

Lecture Notes

Year 1 Semester 1
MSci Physics with Particle Physics and Cosmology

Ash Stewart
University of Birmingham

Lectures Index

| | |
|--|-----------|
| LC Classical Mechanics and Relativity 1 | 4 |
| Lecture 1: Orders of Magnitude and Dimensional Analysis | 5 |
| Lecture 2: Dimensional Analysis (contd.) and Vectors | 6 |
| Lecture 3: Kinematics Introduction | 9 |
| Lecture 4: Projectile Motion and Relativity Reference Frames | 11 |
| Lecture 5: End of Kinematics and Special Relativity I | 14 |
| Lecture 6: Special Relativity II | 17 |
| Lecture 7: Special Relativity III | 21 |
| Lecture 8: Special Relativity IV and Intro to Dynamics | 25 |
| Lecture 9: Dynamics Exercises | 29 |
| Lecture 10 | 32 |
| Lecture 11 | 33 |
| Lecture 12 | 34 |
| Lecture 13 | 35 |
| Lecture 14 | 36 |
| Lecture 15 | 37 |
| Lecture 16 | 38 |
| Lecture 17: Gravitation | 39 |
| Lecture 18 | 41 |
| Lecture 19 | 42 |
| Lecture 20 | 43 |
| | |
| LC Introduction to Probability and Statistics | 44 |
| Lecture S1: Start of Stats: Introduction and Descriptive Statistics | 45 |
| Lecture S2: Population Statistics | 46 |
| Lecture S3: Error Propogation and Combinations of Variables | 47 |
| Lecture S4: Covariance and Correlation | 49 |
| Lecture S5: Distributions | 50 |
| Lecture S6: Likelihood and Log Likelihood | 54 |
| Lecture S7: Fitting a Straight Line 1 | 60 |
| Lecture S8: Fitting a Straight Line 2 | 63 |
| Lecture S9: Linear Regression | 67 |
| Lecture S10: Goodness of Fit | 72 |
| Lecture S11: End of Stats: Revision | 76 |
| Lecture P1: Start of Probability: Introduction | 77 |
| Lecture P2: Combinatorics and Counting | 79 |
| Lecture P3: Combining Probabilities | 82 |
| Lecture P4: Conditional Probability | 85 |
| Lecture P5: Law of Total Probability and Bayes Theorem | 88 |
| Lecture P6: Ordered Events and Expectation Values | 91 |
| Lecture P7: Discrete Distributions | 94 |

| | |
|---|------------|
| Lecture P8: Multivariate Distributions | 97 |
| Lecture P9: Continuous Probability | 101 |
| Lecture P10: Example Distributions and Central Limit Theorem | 104 |
| Lecture P11: End of Probability: Variance Propagation | 108 |
| LC Mathematics for Physicists 1A | 111 |
| Lecture 1: Course Welcome and Introduction to Vectors | 112 |
| Lecture 2 | 115 |
| Lecture 3 | 116 |
| Lecture 4 | 117 |
| Lecture 5 | 118 |
| Lecture 6 | 119 |
| Lecture 7 | 120 |
| Lecture 8 | 121 |
| Lecture 9 | 122 |
| Lecture 10 | 123 |
| Lecture 11 | 124 |
| Lecture 12 | 125 |
| Lecture 13 | 126 |
| Lecture 14 | 127 |
| Lecture 15 | 128 |
| Lecture 16 | 129 |
| Lecture 17 | 130 |
| Lecture 18 | 131 |
| Lecture 19 | 132 |
| Lecture 20 | 133 |
| Lecture 21 | 134 |
| Lecture 22 | 135 |
| Lecture 23 | 136 |
| Lecture 24 | 137 |
| Lecture 25 | 138 |
| Lecture 26 | 139 |
| Lecture 27 | 140 |
| Lecture 28 | 141 |
| Lecture 29 | 142 |
| Lecture 30 | 143 |
| Lecture 31 | 144 |
| Lecture 32 | 145 |
| Lecture 33 | 146 |
| LC Optics and Waves | 147 |
| Lecture 1: Intro to Waves and SHM Recap | 148 |
| Lecture 2: Wave Functions | 150 |
| Lecture 3: Generalised Wavefunctions | 152 |
| Lecture 4: Waves at Boundaries | 156 |
| Lecture 5: Wave Applications and Introduction to Standing Waves | 162 |
| Lecture 6: Standing Waves 2 Electric Boogaloo | 166 |
| Lecture 7: Energy and Power of Waves | 168 |
| Lecture 8: EM Standing Waves and Lasers | 172 |

| | |
|--|------------|
| Lecture 9: Doppler Shift, Shockwaves and Optics I | 175 |
| Lecture 10: Optics II | 180 |
| Lecture 11: Reflection and Refraction Examples I | 185 |
| Lecture 12: Reflection and Refraction Examples II | 190 |
| Lecture 13: Image Formation | 191 |
| LC Quantum Mechanics 1 | 192 |
| Lecture 1: Atomic Structure | 193 |
| Lecture 2: The Ultraviolet Catastrophe | 199 |
| Lecture 3: Particle Nature of Light | 203 |
| Lecture 4: Atomic Energy Levels and Spectra | 210 |
| Lecture 5: X-Ray Production and Diffraction | 215 |
| Lecture 6: X-Ray Spectra | 219 |
| Lecture 7: “Matter Waves” | 225 |
| Lecture 8: Wave-Particle Duality | 230 |
| Lecture 9: Wavefunctions for Quantum Particles | 235 |
| Lecture 10: A Quantum Mechanical Wave Equation | 240 |
| Lecture 11: Applications of The Schrodinger Equation | 243 |

LC Classical Mechanics and Relativity 1

Tue 30 Sep 2025 12:00

Lecture 1 - Orders of Magnitude and Dimensional Analysis

Given some equation, for example $E = mv^2$, we can decompose it into the basic physical quantities that make it up, for example in terms of Mass, Time and Length. We can denote the dimensions of some quantity by wrapping it in square brackets.

$$E = \frac{1}{2}mv^2$$

$$[E] = ML^2T^{-2}$$

Example:

$$\text{Pressure} \equiv \frac{\text{Force}}{\text{Area}}$$

Suppose we want to test whether pressure and linear momentum flux (amount of linear momentum per unit time, per unit surface) were equivalent quantities, we could do this using dimensional analysis:

$$[P] = \frac{[F]}{[A]}$$

$$[P] = \frac{M \times LT^{-2}}{L^2}$$

$$= M/LT^2$$

And for linear momentum flux ($\Phi(p)$ where lowercase p is momentum):

$$\Phi(p) = \frac{[p]}{[A][\text{time}]}$$

$$= \frac{MLT^{-1}}{L^2T}$$

$$= \frac{M}{LT^2}$$

So yes, they seem to be (at least dimensionally) equivalent.

0.1 Challenging the LHC

We want to use orders of magnitude calculations to challenge the idea that the LHC is the “Big Bang Machine”.

The LHC operates on the order of magnitude of approx 10TeV. The age of the universe is approx 13.7Bn Years, or (in orders of magnitude) 10^{10} yrs.

What time was the big bang? The Big Bang started the universe, but we can't really say it happened at 0s, because that doesn't really make sense. What about 1sec? or 1ms? Well it's clearly less than both of those, so we want to find the smallest possible increment of time “Plank Second” and say it happened after one of them.

Thu 02 Oct 2025 15:00

Lecture 2 - Dimensional Analysis (contd.) and Vectors

0.1 Continuation of Dimensional Analysis

What if, in theory, we could build a system of units entirely from c , the speed of light, G , Newton's constant and h , the Plank Constant?

Cont. from Lec01, we can try to use this to work out the earliest possible cosmic time.

$$\begin{aligned} h &= 6.6 \times 10^{-34} \text{ Js} \\ G &= 6.67 \times 10^{-11} \text{ Nm}^2/\text{kg}^2 \text{ nm} \text{ in stall 20} \\ c &= 3 \times 10^8 \text{ m/s} \end{aligned}$$

Dimensionally:

$$\begin{aligned} [h] &= \frac{ML^2}{T} \\ [G] &= \frac{L^3}{T^2 M} \\ [c] &= \frac{L}{T} \end{aligned}$$

We want to use these to build out a time unit, so:

$$\begin{aligned} [h^u G^v c^z] &= T \\ \left(\frac{ML^2}{T} \right)^u \left(\frac{L^3}{T^2 M} \right)^v \left(\frac{L}{T} \right)^z &= T \\ M^{u-v} L^{2u+3v+z} T^{-u-2v-z} &= T \end{aligned}$$

Solving for:

$$u - v = 0$$

$$2u + 3v + z = 0$$

$$-u - 2v - z = 1$$

Gives us:

$$u = \frac{1}{2} \quad (2.1)$$

$$v = \frac{1}{2} \quad (2.2)$$

$$z = \frac{-5}{2} \quad (2.3)$$

$t_p = \sqrt{\frac{Gh}{c^5}}$ and plugging in the values for G , h , c gives us a value of time, which the earliest possible cosmic time equal to about 10^{-43} s

0.2 Plank Energy

Doing the same process for energy gives us (this time, the plank energy is the energy at which traditional theories of physics break down):

$$E_p = \frac{hc^5}{G}^{0.5} \approx 10^9 J$$

On the other hand, the LHC manages about 10TeV, which is orders of magnitude smaller than this, so the LHC cannot accurately simulate energies of this magnitude.

0.3 More Vectors

Again, vector notation will be \vec{a} . We define the x, y, z unit vectors as $\hat{e}_x, \hat{e}_y, \hat{e}_z$.

We can therefore define any vector as:

$$\vec{a} = a_x \hat{e}_x + a_y \hat{e}_y + a_z \hat{e}_z$$

The length of a vector is again $|\vec{a}|$.

0.4 Vector Multiplication

Given \vec{a} and \vec{b} we can define the dot (scalar) product and the cross (vector) product

$$\vec{a} \cdot \vec{b} = |a||b| \cos \theta$$

Say we want to know the component of a vector along an axis, we can do the following (eg for x):

$$\vec{a} \cdot \hat{e}_x = a_x$$

For the vector product, we can define:

$$\vec{a} \times \vec{b} = |a||b| \sin(\theta) \hat{j}$$

As the vector perpendicular to the plane containing a and b. It is in the direction given by the *right hand rule*, where curling four fingers into a fist, and orienting your fist so these fingers sweep from \vec{a} to \vec{b} , the new vector will point in the direction of an extended thumb. Theta is the angle between a and b, while j is the unit vector in the direction the new vector will point.

0.5 Solar Energy Example

The world yearly energy usage is about 180,000TWh, which is about $5 \times 10^{20} J$ total. Is it (theoretically) possible to get this all from solar energy? We can check using an approximate order of magnitude calculation.

The Sun's total luminosity is $L_\odot = 3.8 \times 10^{26}$. This energy is radiated in a spherically symmetric way (we assume). Therefore the energy per time, per unit surface is (using 1AU for distance):

$$\frac{L_\odot}{4\pi \times (1.5 \times 10^6)^2}$$

Which is approximately (using order of magnitude):

$$\frac{3.8 \times 10^{26} W}{10 \times 10^{22} m^2} \approx \frac{1 kW}{m^2}$$

This is true in ideal conditions, and real energy supply is lower (due to clouds, atmosphere etc).

