# Physical Mathematics 2011

Dr Peter Boyle
School of Physics and Astronomy (JCMB 4407)
paboyle@ph.ed.ac.uk

This version: September 23, 2011

**Abstract**

These are the lecture notes to accompany the Physical Mathematics lecture course.

# Contents

# Chapter 1

# Introduction

## 1.1 Organisation

Online notes & tutorial sheets

`www.ph.ed.ac.uk/~paboyle/`

### 1.1.1 Books

There are many books covering the special functions material in this course. Good ones include:

- "Mathematical Methods for Physics and Engineering", K.F. Riley, M.P. Hobson and S.J. Bence (Cambridge University Press)
- "Mathematical Methods in the Physical Sciences", M.L. Boas (Wiley)
- "Mathematical Methods for Physicists", G. Arfken (Academic Press)

These, and plenty more besides, are available **for free** in the JCMB library.

### 1.1.2 On the web

There are some useful websites for quick reference, including:

- `http://mathworld.wolfram.com`,
- `http://en.wikipedia.org`,
- `http://planetmath.org`.
- Numerical Recipes: `http://apps.nrbook.com/c/index.html`.

### 1.1.3 Workshops

Workshops run from week 2 through week 11.
There are two sessions:

- Tuesday 11:10-12:00 (room 1206C JCMB)
- Tuesday 14:00-15:50 (room 3217 JCMB)

### 1.1.4 Feedback

In week 8, I will hand out a 60 minute mock exam, and example solutions.

Anyone who wishes to have their script marked for feedback can hand this in. **The mark will *not* contribute to your course mark, but serves as useful practice and diagnostic.**

### 1.1.5 Structure

This brief introduction is Chapter 1. The rest of the course is composed of two parts.

Chapter 2 covers techniques for the solution of the partial differential equations (PDE's) of physics.

Chapter 3 covers probability, statistics and the fitting of data.

The structure of the course is different compared to previous years, due to the reorganisation of MFP in the second year.

- We retain the material on special functions and PDEs in curvilinear coordinate systems.
- We add material on probability and statistics.

You will note however from past papers that previous years contained substantial emphasis on Fourier series and Fourier transforms, topics which it is now expected that you already know and are skilled in using.

# Chapter 2

# Generalised Fourier series & special functions

## 2.1 PDE's and physics

Physics involves the description of behaviour of the universe with partial differential equations. The main PDEs in physics are:

| | |
|---|---|
| Poisson equation (electrostatics) | Laplace equation |
| Wave equation | Schrödinger equation |
| Navier-Stokes equations | Maxwell's equations |

Some common ones are summarised in the following table.

| Name | Equation | Physical context |
|---|---|---|
| Poisson | $\nabla^2\phi(\boldsymbol{r}) = -\frac{\rho(\boldsymbol{r})}{\epsilon_0}$ | Electrostatics:<br>$\phi(\boldsymbol{r})$ = potential;<br>$\rho(\boldsymbol{r})$ = charge density. |
| Wave | $\nabla^2 u(\boldsymbol{r},t) = \frac{1}{v^2}\frac{\partial^2}{\partial t^2}u(\boldsymbol{r},t)$ | All areas:<br>$v$ = wave speed;<br>$u(\boldsymbol{r},t)$ = 'displacement' from equilibrium. |
| Laplace | $\nabla^2\phi(\boldsymbol{r}) = 0$ | Special cases of above. |
| Schrödinger | $\left(-\frac{\hbar^2}{2m}\nabla^2 + U(\boldsymbol{r})\right)\psi(\boldsymbol{r}) = E\psi(\boldsymbol{r})$ | Quantum mechanics:<br>$\psi(\boldsymbol{r})$ = wave function. |

Solving these equations for two and three dimensional problems can be challenging, and the first half of this course addresses techniques for the solution of PDEs in common situations.

This general approach involves several concepts which we will cover in more detail later.

1. **Differential operators**
   Gradient, Divergence and Curl are the building blocks for three dimensional equations.

2. **Separation of variables**
   the problem can be simplified to independent one dimensional ODE's by seeking solutions of a particular form.

3. **Solution of the separated ODEs**
   Recognising solution (e.g. wave equation)
   Substitute a power series (Method of Froebenius)

4. **Reconstruct general solution**
   The *orthogonality* and *completeness* of the solutions of an ODE allow us to write any solution as a linear combination of the normal mode solutions.
   It is this property that allows us to represent general functions by Fourier series.

We shall see that the normal strategy for dealing with several dimensions in analytical calculation is to duck the issue, and reduce it to one dimensional problems. This works for symmetrical problems if we choose coordinate systems that possess the same symmetries.

For non-symmetrical situations, one often resorts to using numerical approaches. For example, computational fluid dynamics is used to optimise the shapes of complex objects like aeroplanes and cars.

## 2.2 The wave equation

We can describe the transverse displacement of a stretched string using a function $u(x,t)$ which tells us how far the infinitesimal element of string at (longitudinal) position $x$ has been (transversely) displaced at time $t$. The function $u(x,t)$ satisfies a partial differential equation (PDE) known as the wave equation:

$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{c^2}\,\frac{\partial^2 u}{\partial t^2} \tag{2.1}$$

where $c$ is a constant,and has units of length over time (i.e. of velocity) and is, in fact, the speed of propagation of travelling waves on the string.

In the absence of boundaries, the general solution can be seen by noting:

$$\begin{aligned}\frac{\partial^2 u}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} &= \left(\frac{\partial}{\partial x} - \frac{1}{c}\frac{\partial}{\partial t}\right)\left(\frac{\partial}{\partial x} + \frac{1}{c}\frac{\partial}{\partial t}\right) u \\ &= \left(\frac{\partial}{\partial x} + \frac{1}{c}\frac{\partial}{\partial t}\right)\left(\frac{\partial}{\partial x} - \frac{1}{c}\frac{\partial}{\partial t}\right) u\end{aligned}$$

This is solved by

$$u(x,t) = f(x-ct) + g(x+ct)$$

where $f$ and $g$ are arbitrary functions of a single variable. This represents the superposition of arbitrary left and right propagating waves.

### 2.2.1 Separation of variables

Our equation of motion in Eqn. (2.1) is perhaps the simplest second order partial differential equation (PDE) imaginable – it doesn't contain any mixed derivatives (e.g. $\frac{\partial^2 u}{\partial x \partial t}$). We call such a differential equation a *separable* one, or say that it is of *separable form.*

We can seek particular solutions in which variations with space and time are *independent.* Such *standing waves* are of the *seperable* form:

$$u(x,t) = X(x)\,T(t)\ .$$

This really is a *restriction* of the class of possible solutions and there are certainly solutions to the wave equation that are *not* of separated form (e.g. travelling waves as above).

However, we shall see *all* solutions of the wave equation (separated form or not) can be written as a linear combination of solutions of separated form, so this restriction is not a problem.

Differentiating, we get

$$\begin{aligned}\frac{\partial u}{\partial x} &= \frac{dX}{dx}\,T \equiv X'T \qquad &\Rightarrow \qquad \frac{\partial^2 u}{\partial x^2} &= X''T \\ \frac{\partial u}{\partial t} &= X\,\frac{dT}{dt} \equiv X\dot{T} \qquad &\Rightarrow \qquad \frac{\partial^2 u}{\partial t^2} &= X\ddot{T}\end{aligned}$$

Substituting this into the PDE:

$$X''(x)T(t) = \frac{1}{c^2}X(x)\ddot{T}(t) \ ,$$

Thus,

$$\frac{X(x)''}{X(x)} = \frac{1}{c^2}\,\frac{\ddot{T}(t)}{T(t)}$$

Now

$$\frac{\partial}{\partial t}LHS = \frac{\partial}{\partial x}RHS = 0$$

Hence both LHS and RHS must be equal to the same constant and we may write

$$\frac{X''}{X} = \frac{1}{c^2}\,\frac{\ddot{T}}{T} = -k^2 \qquad \text{(say)},$$

where $-k^2$ is called the **separation constant**.

Now we have *separated* our PDE in two variables into two simple second order ordinary differential equations (ODEs) in one variable each:

$$\begin{aligned}\frac{d^2X}{dx^2} &= -\,k^2X(x)\\ \frac{d^2T}{dt^2} &= -\,\omega_k^2T(t)\end{aligned}$$

where the angular frequency $\omega_k = ck$. This is the interpretation of $c$ for standing waves: it is the constant of proportionality that links the wavenumber $k$ to the angular frequency $\omega_k$.

**Quantum mechanics terminology**

These have the form of an *eigenvalue* problem, where $X(x)$ must be an *eigenfunction* of the *differential operator* $\frac{d^2}{dx^2}$ with *eigenvalue* $-k^2$. Similarly $T(t)$ must be an eigenfunction of $\frac{d^2}{dt^2}$ with eigenvalue $-\omega_k^2 = -c^2k^2$.

### 2.2.2 Solving the ODE's

We can now solve the two ODEs separately. The solutions to these are familiar from simple harmonic motion, and we can just write down the solutions:

$$\begin{aligned}X(x) =& A_k \sin kx + B_k \cos kx\\ T(t) =& C_k \sin \omega_k t + D_k \cos \omega_k t\\ \Rightarrow u(x,t) =& \left(A_k \sin kx + B_k \cos kx\right)\left(C_k \sin \omega_k t + D_k \cos \omega_k t\right)\end{aligned}$$

where $A_k$, $B_k$, $C_k$, and $D_k$ are arbitrary constants. The subscript denotes that they can take different values for different values of $k$. At this stage there is no restriction on the values of $k$: each values provides a separate solution to the ODEs.

### 2.2.3 Boundary conditions

The details of a specific physical system may involve the boundary conditions (BCs) solutions must satisfy. For example, what happens at the ends of the string and what were the initial conditions.

- The string weight & tension on a guitar determine $c$.
- The length (& frets) of a guitar determine the boundary conditions.
- The plucking of the guitar determines the initial conditions.

Assume the string is stretched between $x = -L$ and $x = L$, then the BCs in this case are that

$$u(x = -L, t) = u(x = L, t) = 0$$

for all $t$. Because these BCs hold for all times at specific $x$, they affect $X(x)$ rather than $T(t)$. We find

$$\begin{aligned} u(0,t) = 0 \quad &\Rightarrow \quad B_k = 0 \,, \\ u(L,t) = 0 \quad &\Rightarrow \quad k_n = n\pi/L \,, \qquad n = 0, 1, 2 \ldots \end{aligned}$$

Here, BCs have restricted the allowed values of $k$ and thus the allowed frequencies of oscillation. Different boundary conditions will have different allowed values. Restriction of eigenvalues by boundary conditions is a very general property in physics:

| finite boundaries ⇒ discrete (quantised) eigenvalue spectrum ⇒ allowable separation constants. |
|---|

Each $n$ value corresponds to a *normal mode* of the string:

$$u(x,t) = A_n \sin k_n x \{C_n \sin \omega_n t + D_n \cos \omega_n t\}$$

A normal mode is an excitation of the string that obeys the BCs and oscillates with a single, normal mode frequency. We sometimes call these *eigenmodes* of the system, with associated *eigenfrequencies* $\omega_n = \omega_{k_n}$.

#### Completeness

Just like any vector can be represented as a linear combination of basis vectors, so the *general solution* to the wave equation is a linear superposition of (normal) eigenmode solutions:

$$\begin{aligned} u(x,t) &= \sum_{n=1}^{\infty} A_n \sin k_n x \{C_n \sin \omega_n t + D_n \cos \omega_n t\} \\ &\equiv \sum_{n=1}^{\infty} \sin k_n x \{E_n \sin \omega_n t + F_n \cos \omega_n t\} \end{aligned} \tag{2.2}$$

This normal mode decomposition *not* obvious and the proof is beyond the scope of this course. We will simply assume this to be true.

In fact, almost *any* function can be described by such a linear combination of normal modes.

| Completeness of the normal modes is general and applies to all "Sturm Liouville" ODE's |
|---|

As before $\omega_n = ck_n$. We sum only from $n = 1$ because $\sin k_0 x = 0$, and we do not need to include negative $n$ because $\sin \frac{-n\pi x}{L} = -\sin \frac{n\pi x}{L}$. Constants $A_n$, $C_n$, $D_n$ are all unknown, so we can merge them together to give $E_n = A_n C_n$ and $F_n = A_n D_n$.

We also see that the way we have ensured that $u(0,t) = 0$ is by making it an *odd function* in $x$: $u(-x,t) = -u(x,t) \Rightarrow u(0,t) = 0$.

### 2.2.4 Initial conditions

As we have a second order temporal ODE, we need two sets of initial conditions to solve the problem. Typically these are the the shape $f(x)$ and velocity profile $g(x)$ of the string at $t = 0$:

$$u(x,0) = f(x) = \sum_{n=1}^{\infty} F_n \sin k_n x$$
$$\dot{u}(x,0) = g(x) = \sum_{n=1}^{\infty} \omega_n E_n \sin k_n x$$

These conditions determine unique values for each of the $E_n$ and $F_n$. Having got these, we can substitute them back into the general solution to obtain $u(x,t)$ and thus describing the motion for all times.

Consider the equation for $F_n$. Let's choose to calculate one, specific constant out of this set i.e. $F_m$ for some specific $m$. To do this, multiply both sides by $\sin k_m x$ and integrate over the whole string (in this case $x = 0...L$) giving:

$$\int_0^L dx\ f(x) \sin k_m x = \sum_{n=1}^{\infty} F_n \int_0^L dx\ \sin k_n x\ \sin k_m x\ .$$

Now we note that the sine functions form an *orthogonal set*:

$$\begin{aligned}
\int_0^L dx\ \sin k_n x\ \sin k_m x &= \frac{1}{2}\int_0^L dx\ [\cos(k_n x - k_m x) - \cos(k_n x + k_m x)] \\
&= \frac{1}{2}\begin{cases} \left[\frac{\sin(k_n x - k_m x)}{k_n - k_m} - \frac{\sin(k_n x + k_m x)}{k_n + k_m}\right]_0^L & ;\quad n \neq m \\ \left[x - \frac{\sin(k_n x + k_m x)}{k_n + k_m}\right]_0^L & ;\quad n = m \end{cases} \\
&= \frac{L}{2}\delta_{mn}
\end{aligned}$$

where $\delta_{mn}$ is the Kronecker delta, giving zero for $m \neq n$

| The *orthogonality* of normal modes is general and applies to all "Sturm Liouville" ODE's. |
|---|

So:

$$\begin{aligned}\int_0^L dx\ f(x)\sin k_m x =& \sum_{n=1}^{\infty} F_n \int_0^L dx\ \sin k_n x\ \sin k_m x \\ =& \frac{L}{2}\sum_{n=1}^{\infty} F_n \delta_{mn} \\ =& \frac{L}{2} F_m\end{aligned}$$

using the sifting property. Therefore, after relabelling $m \to n$:

$$\begin{aligned}F_n =& \frac{2}{L}\int_0^L dx\ f(x)\sin k_n x \\ E_n =& \frac{2}{L\omega_n}\int_0^L dx\ g(x)\sin k_n x\ .\end{aligned} \tag{2.3}$$

We are given $f$ and $g$, so as long as we can do the integrals on the RHS, we have determined all the unknown constants and therefore know the motion for all times.

The solution written as a sum of sine waves is an example of a Fourier series.

### A quick example

Suppose the initial conditions are that the string is initially stretched into a sine wave $f(x) = a\sin(3\pi x/L)$ (for some $a$) and at rest, i.e. $g(x) = 0$.

The latter immediately gives $E_n = 0$ for all $n$. The former gives:

$$\begin{aligned}F_n =& \frac{2}{L}\int_0^L dx\ f(x)\sin k_n x \\ =& \frac{2a}{L}\int_0^L dx\ \sin\frac{3\pi x}{L}\sin\frac{n\pi x}{L} = \frac{2a}{L}\times\frac{L}{2}\delta_{n3}\end{aligned}$$

using the above relation. So all the $F_n$ are zero except $F_3 = a$. So the motion is described by

$$u(x,t) = a\sin\frac{3\pi x}{L}\cos\frac{3\pi ct}{L}\ .$$

The answer is very simple. If the system starts as a pure *normal mode* of the system, it will remain as one.

## 2.3 Method of Froebenius

A general method that works for more complicated equations can be illustrated by pretending we do not know the solution to the wave equation.

### 2.3.1 Bill and Ted's excellent misadventure

Bill and Ted have brought Pythagoras to the future and lost him in a night club. We now live in a world without sin and cos. To rectify this we will use the method of Froebenius to rediscover these precious functions

1. Substitute the infinite series $y(x) = \sum_{n=0}^{\infty} C_n x^n$ to the differential equation

$$y'' + y = 0$$

We end up with two sums.

$$\sum_{n=0}^{\infty} c_n n(n-1)x^{n-2} + \sum_{n=0}^{\infty} c_n x^n = 0$$

2. Relabel the summation using $m = n - 2$ on the $y''$ term obtaining

$$\sum_{m=-2}^{\infty} c_{m+2}(m+2)(m+1)x^m + \sum_{n=0}^{\infty} c_n x^n = 0$$

3. Use a notation where $C_i = 0$ for $i < 0$ to sum the $y$ term over the range $\sum_{n=-2}^{\infty}$

$$\sum_{m=-2}^{\infty} c_{m+2}(m+2)(m+1)x^m + \sum_{m=-2}^{\infty} c_m x^m = \sum_{m=-2}^{\infty} [c_{m+2}(m+2)(m+1) + c_m]x^m = 0$$

4. As this is true for all values of $x$, each Hence we obtain the *indicial equation*

$$C_{m+2}(m+1)(m+2) = -C_m$$

This relates every other coefficient in a *recurrence relation.*

5. Deduce that $C_0$ can be non-zero even though $C_{-2} = 0$, and that $C_1$ can be non-zero even though $C_{-1} = 0$ because

$$(m+2)(m+1) = 0$$

for $m = -1, -2$

We therefore have two independent series, and two free parameters $C_0$ and $C_1$ as should be the case for a 2nd order ODE.

6. We therefore find the series with

   (a) $C_0 = 1$, $C_1 = 0$

   $$\sum_{n=0}^{\infty}(-1)^{(n)}\frac{x^{2n}}{(2n)!} = 1 - \frac{x^2}{2.1} + \frac{x^4}{4.3.2.1}\ldots$$

   (b) $C_0 = 0$, $C_1 = 1$

   $$\sum_{n=0}^{\infty}(-1)^{(n)}\frac{x^{(2n+1)}}{(2n+1)!} = x - \frac{x^3}{3.2} + \frac{x^5}{5.4.3.2}\ldots$$

   Giving these two independent series their names we recognise

   $$\cos x \equiv 1 - \frac{x^2}{2!} + \frac{x^4}{4!}\ldots$$

   $$\sin x \equiv x - \frac{x^3}{3!} + \frac{x^5}{5!}\ldots$$

7. We can now make up the world's first table of sinusoids by summing the series to high order!

   Towards the end of the 1800's enormous effort was expended computing special functions by hand to high order in the Taylor expansion, and tabulating the values as a function of $x$.

   In fact, prior to scientific calculators it was common for laboriously computed Tables of Sines to be handed out in mathematics examinations, even in the 1980's.

## 2.4 Fourier Series

In the previous section we made things easy by considering the stretched string. The boundary conditions were deliberately chosen to give us only sine solutions. Now we will consider the more general case.

### 2.4.1 Overview

Fourier series are a way of decomposing a function as a sum of sine and cosine waves. We say that the solutions of an ODE are *complete* because as the number of Fourier modes included is taken to $\infty$ the Fourier series will completely describe any function (there is a mathematically precise statement of this).

Fourier series are particularly useful if we are looking at system that satisfies a wave equation, because sinusoids are the *normal mode* oscillations which have simple time dependence. We can use this decomposition to understand more complicated excitations of the system.

Fourier series describe a *finite interval* of a function, typically $-L \to L$ or $0 \to L$. If the size of the system is infinite we instead need to use Fourier transforms.

Outside this range a Fourier series is periodic (repeats itself) because all the sine and cosine waves are themselves periodic. The Fourier series *periodically extends* the function outside the range.

Fourier series are useful in the following contexts:

- The function really is periodic e.g. a continuous square wave
- We are only interested in what happens inside the expansion range.

**Fourier modes** Consider the wave equation on an interval $[-L, L]$ with periodic boundary conditions

$$\frac{d^2X(x)}{dx^2} = -k^2X(x)$$
$$X(x+2L) = X(x)$$

The solutions look like

$$\begin{aligned} & X(x) = a_k\psi_k(x) + b_k\phi_k(x) \ , \\ \text{where} \quad & \psi_k(x) \equiv \cos kx \ , \\ & \phi_k(x) \equiv \sin kx \ . \end{aligned}$$

$a_k$, $b_k$ are unknown constants. Now, periodicity condition means $\cos k(x+2L) = \cos kx$ and $\sin k(x+2L) = \sin kx$. This is satisfied if $2kL = 2n\pi$ or

$$k = \frac{n\pi}{L}$$

for $n$ an integer.

### 2.4.2 The Fourier expansion

The set of Fourier modes $\{\psi_{n\geq 0}(x), \phi_{n\geq 1}(x)\}$ are therefore defined as:

$$\begin{aligned}\psi_n(x) &= \cos\frac{n\pi x}{L}\ ,\\ \phi_n(x) &= \sin\frac{n\pi x}{L}\end{aligned} \tag{4.1}$$

Between $-L \leq x \geq L$ we can write a general real function as a *linear combination* of these Fourier modes:

$$\begin{aligned}f(x) &= \sum_{n=0}^{\infty} a_n\psi_n(x) + \sum_{n=1}^{\infty} b_n\phi_n(x)\\ &= a_0 + \sum_{n=1}^{\infty}\left(a_n\psi_n(x) + b_n\phi_n(x)\right)\end{aligned} \tag{4.2}$$

where $a_n$ and $b_n$ are (real-valued) *Fourier coefficients*.

