{cases} P(x) = \hbar^2/(2m) \\ Q(x) = V(x) \\ \rho(x) = 1 \\ \lambda = E \end{cases}. \quad (5.18)$$

Now, remember the definition of the probability current (4.27)

$$J(x, t) = -\frac{i\hbar}{2m} \left(\bar{\psi}(x, t) \frac{d}{dx} \psi(x, t) - \psi(x, t) \frac{d}{dx} \bar{\psi}(x, t) \right). \quad (5.19)$$

Let us substitute our factorised solution in the above expression⁷

$$\begin{aligned} J(x, t) \equiv J(x) &= \frac{i}{\hbar} \left(\bar{u}'(x) \frac{\hbar^2}{2m} u(x) - \bar{u}(x) \frac{\hbar^2}{2m} u'(x) \right) = \\ &= \frac{i}{\hbar} \left(\bar{u}'(x) P(x) u(x) - \bar{u}(x) P(x) u'(x) \right). \end{aligned} \quad (5.20)$$

The Sturm-Liouville condition (5.5) reads

$$\left[\bar{v}'(x) P(x) u(x) - \bar{v}(x) P(x) u'(x) \right]_{x=a}^{x=b} = 0, \quad (5.21)$$

so it immediately implies

$$J(b) - J(a) = 0. \quad (5.22)$$

But this is very interesting! In fact recall that the probability current enters the continuity equation for the probability density $P(x, t) = |\psi(x, t)|^2$

$$\frac{\partial}{\partial t} P(x, t) + \frac{d}{dx} J(x) = 0. \quad (5.23)$$

As we have discussed, the probability $P_{[a,b]}(t)$ to find the particle inside the interval $[a, b]$

$$P_{[a,b]}(t) = \int_a^b dx P(x, t), \quad (5.24)$$

varies in time in a way that is precisely compensated by the flux of probability across the boundary of the interval

$$\frac{d}{dt} P_{[a,b]}(t) = \int_a^b dx \frac{\partial}{\partial t} P(x, t) = - \int_a^b dx \frac{d}{dx} J(x) = J(b) - J(a) = 0! \quad (5.25)$$

So Sturm-Liouville boundary conditions are all those for which the probability to find the particle in the interval $[a, b]$ remains constant in time.

⁷Notice that for stationary states the probability current is time-independent!

5.2.3 Regular Sturm-Liouville problems

When we studied waves on a string, we mentioned the fact that any solution can be represented as a superposition of standing waves. In fact, more is true. Any function with definite, say D-D, boundary conditions on an interval can be written as a (possibly infinite) sum of sines and cosines: the Fourier series. Furthermore, this sum (or series) can be inverted – yielding the coefficients – thanks to an orthogonality relation satisfied by sinusoidal functions. Now, we have just seen that the eigenfunctions of Sturm-Liouville problems also enjoy an orthogonality relation. It is natural to ask whether they as can be used to reconstruct any function in the defining interval $[a, b]$. In jargon, what we are asking here is whether the set of eigenvectors forms a *complete orthonormal system*⁸.

Sturm-Liouville problems often produce complete orthonormal systems. Often, but not always. Those that do are known as *regular Sturm-Liouville problems* and are determined by the following conditions

- The functions $P(x)$, $P'(x)$, $Q(x)$ and $\rho(x)$ are real and continuous in $[a, b]$;
- $P(x)$ and $\rho(x)$ are strictly positive in $[a, b]$;
- a and b are finite;

These conditions makes us very happy because any time-independent Schrödinger equation on a finite interval satisfies them.

Theorem 3. Complete orthonormal systems in regular S-L problems

For any regular Sturm-Liouville problem the following properties hold true

1. The eigenvalues are infinitely, countably many and constitute an increasing sequence

$$\lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \dots ,$$

with the limiting behavior

$$\lambda_n \xrightarrow[n \rightarrow \infty]{} \infty .$$

2. Associated to each eigenvalue λ_n is a unique (up to multiplication by constants) eigenfunction $u_n(x)$ having exactly $n - 1$ zeroes in the fundamental interval $[a, b]$.
3. The set of all (normalised) eigenfunctions forms a complete orthonormal^a system for the interval $[a, b]$.

^aThe orthonormality is, clearly, considered with respect to the scalar product weighted by $\rho(x)$.

Proof 3.1. Complete orthonormal systems in regular S-L problems

Tough luck, the proof requires functional analysis, so it is well above our paygrade.

The consequences of this theorem for quantum mechanics are deep. What we (did not actually) proved is that

⁸We threw in the normalisation condition, since it amounts to a rescaling of the eigenvectors.

Any one-dimensional quantum system on a finite interval whose total probability is preserved in time possesses a countably infinite number of energy levels (*quantisation of energy*). Additionally, any function $u(x)$ on that interval – not just any solution, any function – with $\langle u, u \rangle < \infty$ can be expanded in the basis of eigenfunctions pertaining to the quantum system

$$u(x) = \sum_{n=1}^{\infty} \langle u_n, u \rangle u_n .$$

Outro: takeaway points

In this chapter, we introduced the *eigenvalue problems of Sturm-Liouville type* and presented their main features. In particular:

- We defined a general *spectral* (or *eigenvalue*) problem as the following combination of ODE and boundary conditions

$$\begin{cases} u''(x) + p(x)u'(x) + q(x)u(x) = \lambda\rho(x)u(x) , & \forall x \in (x_L, x_R) , \\ \mathcal{B}_L(u(x), u'(x)) = 0 , \\ \mathcal{B}_R(u(x), u'(x)) = 0 . \end{cases}$$

Here \mathcal{B}_L and \mathcal{B}_R are some operators that enforce the boundary conditions at the left and right edges of the interval (x_L, x_R) . The problem requires to find the set of complex *eigenvalues* λ such that a solution to the above system exists. Such a solution $u_\lambda(x)$, associated to a given eigenvalue λ , is called an *eigenfunction*. The collection of all eigenvalues of a spectral problem is called *spectrum*.

- We identified a subclass of spectral problems called *Sturm-Liouville problems*. Their form is the following

$$\begin{cases} (L \cdot u)(x) \doteq -\frac{1}{\rho(x)} \frac{d}{dx} \left(P(x)u'(x) \right) + \frac{Q(x)}{\rho(x)}u(x) = \lambda u(x) , & \forall x \in (a, b) \subset \mathbb{R} \\ \mathcal{B}_a(u, u') = 0 , \\ \mathcal{B}'_b(u, u') = 0 , \end{cases} ,$$

where the both a and b can be infinite and following conditions need to be satisfied

- $P(x)$, $Q(x)$ and $\rho(x)$ are real-valued functions;
- the operators \mathcal{B}_a and \mathcal{B}'_b are linear;
- given any pair of C^1 (continuously once-differentiable) function $u(x)$ and $v(x)$ that satisfy both boundary conditions, the following equality must be obeyed

$$\left[\bar{v}'(x)P(x)u(x) - \bar{v}(x)P(x)u'(x) \right]_{x=a}^{x=b} = 0 .$$

A consequence of the last condition is the following identity

$$\int_a^b dx \rho(x) \bar{v}(x) (L \cdot u)(x) = \int_a^b dx \rho(x) \overline{(L \cdot v)}(x) u(x) .$$

- We drew a parallel between operators and function on a side and matrices and vectors on the other. In fact we introduced an *inner product* of functions

$$\langle v, u \rangle \doteq \int_a^b dx \rho(x) \bar{v}(x) u(x) ,$$

which depends on the data of the Sturm-Liouville problem under scrutiny. This operation satisfies all the conditions of an inner product. We can then rewrite the last identity of the previous point as

$$\langle v, L \cdot u \rangle = \langle L \cdot v, u \rangle .$$

This condition is summed up in saying that the linear map L appearing in a Sturm-Liouville problem is *hermitian*. This suggests that, like it is the case for matrices, its *eigenvalues must be real* and *eigenfunctions of different eigenvalues must be orthogonal* w.r.t. the inner product.

- We checked that Dirichlet and Neumann boundary conditions satisfy the requirements in the definition of Sturm-Liouville problems. Additionally, we saw that the Schrödinger equation, supplied with appropriate boundary conditions, is a Sturm-Liouville problem.
- We proved our hunch on the properties of the spectrum and eigenfunctions of Sturm-Liouville problems:
 - the eigenvalues of Sturm-Liouville problems are real;
 - eigenfunctions associated to distinct eigenvalues are orthogonal w.r.t. the inner product with weight $\rho(x)$.
- In the case of the Schrödinger equation, we related the Sturm-Liouville constraint on boundary conditions

$$\left[\bar{v}'(x) P(x) u(x) - \bar{v}(x) P(x) u'(x) \right] ,$$

to the request that the total probability to find the particle in the interval $[a, b]$ is constant in time

$$\frac{d}{dt} P_{[a,b]} = 0 , \quad P_{[a,b]} \doteq \int_a^b dx |\psi(x, t)|^2 .$$

This provides a nice physical interpretation: all quantum systems for which the probability is conserved in time have a stationary Schrödinger equation which is a Sturm-Liouville problem.

- We introduced a special subclass of Sturm-Liouville problems, deemed *regular* ones. In addition to the above conditions, they need to satisfy the following
 - $P(x)$, $P'(x)$, $Q(x)$ and $\rho(x)$ are real and continuous in $[a, b]$;
 - $P(x)$ and $\rho(x)$ are strictly positive in $[a, b]$;
 - a and b are both finite.

Any stationary Schrödinger equation on a finite interval whose probability is conserved in time satisfies these requirements.

Regular Sturm-Liouville problems enjoy the following nice properties

- Its eigenvalues are infinitely, countably many and can be arranged in an increasing sequence asymptoting to infinity

$$\lambda_1 \leq \lambda_2 \leq \cdots, \quad \lambda_n \xrightarrow[n \rightarrow \infty]{} \infty.$$

- Each eigenvalue λ_n has a unique (up to normalisation) eigenfunction $u_n(x)$ which has exactly $n - 1$ zeroes in the interval $[a, b]$.
- The set of all normalised eigenfunctions forms a complete orthonormal system for the interval $[a, b]$.

The last property refers to the fact that all eigenfunctions are orthonormal w.r.t. the inner product and, additionally, that any function satisfying the boundary conditions on $[a, b]$ can be written as a superposition of eigenfunctions. This is an extension of Fourier theorem.

Physically, this means that any quantum system on a finite interval, whose probability is conserved in time, possesses a countably infinite number of energy levels (quantisation of energy). Additionally the eigenfunctions can be used to reconstruct any function (not just any solution) on the interval that satisfies the boundary conditions.

Chapter 6

Power series method

Now that we have gone through some general and formal properties of Sturm-Liouville problems, it is time to be more practical and begin solving them. Sure we did see some examples in Chapter 4, but, although we extracted some pretty interesting physics from them, they were all quite trivial, having piece-wise constant potentials. What about systems whose potentials are proper functions like, say, $V(x) = x^2$ or $V(x) = (x^2 + 1)^{-2}$ or, maybe, even more complicated ones such as $V(x) = 1/\cosh(x)$? In this chapter we will introduce a fairly general method to solve second-order differential equations¹: the Frobenius method, which relies on power series expansions about some point x_0 in the fundamental interval of the system under consideration.

6.1 The harmonic oscillator

Sidney Coleman² once said: “The career of a young theoretical physicist consists of treating the harmonic oscillator in ever-increasing levels of abstraction”. I believe this quote applies very much also to mathematical physicists. In fact here we are going to have a first look at the quantum harmonic oscillator. Later in your academic path, you will probably see this system treated in several different ways.

There are, at least, two reasons why the harmonic oscillator is so ubiquitous in theoretical and mathematical physics. The first arises from Taylor theorem: expand any potential $V(x)$ close to a minimum and, at leading order, you will get $V(x) \sim V_0 + \omega(x - x_0)^2$, which is precisely the potential for the harmonic oscillator. This means that small perturbations of more or less any system³ in Nature is described by a harmonic oscillator.

The second reason is more practical: we human beings are not particularly good at solving equations. This might not be so apparent as you first learn mathematics and physics since your teachers are reasonably unwilling to continually repeat “yeah, we don’t know how to solve this one either”. Instead we shine a light on our successes. But as theoretical physics gets more advanced, these successes become harder and harder to find and, at some point, the only system that we can actually solve is the harmonic oscillator. Or, more precisely,

¹The Frobenius method is extendable to ODEs of arbitrary order.

²He was one of the giants of 20th Century theoretical physics. Look him up.

³There are exceptions. Occasionally, you might find that the quadratic term vanishes – i.e. $\omega = 0$ – and you have to look at the next term in the Taylor expansion. This situation is rare but interesting.

things that can be made to look like the harmonic oscillator. Theoretical physicists are uncannily skilful at making everything look like a harmonic oscillator. Take your pick on the coolest result in physics – say, the Higgs boson, or topological insulators, or maybe gravitational waves. Chances are, the underlying theory is primarily constructed with harmonic oscillators.

So, without further ado, let us study the spectral problem for the potential $V(x) = m\omega^2x^2/2$: the quantum harmonic oscillator

$$-\frac{\hbar^2}{2m}u''(x) + \frac{1}{2}m\omega^2x^2u(x) = Eu(x). \quad (6.1)$$

The choice of constant $m\omega^2/2$ is to simplify later manipulations.

Actually, no. Before jumping in, let us briefly recap the main features of Taylor series.

6.1.1 Taylor series and analytic functions

Consider a smooth – i.e. infinitely, continuously differentiable – function $f(x)$. Selecting a point x_0 , we can associate to such function its *Taylor expansion*

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n, \quad (6.2)$$

where $f^{(n)}(x) = \frac{d^n}{dx^n}f(x)$. A priori, the above series might not converge and, even if it did, it may not be equal to the function $f(x)$.

Those functions that are equal to their convergent Taylor expansion in an interval about x_0 are said to be *analytic* in that interval. Simple examples of functions that are analytic in \mathbb{R} are

$$\begin{aligned} e^x &= \sum_{n=0}^{\infty} \frac{1}{n!}x^n, & \sin(x) &= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!}x^{2n}, & \cos(x) &= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!}x^{2n+1}, \\ \sinh(x) &= \sum_{n=0}^{\infty} \frac{1}{(2n)!}x^{2n}, & \cosh(x) &= \sum_{n=0}^{\infty} \frac{1}{(2n+1)!}x^{2n+1}. \end{aligned}$$

Functions can be analytic in other intervals, for example

$$\begin{aligned} \log(1+x) &= -\sum_{n=1}^{\infty} \frac{(-1)^n}{n}x^n, & x \in (-1, 1], \\ \log(1-x) &= -\sum_{n=1}^{\infty} \frac{1}{n}x^n, & x \in [-1, 1), \\ (1+x)^{\alpha} &= \sum_{n=0}^{\infty} \binom{\alpha}{n}x^n, & x \in (-1, 1), \quad \forall \alpha \in \mathbb{C}. \end{aligned}$$

Consider two functions $f(x)$ and $g(x)$, both analytic in an interval $I \subseteq \mathbb{R}$ about x_0 . Then the following statements hold

- Any linear combination of products of $f(x)$ and $g(x)$ is analytic in I .

- If $g(x_0) \neq 0$, then $f(x)/g(x)$ is a function analytic in the interval $I' \subseteq I$ about x_0 . I' is such that $g(x) \neq 0, \forall x \in I'$.
- The derivatives of all orders of $f(x)$ and $g(x)$ are analytic in I .

Consider a series

$$\sum_{n=0}^{\infty} a_n (x - x_0)^n , \quad (6.3)$$

whose coefficients a_n are such that

$$\lim_{n \rightarrow \infty} \left| \frac{a_n}{a_{n+1}} \right| = R \text{ exists or is infinite .} \quad (6.4)$$

We say that R is the *radius of convergence* of the series (which can be infinite). Such a series defines a function $f(x)$ smooth and analytic in the interval $(x_0 - R, x_0 + R)$. As long as we restrict to the analyticity interval, derivatives can bypass the summation symbol

$$\frac{d^m}{dx^m} f(x) = \frac{d^m}{dx^m} \sum_{n=0}^{\infty} a_n (x - x_0)^n = \sum_{n=0}^{\infty} a_n \frac{d^m}{dx^m} (x - x_0)^n , \quad \forall x \in (x_0 - R, x_0 + R) . \quad (6.5)$$

Note that if $f(x) = \sum_{n=0}^{\infty} a_n (x - x_0)^n$ is analytic in an interval about x_0 and $a_n = 0, \forall n \in \mathbb{N} \cup \{0\}$, then $f \equiv 0$ in that interval.

Let us look at a pair of examples. Consider the series

$$f(x) = \sum_{n=0}^{\infty} \frac{1}{n!} x^n . \quad (6.6)$$

To derive the radius of convergence we compute

$$\lim_{n \rightarrow \infty} \left| \frac{a_n}{a_{n+1}} \right| = \lim_{n \rightarrow \infty} \left| \frac{1/n!}{1/(n+1)!} \right| = \lim_{n \rightarrow \infty} |n+1| \rightarrow \infty . \quad (6.7)$$

Thus R is infinite and the function $f(x)$ is analytic on the whole real line \mathbb{R} , which agrees with the fact that $f(x) = e^x$. Another example is the following

$$f(x) = \sum_{n=0}^{\infty} x^n . \quad (6.8)$$

In this case

$$\lim_{n \rightarrow \infty} \left| \frac{a_n}{a_{n+1}} \right| = \lim_{n \rightarrow \infty} |1| = 1 , \quad (6.9)$$

and the radius of convergence is $R = 1$. In fact the function $f(x)$ is $1/(1-x)$.