If we totally covered the earth's surface area ($A_{\text{surface}} \approx \pi R_\odot^2$) which is approximately:

$$A_{\text{surface}} \approx \pi \times (6 \times 10^3 \times 10^3 m)^2 \approx 10^{14} m^2$$

Therefore total energy received is approximately:

$$P = \frac{1kW}{m^2} \times 10^{14} m^2 \approx 10^{17} W$$

And to power the world:

$$E = \frac{5 \times 10^{20} J}{3 \times 10^7 s} \approx 10^{13} W$$

So, it's theoretically possible, if we could cover enough of the world in solar panels and if we could perfectly capture the sun's energy without losing some to sources such as clouds, atmosphere, areas of the ocean we cannot cover in solar panels etc.

Tue 07 Oct 2025 12:00

Lecture 3 - Kinematics Introduction

For kinematics, we'll treat all objects as points and disregard aspects like rotation/the physical size of the body etc.

Given some point, we can define its position as a function of time $\vec{r}(t)$, and velocity as the derivative wrt time of this:

$$\vec{v}(t) = \frac{d\vec{r}}{dt}$$

And acceleration:

$$\vec{a}(t) = \frac{d\vec{v}}{dt} = \frac{d^2\vec{r}}{dt^2}$$

0.1 Position from Unit Vectors

We can define:

$$\vec{r}(t) = r_x(t)\hat{e}_x + r_y(t)\hat{e}_y + r_z(t)\hat{e}_z = \sum_{j=1}^3 r_j(t)\hat{e}_j$$

So:

$$\begin{aligned} \frac{d\vec{r}}{dt} &= \frac{d}{dt} \left(\sum_{j=1}^3 r_j \hat{e}_j \right) \\ &= \sum_j \frac{d}{dt} (r_j \hat{e}_j) \\ &= \sum_j \frac{dr_j}{dt} \hat{e}_j \\ \vec{v} &= \sum_j v_j \hat{e}_j \end{aligned}$$

And:

$$\vec{a} = \frac{d\vec{v}}{dt} = \sum_{j=1}^3 a_j \hat{e}_j$$

Note: Taking the derivative of a vector wrt time is looking at how the variable changes in some infinitesimal time. This can be a change in direction, and/or a change in magnitude. To differentiate a vector we can differentiate it component-wise.

0.2 Cartesian and Polar

Instead of representing a point as x and y components (in 2D), we can instead define it as a distance from the origin r and the angle this distance line forms with the positive x-axis θ .

Therefore (by basic right angle trig) $x = r\cos\theta$, $y = r\sin\theta$, and hence:

$$\vec{r} = r \cos \theta \hat{e}_x + r \sin \theta \hat{e}_y$$

So:

$$\vec{u}(t) = \frac{d\vec{r}}{dt} = \frac{d}{dt}(r \cos \theta) \hat{e}_x + \frac{d}{dt}(r \sin \theta) \hat{e}_y$$

$$\begin{aligned} &= (\dot{r} \cos \theta + r(-\sin \theta)\dot{\theta}) \hat{e}_x + (\dot{r} \sin \theta + r(\cos \theta)\dot{\theta}) \hat{e}_y \\ &= \dot{r} (\cos \theta \hat{e}_x + \sin \theta \hat{e}_y) + r \dot{\theta} (-\sin \theta \hat{e}_x + \cos \theta \hat{e}_y) \end{aligned}$$

0.3 Example

Lets model a particle, in a single dimension moving with constant acceleration (a_0) along a line. What is $x(t)$?

The introduction of τ here was generally poorly understood by the class at the time. Please see Lec 05 for a more thorough explanation.

$$a = \text{constant} = a_0$$

$$a = a_0 = \frac{dv}{dt}$$

So we can simply integrate to get $v(t)$ and again to get $x(t)$.

What if a is not constant? Consider $a(t) = kt^3$. We begin by redefining $a(t)$ as the following, where tau is a time constant representing one time unit. This could be one second, one year etc.

$$a(t) = \tau^3 k \left(\frac{t}{\tau} \right)^3$$

$$\text{let } a_* \equiv k\tau^3$$

$$a(t) \equiv a_* \left(\frac{t}{\tau} \right)^3$$

$$\int dv = \int a_* \tau \left(\frac{t}{\tau} \right)^3 d \frac{t}{\tau}$$

$$v = \frac{1}{4} a_* \tau \left(\frac{t}{\tau} \right)^4 + v_0$$

I've removed the rest here, as the whole τ thing was confusing in this lecture. Again, please see Lec 05, which re-does this section in a better manner.

Thu 09 Oct 2025 15:00

Lecture 4 - Projectile Motion and Relativity Reference Frames

Projectile Motion: The motion of a particle subject to gravitational acceleration, $g \approx 9.81\text{m/s}^2$

1 Projectile Motion

For this to hold, the height of the particle above the ground must be $m << R_e \approx 6 \times 10^3\text{km}$.

$$x(t) = x_0 + v_0(t - t_0) + \frac{1}{2}a_0(t - t_0)^2$$

Lets begin solely by considering motion in the vertical axis (called z here, for some strange reason). This particle is falling from height $h\text{m}$, to the ground at $h = 0$, with constant acceleration gm/s^2 . It has been dropped at time $t = t_0$

At time t_0 , $v = 0, z = h$

$$\begin{aligned} z(t) &= h + 0 - \frac{1}{2}gt^2 \\ \frac{1}{2}gt^2 &= h \\ \implies t &= \sqrt{\frac{2h}{g}} \end{aligned}$$

1.1 What about 2D?

Now we can expand our example to (rather than drop the particle from rest) give the particle some initial velocity $v_0\text{m/s}$ parallel to the ground. We now want two position functions, $x(t)$ and $z(t)$. As previously calculated:

$$z(t) = h + 0 + \frac{1}{2}(-g)t^2$$

And horizontally:

$$x(t) = 0 + v_0(t) + 0$$

So:

$$\begin{cases} z = h - \frac{1}{2}gt^2 \\ x = v_0t \end{cases}$$

Rearranging:

$$\begin{aligned} t &= \frac{x}{v_0} \\ z &= h - \frac{1}{2}g\left(\frac{x}{v_0}\right)^2 \end{aligned}$$

Since h, g, v_0 are all constants, this is an x^2 parabola.

1.2 Interplanetary Example

Lets consider some planet, with $g_{\text{planet}} = 5 \text{ m/s}^2$. You (denoted Y) fall into the atmosphere at some distance h from the ground, and some horizontal distance d from $O(x = 0)$. There is an alien who wants to kill you, by shooting you down. This “gun” can throw pebbles at some constant speed v_0 . The only degree of freedom the alien has to target you is change the shooting angle wrt the horizontal, θ . From the alien’s perspective, what is the required θ to hit the incoming spacecraft?

To hit you, there is some time t , when the position of the bullet B , with initial velocity v where B is in the same position as Y

Consider B

$$\begin{aligned}x_B(t) &= v_0 \cos(\theta)t \\z_B(t) &= v_0 \sin(\theta)t - \frac{1}{2} g_p t^2\end{aligned}$$

Consider Y

$$\begin{aligned}x_Y(t) &= d \\z_Y(t) &= h - \frac{1}{2} g_p t^2\end{aligned}$$

We want to find a θ where $x_B = x_Y$ and $z_B = z_Y$ at the same t :

$$v_0 \cos(\theta)t = d \quad (4.1)$$

$$v_0 \sin(\theta)t - \frac{1}{2} g_p t^2 = h - \frac{1}{2} g_p t^2 \quad (4.2)$$

From 2:

$$\begin{aligned}v_0 \sin(\theta)t &= h \\ \implies t &= \frac{h}{v_0 \sin \theta}\end{aligned}$$

And substituting:

$$\begin{aligned}v_0 \cos(\theta) \left(\frac{h}{v_0 \sin \theta} \right) &= d \\ \frac{\cos(\theta)h}{\sin(\theta)} &= d \\ \frac{\cos \theta}{\sin \theta} &= \frac{d}{h} \\ \tan \theta &= \frac{h}{d}\end{aligned}$$

Since we have the value of θ in terms of two constants, yes, the alien can always hit the spaceship provided it correctly selects the angle corresponding to the value of these two constants (excluding cases where the particle is too far to the left to possibly be hit regardless of angle). This means that the required angle does not depend on velocity, in this example.

2 Frames of Reference

“Observer” represents a frame of reference. The way that one person sees the world (in terms of relative positions and velocities) is different to how another person may see the world. We observe the same core physics, but need to do coordinate translations to go from one reference frame to another.

Say we have two reference frames, A and B . We can represent the translation from A to B as a vector, denoted \vec{r} . Some vector \vec{b}_B in B ’s frame of reference is therefore equal to:

$$\vec{b}_B = \vec{r} + \vec{b}_A$$

Assume that the frames are moving with a constant uniform velocity u with respect to each other:

$$\frac{d}{dt}(\vec{b}_B) = \frac{d}{dt}(\vec{r}) + \frac{d}{dt}(\vec{b}_A)$$

$$\vec{v}_b = \frac{d\vec{r}}{dt} + \vec{v}_r = u + \vec{v}_r$$

This is known as the “Galilean Transformation”.

Tue 14 Oct 2025 12:00

Lecture 5 - End of Kinematics and Special Relativity I

In this lecture:

- Vecchio clarifying things he'd been asked from Kinematics.
- The start of Special Relativity.

1 Use of Tau

This caused quite a bit of confusion for people in Lec 03. We have a particle subject to constant acceleration $\mathbf{a} = a_0 \mathbf{\hat{a}}$.

The displacement of a particle at time t is given by:

$$x(t) = x_0 + v_0(t - t_0) + \frac{1}{2}a_0(t - t_0)^2$$

This is only true for a constant acceleration. More generally, we have:

$$a(t) = \frac{dv}{dt}$$

So we can integrate twice to get $x(t)$. Consider the example where $a(t) = kt^3$. This is non-constant acceleration. We assume that $t_0 = 0$ to simplify things a little. We further assume that $v(t = t_0) = v(t = 0) = 0$ and $x(t = t_0) = x(t = 0) = 0$.

We want to determine $x(t)$.

$$\begin{aligned} \frac{dv}{dt} &= kt^3 \implies dv = kt^3 dt \\ v - v_0 &= \frac{kt^4}{4} \Big|_{t_0}^t \end{aligned}$$

Since we have $v_0 = t_0 = 0$, we have:

$$v(t) = \frac{k}{4}t^4$$

And integrating again:

$$dx = v dt$$

$$x - x_0 = \frac{k}{4} \frac{t^5}{5} \Big|_0^t = \frac{k}{20}t^5$$

Again, $x_0 = 0$ so we finally get:

$$x(t) = \frac{k}{20}t^5$$

Note we have simplified by assuming the initial conditions are all 0, hence we can disregard v_0 etc. If we didn't have this, we'd have to include them in the integration all the way down.

The goal of using τ is to make the problem clearer and easier to understand. Going back to $a(t) = kt^3$, we can tell by dimensions that k must have units of an acceleration divided by a time cubed. This is a messy constant with dimensions then of $[k] = L/T^5$. It is therefore difficult to see what an increase in one time unit actually causes $a(t)$ to do.