### 2.4.3 Orthogonality

Having written a function as a sum of Fourier modes, we would like to be able to calculate the components. This is made easy because the Fourier mode functions are *orthogonal* i.e.

$$\begin{aligned}\int_{-L}^{L} dx\ \psi_m(x)\psi_n(x) &= N_n^{\psi}\delta_{mn}\ ,\\ \int_{-L}^{L} dx\ \phi_m(x)\phi_n(x) &= N_n^{\phi}\delta_{mn}\ ,\\ \int_{-L}^{L} dx\ \psi_m(x)\phi_n(x) &= 0\ .\end{aligned}$$

$N_{n\geq 0}^{\psi}$ and $N_{n\geq 1}^{\phi}$ are *normalisation constants* which we find by doing the integrals using the trig. identities in Eqn. (4.3) below. It turns out that $N_n^{\psi} = N_n^{\phi} = L$ for all $n$, except $n = 0$ when $N_0^{\psi} = 2L$.

**ASIDE: useful trig. relation** To prove the orthogonality, the following double angle formulae are useful:

$$\begin{aligned}2\cos A\cos B &= \cos(A+B) + \cos(A-B)\\ 2\sin A\cos B &= \sin(A+B) + \sin(A-B)\\ 2\sin A\sin B &= -\cos(A+B) + \cos(A-B)\\ 2\cos A\sin B &= \sin(A+B) - \sin(A-B)\end{aligned} \tag{4.3}$$

### 2.4.4 Calculating the Fourier coefficients

The orthogonality proved above allows us to calculate the Fourier coefficients as follows:

$$a_m = \begin{cases} \dfrac{1}{2L}\displaystyle\int_{-L}^{L} dx\ \psi_m(x) f(x) & m = 0\ , \\ \dfrac{1}{L}\displaystyle\int_{-L}^{L} dx\ \psi_m(x) f(x) & m > 0\ , \end{cases}$$

$$\text{Similarly,} \quad b_m = \frac{1}{L}\int_{-L}^{L} dx\ \phi_m(x) f(x)\ .$$

Proof: suppose $f(x) = \sum_k a_k \psi_k + b_k \phi_k$, then

$$\begin{aligned} \frac{1}{N_j^\psi}\int_{-L}^{L} \psi_j f(x) dx &= \sum_k \frac{a_k}{N_j^\psi}\int_{-L}^{L} \psi_j \psi_k + \frac{b_k}{N_j^\psi}\int_{-L}^{L} \psi_j \phi_k dx \\ &= \sum_k a_k \delta_{jk} + b_k \times 0 \\ &= a_j \end{aligned}$$

The proof for $b_k$ is very similar.

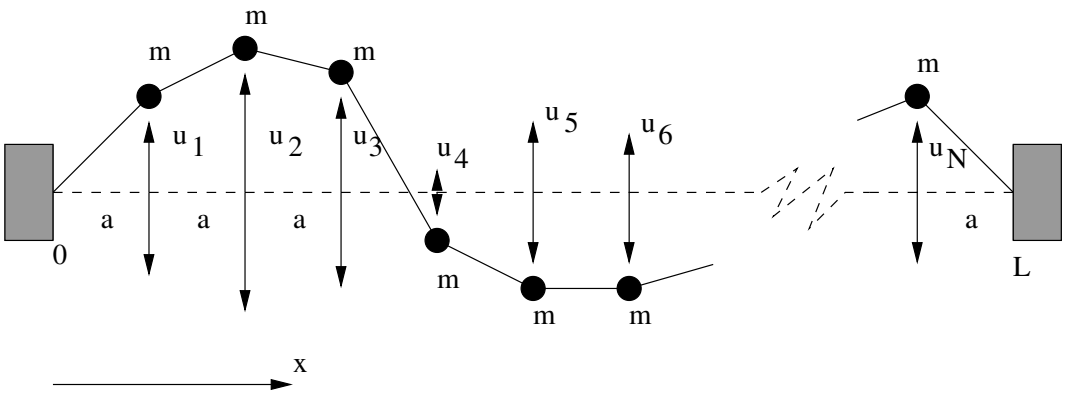



Figure 5.1: Transverse vibrations of $N$ masses $m$ attached to a stretched string of length $L$.

## 2.5 Functions as vectors

### 2.5.1 Scalar product of functions

In physics we often make the transition:

$$\text{discrete picture} \rightarrow \text{continuous picture.}$$

Take $N$ particles of mass $m = \frac{M}{N}$ evenly spaced on a massless string of length $L$, under tension $T$. The transverse displacement of the $n^{\text{th}}$ particle is $u_n(t)$. As there are $N$ coordinates we can think of the motion as occurring in an $N$-dimensional space.

The gap between particles is $a = L/(N+1)$, and we can label the $n^{\text{th}}$ particle with $x = x_n \equiv na$. We can also write $u_n(t)$, the transverse displacement of the $n^{\text{th}}$ particle, as $u(x_n, t)$.

**Transition to the continuum** Let the number of masses $N$ become infinite, while holding $L = Na$ and $M = Nm$ fixed. We go from thinking about motion of masses on a massless string to oscillations of a massive string. As $N \rightarrow \infty$, we have $a \rightarrow 0$ and $x_n \equiv na \rightarrow x$: a continuous variable. The $N$-component displacement vector $u(x_n, t)$ becomes a continuous function $u(x, t)$.

**Scalar product $\rightarrow$ integral**

In $N$-dimensions the inner product of vectors $\boldsymbol{f} = \{f_1, f_2, \ldots f_N\}$ and $g = \{g_1, g_2, \ldots g_N\}$, is:

$$\boldsymbol{f} \cdot \boldsymbol{g} = f_1^* g_1 + f_2^* g_2 + \ldots + f_N^* g_N = \sum_{n=1}^{N} f_n^* g_n$$

$$\equiv \sum_{n=1}^{N} f(x_n)^* g(x_n)$$

Again, we have moved the $n$ dependence inside the brackets for convenience.

In an interval $x \to x + dx$ there are $\Delta n = dx/a$ particles. So, for large $N$ we can replace a sum over $n$ by an integral with respect to $dx/a$: the sum becomes a definite integral. Multiplying through by this factor of $a$ ,

$$a\boldsymbol{f} \cdot \boldsymbol{g} \equiv a \sum_{n=1}^{N} f(x_n)^* g(x_n) \underset{N\to\infty}{\longrightarrow} \int_0^L dx\ f(x)^*\ g(x)$$

The conclusion is that there is a strong link between the inner product of two vectors and the inner product of two functions.

### 2.5.2 Inner products, orthogonality, orthonormality and Fourier series

We want the inner product of a function with itself to be *positive-definite*, i.e. $f \cdot f \geq 0$ meaning it is a real, non-negative number and that $|f|^2 = f \cdot f = 0 \Rightarrow f(x) = 0$.

That is, the *norm* of a function is only zero if the function $f(x)$ is zero everywhere.

For real-valued functions of one variable (e.g $f(x)$, $g(x)$) we *choose to define* the inner product as

$$\begin{aligned} f \cdot g &\equiv \int dx\ f(x).g(x) \\ &= g \cdot f \end{aligned}$$

and for complex-valued functions

$$\begin{aligned} f \cdot g &\equiv \int dx\ f(x)^*.g(x) \\ &= (g \cdot f)^* \neq g \cdot f \end{aligned}$$

The integration limits are chosen according to the problem we are studying. For Fourier analysis we use $-L \to L$. For the special case of waves on a string we used $0 \to L$.

**Normalised:** We say a function is *normalised* if the inner product of the function with itself is 1, i.e. if $f \cdot f = 1$.

**Orthogonal:** We say that two functions are *orthogonal* if their inner product is zero, i.e. if $f \cdot g = 0$. If we have a set of functions for which any pair of functions are orthogonal, we call it an *orthogonal set*, i.e. if $\phi_m \cdot \phi_n \propto \delta_{mn}$ for all $m$, $n$.

**Orthonormal:** If all the functions in an orthogonal set are normalised, we call it an *orthonormal set* i.e. if $\phi_m \cdot \phi_n = \delta_{mn}$.

#### Example: complex Fourier series

An example of an orthogonal set is the set of complex Fourier modes $\{\varphi_n\}$. We can decompose any function $f(x)$ as

$$f(x) = \sum_{n=-\infty}^{\infty} c_n \varphi_n(x) \quad \text{where} \quad \varphi_n(x) = e^{-ik_n x} = e^{-in\pi x/L} \tag{5.1}$$

We define the inner product as

$$f \cdot g = \int_{-L}^{L} dx\ f(x)^* g(x)$$

and we see the orthogonality:

$$\varphi_m \cdot \varphi_n \equiv \int_{-L}^{L} dx\ \varphi_m(x)^* \varphi_n(x) = N_n\ \delta_{mn}$$

with normalisation $N_n \equiv \varphi_n \cdot \varphi_n = 2L$. Then

$$f(x) = \sum_{n=-\infty}^{\infty} c_n \varphi_n(x) \qquad \Rightarrow \quad c_n = \frac{\varphi_n \cdot f}{\varphi_n \cdot \varphi_n}$$

In more detail, the numerator is

$$\varphi_n \cdot f \equiv \int_{-L}^{L} dx\ \varphi_n^*(x) f(x)$$

Note that the order of the functions is *very important*: the basis function comes first.

In each case, we can *projected out* the relevant Fourier component by exploiting the fact that the *basis functions* formed an *orthogonal set*. The same can be done for the real Fourier series which we leave as an exercise.

### Normalised basis functions

Consider the complex Fourier series. If we define some new functions

$$\widehat{\varphi}_n(x) \equiv \frac{1}{\sqrt{\varphi_n \cdot \varphi_n}}\ \varphi_n(x) = \frac{\varphi_n(x)}{\sqrt{2L}}$$

then it should be clear that

$$\widehat{\varphi}_m \cdot \widehat{\varphi}_n = \delta_{mn}$$

giving us an *orthonormal set*. The inner product is defined exactly as before.

We can *choose* to use these *normalised* functions as a Fourier basis, with new expansion coefficients $C_n$:

$$f(x) = \sum_{n=-\infty}^{\infty} C_n \widehat{\varphi}_n(x) \qquad \Rightarrow \quad C_n = \widehat{\varphi}_n \cdot f$$

because the denominator $\widehat{\varphi}_n \cdot \widehat{\varphi}_n = 1$.

Coefficients $C_n$ and $c_n$ are closely related:

$$C_n = \widehat{\varphi}_n \cdot f = \frac{\varphi_n \cdot f}{\sqrt{\varphi_n \cdot \varphi_n}} = \sqrt{\varphi_n \cdot \varphi_n}\ \frac{\varphi_n \cdot f}{\varphi_n \cdot \varphi_n} = \sqrt{\varphi_n \cdot \varphi_n}\ c_n = \sqrt{2L}\ c_n$$

We can do the same for real Fourier series, defining $\widehat{\psi}_n$ and $\widehat{\phi}_n$ and associated Fourier coefficients $A_n$ and $B_n$. The relationship to $a_n$ and $b_n$ can be worked out in exactly the same way.

## 2.6 Differential operators

### 2.6.1 Partial derivatives

If $f(x, y)$ is a function of two variables, we can define partial derivatives with respect to each variable.

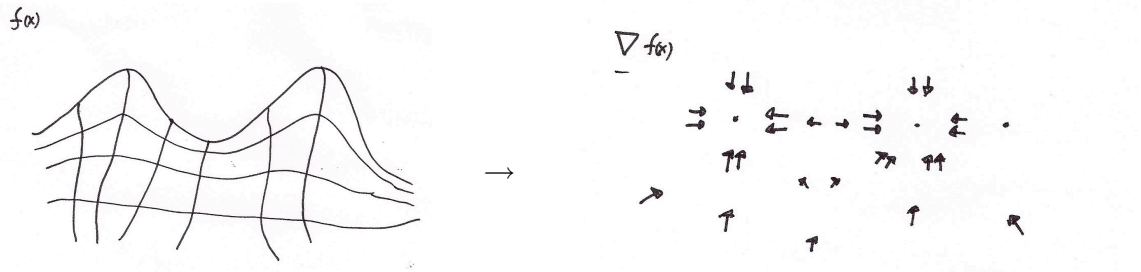
$$\begin{aligned} \frac{\partial}{\partial_x} f(x,y) &= \lim_{\epsilon \to 0} \frac{f(x+\epsilon, y) - f(x,y)}{\epsilon} \\ \frac{\partial}{\partial_y} f(x,y) &= \lim_{\epsilon \to 0} \frac{f(x, y+\epsilon) - f(x,y)}{\epsilon} \end{aligned} \tag{6.2}$$

### 2.6.2 Gradient operator

We define the gradient operator *grad* as:

$$\boldsymbol{\nabla} f(\boldsymbol{x}) = \boldsymbol{e}_i \partial_i f(\boldsymbol{x}) = \hat{x} \frac{\partial}{\partial_x} f(\boldsymbol{x}) + \hat{y} \frac{\partial}{\partial_y} f(\boldsymbol{x}) + \hat{z} \frac{\partial}{\partial_z} f(\boldsymbol{x}) \tag{6.3}$$

In two dimensions this is very familiar. If $f(x, y)$ is a function of two variables (a height as a function of map coordinates), then $\boldsymbol{\nabla} f(x, y)$ is just a two component vector indicating the steepness and direction of slope.




**Example: Electrostatic field**

The electric field in one dimension is given by $E(x) = -\frac{dV(x)}{dx}$.

In three dimensions this clearly generalises as follows:

$$\begin{aligned} E_x(x,y,z) &= -\frac{\partial V(x,y,z)}{\partial_x} \\ E_y(x,y,z) &= -\frac{\partial V(x,y,z)}{\partial_y} \\ E_z(x,y,z) &= -\frac{\partial V(x,y,z)}{\partial_z}. \end{aligned} \tag{6.4}$$

This is rather succintly written as

$$\boldsymbol{E} = -\boldsymbol{\nabla} V = -(\frac{\partial V(x,y,z)}{\partial_x}, \frac{\partial V(x,y,z)}{\partial_y}, \frac{\partial V(x,y,z)}{\partial_z})$$

Here, $V$ is a scalar, $\boldsymbol{\nabla}$ is a vector indexed differential operator, and the result $\boldsymbol{E}$ is vector.

The direction of the gradient $\boldsymbol{\nabla}$ of a function picks out the direction of steepest upwards slope (i.e. the opposite direction to a skiers "fall-line") automatically.

### 2.6.3 Divergence of a vector function

Consider a *vector field* $\boldsymbol{v}(x,y,z) = \boldsymbol{v}(\boldsymbol{x})$.

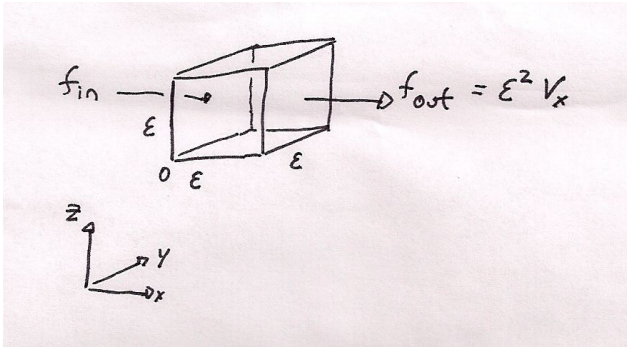
The divergence of a vector field is the scalar product

$$\boldsymbol{\nabla} \cdot \boldsymbol{v}(\boldsymbol{x}) = \partial_i v_i = \frac{\partial}{\partial_x} v_x(\boldsymbol{x}) + \frac{\partial}{\partial_y} v_y(\boldsymbol{x}) + \frac{\partial}{\partial_z} v_z(\boldsymbol{x}).$$

**Example: Oil from a well**

To illustrate the meaning of *divergence* consider an idealised blow-out preventer as a small pipe releasing crude oil into the Gulf stream at huge environmental cost.

We denote the velocity field $\boldsymbol{v}(\boldsymbol{x})$, and consider a cubical box of side $\epsilon$, volume $\epsilon^3$, containing the outlet. The box is taken sufficiently small that Taylor expansion of the $\boldsymbol{v}(\boldsymbol{x})$ works.




If we consider the $\hat{x}$ faces first, the inflow (for positive $v_x$), in $m^3 s^1$, through face-A is

$$f_{\text{in}} = \epsilon^2 v_x(0, \frac{\epsilon}{2}, \frac{\epsilon}{2}) + O(\epsilon)^4).$$

We can always take $\epsilon$ small enough that this is a good approximation. The net outflow through face-B is is

$$f_{\text{out}} = \epsilon^2 v_x(\epsilon, \frac{\epsilon}{2}, \frac{\epsilon}{2}) + O(\epsilon)^4) = \epsilon^2 [v_x(0, \frac{\epsilon}{2}, \frac{\epsilon}{2}) + \epsilon \frac{\partial}{\partial_x} v_x(0, \frac{\epsilon}{2}, \frac{\epsilon}{2})]$$

Thus the net outflow from these faces is

$$f_{\text{out}} - f_{\text{in}} = \epsilon^3 \frac{\partial}{\partial_x} v_x(0, \frac{\epsilon}{2}, \frac{\epsilon}{2}).$$

The other directions are similar and the net outflow per unit volume is

$$\frac{\partial}{\partial_x}v_x + \frac{\partial}{\partial_y}v_y + \frac{\partial}{\partial_z}v_z = \boldsymbol{\nabla}\cdot\boldsymbol{v}.$$

Thus we can identify the divergence of a vector field as the *outflow per unit volume*.

### 2.6.4 Curl of a vector function

The vector cross product can also be used to define the *curl* of a vector. The name *curl* was coined by James Clerk Maxwell.

$$\boldsymbol{\nabla}\times\boldsymbol{v}(\boldsymbol{x}) = (\frac{\partial}{\partial_y}v_z(\boldsymbol{x}) - \frac{\partial}{\partial_z}v_y(\boldsymbol{x}), \frac{\partial}{\partial_z}v_x(\boldsymbol{x}) - \frac{\partial}{\partial_x}v_z(\boldsymbol{x}), \frac{\partial}{\partial_x}v_y(\boldsymbol{x}) - \frac{\partial}{\partial_y}v_x(\boldsymbol{x}))$$

**Curl on a square**

Consider an infinitessimal square in the $x-y$ plane. The $z$-component of curl is

$$\frac{\partial}{\partial_x}v_y(\boldsymbol{x}) - \frac{\partial}{\partial_y}v_x(\boldsymbol{x}) \simeq \frac{1}{\epsilon}\left[v_y(\epsilon,0,0) - v_y(0,0,0) + v_x(0,\epsilon,0) - v_x(0,0,0)\right]$$

This just measures the difference in the parallel component of $\boldsymbol{v}$ between opposite sides, and is thus a measure of how our square would rotate if it were suspended in a fluid flow $\boldsymbol{v}$.

### 2.6.5 Laplacian

The Laplacian operator acting on a scalar function $f$ is

$$\boldsymbol{\nabla}^2 f(\boldsymbol{x}) = \boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}f(\boldsymbol{x})) = (\frac{\partial^2}{\partial_x^2} + \frac{\partial^2}{\partial_y^2} + \frac{\partial^2}{\partial_z^2})f(\boldsymbol{x}),$$

.

We define the Laplacian of a vector function $\boldsymbol{v}$ as

$$\boldsymbol{\nabla}^2\boldsymbol{v}(\boldsymbol{x}) = (\boldsymbol{\nabla}^2\boldsymbol{v}_x(\boldsymbol{x}), \boldsymbol{\nabla}^2\boldsymbol{v}_y(\boldsymbol{x}), \boldsymbol{\nabla}^2\boldsymbol{v}_z(\boldsymbol{x}))$$

The Laplacian is a measure of the total curvature, summed across all three dimensions.

- A (symmetrical) saddle point has zero Laplacian.
- A minimum has positive Laplacian
- A maximum has negative Laplacian

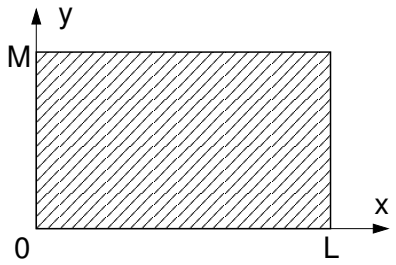



Figure 7.1: A vibrating rectangular membrane with displacement normal to the page.

## 2.7 Multi-dimensional differential equations

In one (spatial) dimension, the wave equation read

$$\frac{\partial^2}{\partial x^2}u(x,t) = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}u(x,t)\ . \tag{7.1}$$

In a general number of dimensions, it can be written in a coordinate independent way as

$$\nabla^2 u(\boldsymbol{r},t) = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}u(\boldsymbol{r},t)\ , \tag{7.2}$$

where $\nabla^2 \equiv$ div grad is known as the **Laplacian operator** and $\boldsymbol{r}$ is a position vector.

If we are working in Cartesian coordinates $\boldsymbol{r} = (x, y, z, \ldots)$ the Laplacian takes the form

$$\nabla^2 = \begin{cases} \frac{\partial^2}{\partial x^2} & d = 1\ , \\ \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} & d = 2\ , \\ \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} & d = 3\ , \end{cases}$$

and so on. Later we will see what form it takes in other, curvilinear coordinate systems.

### 2.7.1 Waves in a rectangular membrane

| Given a physical problem, we should always choose a coordinate system that best matches the boundary conditions. |
|---|

For a rectangular membrane, a Cartesian coordinate system is best. (For a circular membrane, circular polar coordinates are better. We shall see later that they bring their own complications.)

The membrane (of size $L \times M$) is shown in Fig. 7.1, with edges fixed and oscillations in and out of the page. The displacement from equilibrium $u(x, y, t)$ satisfies the wave equation:

$$\nabla^2 u \equiv \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2}\ .$$

To solve this, we consider solutions of separated form:

$$u(x,y,t) = X(x)\,Y(y)\,T(t)\ .$$

Substituting this into the wave equation, dividing by $u = XYT$ and rearranging, we get

$$\frac{1}{X}\frac{d^2X}{dx^2} + \frac{1}{Y}\frac{d^2Y}{dy^2} = \frac{1}{c^2T}\frac{d^2T}{dt^2} = -k^2 \ .$$

We have used the same argument as before that both sides are constant: the LHS is independent of $t$, whereas the RHS depends only on $t$. As we expect the solutions to have an oscillatory form, we have chosen the separation constant to be negative and squared.