6.1.2 A reduction of the harmonic oscillator equation

All right, now we can start. The first step, as always, is to clean up our expression as much as possible. With this in mind, we multiply the sides of equation (6.1) by $-2m/\hbar^2$

$$u''(x) - \frac{m^2 \omega^2}{\hbar^2} x^2 u(x) = -2 \frac{mE}{\hbar^2} u(x) . \quad (6.10)$$

Now, let us do a bit of dimensional analysis. The potential $V(x)$ has the same dimensions as an energy $[E]$. Since $[E] = [M][L]^2/[T]^2$ (energy times time), we have

$$[E] = [V(x)] = [M][\omega]^2[L]^2 \implies [\omega]^2 = \frac{[E]}{[M][L]^2} = \frac{1}{[T]^2}, \quad (6.11)$$

which means the constant ω is a frequency, as expected from the Hooke's form of the potential. Remembering that $[\hbar] = [E][T] = [M][L]^2/[T]$, we see that

$$\left[\frac{m\omega}{\hbar} \right] = \frac{[M]/[T]}{[M][L]^2/[T]} = [L]^{-2}, \quad (6.12)$$

is the inverse of a length squared. Also

$$\left[\frac{mE}{\hbar^2} \right] = \frac{[M]}{[E][T]^2} = \frac{[M]}{[M][L]^2} = [L]^{-2}, \quad (6.13)$$

is an inverse length squared. All right, so? What do we do with this knowledge? The answer is: we make everything a-dimensional. It turns out that this step greatly simplifies the equations one is dealing with, so it is always a good first step. In the present case, we define a new variable z and a new parameter ϵ as follows⁴

$$z = \sqrt{\frac{m\omega}{\hbar}}x, \quad \epsilon = 2\frac{mE}{\hbar^2}\frac{\hbar}{m\omega} = 2\frac{E}{\hbar\omega}, \quad v(z) \doteq u(x(z)). \quad (6.14)$$

Now we substitute inside (6.10), obtaining

$$\frac{m\omega}{\hbar}v''(z) - \frac{m\omega}{\hbar}z^2v(z) = -\frac{m\omega}{\hbar}\epsilon v(z). \quad (6.15)$$

Eliminating the common factor $m\omega/\hbar$ we finally arrive at the simple expression

$$v''(z) + (\epsilon - z^2)v(z) = 0. \quad (6.16)$$

If you stare long enough at this equation you might get a headache, but you might also notice that it is quite similar to the equation satisfied by the gaussian function $e^{-z^2/2}$:

$$\frac{d^2}{dz^2}e^{-z^2/2} = (z^2 - 1)e^{-z^2/2}. \quad (6.17)$$

This means that we already have a solution $v(z; \epsilon = 1) = e^{-z^2/2}$. For the other potential solutions we need to work a bit harder. We can study what happens as $|z| \rightarrow \infty$. In this case, ϵ is negligible compared to z^2 , so we conclude that all solutions should behave as⁵ $e^{-z^2/2}$ for large $|z|$. This motivates us to strip off this exponential behaviour from the function $v(z)$, defining

$$h(z) = \frac{v(z)}{e^{-z^2/2}}. \quad (6.18)$$

Now the equation becomes

$$h''(z) - 2zh'(z) + (\epsilon - 1)h(z) = 0. \quad (6.19)$$

Now you might argue that this equation looks more complicated than (6.10) and you would be true. However, here we introduced an additional information on the asymptotic behaviour of $v(z)$, which makes us hope that the solution of (6.19) will be simpler.

⁴The factor 2 in the definition of ϵ cannot be determined by dimensional analysis. One might leave it as an undetermined pure number, to be fixed later so that it simplifies the expressions.

⁵One might also choose $e^{+z^2/2}$, but this would make the wavefunction non-normalisable.

6.1.3 Power series solution and the energy spectrum

Our strategy is to search for a solution in the form of a power series about $z = 0$

$$h(z) = \sum_{n=0}^{\infty} h_n z^n . \quad (6.20)$$

Are we sure that such an expansion converges? And if so, what about its radius of convergence? There exist a theorem that reassures us. We need some definitions though. Consider the following second-order linear ODE

$$u''(z) + p(z)u'(z) + q(z)u(z) = 0 . \quad (6.21)$$

We say that the point $z = z_0$ is

- an *ordinary point*, if $p(z)$ and $q(z)$ are analytic about z_0 ;
- a *regular singular point*, if $(z - z_0)p(z)$ and $(z - z_0)^2q(z)$ are analytic about z_0 ;
- an *irregular singular point* otherwise.

With these definitions we can present, without a proof, the following theorem

Theorem 4. Cauchy theorem

Let z_0 be an ordinary point of the equation

$$u''(z) + p(z)u'(z) + q(z)u(z) = 0 ,$$

and let the functions $p(z)$ and $q(z)$ have Taylor's series about z_0 with radii of convergence R_p and R_q . Then there exist two linearly independent solutions $u_i(z)$, $i = 1, 2$, with power series expansion

$$u_i(z) = \sum_{n=0}^{\infty} u_{i,n}(z - z_0)^n ,$$

with radii of convergence $R_i \geq \min(R_p, R_q)$.

In equation (6.19) $p(z) = -2z$ and $q(z) = \epsilon - 1$, whose radii of convergence are both infinite, ensuring that $h(x)$ is analytic on \mathbb{R} and, thus, admits an everywhere convergent Taylor series about $x = 0$. Let us then work out the expressions of the coefficients h_n . We substitute the Taylor expansion inside (6.19) (remember that for analytic functions we can swap the differentiation and the summation signs)

$$\sum_{n=2}^{\infty} n(n-1)h_n z^{n-2} - 2 \sum_{n=1}^{\infty} nh_n z^n + (\epsilon - 1) \sum_{n=0}^{\infty} h_n z^n = 0 . \quad (6.22)$$

The above equation must hold for any value of z , meaning we need to cancel independently all the coefficients of z^n . First though we need to have the powers inside all sums agree with each other. We do so by shifting the summation index:

$$\sum_{n=2}^{\infty} n(n-1)h_n z^{n-2} = \sum_{n=0}^{\infty} (n+2)(n+1)h_{n+2} z^n . \quad (6.23)$$

So we can write (note that $\sum_{n=1}^{\infty} 2nh_n z^n \equiv \sum_{n=0}^{\infty} 2nh_n z^n$)

$$\sum_{n=0}^{\infty} [(n+2)(n+1)h_{n+2} + (\epsilon - 2n - 1)h_n] z^n = 0 , \quad (6.24)$$

which gives us a *second-order recurrence relation* between the coefficients

$$h_{n+2} = \frac{2n+1-\epsilon}{(n+1)(n+2)} h_n , \quad \forall n \geq 0 . \quad (6.25)$$

In agreement with Cauchy theorem, we have two solutions, since the above recurrence leaves us with two undetermined constants h_0 and h_1 .

One notable property of (6.25) is that it connects even-indexed coefficients to even-indexed coefficients only and similarly for odd-indexed ones. This implies that we have two independent classes of solutions which, again, agrees with Cauchy theorem. Additionally, the relation is homogeneous, meaning that its solution has the form

$$h_{2n} = H_e(n)h_0 , \quad h_{2n+1} = H_o(n)h_1 , \quad (6.26)$$

for some functions $H_e(n)$ and $H_o(n)$. This means that setting $h_1 = 0$ will yield an even function $h(-z) = h(z)$ (since only even powers of z will appear in the series) and, similarly, setting $h_0 = 0$ will yield an odd function $h(-z) = -h(z)$. Another important thing we can extract directly from (6.25) is the radius of convergence of the series (6.20). In fact, consider the following identities

$$R^2 = \lim_{n \rightarrow \infty} \left| \frac{h_n}{h_{n+1}} \right| \left| \frac{h_{n+1}}{h_{n+2}} \right| = \lim_{n \rightarrow \infty} \left| \frac{h_n}{h_{n+2}} \right| = \lim_{n \rightarrow \infty} \left| \frac{(n+1)(n+2)}{2n+1-\epsilon} \right| \rightarrow \infty . \quad (6.27)$$

So, the function $h(z)$ is analytic on the whole real line \mathbb{R} .

Now, let us look at what happens to h_n themselves for large n . We have two possibilities

1. $h_n = 0, \forall n > N$.

In this case, the function $h(z)$ is a polynomial, since the series truncates to a finite sum. The corresponding wavefunction $v(z)$ remains normalisable, since the asymptotic behaviour is still dominated by $e^{-z^2/2}$.

You can easily convince yourselves that this case can only happen for $\epsilon = 2N + 1$.

2. $h_n \neq 0, \forall n \in \mathbb{N} \cup \{0\}$.

Now the large $|z|$ behaviour of the function $h(z)$ might interfere with the exponential decay we extracted from $v(z)$. From the recursion relation, we see that for large n

$$h_{n+2} \approx \frac{2}{n} h_n , \quad (6.28)$$

and this is bad, since it is the same large-index behaviour that gaussians with positive exponents have. You can verify this easily, by looking at the expansions

$$\begin{cases} e^{y^2} = \sum_{n=0}^{\infty} \frac{y^{2n}}{n!} & \Rightarrow a_n = \begin{cases} \frac{1}{(n/2)!} & n \text{ even} \\ 0 & n \text{ odd} \end{cases} \\ ye^{y^2} = \sum_{n=0}^{\infty} \frac{y^{2n+1}}{n!} & \Rightarrow a_n = \begin{cases} 0 & n \text{ even} \\ \frac{1}{(\frac{n-1}{2})!} & n \text{ odd} \end{cases} \end{cases} \Rightarrow a_{n+2} \approx \frac{2}{n} a_n . \quad (6.29)$$

The upshot is that if the series (6.20) does not truncate, then the wavefunction will have a non-normalisable asymptotic behaviour $v(z) \xrightarrow[|z| \rightarrow \infty]{} e^{-y^2/2} e^{y^2} = e^{y^2/2}$.

Thus we want the series (6.20) to truncate – i.e. we want to fall in the case #1 above. This requirement on the normalisability of the wavefunction is what produces the quantisation of the energy spectrum. In fact the series can only truncate if the rescaled energy ϵ is an odd positive integer

$$\epsilon = 2N + 1, \quad N \in \mathbb{N} \quad (6.30)$$

The integer N correspond with the degree of the polynomial $h(z)$. Recalling how we defined ϵ in (6.14), we find

$$E_N = \hbar\omega \left(N + \frac{1}{2} \right). \quad (6.31)$$

All energies are proportional to $\hbar\omega$, with ω the frequency of the harmonic oscillator. The lowest energy is non-zero $E_0 = \hbar\omega/2$. Furthermore the energy levels are equally spaced, with

$$E_{n+1} - E_n = \hbar\omega. \quad (6.32)$$

6.1.4 The wavefunctions

Now that we settled the problem of finding the spectrum, let us look at the wavefunctions. We will start with the low energy ones. For the purpose of normalising the wavefunctions, we will use the general integration formula

$$\int_{\mathbb{R}} dx x^{2n} e^{-Ax^2/2} = \sqrt{\frac{\pi}{A}} A^{-n} \prod_{l=1}^n \left(l - \frac{1}{2} \right). \quad (6.33)$$

- $N = 0 \implies E_0 = \hbar\omega/2$

The polynomial $h(z)$ is of degree $N = 0$, so it is just a constant. Hence the wavefunction is

$$v_0(z) = h_0 e^{-z^2/2} \underset{\text{norm.}}{\implies} u_0(x) = \left(\frac{m\omega}{\pi\hbar} \right)^{1/4} e^{-\frac{m\omega}{2\hbar}x^2}.$$

- $N = 1 \implies E_0 = 3\hbar\omega/2$

All the coefficients h_n with $n > 1$ need to vanish, so this means that $h_0 = 0$, otherwise we will not be able to cancel any of the even ones. Thus $h(z) = h_1 z$ and the coefficient h_1 is, again, an undetermined constant that we fix by normalisation. We find

$$v_1(z) = h_1 z e^{-z^2/2} \underset{\text{norm.}}{\implies} u_1(x) = \left(\frac{4m^3\omega^3}{\pi\hbar^3} \right)^{1/4} x e^{-\frac{m\omega}{2\hbar}x^2}.$$

- $N = 2 \implies E_0 = 5\hbar\omega/2$

By the same argument as above, $h_1 = 0$ to cancel all the odd terms. We are left with $h(z) = h_0 + h_2 z^2$ and h_2 is determined by the recursion (6.25)

$$h_2 = -2h_0.$$

So we find the wavefunction

$$v_2(z) = h_0(1 - 2z^2) e^{-z^2/2} \underset{\text{norm.}}{\implies} u_2(x) = \left(\frac{m\omega}{4\pi\hbar} \right)^{1/4} \left(1 - 2\frac{m\omega}{\hbar}x^2 \right) e^{-\frac{m\omega}{2\hbar}x^2}.$$

- $N = 3 \implies E_0 = 7\hbar\omega/2$

We proceed just as above, this time with $h_0 = 0$ and $v(x) = h_1x + h_3x$. The coefficient h_3 is determined by (6.25)

$$h_3 = -\frac{2}{3}h_1.$$

Thus the wavefunction is

$$v_3(z) = h_1z\left(1 - \frac{2}{3}z^2\right)e^{-z^2/2} \underset{\text{norm.}}{\implies} u_3(x) = \left(\frac{9m^3\omega^3}{\pi\hbar^3}\right)^{1/4} x \left(1 - \frac{2}{3}\frac{m\omega}{\hbar}x^2\right) e^{-\frac{m\omega}{2\hbar}x^2}.$$

Figure 6.1 displays the four wavefunctions we have derived above.

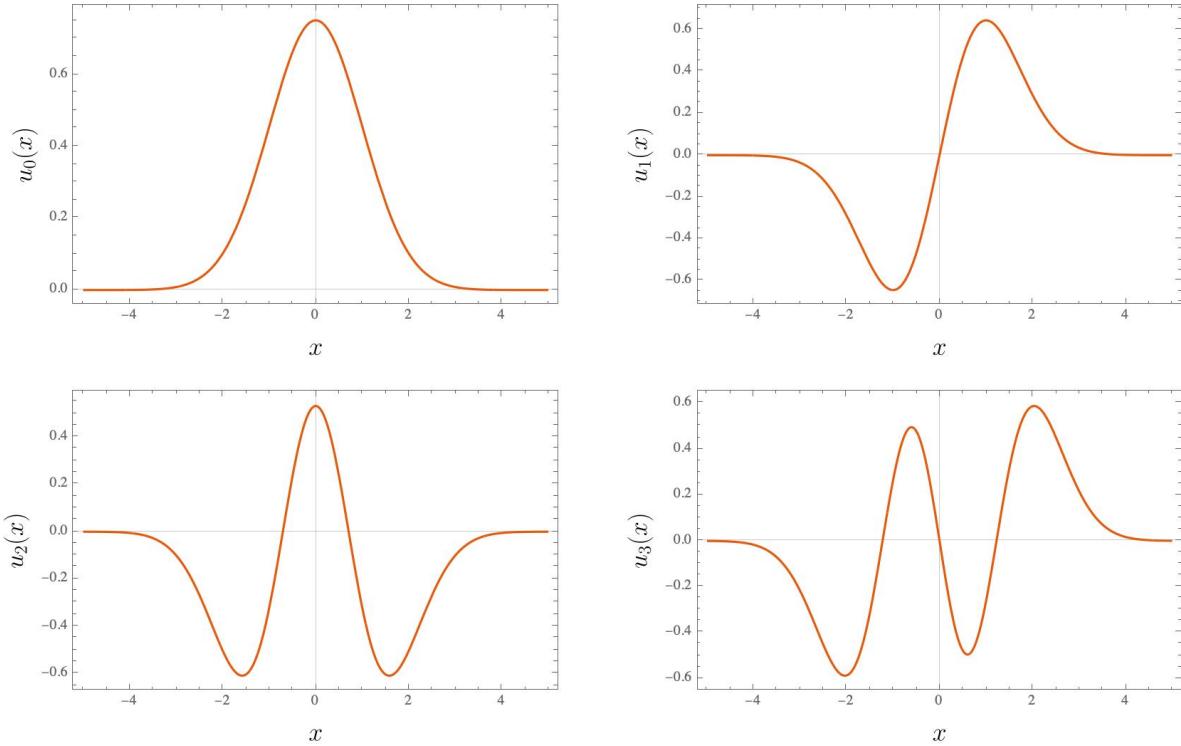


Figure 6.1: First 4 wavefunctions of the quantum harmonic oscillator (6.1). We set $\omega = \hbar/m$ for these specific plots.

More generally, the polynomial solutions $h(z)$ to the equation (6.19), are known as *Hermite polynomials* and are usually denoted with a capital letter $H_n(z)$. They satisfy a number of nice properties that, for the sake of space, we are not going to go through here.

It is interesting to look at the plots of the probability density $|u_n(x)|^2$. They are collected in Figure 6.2 for $n = 0, 1, 2, 3$. We notice that the probability to find the particle becomes larger at the edges of the potential as the energy grows larger. This is perhaps even more clear in the Figure 6.3, where the densities are superimposed with the potential well. This agrees with our classical intuition: a classical particle spends more time at the edge of the potential, simply because it moves more slowly there (near the classical turning points, the momentum is the smallest). Conversely, the ball is less likely to be caught as it whizzes past the bottom of the well. We see that the quantum world is not completely disconnected from the classical. You just have to know where to look.