We can pick a constant timescale called τ . Tau can be whatever we like, one hour, one millisecond, fifteen years etc etc. We rewrite:

$$a(t) = kt^3 = k \frac{t^2}{\tau^3} \tau^3 = k \tau^3 \left(\frac{t}{\tau}\right)^3$$

We now have a new constant with units of acceleration, $k\tau^3$ which we call a_* .

$$a(t) = a_* \left(\frac{t}{\tau}\right)^3$$

This lets us think about the problem a little more clearly, as we know that after one τ has passed, the object will have acceleration a_* . After two τ s of time have passed, the object will have acceleration $2^3 a_* = 8a_*$ etc. The acceleration now nicely scales in a cubic manner.

Reintegrating with τ gives:

$$\begin{aligned} dv &= adt \\ v - v_0 &= a_* \tau \left(\frac{t}{\tau}\right)^4 \frac{1}{4} \end{aligned}$$

In our case for a particle starting at rest:

$$v(t) = \frac{1}{4} a_* \tau \left(\frac{t}{\tau}\right)^4$$

And for x :

$$\begin{aligned} x - x_0 &= \frac{1}{4} a_* \tau^2 \frac{1}{5} \left(\frac{t}{\tau}\right)^5 \\ x &= \frac{1}{20} a_* \tau^2 \left(\frac{t}{\tau}\right)^5 \end{aligned}$$

The benefit of τ for v and x is a bit less stark, but it's still somewhat present. For constant acceleration, we can write either:

$$\begin{aligned} x(t) &= \frac{1}{2} a_0 t^2 \\ x(t) &= \frac{1}{2} a_0 \tau^2 \left(\frac{t}{\tau}\right)^2 \end{aligned}$$

The distance travelled over some time τ is $\frac{1}{2} a_0 \tau^2$. Note that we can compare this to the derived result for non-constant acceleration, so using τ gives us a more comfortable and familiar form even in the non-constant scenario.

2 Normalisation

This is where we described a particle's position not in terms of unit vectors $\hat{e}_x, \hat{e}_y, \hat{e}_z$ and instead using polar form \hat{e}_θ and \hat{e}_r .

Consider a particle in circular motion (i.e. a child on a merry-go-round). Lets say the child wants to accelerate their motion, we want to keep the distance from the origin constant (or the child would fly off!) while increasing the speed around the circle. Doing this with the former notation would change the coefficients all three unit vectors, while using the latter notation allows us to express it as only a single constant multiplied by the unit vector changing unit vector.

3 Special Relativity

We will cover:

- The Lorentz Factor γ .
- Time Dilation
- Length Contraction

Special relativity is about how two different observers observe the kinematics of objects. For special relativity to hold, these observers cannot be accelerating. They must move with constant velocity with respect to each other. For an observer, we describe an event with four coordinates: (x, y, z, t) , where t is time. For a moving observer, it will see the same event, but at a different set of coordinates (x', y', z', t') .

We note that the speed of light c must be constant and independent of any observer. If two observers measure c in a vacuum, they will both determine the same value regardless of motion. This breaks the standard rules of kinematics that we've seen so far, and it means that time and space are both relative - i.e. one second for one observer may be different to one second for another.

We have these assumptions:

- Two inertial observers will observe the same physics.
- Two inertial observers will observe the same speed of light.

Everything in special relativity scales with the 'Lorentz Factor' in some form, given by:

$$\gamma = \frac{1}{\sqrt{1 - \left(\frac{u}{c}\right)^2}}$$

Gamma is always larger than 1, as $u < c$. We have two key results which we will derive later:

- Moving clocks run slow - a moving observer experiences time slower relative to a static observer.
- Moving objects shrink - a static observer will observe that a moving object has shrunk relative to what the moving observer measures about itself.

Thu 16 Oct 2025 15:00

Lecture 6 - Special Relativity II

1 Special Relativity

We consider theoretical observers that are unaccelerated with respect to each other. Either both observers are at rest, or moving with respect to each other at a constant speed. For ease in CMR1, we only consider motion in one dimension.

We also say that:

- The First Law of Dynamics (Newton's First Law) still holds true, so an object at rest will remain at rest, and an object in constant motion will remain in constant motion, unless an external force acts upon it.
- The distance between two points is constant (relative to an observer).
- We can synchronise clocks between two observers, and they will tick at the same rate.
- We only deal with Euclidean geometry.

We have two key postulates:

1. The laws of physics are the same for every inertial observer.
2. The speed of light in a vacuum is constant for every inertial observer. It is independent of any motion of the source or the observer. Even if a source travelling at $0.5c$ shines a laser facing forward, that light will still travel at c , and not $1.5c$.

1.1 Lorentz Factor

We have some stationary observer A, and a second observer B which is moving at velocity um/s relative to A. The Lorentz Factor γ is defined as:

$$\gamma = \frac{1}{\sqrt{1 - \left(\frac{u}{c}\right)^2}}$$

We may also see it written as:

$$\gamma = \frac{1}{\sqrt{1 - (u/c)^2}}$$

Which holds only if u is already measured in units of the speed of light. For this course, we use the first definition. This is also known as the "Relativistic Factor". Note that it is dimensionless and is a positive number $\gamma > 1$, as $u < c$.

Taking Limits: We take limits of γ to see its behaviour as u changes relative to the speed of light.

If $u \ll c$:

$$\frac{u}{c} \ll 1$$

We use ϵ to denote a very small value. Let $\epsilon \equiv u/c$.

$$\gamma = \frac{1}{\sqrt{1-\epsilon^2}}$$

We expand this using a Taylor Series:

$$f(x) = f(x_0) + (x - x_0)f'(x_0) + \frac{1}{2}(x - x_0)^2 f''(x_0) + \dots$$

$$\gamma = 1 - \left(-\frac{1}{2}\epsilon^2\right) + \dots$$

If u is very fast, say $u = 30\text{km/s}$, then:

$$\epsilon = \frac{3 \times 10 \times 10^3}{3 \times 10^8}$$

$$\epsilon = 10^{-4}$$

Hence:

$$\gamma = 1 + \frac{1}{2}10^{-8} + \dots$$

So even for speeds which are classically extremely fast, $\gamma \approx 1$ and we therefore do not encounter relativistic effects in classical mechanics.

If u/c is 'large', i.e. $u/c \rightarrow 1$:

Now, $1 - u/c$ is small, so we define $\epsilon \equiv 1 - \frac{u}{c} \ll 1$ instead.

$$\frac{u}{c} = 1 - \epsilon$$

$$\gamma = \frac{1}{\sqrt{1 - \left(\frac{u}{c}\right)^2}} = \frac{1}{\sqrt{(1 - \frac{u}{c})(1 + \frac{u}{c})}}$$

$$\gamma = \frac{1}{\sqrt{(\epsilon)(1 + (1 - \epsilon))}}$$

$$\gamma = \frac{1}{\sqrt{2\epsilon - \epsilon^2}}$$

Since $\epsilon \ll 1$, we say that the ϵ^2 term is small enough to disregard, so we have:

$$\gamma = \frac{1}{\sqrt{2}}\epsilon^{-1/2}$$

Again since ϵ is very small, $\epsilon^{1/2}$ tends to infinity, so as $\gamma \propto \epsilon^{-1/2}$, $\gamma \rightarrow \infty$. For non-relativistic objects, we therefore treat $\gamma = 1$, but a curve of γ against u/c has an asymptote at $u/c = 1$, hence γ rapidly increases unbounded as $u \rightarrow c$.

2 Einstein's Thought Experiment

We want to design a clock. We do so by creating a perfect cylinder, with a perfectly reflective top and bottom.

We place a light source (laser) at the bottom, and we shine this laser up towards the top of the cylinder. The photons travel to the top, hit the ceiling, which is perfectly reflective, so travels back down.

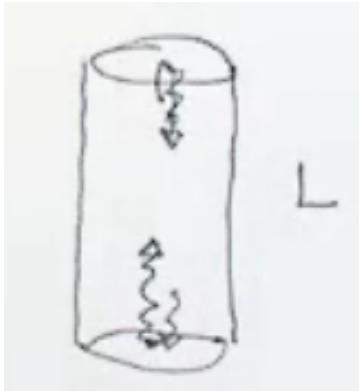


Figure 6.1

We also have a perfect clock, which measures the round-trip time for the photon to go up, hit the ceiling and hit the bottom again. The cylinder is L distance units high, so the total photon path is $2L$ for the round trip. The time taken is therefore:

$$\Delta t = \frac{2L}{c}$$

We add two observers, (B) who is fixed to the top of the cylinder (and is travelling with it). We have some other observer (A) who has designed the problem to place the whole cylinder on a moving trolley, moving in 1D with speed u . Observer A is standing stationary on the ground as the trolley speeds past them.

Observer (B) while moving sees a photon emitted at the bottom, travel up and reflect back down, with no issues.

However, Observer (A) sees the whole setup moving. It sees a photon emitted and travel up, and while it travels up the trolley has moved some distance. The trolley (and photon) have moved some distance when the photon strikes the top and reflects. As the photon travels back down, the trolley (and photon) have moved some distance again.

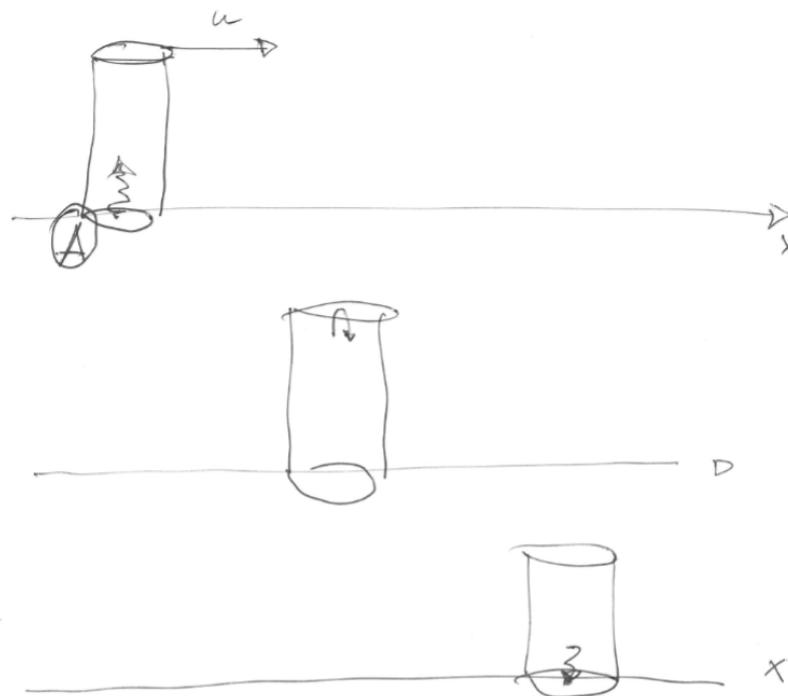


Figure 6.2: The trolley at point of photon emission, point of reflection, and point of detection. Note the photon has moved with the trolley.

We note that the height of the cylinder is not affected by the motion of the trolley, as the motion is perpendicular to this length. We are given this as fact. The only lengths which may be affected are the lengths with components in the direction of motion.

From the perspective of (B), the photon has taken the standard and simple up-down path, in some time Δt . However, from (A)'s perspective, the photon has taken a much longer path which includes horizontal motion:



Figure 6.3

In Observer (A)'s reference frame:

The two (equal) lengths that form the bases of the two right-angled triangles have length $u \frac{\Delta t}{2}$, and the two hypotenuses have length $c \frac{\Delta t}{2}$.