We can further rearrange the LHS equation to read

$$\frac{1}{X}\frac{d^2X}{dx^2} = -k^2 - \frac{1}{Y}\frac{d^2Y}{dy^2} = -p^2$$

Again, the LHS depends only on $x$ and the RHS only on $y$, so they must both be equal to a second separation constant. We now have three separate differential equations to solve:

$$\frac{d^2X}{dx^2} = -p^2X \ , \qquad \frac{d^2Y}{dy^2} = -q^2Y \ , \qquad \frac{d^2T}{dt^2} = -\omega_k^2T \ .$$

with $k^2 = p^2 + q^2$ and $\omega_k = ck$. The solution may be written as:

$$u(x, y, t) = (A\cos px + B\sin px)\,(C\cos qy + D\sin qy)\,(G\cos\omega_k t + H\sin\omega_k t) \ .$$

**Spatial boundary conditions: the normal modes**

The membrane is fixed at the boundary, so:

- $u(x = 0, y, t) = 0$ for all $y, t$, implying $A = 0$.
- $u(x = L, y, t) = 0$ for all $y, t$, implying $\sin pL = 0$ so $p = p_m = m\pi/L$ for integer $m$.
- $u(x, y = 0, t) = 0$ for all $x, t$, implying $C = 0$.
- $u(x, y = M, t) = 0$ for all $x, t$, implying $\sin qM = 0$ so $q = q_n = n\pi/M$ for integer $n$.

The normal mode vibrations therefore take the form

$$u(x, y, t) = \sin\left(\frac{m\pi x}{L}\right)\,\sin\left(\frac{n\pi y}{M}\right)\,[E_{mn}\cos\omega_{mn}t + F_{mn}\sin\omega_{mn}t] \,,$$

where $E_{mn}$ and $F_{mn}$ are unknown constants that will be determined from the initial conditions (i.e. the temporal boundary conditions).

Each normal mode is labelled by integers $(m, n)$, with an oscillation frequency that increases with both $m$ and $n$

$$\omega_{mn} = ck_{mn} = c\sqrt{p_m^2 + q_n^2} = c\sqrt{\left(\frac{m\pi}{L}\right)^2 + \left(\frac{n\pi}{M}\right)^2}$$

In the $(m, n)$ mode:

- lines of zero displacement are *nodal lines*

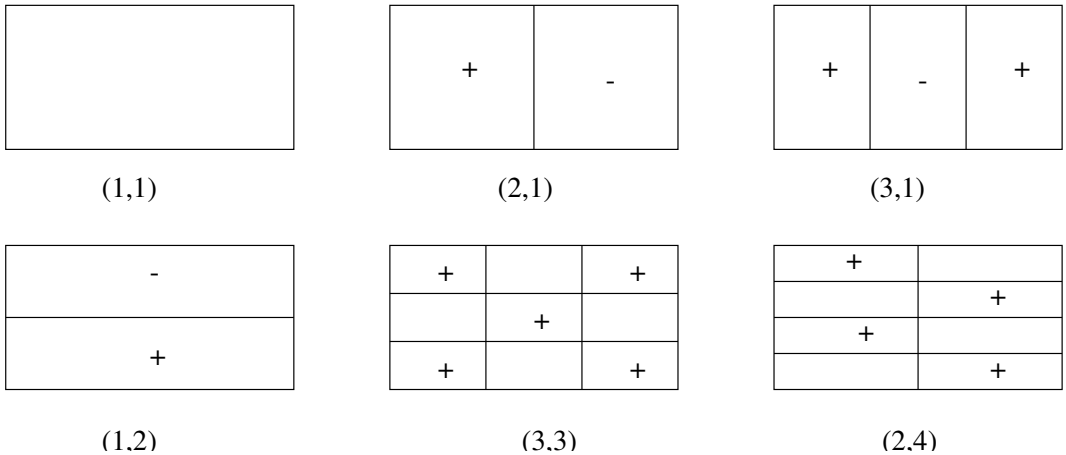



Figure 7.2: Rectangular membrane: normal modes of vibration.

- there are nodal lines at $x = 0, L/m, 2L/m, \ldots L$
- there are nodal lines at $y = 0, M/n, 2M/n, \ldots M$
- on opposite sides of any nodal line, the amplitude has opposite sign.

Fig. 7.2 shows some of the low-lying normal modes. "±" denotes whether the membrane is above or below its equilibrium position. Only the relative sign matters, of course: the whole membrane is oscillating up and down.

**Initial conditions: the rectangular harmonics**

The general solution is a linear superposition of normal modes, which looks just like a two dimensional Fourier series:

$$u(x, y, t) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} R_{mn}(x, y) \left[ E_{mn} \cos \omega_{mn} t + F_{mn} \sin \omega_{mn} t \right] \tag{7.3}$$

where $R_{mn}$ are what we might call the **rectangular harmonics**:

$$R_{mn}(x, y) = \sin\left(\frac{m\pi x}{L}\right) \sin\left(\frac{n\pi y}{M}\right) . \tag{7.4}$$

They form a complete, orthogonal basis set, satisfying the orthonormality condition:

$$R_{mn} \cdot R_{uv} \equiv \int_0^L dx \int_0^M dy \; R_{mn}(x, y)^* \; R_{uv}(x, y) = R_{mn} \cdot R_{mn} \; \delta_{mu} \; \delta_{nv} = \frac{LM}{4} \delta_{mu} \; \delta_{nv} . \tag{7.5}$$

Consider an example where we are told that initially the membrane is at rest (implying $F_{mn} = 0$ for all modes) and stretched into the form of a given function $p(x, y)$:

$$u(x, y, t = 0) = p(x, y) = \sum_{m,n} E_{mn} R_{mn}(x, y) .$$

Using the orthonormality, we can project out the components:

$$R_{uv} \cdot p \equiv \int_0^L dx \int_0^M dy\; R_{uv}(x,y)\; p(x,y) = \sum_{mn} E_{mn}\; (R_{uv} \cdot R_{mn}) = \sum_{mn} E_{mn}\; (R_{uv} \cdot R_{uv})\; \delta_{um}\; \delta_{vn}$$
$$= \frac{LM}{4} \sum_{mn} E_{mn}\; \delta_{um}\; \delta_{vn} = \frac{LM}{4} E_{uv}$$
$$\Rightarrow \quad E_{uv} \equiv \frac{R_{uv} \cdot p}{R_{uv} \cdot R_{uv}} = \frac{4}{LM} \int_0^L dx \int_0^M dy\; R_{uv}(x,y)\; p(x,y)$$

and with these coefficients and Eqn. (7.3) we know the motion for all subsequent times.

### 2.7.2 More dimensions

If we solve the wave equation in higher dimensional systems with rectangular boundaries, we find a quantised separation constant for each dimension. The squared frequency is proportional to the sum of the squares of these constants. Try it for three dimensions to see how it works.

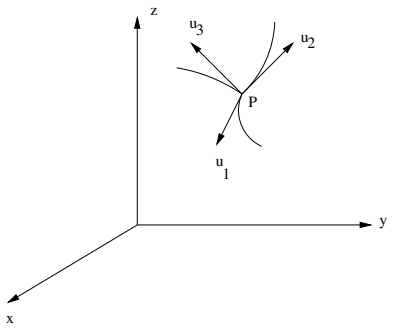



Figure 8.1: Orthogonal curvilinear coordinates in three dimensions.

## 2.8 Curvilinear coordinate systems

We now consider vector calculus in alternate coordinates systems which use some combination angles and distances.

Specifically we consider orthogonal curvilinear coordinate systems (fig. 8.1). Such coordinate systems can be particularly useful when the functions we are considering have symmetries, such as cylindrical or spherical symmetry.

In plain English these are orthogonal curvilinear coordinate systems systems that have perpendicular directions $\boldsymbol{e}_i$, but which rotate in some position dependent way, which we choose to track some symmetry of the physics.

The directions at each point are selected by the infinitessimal change in $\boldsymbol{x}$ generated by an infinitessimal change in each of the *curvilinear coordinates*.

We must be able to translate the differential equations of physics appropriately. If we label the curvilinear coordinates $(\xi_1, \xi_2, \xi_3)$ the local axes are given by

$$\frac{\partial \boldsymbol{x}}{\partial \xi_i} = h_i \boldsymbol{e}_i$$

Here, $h_i = \left| \frac{\partial \boldsymbol{x}}{\partial \xi_i} \right|$ is a scale factor with that ensures $\boldsymbol{e_i}$ is a unit vector.

It is useful to consider a *locally* defined cartesian coordinate system

$$\tilde{\boldsymbol{x}}_i = \boldsymbol{x} \cdot \boldsymbol{e}_i$$

### 2.8.1 Circular (or plane) polar coordinates

Plane polar coordinates $(r, \phi)$ are defined by:

$$x = r\cos\phi\ , \qquad y = r\sin\phi\ , \tag{8.1}$$

The radial coordinate $r$ (sometimes written as $r$) can range from 0 to $\infty$. The angular coordinate $\phi$ (sometimes written as $\theta$) ranges from 0 to $2\pi$.

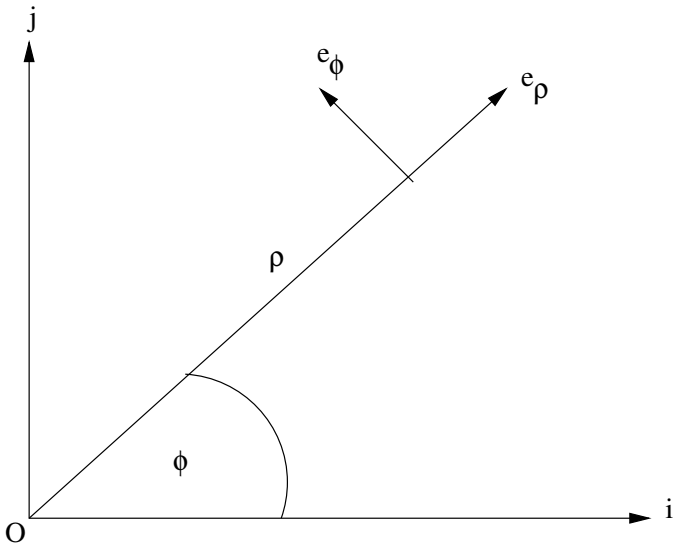



Figure 8.2: Plane polar coordinates.

The *local* orthonormal basis generated by polar coordinates is:

$$\begin{aligned}
\frac{\partial \boldsymbol{x}}{\partial r} &= (\cos\phi, \sin\phi) \\
\frac{\partial \boldsymbol{x}}{\partial \phi} &= r(-\sin\phi, \cos\phi) \\
h_r &= 1 \\
h_\phi &= r \\
e_r &= (\cos\phi, \sin\phi) \\
e_\phi &= (-\sin\phi, \cos\phi)
\end{aligned} \tag{8.2}$$

**Exercise** Show that this system is *orthogonal* by verifying that $(\frac{\partial \boldsymbol{x}}{\partial r}) \cdot (\frac{\partial \boldsymbol{x}}{\partial \phi}) = 0$

### Area integrals

When we change coordinates in an integral, we have to include a scale factor in the integration variable.

As we increase each coordinate by an infinitesimal amount, we sweep out a small area which we call $d\mathcal{A}$. Now, as we chose *orthogonal* curvilinear coordinates we know $\boldsymbol{e}_r$ and $\boldsymbol{e}_\phi$ are perpendicular, and the area is

$$\begin{aligned}
d\mathcal{A} &= d\tilde{x}_r d\tilde{x}_\phi \\
&= h_r dr h_\phi d\phi \\
&= r dr d\phi
\end{aligned} \tag{8.3}$$

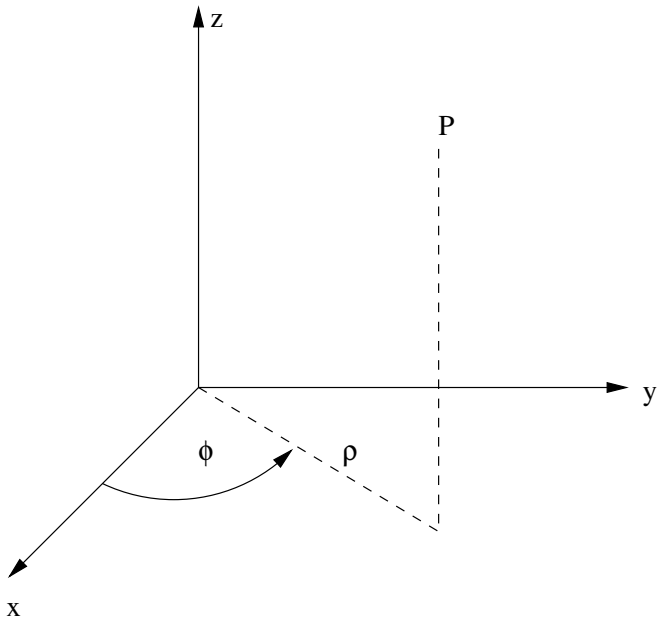



Figure 8.3: Cylindrical polar coordinates.

**Example: area of circle** Consider

$$\int_0^R dr \int_0^{2\pi} r d\phi = 2\pi \int_0^R r dr = \pi R^2$$

### 2.8.2 Cylindrical polar coordinates

An simple extension of plane polar coordinates into the third dimension: see Fig. 8.3.

$$x = r\cos\phi\,, \qquad y = r\sin\phi\,, \qquad z = z \tag{8.4}$$

- radius $r \in [0, \infty]$, angle $\phi \in [0, 2\pi]$, and $z$ coordinate $z \in [-\infty, \infty]$.
- Scale factors: $h_r = 1$, $h_\phi = r$, $h_z = 1$.

**Example: volume of a cylinder** Consider

$$\int_0^L dz \int_0^R dr \int_0^{2\pi} r d\phi = 2\pi L \int_0^R r dr = \pi R^2 L$$

### 2.8.3 Spherical polar coordinates

Useful when there is spherical symmetry: see Figure 8.4.

$$x = r\sin\theta\cos\phi, \qquad y = r\sin\theta\sin\phi, \qquad z = r\cos\theta \tag{8.5}$$

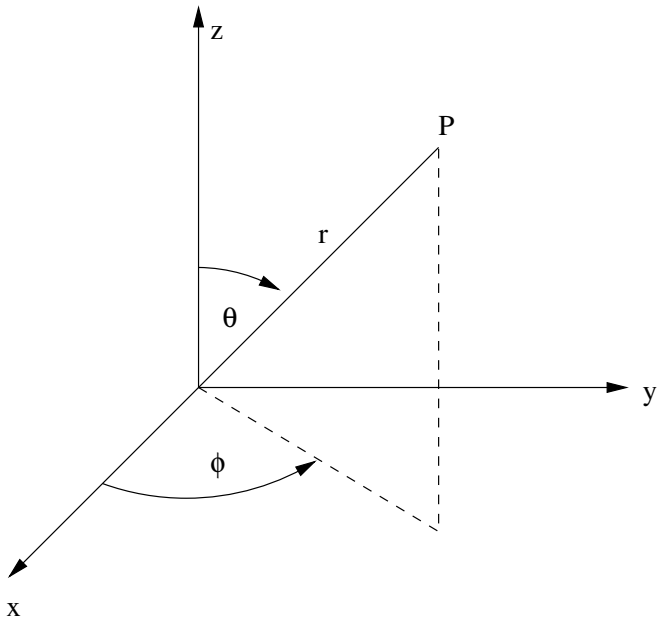



Figure 8.4: Spherical polar coordinates.

- radius $r \in [0, \infty]$, angle $\theta \in [0, \pi]$, angle $\phi \in [0, 2\pi]$.

$$\begin{aligned}
\frac{\partial \boldsymbol{x}}{\partial r} &= (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta) \\
\frac{\partial \boldsymbol{x}}{\partial \theta} &= r(\cos\theta\cos\phi, \cos\theta\sin\phi, -\sin\theta) \\
\frac{\partial \boldsymbol{x}}{\partial \phi} &= r(-\sin\theta\sin\phi, \sin\theta\cos\phi, 0) \\
h_r &= 1 \\
h_\theta &= r \\
h_\phi &= r\sin\theta
\end{aligned} \tag{8.6}$$

**Example : volume of sphere**

$$\begin{aligned}
\int_0^R dr \int_0^\pi r d\theta \int_0^{2\pi} r\sin\theta d\phi &= 2\pi \int_0^R r^2 dr \int_0^\pi \sin\theta d\theta \\
&= 2\pi \left[\frac{r^3}{3}\right]_0^R [-\cos\theta]_0^\pi \\
&= \frac{4}{3}\pi R^3
\end{aligned} \tag{8.7}$$

**Example : area of sphere**

$$\begin{aligned}
\int_0^\pi R d\theta \int_0^{2\pi} R\sin\theta d\phi &= 2\pi R^2 \int_0^\pi \sin\theta d\theta \\
&= 2\pi R^2 [-\cos\theta]_0^\pi \\
&= 4\pi R^2
\end{aligned} \tag{8.8}$$

### 2.8.4 Gradient

In the *local* coordinate system the gradient operator is

$$\sum_i \frac{\partial}{\partial \tilde{x}_i}\boldsymbol{e}_i = \sum_i \frac{1}{h_i}\frac{\partial}{\partial \tilde{\xi}_i}\boldsymbol{e}_i$$

**Circular polars**

$$\begin{aligned}\boldsymbol{\nabla} f &= \frac{\partial f}{\partial \tilde{x}_r}\boldsymbol{e}_r + \frac{\partial f}{\partial \tilde{x}_\phi}\boldsymbol{e}_\phi \\ &= \frac{\partial f}{h_r \partial r}\boldsymbol{e}_r + \frac{\partial f}{h_\phi \partial \phi}\boldsymbol{e}_\phi \\ &= \frac{\partial f}{\partial r}\boldsymbol{e}_r + \frac{1}{r}\frac{\partial f}{\partial \phi}\boldsymbol{e}_\phi \end{aligned} \tag{8.9}$$

**Cylindrical polars**

$$\boldsymbol{\nabla} = \frac{\partial f}{\partial r}\boldsymbol{e}_r + \frac{1}{r}\frac{\partial f}{\partial \phi}\boldsymbol{e}_\phi + \frac{\partial f}{\partial z}\hat{z}$$

**Spherical polars**

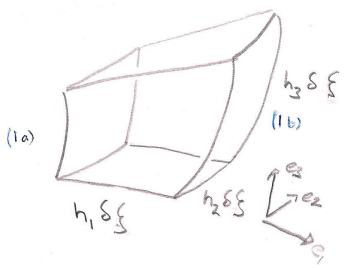
$$\boldsymbol{\nabla} = \frac{\partial f}{\partial r}\boldsymbol{e}_r + \frac{1}{r}\frac{\partial f}{\partial \theta}\boldsymbol{e}_\theta + \frac{1}{r\sin\theta}\frac{\partial f}{\partial \phi}\boldsymbol{e}_\phi$$

### 2.8.5 Divergence

For the divergence we must be a little bit more careful. The coordinate dependence of the scale factors themselves must be taken into account.

$$\boldsymbol{\nabla}\cdot\boldsymbol{v} = \lim_{V\to 0}\frac{\int_{\mathcal{A}} \boldsymbol{v}\cdot\hat{\boldsymbol{n}}d\mathcal{A}}{V}$$

Consider the infinitessimal cube




Flux through (1a) is $F_1 = v_1\, h_2\, d\xi_2\, h_3\, d\xi_3$.

Net flux difference between (1a) and (1b) is $\frac{\partial F_1}{\partial \xi_1}\delta\xi_1$.

Thus, summing over all pairs of faces

$$\begin{aligned}\boldsymbol{\nabla}\cdot\boldsymbol{v} &= \frac{1}{h_1h_2h_3\delta\xi_1\delta\xi_2\delta\xi_3}\left[\frac{\partial F_1}{\partial\xi_1}\delta\xi_1+\frac{\partial F_2}{\partial\xi_2}\delta\xi_2+\frac{\partial F_3}{\partial\xi_3}\delta\xi_3\right]\\ &= \frac{1}{h_1h_2h_3}\left[\frac{\partial h_2h_3v_1}{\partial\xi_1}+\frac{\partial h_3h_1v_2}{\partial\xi_2}+\frac{\partial h_1h_2v_3}{\partial\xi_3}\right]\end{aligned} \tag{8.10}$$

**Circular polars**

$$\boldsymbol{\nabla}\cdot\boldsymbol{v}=\frac{1}{r}\frac{\partial}{\partial r}(rv_r)+\frac{1}{r}\frac{\partial}{\partial\phi}v_\phi$$

**Cylindrical polars**

$$\boldsymbol{\nabla}\cdot\boldsymbol{v}=\frac{1}{r}\frac{\partial}{\partial r}(rv_r)+\frac{1}{r}\frac{\partial}{\partial\phi}v_\phi+\frac{\partial}{\partial z}v_z$$

**Spherical polars**

$$\boldsymbol{\nabla}\cdot\boldsymbol{v}=\frac{1}{r^2}\frac{\partial}{\partial r}(r^2v_r)+\frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta v_\theta)+\frac{1}{r\sin\theta}\frac{\partial}{\partial\phi}v_\phi$$

### 2.8.6 Laplacian

We can now form the Laplacian as simply the divergence of the gradient, combining the results of the previous two subsections:

$$\boldsymbol{\nabla}^2f=\boldsymbol{\nabla}\cdot\boldsymbol{\nabla}f$$

leading to:

**Circular polars**

$$\nabla^2f=\frac{1}{r}\frac{\partial}{\partial r}(r\frac{\partial f}{\partial r})+\frac{1}{r^2}\frac{\partial^2f}{\partial\phi^2}$$

**Cylindrical polars**

$$\nabla^2f=\frac{1}{r}\frac{\partial}{\partial r}(r\frac{\partial f}{\partial r})+\frac{1}{r^2}\frac{\partial^2f}{\partial\phi^2}+\frac{\partial^2f}{\partial z^2}$$

**Spherical polars**

$$\nabla^2f=\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial f}{\partial r}\right)+\frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\,\frac{\partial f}{\partial\theta}\right)+\frac{1}{r^2\sin^2\theta}\frac{\partial^2f}{\partial\phi^2}. \tag{8.11}$$

Note, that there are no mixed second order derivatives and that these equations are seperable.