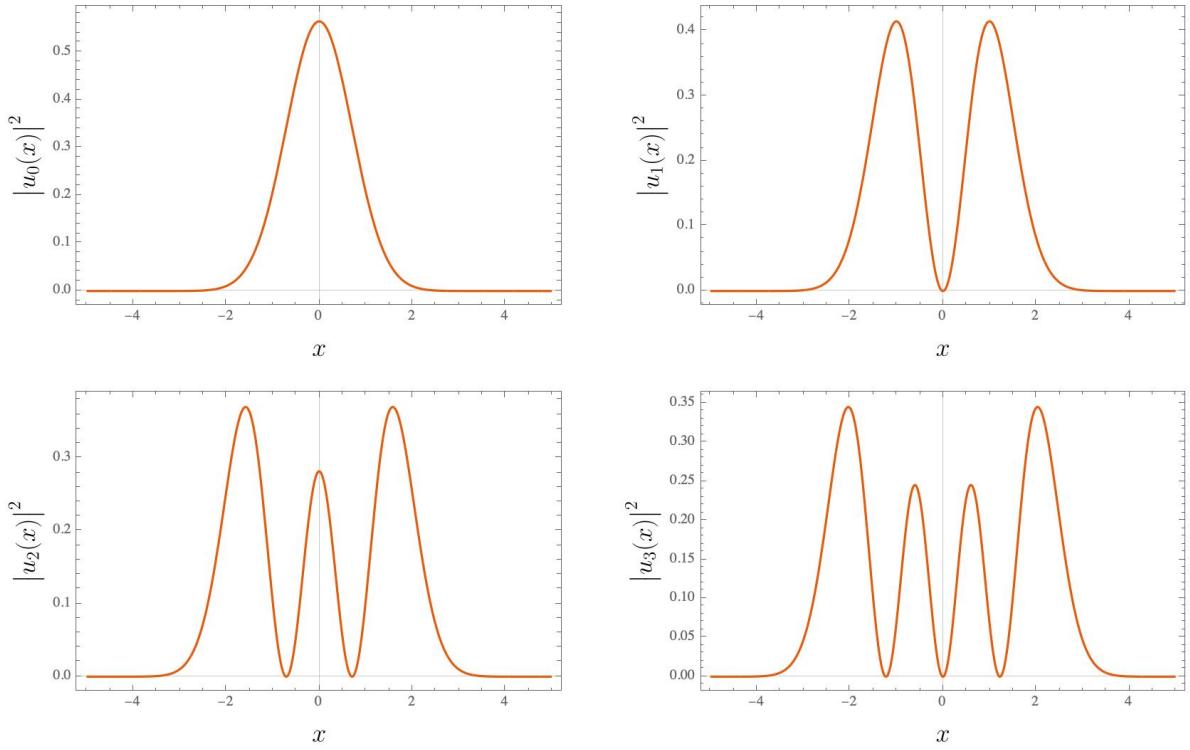


Figure 6.2: Probability densities $|u_n(x)|^2$ for the first 4 states of the harmonic oscillator. We set $\omega = \hbar/m$ for these specific plots.

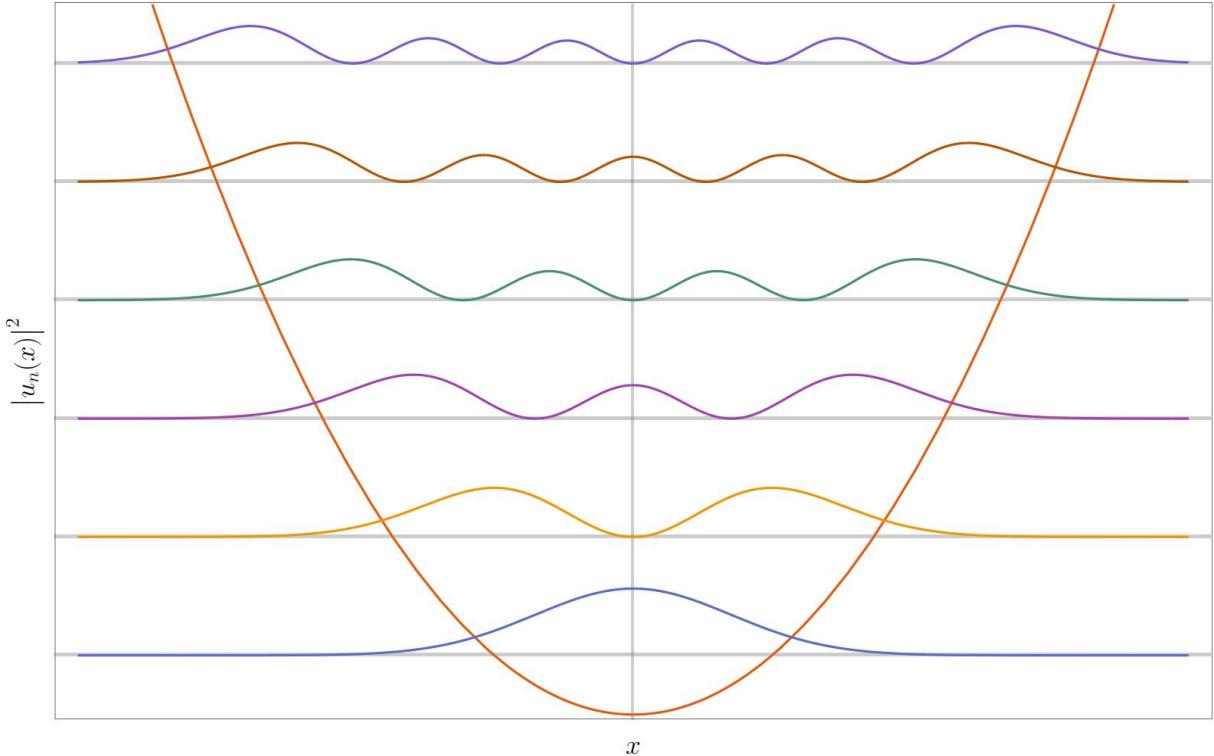


Figure 6.3: Probability densities $|u_n(x)|^2$ for the first 6 states of the harmonic oscillator, superimposed on the potential well. The densities are shifted vertically by $2n + 1$ for clarity. Note that this plot is not in scale.

6.2 The method of Frobenius

In the study of the harmonic oscillator, we expanded in series the solution of a second-order, linear ODE about the regular point $x = 0$. We saw that this expansion is rather nice and we were able to derive exactly the energy spectrum and the wavefunctions. However, the goal of this module is to provide you with the tools to deal with mathematical models that describe real-world physics. In quantum physics, the simplest possible real-world system is the hydrogen atom: one proton together with one electron orbiting it. Its mathematical description and analysis is the ultimate goal of these lectures. Before tackling it, we need to understand how to deal with differential equations that display regular singular points. This is because the force that holds the electron in orbit at a distance r around an atom is the Coulomb one $F \propto -1/r^2$, which arises from a potential $V \propto 1/r$. Thus, in the Schrödinger equation the origin $r = 0$ will be a regular singular point. In this section we are going to see how the series expansion approach we used for the harmonic oscillator extends in the presence of regular singularities. This takes the name of *method of Frobenius*.

Consider the second-order, linear ODE

$$u''(x) + p(x)u'(x) + q(x)u(x) = 0. \quad (6.34)$$

If either of the functions $p(x)$ or $q(x)$ is not analytic at $x = x_0$, then Cauchy theorem does not apply anymore and it seems we cannot rely on power series expansions about $x = x_0$. However, if x_0 is a regular singularity – i.e. if $(x - x_0)p(x)$ and $(x - x_0)^2q(x)$ are analytic – then Ferdinand Georg Frobenius tells us that we can try to find a solution in the following form

$$u(x) = \sum_{n=0}^{\infty} c_n(x - x_0)^{n+\sigma}, \quad (6.35)$$

where $\sigma \in \mathbb{R}$ is a parameter that we will fix later. In truth, it was not Frobenius who came up with the above form of the solutions, it was Lazarus Fuchs that established its existence through his famous theorem⁶.

Theorem 5. Fuchs theorem

Let x_0 be a regular singular point of the second-order, linear ODE

$$u''(x) + p(x)u'(x) + q(x)u(x) = 0.$$

Then a solution $u(x)$ always exists which has the form of a Taylor series times a general power of $x - x_0$:

$$u(x) = \sum_{n=0}^{\infty} c_n(x - x_0)^{n+\sigma}, \quad \sigma \in \mathbb{R}.$$

Hence, we are assured that we will be able to find, at least, one solution in the form (6.35). Let us see this in action.

⁶The contributions of Frobenius to the eponymous method was to show that the coefficients c_n always obey a recurrence relation, to determine the form of the second, linearly independent solution, to prove the convergence of the series and computing their radii of convergence.

Example 6.2.1 Consider the equation

$$2x^2u''(x) + x(2x+1)u'(x) - u(x) = 0. \quad (6.36)$$

Multiplication by $2x^2$ shows that it has the form (6.34), with $p(x) = 1 + 1/(2x)$ and $q(x) = -1/(2x^2)$. So $x = 0$ is clearly a singularity. It is in fact a regular singularity, since $xp(x) = x + 1/2$ and $x^2q(x) = -1/2$ are both analytic about $x = 0$. Let us try to find a solution in the form (6.35) with $x_0 = 0$. The three terms in the equation look as follows when we substitute the series

$$\begin{aligned} 2x^2u''(x) &= 2 \sum_{n=0}^{\infty} (n+\sigma)(n+\sigma-1)c_n x^{n+\sigma}, \\ x(2x+1)u'(x) &= 2 \sum_{n=0}^{\infty} (n+\sigma)c_n x^{n+\sigma+1} + \sum_{n=0}^{\infty} (n+\sigma)c_n x^{n+\sigma}, \\ -u(x) &= - \sum_{n=0}^{\infty} c_n x^{n+\sigma}. \end{aligned} \quad (6.37)$$

We need to shift the first series in the second expression above, so that the powers of x agree between all the terms

$$2 \sum_{n=0}^{\infty} (n+\sigma)c_n x^{n+\sigma+1} = 2 \sum_{n=1}^{\infty} (n+\sigma-1)c_{n-1} x^{n+\sigma}. \quad (6.38)$$

Now we extract from the sums the $n = 0$ terms, so that, when we sum the three terms, we can clump the summation signs all up together

$$\begin{aligned} 2x^2u''(x) &= 2\sigma(\sigma-1)c_0 x^\sigma + 2 \sum_{n=1}^{\infty} (n+\sigma)(n+\sigma-1)c_n x^{n+\sigma}, \\ x(2x+1)u'(x) &= \sigma c_0 x^\sigma + \sum_{n=1}^{\infty} \left[2(n+\sigma-1)c_{n-1} + (n+\sigma)c_n \right] x^{n+\sigma}, \\ -u(x) &= -c_0 x^\sigma - \sum_{n=1}^{\infty} c_n x^{n+\sigma}. \end{aligned} \quad (6.39)$$

The sum of these terms – which gives 0 due to the differential equation – reads

$$(2\sigma+1)(\sigma-1)c_0 x^\sigma + \sum_{n=1}^{\infty} \left[2(n+\sigma-1)c_{n-1} + (2n+2\sigma+1)(n+\sigma-1)c_n \right] x^{n+\sigma} = 0. \quad (6.40)$$

Since x is generic, all the coefficients must vanish independently and we arrived at the following equations

$$\begin{aligned} (2\sigma+1)(\sigma-1)c_0 &= 0, \\ 2(n+\sigma-1)c_{n-1} &= -(2n+2\sigma+1)(n+\sigma-1)c_n. \end{aligned} \quad (6.41)$$

Note that the recursion relation above is next-neighbour. This means that if $c_0 = 0$, then $c_n = 0, \forall n$. We do not want that, since it yields the trivial solution $u(x) = 0$. Hence we need to impose

$$(2\sigma+1)(\sigma-1) = 0 \implies \sigma = \begin{cases} 1 \\ -1/2 \end{cases}. \quad (6.42)$$

The above equation for σ is known as *indicial equation* and, usually, yields two distinct values for σ , one for each solution. We return to the linear independence of the solutions momentarily. Let us conclude our work here.

The recursion relation can be simplified

$$c_n = -\frac{1}{n + \sigma + 1/2} c_{n-1}, \quad \forall n \in \mathbb{N}. \quad (6.43)$$

It is easy to solve: we just iterate

$$\begin{aligned} c_n &= \left(-\frac{1}{n + \sigma + 1/2}\right) \left(-\frac{1}{n + \sigma - 1/2}\right) c_{n-2} = \\ &= \left(-\frac{1}{n + \sigma + 1/2}\right) \left(-\frac{2}{n + \sigma - 1/2}\right) \left(-\frac{2}{n + \sigma - 3/2}\right) c_{n-3} = \dots, \end{aligned} \quad (6.44)$$

until we reach c_0

$$c_n = \frac{(-1)^n}{(\sigma + 3/2)(\sigma + 3/2 + 1) \cdots (\sigma + 3/2 + n - 1)} c_0. \quad (6.45)$$

Note that the radius of convergence is easily extracted from the recursion relation to be infinite

$$\lim_{n \rightarrow \infty} \left| \frac{c_{n-1}}{c_n} \right| = \lim_{n \rightarrow \infty} \left| n + \sigma + \frac{1}{2} \right| = \infty \quad (6.46)$$

Now let us look at the two cases $\sigma = -1/2$ and $\sigma = 1$.

- $\sigma = -1/2$

Substituting, we find

$$c_n = \frac{(-1)^n}{1 \cdot 2 \cdot 3 \cdots n} c_0 = \frac{(-1)^n}{n!} c_0, \quad (6.47)$$

which gives us immediately the solution

$$u_1(x) = c_0 x^{-1/2} \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} x^n = c_0 x^{-1/2} e^{-x}. \quad (6.48)$$

- $\sigma = 1$

Again, we simply substitute

$$c_n = \frac{(-1)^n}{5/2 \cdot 7/2 \cdot 9/2 \cdots \frac{2n+3}{2}} c_0 = \frac{3}{3} \frac{2^n (-1)^n}{5 \cdot 7 \cdot 9 \cdots (2n+3)} c_0 = \frac{3 \times (-2)^n}{(2n+3)!!} c_0, \quad (6.49)$$

where we introduced the *double factorial function*

$$(2k-1)!! \doteq (2k-1)(2k-3)(2k-5) \cdots 1, \quad \forall k \in \mathbb{N}. \quad (6.50)$$

Hence the second solution takes the following form

$$u_2(x) = 3c_0 \sum_{n=0}^{\infty} \frac{(-2)^n}{(2n+3)!!} x^{n+1}. \quad (6.51)$$

This series can in principle be resummed to a closed form function, though you would not recognise it as it is not a common one⁷.

⁷For the curious ones, it involves the *error function* which is the integral of a gaussian exponential on an interval $[0, x]$.

This example showed us that not only does the series expansion (6.35) constitutes a sensible solution of (6.34) for some specific value of σ – which is what Fuchs theorem ensures us of – but it also allows us to find a second *linearly independent* solution. This was, as it turns out, somewhat a stroke of luck. In fact, the following statement holds true

Given a second-order linear ODE (6.34) with regular singularity at x_0 , if the roots σ_1 and σ_2 of the indicial equation are such that $\sigma_1 - \sigma_2 \notin \mathbb{Z}$, then there exist two linearly independent solution admitting a series expansion (6.35).

If, otherwise, $\sigma_1 - \sigma_2 \in \mathbb{N} \cup \{0\}$ – where we chose to order the roots such that $\sigma_1 \geq \sigma_2$ – then there exist a power series solution

$$u_1(x) = \sum_{n=0}^{\infty} c_n (x - x_0)^{n+\sigma_1},$$

together with a second solution with the following form

$$u_2(x) = \alpha u_1(x) \log(x) + \sum_{n=0}^{\infty} \tilde{c}_n (x - x_0)^{n+\sigma_2},$$

where $\alpha \in \mathbb{R}$ can be determined as a function of $\tilde{c}_0 \neq 0$ and $c_0 \neq 0$.

Let us look at another example.