We therefore can simply use Pythagoras to get:

$$\left(u \frac{\Delta t}{2}\right)^2 + L^2 = \left(c \frac{\Delta t}{2}\right)^2$$

In Observer (B)'s reference frame:

$$\Delta t = \frac{2L}{c}$$

We denote the clocks held by (A) to give measurements:

$$\Delta t_B = 2 \frac{L_B}{c_B}$$

and for (A):

$$\left(u \frac{\Delta t_A}{2}\right)^2 + L_A^2 = \left(c_A \frac{\Delta t_A}{2}\right)^2$$

We know that the speed of light is identical in every reference frame, so $c_A = c_B = c$. We have been told that L is unaffected, since it is perpendicular to the direction of motion, so $L_A = L_B = L$.

Hence:

$$\Delta t_B = \frac{2L}{c}$$

Which we can rearrange and substitute to get:

$$\begin{aligned} \left(u \frac{\Delta t_A}{2}\right)^2 + \left(\frac{c \Delta t_B}{2}\right)^2 &= \left(c \frac{\Delta t_A}{2}\right)^2 \\ (c \Delta t_B)^2 &= (c \Delta t_A)^2 - (u \Delta t_A)^2 \end{aligned}$$

For this to be true, $t_B \neq t_A$, so the two observers can no longer agree on the time the photon took. This gives us time dilation, where moving clocks (i.e. the clock used by (A)) run slower, and record a longer time between two events compared to a stationary observer.

Tue 21 Oct 2025 12:00

Lecture 7 - Special Relativity III

1 Time Dilation

We concluded the previous lecture with:

$$(c\Delta t_B)^2 = (c\Delta t_A)^2 - (u\Delta t_A)^2$$

We rearrange to get:

$$\begin{aligned} c^2 \Delta t_B^2 &= (c^2 - u^2) \Delta t_A^2 \\ \Delta t_B^2 &= \left(1 - \left(\frac{u}{c}\right)^2\right) \Delta t_A^2 \\ \Delta t_B &= \sqrt{1 - \left(\frac{u}{c}\right)^2} \Delta t_A \end{aligned}$$

Hence:

$$\begin{aligned} \Delta t_A &= \frac{1}{\sqrt{1 - \left(\frac{u}{c}\right)^2}} \Delta t_B \\ \boxed{\Delta t_A = \gamma \Delta t_B} \end{aligned}$$

This tells us that the time recorded for the photon to travel by the two observers is different. The moving clock runs slower, so the stationary observer measures a longer duration than the moving clock records. For non-relativistic speeds, $\gamma \approx 1$ so the difference is negligible; however, at larger speeds the disparity becomes much larger and grows without bounds. This makes physical sense, as for faster speeds, the trolley will have travelled a larger horizontal distance, therefore (A) will measure a longer path, and hence require a larger time.

Generally, we have:

$$\Delta T = \gamma \Delta t_0$$

Where t_0 is the “proper time” and is defined as the time interval taken between two events that take place in the same frame, by an observer in that frame.

2 Length Contraction

We have the same identical setup, except the setup is now horizontally on the trolley. Observer (B) is again attached to the cylinder, with photons again bouncing along the length of the cylinder, just with the left/right instead of top/bottom surfaces. The cylinder is still moving on a trolley, and observer (A) is still stationary.



Figure 7.1

We define Δt as the interval between emission and detection, again with a subscript to denote who is making the measurement.

$$\Delta t_B = \frac{2L_B}{c}$$

In (B)'s frame of reference:

$$c\Delta t_B = 2L_B$$

In (A)'s frame of reference:

The cylinder has moved to the right as the photon travels, this adds some extra length that the photon must travel.

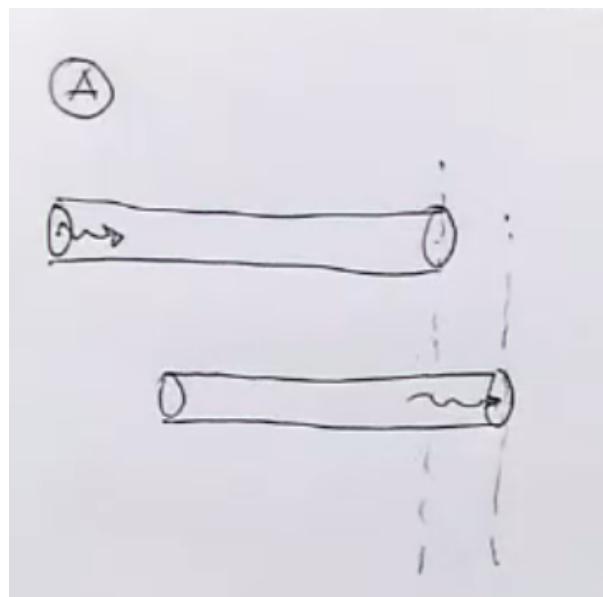


Figure 7.2

This extra length (between the two dotted lines) is $u\Delta t_1$ where Δt_1 is the time for the photon to hit the right wall. The photon now hits the wall and bounces back, while the cylinder is still moving to the right. The cylinder will move $u\Delta t_2$, where Δt_2 is the time taken for the photon to travel back and hit the left wall.

The time taken between emission and detection Δt is given by:

$$\Delta t = \Delta t_1 + \Delta t_2$$

For the first part of the trip, the distance is $L_A + u\Delta t_1$, so:

$$\begin{aligned} c\Delta t_1 &= L_A + u\Delta t_1 \\ \Delta t_1 &= \frac{L_A}{c-u} \end{aligned}$$

For the second part of the trip, the distance is less as the cylinder “catches up” with the photon as it moves, giving us a distance of $L_A - u\Delta t_2$. This gives us:

$$\begin{aligned} c\Delta t_2 &= L_A - u\Delta t_2 \\ \Delta t_2 &= \frac{L_A}{c+u} \end{aligned}$$

Hence the round-trip time is:

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

So:

$$\Delta t_A = \frac{2L_A}{c} \left(\frac{1}{1-\frac{u^2}{c^2}} \right)$$

And we know that:

$$\Delta t_B = \frac{2L_B}{c}$$

The fact A and B don't agree is fine, we can apply the time dilation formula:

$$\begin{aligned} \Delta t_A &= \gamma \Delta t_B \\ \frac{2L_A}{c} \left(\frac{1}{1-\frac{u^2}{c^2}} \right) &= \frac{1}{\sqrt{1-(\frac{u^2}{c^2})}} \frac{2L_B}{c} \\ L_A \left(\frac{1}{1-\frac{u^2}{c^2}} \right) &= \frac{1}{\sqrt{1-(\frac{u^2}{c^2})}} L_B \\ L_A \left(\frac{\sqrt{1}}{\sqrt{1-\frac{u^2}{c^2}}} \right)^2 &= \frac{1}{\sqrt{1-(\frac{u^2}{c^2})}} L_B \\ L_A \left(\frac{1}{\sqrt{1-\frac{u^2}{c^2}}} \right) &= L_B \\ L_A &= \frac{L_B}{\gamma} \end{aligned}$$

This tells us that measurements of lengths for a stationary vs moving observer also do not agree. Just like time runs slower for a moving object, a moving object will be measured to be smaller by a stationary observer. Effectively, moving lengths shrink. This is called length contraction.

Generally, we have:

$$\boxed{\Delta L = \frac{\Delta L_0}{\gamma}}$$

Where ΔL_0 is the “proper length” of an object, i.e. the length of an object measured by an observer at rest relative to it.

3 Example

We have someone on a spaceship, coming back to earth. The spaceship is travelling rightwards, directly towards the earth in 1D.

The spaceship has an astronaut, and we consider one person on the earth in Mission Control. MC spots the spaceship at distance $l = 3,000\text{km}$ and is at speed u corresponding to $\gamma = 10$ (so $u \approx c$). The earth is stationary.

When the spaceship is at this distance l , the spaceship sends earth a distress signal saying $\Delta t_r = 10^{-3}\text{s}$, where Δt_r is how long the spaceship's oxygen supply lasts. We assume the astronaut cannot hold their breath and dies immediately if the oxygen supply runs out before the ship makes it back to Earth. Will the astronaut survive?

MC's child, who has no knowledge of relativity, calculates the travel time:

$$\Delta t = \frac{l}{u} \approx \frac{l}{c} = \frac{3 \times 10^3 \times 10^3}{3 \times 10^8} = 10^{-2}\text{s}.$$

The child compares this to the oxygen supply (10^{-3}s) and concludes the astronaut dies. This isn't quite accurate however, as we need to treat the oxygen time relativistically. To MC, the clock on the spaceship runs slow (Time Dilation):

$$\Delta t_{MC} = \gamma \Delta t_0$$

$$\Delta t_{MC} = 10 \times 10^{-3}\text{s} = 10^{-2}\text{s}$$

Comparing the relativistic oxygen duration (10^{-2}s) to the travel time (10^{-2}s), we see that the astronaut (just barely!!) survives.

We solved this in the Earth frame, where the distance l is a proper length. We could alternatively solve this in the Astronaut's frame, where the distance to Earth is length contracted to $L = l/\gamma = 300\text{km}$. In that frame, the travel time is 10^{-3}s , which matches the proper time of the oxygen supply, so gives the same result.

Thu 23 Oct 2025 15:00

Lecture 8 - Special Relativity IV and Intro to Dynamics

1 The Relativistic Doppler Effect

For an emitted frequency f_0 , emitted by an object moving with velocity (in 1D) u relative to an observer, the received frequency f is given by:

$$f = \sqrt{\frac{1+u/c}{1-u/c}} f_0$$

For a relativistic speed u . Note that we cannot use the standard Doppler formula for relativistic speeds. Also note the lack of \pm , as we encode this into u . If the object moves towards the observer, u is positive, and if the object is moving away u is negative.

If u is non-relativistic, we assume that $u/c \ll 1$:

$$f = \left(1 + \frac{u}{c}\right)^{1/2} \left(1 - \frac{u}{c}\right)^{-1/2} f_0$$

And Taylor Series expanding:

$$= \left(1 + \frac{1}{2} \frac{u}{c} + \dots\right) \left(1 - \frac{-1}{2} \frac{u}{c} + \dots\right) f_0$$

Since u/c is small, we ignore any quadratic, cubic etc terms of u/c , as these are very small.

$$= \left(1 + \frac{1}{2} \frac{u}{c} + \frac{1}{2} \frac{u}{c}\right) f_0$$

Hence:

$$f \approx \left(1 + \frac{u}{c}\right) f_0$$

Since we assume $u/c \ll 1$:

$$\begin{aligned} f &= f_0 + \frac{u}{c} f_0 \\ f - f_0 &= \Delta f = \frac{u}{c} f_0 \\ \frac{\Delta f}{f_0} &= \frac{u}{c} \quad \text{plus higher order terms we ignore} \end{aligned}$$

This is the classical result that we're familiar with, for non-relativistic speeds.

2 Lorentz Transformation

Say we have a reference frame s' which is moving along the x -direction relative to a static reference frame s .