### 2.8.7 Curl

## 2.9 Wave equation in circular polars

Equivalently, solving the wave equation for a circular drum

$$\nabla^2 u = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} \ .$$

In this section we shall use $r$ for the radial coordinate. As before, we consider solutions of separated form: $u(r,\phi,z,t) = R(r)\Phi(\phi)T(t)$. Substitute into wave equation and divide across by $u = R\Phi T$.

$$\frac{1}{R}\frac{\partial^2 R}{\partial r^2} + \frac{1}{rR}\frac{\partial R}{\partial r} + \frac{1}{r^2\Phi}\frac{\partial^2 \Phi}{\partial \phi^2} = \frac{1}{c^2 T}\frac{\partial^2 T}{\partial t^2} \ .$$

*First separation*: time equation: $\mathrm{LHS}(r,\phi,z) = \mathrm{RHS}(t) = \text{constant}$

$$\frac{1}{c^2}\frac{1}{T}\frac{d^2 T}{dt^2} = -k^2.$$

The solutions to this are of the form $T(t) = G_k \cos\omega_k t + H_k \sin\omega_k t$ with $\omega_k \equiv ck$.

*Second separation*: Multiply through by $r^2$ and separate again:

$$\mathrm{LHS}(r) = \mathrm{RHS}(\phi) = \text{a constant}.$$

For the angular dependence:

$$\frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2} = -n^2;$$

The solution is $\Phi = C\cos n\phi + D\sin n\phi$.

We want the solution to the wave equation to be *single valued*, so $\Phi(\phi + 2\pi) = \Phi(\phi)$, forcing $n$ to be integer-valued: $n = 0, \pm 1, \pm 2 \ldots$

The equation describing the radial dependence is the only difficult one to solve:

$$\frac{d^2 R}{dr^2} + \frac{1}{r}\frac{dR}{dr} - \frac{n^2}{r^2}R + k^2 R = 0 \ .$$

Multiply across by $r^2$ and rewrite

$$r^2 R'' + rR' + (k^2 r^2 - n^2)R = 0 \ . \tag{9.12}$$

This is known as *Bessel's equation of order n.* The solutions are known as Bessel functions. Being a quadratic ODE, there are two independent solutions called $J_n(kr)$ and $Y_n(kr)$. Note we have labelled the solutions with integer $n$.

#### Method of Froebenius

We can solve Bessel's equation by substituting a general *Laurent series* as a trial solution. A Laurent series is a generalisation of a Taylor series to possibly include negative power terms (called poles).

We try a solution

$$R(r) = \sum_{i=0}^{\infty} C_i r^{i+m}$$

where $c_i$ and $m$ are unknowns. $m$ represents the lowest power of $r$ that occurs in the solution, and where it arises $c_i = 0$ for $i < 0$ because otherwise $m$ would *not* represent the lowest power of $r$.

Differentiating we get

$$R'(r) = \sum_{i=0}^{\infty} (i+m) C_i r^{i+m-1}$$

$$rR'(r) = \sum_{i=0}^{\infty} (i+m) C_i r^{i+m}$$

$$r^2 R''(r) = \sum_{i=0}^{\infty} (i+m)(i+m-1) C_i r^{i+m}$$

Bessel's equation becomes a relation between coefficients:

$$\sum_{i=0}^{\infty} (i+m) C_i r^{i+m} + (i+m)(i+m-1) C_i r^{i+m} - n^2 C_i r^{i+m} + k^2 C_{i-2} r^{i+m}$$

Since this must be true for all $r$, then we have the *indicial equation*

$$\left[(i+m) + (i+m)(i+m-1) - n^2\right] C_i + k^2 C_{i-2} = 0$$

The series *switches on* when $C_{-2} = 0$ and $C_0 \neq 0$. Then,

$$m^2 = n^2.$$

We are interested in the case where $m \geq 0$ so that the solutions are finite at $r = 0$.

Next, we are interested in forming a recurrence relation between coefficients. The above indicial equation suggests

$$C_i = \frac{k^2}{n^2 - (i+m)^2} C_{i-2}$$

If we consider the Bessel function $J_0(r)$, we take $n = m = 0$ and have (up to normalisation)

$$\begin{aligned} C_0 &= 1 \\ C_2 &= -\frac{k^2}{4} \\ C_4 &= +\frac{k^4}{4.16} \\ C_6 &= -\frac{k^6}{4.16.36} \\ &\ldots \end{aligned} \tag{9.13}$$

The series is

$$1 - \frac{(kr)^2}{4} + \frac{(kr)^4}{4.16} - \frac{(kr)^6}{4.16.36} \ldots$$

and is purely a function of $(kr)$. As the sign oscillates we have many turning points.

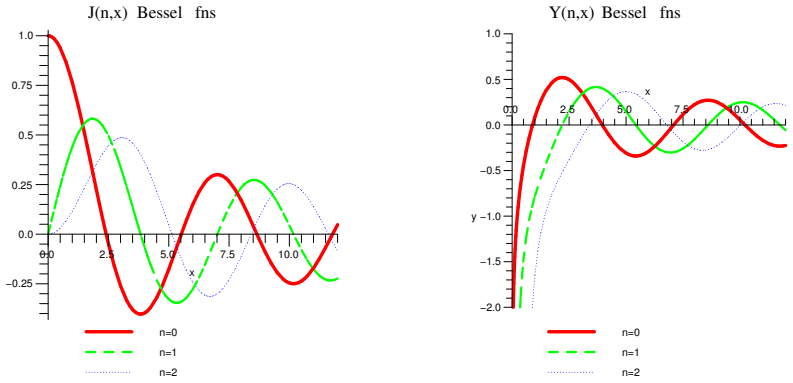



Figure 9.5: The first three Bessel functions of integral order.

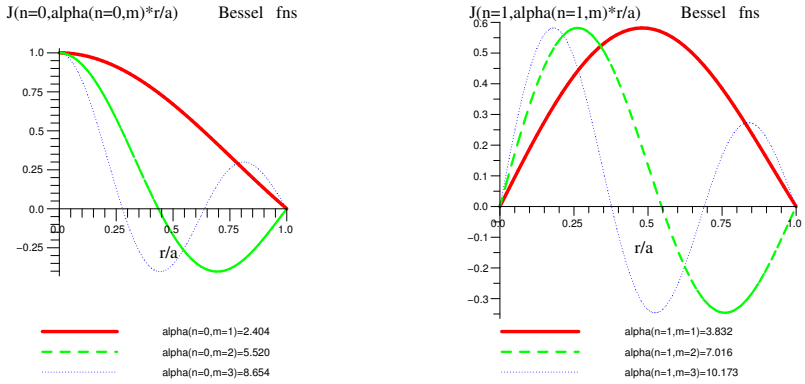



Figure 9.6: Rescaling the $J_0$ and $J_1$ Bessel functions so that one of the nodes lies at $r = a$

### Roots of Bessel functions

The first few $J_n$ and $Y_n$ functions are plotted in Fig. 9.5. The $Y_n$ functions diverge at the origin and so are not suitable for describing oscillations of a drumskin.

The Bessel functions $J_n(x)$ have a series of zeros ("nodes" or "roots") which we label $\alpha_{n1}$, $\alpha_{n2}$, $\alpha_{n3}$.

For the function $\sin nx$, the nodes occur at $x = \alpha_{nm} = m\pi$ are equally spaced. For Bessel functions, however, they are not. The nodes must be found numerically, in practice either looked up in tables or calculated using packages such as Maple.

### Spatial BCs and normal modes

Our solution can be written as

$$u(r, \phi, t) = J_n(kr)\ (C \cos n\phi + D \sin n\phi)\ (G \cos \omega_k t + H \sin \omega_k t)$$

with $\omega_k = ck$ and currently no restriction on $k$.

We now apply spatial boundary conditions. Recall periodicity in $\phi$ quantised $n$. In the radial direction we require that the drumskin does not move at the rim:

$$u(r = a, \phi, t) = 0 \text{ for all } \phi \text{ and } t.$$

We therefore want the edge of the drum to coincide with one of the nodes of the Bessel function. The $m^{th}$ node of the Bessel function of order $n$ occurs when the argument of the Bessel function takes value $\alpha_{nm}$, and we rescale the Bessel function so that one of these zeros coincides with $r = a$.

It doesn't matter which node we choose to lie at $r = a$, so we have different normal mode solutions depending on which $m$ we choose. The allowed values of $k$ are therefore

$$k_{nm}a = \alpha_{nm}.$$

Quantising $k$ also quantises $\omega_k$. This is like we fond for the harmonics, but the normal mode frequencies are here not equally spaced (because the $\alpha_{nm}$ are not evenly spaced). This proves why the drum is not as harmonious as the guitar.

Some examples of rescaling for $n = 0$ and $n = 1$ are shown in Fig. 9.6.

Our normal mode solutions are therefore

$$u(r, \phi, t) = J_n\left(\alpha_{nm}\frac{r}{a}\right)\ (C_{nm}\cos n\phi + D_{nm}\sin n\phi)\ (G_{nm}\cos\omega_{nm}t + H_{nm}\sin\omega_{nm}t)\ .$$

Each normal mode is labelled by $n$ and $m$ and will have different constants so we label them appropriately. $k$ depends on $n$ and $m$ via $\alpha_{nm}$, so we also change the label on $\omega$.

### Zeros and nodal lines

Only $J_0$ is zero at the origin (e.g. Fig. 9.5) so $u = 0$ at $r = 0$ for all $t$ if $n > 0$.

Nodal lines are other points on the drumskin that remain stationary for this normal mode:

- We find $(m-1)$ nodal lines in $r$: they occur at $\alpha_{nm}r/a = \alpha_{nm'} \implies r = a\alpha_{nm'}/\alpha_{nm}$ for $m' = 1, 2 \ldots (m-1)$. (See Fig. 9.6.)
- $2n$ nodal lines in $\phi$: occur at intervals $\delta\phi = \pi/n$ for $n \neq 0$. N.B. do not need to start at $\phi = 0$.

Some low-lying modes are shown in Fig. 9.7. Note that the wave equation had rotational symmetry. This does not mean that the solutions have to have rotational symmetry ($n = 1, 2 \ldots$ do not). It means that if we take normal mode solution and then rotate it, it is still a solution of the wave equation.

### The general solution

The general solution is a linear superposition of all allowed modes:

$$u(r, \phi, t) = \sum_{n=0}^{\infty}\sum_{m=1}^{\infty} J_n\left(\alpha_{nm}\frac{r}{a}\right)\ (C_{nm}\cos n\phi + D_{nm}\sin n\phi)\ (G_{nm}\cos\omega_{nm}t + H_{nm}\sin\omega_{nm}t) \tag{9.14}$$

Each $(n, m)$ term contains *two* normal modes ($\cos n\phi$ and $\sin n\phi$), and there are four unknown constants. Two unknowns per mode is what we expect for a second order differential equation.

We will use initial conditions to fix the unknowns, but before that we need to learn a bit more about the properties of Bessel functions. In particular we need an orthogonality relation.

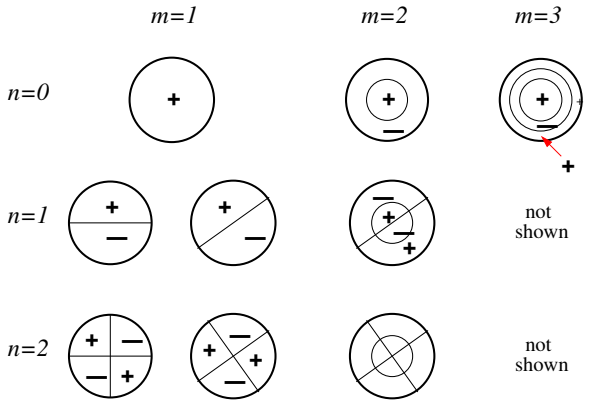



Figure 9.7: Some normal modes for a round drum.

### Orthogonality and completeness

We state one orthogonality relation without proof. For given, fixed $n$

$$\int_0^a dr\ J_n\left(\alpha_{nm}\frac{r}{a}\right) J_n\left(\alpha_{nl}\frac{r}{a}\right) r = \frac{a^2}{2}[J_{n+1}(\alpha_{nm})]^2\delta_{m,l}\ , \tag{9.15}$$

where $J_n(\alpha_{nm}) = 0$, i.e. $\alpha_{nm}$ is the $m^{\text{th}}$ root of the Bessel function of integral order $n$.

The extra factor of $r$ compared with Fourier orthogonality arises *mathematically* because Bessel's equation contained first order derivatives in $r$.

The extra factor of $r$ compared with Fourier orthogonality arises *physically* because Bessel's equation arose in the radial direction of two dimensional wave equation. It is a *vestigial circumference* factor $2\pi r$, turning a line integral into an area integral that is appropriate for a 2D orthogonality relation.

The set of Bessel functions $\{J_n(\alpha_{nm}x)\ ;\ m = 1\ldots\infty\}$ for fixed $n$ form a complete set, so any function can be expanded in the interval $0 \le r \le a$ as a *Bessel* (or Fourier-Bessel) *series*:

$$f(r) = \sum_{m=1}^{\infty} A_{nm} J_n\left(\alpha_{nm}\frac{r}{a}\right), \tag{9.16}$$

The coefficients are determined using the orthogonality condition in the usual way, as we shall now see.

### Initial conditions for the drumskin

The general solution for the displacement of a circular drumskin of radius $a$ was given in Eqn. (9.14). Combining constants we have:

$$f(r,\phi,t) = \sum_{n=0}^{\infty}\sum_{m=1}^{\infty} J_n\left(\frac{\alpha_{nm}r}{a}\right)(A_{nm}\cos n\phi + B_{nm}\sin n\phi)\cos(\omega_{nm}t + \varepsilon_{nm})$$

Typical initial conditions are that the drumskin is initially at rest (implying $\varepsilon_{nm} = 0$) and described by given function $p(r, \phi)$:

$$f(r, \phi, t = 0) \equiv \sum_{n=0}^{\infty} \sum_{m=1}^{\infty} A_{nm} J_n \left( \frac{\alpha_{nm} r}{a} \right) \cos n\phi + B_{nm} J_n \left( \frac{\alpha_{nm} r}{a} \right) \sin n\phi = p(r, \phi) \ . \quad (9.17)$$

Given initial conditions $p(r, \phi)$ we can find the coefficients $A_{nm}$ and $B_{nm}$ via our orthogonality relations. We rewrite the initial condition equation as

$$\sum_{n=0}^{\infty} \sum_{m=1}^{\infty} (A_{nm} \Psi_{nm}(r, \phi) + B_{nm} \Phi_{nm}(r, \phi)) = p(r, \phi) \ . \quad (9.18)$$

where we have basis functions

$$\Psi_{nm}(r, \phi) = J_n \left( \frac{\alpha_{nm} r}{a} \right) \cos n\phi \ , \qquad \Phi_{nm}(r, \phi) = J_n \left( \frac{\alpha_{nm} r}{a} \right) \sin n\phi \ . \quad (9.19)$$

Define the inner product of two of these basis functions as

$$S \cdot T \equiv \int dA \ S \ T = \int dr \ \int d\phi \ r \ S(r, \phi) \ T(r, \phi) \ ,$$

and we find that $\{\Psi_{nm}, \Phi_{nm}\}$ form an orthogonal set:

$$\begin{aligned} \Psi_{uv} \cdot \Psi_{nm} &= \frac{a^2 \pi}{2} (1 + \delta_{u0}) [J_u(\alpha_{uv})]^2 \ \delta_{un} \ \delta_{vm} \\ \Psi_{uv} \cdot \Phi_{nm} &= 0 \\ \Phi_{uv} \cdot \Phi_{nm} &= \frac{a^2 \pi}{2} [J_u(\alpha_{uv})]^2 \ \delta_{un} \ \delta_{vm} \end{aligned}$$

We might term such functions the (unnormalised) *Circular Harmonics*. They are orthogonal, but not orthonormal. Note that the extra $r$ is just what we get when we transform an area integral from Cartesian to circular polar coordinates. Note also that the angular orthogonality ensures we only compare Bessel functions of the same order (which is all Eqn. (9.15) covered).

We can use this orthogonality to obtain the coefficients from Eqn. (9.17). Applying $(\Psi_{uv} \cdot)$ or $(\Phi_{uv} \cdot)$ to both sides we get

$$\begin{aligned} A_{uv} &= \frac{2}{a^2 \pi (1 + \delta_{u0}) [J_u(\alpha_{uv})]^2} \times \int_0^a dr \ \int_0^{2\pi} d\phi \ r \ \Psi_{uv}(r, \phi) \ p(r, \phi) \\ B_{uv} &= \frac{2}{a^2 \pi [J_u(\alpha_{uv})]^2} \times \int_0^a dr \ \int_0^{2\pi} d\phi \ r \ \Phi_{uv}(r, \phi) \ p(r, \phi) \end{aligned}$$

We probably have to do these integrals numerically.

## 2.10 Wave equation in spherical polar coordinates

We now look at solving problems involving the Laplacian in spherical polar coordinates. The angular dependence of the solutions will be described by *spherical harmonics*.

We take the wave equation as a special case:

$$\nabla^2 u = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2}$$

The Laplacian given by Eqn. (8.11) can be rewritten as:

$$\nabla^2 u = \underbrace{\frac{\partial^2 u}{\partial r^2} + \frac{2}{r}\frac{\partial u}{\partial r}}_{\text{radial part}} + \underbrace{\frac{1}{r^2 \sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial u}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 u}{\partial\phi^2}}_{\text{angular part}} . \tag{10.1}$$

### 2.10.1 Separating the variables

We consider solutions of separated form

$$u(r,\theta,\phi,t) = R(r)\,\Theta(\theta)\,\Phi(\phi)\,T(t) .$$

Substitute this into the wave equation and divide across by $u = R\Theta\Phi T$:

$$\frac{1}{R}\frac{d^2R}{dr^2} + \frac{2}{rR}\frac{dR}{dr} + \frac{1}{r^2}\frac{1}{\Theta\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2} = \frac{1}{c^2}\frac{1}{T}\frac{d^2T}{dt^2}.$$

**First separation: $r$, $\theta$, $\phi$ versus $t$**

$$\text{LHS}(r,\theta,\phi) = \text{RHS}(t) = \text{constant} = -k^2.$$

This gives the $T$ equation:

$$\frac{1}{c^2}\frac{1}{T}\frac{d^2T}{dt^2} = -k^2 \tag{10.2}$$

which is easy to solve.

**Second separation: $\theta$, $\phi$ versus $r$**

Multiply LHS equation by $r^2$ and rearrange:

$$-\frac{1}{\Theta\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right) - \frac{1}{\sin^2\theta}\frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2} = \frac{r^2}{R}\frac{d^2R}{dr^2} + \frac{2r}{R}\frac{dR}{dr} + k^2r^2 . \tag{10.3}$$

$$\text{LHS}(\theta,\phi) = \text{RHS}(r) = \text{constant} = \lambda$$

We choose the separation constant to be $\lambda$. For later convenience, it will turn out that $\lambda = l(l+1)$ where $l$ has to be integer.

Multiplying the RHS equation by $R/r^2$ gives the $R$ equation:

$$\frac{d^2R}{dr^2} + \frac{2}{r}\frac{dR}{dr} + \left[k^2 - \frac{\lambda}{r^2}\right]R = 0. \tag{10.4}$$

This can be turned into Bessel's equation; we'll do this later.

**Third separation: $\theta$ versus $\phi$**

Multiply LHS of Eqn. (10.3) by $\sin^2\theta$ and rearrange:

$$\frac{\sin\theta}{\Theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)+\lambda\sin^2\theta=-\frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2}=m^2$$

$$\text{LHS}(\theta)=\text{RHS}(\phi)=\text{constant}=-m^2\ .$$

The RHS equation gives the $\Phi$ equation without rearrangement:

$$\frac{d^2\Phi}{d\phi^2}=-m^2\Phi\ . \tag{10.5}$$

Multiply the LHS by $\Theta/\sin^2\theta$ to get the $\Theta$ equation:

$$\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)+\left[\lambda-\frac{m^2}{\sin^2\theta}\right]\Theta=0\ . \tag{10.6}$$

## 2.10.2 Solving the separated equations

Now we need to solve the ODEs that we got from the original PDE by separating variables.

**Solving the $T$ equation**

Eqn. (10.2) is of simple harmonic form and solved as before, giving sinusoids as solutions:

$$\frac{d^2T}{dt^2}=-c^2k^2T\equiv-\omega_k^2T\ ,$$

with $\omega_k=ck$.

**Solving the $\Phi$ equation**

Eqn. (10.5) is easily solved. Rather than using cos and sin, it is more convenient to use complex exponentials:

$$\Phi(\phi)=e^{\pm im\phi}$$

Note that we have to include both positive and negative values of $m$.

As $\phi$ is an angular coordinate, we expect our solutions to be *single-valued*, i.e. unchanged as we go right round the circle $\phi\to\phi+2\pi$:

$$\Phi(\phi+2\pi)=\Phi(\phi)\qquad\Rightarrow\qquad e^{i2\pi m}=1\qquad\Rightarrow\qquad m=\text{integer.}$$

This is another example of a BC (periodic in this case) quantising a separation constant.

In principle $m$ can take any integer value between $-\infty$ and $\infty$.

It turns out in Quantum Mechanics that

$$\boxed{m\text{ is the integer magnetic quantum number and }-l\le m\le l}$$

for the $z$-component of angular momentum. In that context we will see that it is restricted to the range $-l\le m\le l$.