Example 6.2.2 Now consider the following equation

$$xu''(x) + 2u'(x) + u(x) = 0, \quad (6.52)$$

for which $p(x) = 2/x$ and $q(x) = 1/x$. Again, $x = 0$ is a regular singularity, so we search for a solution in the form (6.35). Substitution yields (we multiply every term by x , to make things somewhat simpler)

$$\begin{aligned} x^2 u''(x) &= \sum_{n=0}^{\infty} (n+\sigma)(n+\sigma-1) c_n x^{n+\sigma}, \\ 2xu'(x) &= 2 \sum_{n=0}^{\infty} (n+\sigma) c_n x^{n+\sigma}, \\ xu(x) &= \sum_{n=0}^{\infty} c_n x^{n\sigma+1} = \sum_{n=1}^{\infty} c_{n-1} x^{n+\sigma}. \end{aligned} \quad (6.53)$$

As before, we isolate the $n = 0$ terms and combine the summations in a single one

$$\sigma(\sigma+1)c_0x^\sigma + \sum_{n=1}^{\infty} [(n+\sigma)(n+\sigma+1)c_n + c_{n-1}] x^{n+\sigma} = 0. \quad (6.54)$$

So we arrive at the indicial equation

$$\sigma(\sigma+1) = 0, \quad (6.55)$$

whose roots $\sigma = 0, -1$ differ by an integer. We should then fall in the second case of the above statement: the second solution should have a logarithmic term! Before dealing with this, let us look at the first solution, with $\sigma_1 = 0$. The recursion relation is

$$c_{n+1} = -\frac{1}{(n+1)(n+2)} c_n , \quad \forall n \geq 0 . \quad (6.56)$$

It is easy to see that the radius of convergence is infinite and that the solution $u_1(x)$ is⁸

$$u_1(x) = c_0 \sum_{n=0}^{\infty} \frac{(-1)^n}{n!(n+1)!} x^n . \quad (6.57)$$

Now, we can immediately see that, when $\sigma = \sigma_2 = -1$, the recursion becomes ill-defined at $n = 0$:

$$c_{n+1} = -\frac{1}{n(n+1)} c_n . \quad (6.58)$$

Setting $c_0 = 0$ and taking c_1 as the normalisation constant only reproduces $u_1(x)$, so that is not good. Let us try to work out the expression with a logarithm we saw above. In other words, take the ansatz

$$u_2(x) = \alpha \log(x) u_1(x) + v_2(x) , \quad (6.59)$$

where $v_2(x)$ is a power series (6.35) with $\sigma = \sigma_2 = -1$. Substitution in the differential equation gives

$$\begin{aligned} & \alpha \log x \left[x^2 u_1''(x) + 2xu_1'(x) + xu_1(x) \right] + 2\alpha x u_1'(x) + \alpha u_1(x) + \\ & + x^2 v_2''(x) + 2xv_2'(x) + xv_2(x) = 0 . \end{aligned} \quad (6.60)$$

The sum in square bracket vanishes since we took $u_1(x)$ to be a solution to the differential equation, so we are left with the non-homogeneous equation for $v_2(x)$

$$x^2 v_2''(x) + 2xv_2'(x) + xv_2(x) + 2\alpha x u_1'(x) + \alpha u_1(x) = 0 . \quad (6.61)$$

Now, let us insert in v_2 and u_1 their respective series expansions. After some simple manipulations, we arrive at

$$\sum_{n=0}^{\infty} \left[n(n+1)\tilde{c}_{n+1} + \tilde{c}_n \right] x^n + \sum_{n=0}^{\infty} \alpha(2n+1) \frac{(-1)^n}{n!(n+1)!} c_0 x^n = 0 . \quad (6.62)$$

Hence the recursion relation now splits into an equation for $n = 0$ and another valid for $n > 0$

$$\begin{aligned} \tilde{c}_0 &= -\alpha c_0 , \\ \tilde{c}_{n+1} &= -\frac{1}{n(n+1)} \tilde{c}_n - \alpha c_0 (-1)^n \frac{1}{n(n+1)} \frac{2n+1}{n!(n+1)!} , \quad \forall n > 0 . \end{aligned} \quad (6.63)$$

We see that $\alpha = -\tilde{c}_0/c_0$ is determined, as predicted, in terms of c_0 and \tilde{c}_0 . We will not be solving this new recursion.

⁸For the curious ones, this is the function $J_1(2\sqrt{x})/\sqrt{x}$, where $J_1(x)$ is the Bessel J function of order 1 we met when studying vibrations on a circular membrane.

Outro: takeaway points

In this chapter, we looked at power series methods to solve second-order, linear ODEs. In particular

- We looked at the prototypical example of the harmonic oscillator, whose quantum mechanical version is described by the ODE

$$-\frac{\hbar^2}{2m}u''(x) + \frac{1}{2}m\omega^2x^2u(x) = Eu(x) .$$

Thanks to some re-definitions

$$z = \sqrt{\frac{m\omega}{\hbar}}x , \quad \epsilon = 2\frac{mE}{\hbar^2}\frac{\hbar}{m\omega} = 2\frac{E}{\hbar\omega} , \quad v(z) \doteq u(x(z)) ,$$

we transformed the ODE into the following

$$v''(z) + (\epsilon - z^2)v(z) = 0 .$$

We saw that solutions which are normalisable take the form

$$v(z) = e^{-z^2/2}h(z) ,$$

where $h(z)$ is a polynomial satisfying the equation

$$h''(z) - 2zh'(z) + (\epsilon - 1)h(z) = 0 .$$

- We reminded the concept of an *analytic function* about a point $x = x_0$ as a function which is equal to its convergent Taylor expansion in an interval about the point x_0 . Then we introduced a classification of the singularities of an ODE

$$u''(x) + p(x)u'(x) + q(x)u(x) = 0 ,$$

as follows. A point $x = x_0$ is

- an *ordinary point* if $p(x)$ and $q(x)$ are analytic about x_0 ;
- a *regular singular point* if $(x - x_0)p(x)$ and $(x - x_0)^2q(x)$ are analytic about x_0 ;
- an *irregular singular point* otherwise.

In the neighbourhood of an ordinary point, a second-order, linear ODE admits two independent solutions $u_i(x)$ with convergent power series expansion

$$u_i(x) = \sum_{n=0}^{\infty} u_{i,n}(x - x_0)^n ,$$

having radii of convergence $R_i \geq \min(R_p, R_q)$ (where R_p and R_q are, respectively, the radii of convergence of the Taylor expansions of $p(x)$ and $q(x)$). This is *Cauchy theorem*.

- Equipped with the above machinery, we looked at the power series expansion of $h(z)$ about the regular point $z = 0$

$$h(z) = \sum_{n=0}^{\infty} h_n z^n ,$$

finding the recursion relation for the coefficients

$$h_{n+2} = \frac{2n+1-\epsilon}{(n+1)(n+2)} h_n , \quad \forall n \geq 2 .$$

With this we determined the radius of convergence to be infinite and, requiring the series to truncate to some integer N , we determined the quantisation of the energy

$$\epsilon = 2N + 1 \implies E_N = \hbar\omega \left(N + \frac{1}{2} \right) , \quad N \in \mathbb{N} .$$

We also computed the first few wavefunctions and checked that they display exactly N zeros in the interval $x \in (-\infty, \infty)$. The polynomial solutions of the equation for $h(z)$ are called *Hermite polynomials* and are usually denoted by $H_n(z)$.

- Then we moved to the analysis of solutions to second-order, linear ODEs in the vicinity of a regular singular point. We introduced the method of Frobenius which relies on Fuchs theorem. This states that if x_0 is a regular singular point of

$$u''(x) + p(x)u'(x) + q(x)u(x) = 0 ,$$

then a solution $u(x)$ always exists that takes the power series expansion

$$u(x) = \sum_{n=0}^{\infty} c_n (x - x_0)^{n+\sigma} , \quad \sigma \in \mathbb{R} .$$

- We checked the Fuchs theorem in an example equation, finding that the value of σ is determined by an *indicial equation* of order 2, having, consequently, 2 solutions. In our example, the two choices of σ led to two independent solutions of the ODE. However this is not the usual situation. We stated a theorem without proof: if the solutions σ_1 and σ_2 to the indicial equation are such that $\sigma_1 - \sigma_2 \notin \mathbb{Z}$, then the functions

$$u_i(x) = \sum_{n=0}^{\infty} c_{i,n} (x - x_0)^{n+\sigma_i} ,$$

are independent (note that the coefficients depend on σ ! Hence the added index i). Otherwise, if $\sigma_1 - \sigma_2 \in \mathbb{N} \cup \{0\}$, one solution is

$$u_1(x) = \sum_{n=0}^{\infty} c_n (x - x_0)^{n+\sigma_1} ,$$

and an independent solution has the form

$$u_2(x) = \alpha u_1(x) \log(x) + \sum_{n=0}^{\infty} \tilde{c}_n (x - x_0)^{n+\sigma_2} .$$

Chapter 7

A quantum particle in 3 dimensions

In this last chapter on quantum dynamics, we are going to study the (time-independent) Schrödinger equation in 3 dimensions. Its form is the same as the one we studied up to now, only with a $\vec{\nabla}$ differential operator in place of the derivative $\partial/\partial x$

$$-\frac{\hbar^2}{2m}\nabla^2 u(\vec{x}) + V(\vec{x})u(\vec{x}) = Eu(\vec{x}) . \quad (7.1)$$

We will limit our attention to *central potentials*, i.e. such that

$$V(\vec{x}) = V(r) , \quad r \doteq |\vec{x}| , \quad (7.2)$$

since they are the simplest ones to study and, thankfully, also the most useful in physics.

It is worth to spend a moment to recall how such potentials are approached in classical mechanics: everything becomes much simpler when the *angular momentum* \vec{L} is employed

$$\vec{L} = \vec{x} \times \vec{p} . \quad (7.3)$$

The reason for such a simplification is that \vec{L} is a *conserved quantity*, meaning it does not change with time¹. In classical mechanics, one makes use of the conservation of \vec{L} twice. First, the fact that the direction of \vec{L} is fixed allows the reduction of the problem from 3 to 2 dimensions, since the particle only moves on the plane determined by $\vec{L} \cdot \vec{x} = 0$. Then one uses the conservation of the magnitude $|\vec{L}|$ to further reduce the problem to a single dimension: the motion in the radial direction. So the fact that \vec{L} is a conserved quantity is such a powerful feature that allows the reduction of a complicated-looking 3-dimensional problem to a much simpler – often solvable – 1-dimensional one. As we will see, this strategy works, with some modifications, in quantum mechanics as well.

7.1 Separation of variables, but quantum

Remember when we dealt with vibrating membranes, we used the separation of variables method to reduce a 3-dimensional PDE to 3 independent ODEs, one for each variable t ,

¹If you do not remember how this goes, here is a refresher. $\dot{\vec{L}} = \dot{\vec{x}} \times \vec{p} + \vec{x} \times \dot{\vec{p}}$. But $\vec{p} = m\vec{v} = m\dot{\vec{x}}$, meaning \vec{p} is parallel to \vec{x} , so their cross product vanishes. Similarly, $\dot{\vec{p}}$ is parallel to \vec{x} , since $\dot{\vec{p}} = -\vec{\nabla}V(r)$ and the latter is parallel to \vec{x} since the potential is central.

x and y . Here we are going to do essentially the same thing for the Schrödinger equation (7.1). However, we will follow a different route that will highlight the physical principles allowing such a separation to happen in the first place.

It all starts, again, with the *angular momentum*. We are now in the realm of quantum mechanics, so we must remember that observables are operators. We follow the recipe (4.38) and write

$$\hat{\vec{L}} = \hat{\vec{x}} \times \hat{\vec{p}}, \quad (7.4)$$

and remember that the action on wavefunctions is

$$\hat{\vec{x}}u(\vec{x}) = \vec{x}u(\vec{x}), \quad \hat{\vec{p}}u(\vec{x}) = -i\hbar\vec{\nabla}u(\vec{x}). \quad (7.5)$$

We can then easily compute the components of the operator $\hat{\vec{L}}$

$$\begin{aligned} \hat{L}_1 u(\vec{x}) &= -i\hbar \left(x_2 \frac{\partial}{\partial x_3} - x_3 \frac{\partial}{\partial x_2} \right) u(\vec{x}), \\ \hat{L}_2 u(\vec{x}) &= -i\hbar \left(x_3 \frac{\partial}{\partial x_1} - x_1 \frac{\partial}{\partial x_3} \right) u(\vec{x}), \\ \hat{L}_3 u(\vec{x}) &= -i\hbar \left(x_1 \frac{\partial}{\partial x_2} - x_2 \frac{\partial}{\partial x_1} \right) u(\vec{x}). \end{aligned} \quad (7.6)$$

Usually these relations are summarised with the help of the *epsilon symbol* ϵ_{ijk} , an object with three indices taking values in the set $\{1, 2, 3\}$, whose definition is

$$\epsilon_{123} = 1, \quad \epsilon_{ijk} = -\epsilon_{jik} = -\epsilon_{kji} = -\epsilon_{ikj}. \quad (7.7)$$

In words, it is totally a-symmetric in its indices and its entry 1, 2, 3 is conventionally taken to be 1. The action of the angular momentum operator is easily seen to be writable as

$$\hat{L}_i u(\vec{x}) = -i\hbar \sum_{j,k=1}^3 \epsilon_{ijk} x_j \frac{\partial}{\partial x_k} u(\vec{x}) \underset{\text{convention}}{\equiv} -i\hbar \epsilon_{ijk} x_j \frac{\partial}{\partial x_k} u(\vec{x}). \quad (7.8)$$

In the right-most side above, we introduced another common piece of notation: the *Einstein summation convention*. Whenever you see two repeated indices in a product, you should assume that the product is to be summed over the values of that repeated index. Let us make some examples:

$$x_i x_j = x_i x_j, \quad x_i p_k Q_i^k = \sum_{i,k=1}^3 x_i Q_i^k p_k, \quad A_3 B_3 x_k \alpha_k \gamma_j^i q_j = A_3 B_3 \sum_{k=1}^3 \alpha_k x_k \sum_{j=1}^3 \gamma_j^i q_j.$$

In the third example $A_3 B_3$ are not to be summed because the indices is fixed to the value 3, it just so happen to be the same for A and B . For all this chapter, indices i, j, k, l will only take values in the set $\{1, 2, 3\}$.

You might complain that the definition (7.4) suffers from ordering ambiguities, the kind we mentioned in §§4.2.2. What if we define the angular momentum operator as

$$\hat{\vec{L}} = -\hat{\vec{p}} \times \hat{\vec{x}}? \quad (7.9)$$

Well, as you may check yourselves, $\hat{\vec{x}} \times \hat{\vec{p}} \equiv -\hat{\vec{p}} \times \hat{\vec{x}}$, unlike what happens with, say, the scalar product (4.40). So the above definition is identical to (7.4).

7.1.1 Commutation relations

The components of the angular momentum operator (7.6) do not commute amongst themselves²

$$[\hat{L}_i, \hat{L}_j] \doteq \hat{L}_i \hat{L}_j - \hat{L}_j \hat{L}_i = i\hbar \epsilon_{ijk} \hat{L}_k . \quad (7.10)$$

Here we defined the *commutator* $[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A}$ of two operators \hat{A} and \hat{B} . From a physical perspective, the fact that the components of the angular momentum fail to commute means that a quantum particle cannot have a well defined angular momentum in all three directions simultaneously. If, for example, you know the angular momentum L_1 then there will necessarily be some uncertainty in the angular momentum in the other two directions. This is precisely the same thing that happens with position and momentum. In that case the commutator $[\hat{x}_i, \hat{p}_j] = i\hbar \delta_{ij}$ is telling us that we cannot know with infinite precision both, say, x_1 and p_1 ; the uncertainty principle (4.42) kicks in.

The commutation relations (7.10) are simple to prove. Let us look at $i = 1$ and $j = 2$, so we want to prove

$$[\hat{L}_1, \hat{L}_2] = i\hbar \epsilon_{123} \hat{L}_3 \equiv i\hbar \hat{L}_3 . \quad (7.11)$$

Let us act with the right-hand side on a wavefunction

$$\begin{aligned} [\hat{L}_1, \hat{L}_2] u(\vec{x}) &= (-i\hbar)^2 \left(x_2 \frac{\partial}{\partial x_3} - x_3 \frac{\partial}{\partial x_2} \right) \left(x_3 \frac{\partial}{\partial x_1} - x_1 \frac{\partial}{\partial x_3} \right) u(\vec{x}) + \\ &\quad - (-i\hbar)^2 \left(x_3 \frac{\partial}{\partial x_1} - x_1 \frac{\partial}{\partial x_3} \right) \left(x_2 \frac{\partial}{\partial x_3} - x_3 \frac{\partial}{\partial x_2} \right) u(\vec{x}) = \\ &= -\hbar^2 \left(x_2 \frac{\partial}{\partial x_3} x_3 \frac{\partial u(\vec{x})}{\partial x_1} - x_1 x_2 \cancel{\frac{\partial^2 u(\vec{x})}{\partial x_3^2}} - x_3^2 \cancel{\frac{\partial^2 u(\vec{x})}{\partial x_1 \partial x_2}} + x_1 x_3 \frac{\partial^2 u(\vec{x})}{\partial x_2 \partial x_3} \right) + \\ &\quad + \hbar^2 \left(x_2 x_3 \frac{\partial^2 u(\vec{x})}{\partial x_1 \partial x_3} - \cancel{x_3^2 \frac{\partial^2 u(\vec{x})}{\partial x_1 \partial x_2}} - x_1 x_2 \cancel{\frac{\partial^2 u(\vec{x})}{\partial x_3^2}} + x_1 \frac{\partial}{\partial x_3} x_3 \frac{\partial u(\vec{x})}{\partial x_2} \right) = \quad (7.12) \\ &= -\hbar^2 \left(x_2 \frac{\partial u(\vec{x})}{\partial x_1} + x_2 x_3 \cancel{\frac{\partial^2 u(\vec{x})}{\partial x_1 \partial x_3}} + x_1 x_3 \cancel{\frac{\partial^2 u(\vec{x})}{\partial x_2 \partial x_3}} \right) + \\ &\quad + \hbar^2 \left(x_2 x_3 \cancel{\frac{\partial^2 u(\vec{x})}{\partial x_1 \partial x_3}} + x_1 x_3 \cancel{\frac{\partial u(\vec{x})}{\partial x_2 \partial x_3}} + x_1 \frac{\partial u(\vec{x})}{\partial x_2} \right) = \\ &= i\hbar(-i\hbar) \left(x_1 \frac{\partial}{\partial x_2} - x_2 \frac{\partial}{\partial x_1} \right) u(\vec{x}) \equiv i\hbar \hat{L}_3 u(\vec{x}) . \end{aligned}$$

We did not specify the wavefunction $u(\vec{x})$ so this result must hold in full generality, confirming the commutation relations (7.10). You can verify the other ones yourselves.