An event in the s frame has coordinates (x, y, z, t) and the same event in the s' frame has coordinates (x', y', z', t') . We have the following transformations:

$$t' = \gamma \left(t - \frac{u}{c^2} x \right)$$

$$x' = \gamma(x - ut)$$

Noting that $y = y'$, and $z = z'$ as these are orthogonal to the direction of motion. We also have:

$$u'_x = \frac{u_x - u}{1 - \frac{u}{c^2}u_x}$$

Please note that CMR1 does not include derivations of these equations (which collectively form the Lorentz Transformations), however for understanding's sake I'll include them here regardless.

2.1 Derivations

We want a transformation in time and space between a stationary frame S and the moving frame S' . We have three postulates to do this:

- **Linearity:** The transformation must be linear, i.e. a straight line in S must map to a straight line in S' .
- **Standard rule for c :** c is invariant and has the same velocity in all frames, regardless of motion.
- **Inverse symmetry:** The inverse transformation ($S' \rightarrow S$) is the same, but with $u \rightarrow -u$, as in S' 's reference frame, it is static with S moving with speed $-u$.

Deriving x' : We derive the transformation in position using length contraction:

Imagine a ruler at rest in the moving frame S' . It has one end on the origin O' and the right end at some coordinate x' . The ruler therefore has proper length $L_0 = x'$.

From the stationary frame S :

- The origin O' has moved distance ut after some time t .
- The ruler is moving, so has been length contracted to the observer in S . The ruler now appears to have length x'/γ .

The total coordinate as seen by S is therefore:

$$x = ut + \frac{x'}{\gamma}$$

And rearranging gives:

$$x' = \gamma(x - ut)$$

Deriving t' : We derive the transformation for time using the third postulate above:

If the transformation from $S \rightarrow S'$ is:

$$x' = \gamma(x - ut)$$

Then the transformation from $S' \rightarrow S$ is:

$$x = \gamma(x' + ut')$$

As the scenario is the same: S' is moving with speed u relative to S , if we instead consider S' 's reference frame then it is static, and S is moving in the opposite direction with the same magnitude of velocity, so the transformation must be the same with $u \rightarrow -u$.

Substituting the position transformation for x' into this:

$$x = \gamma[\gamma(x - ut) + ut']$$

$$\frac{x}{\gamma} = \gamma x - \gamma ut + ut'$$

Rearranging for t' :

$$t' = \gamma t - \frac{x}{u} \left(\gamma - \frac{1}{\gamma} \right)$$

Using the identity $\gamma - 1/\gamma = \beta^2\gamma = \frac{u^2}{c^2}\gamma$:

$$t' = \gamma t - \frac{x}{u} \left(\gamma \frac{u^2}{c^2} \right)$$

$$t' = \gamma \left(t - \frac{u}{c^2} x \right)$$

Which gives us the time transformation.

Deiving u'_x : We find velocity in the moving frame as dx'/dt' .

The velocity in the moving frame S' is defined as $u'_x = \frac{dx'}{dt'}$. Taking these derivatives

$$dx' = \gamma(dx - u dt) \quad \text{and} \quad dt' = \gamma \left(dt - \frac{u}{c^2} dx \right)$$

Substituting these into the definition of velocity:

$$u'_x = \frac{\gamma(dx - u dt)}{\gamma \left(dt - \frac{u}{c^2} dx \right)} = \frac{dx - u dt}{dt - \frac{u}{c^2} dx}$$

Dividing through by dt :

$$u'_x = \frac{\frac{dx}{dt} - u}{1 - \frac{u}{c^2} \frac{dx}{dt}}$$

And using $u_x = \frac{dx}{dt}$:

$$u'_x = \frac{u_x - u}{1 - \frac{uu_x}{c^2}}$$

3 Dynamics

Kinematics is effectively looking at objects in motion. Dynamics is effectively “why” they move (Newton’s laws, static friction etc).

3.1 Newton’s Laws

We have three:

1. If there is no resultant force acting upon an object, there are two possibilities:
 - The object was initially at rest, and stays at rest.
 - The object moves at a constant speed with no acceleration.
2. $\vec{F} = m\vec{a}$. Force and acceleration are proportional with a constant of proportionality m , the “inertial mass”. We call it inertial mass because this is theoretically distinct from gravitational mass, however all experiments give them as having the same value.
3. The Reaction Principle. If a body A is producing a force \vec{F} on a body B , then B acts back upon A with a force of the same magnitude but the opposite direction “Every action has an equal and opposite reaction”.

3.2 Superposition Principle

If we have some body with N forces acting upon it, the final resultant force that acts upon an object is a vector sum of these forces:

$$\vec{F} = \vec{F}_1 + \vec{F}_2 + \cdots + \vec{F}_n = \sum_{i=1}^N \vec{F}_i = m\vec{a}$$

3.3 Example

Consider an object of mass m hanging from the ceiling with an “ideal string”. This means that string is inextensible and is massless. The body is initially at rest.

Two forces act upon this body:

- The weight force due to local gravitational acceleration: $w = mg$.
- The force produced by the string (tension).

Since the body is at rest, the resultant force must be zero and:

$$T - mg = 0$$

$$T = mg$$

3.4 Example II

Consider a body on a horizontal surface. We laterally pull the object with force F . We have these forces:

- Again a weight force: $w = mg$.
- The normal force produced by the table acting back upon the body iaw Newton's Third Law.
- A frictional force acting in opposition to the direction of motion, F_{fric}

The frictional force is proportional to the Normal force with a coefficient depending on the materials used:

$$F_{\text{fric}} = \mu N$$

When moving an object there are two stages:

- Attempting to take an object from stationary to actually moving.
- Continuing the motion of the object once it's moving (this is easier).

We therefore have multiple coefficients of friction. Here we consider the coefficient of static friction μ_s and the coefficient of kinetic friction μ_k . There is also the coefficient of rolling friction μ_r seen in labs. Generally, $\mu_s > \mu_k$.

Tue 28 Oct 2025 12:00

Lecture 9 - Dynamics Exercises

1 Exercise I

We have a (very heavy) book on a table. We pull that book at some angle θ from the horizontal with constant force \vec{F} . The book has $m = 10\text{kg}$ and there is some coefficient of **kinetic** friction μ .

What is the minimum \vec{F} so that the book moves with constant velocity?

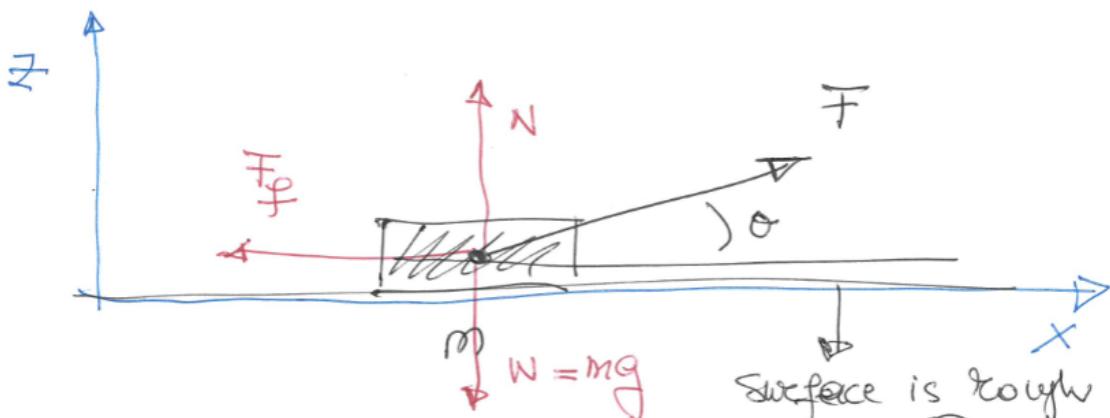


Figure 9.1

We have the following forces acting on the object.

Vertical (in z):

- The Normal reaction force N acting perpendicular to the table (directly up).
- The weight force $w = mg$ acting perpendicular to the table (directly down).
- A vertical component of \vec{F} given by $F \sin \theta$.

Horizontal (in x):

- The frictional force $-\mu N$.
- A horizontal component of the pulling force \vec{F} given by $F \cos \theta$.

The velocity is constant, so $dv/dt = a$. Hence $v = \int a dt$. For v to be constant, a must be zero. Hence there is no resultant force. Consider the total forces with $F = ma$:

$$\text{For } T_x : -\mu N + F \cos \theta = ma_x = 0$$

$$F \cos \theta - \mu N = 0$$

$$\text{For } T_z : N - mg + F \sin \theta = 0$$

Therefore solving for N :

$$\mu N = F \cos \theta \implies N = \frac{F \cos \theta}{\mu}$$

And substituting into z:

$$\begin{aligned} \frac{F \cos \theta}{\mu} - mg + F \sin \theta &= 0 \\ \implies F(\cos \theta + \mu \sin \theta) - \mu mg &= 0 \\ \implies F &= \frac{\mu mg}{\cos \theta + \mu \sin \theta} \end{aligned}$$

We can now ask "what is the right choice of θ to minimise the required F ?". We take the derivative of F wrt θ so that we can solve for a minima.

$$\frac{dF(\theta)}{d\theta} = 0$$

This will give us stationary points, and we can classify the stat points to ensure we find a minima. Alternatively, we can save a little bit of faff and do this quicker by recognising that the numerator is a constant. The minimum of F is therefore the maximum of $\cos \theta + \mu \sin \theta$.

$$\begin{aligned} \frac{d}{d\theta} (\cos \theta + \mu \sin \theta) &= 0 \\ -\sin \theta + \mu \cos \theta &= 0 \\ \mu = \frac{\sin \theta}{\cos \theta} &= \tan \theta \implies \theta = \arctan \mu \end{aligned}$$

So the optimum angle depends on the coefficient of friction.

2 Exercise II

We have an inclined plane at angle θ . There is a pulley at the top of plane and one block with mass m_1 on the plane, connected by an ideal string over the pulley to another block. This block has mass m_2 and is hanging off the plane. The plane is frictionless.

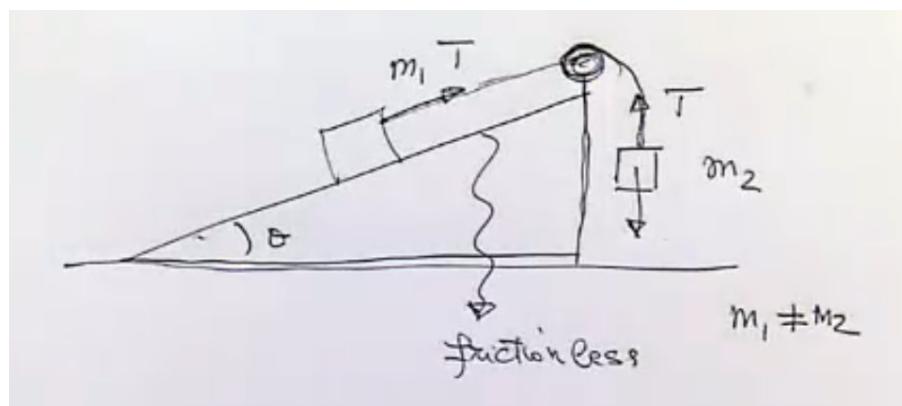


Figure 9.2

What is θ such that the two masses move together with constant speed?

Both objects will move together, either up or down (depending on which mass is greater).

For m_2 :

- The weight $w = m_2 g$ acting downwards.
- The tension force T acting upwards.

For m_2 :

- The normal force acting perpendicular to the plane.
- The tension force T acting in the direction of motion.
- The weight force $w = m_1 g$ acting immediately down.