**Solving the $\Theta$ equation**

Starting from Eqn. (10.6), make a change of variables $w = \cos\theta$:

$$\frac{d}{dw} = \frac{d\theta}{dw}\,\frac{d}{d\theta} = \left(\frac{dw}{d\theta}\right)^{-1}\frac{d}{d\theta} = -\frac{1}{\sin\theta}\,\frac{d}{d\theta}\,,$$
$$(1-w^2)\frac{d}{dw} = -\frac{1-\cos^2\theta}{\sin\theta}\,\frac{d}{d\theta} = -\frac{\sin^2\theta}{\sin\theta}\,\frac{d}{d\theta} = -\sin\theta\,\frac{d}{d\theta}\,,$$
$$\frac{d}{dw}(1-w^2)\frac{d}{dw} = -\frac{1}{\sin\theta}\,\frac{d}{d\theta}\left[-\sin\theta\,\frac{d}{d\theta}\right] = \frac{1}{\sin\theta}\,\frac{d}{d\theta}\left[\sin\theta\,\frac{d}{d\theta}\right]\,.$$

Eqn. (10.6) becomes

$$\left(\frac{d}{dw}(1-w^2)\frac{d}{dw} + \lambda - \frac{m^2}{1-w^2}\right)\Theta(w) = 0\,.$$

which is known as the *Associated Legendre Equation*. Solutions of the Associated Legendre Equation are the Associated Legendre Polynomials. Note that the equation depends on $m^2$ and the equation and solutions are the same for $+m$ and $-m$.

It will turn out that there are smart ways to generate solutions for $m \neq 0$ from the solutions for $m = 0$ using angular momentum ladder operators (see quantum mechanics of hydrogen atom). So it would be unnecessarily "heroic" to directly solve this equation for $m \neq 0$.

| In this course we will only solve this equation for $m = 0$. |
|---|

**Solving the Legendre equation**

For $m = 0$ we can write the special case as the *Legendre Equation*:

$$\left((1-w^2)\frac{d^2}{dw^2} - 2w\frac{d}{dw} + \lambda\right)\Theta(w) = 0\,.$$

We apply the method of Froebenius by taking

$$\Theta(w) = \sum_{i=0}^{\infty} c_i w^i$$

Then

$$\sum_{i=0}^{\infty} c_i i(i-1)(w^{i-2} - w^i) - 2c_i i w^i + \lambda c_i w^i = 0$$

and rearranging the series to always refer the power $w^i$,

$$\sum_{i=0}^{\infty} \left[c_{i+2}(i+2)(i+1) + c_i(\lambda - i(i-1) - 2i\right] w^i = 0$$

Since this is true for all $w$, it is true term by term, and the indicial equation is

| $c_{i+2}\left((i+2)(i+1)\right) = c_i\left(i(i+1) - \lambda\right)$ |
|---|

Start The series "switches on" when $c_0 \times 0 = 0$ admits $c_0 \neq 0$ and $c_{-2} = 0$
Also when $c_1 \times 0 = 0$ admits $c_1 \neq 0$ and $c_{-1} = 0$.

Termination Note, however $c_{i+2} \simeq c_i$ for large $i$. This gives an ill convergent series and for finite solutions the series must terminate at some value of i, which we call $l$. Thus,

$$\lambda = l(l+1)$$

for some (quantised) integer value $l$.

It will turn out in quantum mechanics that $l$ is the orbital angular momentum quantum number.

**Legendre polynomials**

We denote the solutions the Legendre polynomials

$$P_l(w) \equiv P_l(\cos\theta)$$

**For example:** $P_0$ starts, and terminates with a single term $C_0$.
$P_1$ starts, and terminates with a single term $C_1$.
$P_2$ starts, with $C_0$ and terminates on $C_2$.
etc...

The first few are

$$\begin{aligned} P_0(w) &= 1 \\ P_1(w) &= w \\ P_2(w) &= \frac{1}{2}(3w^2 - 1) \\ P_3(w) &= \frac{1}{2}(5w^3 - 3w) \end{aligned}$$

**Exercise:** use the recurrence relation

$$c_{i+2}\left((i+2)(i+1)\right) = c_i\left(i(i+1) - l(l+1)\right)$$

to verify that these are our series solutions of Legendre's equation.

**Orgthogonality**

The orthogonality relation is

$$\int_{-1}^{1} P_m(w)P_n(w)dw = \int_0^{\pi} P_m(\cos\theta)P_n(\cos\theta)\sin\theta d\theta = N_m\delta_{mn}$$

where $N_m$ is a normalisation factor that we do not need here.

In quantum mechanics this is already sufficient to cover S, P , D and F orbitals.

**Associated Legendre polynomials**

As mentioned the associated Legendre polynomials can be produced from Legendre polynomials in quantum mechanics using angular momentum ladder operators. Firstly,

$$P_l^0(w) = P_l(w)$$

Without proof, we can note that it can be shown that if $P_l(w)$ satisfies Legendre's equation, then

$$P_l^{|m|}(w) = (1-w^2)^{|m|/2}\frac{d^{|m|}}{dw^{|m|}}P_l(w)$$

will satisfy the associated Legendre polynomial for magnetic quantum number $m$.

As $P_l$ is a polynomial of order $l$, then the above $m$-th derivative vanishes for $|m| > l$ and thus $m = -l, -l+1, \ldots, 0, \ldots l-1, l$.

**General angular solution**

Putting aside the radial part for the moment, the rest of the general solution is:

$$\Theta(\theta)\Phi(\phi)T(t) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} P_l^m(\cos\theta)\, e^{im\phi}\ (E_{ml}\cos\omega_k t + F_{ml}\sin\omega_k t)$$

The angular dependence is given by the combination:

$$P_l^m(\cos\theta)\, e^{im\phi} \propto Y_m^l(\theta,\phi)$$

These are known as the *spherical harmonics* (once we include a normalisation constant). We'll discuss these more in Sec. 2.10.3. What we have not yet established is the link between the value of $k$ (and hence $\omega_k$) and the values of $m$ and $l$. To do this, we would need to solve the radial equation for various special cases.

### 2.10.3 The spherical harmonics

Spherical harmonics $\{Y_l^m(\theta,\phi)\}$ provide a complete, orthonormal basis for expanding the angular dependence of a function. They crop up a lot in physics because they are the normal mode solutions to the angular part of the Laplacian. They are defined as:

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

The extra factor of $(-1)^m$ introduced is just a convention and does not affect the orthonormality of the functions.

The spherical harmonics satisfy an orthogonality relation:

$$\int_0^{2\pi} d\phi \int_0^{\pi} d\theta\ \sin\theta\ \left[Y_{l_1}^{m_1}(\theta,\phi)\right]^* Y_{l_2}^{m_2}(\theta,\phi) = \delta_{l_1,l_2}\delta_{m_1,m_2}\ .$$

Note that they are orthonormal, not just orthogonal, as the constant multiplying the product of Kronecker deltas is unity.

### Completeness and the Laplace expansion

The completeness property means that any function $f(\theta, \phi)$ evaluated over the surface of the unit sphere can be expanded in the double series known as the *Laplace series*:

$$f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} a_{lm} Y_l^m(\theta, \phi) \ ,$$

$$\Rightarrow \qquad a_{lm} = \int_0^{\pi} d\theta \int_{-\pi}^{\pi} d\phi \ \sin\theta \ [Y_l^m(\theta, \phi)]^* f(\theta, \phi) \ .$$

Note that the sum over $m$ only runs from $-l$ to $l$, because the associated Laplace polynomials $P_l^m$ are zero outside this range.

### 2.10.4 Time independent Schroedinger equation in central potential

Consider

$$-\frac{\hbar^2}{2m}\nabla^2\psi(\boldsymbol{x})+\tilde{V}(r)\psi(\boldsymbol{x})=\tilde{E}\psi(\boldsymbol{x}).$$

We consider solutions of separated form: $\psi(r,\theta,\phi)=R(r)\Theta(\theta)\Phi(\phi)$. Substitute into Schroedinger equation and divide across by $\psi=R\Theta\Phi$.

$$\frac{2m}{\hbar^2}\left(V(r)-E\right)-\frac{1}{R}\frac{1}{r^2}\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r}R=\frac{1}{\Theta}\frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta}\Theta+\frac{1}{\Phi}\frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\Phi$$

Multiplying through by $r^2$

$$r^2\frac{2m}{\hbar^2}\left(V(r)-E\right)-\frac{1}{R}\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r}R=\frac{1}{\Theta}\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta}\Theta+\frac{1}{\Phi}\frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\Phi$$

*First separation*: radial & angular dependence

$$\mathrm{LHS}(r)=\mathrm{RHS}(\theta,\phi)=\text{constant}=-l(l+1).$$

**Radial equation**

$$\left[-\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r}+l(l+1)+r^2\frac{2m}{\hbar^2}(V(r)-E)\right]R=0$$

The differential equation is simplified by a substitution,

$$\begin{aligned}u(r)&=rR(r)\\u'(r)&=R(r)+rR'(r)\\u''(r)&=2R'(r)+rR''(r)=\frac{1}{r}\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r}R\end{aligned}$$

and so

$$\left[-\frac{d^2}{dr^2}+\frac{l(l+1)}{r^2}+\frac{2m}{\hbar^2}(V(r)-E)\right]u(r)=0$$

We take a Coulomb potential and will be considering bound states, with $E<0$. It is convenient to rewrite in terms of the modulus $|E|$ and introduce explicit negative sign. We also change variables to $\rho=\frac{\sqrt{8m|E|}}{\hbar}r$

$$V(r)=\frac{-e^2}{4\pi\epsilon_0 r}=\frac{-e^2\sqrt{8m|E|}}{4\pi\epsilon_0\hbar\rho},$$

and so multiplying by $\frac{1}{r}$ and expressing in terms of $u$

$$\left\{\frac{8m|E|}{\hbar^2}\left[-\frac{d^2}{d\rho^2}+\frac{l(l+1)}{\rho^2}\right]+\frac{2m}{\hbar^2}\left[|E|-\frac{e^2}{4\pi\epsilon_0\rho}\sqrt{\frac{8m|E|}{\hbar^2}}\right]\right\}u(\rho)=0$$

We define $\lambda=\frac{e^2}{4\pi\epsilon_0\hbar}\sqrt{\frac{m}{2|E|}}=\alpha\sqrt{\frac{mc^2}{2|E|}}$, where $\alpha=\frac{e^2}{4\pi\epsilon_0\hbar c}\simeq\frac{1}{137}$ is the fine structure constant. This gives us

$$\left[\frac{d^2}{d\rho^2}-\frac{1}{4}-\frac{l(l+1)}{\rho^2}+\frac{\lambda}{\rho}\right]u(\rho)=0$$

**Solution by method of Froebenius**

We are now (almost!) ready to apply the method of Froebenius. In principle it could immediately be applied and we would get a an infinite Taylor series that indeed solves the equation.

However, a closed form solution can be obtained with one extra transformation that removes an over all exponential dependence on $\rho$. Observe that this equation for large $\rho$ tends to

$$\rightarrow \left[\frac{d^2}{d\rho^2} - \frac{1}{4}\right] u(\rho)$$

which has as a normalisable solution $u(\rho) \rightarrow e^{-\frac{\rho}{2}}$ (and in addition a non-normalisable solution $u(\rho) \rightarrow e^{+\frac{\rho}{2}}$ which we ignore).

We can do rather better by taking a trial solution:

$$u(\rho) = e^{-\frac{\rho}{2}} f(\rho).$$

Then,

$$u' = e^{-\frac{\rho}{2}} \left[f'(\rho) - \frac{1}{2} f(\rho)\right].$$

$$u'' = e^{-\frac{\rho}{2}} \left[f''(\rho) - f'(\rho) + \frac{1}{4} f(\rho)\right].$$

Now,

$$\left[\frac{d^2}{d\rho^2} - \frac{d}{d\rho} + \cancel{\frac{1}{4}} \cancel{- \frac{1}{4}} - \frac{l(l+1)}{\rho^2} + \frac{\lambda}{\rho}\right] f(\rho) = 0$$

We now apply the method of Froebenius for a series substituting $f(\rho) = \sum_{i=0}^{\infty} c_i \rho^i$,

$$\sum_{i=0}^{\infty} c_i(i)(i-1)\rho^{i-2} - c_i(i)\rho^{i-1} - l(l+1)c_i\rho^{i-2} + \lambda c_i \rho^{i-1}$$

Thus, reexpressing so that all terms are of equal power $\rho^{i-1}$

$$\sum_{i=-1}^{\infty} c_{i+1}(i+1)(i)\rho^{i-1} - c_i(i)\rho^{i-1} - l(l+1)c_{i+1}\rho^{i-1} + \lambda c_i \rho^{i-1}$$

and we have the indicial equation,

$$\boxed{c_{i+1}\left[i(i+1) - l(l+1)\right] = c_i\left[i - \lambda\right]}$$

**Series start:** The series "switches on" for $c_k \equiv c_{i+1}$ when $i(i+1) = (k-1)k = l(l+1)$.

The first term $c_k$ has $k = l + 1$.

**Series termination:** If the series does not terminate, then $c_{i+1} \to \frac{c_i}{i}$, and $f \to \sum \frac{\rho^i}{i!}$. This looks like the other solution that is a non-normalisable exponential $u(\rho) \simeq e^{+\frac{\rho}{2}}$ which we do not seek.

Only if $\lambda = i = n$ then the series "switches off" after $n - l$ terms.

$$c_{i+1} = c_i \frac{i - \lambda}{i(i+1) - l(l+1)}$$

We call $n$ the principal quantum number. Note that for any given $l$, then $n \geq l + 1$ as the series commences at $k = l + 1$. The energy is

$$\alpha \sqrt{\frac{mc^2}{2|E|}} = n$$

Thus

$$|E| = \frac{\alpha^2 mc^2}{2n^2}$$

This energy is consistent with the Hydrogen spectrum (Lymann, Balmer series etc...).

### Wavefunctions

We denote the radial solution for $L = l$, and principle quantum number $n \geq l + 1$ as $R_{nl}$.

Using our recurrence relation

$$c_{i+1} = c_i \frac{i - n}{i(i+1) - l(l+1)}$$

we have

| | $n = 1, l = 0$ | $n = 2, l = 0$ | $n = 2, l = 1$ |
|---|---|---|---|
| $c_0$ | 0 | 0 | 0 |
| $c_1$ | 1 | 1 | 0 |
| $c_2$ | 0 | $-\frac{1}{2}$ | 1 |
| $c_3$ | 0 | 0 | 0 |

The above energy relation gives us that for each $n$

$$\rho = \frac{\sqrt{8m|E|}}{\hbar} r \tag{10.7}$$

$$= \frac{2}{n} \frac{\alpha mc}{\hbar} r \tag{10.8}$$

$$= \frac{2}{n} \frac{r}{a_0} \tag{10.9}$$

where $a_0 = \frac{\hbar}{\alpha mc}$ is the usual Bohr radius.

The solutions are then

$$
\begin{aligned}
R_{1S} &\propto \frac{1}{\rho} e^{-\frac{\rho}{2}} \rho && (10.10)\\
&= e^{-\frac{\rho}{2}} && (10.11)\\
&= e^{-\frac{r}{a_0}} && (10.12)\\
R_{2S} &\propto \frac{1}{\rho} e^{-\frac{\rho}{2}} \left( \rho - \frac{1}{2}\rho^2 \right) && (10.13)\\
&= e^{-\frac{\rho}{2}} \left( 1 - \frac{1}{2}\rho \right) && (10.14)\\
&= e^{-\frac{r}{2a_0}} \left( 1 - \frac{1}{2}\frac{r}{a_0} \right) && (10.15)\\
R_{2P} &\propto \frac{1}{\rho} e^{-\frac{\rho}{2}} \rho^2 && (10.16)\\
&= e^{-\frac{\rho}{2}} \rho && (10.17)\\
&= e^{-\frac{r}{2a_0}} \frac{r}{a_0} && (10.18)
\end{aligned}
$$

Note, we have not carefully normalised these radial wavefunctions. For each $l$ the different wavefunctions for the various $n > l$ are orthogonal to each other. The orthogonality relation contains a residual $r^2$ factor corresponding to a vestige of the required orthogonality of wavefunctions under the 3d volume integral.

$$\int_0^\infty R_{nl}(r) R_{ml}(r) r^2 dr = N_n N_m \delta_{nm}$$

A complete orthonormal set can of course be formed as usual but is beyond the scope of the course.

## 2.11 Generalised Fourier Series

Some common themes have run through the course so far. We have solved wave and diffusion equations in a variety of numbers of dimensions and using a variety of coordinate sets. In each case we have found the solution as a superposition of normal modes, and found that we can define an inner product so that these normal modes are orthogonal. In some cases this inner product has needed a weight function (e.g. a factor of $r$ [or $\rho$] for Bessel functions). We have also used without proof the statement that the normal modes form a complete set of functions.

It turns out that these common themes are not coincidental, and Sturm-Liouville theory explains why not. Sturm–Liouville theory was originally developed to describe waves on a string of variable density, but it applies to a far wider class of physical problems. The beauty of it is that it looks at properties of the differential equation and predicts properties of the normal modes (like orthogonality and completeness) *before we even start to solve the equation*. It also tells us exactly how to define our inner product (i.e. what weight function to include) in each case so we get the orthogonality that is essential if we are to project out the expansion coefficients in the normal mode expansion.

### 2.11.1 The Sturm-Liouville problem

Sturm-Liouville (S-L) theory makes predictions for a set of 1-dimensional Ordinary Differential Equations, such as we might obtain after separation of a multidimensional Partial Differential Equation in a particular choice of coordinate set.

Sturm-Liouville theory tells us that if our differential equation can be rearranged to have a certain form, then we are guaranteed that the solutions will have some useful properties.

We require that our differential equation can be written in the form

$$\left[\frac{d}{dx}\left(P(x)\frac{d}{dx}\right) + Q(x)\right]\phi_i(x) = -\lambda_i\,\rho(x)\,\phi_i(x) \tag{11.1}$$

and that we are solving for functions $\phi_i(x)$ in the region $a \le x \le b$. We can think of $i$ as labelling the different normal modes of our system.

The exact forms of the functions $\rho(x)$ (known as the weight function), $P(x)$ and $Q(x)$ depend on the problem that we are studying, and our choice of coordinate set. Our choice of BCs will be dictated by the physics of the problem, which also influences our choice of coordinate set. S-L theory only addresses problems for which the following conditions are met. Fortunately, this covers most common physical systems.

The necessary conditions are:

1. $\rho(x)$, $P(x)$ and $Q(x)$ should be real functions
2. Neither $\rho(x)$ nor $P(x)$ should change sign in the interval $a \le x \le b$
3. The Boundary Conditions at $x = a$ and $x = b$ should lead to

$$\left[P(x)\left(\phi_i^*\frac{d\phi_j}{dx} - \phi_j\frac{d\phi_i^*}{dx}\right)\right]_a^b = 0 \tag{11.2}$$

(i.e. [...] evaluated at $x = b$ minus [...] evaluated at $x = a$) for all combinations of $i$ and $j$.

If these conditions are satisfied, S-L theory predicts (even before we try and solve the differential equation) that the solutions will have the following properties:

1. The eigenvalues $\lambda_i$ are real (and there are an infinite number)
2. The set of eigenfunctions $\{ \phi_i(x) \}$ is orthogonal: $\phi_i \cdot \phi_j \propto \delta_{ij}$
3. The inner product should be defined as $f \cdot g \equiv \int_a^b dx\ \rho(x)\ f(x)^*\ g(x)$
4. The eigenfunctions form a complete set.
   So we can expand a general function $f(x)$ in the interval $a \le x \le b$:

$$f(x) = \sum_i a_i\ \phi_i(x) \qquad \Rightarrow \qquad a_i = \frac{\phi_i \cdot f}{\phi_i \cdot \phi_i}$$

Note that S-L theory tells us nothing about the normalisation of $\phi_i(x)$; this is our choice. It also doesn't tell us that $\phi_i$ are real functions; in many cases they are not.

**Example 1: waves on a string**

To avoid confusion with the square root of $-1$, let's temporarily use $n$ instead of $i$. For waves on a string $0 \le x \le L$, we solved a spatial equation

$$\frac{d^2X}{dx^2} = -k^2 X(x)\ .$$

Comparing to Sturm-Liouville form, $P(x) = 1$, $Q(x) = 0$, $\rho(x) = 1$ with $a = 0$, $b = L$. The solutions were $X(x) = \phi_n(x) = \sin\frac{n\pi x}{L}$ (i.e. normal modes labelled by $n$), with $\lambda_n = k_n^2 = (n\pi/L)^2$. Our boundary conditions were that $\phi_n(x = 0) = \phi_n(x = L) = 0$, which automatically satisfies Eqn. (11.2) because (...) = 0 at both $a$ and $b$, so their difference is zero.

S-L theory predicts orthogonality

$$\int_a^b dx\ \rho(x)\phi_m(x)^*\phi_n(x) = \int_0^L dx\ \sin\frac{m\pi x}{L}\ \sin\frac{n\pi x}{L} \propto \delta_{mn}$$

which we have already shown explicitly in this course.

**Example 2: Bessel's equation**

In circular polar coordinates, our radial equation becomes (after rearrangement)

$$\frac{d}{dr}\left(r\frac{dR}{dr}\right) - \frac{n^2}{r}R = -k^2 rR\ .$$

Instead of $x$ we are using radial coordinate $r$. For a circular drum of radius $r_0$ (to avoid "$a$-confusion"), the BC was $R(r = r_0) = 0$ which led to solutions

$$R(r) = \phi_m(r) = J_n\left(\frac{\alpha_{nm} r}{r_0}\right) .$$

So, comparing all this to S-L form: $P(r) = r$, $Q(r) = -n^2/r$, $\rho(r) = r$ with $a = 0$, $b = r_0$. The solutions are labelled by $m$ instead of $i$, with $\lambda_m = k_m^2 = (\alpha_{nm}/r_0)^2$.

Eqn. (11.2) is satisfied by (...) being 0 at $r = b = r_0$ and by $P(r) = 0$ at $r = a = 0$, so their difference is zero.