Another very important quantity is the magnitude of the angular momentum. In quantum mechanics it is more practical to work with \hat{L}^2 rather than $|\hat{L}|$ and one refers to \hat{L}^2 as

²These commutation relations are very famous in mathematics. They appear in many places that have nothing to do with quantum mechanics and define what is known as a *Lie algebra*, in this case the algebra \mathfrak{su}_2 , which, in some sense, distil the “essence of rotations in 3D”. There are some subtleties involved here that open up a number of weird and wonderful loopholes that the quantum world exploits. For example, the fact that there are particles called fermions that don’t come back to themselves when rotated by 2π .

the *total angular momentum operator*

$$\hat{L}^2 = \hat{L}_1^2 + \hat{L}_2^2 + \hat{L}_3^2 . \quad (7.13)$$

Perhaps surprisingly, it commutes with all the angular momentum components

$$[\hat{L}^2, \hat{L}_i] = 0 , \quad \forall i \in \{1, 2, 3\} . \quad (7.14)$$

This means that a quantum system can have a definite total angular momentum along with a definite component of the angular momentum in some chosen reference direction, usually taken to be L_3 .

We can prove (7.14) rather straightforwardly thanks to the following identity, valid for any pair of operators \hat{A} and \hat{B}

$$[\hat{A}^2, \hat{B}] = \hat{A} [\hat{A}, \hat{B}] + [\hat{A}, \hat{B}] \hat{A} , \quad (7.15)$$

the proof of which is left to you as an exercise. With this, we can prove, e.g.

$$[\hat{L}^2, \hat{L}_1] = [\hat{L}_1^2, \hat{L}_1] + [\hat{L}_2^2, \hat{L}_1] + [\hat{L}_3^2, \hat{L}_1] = 0 . \quad (7.16)$$

The first term in the middle is automatically zero since, clearly, any object commutes with itself. For the second we see that

$$\begin{aligned} [\hat{L}_2^2, \hat{L}_1] &= \hat{L}_2 [\hat{L}_2, \hat{L}_1] + [\hat{L}_2, \hat{L}_1] \hat{L}_2 = -\hat{L}_2 [\hat{L}_1, \hat{L}_2] - [\hat{L}_1, \hat{L}_2] \hat{L}_2 = \\ &= -i\hbar (\hat{L}_2 \hat{L}_3 + \hat{L}_3 \hat{L}_2) , \end{aligned} \quad (7.17)$$

where we used the obvious fact that $[\hat{A}, \hat{B}] = -[\hat{B}, \hat{A}]$. We also see that

$$\begin{aligned} [\hat{L}_3^2, \hat{L}_1] &= \hat{L}_3 [\hat{L}_3, \hat{L}_1] + [\hat{L}_3, \hat{L}_1] \hat{L}_3 = i\hbar \epsilon_{312} (\hat{L}_3 \hat{L}_2 + \hat{L}_2 \hat{L}_3) = \\ &= i\hbar (\hat{L}_3 \hat{L}_2 + \hat{L}_2 \hat{L}_3) = \\ &= -[\hat{L}_2^2, \hat{L}_1] , \end{aligned} \quad (7.18)$$

since $\epsilon_{312} = -\epsilon_{132} = \epsilon_{123} = 1$. Hence, we verified (7.16). The other commutation relations are proved in the same exact way.

There is one last commutation relation we want to look at: the angular momentum vs. the Hamiltonian. Remember, the Hamiltonian operator is, in our case, the left-hand side of the Schrödinger equation (7.1)

$$\hat{H} u(\vec{x}) = \left(-\frac{\hbar^2}{2m} \nabla^2 + V(r) \right) u(\vec{x}) . \quad (7.19)$$

This operator commutes with the angular momentum operator's components

$$[\hat{H}, \hat{L}_i] = [\hat{H}, \hat{L}^2] = 0 , \quad (7.20)$$

and it does so because the potential is central. If this was not the case, the Hamiltonian would fail to commute with the angular momentum³. The proof of (7.20) is not hard: it follows the same pattern we have seen above. We can proceed in steps. First you can prove that

$$[\hat{L}_i, \hat{x}_j] = i\hbar\epsilon_{ijk}\hat{x}_k, \quad [\hat{L}_i, \hat{p}_j] = i\hbar\epsilon_{ijk}\hat{p}_k. \quad (7.21)$$

and then using the identity (7.15) you can show

$$\begin{aligned} [\hat{L}_i, \hat{\vec{x}}^2] &= [\hat{L}_i, \hat{x}_1^2] + [\hat{L}_i, \hat{x}_2^2] + [\hat{L}_i, \hat{x}_3^2] = 0, \\ [\hat{L}_i, \hat{\vec{p}}^2] &= [\hat{L}_i, \hat{p}_1^2] + [\hat{L}_i, \hat{p}_2^2] + [\hat{L}_i, \hat{p}_3^2] = 0. \end{aligned} \quad (7.22)$$

These relations are all we need, since the hamiltonian

$$\hat{H} = \frac{\hat{\vec{p}}^2}{2m} + V(\hat{\vec{x}}^2) \quad (7.23)$$

only depends on $\hat{\vec{p}}^2$ and $\hat{\vec{x}}^2$.

Let us pause a moment and think of what all these commutation mean. Between the Hamiltonian and the angular momentum, we have at most three operators that all commute amongst themselves: the Hamiltonian \hat{H} , the angular momentum squared $\hat{\vec{L}}^2$ and one angular momentum component, conventionally taken to be \hat{L}_3 . Why do we care so much that these operator commute? It has to do with the following theorem

Theorem 6. Common eigenfunctions of commuting operators

Two operators commute if and only if they share the same eigenfunctions.

We are not going to prove this. But now it is clearer why commutations matter, especially with the Hamiltonian. In fact, we want a solution of the Schrödinger equation (7.1) that, in terms of the Hamiltonian, reads

$$\hat{H}u(\vec{x}) = Eu(\vec{x}). \quad (7.24)$$

Hence any solution is automatically an eigenfunction of the Hamiltonian. But since \hat{H} commutes with both $\hat{\vec{L}}^2$ and with \hat{L}_3 and these commute amongst themselves, the solutions we are looking for can be taken to be simultaneous eigenfunctions of these three operators

$$\begin{aligned} \hat{H}u_{m,l,E}(\vec{x}) &= Eu_{m,l,E}(\vec{x}), \\ \hat{\vec{L}}^2u_{m,l,E}(\vec{x}) &= l(l+1)\hbar^2u_{m,l,E}(\vec{x}), \\ \hat{L}_3u_{m,l,E}(\vec{x}) &= m\hbar u_{m,l,E}(\vec{x}). \end{aligned} \quad (7.25)$$

The reason why we chose to parametrise the eigenvalues of $\hat{\vec{L}}^2$ as $l(l+1)$ will be clearer momentarily.

³This parallels the situation in classical mechanics. There the angular momentum is only useful when dealing with central potentials because only then is the angular momentum conserved in time.

7.1.2 The eigenfunctions

The task is now to find these eigenfunctions. Since we are working in a setting that enjoys spherical symmetry, it is only natural to use spherical coordinates

$$\begin{cases} x_1 = r \sin \theta \cos \phi \\ x_2 = r \sin \theta \sin \phi \\ x_3 = r \cos \theta \end{cases}, \quad \begin{cases} \theta \in [0, \pi] \\ \phi \in [0, 2\pi] \end{cases}. \quad (7.26)$$

We need to convert the expressions (7.6) in terms of the coordinates (ρ, θ, ϕ) and to do so, we simply apply the chain rule. For example

$$\frac{\partial}{\partial x_1} f(r, \theta, \phi) = \frac{\partial r}{\partial x_1} \frac{\partial}{\partial r} f(r, \theta, \phi) + \frac{\partial \theta}{\partial x_1} \frac{\partial}{\partial \theta} f(r, \theta, \phi) + \frac{\partial \phi}{\partial x_1} \frac{\partial}{\partial \phi} f(r, \theta, \phi). \quad (7.27)$$

The partial derivatives $\partial r / \partial x_1$ and so on can be found by deriving the definitions (7.26) with respect to x_1

$$\begin{aligned} (A) \quad 1 &= \sin \theta \cos \phi \frac{\partial r}{\partial x_1} + r \cos \theta \cos \phi \frac{\partial \theta}{\partial x_1} - r \sin \theta \sin \phi \frac{\partial \phi}{\partial x_1}, \\ (B) \quad 0 &= \sin \theta \sin \phi \frac{\partial r}{\partial x_1} + r \cos \theta \sin \phi \frac{\partial \theta}{\partial x_1} + r \sin \theta \cos \phi \frac{\partial \phi}{\partial x_1}, \\ (C) \quad 0 &= \cos \theta \frac{\partial r}{\partial x_1} - r \sin \theta \frac{\partial \theta}{\partial x_1}. \end{aligned} \quad (7.28)$$

Taking the combination

$$[(A) \cos \phi + (B) \sin \phi] \sin \theta + (C) \cos \theta, \quad (7.29)$$

we obtain

$$\frac{\partial r}{\partial x_1} = \sin \theta \cos \phi. \quad (7.30)$$

The other partial derivatives can be found in very similar way. The end result is

$$\begin{aligned} \hat{L}_1 &= i\hbar \left(\cot \theta \cos \phi \frac{\partial}{\partial \phi} + \sin \phi \frac{\partial}{\partial \theta} \right), \\ \hat{L}_2 &= i\hbar \left(\cot \theta \sin \phi \frac{\partial}{\partial \phi} - \cos \phi \frac{\partial}{\partial \theta} \right), \\ \hat{L}_3 &= -i\hbar \frac{\partial}{\partial \phi}. \end{aligned} \quad (7.31)$$

Note that the angular momentum operators care nothing for how the wavefunctions depend on the radial coordinate r . The angular momentum of a state is encoded in how the wavefunction varies in the angular directions θ and ϕ . Furthermore, \hat{L}_3 takes a particularly simple form, and that is because x_3 is singled out as a special direction in spherical polar coordinates. The total angular momentum operator \hat{L}^2 is then computed from the above expressions to be

$$\hat{L}^2 = \hat{L}_1^2 + \hat{L}_2^2 + \hat{L}_3^2 = -\frac{\hbar^2}{\sin^2 \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \sin \theta \frac{\partial}{\partial \theta} + \frac{\partial^2}{\partial \phi^2} \right). \quad (7.32)$$

One should be careful about what the above expression means: the first $\partial/\partial\theta$ is supposed to act both on the following $\sin\theta$ and on whatever function \hat{L}^2 is acting on.

We can now get down to finding the simultaneous eigenfunctions of \hat{L}_3 and \hat{L}^2 . We will drop the explicit dependence on r for the time being. Let us start with the simplest

$$\hat{L}_3 u(\theta, \phi) = m\hbar u(\theta, \phi). \quad (7.33)$$

Given the very simple form (7.31) we immediately find a solution

$$-i\hbar \frac{\partial}{\partial\phi} u(\theta, \phi) = m\hbar u(\theta, \phi) \implies u(\theta, \phi) = \Theta(\theta) e^{im\phi}, \quad (7.34)$$

for any function $\Theta(\theta)$. Notice that the wavefunction cannot take different values at the same point – in jargon, it is a *single-valued* function – so we must impose $u(\theta, \phi + 2\pi) = u(\theta, \phi)$. This forces $m \in \mathbb{Z}$. Thus we have learned that the third component of the angular momentum is quantised in units of \hbar . Of course, this would be true had we chosen any other component of \hat{L} . In fact it would be true for any direction $\vec{n} \cdot \hat{L}$ – where \vec{n} is some vector of unit norm in 3 dimensions. We simply chose a preferred direction x_3 and adapted the spherical coordinates accordingly.

Now it is the turn of \hat{L}^2

$$\hat{L}^2 u(\theta, \phi) = l(l+1)\hbar^2 u(\theta, \phi). \quad (7.35)$$

We already solved the dependence on ϕ , so we are searching for solutions $\Theta(\theta)$. Using the expression (7.32) we find

$$-\frac{\hbar^2}{\sin^2\theta} \left[\sin\theta \frac{\partial}{\partial\theta} \sin\theta \frac{\partial}{\partial\theta} + \frac{\partial^2}{\partial\phi^2} \right] \Theta(\theta) e^{im\phi} = l(l+1)\hbar^2 \Theta(\theta) e^{im\phi}. \quad (7.36)$$

Slightly working this out, we find

$$-\frac{1}{\sin^2\theta} \left[\sin\theta \frac{d}{d\theta} \sin\theta \frac{d}{d\theta} - m^2 \right] \Theta(\theta) = l(l+1)\Theta(\theta). \quad (7.37)$$

Let us perform one last change of variables, setting

$$\chi = \cos\theta, \quad (7.38)$$

so that we have

$$\sin\theta \frac{d}{d\theta} = \sin\theta \frac{d\chi}{d\theta} \frac{d}{d\chi} = -\sin^2\theta \frac{d}{d\chi} = (\cos^2\theta - 1) \frac{d}{d\chi} = (\chi^2 - 1) \frac{d}{d\chi}. \quad (7.39)$$

In terms of χ – and writing $P(\chi(\theta)) = \Theta(\theta)$ – we get the equation

$$\frac{1}{\chi^2 - 1} \left[(\chi^2 - 1) \frac{d}{d\chi} (\chi^2 - 1) \frac{d}{d\chi} - m^2 \right] P(\chi) = l(l+1)P(\chi). \quad (7.40)$$

We are almost there, promise. One last round of manipulations brings us to the so-called *associated Lagrange equation*

$$\frac{d}{d\chi} \left[(1 - \chi^2) P'(\chi) \right] + l(l+1)P(\chi) - \frac{m^2}{1 - \chi^2} P(\chi) = 0.$$

(7.41)

The name *Legendre equation* – without the connotation “associated” – is reserved for the case in which $m = 0$

$$\boxed{\frac{d}{d\chi} \left[(1 - \chi^2) P'(\chi) \right] + l(l+1) P(\chi) = 0} . \quad (7.42)$$

Both these equations are very important in mathematics, so it is worth spending some time on them. Actually, we will be looking exclusively on (7.42) due to the following, very helpful, proposition

Proposition 7. Solutions to the associated Legendre equation

Let $P_l(\chi)$ be a solution of the Legendre equation (7.42) with eigenvalue $l(l+1)$. Then, for any integer $m \in \mathbb{Z}$, the following function

$$P_l^m(\chi) \doteq (1 - \chi^2)^{|m|/2} \frac{d^{|m|}}{d\chi^{|m|}} P_l(\chi) , \quad (7.43)$$

solves the associated Legendre equation (7.41).

We will not delve in the proof of this proposition, although it is not excessively hard. For example you might proceed by induction (assuming $m > 0$), or directly differentiate the Legendre equation (7.42) $m > 0$ times and use the generalised Leibnitz rule for the m -the derivative $(f(x)g(x))^{(m)} = \sum_{k=0}^m \binom{m}{k} f^{(k)}(x)g^{(m-k)}(x)$.

So a common eigenfunction of \hat{L}^2 and \hat{L}_3 takes the form

$$u(r, \theta, \phi) = R(r) P_l^m(\cos \theta) e^{im\phi} , \quad (7.44)$$

where $P_l^m(x)$ are given by (7.43) and the overall constant R is allowed to depend on the remaining variable r , although in this subsection we omitted the explicit dependence.

7.1.3 The Legendre equation

So all we need now, is to solve (7.42). Let us rewrite it in the following form

$$P''(\chi) - \frac{2\chi}{1 - \chi^2} P'(\chi) + \frac{l(l+1)}{1 - \chi^2} P(\chi) = 0 , \quad \chi \in [-1, 1] . \quad (7.45)$$

Here, the condition on the range of χ follows from the fact that $\chi = \cos \theta$. We can appreciate immediately that $\chi = 0$ is an ordinary point, so we may expand in series without having to use the Frobenius method. Hence, we write

$$P(x) = \sum_{n=0}^{\infty} p_n \chi^n , \quad (7.46)$$

and substitute into the equation. After some purely algebraic manipulations we find

$$\sum_{n=0}^{\infty} \left[(n+2)(n+1)c_{n+2} - (n(n+1) - l(l+1))c_n \right] \chi^n = 0 , \quad (7.47)$$

which gives the recurrence relation

$$c_{n+2} = \frac{(n-l)(n+1+l)}{(n+2)(n+1)} c_n . \quad (7.48)$$

This tells us immediately that the radius of convergence is $R = 1$

$$\lim_{n \rightarrow \infty} \left| \frac{c_n}{c_{n+2}} \right| = \lim_{n \rightarrow \infty} \left| \frac{(n+2)(n+1)}{(n-l)(n+1+l)} \right| = 1 = R^2 . \quad (7.49)$$

In order for our power series solution to be finite at $\chi = \pm 1$ – which, remember, correspond to $\theta = 0, \pi$ – we must ask for the series to truncate at some integer $N \in \mathbb{N}$. In other words, we are limiting our attention to polynomial solutions to the Legendre equation (7.42). However, in order for the truncation to be possible, the parameter l must be such that $l = N$ or $l = -N - 1$. Notice that the eigenvalue $l(l+1)$ is the same for both choices, so we restrict to $l = N \in \mathbb{N}$.