We ignore the normal force and the component of the weight acting perpendicular to the plane, as these are not in the direction of motion (and we have no frictional force that relies on them).

Therefore (taking the vertical axis positive down):

$$m_2 g - T = 0$$

And (taking the horizontal axis positive right parallel to the direction of motion, and the vertical axis now positive up perpendicular to the direction of motion):

$$T - m_1 g \sin \theta = 0 \implies T = m_1 g \sin \theta$$

Substituting the second into the first:

$$m_2 g - m_1 g \sin \theta = 0$$

$$m_2 - m_1 \sin \theta = 0$$

$$\sin \theta = \frac{m_2}{m_1} \implies \theta = \arcsin \frac{m_2}{m_1}$$

Thu 30 Oct 2025 15:00

Lecture 10

Tue 04 Nov 2025 12:00

Lecture 11

Thu 06 Nov 2025 15:00

Lecture 12

Tue 11 Nov 2025 12:00

Lecture 13

Thu 13 Nov 2025 15:00

Lecture 14

Tue 18 Nov 2025 12:00

Lecture 15

Thu 20 Nov 2025 15:00

Lecture 16

Tue 25 Nov 2025 12:00

Lecture 17 - Gravitation

1 Gravitation

Consider two particles, with masses m_1 and m_2 at distance \vec{r} . We know there is an attractive force between the two masses:

$$\vec{F} = -G \frac{m_1 m_2}{r^2} \hat{e}_r \quad (17.1)$$

$$G = 6.67 \times 10^{-11} \text{ m}^3/\text{kg s}^2$$

1.1 Key Properties

- “Long Range” force. A force will exist between any two masses anywhere in the universe, regardless of distance and cannot be cancelled. Negligible at large distances, as force quickly tends to zero as distance increases - but never zero.
- Weak.
- $\vec{F} \propto \frac{1}{r^2}$

1.2 Mass Caveats

While we don't practically make a distinction between them, m in gravity refers to ‘gravitational mass’, m_g , while mass in Newton's Second Law $\vec{F} = m\vec{a}$ is ‘inertial mass’, m_i .

Einstein's Equivalence Principle says that they're equal, i.e:

$$\frac{dm_g}{dm_i} = 1$$

Why ‘they could be different but they fundamentally are not’ is important is beyond me.

2 Freefall

Consider an object of mass m at height h from the ground. It has force:

$$F = mg$$

And potential energy:

$$\Delta U(h) = mgh$$

However this first expression looks quite different to our definition of force.

2.1 Derivation

Gravitational force is conservative, and there is a potential energy at all points in the gravitational field, associated with this force.

Give two bodies, M and m . m is moved away from M from r_1 to r_2 along the radial direction (preserving angle, i.e. moving only in \hat{e}_r).

$$\vec{F} = -G \frac{Mm}{r^2} \hat{e}_r$$

This is the negative gradient of the potential (as potential is area under a force curve):

$$\vec{F} = -\nabla U = -\frac{dU}{dr} \hat{e}_r$$

$$\vec{F} dr = -dU \hat{e}_r$$

The work done from r_1 to r_2 is:

$$w_{1 \rightarrow 2} = \int_{r_1}^{r_2} \vec{F} \cdot \hat{e}_r dx$$

Thu 27 Nov 2025 15:00

Lecture 18

Tue 02 Dec 2025 12:00

Lecture 19

Thu 04 Dec 2025 15:00

Lecture 20

LC Introduction to Probability and Statistics

Wed 01 Oct 2025 12:00

Lecture S1 - Start of Stats: Introduction and Descriptive Statistics

0.1 Course Welcome

- First half of the semester: Statistics
- Second half of the semester: Probability
- All slides and notes on Canvas.

Why Descriptive Statistics? If we want to share an interesting bit of data, sharing the whole data is going to be confusing. Instead, we can share a small number of stats which describe and summarise the data.

0.2 Sample Statistics

One of the most simple is the number of samples (N), and the sample mean:

$$\text{Sample Mean: } \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

We can also calculate the sample standard deviation as the average of mean squared error across the points in the sample:

$$\text{Sample STDev: } s_n^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

We can also use median or mode as measures of central tendency. The mode is a poor estimator however (as it massively depends on how binning is done, for a continuous measurement), while the median is more resistant to outliers.

Thu 02 Oct 2025 09:00

Lecture S2 - Population Statistics

0.1 Accuracy and Precision

We usually take measurements to determine some kind of true value. Usually, we can't actually know what this true value is, but if we could there are two bits of terminology that is particularly important:

Accuracy: Accuracy is the 'closeness' between our value and the 'true' value.

Precision: Precision is the 'closeness' between our measurements, i.e. how spread out are our various measurements.

0.2 Error

Random Error: is uncertainty related to the fact that our measurements are only a finite sample, so is not going to be immediately representative of the true value. The smaller this error, the more precise the measurement is.

Systematic Error: is related to some kind of issue with the measurement or the equipment. This shifts all values, and negatively affects accuracy (but leaves precision unchanged)

Taking many repeat measurements decreases the effects of random error, but the effects of systematic error are much harder to combat.

Ideally, we want to be both precise and accurate, however accuracy is arguably more important. This is because a value which is precise, but not accurate may lead to false conclusions around the inaccurate value.

Wed 08 Oct 2025 12:02

Lecture S3 - Error Propogation and Combinations of Variables

Office Hours: 11:00 to 13:00 Thursdays, Physics West Rm 122

1 Types of Error

Broadly two types of error: Statistical/Random Error (resulting from low precision) and Systematic Error (from Low Accuracy).

Random error widens the distribution, while systematic error shifts the whole distribution up or down, meaning no matter how many repeats you take and how precise you think you are, the value is still nonsense as all datapoints have been equally shifted (i.e. by a poor experimental setup).

For example, you are trying to measure the length of an object using a ruler that has been unknowingly stretched. You cannot get a true value no matter the number of repeats or degree of precision.

1.1 Accuracy vs Precision

High accuracy is preferable to high precision - having high precision but low accuracy can lead to false conclusions (as an incorrect value appears confidently correct). Accuracy is more difficult to improve - precision can be improved by gathering more data, while higher accuracy can only be improved by a better experimental design.

2 Error Propagation

If we take a distribution, and add a constant value to all points, the distribution is shifted up/down without changing the variance.

$$\langle x + k \rangle = \langle x \rangle + k$$

$$Var(x + k) = Var(x)$$

If we multiply by a constant value, the mean is multiplied by this value, but the distribution becomes stretched and the variance grows:

$$\langle xk \rangle = k\langle x \rangle$$

$$Var(kx) = k^2 Var(x)$$

Or taking the natural log:

$$\langle \ln x \rangle \approx \ln \langle x \rangle$$

$$Var(\ln x) \approx \frac{Var(x)}{x^2}$$

As this is a non-linear operator, these become good approximations rather than strict rules of equivalence.

And another example:

$$\langle e^x \rangle \approx e^{\langle x \rangle}$$

$$Var(e^x) \approx (e^{\langle x \rangle})^2 Var(x)$$

Note here, even though our underlying distribution is Normal and symmetric, the new distribution after e^x is neither, and these are an even worse approximation than before.

2.1 Combining Operators

We can apply some linear transformation $mx + c$, we can chain these rules together by doing the multiplicative transformation m first, then the linear scale c.

$$\langle mx + c \rangle = m\langle x \rangle + c$$

$$Var(mx + c) = m^2 Var(x)$$

2.2 Multiple Variables

What if we have multiple distributed variables we want to add?

$$\langle A + B \rangle = \langle A \rangle + \langle B \rangle$$

$$Var(A + B) = Var(A) + Var(B)$$

And multiplying them (again this are now approximations)?

$$\langle AB \rangle \approx \langle A \rangle \langle B \rangle$$

$$Var(AB) \approx \langle B \rangle^2 Var(A) + \langle A \rangle^2 Var(B)$$

$$\frac{Var(AB)}{\langle AB \rangle^2} \approx \frac{Var(A)}{\langle A \rangle^2} + \frac{Var(B)}{\langle B \rangle^2}$$

Or division?

$$\left\langle \frac{A}{B} \right\rangle \approx \frac{\langle A \rangle}{\langle B \rangle}$$

$$Var\left(\frac{A}{B}\right) = \frac{Var(A)}{\langle B \rangle^2} + \frac{Var(B)}{\langle A \rangle^2}$$

3 One Rule to Rule Them All

This single rule allows us to propagate error in any situation, assuming the two variables are uncorrelated:

$$Var(f) \approx \left(\frac{\partial f}{\partial A} \Big|_{A=\langle A \rangle, B=\langle B \rangle} \right)^2 Var(A) + \left(\frac{\partial f}{\partial B} \Big|_{A=\langle A \rangle, B=\langle B \rangle} \right)^2 Var(B)$$

Thu 09 Oct 2025 09:00

Lecture S4 - Covariance and Correlation

Office Hours: Thursday 11am - 1pm, Physics West Rm 222 (b.becsy@bham.ac.uk)

Previously, when looking at two or more variables for error propagation/combinations etc, we assumed that they were independent of one another. Today we look at how to handle multiple variables which may be correlated.

1 Covariance

Covariance is a measure that indicates how much two variables fluctuate together:

$$\text{Cov}(x, y) = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$$

Covariance matrices represent all combinations of covariance (noting $\text{Cov}(x, y) = \text{Cov}(y, x)$ and $\text{Cov}(x, y) = \text{Var}(x)$)

$$\Sigma = \begin{pmatrix} \text{Cov}(x, x) & \text{Cov}(x, y) \\ \text{Cov}(y, x) & \text{Cov}(y, y) \end{pmatrix}$$

We can then define correlation:

$$\text{Corr}(x, y) = \frac{\text{Cov}(x, y)}{\sqrt{\text{Var}(x)\text{Var}(y)}} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}}$$

This is bounded between -1 ($x = -y$), 1 ($x = y$) and zero for no correlation. We can again put this in a matrix, noting it is symmetrical:

$$\begin{pmatrix} 1 & \text{Corr}(x, y) \\ \text{Corr}(y, x) & 1 \end{pmatrix}$$

1.1 Variable Combinations

Now, with correlated variables, we can say:

$$\langle x + y \rangle = \langle x \rangle + \langle y \rangle$$

$$\text{Var}(x, y) = \text{Var}(x) + \text{Var}(y) + 2\text{Cov}(x, y)$$

And (noting the mean slightly increases with correlated variables):

$$\langle xy \rangle = \langle x \rangle \langle y \rangle + \text{Cov}(x, y)$$

And the one formula to rule them all, taking correlation into account:

$$\text{Var}(f) \approx \frac{\partial f}{\partial A}^2 \text{Var}(A) + \frac{\partial f}{\partial B}^2 \text{Var}(B) + 2 \frac{\partial f}{\partial A} \frac{\partial f}{\partial B} \text{Cov}(A, B)$$

Wed 15 Oct 2025 12:00

Lecture S5 - Distributions

1 Coin Flips and Probability Recap

Flipping a coin is one of the simplest distributions we can create.

Given a:

$$\begin{aligned} P(H) &= 0.5 \\ P(T) &= 0.5 \end{aligned}$$

We know that $P(HHHH) = 0.5^4$.