S-L theory predicts orthogonality

$$\int_a^b dr\ \rho(r)\phi_m(r)^*\phi_l(r) = \int_0^{r_0} dr\ r\ J_n\left(\frac{\alpha_{nm} r}{r_0}\right)\ J_n\left(\frac{\alpha_{nl} r}{r_0}\right) \propto \delta_{ml}$$

which we have already seen in this course.

ASIDE: At $r = 0$ we have $P(r) = 0$, but to satisfy Eqn. (11.2) we also need $(\ldots)$ finite. This is why we rejected solution $Y_n(kr)$.

### 2.11.2 Boundary condition choices

Our choice of spatial BCs is dictated by the physics of the problem that we are studying, but in most cases they are one of the following set. In each case, these BCs are sufficient to satisfy Eqn. (11.2) and allow us to make use of the powerful predictions from S-L theory.

1. Fixed BCs (a.k.a. Dirichlet BCs):
   $\phi_i(x = a) = \phi_i(x = b) = 0$ for all modes labelled by different $i$.
   e.g. a drumskin fixed at its edge

2. Open BCs (a.k.a. Neumann BCs):
   $\left.\frac{d\phi_i}{dx}\right|_{x=a} = \left.\frac{d\phi_i}{dx}\right|_{x=b} = 0$
   e.g. no ink flows out of the edge of a water tank i.e. no concentration gradient.

3. Mixed BCs:
   $\left.\left(\phi_i + c\frac{d\phi_i}{dx}\right)\right|_{x=a} = \left.\left(\phi_i + c\frac{d\phi_i}{dx}\right)\right|_{x=b} = 0$
   For some real value of $c$. Not so common.

4. Periodic BCs:
   $\phi_i(x = b) = \phi_i(x = a)$, $\left.\frac{d\phi_i}{dx}\right|_{x=b} = \left.\frac{d\phi_i}{dx}\right|_{x=a}$ and $P(x = a) = P(x = b)$.
   e.g. where $x$ is an angular variable and $a = 0$, $b = 2\pi$.

N.B. For cases (1)-(3), we can make different choices of BCs et each end and still satisfy the S-L condition. So we might talk about Fixed-Open or Dirichlet-Neumann BCs, for instance.

The remainder of this section is more mathematical, and proves some of the properties that we have quoted above. Still, these proofs are a favourite examination question, so you should make sure you know how to do it.

### 2.11.3 Showing the S-L operator is Hermitian

We define a S-L operator

$$\mathcal{L}(x) = \left[\frac{d}{dx}\left(P(x)\frac{d}{dx}\right) + Q(x)\right] .$$

The S-L operator is said to be **Hermitian** or **self-adjoint**. By this we mean that if we define a **matrix element**

$$M_{ij} \equiv \int_a^b dx\ \phi_i^*(x)\ \mathcal{L}(x)\ \phi_j(x)$$

then $M_{ij} = (M_{ji})^*$ (just like we would say for a Hermitian matrix), or

$$\int_a^b dx\ \phi_i^*\mathcal{L}\phi_j = \left(\int_a^b dx\ \phi_j^*\mathcal{L}\phi_i\right)^* = \int_a^b dx\ \phi_j\mathcal{L}^*\phi_i^*$$

Proof: The proof is quite easy. We show below that

$$\int_a^b dx\ \phi_i^*\ \mathcal{L}\ \phi_j = \int_a^b dx\ \phi_j\ \mathcal{L}\ \phi_i^* + \left[P(x)\left(\phi_i^*\frac{d\phi_j}{dx} - \phi_j\frac{d\phi_i^*}{dx}\right)\right]_a^b \tag{11.3}$$

So, when we reorder the functions, we get an extra term. This is called a **surface term** because its value depends only on the value of [...] evaluated at the two boundaries of the problem. The points $x = a, b$ form the zero-dimensional surface of the 1-dimensional line element $a \leq x \leq b$.

The surface term that we get is precisely that in Eqn. (11.2). So, the S-L restriction on BCs is exactly so that we set the surface term to zero, and we then have

$$\int_a^b dx\ \phi_i^*\ \mathcal{L}\ \phi_j = \int_a^b dx\ \phi_j\ \mathcal{L}\ \phi_i^* = \int_a^b dx\ \phi_j\ \mathcal{L}^*\ \phi_i^*$$

because S-L theory only works if $P$ and $Q$ are real, so $\mathcal{L} = \mathcal{L}^*$.

This proves the S-L is Hermitian (and explains some of the restrictions we made earlier).

#### Aside: proving the reordering relation

In this section we prove Eqn. (11.3).

The integrals are linear in $\mathcal{L}$, so we can break $\mathcal{L}$ into its two parts that we integrate separately. First the part $Q(x)$:

$$\int_a^b dx\ \phi_i^*Q(x)\phi_j = \int_a^b dx\ \phi_jQ(x)\phi_i^* \tag{11.4}$$

by simple rearrangement ($Q(x)$ is just a function).

The term in $P(x)$ is more complicated, as it contains derivatives and we cannot just take functions in and out of these. Instead:

$$\int_a^b dx\ \phi_i^*\frac{d}{dx}\left(P(x)\frac{d\phi_j}{dx}\right) = \left[\phi_i^*P(x)\frac{d\phi_j}{dx}\right]_a^b - \int_a^b dx\ \frac{d\phi_i^*}{dx}P(x)\frac{d\phi_j}{dx} .$$

To do this, we have integrated by parts, identifying $u = \phi_i^*$ and $dv = \frac{d}{dx}\left(P(x)\frac{d\phi_j}{dx}\right)$ (so $v = P(x)\frac{d\phi_j}{dx}$).

Now we integrate by parts again, identifying this time $u = \frac{d\phi_i^*}{dx}P(x)$ and $dv = \frac{d\phi_j}{dx}$ (so $v = \phi_j$). This gives

$$\begin{aligned}\int_a^b dx\ \phi_i^* \frac{d}{dx}\left(P(x)\frac{d\phi_j}{dx}\right) &= \left[\phi_i^* P(x)\frac{d\phi_j}{dx}\right]_a^b - \left[\frac{d\phi_i^*}{dx}P(x)\phi_j\right]_a^b + \int_a^b dx\ \frac{d}{dx}\left(\frac{d\phi_i^*}{dx}P(x)\right)\phi_j \\ &= \int_a^b dx\ \phi_j \frac{d}{dx}\left(P(x)\frac{d\phi_i^*}{dx}\right) + \left[P(x)\left(\phi_i^*\frac{d\phi_j}{dx} - \phi_j\frac{d\phi_i^*}{dx}\right)\right]_a^b \quad (11.5)\end{aligned}$$

In the final line, we have just rearranged stuff, but without changing what the derivatives act upon. Adding Eqns. (11.4) and (11.5), we get Eqn. (11.3).

**More technical aside**

We do not need to use the eigenfunctions to define the matrix elements of an operator $H(x)$. In general we can use any complete set of functions. Call them $\psi_i(x)$. The matrix elements are

$$M_{ij} = \int_a^b dx\ \psi_i^*(x)\ H(x)\ \psi_j(x)$$

The actual value of a given matrix element (i.e. specific values for $i$ and $j$) is basis dependent and will change if we use a different set of basis functions.

Nonetheless, if we can show that an operator is Hermitian in one basis, we know it will be true in all bases.

If we choose as our basis the normalised eigenfunctions of $H$, it is easy to show that $M_{ij} \propto \delta_{ij}$.

All of this is very analogous to matrix manipulations.

### 2.11.4 S-L as an eigenvalue problem

S-L theory tells us about the properties of solutions to $\mathcal{L}(x)\phi_i(x) = -\lambda_i\rho(x)\phi_i(x)$. This is a generalised eigenvalue problem, with $\lambda_i$ the eigenvalues and $\phi_i(x)$ the eigenfunctions (equivalent of eigenvectors). The extra $\rho$ function is why we call this "generalised".

We have shown that $\mathcal{L}$ is Hermitian. In this section we'll show that this ensures both that the eigenvalues are real and that the eigenfunctions form a complete set.

Different solutions to the S-L problem are labelled by different $i$ values. Let's choose two solutions, one labelled by some specific value $i$ and the other by some specific value $j$:

$$\mathcal{L}(x)\ \phi_i(x) = -\lambda_i\ \rho(x)\ \phi_i(x) \quad (11.6)$$
$$\mathcal{L}(x)\ \phi_j(x) = -\lambda_j\ \rho(x)\ \phi_j(x) \quad (11.7)$$

If we complex conjugate the first, Eqn. (11.6) we get (remembering $\mathcal{L}^* = \mathcal{L}$)

$$\mathcal{L}(x)\ \phi_i^*(x) = -\lambda_i^*\ \rho(x)\ \phi_i^*(x) \quad (11.8)$$

Now, pre-multiply (i.e. "multiply on the LHS") both sides of Eqn. (11.7) by $\phi_i^*$, and pre-multiply Eqn. (11.8) by $\phi_j$. Integrate both sides of both equations from $x = a$ to $x = b$ to give (respectively):

$$\int_a^b dx\ \phi_i^*(x)\ \mathcal{L}(x)\ \phi_j(x) = -\lambda_j \int_a^b dx\ \rho(x)\ \phi_i^*(x)\ \phi_j(x)$$

$$\int_a^b dx\ \phi_j(x)\ \mathcal{L}(x)\ \phi_i^*(x) = \int_a^b dx\ \phi_i^*(x)\ \mathcal{L}(x)\ \phi_j(x) = -\lambda_i^* \int_a^b dx\ \rho(x)\ \phi_i^*(x)\ \phi_j(x)$$

We have used the Hermiticity of $\mathcal{L}$ (that we proved above) to rearrange the LHS of the second equation. The LHSs of the two equations are now the same, so we can subtract one equation from the other to give:

$$(\lambda_j - \lambda_i^*) \int_a^b dx\ \rho(x)\ \phi_i^*(x)\ \phi_j(x) = 0\ .$$

We are free to choose $i$ and $j$ as we please. There are two cases:

1. $i = j$:
   In general, $\int_a^b dx\ \rho\ \phi_i^*\ \phi_j \neq 0$, as otherwise the eigenfunctions would not be normalisable (which is unphysical). So
   $$\lambda_i = \lambda_i^* \qquad \text{implying the eigenvalues are real}$$
2. $i \neq j$:
   In general $\lambda_i \neq \lambda_j$ so
   $$\int_a^b dx\ \rho(x)\ \phi_i^*(x)\ \phi_j(x) \equiv \phi_i(x) \cdot \phi_j(x) = 0$$
   and the eigenfunctions are orthogonal.

ASIDE: there is the additional case $i \neq j$ but $\lambda_i = \lambda_j$, in this case we are are not guaranteed that $\phi_i \cdot \phi_j = 0$. We are guaranteed, however, that we can pick two linear combinations of $\phi_i$ and $\phi_j$ that are orthogonal to each other and all other eigenfunctions. So, we can always arrange orthogonality.

Hermiticity also guarantees that the eigenfunctions $\{\ \phi_i(x)\ \}$ form a complete set of functions in the interval $a \leq x \leq b$, but we will not prove that here.

In the remainder of this course, we will use the results of Sturm-Liouville theory quite a lot.

### 2.11.5 Orthogonality of normal modes

Physicists tend to be somewhat lazy (or perhaps unclear?) about decomposing general functions into eigenfunctions of a given differential operator.

Sturm–Liouville theory describes a class of problem and gives some useful results. We will not go into a complete discussion of Sturm-Liouville problems, but note that the ODE equations

we consider in this course in fact satisfy the Sturm-Liouville conditions. We summarise some of the precise statements made possible.

Suppose $X_1$ and $X_2$ are eigenfunctions of $\frac{d^2}{dx^2}$, with eigenvalues $-\lambda_1$ and $-\lambda_2$, and satisfy some BCs at $x = a$ and $x = b$.

$$\ddot{X}_1 = -\lambda_1 X_1$$
$$\ddot{X}_2 = -\lambda_2 X_2$$

Observe:

$$\frac{d}{dx}(\dot{X}_1 X_2 - X_1 \dot{X}_2) = \ddot{X}_1 X_2 - X_1 \ddot{X}_2$$

Thus,

$$\begin{aligned} \int_a^b \left(\ddot{X}_1 X_2 - X_1 \ddot{X}_2\right) dx &= \left[\dot{X}_1 X_2 - X_1 \dot{X}_2\right]_a^b \\ &= (\lambda_2 - \lambda_1) \int_a^b X_1 X_2 dx \end{aligned} \tag{11.9}$$

Now, for many standard boundary conditions the *Wronskian* $\left[\dot{X}_1 X_2 - X_1 \dot{X}_2\right]_a^b = 0$, for example:

Dirichlet $X_1(a) = X_1(b) = X_2(a) = X_2(b) = 0$

Neumann $\dot{X}_1(a) = \dot{X}_1(b) = \dot{X}_2(a) = \dot{X}_2(b) = 0$

Periodic $\dot{X}_1(a) = \dot{X}_1(b)\ ; \dot{X}_2(a) = \dot{X}_2(b)\ ; X_1(a) = X_1(b)\ ; X_2(a) = X_2(b)$

In these boundary conditions, eigenfunctions with distinct eigenvalues are *orthogonal* under the scalar (dot) product defined by:

$$X_1 \cdot X_2 = \int_a^b X_1 X_2 dx$$

If $\lambda_1 = \lambda_2$, then we are not guaranteed orthogonality, however, if $X_1$ and $X_2$ are genuinely different (linearly independent) then the Gramm-Schmidt procedure in section **??** allows us to construct an orthonormal basis in any case.

### 2.11.6 Completeness of normal modes

*Proof* of completeness is beyond the scope of this course. However, we can state two theorems regarding different *meanings* of "completeness":

**Uniform convergence:**

if a function $f$ has continuous first and second derivatives on $[a, b]$ and $f$ satisfies the boundary conditions then $f$ can be exactly represented by a sum of eigenmodes: it will match at *every* point.

That is the *maximum* deviation between $f(x)$ and the sum $S(x) = \sum_n a_n X_n$ of eigenmodes becomes zero as the number of modes included tends to $\infty$.

$$\max_{x \in a,b} |f(x) - S(x)|^2; \to 0 \tag{11.10}$$

| Any solution of the differential equation satisfying the boundary conditions can be written as a sum of eigenmodes. |
|---|

**$L_2$ convergence:**

If the function $f(x)$ has $\int_a^b |f(x)|^2 dx$ finite it can be approximated by a sum $S(x) = \sum_n a_n X_n$ of eigenmodes in the weaker sense

$$\int_a^b |f(x) - S(x)|^2 dx \to 0 \tag{11.11}$$

(11.11) means that the sum $S(x)$ can deviate from $f(x)$ at certain points, and the maximum error can remain non-zero. However, not for anything other than an *inifinitessimal* distance asotherwise it would contribute something to the integral.

Thus the basis is complete it can describe and only differs infinitessimally close to discontinuities and violations of the boundary conditions.

| any bounded function on the interval $[a, b]$ can be described away from discontinuities & boundaries |
|---|

## 2.12 Generating functions and Ladder operators

Introduce some generating functions and ladder operators

# Chapter 3

# Statistical analysis and fitting data

## 3.1 Probability distributions

In this part of the course we will consider the analysis and fitting of data. This requires a foundation in the study of probability and statistics. We shall see that Gaussian distributions play a particular role in this study.

### 3.1.1 Random variables

A *random variable* $X$ is a measurent or process that gives you a different result each time. The results are distributed randomly according to a *probability density* $P_X$.

$$P_X(x)dx$$

is the probability that any measurement of variable $X$ will have a value between $x$ and $x+dx$.

As all measurements return a value between $-\infty$ and $\infty$ we have

$$\int_{-\infty}^{\infty} P_X(x)dx = 1$$

The average, or mean, value is the first moment of this distribution

$$\bar{X} = \int_{-\infty}^{\infty} xP_X(x)dx$$

The variance is the mean squared deviation from the mean

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x-\bar{X})^2 P_X(x)dx$$

### 3.1.2 Zero mean, unit variance

Any random variable $X$ can be brought into convenient zero mean, unit variance form by considering in its place the scaled and shifted variable

$$X' = b(X-a)$$

where $a = \bar{X}$, and $b = \frac{1}{\sqrt{\mathrm{Var}X}}$. Then $X'$ has zero mean, and unit variance.

### 3.1.3 Independent random variables

Two random variables are *independent* if the value obtained from one does not affect the value obtained from another.

For example, when tossing coins, *even if you have already obtained five heads in a row.* you are equally likely to obtain heads or tails on the sixth try. If you collect statistics on the number of heads, each attempt is independent.

In contrast, if you are drawing cards from a single deck these are not independent as once the Ace of Spades is drawn, all players are very unlikely to receive another (unless you are in an establishment of particularly ill-repute).

### 3.1.4 Adding random variables

We consider adding two independent random variables $X$ and $Y$ together. This is a key to understanding the process of averaging results. Because the variables (measurements) are independent, we can consider measuring one first without affecting the distribution of the other. The probability distribution of the sum $s_{xy}$ is just the *convolution* of the two probability distributions.

$$P_{X+Y}(s_{xy}) = \int_{-\infty}^{\infty} P_X(x)P_Y(s_{xy} - x)dx$$

This can be seen by noting that the probability for variable $X$ to give a result in band $x \to x + dx$ is $P_X(x)dx$. The probability density in the sum $s_{xy} = x + y$ that variable $Y$ then gives a value between value $y = s_{xy} - x$ and $y = s_{xy} - x + ds$ is $P_Y(s_{xy} - x)$.

This can be repeated, adding a third random variable gives

$$P_{X+Y+Z}(s_{xyz}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P_X(x)P_Y(s_{xy} - x)P_Z(s_{xyz} - s_{xy})dxds_{xy}$$

and so on.

### 3.1.5 Scaling random variables

If we scale the random variable by a constant $C$, the probability of having a result between $u = Cx$ and $u + du = Cx + Cdx$ must equal the orginal probability between $x$ and $x + dx$. Thus

$$P_{CX}(u)du = P_{CX}(Cx)Cdx = P_X(x)dx$$

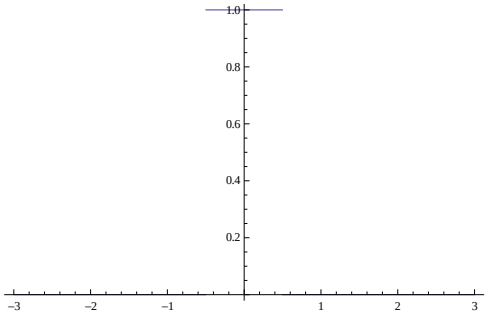
so that

$$P_{CX}(u) = \frac{1}{C}P_X(\frac{u}{c}).$$

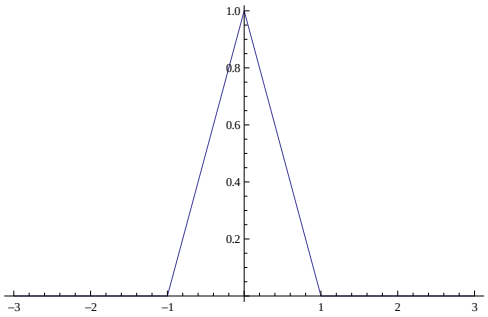
### 3.1.6 Some examples

Now let's add something simple like the top function to itself a number of times. It is well known, and easy to show that performing this operation once produces a triangular function. Beyond this gets tricky due to the piecewise bounds on integrations. However, we can continue to make progress using Mathematica to illustrate our point:

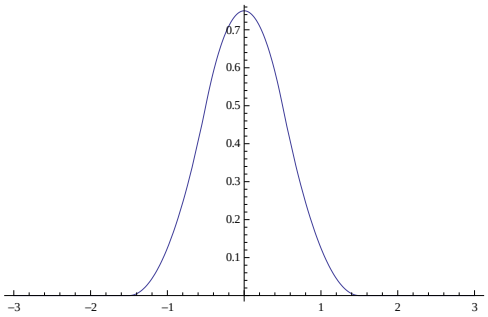
```
h[x] = PieceWise[ {0,x < -0.5}, {1, -0.5<=x<=0.5}, {0,x>0.5}]
Plot[ h[x] , {x,-3,3}]
```



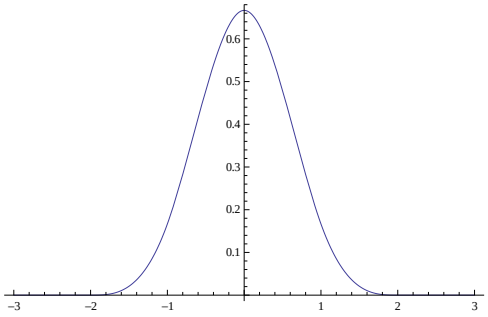

```
Plot[ Convolve[ h[y],h[y],y,x]], {x,-3,3}]
```




```
Plot[ Convolve[Convolve[ h[y],h[y],y,z],h[z],z,x]], {x,-3,3}]
```




```
Plot[ Convolve[Convolve[Convolve[ h[y],h[y],y,z],h[z],z,w],h[w],w,x], {x,-3,3}]
```



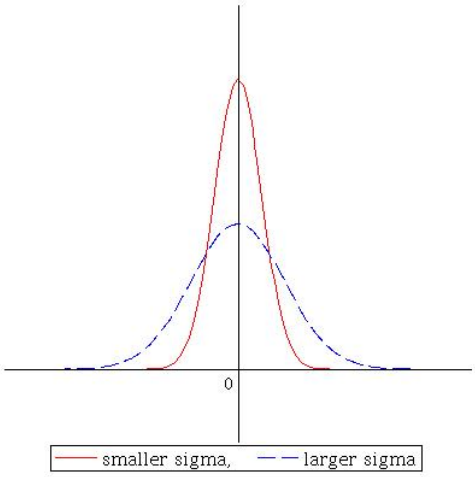

A curious thing happens to the distribution of the average. It is looking more and more bell shaped!

## 3.2 Gaussian distributions

A particularly common form of distribution is the Gaussian, or normal distribution $\mathcal{N}$

$$P_{\mathcal{N}}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

The Gaussian curve is also known as the bell-shaped or normal curve.