The solutions $P_l(\chi)$ to the Legendre equation (7.42) with $l \in \mathbb{N}$ are known as *Legendre polynomials*. Below we list the first few (the normalisation constant is a convention)

$$\begin{aligned} P_0(\chi) &= 1 , \\ P_1(\chi) &= \chi , \\ P_2(\chi) &= -\frac{1 - 3\chi^2}{2} , \\ P_3(\chi) &= -\chi \frac{3 - 5\chi^2}{2} , \\ P_4(\chi) &= \frac{3 - 30\chi^2 + 35\chi^4}{8} . \end{aligned} \quad (7.50)$$

Notice that polynomial with even l are even and conversely for odd l . This is because, if l is an even integer, it can only be truncate the sequence of coefficients c_n with even index. Hence for the whole series to truncate to a polynomial, we must also require $c_1 = 0$, which eliminates all the odd terms in the sum.

Notice that the Legendre equation (7.42) is a Sturm-Liouville problem on the interval $\chi \in [-1, 1]$ with weight function $\rho(\chi) = 1$ and $Q(\chi) = 0$. However it is a *singular* – or non-regular – problem, since $P(\chi) = 1 - \chi^2$ vanishes at the boundaries of the interval. Nonetheless the Legendre polynomials – properly normalised – constitute a complete orthonormal system on $\chi \in [-1, 1]$ with weight $\rho(\chi) = 1$. Their orthogonality relation reads

$$\int_{-1}^1 d\chi P_l(\chi) P_n(\chi) = \frac{2}{2l-1} \delta_{ln} . \quad (7.51)$$

Our final comment here concerns the expression (7.43) for the solutions to the associated Legendre equation. If $P_l(\chi)$ is a polynomial solution to the Legendre equation, then its degree is l . Since $P_l^m(\chi)$ is obtained by deriving $P_l(\chi)$ m times, we only get non-vanishing polynomial solutions to the associated Legendre equation (7.41) if $|m| \leq l$.

7.1.4 Spherical harmonics

Let us recapitulate what we saw up to now. We started with a 3-dimensional, time-independent Schrödinger equation in a central – i.e. spherically symmetric – potential

$$\hat{H}u(\vec{x}) = -\frac{\hbar^2}{2m}\nabla^2 u(\vec{x}) + V(|\vec{x}|)u(\vec{x}) = Eu(\vec{x}) . \quad (7.52)$$

By virtue of the spherical symmetry of the potential, we saw that the total angular momentum $\hat{\vec{L}}^2$ and one of the angular momentum components – traditionally chosen to be the 3rd one – form, together with the Hamiltonian, a triple of commuting operators

$$[\hat{H}, \hat{\vec{L}}^2] = 0 , \quad [\hat{H}, \hat{L}_3] = 0 , \quad [\hat{\vec{L}}^2, \hat{L}_3] = 0 . \quad (7.53)$$

Consequently, these three operators share the same eigenfunctions. Then, using spherical coordinates (7.26), we asked $u(\vec{x})$ to be a simultaneous eigenfunction of $\hat{\vec{L}}^2$ and \hat{L}_3

$$\begin{aligned} \hat{\vec{L}}^2 u(r, \theta, \phi) &= l(l+1)\hbar^2 u(r, \theta, \phi) , \\ \hat{L}_3 u(r, \theta, \phi) &= m\hbar u(r, \theta, \phi) . \end{aligned} \quad (7.54)$$

finding that it needs to take the following factorised form

$$u(r, \theta, \phi) = R(r)Y_l^m(\theta, \phi) . \quad (7.55)$$

The angular dependence is contained in a family of functions $Y_l^m(\theta, \phi)$, parametrised by two integers: the *total angular momentum quantum number* $l \in \mathbb{N}$ and the *azimuthal angular momentum quantum number* $m \in \mathbb{Z}$, with the constraint⁴

$$|m| \leq l . \quad (7.56)$$

These functions are known by the name of *spherical harmonics*. They further factorise as

$$Y_l^m(\theta, \phi) = N_{l,m} P_l^m(\cos \theta) e^{im\phi} , \quad P_l^m(\chi) = (1 - \chi^2)^{|m|/2} \frac{d^{|m|}}{d\chi^{|m|}} P_l(\chi) , \quad (7.57)$$

with $P_l(\chi)$ being a polynomial – of order l – solution of the Legendre equation (7.42). The normalisation constant $N_{l,m}$ is chosen in such a way that the orthonormality relation holds

$$\int_0^\pi d\theta \sin \theta \int_0^{2\pi} d\phi \overline{Y_l^m}(\theta, \phi) Y_{l'}^{m'}(\theta, \phi) \equiv \langle Y_l^m, Y_{l'}^{m'} \rangle = \delta_{l,l'} \delta_{m,m'} . \quad (7.58)$$

⁴The requirement that $|m| \leq l$ is simply telling us that the angular momentum in any given direction – in this case the 3rd – can't be more than the total angular momentum. When $m = +l$ (or $m = -l$) you should think of the object as having angular momentum as maximally aligned (or anti-aligned) with the 3rd axis. When $|m| < l$, you should think of the angular momentum as aligned in some intermediate direction. Of course, all of these thoughts should be quantum: even when $m = l$, if you measure that angular momentum in a perpendicular direction like \hat{L}_1 then you're not guaranteed to get zero. Instead there will be a probability distribution of answers centred around zero.

There exists a general expression

$$N_{l,m} = \begin{cases} (-1)^m \sqrt{\frac{2l+1}{4\pi} \frac{(l-|m|)!}{(l+|m|)!}} & m > 0 \\ \sqrt{\frac{2l+1}{4\pi} \frac{(l-|m|)!}{(l+|m|)!}} & m < 0 \end{cases}. \quad (7.59)$$

Let us look at the explicit expressions of the first few spherical harmonic functions

- $l = 0$

In this case we only have a single function

$$Y_0^0(\theta, \phi) = \frac{1}{\sqrt{4\pi}}.$$

This is the case in which the particle has no angular momentum and, thus, the wavefunction exhibit no angular dependence. In atomic physics this case is referred to as *s-wave*.

- $l = 1$

Now we have three functions

$$Y_1^{\mp 1}(\theta, \phi) = \pm \sqrt{\frac{3}{8\pi}} \sin \theta e^{\mp i\phi}, \quad Y_1^0(\theta, \phi) = \sqrt{\frac{3}{4\pi}} \cos \theta.$$

These are called *p-waves*.

- $l = 2$

At this level we meet the 5 *d-waves*

$$Y_2^{\mp 2}(\theta, \phi) = \sqrt{\frac{15}{32\pi}} \sin^2 \theta e^{\mp 2i\phi}, \quad Y_2^{\mp 1}(\theta, \phi) = \pm \sqrt{\frac{15}{32\pi}} \sin(2\theta) e^{\mp i\phi},$$

$$Y_2^0(\theta, \phi) = \sqrt{\frac{5}{16\pi}} (3 \cos^2 \theta - 1).$$

In Figure 7.1 are plotted the first few spherical harmonics.

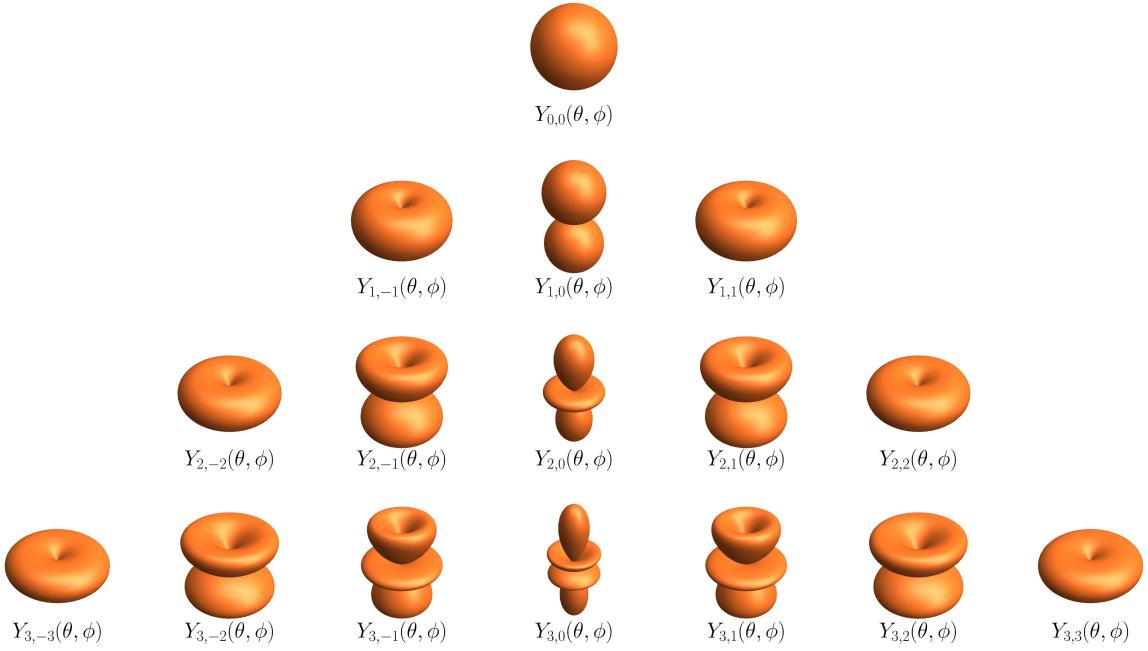


Figure 7.1: A visualisation of the first few spherical harmonics. Here the absolute value $|Y_l^m(\theta, \phi)|$ is taken to be the distance r from the origin in the (θ, ϕ) direction.

There remains one undetermined function $R(r)$ in the complete wavefunction. This will be determined by the remaining eigenvalue equation

$$\hat{H}u(r, \theta, \phi) = -\frac{\hbar^2}{2m}\nabla^2u(r, \theta, \phi) + V(r)u(r, \theta, \phi) = Eu(r, \theta, \phi) . \quad (7.60)$$

Thus the information about the specific potential $V(r)$ in which our particle sits is completely encoded in the radial part of the wavefunction and the angular part is universal, in the sense that it is always given by spherical harmonics for any centrally symmetric system. Let us see how this eigenvalue equation might look like in general. Remember the expression of the Laplacian in spherical coordinates

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r} + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta} + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2} , \quad (7.61)$$

where, as in (7.32), the derivatives act on everything that stands on their right, included the function on which ∇^2 is supposed to act. While we are looking at (7.32), we might notice that the last two terms of ∇^2 above are exactly the same as those entering $\hat{\vec{L}}^2$. In fact, we can write

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r} - \frac{\hat{\vec{L}}^2}{\hbar^2r^2} . \quad (7.62)$$

Thus, when we act on the factorised wavefunction (7.55), the angular part simply drops out, leaving an additional term proportional to the eigenvalue of $\hat{\vec{L}}^2$

$$-\frac{\hbar^2}{2mr^2}\frac{d}{dr}\left(r^2R'(r)\right) + V_{\text{eff}}(r)R(r) = ER(r) , \quad V_{\text{eff}}(r) = V(r) + \hbar^2\frac{l(l+1)}{2mr^2} . \quad (7.63)$$

Thus, the radial dependence of the wavefunction feels a leftover contribution coming from the angular dynamics in the form of a $1/r^2$ term which we absorbed inside an *effective* potential V_{eff} . This term is known as *angular momentum barrier*.

Note that the azimuthal angular momentum quantum number m does not appear in the Hamiltonian.

Sadly for us, there is an m sitting in both the kinetic term and in the effective potential, but this is the mass, not the azimuthal angular momentum quantum number! The canonical name for both is, unfortunately, m . Thankfully, this rarely leads to confusion precisely because the azimuthal quantum number does not appear in the radial equation of motion.

Since the radial equation is independent of the azimuthal quantum number, so will be the quantisation condition for the energy. This means that for each energy we may find multiple eigenfunctions with different eigenvalues of \hat{L}_3 . This situation is known with the name of *degenerate spectrum* and the number of distinct eigenfunctions having the same energy level is called *degeneracy*. Clearly, the energy levels will depend on the total angular momentum quantum number l and for a fixed value of this, we have $2l+1$ distinct possible eigenvalues of \hat{L}_3 . So we conclude that in any spherically symmetric quantum system, the energy levels E_l will exhibit a degeneracy $2l+1$.

7.2 The Hydrogen atom

Finally, we are going to look at one concrete, physical system: the hydrogen atom. This is the simplest amongst all atoms: a single electron held in orbit around a proton by the Coulomb force

$$F(r) = -\frac{\partial}{\partial r}V(r) = -\frac{e^2}{4\pi\epsilon_0}\frac{1}{r^2}. \quad (7.64)$$

Here e is the electric charge of the proton with the electron having opposite charge $-e$. The constant ϵ_0 is called *electric constant* or *vacuum permittivity*. It characterises the strength of the force. Both e and ϵ_0 are universal constants of physics, with well-determined values. We will present and use them later, when looking at the hydrogen atom's spectrum. The Coulomb potential is (minus) the primitive of the force

$$V(r) = -\frac{e^2}{4\pi\epsilon_0}\frac{1}{r}. \quad (7.65)$$

Its shape is plotted in Figure 7.2.

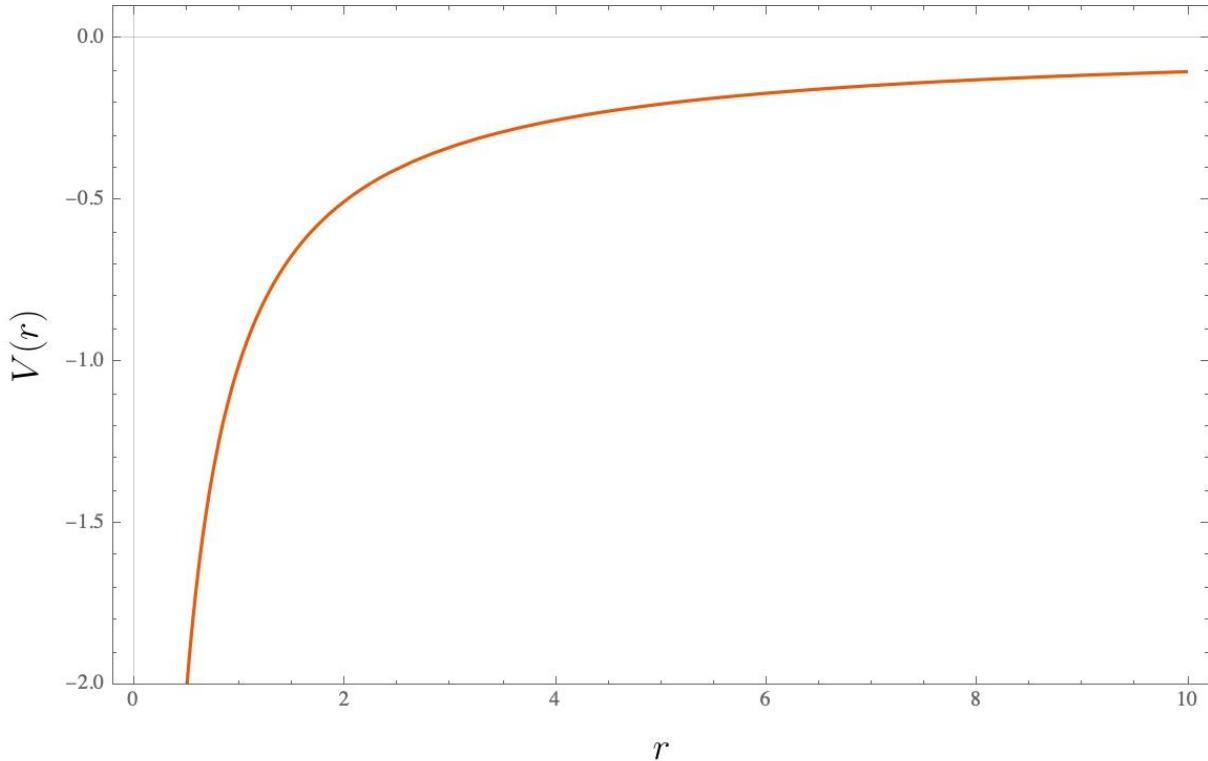


Figure 7.2: A plot of the Coulomb potential (7.65), where we fixed $e^2 = 4\pi\epsilon_0$.

Notice that this potential is negative everywhere and approaches 0 for large distances and $-\infty$ around the origin. From our experience in 1 dimension, we expect such a potential to behave somewhat like an infinitely deep well and allow a discrete, infinite set of energy levels. There are some subtleties in this argument that we are sweeping under the rug, but the bottom line is that the energy levels are indeed discrete and infinite in number.