And $P(\text{Three Heads and One Tail}) = P(\text{HHHT or HHTH or THHH or HTTH}) = 4 \times 0.5^4$. We will take 4 coins, A, B, C, D. We denote a single result as A_{Heads} or C_{Tails} etc.

We can also say that the coins are independent, i.e. the probability of one result given another result is equal to just the probability of the first result:

$$P(A_h | B_h) = 0.5$$

The chance of A and B being heads is:

$$P(A_h \text{ and } B_h) = P(A_h) \times P(B_h) = P(HH)$$

The chance of A or B being heads is (noting *or* excludes the case where both are true):

$$P(A_h \text{ or } B_h) = P(A_h) + P(B_h) - P(A_h \text{ and } B_h)$$

1.1 Discrete Distribution

Lets consider flipping 4 coins and counting the number of heads. This forms a discrete distribution (where only 5 possible values are possible, 0, 1, 2, 3, 4). This distribution must be normalised (sum to 1), so:

$$\sum_r P(r) = 1$$

We can also consider the mean (expected) number of heads:

$$\langle r \rangle = \sum_r r P(r)$$

This function, $P(x)$ is called a *probability mass function*, and the sum of all values must be 1.

1.2 Continuous Distributions

Continuous distributions have similar conditions:

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

$$\langle x \rangle = \int_{-\infty}^{\infty} x P(x) dx$$

And for the probability of the result lying between a and b:

$$\int_a^b P(x) dx$$

We cannot, in a continuous distribution consider the probability of an exact result, i.e. $P(x = a)$, $a \in \mathbb{R}$. As there are infinitely many possible values, the probability of any precise one is not meaningful (always zero). We therefore must always consider the probability of the result lying in some non-zero range.

$P(x)$ in this case is called a *probability density function* and the area under the PDF curve must sum to one. Note that this means that $P(x)$ at any point may exceed one, so long as the overall area is equal to 1. For some probability $a < P(x) < b$ (noting that since the $P(a)$ for any precise a is zero, the equalities being strict or not is meaningless), the probability is the area of the curve between a and b.

2 Binomial Distribution

The Binomial Distribution represents a scenario where we conduct some number of identical trials, where each trial has two possible outcomes (which we denote success and failure). For example, flipping a coin. Here:

- n - The number of trials.
- p - The probability of success.
- q - The probability of failure ($q = 1 - p$).
- r - The number of successes.

This has probability mass function:

$$P(r; n, p) = \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r}$$

And has the following properties:

$$\langle r \rangle = np$$

$$Var(r) = np(1-p)$$

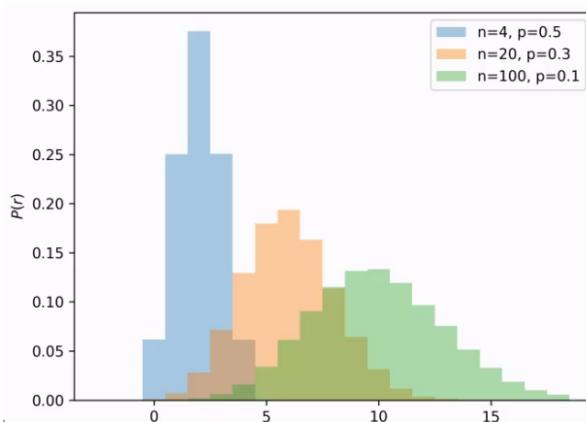


Figure 25.1: The Binomial Distribution

3 Poisson Distribution

This describes the number of events (i.e. the number of neutrinos detected by a neutrino detector) occurring in some time interval, given:

- The mean rate of events is constant.
- Each event occurs independently from the last.

This is created by taking the limit of a Binomial distribution, as:

- The number of trials tends to infinity ($n \rightarrow \infty$)
- The mean number of successes remains fixed ($np = \lambda = \text{constant}$)

Given λ as the mean number of expected events (per unit time) and r as the number of events occurring in that time, it has PMF:

$$P(r; \lambda) = \frac{\exp(-\lambda)\lambda^r}{r!}$$

And has the following properties:

$$\langle r \rangle = \lambda$$

$$\text{Var}(r) = \lambda$$

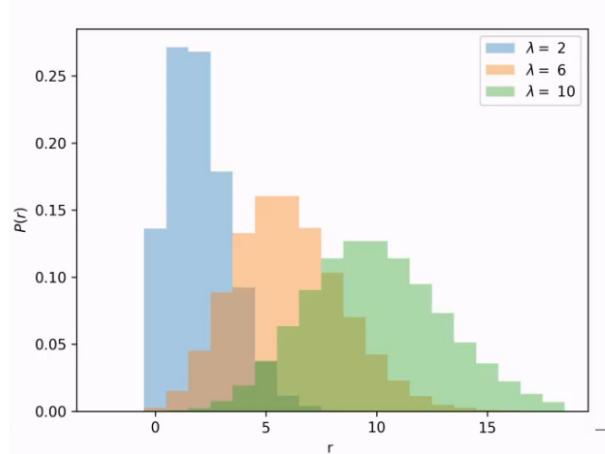


Figure 25.2: The Poisson Distribution

4 Normal Distribution

A.K.A. The Gaussian distribution. This is the most well known and most useful distribution. Given a mean μ and a standard deviation σ , the probability density function is:

$$P(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$

It is a very important distribution as it arises as a result of the Central Limit Theorem, which we will cover properly in the probability section of the course. It looks like this, noting it is symmetric and forms a “bell-shaped curve”:

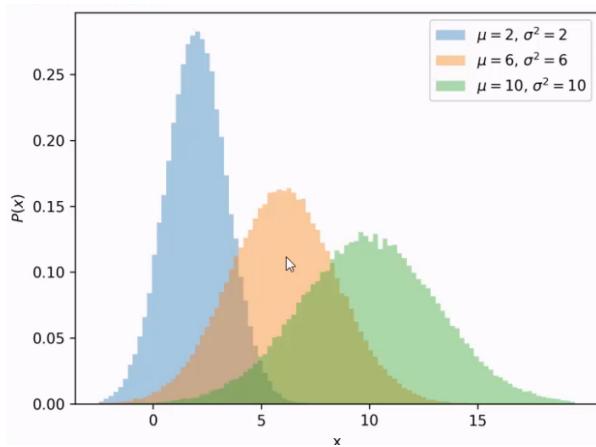


Figure 25.3: The Normal Distribution

Note that since the tails are logarithmic, they tend to zero, but never reach it truly. A Poisson distribution approaches a Normal distribution as $\lambda \rightarrow \infty$. It is generally a good approximation for $\lambda > 30$ but this depends on the application being used.

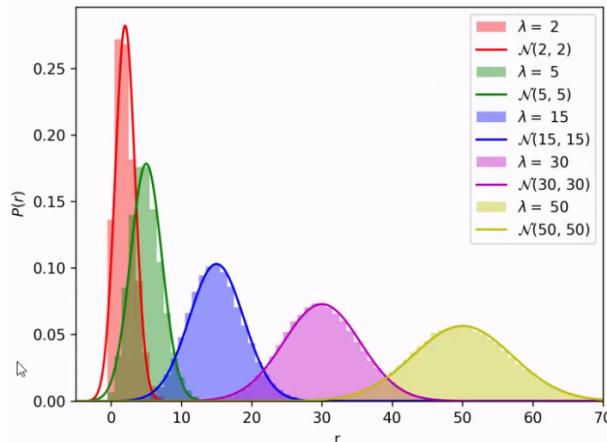


Figure 25.4: Poisson approximations to a Normal

Thu 16 Oct 2025 09:00

Lecture S6 - Likelihood and Log Likelihood

1 Likelihood

We want to fit a model to our data. We want some kind of function to specify how well this model fits the data, so that we can optimise to find the best. This is the likelihood function. There are many different ways to formulate it, but we denote it:

$$P(D | \theta)$$

Where D is our data, and θ is our model parameters. This is the probability of the data, given some parameters.

2 An Example

Lets say we have this model:

$$T(t) = T_{\text{env}} + (T_0 - T_{\text{env}}) \exp(-t/\tau)$$

Which represents the cooling of an object, where τ is a constant of cooling. We think we will observe some additive, normally distributed noise on these measurements, giving us:

$$T_{\text{obs}}(t) = T(t) + \epsilon$$

We may observe something like this, where the blue dots are the model-predicted values and the observed data with error noise is in black:

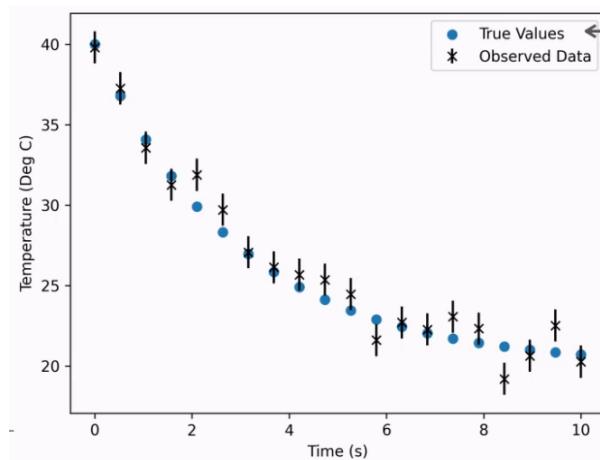


Figure 26.1: Simulated Data

Given the noise is normally distributed, we would expect the true values to lie within the error bars of our observation about 68% of the time. We see this approximately here. How can we then fit a model to this data? We need to:

1. Formulate a model.
2. Estimate a probability that the model is correct.

Given we now have data, and some model we would like to try to fit the data to (we want to fit it to Newton's Law of Cooling, the model previously, and determine an appropriate value of parameters and τ). We therefore want to find a 'merit function' to describe how good a fit any model we might create is. We start from the probability of getting some value of the noise.

As a reminder, our model is, noting we are treating time as a discrete set of times, indexed by i :

$$M(t_i, \theta) = T_{\text{env}} + (T_0 - T_{\text{env}}) \exp(-t_i/\tau)$$

And the probability of getting some value of the noise on the i th measurement is (note the first equality, where we can also write it ignoring theta, because noise is independent of the parameters):

$$P(\epsilon_i | \theta) = P(\epsilon_i) = \frac{1}{\sigma_{D_i} \sqrt{2\pi}} \exp\left(\frac{-\epsilon_i^2}{2\sigma_{D_i}^2}\right)$$

This is a Normal distribution with a mean of zero, and a standard deviation of σ_{D_i}

We cannot directly measure ϵ_i , but we know it is the difference between the measured value in the data and the 'true' value predicted by our model:

$$\epsilon_i = D_i - M(t_i, \theta)$$

Since the noise is additive and Normal, we can combine these two equations to get our merit function - the probability of a single observed data point given the parameters as:

$$P(D_i | \theta) = \frac{1}{\sigma_{D_i} \sqrt{2\pi}} \exp\left(\frac{-(D_i - M(t_i, \theta))^2}{2\sigma_{D_i}^2}\right)$$

Where θ is our list of parameters $\theta = [T_0, T_{\text{env}}, \tau]$. Crucially, this is a Normal distribution where the mean is our model's prediction given the parameters, and the standard deviation is the uncertainty on the error point. Assuming we have multiple uncorrelated data points, the total likelihood function is:

$$P(D | \theta) = \prod_{i=1}^n P(D_i | \theta)$$

Considering τ as the variable we actually change, we get:

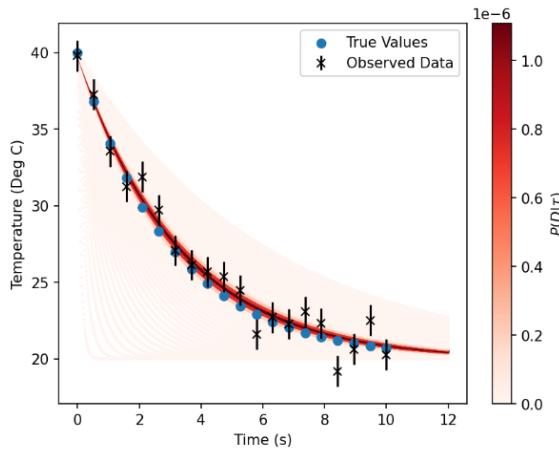


Figure 26.2: A heatmap of the likelihood function overlaid on the data.