Sketch of normalised Gaussians. The intercepts are $P_{\mathcal{N}}(0) = \frac{1}{\sqrt{2\pi\sigma^2}}$.

In order to check this probability is indeed normalised we must evaluate a Gaussian integral. Under a change of variables to $\tilde{x} = \frac{x}{\sqrt{2}\sigma}$ the integrated probability becomes.

$$\int_{-\infty}^{\infty} P_{\mathcal{N}}(x)dx = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-\tilde{x}^2} d\tilde{x}$$

### 3.2.1 Gaussian integral

Gaussian integrals occur regularly in physics. We are interested in

$$I = \int_{-\infty}^{\infty} \exp -x^2 dx$$

This can be solved via a trick in polar coordinates: observe

$$\begin{aligned} I^2 &= \int_{-\infty}^{\infty} \exp -x^2 dx \int_{-\infty}^{\infty} \exp -y^2 dy \\ &= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \exp -(x^2+y^2) \\ &= \int_0^{\infty} dr 2\pi r \exp -r^2 \\ &= \pi \int_0^{\infty} du \exp -u \\ &= \pi \left[ -\exp -u \right]_0^{\infty} \\ &= \pi \end{aligned}$$

Thus,

$$I = \int_{-\infty}^{\infty} \exp -x^2 dx = \sqrt{\pi}$$

Thus, we see

$$\int_{-\infty}^{\infty} P_{\mathcal{N}}(x) dx = 1$$

### 3.2.2 Cumulative distribution, and the error function

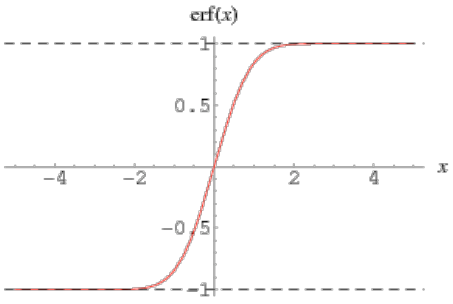
We are often interested in the probability of having a result deviating from its mean by some amount. The question of "what was the chance of that" is measured by the cumulative probability to have a result deviating by an amount measured in standard deviations $a = \frac{x}{\sigma}$ from its mean and is given by the error function

$$\begin{aligned} \mathrm{erf}(\frac{a}{\sqrt{2}}) &= \frac{1}{\sqrt{2\pi}} \int_{-a}^{a} e^{\frac{-x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\frac{a}{\sqrt{2}}}^{\frac{a}{\sqrt{2}}} e^{-\tilde{x}^2} d\tilde{x} \end{aligned}$$

where $\tilde{x} = \frac{x}{\sqrt{2}}$. We generally have to look up this function in tables, or use numerical software to evaluate.

You will notice, unfortunately, there is a $\frac{1}{\sqrt{2}}$ factor between the argument of $\mathrm{erf}(\frac{a}{\sqrt{2}})$ and the deviation from the mean $a$ measured in standard deviations. Thus, the probability of deviating by less than $1\sigma$ is $\mathrm{erf}(0.707) \sim 68\%$.

| Deviation less than | | Percent |
|---|---|---|
| $1\sigma$ | $\mathrm{erf}(0.707)$ | 68% |
| $2\sigma$ | $\mathrm{erf}(1.414)$ | 95% |
| $3\sigma$ | $\mathrm{erf}(2.12)$ | 99.7% |
| $4\sigma$ | $\mathrm{erf}(2.83)$ | 99.992% |
| $5\sigma$ | $\mathrm{erf}(2.12)$ | 99.99993% |




The process of integrating a probability distribution function (PDF), such as the Gaussian, to find the total probability of having a deviation below the observed deviation is common and useful to estimate the likelihood of our having observed something.

This integrated version of the probability, as a function of this threshold, is known as the cumulative distribution function (CDF).

- erf(x) is the CDF corresponding to the Gaussian PDF.
- $1, 2, 3\sigma$ correspond to 68%, 95%, 99.7%
  - Enshrined/rote learned as the "68-95-99.7" rule in statistics
- *Warning:* conventions vary w.r.t. the CDF being integrated from $[-a, a]$ or from $[-\infty, a]$.

### Life and death safety margins ($3\sigma$)

Under EU law, rock climbing carabiners are tested to ensure they can withstand a load of around $21KN$ (around 2 tonnes) with a 3 sigma margin.

This means, worryingly that around 3 in every 1000 *cannot* withstand the 2 tonne rating, however the $\sigma$ for the distribution is never disclosed to the climbers. Your lecturer very much hopes that a $5\sigma$ bound that includes even these weak ones is still more than fit for purpose!

### Experimental discovery standards (5 $\sigma$)

Another example is that standards to which particle physics adheres for particle discovery. When new particles, such as the Higgs boson, are discovered, it no longer always possible to simply "spot it" as a particle track because they decay so quickly.

Rather the theoretical prediction for their contributions to decays must be compared to experimental observation, involving fitting mathematical expressions to data.

The significance of the contribution from a Higgs boson, compared to not having the Higgs boson is critically important. Particle Physicists hold themselves to a $5\sigma$ criterion, which means that there can only be a 0.00007% chance they were simply fooled by luck if or when they claim LHC has "discovered" the Higgs boson.

### 3.2.3 Fourier transform of gaussian

A function $f(x)$ and its Fourier transform $F(k)$ are related by:

$$F(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx\ f(x)\ e^{ikx}\ , \tag{2.1}$$

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk\ F(k)\ e^{-ikx}\ . \tag{2.2}$$

The FT of the Gaussian is

$$\begin{aligned}
\tilde{P}_{\mathcal{N}}(k) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx P_{\mathcal{N}}(x) e^{ikx} \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx\ \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right)\ e^{ikx} \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{k^2\sigma^2}{2}\right)\ ,
\end{aligned}$$

i.e. the FT of a Gaussian is another Gaussian (this time as a function of $k$).

**Deriving the FT** For notational convenience, let's write $a = \frac{1}{2\sigma^2}$, so

$$\tilde{P}_{\mathcal{N}}(k) = \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx\ \exp\left(-\left[ax^2 - ikx\right]\right)$$

Now we can complete the square inside $[\ldots]$:

$$-ax^2 + ikx = -\left(x\sqrt{a} - \frac{ik}{2\sqrt{a}}\right)^2 - \frac{k^2}{4a}$$

giving

$$\tilde{P}_{\mathcal{N}}(k) = \frac{1}{2\pi\sigma} e^{-k^2/4a} \int_{-\infty}^{\infty} dx\ \exp\left(-\left[x\sqrt{a} - \frac{ik}{2\sqrt{a}}\right]^2\right)\ .$$

We then make a change of variables:

$$u = x\sqrt{a} - \frac{ik}{2\sqrt{a}}\ .$$

This does not change the limits on the integral, and the scale factor is $dx = du/\sqrt{a}$, giving

$$\tilde{P}_{\mathcal{N}}(k) = \frac{1}{2\pi\sigma} \frac{1}{\sqrt{a}}\ e^{-k^2/4a} \int_{-\infty}^{\infty} du\ e^{-u^2} = \frac{1}{2\sqrt{\pi}\sigma} \frac{1}{\sqrt{a}} \cdot e^{-k^2/4a}\ . \tag{2.3}$$

Finally, we change back from $a$ to $\sigma$.

### 3.2.4 Convolution of Gaussians

We can use the convolution theorem to convolve together two Gaussians $P_{\mathcal{N}_1}(x)$ and $P_{\mathcal{N}_2}(x)$ of width $\sigma_1$ and $\sigma_2$ to represent the distribution of the sum using the convolution theorem:

$$P_{\mathcal{N}_1}(x) * P_{\mathcal{N}_2}(x) = \sqrt{2\pi}\left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dk \tilde{P}_{\mathcal{N}_1}(k)\tilde{P}_{\mathcal{N}_2}(k)e^{-ikx}\right)$$

$$\begin{aligned} P_{\mathcal{N}_1}(x) * P_{\mathcal{N}_2}(x) &= \sqrt{2\pi}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dk e^{-ikx}\frac{1}{\sqrt{2\pi}}e^{-\frac{k^2\sigma_1^2}{2}}\frac{1}{\sqrt{2\pi}}e^{-\frac{k^2\sigma_2^2}{2}} \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dk e^{-ikx}\frac{1}{\sqrt{2\pi}}e^{-\frac{k^2(\sigma_1^2+\sigma_2^2)}{2}} \end{aligned}$$

Which we can recognise from Eq. 2.3 as a normalised Gaussian of width $\sigma^2 = \sigma_1^2 + \sigma_2^2$.

- This is important - Gaussian distributed random variables remain Gaussian distributed under addition (also their widths add in quadrature).
- Further, we just saw in a handwaving sense that random variables of other distributions become more Gaussian-like after addition.

It appears that the Gaussian distribution "attracts" all other distributions under the avergaing process, and we shall see in the next section that this is indeed true.

## 3.3 Central limit theorem

The Central Limit Theorem (CLT) is of fundamental importance to almost all experimental science.

Take $X$ as a zero mean, unit variance random variable with *any* distribution $P_X(x)$.

The random variable for the average of $N$ of these random variables

$$S_N = \frac{1}{N}\left(X_1 + X_2 + \ldots X_N\right)$$

and in the limit of large $N$, this is distributed according to

$$\lim_{N\to\infty} P_{S_N}(x) \to \frac{\sqrt{N}}{\sqrt{2\pi}} e^{-\frac{Nx}{2}}$$

**Proof**

The key point is that the average of many measurements becomes increasingly dominated by the lowest frequencies in the probability distribution.

In other words, small wiggles in the probability distribution will cancel and not affect the average of many measurements; sometimes you get lucky, sometimes unlucky, but on average small wiggles disappear rapidly.

1. First we represent sum of $N$ independent random variables drawn with probability $P_X(x)$ as a convolution.

$$P_{NX}(x') = \int_{-\infty}^{\infty} dx_1 \ldots \int_{-\infty}^{\infty} dx_{N-1} P_X(x_1) P_X(x_2 - x_1) P_X(x_3 - x_2) \ldots P_X(x' - x_{N-1})$$

2. We can apply the convolution theorem $N-1$ times obtaining

$$P_{NX}(x') = \frac{1}{\sqrt{2\pi}}(\sqrt{2\pi})^{N-1} \int_{-\infty}^{\infty} dk e^{-ikx'} \left(\tilde{P}_X(k)\right)^N dk$$

3. Taking the factor of $\frac{1}{N}$ for the average and applying our scaling rule Section 3.1.5 gives

$$\begin{aligned} P_{S_N}(u) &= N\frac{1}{\sqrt{2\pi}}(\sqrt{2\pi})^{N-1} \int_{-\infty}^{\infty} dk e^{-ikNu} \left(\tilde{P}_X(k)\right)^N dk \\ &= \frac{1}{\sqrt{2\pi}}(\sqrt{2\pi})^{N-1} \int_{-\infty}^{\infty} dk' e^{-ik'u} \left(\tilde{P}_X(\frac{k'}{N})\right)^N dk' \end{aligned}$$

4. In the limit of large $N$, the lowest frequency terms of $\tilde{P}_X(\frac{k'}{N})$ dominate, and these correspond to the lowest *moments* of the distribution.

We can Taylor expand

$$\begin{aligned}\tilde{P}_X(\frac{k'}{N}) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{i\frac{k'x}{N}} P_X(x)dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\left[1+\frac{ik'x}{N}-\frac{k^2x^2}{2N^2}+\ldots\right]P_X(x)dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}P_X(x)dx+\frac{ik'}{N}\int_{-\infty}^{\infty}xP_X(x)dx-\frac{k^2x^2}{2N^2}\int_{-\infty}^{\infty}x^2P_X(x)dx \\ &= \frac{1}{\sqrt{2\pi}}\left(1-\frac{k^2}{2N^2}+\ldots\right)\end{aligned}$$

where we have used the fact that $P_X(x)$ is normalised, zero mean, and unit variance.

5. Inserting this into the full expression for $P_{S_N}(u)$ we have

$$\begin{aligned}P_{S_N}(u) &= \frac{1}{\sqrt{2\pi}}(\sqrt{2\pi})^{N-1}(\frac{1}{\sqrt{2\pi}})^N\int_{-\infty}^{\infty}dk'e^{-ik'u}\left(1-\frac{k^2}{2N}\frac{1}{N}\right)^N dk' \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}dk'e^{-ik'u}\left(1-\frac{k^2}{2N}\frac{1}{N}\right)^N dk'\end{aligned}$$

6. The final significant step is to note that with $b=-\frac{k^2}{2N}$

$$\begin{aligned}\left(1+\frac{b}{N}\right)^N &= \sum_{r=0}^{N}\frac{(N!)}{r!(N-r)!}b^r \\ &= \sum_{r=0}^{N}\frac{N(N-1)\ldots(N-r+1)}{N^r}\frac{b^r}{r!} \\ &\to \sum_{r=0}^{N}\frac{b^r}{r!} \\ &\to e^b\end{aligned}$$

Thus,

$$P_{S_N}(u)\to\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}dk'e^{-ik'u}\frac{1}{\sqrt{2\pi}}e^{-\frac{k^2}{2N}}dk'$$

which we can recognise from Eq 2.3 as a normalised Gaussian of width $\sigma_S^2=\frac{1}{N}$

$$\begin{aligned}P_{S_N}(u) &= \frac{\sqrt{N}}{\sqrt{2\pi}}e^{-\frac{u^2N}{2}} \\ &= \frac{1}{\sqrt{2\pi}\sigma_S}e^{-\frac{u^2}{2\sigma_S^2}}\end{aligned}$$

**Consequences**

This theorem has very important consequences that permeate most of science!

- If we measure something many times, we can assume the average has a *Gaussian* distribution.
- If the original non-Gaussian distribution had unit variance, the variance of the average is $\frac{1}{\sqrt{N}}$.
- If original distribution was not unit variance, the variance of the mean is reduced by $\frac{1}{\sqrt{N}}$

## 3.4 Analysing data

We now consider how to analyse and/or fit experimental (or numerical) data. Each data point can be measured a number of times, and the mean and the error on the mean can be computed as follows.

If we take $N$-measurements $\{x_1, \ldots, x_N\}$, sampled randomly from a (non-Gaussian) probability distribution $P_X(x)$

The sample mean is

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

This differs from the true mean

$$\bar{X} = \int_{-\infty}^{\infty} x P_X(x) dx$$

but will become equal in the limit of infinite $N$.

The sample mean is the best estimate of the mean of the non-Gaussian distribution $P_X$ one can make given the finite number of measurements.

### 3.4.1 Variance

The variance of the sample is

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2$$

This is a best estimate of the variance of the non-Gaussian distribution $P_X$ one can make given the finite number of measurements.

The $(N-1)$ arises from an adjustment to prevent underestimating the variance with a finite sample size. Practically, if the change from $N$ to $(N-1)$ makes a worrying difference, your Lecturer recommends you get more data.

### 3.4.2 Standard deviation

An important question that arises is how incorrect is the sample mean *likely* to be. For a finite sample size, this question cannot be answered without knowing the details of the original distribution $P_X(x)$.

However, if we are prepared to assume that we have sufficiently many measurements that our finite sample yields a good estimate of variance of $P_X(x)$, and that the distributions of averages will approach its asymptotic Gaussian form we can proceed.

The variance of the Gaussian distributed mean will be $\frac{1}{\sqrt{N}}\sigma$.

For moderate $N$ it is more conventional to take the standard error of the mean as

$$\text{std.err.} = \sqrt{\frac{1}{N(N-1)}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

## 3.5 $\chi^2$/dof and the $\chi^2$ distribution

If you roll a dice many times, you are bound to eventually get lucky or unlucky. If we make a million measurements we expect to see one in a million probability events.

So, when we make many measurements, we should ask how likely we were to have seen a certain set of deviations: me must ask if, on the balance of probability, is this *consistent* with the errors we have calculated for each data point.

- Suppose we take N-samples $\{x_i\}$ from normally distributed variables of unit variance, zero mean, each with probability $P_{\mathcal{N}}(x)$.

  We consider these N-samples as be a single sample from a N-dimensional space.

- The probability distribution for this quantity $\chi^2$ is determined by the normal distributions of the individual variables.

  Since the variables are independent, they probability density for the N-samples is just the product of the individual variables probabilities:

$$P(x_1, \ldots, x_N)dx_1 \ldots dx_N = P_{\mathcal{N}}(x_1) \ldots P_{\mathcal{N}}(x_N)dx_1 \ldots dx_N$$

- It is simpler to reduce our measurement of the vector $\{x_1, \ldots x_N\}$ to a single number. To do this, we define the quantity $\chi^2$ to be the sum of the deviations from zero:

$$\chi^2 = \sum_i x_i^2$$

- The expectation value of $\chi^2$ is given by

$$\begin{aligned}
\langle\chi^2\rangle &= \int \chi^2 P_{\chi^2}(\chi^2)d(\chi^2) \\
&= \int (x_1^2 + \ldots + x_N^2) P_{\mathcal{N}}(x_1) \ldots P_{\mathcal{N}}(x_N)dx_1 \ldots dx_N \\
&= \int x_1^2 P_{\mathcal{N}}(x_1)dx_1 \int P_{\mathcal{N}}(x_2)dx_2 \ldots \int P_{\mathcal{N}}(x_N)dx_N \\
&+ \int P_{\mathcal{N}}(x_1)dx_1 \int x_2^2 P_{\mathcal{N}}(x_2)dx_2 \ldots \int P_{\mathcal{N}}(x_N)dx_N \\
&\ldots \\
&+ \int P_{\mathcal{N}}(x_1)dx_1 \int P_{\mathcal{N}}(x_2)dx_2 \ldots \int x_N^2 P_{\mathcal{N}}(x_N)dx_N \\
&= 1 + \ldots + 1 \\
&= N
\end{aligned}$$

where we have made use of the fact that the distribution $P_{\mathcal{N}}(x)$ had unit variance, and unit total probability.

- The average value for $\chi^2$ is $N$, the number of datapoints.
  - A value of $\chi^2/N \gg 1$ will indicate non-Gaussian, or significantly underestimated errors.
  - A value of $\chi^2/N \ll 1$ will indicate significantly overestimated errors.
  - One does not expect to be either abnormally lucky, or unlucky and having results that appear so may indicate other problems in data or model.

### 3.5.1 Some detail on the distribution of $\chi^2$

We wish to obtain the probability distribution $P^N_{\chi^2}(\chi^2)$ of $\chi^2$ for $N$ variables in more detail than simply its mean value.

We must use theta functions to evaluate the integral of probability to find $\chi^2$ in the spherical shell corresponding to radius $r \equiv \sqrt{\chi^2}$ between $r = \sqrt{\chi^2}$ and $r + \delta r = \sqrt{\chi^2 + \delta(\chi^2)}$ in the N-dimensional space of deviations $\{x_i\}$.

We need to do this to identify the 68th percentile in the cumulative distribution of $\chi^2$ for $n$ independent variables (for example). This is critical to estimating the uncertainty in parameters after a fit.

$$\begin{aligned}
P^N_{\chi^2}(\chi^2)\delta(\chi^2) &= \frac{1}{(\sqrt{2\pi})^N}\int \theta(|X|^2-\chi^2)\theta(\chi^2+\delta(\chi^2)-|X|^2)e^{-\frac{x_1^2}{2}}\dots e^{-\frac{x_N^2}{2}}dx_1\dots dx_N \\
&= \frac{1}{(\sqrt{2\pi})^N}\int \theta(|X|^2-\chi^2)\theta(\chi^2+\delta(\chi^2)-|X|^2)e^{-\frac{\chi^2}{2}}dx_1\dots dx_N \\
&= \frac{1}{(\sqrt{2\pi})^N}e^{-\frac{r^2}{2}}S^N(r)\delta r
\end{aligned}$$

Here $S^N(r)$ is the surface area of an N-dimensional ball of radius $r$ for general dimensions (without proof here), this is

$$S^N(r) = \frac{Nr^{N-1}\pi^{\frac{N}{2}}}{\Gamma(N/2+1)}$$

Recall, $\Gamma(x) = (x-1)\Gamma(x-1)$, $\Gamma(N) = (N-1)!$, and $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. It can be easily verified for two and three dimensions that this produces formulae we know and love!

$$\begin{aligned}
S_2(r) &= 2\pi r \\
S_3(r) &= 4\pi r^2 \\
S_4(r) &= 2\pi^2 r^3
\end{aligned}$$

Note that

$$\delta(\chi^2) = \delta(r^2) = 2r\delta r,$$

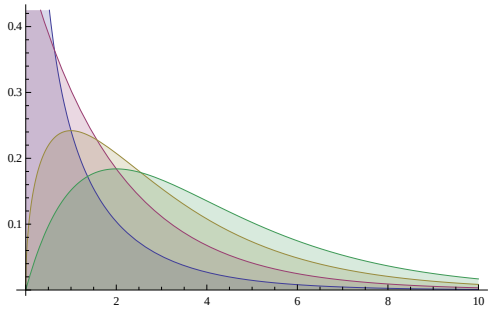
and so

$$\delta r = \frac{\delta\chi^2}{2r}$$

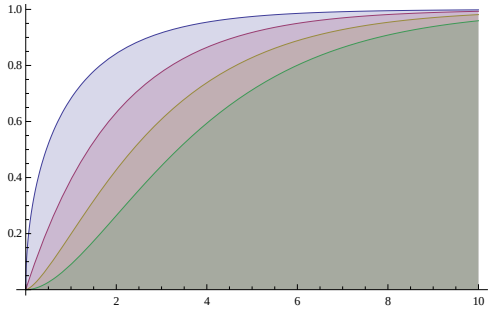
Thus,

$$P^N_{\chi^2}(\chi^2)\delta(\chi^2) = \frac{1}{(\sqrt{2})^N}\frac{\delta(\chi^2)}{2\sqrt{\chi^2}}e^{-\frac{\chi^2}{2}}\frac{N(\chi^2)^{\frac{N-1}{2}}}{\Gamma(N/2+1)}$$
$$= \frac{(\chi^2)^{\frac{N}{2}-1}}{2^{\frac{N}{2}}\Gamma(N/2)}e^{-\frac{\chi^2}{2}}\delta(\chi^2)$$

The prefactors are painful normalisation. However, the change in the power of $\chi^2$ in front of the exponential as $N$ is varied is much more important as these change the shape of the $\chi^2$ distribution as a function of $N$.