Thus our task is to solve the Schrödinger equation

$$\hat{H}u(\vec{x}) \equiv -\frac{\hbar^2}{2m}\nabla^2u(\vec{x}) - \frac{e^2}{4\pi\epsilon_0}\frac{1}{r}u(\vec{x}) = Eu(\vec{x}) . \quad (7.66)$$

Following the discussion in the previous section, we already know the following

- The states of the hydrogen atom will be characterised by three quantum numbers
 - n , the *principal* quantum number, labelling the energy state;
 - l , the *total angular momentum* quantum number, appearing as the eigenvalue of $\hat{\vec{L}}^2$ in the form $l(l+1)\hbar^2$;
 - m , the *azimuthal momentum* quantum number, appearing as the eigenvalue of \hat{L}_3 in the form $m\hbar$.
- There will be a degeneracy of states of size $l(l+1)$ for any fixed l .
- The wavefunction factorises as

$$u(\vec{x}) = R(r)Y_l^m(\theta, \phi) ,$$

with $Y_l^m(\theta, \phi)$ being the spherical harmonic functions we studied previously.

- The radial part of the wavefunction satisfies the equation

$$-\frac{\hbar^2}{2m} \frac{1}{r^2} \frac{d}{dr} \left(r^2 R'(r) \right) + V_{\text{eff}}(r) R(r) = E R(r) , \quad (7.67)$$

where the effective potential is

$$V_{\text{eff}} = -\frac{e^2}{4\pi\epsilon_0 r} + \frac{\hbar^2 l(l+1)}{2m} \frac{1}{r^2} .$$

7.2.1 Solving the radial equation

So the task at hand is reduced to finding the radial wavefunction and the expression for the energy levels. As always, it is good to work with dimensionless quantities. We notice that⁵

$$\left[\frac{mE}{\hbar^2} \right] = \mathsf{L}^{-2} , \quad \left[\frac{me^2}{\epsilon_0 \hbar^2} \right] = \mathsf{L}^{-1} . \quad (7.68)$$

Then, let us denote

$$a^2 = -8 \frac{mE}{\hbar^2} , \quad n = \frac{me^2}{2\pi\epsilon_0 \hbar^2 a} , \quad s = ar , \quad v(s) = R(s/a) . \quad (7.69)$$

The dimensional analysis bit we just performed tells us that a is an inverse length, while n and s are pure numbers. In terms of these, the radial equation takes the following cleaner form

$$\frac{1}{s^2} \frac{d}{ds} \left(s^2 v'(s) \right) + \left(\frac{n}{s} - \frac{l(l+1)}{s^2} - \frac{1}{4} \right) v(s) = 0 . \quad (7.70)$$

We start the analysis of this equation by studying its asymptotic behaviour. For large s , we can drop all the terms that go as $1/s$ and $1/s^2$, so we are left with

$$v''(s) - \frac{v(s)}{4} \underset{s \rightarrow \infty}{\sim} 0 \implies v(s) \underset{s \rightarrow \infty}{\sim} e^{\pm s/2} . \quad (7.71)$$

Since we wish the wavefunction to be normalisable, we had better choose the negative sign in the exponent. This insight leads us to the following ansatz

$$v(s) = f(s) e^{-s/2} , \quad (7.72)$$

where $f(s)$ is such that its growth at large r should be slower than exponential. In fact we are going to see that $f(s)$ can be chosen to be a polynomial. Plugging (7.72) into the equation for $v(s)$, we find

$$\frac{1}{s^2} \frac{d}{ds} \left[s^2 e^{-s/2} (f'(s) - f(s)/2) \right] + \left(\frac{n}{s} - \frac{l(l+1)}{s^2} - \frac{1}{4} \right) f(s) e^{-s/2} = 0 . \quad (7.73)$$

After some massaging, we bring the equation to the following form

$$f''(s) + \left(\frac{2}{s} - 1 \right) f'(s) + \left(\frac{n-1}{s} - \frac{l(l+1)}{s^2} \right) f(s) = 0 . \quad (7.74)$$

⁵The trick to do this analysis even if you are unfamiliar with electrostatic quantities – such as the charge and the vacuum permittivity – is to see that the term $r^{-2} d/dr(r^2 d/dr R(r))$ has dimensions $\mathsf{L}^{-2}[R]$, where $[R]$ is whatever the dimension of $R(r)$ is (it will not matter, since the equation is linear and homogeneous). But in order to be able to sum this term with $2mV_{\text{eff}}(r)/\hbar^2 R(r)$ and $2mE/\hbar^2 R(r)$, they all need to have the same dimension $\mathsf{L}^{-2}[R]$. With some simple algebra, you arrive at (7.68).

We immediately notice that $s = 0$ is a regular singular point, so we can use the Frobenius method we learned in §6.2. We start from the power series expansion

$$f(s) = \sum_{k=0}^{\infty} c_k s^{k+\sigma}, \quad (7.75)$$

and insert it into the equation (we will have a slightly easier time if we multiply everything by s^2). We need the following terms

$$\begin{aligned} s^2 f''(s) &= \sum_{k=0}^{\infty} (k+\sigma)(k+\sigma-1) c_k s^{k+\sigma}, \\ (2s - s^2) f'(s) &= \sum_{k=0}^{\infty} 2(k+\sigma) c_k s^{k+\sigma} - \sum_{k=0}^{\infty} (k+\sigma) c_k s^{k+1+\sigma} = \\ &= \sum_{k=0}^{\infty} 2(k+\sigma) c_k s^{k+\sigma} - \sum_{k=1}^{\infty} (k-1+\sigma) c_{k-1} s^{k+\sigma}, \\ [(n-1)s - l(l+1)] f(s) &= \sum_{k=0}^{\infty} (n-1) c_k s^{k+1+\sigma} - \sum_{k=0}^{\infty} l(l+1) c_k s^{k+\sigma} = \\ &= \sum_{k=1}^{\infty} (n-1) c_{k-1} s^{k+\sigma} - \sum_{k=0}^{\infty} l(l+1) c_k s^{k+\sigma}. \end{aligned} \quad (7.76)$$

Putting them together and separating the $k = 0$ terms from the other ones we obtain

$$[\sigma(\sigma+1) - l(l+1)] c_0 s^\sigma + \sum_{k=1}^{\infty} \left[\left((k+\sigma)(k+\sigma+1) - l(l+1) \right) c_k - (k+\sigma-n) c_{k-1} \right] s^{k+\sigma} = 0.$$

This equation is valid for any s , meaning the coefficients of each power of s must vanish independently

$$\sigma(\sigma+1) = l(l+1), \quad (7.77a)$$

$$c_k = \frac{k+\sigma-n}{k(k+2\sigma+1)} c_{k-1}, \quad (7.77b)$$

where we already used the first equation to eliminate $l(l+1)$ from the second. The recurrence relation immediately tells us that the radius of convergence is infinite

$$\lim_{k \rightarrow \infty} \left| \frac{c_{k-1}}{c_k} \right| = \left| \frac{k(k+2\sigma+1)}{k+\sigma-n} \right| \rightarrow \infty. \quad (7.78)$$

7.2.2 The wavefunction is finite at the origin

Now, the indicial equation (7.77a) is quadratic and thus possesses two solutions $\sigma = l \geq 0$ and $\sigma = -l-1 < 0$. Clearly if we want the wavefunction to be normalisable, we should choose the positive solution. However, the attentive reader might counter that this argument does not hold for $l = 0$. In fact the total probability associated with the wavefunction $\psi(\vec{x}, t) = u(\vec{x})e^{-iEt} = R(r)Y_l^m(\theta, \phi)e^{-iEt}$ is

$$\int_{\mathbb{R}^3} dV |\psi(\vec{x}, t)|^2 = \int_0^{\infty} dr r^2 |R(r)|^2 \int_0^{\pi} d\theta \sin \theta \int_0^{2\pi} d\phi |Y_l^m(\theta, \phi)|^2, \quad (7.79)$$

where the r^2 and $\sin \theta$ come from the volume element dV once expressed in spherical coordinates. So, in order for the wavefunction to be normalisable, the expression that must be finite is

$$\int_0^\infty dr r^2 |R(r)|^2 = \frac{1}{a} \int_0^\infty ds s^2 |f(s)|^2 e^{-s} , \quad (7.80)$$

and we can allow the behaviour $f(s) \sim s^{-1}$ near the origin. Thus, at least when $l = 0$, it seems that $\sigma = -1 - l = -1$ is admissible. The true, deep reason why we have to discard $\sigma = -1 - l$, also in the case $l = 0$, has to do with the fact that we want the operator \hat{H} for this system – and in general, for any quantum system – to be Hermitian. Stated differently we want our Schrödinger equations, together with boundary conditions, to be Sturm-Liouville problems. In the present case, the equation (7.67) has the form (5.4) with $\rho(r) = r^2$, $P(r) = r^2$, $Q(r) = r^2 V_{\text{eff}}(r)$, $\lambda = E$, $a = 0$ and $b = \infty$. We must then ask

$$\left[s^2 \frac{d\bar{g}(s)}{ds} f(s) - s^2 \frac{df(s)}{ds} \bar{g}(s) \right]_{s=0}^{s=\infty} = 0 , \quad (7.81)$$

for all the functions $g(s)$ and $f(s)$ that satisfy the boundary conditions. The $s \rightarrow \infty$ boundary is trivial since all functions behave as $e^{-s/2}$. Let us look at the $s = 0$ boundary. The request is that

$$\lim_{s \rightarrow 0} s^2 \left[\frac{d\bar{g}(s)}{ds} f(s) - \frac{df(s)}{ds} \bar{g}(s) \right] = 0 . \quad (7.82)$$

But this cannot happen if, say, $g(0)$ is finite and $f(s) \sim s^{-1}$! Consequently we have to discard all solutions that behave near the origin as s^σ with $\sigma < 0$.

7.2.3 The energy spectrum

Hence we choose $\sigma = l \geq 0$ as the solution to the indicial equation. This leaves us with the recurrence relation

$$c_k = \frac{k + l - n}{k(k + 2l + 1)} c_{k-1} . \quad (7.83)$$

Just as was the case for the Legendre equation, we want the series to truncate. That is, we want the function $f(s)$ to be a polynomial in s of some degree S . But this is only possible if $S + 1 + l - n = 0$, for $S \in \mathbb{N} \cup \{0\}$ and $l \in \mathbb{N} \cup \{0\}$, meaning n must be a natural number⁶

$$n \in \mathbb{N} . \quad (7.84)$$

This explains why we called this parameter in this way. Now we return to our physical parameters via (7.69) and obtain the expression for the energy levels E_n

$$-8 \frac{mE}{\hbar^2} = a^2 = \frac{m^2 e^4}{4\pi^2 \epsilon_0^2 \hbar^4} \frac{1}{n^2} \implies \boxed{E_n = -\frac{me^4}{32\pi^2 \epsilon_0^2 \hbar^2} \frac{1}{n^2}} . \quad (7.85)$$

Note that, rather surprisingly, the energy does not depend on the total angular momentum number⁷ l ! There is an indirect dependence, in the sense that, in order for the series

⁶It cannot be 0 since its minimal value $n = 1$ is reached for $S = 0$ and $l = 0$.

⁷There are actually corrections due to special relativity. These corrections split the n^2 degeneracy (see slightly below) so that the states with different l , but the same n , have different energies. These splitting are of order α^4 (again, see below), much smaller than the original energies that are of order α^2 , and are referred to as the *fine structure* of the hydrogen atom.

truncation to take place, one must have $l < n$. So, for each energy E_n , only a finite number of total angular momenta are allowed. Note that this is not a common phenomenon. In fact the hydrogen atom and the 3-dimensional harmonic oscillator are the only systems in which the energy spectrum is independent of the angular momentum. The total degeneracy for a state with energy E_n – that is the number of different states with the same energy – is easily counted. For fixed n we have n choices of $l \in [0, n - 1]$. And for each l we have $2l + 1$ choices of $m \in [-l, l]$. Hence

$$\text{degeneracy}(n) = \sum_{l=0}^{n-1} (2l + 1) = n^2 . \quad (7.86)$$

The ground state $n = 1$ is unique: it sits in a $l = 0$ s-wave. There are four $n = 2$ excited states corresponding to the s-wave $l = 0$ and three $l = 1$ states of the p-wave. There are nine $n = 3$ excited states, corresponding to the s-wave $l = 0$, the three $l = 1$ p-wave states and five $l = 2$ d-wave states. And so on.

The messy cocktail of constants in the expression of E_n can be simplified by introducing the *fine structure constant*

$$\alpha = \frac{e^2}{4\pi\epsilon_0\hbar c} , \quad (7.87)$$

where $c \approx 2.99792458 \times 10^8$ m/s is the speed of light constant. One might question why introducing α is a good idea, since it brings the speed of light in the game out of nowhere. Well, one good reason is that α is a dimensionless constant. The second is that it is a number which is extremely easy to remember approximately

$$\alpha \approx \frac{1}{137} \approx 0.0073 . \quad (7.88)$$

Using the fine structure constant, the energy levels read

$$E_n = -\frac{\alpha^2}{2n^2}m_e c^2 .$$

(7.89)

Here we wrote m_e in place of m to stress the fact that this is the mass of the electron $m_e \approx 9.10938 \times 10^{-31}$ kg. The constant $m_e c^2 \approx 8.1711 \times 10^{-14}$ J is called *rest energy* of the electron and you will learn further in your studies that it is the intrinsic energy associated to an object of mass m_e . It is a concept that arises from special relativity. The rest energy of the electron is what sets the energy scale for the hydrogen atom. The associated unit of measure is called *Rydberg* and denoted by Ry

$$1\text{Ry} = \frac{\alpha}{2}m_e c^2 \approx 2.98799 \times 10^{-16}\text{J} . \quad (7.90)$$

Another important quantity is the scale $1/a$ we used to make the variable r dimensionless

$$\frac{1}{a} = \frac{n}{2}r_B , \quad r_B = \frac{4\pi^2\epsilon_0\hbar^2}{me^2} = \frac{\hbar}{\alpha m_e c} \approx 5.291772 \times 10^{-11}\text{m} , \quad (7.91)$$

where r_B is the *Bohr radius* and represents the characteristic size of the hydrogen atom.

Correct prediction for the spectrum of the hydrogen atom was one of the most spectacular successes of quantum mechanics. Figure 7.3 shows the emission and absorption lines

corresponding to the *Balmer series*: a series of spectral emission lines of the hydrogen atom that result from electron transitions from higher levels $n > 2$ down to the energy level with principal quantum number $n = 2$. The accumulation of lines as the wavelength grows follows exactly the $1/n^2$ behaviour dictated by our computations.

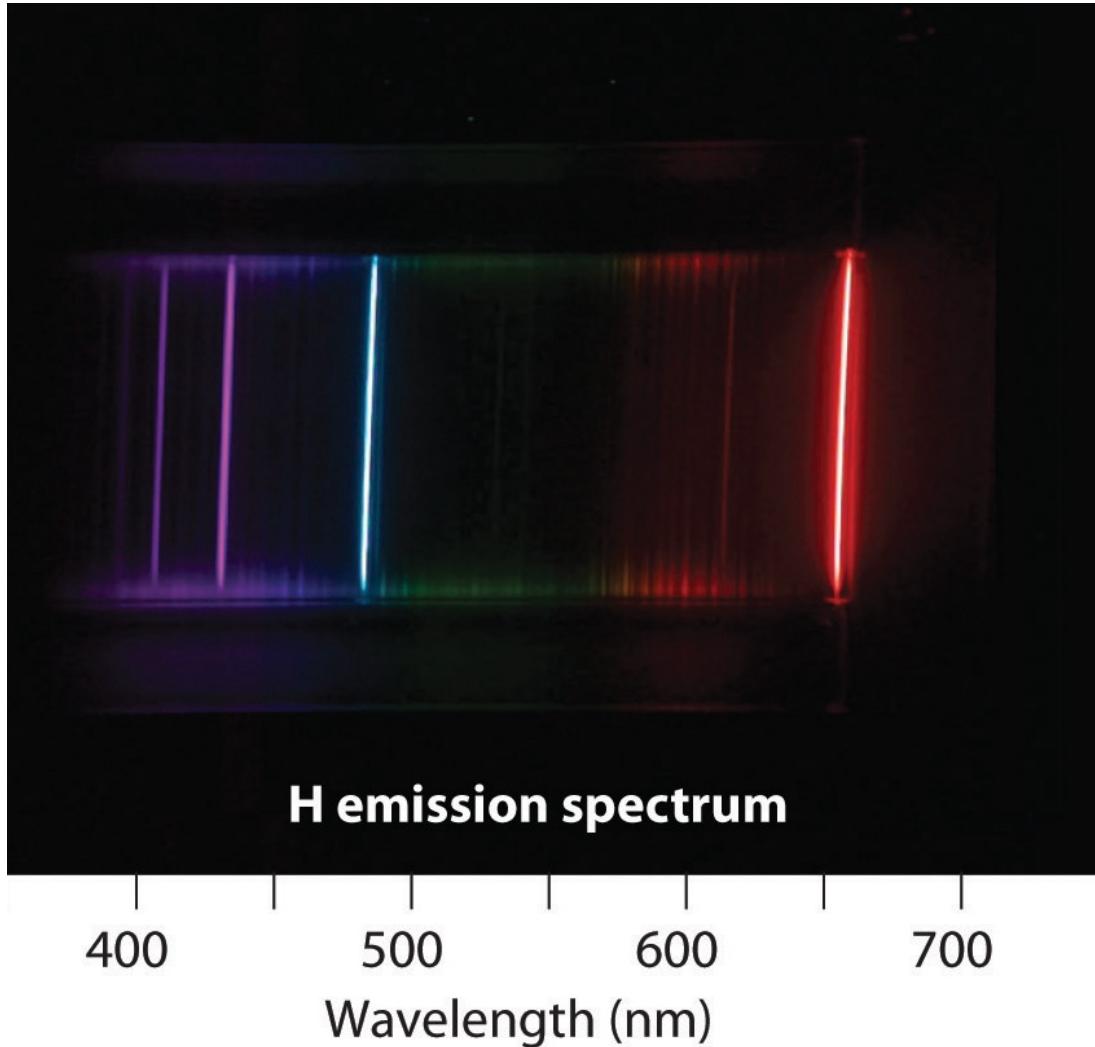


Figure 7.3: Emission lines of the hydrogen atom in the visible spectrum.