The probability distribution function $P^N_{\chi^2}(\chi^2)$ for $\chi^2$ as a function $\chi^2$ for different numbers $n$ of random Gaussian variables that are averaged ($n = 1, 2, 3, 4$).




The cumulative distribution function (CDF)

$$P^{\text{CDF}}_{\chi^2}(N, \chi^2) = \int_0^{\chi^2} P^N_{\chi^2}(Q)dQ$$

for $\chi^2$ as a function $\chi^2$ for different numbers $n$ of random Gaussian variables that are averaged ($n = 1, 2, 3, 4$).

Here (this is *not* proven)

$$P^{\text{CDF}}_{\chi^2}(N, \chi^2) = \frac{1}{\Gamma\frac{N}{2}}\gamma(\frac{N}{2}, \frac{\chi^2}{2})$$

where $\gamma$ is the lower imcomplete gamma function

$$\gamma(s, x) = \int_0^x t^{s-1}e^{-t}dt$$

As we expect from our discussion of the mean value of $\chi^2$, the cumulative distribution grows roughly linearly in $N$.

We will be interested in the value of $\chi^2$ that covers 68.3% of fluctuations (or in other words $1\sigma$). This means the value of $\chi^2$ for which the cumulative distrubution function has value 0.683. This value is dependent on the number of degrees of freedom.

| n | 68th percentile $\chi^2$ |
|---|---|
| 1 | 1 |
| 2 | 2.3 |
| 3 | 3.5 |
| 4 | 4.7 |
| 5 | 5.9 |

We can now meaningfully ask how model parameters could be varied while staying within this bound on $\chi^2$.

## 3.6 $\chi^2$ minimisation

The definition of $\chi^2$ can now be generalised to the case where the i-th variable has non-zero mean and non-unit variance in the obvious way.

$$\chi^2 = \sum_i \frac{(x_i - \bar{x}_i)^2}{\sigma_i^2}$$

All the earlier results remain applicable because the modified variables $\tilde{x} = \frac{x-\bar{x}}{\sigma}$ are normally distributed with zero mean and unit variance, and our generalised $\chi^2$ is just

$$\chi^2 = \sum_i \frac{(x_i - \bar{x}_i)^2}{\sigma_i^2} = \sum_i \tilde{x}_i^2$$

In particular, we still expect that the average generalised $\chi^2$ should still be the number variables $N$.

Thus far we assumed that there was some external, given probability distribution. In the real world, the true distribution will *not* be known.

### 3.6.1 Averaging *is* fitting to a constant

Suppose the data is distributed as a normal distribution around some unknown mean $A$, but all the data has the same width $\sigma$.

$$P_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-A)^2}{2\sigma^2}}$$

As we do not know $A$ independent of the data, we must ask what is the *probability* for having obtained our set of $N$ results $x_i$, for each possible value of $A$. This probability is just

$$P_X(x_1)P_X(x_2)\dots P_X(x_N) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^N e^{-\sum \frac{(x_i-A)^2}{2\sigma^2}}$$

Gaussian statistics tells us the probability of data having been produced from some "externally given" model. The maximum likelihood method involves the reasonable approach of turning this around: use the the Gaussian probability for the data having been produced for each possible parameter to *distinguish* which is the *right* parameter.

Thus, the most *likely* value for $A$ is obtained by maximising this probability. This is achieved by minimising the term in the exponent $\frac{(x_i-A)^2}{2\sigma^2}$, and hence by minimising our $\chi^2$.

The minimum of $\chi^2$ is found by requiring the derivative wrt $A$ be zero:

$$0 = \frac{d}{dA}\chi^2 = -\frac{1}{\sigma^2}\sum_i 2(x_i - A)$$

and so

$$A = \frac{1}{N}\sum_i x_i$$

and unsurprisingly the most likely value for $A$ is the average of the data.

The importance of this very predictable result is that it points the way forward for more complicated situations.

### 3.6.2 Weighted Averages

Suppose we have data distributed as a normal distribution around some unknown mean $A$, but where each measured data point has a different width $\sigma_i$.

$$P_{X_i}(x_i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x_i-A)^2}{2\sigma_i^2}}$$

Now, varying $A$ to maximising the probability of the distribution describing our data yields:

$$0 = \frac{d}{dA}\chi^2 = -2\sum_i \frac{x_i - A}{\sigma_i^2}$$

and so,

$$A = \frac{\sum_i \frac{x_i}{\sigma_i^2}}{\sum_j \frac{1}{\sigma_j^2}}$$

This is the (famous) weighted mean for averaging data points with errors.

At the minimum,

$$\frac{d^2\chi^2}{dA^2} = 2\sum_i \frac{1}{\sigma_i^2}$$

We take the variance of the weighted average as

$$\sigma_A^2 = \frac{1}{\sum_j \frac{1}{\sigma_j^2}}$$

If we vary $A$ by an amount $\sigma_A$, then the change in $\chi^2$ is

$$\Delta\chi^2 = \frac{1}{2} \cdot \frac{d^2\chi^2}{dA^2}\sigma_A^2 = 1$$

If all the $\sigma_i$ are the same, this gives our familiar $\sqrt{N}$ reduction for the error on the mean.

**PDG scale factors**

PDG SCALE FACTOR – WHAT TO DO IF $\chi^2$ WAS BAD

## 3.7 General curve fitting

Suppose we have some number $N_x$ of experimental inputs (x-values, $\{x_i\}$) and experimental results (y-values, $\{y_i\}$), each of which has a random error, $\{\sigma_i\}$.

*The random variables are now the co-ordinates $y_i$, and the $x_i$ are precisely known ordinates. We seek to fit a model curve $y = f_{\{p\}}(x)$*

These which can be plotted on a two dimensional plot, and we have some expected model function $f_{\{p\}}(x)$ depending on some set of $N_p$ parameters $\{p_i\} = \{p_1, \ldots, p_{N_p}\}$.

We expect that, with the "true" parameters,

$$y = f_{\{p\}}(x)$$

will be a curve that well describes the data,

Our expected probability distribution depends on the parameters $\{p\}$

$$P_{Y_i}(y_i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(y_i - f_{\{p\}}(x_i))^2}{2\sigma_i^2}}$$

The definition of $\chi^2$ becomes

$$\chi^2 = \sum_i \frac{(y_i - f_{\{p\}}(x_i))^2}{\sigma_i^2}$$

and we seek the parameters $p$ that minimise $\chi^2$.

We must find parameters $\{p\}$ such that for all $j$

$$0 = \frac{\partial\chi^2}{\partial p_j} = -2\sum_n \frac{\partial f_{\{p\}}(x_n)}{\partial p_j} \frac{(y_n - f_{\{p\}}(x_n))}{\sigma_n^2}$$

For general functions this minimisation has to be performed numerically. This is normally perfromed using the Marquardt-Levenberg algorithm, which itself is an improvement on steepest descent.

### 3.7.1 Quality of fit

A fit has some number $N_{dof} = N_x - N_p$ of degrees of freedom. For example, if fitting 7 data points with a 2 parameter function we have dof $= 5 = 7 - 2$.

As discussed above, we expect $\chi^2 \sim N_{dof}$ for any fit after we have determined the most likely parameters.

For example, a line always fits two data points perfectly and gives zero $\chi^2$ after fitting. It is only once we have three data points for a two parameter linear fit that we gain any measure of the statistical consistency of the data.

We are interested in the value of $\chi^2$ after we have found those parameters that minimise $\chi^2$. If this $\chi^2$ is too high, we have not been successful in fitting the data with our model.

If $\chi^2/N_{dof}$ is too low, we have unbelievably good fit, and have likely overestimated the errors on the individual datapoints.

### 3.7.2 Confidence regions

We can ask how much the fitted parameters are likely to vary.

There is some set of parameters $\{p_{\min}\}$ that minimise $\chi^2$. At this minimum

$$\frac{\partial}{\partial p_j}\chi^2(\{p_{\min}\}) = 0$$

We define a matrix $M_{ij}$ as the second partial derivative

$$M_{ij} = \frac{\partial^2}{\partial p_i \partial p_j}\chi^2(\{p_{\min}\}) = \sum_n \frac{2}{\sigma_n^2}\left[\frac{\partial f_{\{p\}}(x_n)}{\partial p_i}\frac{\partial f_{\{p\}}(x_n)}{\partial p_j} - \frac{\partial^2 f_{\{p\}}(x_n)}{\partial p_i \partial p_j}(y_n - f_{\{p\}}(x_n))\right]$$

$M_{ij}$ is symmetric, so for example, $M_{12} = M_{21}$. In the region of the minimum we define the parameter shifts

$$\{\tilde{p}\} = \{p\} - \{p_{\min}\}$$

we can Taylor expand $\chi^2$ in the region of the minimum as

$$\chi^2(\tilde{p}) = \chi^2_{\min} + \frac{1}{2}\tilde{p}^T M \tilde{p}$$

#### Joint parameter distribution

We define our region of confidence by

$$\chi^2_{\min} - \chi^2(\tilde{p}) = \frac{1}{2}\tilde{p}^T M \tilde{p} \leq \Delta\chi^2 \tag{7.4}$$

The correct approach is to take $\Delta\chi^2$ to cover the 68th percentile of the cumulative distribution

$$P^{\text{CDF}}_{\chi^2}(N_{\text{dof}}, \Delta_{\chi^2}) = 0.68$$

(see table in section 3.5.1).

It is also fairly common for the $\Delta_{\chi^2}$ defining the allowed space to be simply taken as the approximate value $\frac{\Delta_{\chi^2}}{N_{\rm dof}} = 1$ (i.e.$\Delta_{\chi^2} = N_{\rm dof}$).

- With one parameter there is an allowed band

$$\frac{1}{2}\tilde{p}_1^2 M_{11} \leq \Delta_{\chi^2} = 1$$

- With two parameters there is an allowed ellipse [1]

$$\frac{1}{2}[\tilde{p}_1^2 M_{11} + 2\tilde{p}_1\tilde{p}_2 M_{12} + \tilde{p}_2^2 M_{22}] \leq \Delta_{\chi^2} = 2.3$$

**General case**

With $N_p$ parameters we determine an ellipsoid region parameter space. $M_{ij}$ is symmetric because $\frac{\partial^2}{\partial_i \partial_j} = \frac{\partial^2}{\partial_j \partial_i}$. Symmetric matrices are diagonalisable, and thus $M_{ij}$ is diagonalisable with

$$M_{ij} = V^T \text{diag}(\lambda_1, \ldots, \lambda_N) V$$

where V is a unitary matrix representing a change of basis and whose rows contain the $N_p$ eigenvectors $e_k$ of $M_{ij}$, and these are orthogonal to each other.

**Example**

For example, for $N_p = 2$ parameters, V must have the form

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

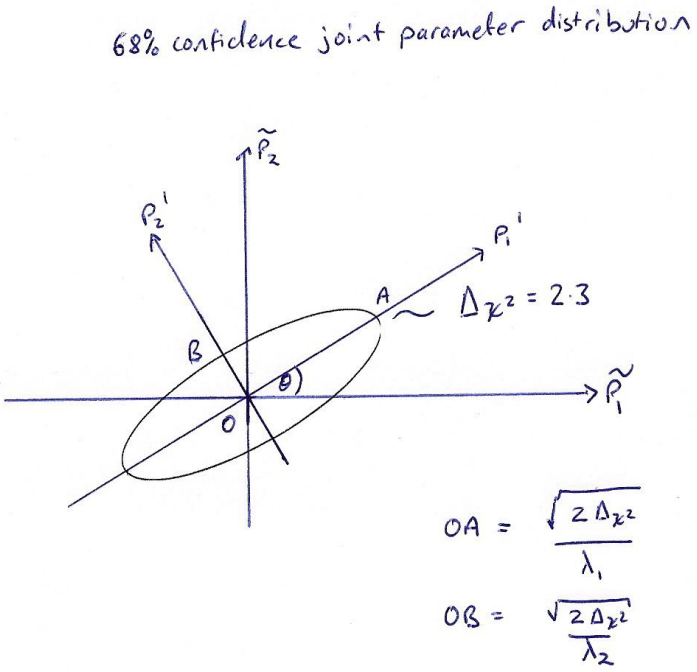
In other words, there are *rotated* parameters

$$p_1' = \cos\theta\tilde{p}_1 + \sin\theta\tilde{p}_2$$
$$p_2' = \cos\theta\tilde{p}_2 - \sin\theta\tilde{p}_1$$

for which the equation $\chi^2 = 2.3$ gives the simple ellipse

$$\lambda_1 p_1'^2 + \lambda_2 p_2'^2 = 2\Delta_{\chi^2}$$

[1] Recall, general ellipse is $ax^2 + by^2 + cxy = d$




The width of the ellipse (major and minor axes) in the $p_1'$ and $p_2'$ directions are $\sqrt{\frac{2\Delta_{\chi^2}}{\lambda_1}}$ and $\sqrt{\frac{2\Delta_{\chi^2}}{\lambda_2}}$ respectively.

We can expand out $M$

$$\begin{aligned} M &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \\ &= \begin{pmatrix} \cos^2\theta\lambda_1 + \lambda_2\sin^2\theta & \sin\theta\cos\theta(\lambda_1 - \lambda_2) \\ \sin\theta\cos\theta(\lambda_1 - \lambda_2) & \sin^2\theta\lambda_1 + \cos^2\theta\lambda_2 \end{pmatrix} \end{aligned}$$

and so,

$$\frac{2M_{12}}{M_{11} - M_{22}} = \frac{2\sin\theta\cos\theta}{\cos^2\theta - \sin^2\theta} = \tan 2\theta$$

Further,

$$\lambda_1 + \lambda_2 = M_{11} + M_{12}$$

and

$$\lambda_1 - \lambda_2 = \frac{M_{12}}{\sin\theta\cos\theta}$$

### 3.7.3 Single parameter errors

Independent standard deviations on *correlated* fit parameters are somewhat ambiguous as the parameters do not independently vary.

The rule is that we take the error on the $j$-th parameter as

$$\sigma_j^2 = 2(M^{-1})_{jj}$$

It is common to define a covariance matrix

$$C_{ij} = 2(M^{-1})_{ij}$$

and the error on the $j$-th parameter is given by

$$\sigma_j^2 = C_{jj}$$

Gnuplot will be introduced in the next section. For completeness, Gnuplot prints the correlation matrix

$$\text{Corr}_{ij} = \frac{C_{ij}}{\sigma_i \sigma_j}$$

In principle this can be used to connect parameter errors $\sigma_j$ to the ellipsoid region of the joint distribution as given above.

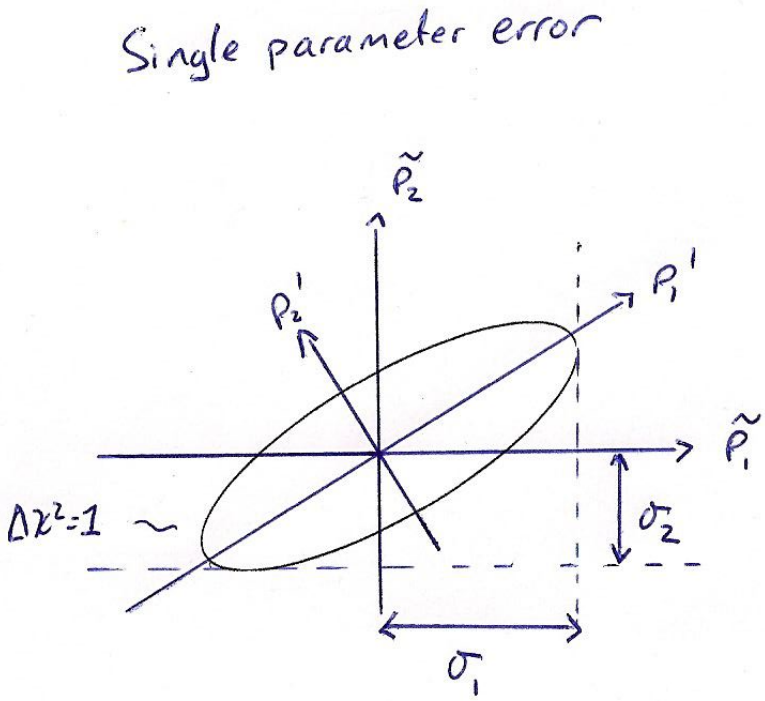
Your lecturer does not believe this is often done in practice because it is involved.

### 3.7.4 Detail on the single parameter error

We must displace one parameter ($\tilde{p}_1$ for example) at a time by an amount $\sigma$, while leaving the other parameters ($\tilde{p}_2 \ldots \tilde{p}_{N_p}$) free to take whatever minimises $\chi^2(\{\tilde{p}\}$.

The error is defined from the 68th percentile of this single degree of freedom $\chi^2$ distribution. That is we find the value of $\tilde{p}_1$ leading to $\delta\chi^2 = 1$ (subject to allowing the other parameters to float to a minimum of $\chi^2$).

Pictorially this means we displace $p_1$ as far as we can while satisfying $\Delta_{\chi^2} = 1$, and select the vertical tangent to the $\Delta_{\chi^2} = 1$ contour




Mathematically, we must minimise

$$\chi^2(\tilde{p}) = \chi^2_{\min} + \frac{1}{2}\tilde{p}^T M \tilde{p}$$

Subject to the *constraint* that $\tilde{p}_1 = \sigma_1$.

### Constrained minimisation: Lagrange multipliers

We introduce a general method for constrained minimisation called the method of Lagrange multipliers. This is generally useful, and not specific to this context.

To minimise $\chi^2(\tilde{p})$ over the surface of $\tilde{p}_1 = \sigma_1$ by varying $\tilde{p}_1 \ldots \tilde{p}_{N_p}$ we require that the components for $\nabla_p \chi^2(\tilde{p})$ parallel to the surface be zero (otherwise, we can reduce $\chi^2$ by moving within the surface).

The normal to this surface is (in our case) $\hat{n} = (1, 0, \ldots, 0)$. [2]

This means that at the minimum $\nabla\chi^2$ must be parallel to the normal and so they are

[2] for a general surface $f(\vec{p}) = C$, the normal is parallel to $\nabla f$

proportional with an unknown constant of proportionality $\lambda$

$$\nabla_p \chi^2(\tilde{p}) = M_{ij}\tilde{p}_j = \lambda\hat{n} = \begin{bmatrix} \lambda \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Where $\lambda$ is an unknown constant called a Lagrange multiplier. Thus,

$$\tilde{p}_j = \lambda M^{-1}_{j1}$$

Now, we chose our surface to have $\tilde{p}_1 = \lambda M^{-1}_{11} = \sigma_1$, and so $\lambda = \frac{\sigma_1}{M^{-1}_{11}}$.

We will then define the value of $\sigma_1$ by requiring that $\chi^2$ hold the value $\chi^2(\tilde{p}) = 1$, in order that we obtain the 68th percentile for a single variable $\chi^2$ distribution. This means that

$$\begin{aligned} \Delta_{\chi^2} = 1 &= \frac{1}{2}\tilde{p}^T M \tilde{p} \\ &= \frac{1}{2}\frac{\sigma_1^2}{(M^{-1}_{11})^2} M^{-1}_{1j} M_{jk} M^{-1}_{k1} \\ &= \frac{1}{2M^{-1}_{11}}\sigma_1^2 \end{aligned}$$

and generalising to any parameter $p_j$

$$\sigma_j^2 = 2(M^{-1})_{jj}$$

## 3.8 gnuplot

The common Linux package gnuplot includes an implementation of the Marquardt-Levenberg algorithm and can plot and fit any function you care to type in.

This is quite useful to demonstrate the topics discussed in this chapter.

First we must prepare our experimental data in a form that gnuplot can read.

1. Use a text editor (emacs or similar) to prepare a file "experiment.dat" containing $x$, $y$ $\sigma_y$ columns

   ```
   # x y dy
   1 1.01 0.01
   2 3.98 0.02
   3 9.02 0.03
   ```

2. start gnuplot in the same directory as the file

   ```
   bash$ gnuplot
   ```

3. First simply plot the data file. For example:

```
gnuplot> plot 'experiment.dat' using 1:2:3 with yerrorbars
gnuplot> set xrange [0:4]
gnuplot> set yrange [0:10]
gnuplot> replot
```

4. We can define a fit function as follows

```
gnuplot> y(x) = A * x*x
gnuplot> A=0.5
gnuplot> replot y(x)
```

5. We can tell gnuplot to perform a fit as follows:

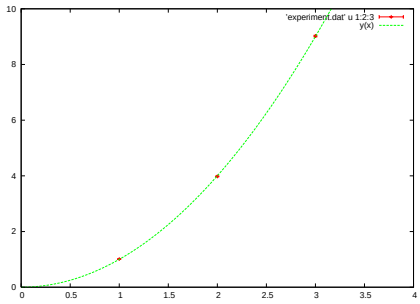
```
gnuplot> fit y(x) 'experiment.dat' using 1:2:3 via A
```

6. We can replot with the fit results

```
gnuplot> replot
```

7. We can save the fit to a postscript file with

```
gnuplot> set terminal postscript landscape color
gnuplot> set output 'experiment.ps'
gnuplot> replot
```




8. We can look at the fitted parameters. Gnuplot prints

```
After 4 iterations the fit converged.
final sum of squares of residuals : 2.37302
rel. change during last iteration : -1.13084e-11

degrees of freedom (ndf) : 2
rms of residuals      (stdfit) = sqrt(WSSR/ndf)      : 1.08927
variance of residuals (reduced chisquare) = WSSR/ndf : 1.18651

Final set of parameters            Asymptotic Standard Error
=======================            ==========================
```

```
A                = 1.00071          +/- 0.002911     (0.2909%)


correlation matrix of the fit parameters:

                A
A                1.000
```

This means it obtained $A = 1.00071$ with $\sigma_A = 0.0029$ (in otherwords $A = 1.001(3)$).