The lines of the Balmer series are not the only ones, they simply are the only ones whose light is in the visible spectrum. The hydrogen atom has emission spectra corresponding to transitions from any level $n > 1$ to any other level $n \geq 1$. Figure 7.4 shows some

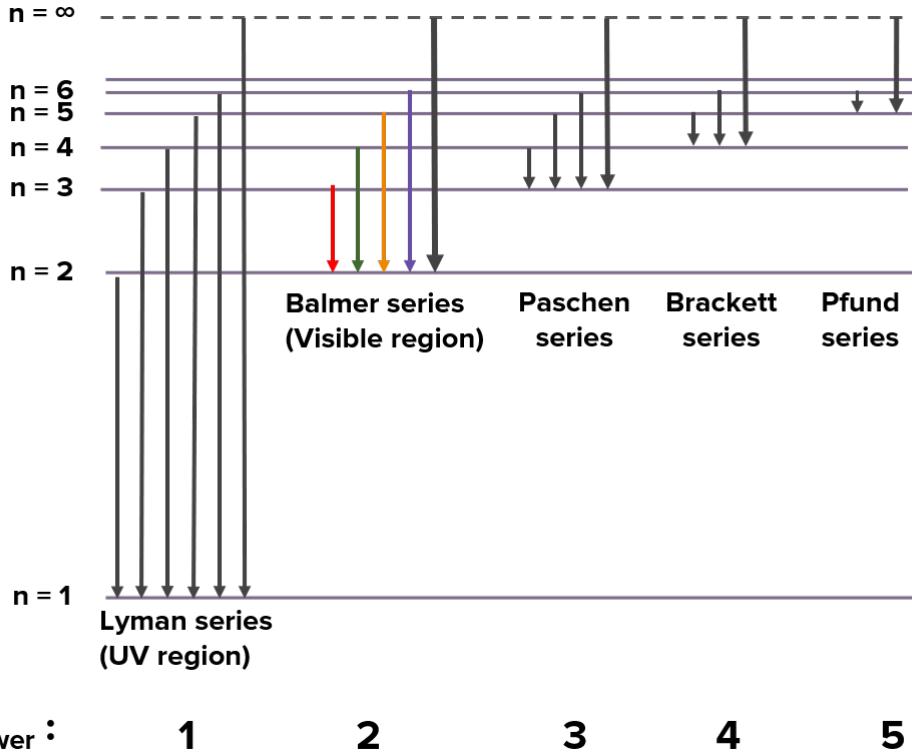


Figure 7.4: Emission lines of the hydrogen atom subdivided in various series indexed by the principal quantum number of the landing state n_{lower} .

7.2.4 The wavefunction

Now let us collect all the pieces of our derivation and write down the wavefunctions. A state of the hydrogen atom with quantum numbers n , l and m has the following wavefunction

$$u(r, \theta, \phi) = M_{n,l} r^l L_{n-l-1}^{2l+1}(2r/(nr_B)) e^{-r/(nr_B)} Y_l^n(\theta, \phi), \quad (7.92)$$

where $M_{n,l}$ is the normalisation constant

$$M_{n,l} = \sqrt{\left(\frac{2m\alpha\hbar}{n}\right)^3 \frac{(n-l-1)!}{2n[(n+l)!]^3}}, \quad (7.93)$$

and $L_{n-l-1}^{2l+1}(x)$ are known as *associated Laguerre polynomials*

$$L_{n-l-1}^{2l+1}(x) = \sum_{k=0}^{n-l-1} c_k x^k, \quad c_k = \frac{k+l-n}{k(k+2l+1)} c_{k-1}. \quad (7.94)$$

Outro: takeaway points

In this chapter, we considered quantum-mechanical particles in 3 dimensions. We focused exclusively on *central* potentials, i.e. potentials that only depend on the radial coordinate. In particular

- We introduced the *angular momentum operator*

$$\hat{\vec{L}} \doteq \hat{\vec{x}} \times \hat{\vec{p}},$$

whose action on wavefunctions $u(\vec{x})$ has the following form (in components)

$$\begin{aligned}\hat{L}_1 u(\vec{x}) &= -i\hbar \left(x_2 \frac{\partial}{\partial x_3} - x_3 \frac{\partial}{\partial x_2} \right) u(\vec{x}), \\ \hat{L}_2 u(\vec{x}) &= -i\hbar \left(x_3 \frac{\partial}{\partial x_1} - x_1 \frac{\partial}{\partial x_3} \right) u(\vec{x}), \\ \hat{L}_3 u(\vec{x}) &= -i\hbar \left(x_1 \frac{\partial}{\partial x_2} - x_2 \frac{\partial}{\partial x_1} \right) u(\vec{x}).\end{aligned}$$

This is written more compactly thanks to the *epsilon symbol* ϵ_{ijk}

$$\epsilon_{123} = 1, \quad \epsilon_{ijk} = -\epsilon_{jik} = -\epsilon_{ikj} = -\epsilon_{kji},$$

with i, j, k taking values in the set $\{1, 2, 3\}$. We then write

$$\hat{L}_i u(\vec{x}) = -i\hbar \epsilon_{ijk} x_j \frac{\partial}{\partial x_k} u(\vec{x}),$$

where we use the *Einstein summation convention* which tells us that anytime we see a repeated index we need to sum over all the values of that integer (in this case 1,2 and 3).

- We saw that the components of the energy momentum operator do not commute amongst themselves

$$[\hat{L}_i, \hat{L}_j] \doteq \hat{L}_i \hat{L}_j - \hat{L}_j \hat{L}_i = i\hbar \epsilon_{ijk} \hat{L}_k.$$

Here $[\hat{A}, \hat{B}]$ is the *commutator* of the operators \hat{A} and \hat{B} . It is generally non-zero for quantum observables. E.g. $[\hat{x}_i, \hat{p}_j] = i\hbar \delta_{ij}$.

The lack of commutation between two observable means that it is impossible for any state to have a well defined expectation value for both: there is an uncertainty relation relating to these observables.

However, the *total angular momentum tensor operator* $\hat{\vec{L}}^2$ commutes with all the components of $\hat{\vec{L}}$

$$[\hat{\vec{L}}^2, \hat{L}_i] = 0, \quad \forall i = 1, 2, 3.$$

This means that we can know both expectation values of $\hat{\vec{L}}^2$ and, say, \hat{L}_3 with arbitrary precision.

Additionally, both $\hat{\vec{L}}^2$ and \hat{L}_i commute with the Hamiltonian operator

$$\hat{H} = \frac{\hat{\vec{p}}^2}{2m} + V(r),$$

where, critically, the potential V only depends on the radial coordinate.

Thus we got ourselves a triple $\{\hat{H}, \hat{\vec{L}}^2, \hat{L}_3\}$ (where we chose by convention the last one amongst the three components) of commuting operators

$$[\hat{H}, \hat{\vec{L}}^2] = 0, \quad [\hat{H}, \hat{L}_3] = 0, \quad [\hat{\vec{L}}^2, \hat{L}_3] = 0.$$

This is important because *two commuting operators share the same eigenfunctions*. As a consequence, the solutions to the stationary Schrödinger equation

$$\hat{H}u(\vec{x}) = Eu(\vec{x}) ,$$

can be taken to be *simultaneous eigenfunctions* of the operators \hat{H} , $\hat{\vec{L}}^2$ and \hat{L}_3

$$\begin{aligned}\hat{H}u_{m,l,E}(\vec{x}) &= Eu_{m,l,E}(\vec{x}) , \\ \hat{\vec{L}}^2u_{m,l,E}(\vec{x}) &= l(l+1)\hbar^2u_{m,l,E}(\vec{x}) , \\ \hat{L}_3u_{m,l,E}(\vec{x}) &= m\hbar u_{m,l,E}(\vec{x}) .\end{aligned}$$

This is the quantum version of the separation of variables. The parameters m and l are known as *azimuthal and total angular momentum quantum numbers*, respectively

- We introduced *spherical coordinates*

$$\begin{cases} x_1 = r \sin \theta \cos \phi \\ x_2 = r \sin \theta \sin \phi \\ x_3 = r \cos \theta \end{cases} , \quad \begin{cases} \theta \in [0, \pi] \\ \phi \in [0, 2\pi] \end{cases} ,$$

and wrote the angular momentum operator components in their terms

$$\begin{aligned}\hat{L}_1 &= i\hbar \left(\cot \theta \cos \phi \frac{\partial}{\partial \phi} + \sin \phi \frac{\partial}{\partial \theta} \right) , \\ \hat{L}_2 &= i\hbar \left(\cot \theta \sin \phi \frac{\partial}{\partial \phi} - \cos \phi \frac{\partial}{\partial \theta} \right) , \\ \hat{L}_3 &= -i\hbar \frac{\partial}{\partial \phi} .\end{aligned}$$

The total angular momentum operator reads

$$\hat{\vec{L}}^2 = -\frac{\hbar^2}{\sin^2 \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \sin \theta \frac{\partial}{\partial \theta} + \frac{\partial^2}{\partial \phi^2} \right) .$$

With these expressions, the eigenfunctions of \hat{L}_3 are easily found to be

$$u(r, \theta, \phi) = R(r)\Theta(\theta)e^{im\phi} .$$

The reason why the wavefunction decomposes in a product of single-variable functions is because the eigenvalue equations for \hat{L}_3 only involves ϕ and the one for $\hat{\vec{L}}^2$ only involves θ and ϕ .

The eigenvalue equation for $\hat{\vec{L}}^2$ takes the form

$$-\frac{1}{\sin^2 \theta} \left[\sin \theta \frac{d}{d\theta} \sin \theta \frac{d}{d\theta} - m^2 \right] \Theta(\theta) = l(l+1)\Theta(\theta) .$$

its normalisable solutions are polynomials of $\cos \theta$ known as *associated Legendre polynomials*

$$\Theta(\theta) = P_l^m(\cos \theta) , \quad P_l^m(\chi) \doteq (1 - \chi^2)^{|m|/2} \frac{d^{|m|}}{d\chi^{|m|}} P_l(\chi) ,$$

with $P_l(\chi)$ being a polynomial solution to the Legendre equation

$$\frac{d}{d\chi} \left[(1 - \chi^2) P_l'(\chi) \right] + l(l+1)P_l(\chi) = 0 .$$

- We studied the Legendre equation and saw that it admits polynomial solutions whenever $l = N \in \mathbb{N}$, with N being the degree of the polynomial. The first few are

$$P_0(\chi) = 1, \quad P_1(\chi) = \chi, \quad P_2(\chi) = -\frac{1 - 3\chi^2}{2}.$$

These polynomials form a complete orthogonal basis for functions on the interval $[-1, 1]$ and

$$\int_{-1}^1 d\chi P_l(\chi) P_n(\chi) = \frac{2}{2l - 1} \delta_{ln}.$$

Since $P_l(\chi)$ is a polynomial of order l and $P_l^m(\chi)$ is obtained by taking $|m|$ derivatives of $P_l(\chi)$, we conclude that

$$|m| \leq l.$$

- The angular dependence of the wavefunction is then encoded into the *spherical harmonics* $Y_l^m(\theta, \phi)$, which are simultaneous eigenfunctions of $\hat{\vec{L}}^2$ and \hat{L}_3

$$\hat{\vec{L}}^2 Y_l^m(\theta, \phi) = l(l+1)\hbar^2 Y_l^m(\theta, \phi), \quad \hat{L}_3 Y_l^m(\theta, \phi) = m\hbar Y_l^m(\theta, \phi),$$

defined as

$$Y_l^m(\theta, \phi) = N_{l,m} P_l^m(\cos \theta) e^{im\phi},$$

where the normalisation is

$$N_{l,m} = \begin{cases} (-1)^m \sqrt{\frac{2l+1}{4\pi} \frac{(l-|m|)!}{(l+|m|)!}}, & m > 0 \\ \sqrt{\frac{2l+1}{4\pi} \frac{(l-|m|)!}{(l+|m|)!}}, & m < 0 \end{cases}.$$

This ensures that the spherical harmonics are a complete orthonormal basis for functions on the sphere

$$\langle Y_l^m, Y_{l'}^{m'} \rangle \equiv \int_0^\pi d\theta \sin \theta \int_0^{2\pi} d\phi \bar{Y}_l^m(\theta, \phi) Y_{l'}^{m'}(\theta, \phi) = \delta_{l,l'} \delta_{m,m'}.$$

- What is left is the radial part of the wavefunction which is constrained by the hamiltonian eigenvalue equation

$$\hat{H}u(r, \theta, \phi) = -\frac{\hbar^2}{2m} \nabla^2 u(r, \theta, \phi) + V(r)u(r, \theta, \phi) = Eu(r, \theta, \phi).$$

In spherical coordinates, we saw that

$$\nabla^2 = \frac{1}{r^2} \frac{\partial}{\partial r} r^2 \frac{\partial}{\partial r} - \frac{\hat{\vec{L}}^2}{\hbar^2 r^2}.$$

Hence, using the expression

$$u(r, \theta, \phi) = R(r)Y_l^m(\theta, \phi),$$

we arrive at the equation

$$-\frac{\hbar^2}{2mr^2} \frac{d}{dr} \left(r^2 R'(r) \right) + V_{\text{eff}}(r) R(r) = ER(r) , \quad V_{\text{eff}}(r) = V(r) + \hbar^2 \frac{l(l+1)}{2mr^2} .$$

Note: here m is the mass, not the azimuthal quantum number.

- We used all of the above to study the hydrogen atom: an electron orbiting a proton, held in place by the Coulomb potential

$$V(r) = -\frac{e^2}{4\pi\epsilon_0 r} .$$

Here e is the electric charge of both the proton and the electron (with a $-$ sign for the latter) while ϵ_0 is a universal constant called the *vacuum permittivity*. Introducing the new parametrisations

$$a^2 = -8 \frac{mE}{\hbar^2} , \quad n = \frac{me^2}{2\pi\epsilon_0\hbar^2} \frac{1}{a} , \quad s = ar , \quad v(s) = R(s/a) ,$$

we rewrote the eigenvalue equation for $R(r)$ as

$$\frac{1}{s^2} \frac{d}{ds} \left(s^2 v'(s) \right) + \left(\frac{n}{s} - \frac{l(l+1)}{s^2} - \frac{1}{4} \right) v(s) = 0 .$$

As in the case of the harmonic oscillator, we chose a normalisable solution in the form

$$v(s) = f(s)e^{-s/2} ,$$

where $f(s)$ is a polynomial, solution to the equation

$$f''(s) + \left(\frac{2}{s} - 1 \right) f'(s) + \left(\frac{n-1}{s} - \frac{l(l+1)}{s^2} \right) f(s) = 0 .$$

- This equation has a regular singular point at the origin, so we studied it with the Frobenius method and found that the series solution truncates only if $n \in \mathbb{N}$ and $l < n$, which implies the quantisation condition on the energy

$$E_n = -\frac{me^4}{32\pi^2\epsilon_0^2\hbar^2} \frac{1}{n^2} , \quad n \in \mathbb{N} .$$

So the energy does not depend on l nor m , meaning that given n we have multiple eigenfunctions with the same energy, one for each allowed pair (l, m) (remember, $0 \leq l < n$ and $|m| \leq l$). This number of states with the same energy is called *degeneracy*

$$\text{degeneracy}(n) = \sum_{l=0}^{n-1} (2l+1) = n^2 .$$

The quantisation of the energy becomes simpler-looking by introducing the *fine-structure constant*, a pure number

$$\alpha \doteq \frac{e^2}{4\pi\epsilon_0\hbar c} \approx \frac{1}{137} ,$$

where $c \approx 2.99792458 \times 10^8$ m/s is the speed of light. With this we see that the energy levels

$$E_n = -\frac{\alpha^2}{2n^2}m_e c^2 ,$$

are multiples of the quantity $m_e c^2$ which is the *rest energy* of the electron (m_e is the mass of the electron).

- The full wavefunction for the hydrogen atom takes the form

$$u(r, \theta, \phi) = M_{n,l} r^l L_{n-l-1}^{2l+1}(2r/(nr_B)) e^{-r/(nr_B)} Y_l^n(\theta, \phi) ,$$

where

$$M_{n,l} = \sqrt{\left(\frac{2m\alpha\hbar}{n}\right)^3 \frac{(n-l-1)!}{2n[(n+l)!]^3}} ,$$

$L_{n-l-1}^{2l+1}(x)$ are the *associated Laguerre polynomials*

$$L_{n-l-1}^{2l+1}(x) = \sum_{k=0}^{n-l-1} c_k x^k , \quad c_k = \frac{k+l-n}{k(k+2l+1)} c_{k-1} ,$$

and r_B is the *Bohr radius* which constitute the characteristic size of the hydrogen atom

$$r_B \doteq \frac{\hbar}{\alpha m_e c} \approx 5.291772 \times 10^{-11} \text{ m} .$$