The high likelihood values of $P(D | \tau)$ are the models which are most likely to generate the observed data, given the parameters. This, therefore, means that they are the models which best fit the data.

If we plot $P(D, \tau)$ against τ , we can see that the likelihood does a reasonable job of giving us a value which is close to true:

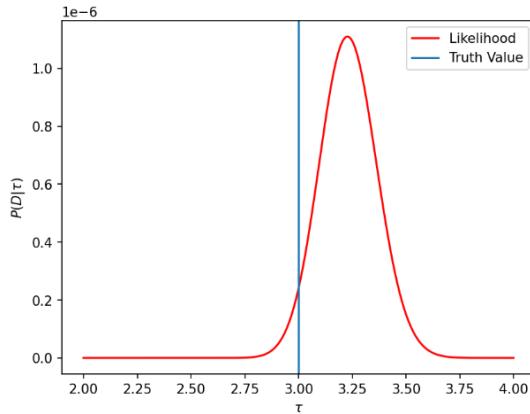


Figure 26.3

We can use Maximum Likelihood Estimation (note that it is just an estimate, the value of tau given by the maximum likelihood and the true value are **not** the same) to estimate the best value of tau for the model. We chose the value of tau that gives the maximum likelihood:

$$\hat{\tau} = \arg \max_{\tau} P(D | \tau)$$

In general, given a set of multiple parameters, we say:

$$\hat{\theta} = \arg \max_{\theta} P(D | \theta)$$

3 Log Likelihood

We still need to estimate the uncertainty on this predicted best value of tau. It turns out that a good way to do this is by taking the log likelihood instead of just the likelihood. We take the natural log of the normal probability density function for a single data point:

$$\begin{aligned} P(D_i | \theta) &= \frac{1}{\sigma_{D_i} \sqrt{2\pi}} \exp\left(\frac{-(D_i - M(t_i, \theta))^2}{2\sigma_{D_i}^2}\right) \\ \ln P(D_i | \theta) &= \ln\left(\frac{1}{\sigma_{D_i} \sqrt{2\pi}} \exp\left(\frac{-(D_i - M(t_i, \theta))^2}{2\sigma_{D_i}^2}\right)\right) \\ \ln P(D_i | \theta) &= \ln\left(\frac{1}{\sigma_{D_i} \sqrt{2\pi}}\right) + \ln\left(\exp\left(\frac{-(D_i - M(t_i, \theta))^2}{2\sigma_{D_i}^2}\right)\right) \\ \ln P(D_i | \theta) &= \left(\frac{-(D_i - M(t_i, \theta))^2}{2\sigma_{D_i}^2}\right) - \ln(\sigma_{D_i}) - \ln(\sqrt{2\pi}) \end{aligned}$$

If we ignore the constant term, as it does not change the results (as we care about the results comparative to each other to find the maximum), and if we assume that the uncertainty is the same on each data point (which we must be careful about, in case uncertainties are variables too):

$$\ln P(D_i | \theta) = \left(\frac{-(D_i - M(t_i, \theta))^2}{2\sigma_{D_i}^2}\right) + \text{constant}$$

This gives the final log likelihood as:

$$\mathcal{L} = \ln P(D | \theta) = \sum_{i=1}^n \ln P(D_i | \theta) = \sum_{i=1}^n \left(\frac{-(D_i - M(t_i, \theta))^2}{2\sigma_{D_i}^2}\right)$$

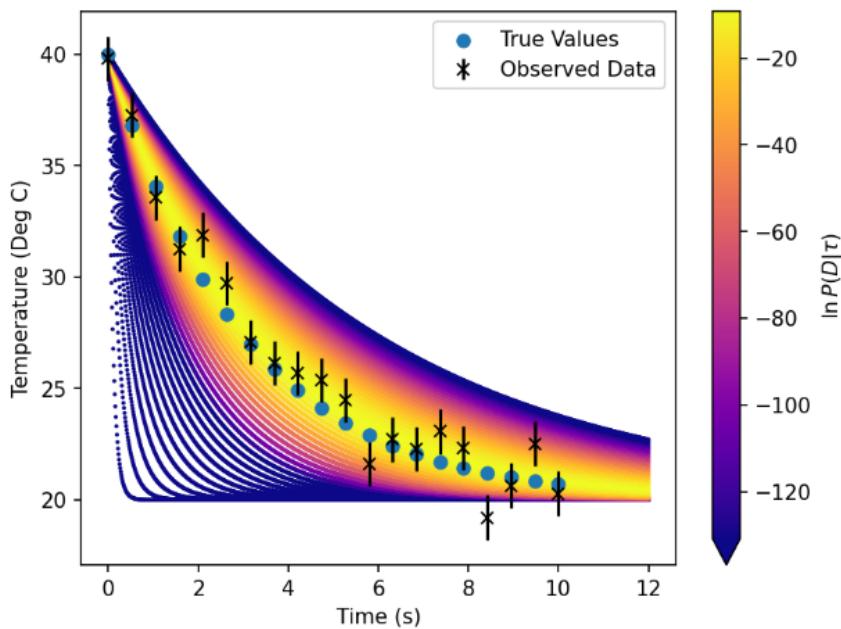


Figure 26.4: The likelihood plot, repeated with log likelihood

We can see this generates similar values to the standard likelihood, but with much friendlier values (-20 to -120, rather than very small numbers). The equation is also nicer to calculate as we're able to get rid of the constant terms. If we again consider this as a function of tau (the parameter we're actually changing to produce a fit):

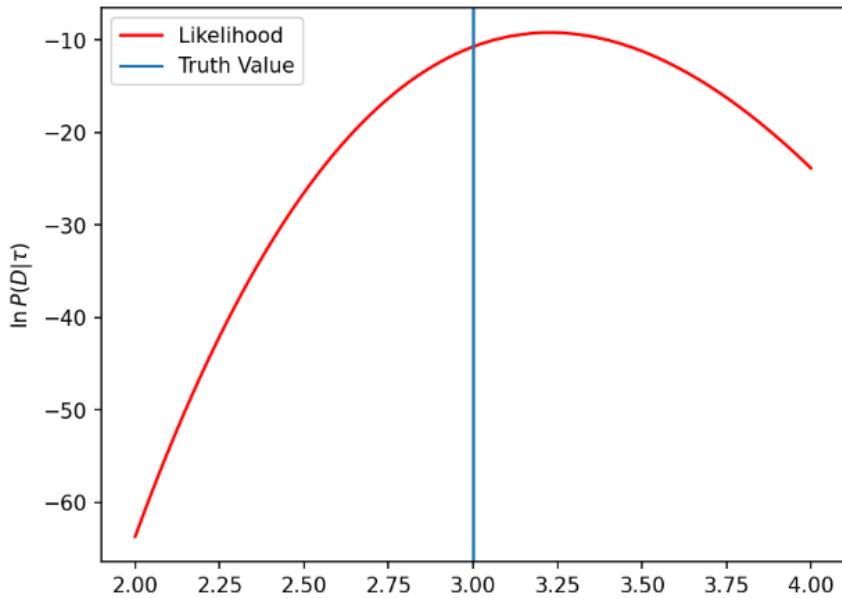


Figure 26.5

This is similar again, but with a very different order of magnitude. We can calculate the predicted optimum τ by determining the maximum point where:

$$\frac{\partial}{\partial \tau} \ln P(D | \tau) = 0$$

Note the swap from theta to tau, this is because tau is the parameter we're actually using to fit, while the other parameters bundled into theta are constant. This gives $\tau \approx 3.23$. Yes we could have done this with

traditional likelihood, but the differentiation is nicer in log form and it becomes relevant later.

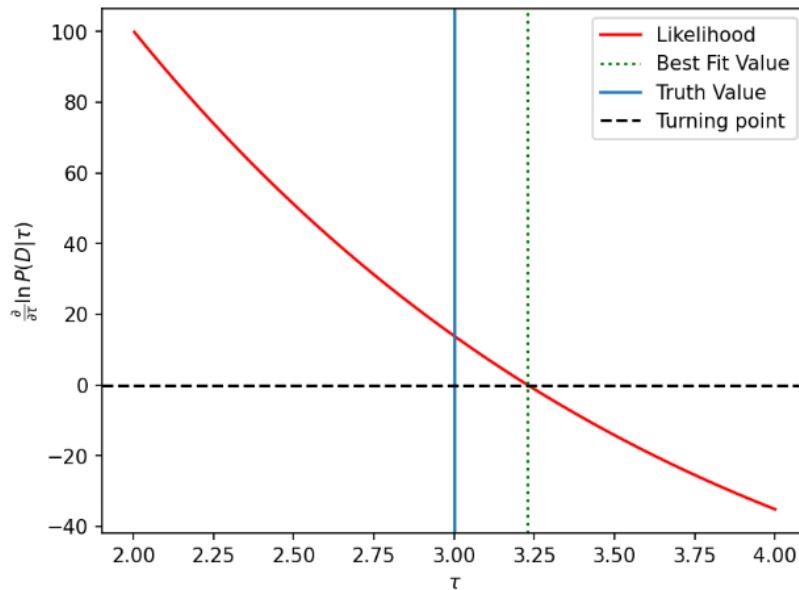


Figure 26.6

4 Uncertainties on Log Likelihood

In our single dimension problem, we'll create a new value called the Hessian (H), we define this as:

$$H = \left. \frac{\partial^2 \mathcal{L}}{\partial \tau^2} \right|_{\tau=\hat{\tau}}$$

Why is this (and the log likelihood) actually useful? It turns out that the curvature of the log likelihood around the maximum point tells us the uncertainty. We can use the Hessian and log likelihood to estimate the uncertainty on the parameter τ , using the inverse of H , H^{-1} where $HH^{-1} = 1$.

$$\sigma_{\hat{\tau}} \approx \sqrt{|H^{-1}|}$$

Since we're in 1D, this value of the Hessian is a scalar, and we can estimate it by either differentiating twice or estimating using two points:

$$f''(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - 2f(x) + f(x-h)}{h^2}$$

And calculating:

$$\sigma_{\hat{\tau}} = \sqrt{\frac{1}{|H|}}$$

But note that this works differently in higher dimensions (more parameters) has H becomes a matrix. This gives us $\hat{\tau} \approx 3.23$ from the previous result, and now $\sigma_{\hat{\tau}} \approx 0.133$. Note that the true value is 1.8 standard deviations away from the best estimate, which is okay - anything larger than 3 sigma away would start to become worrying.

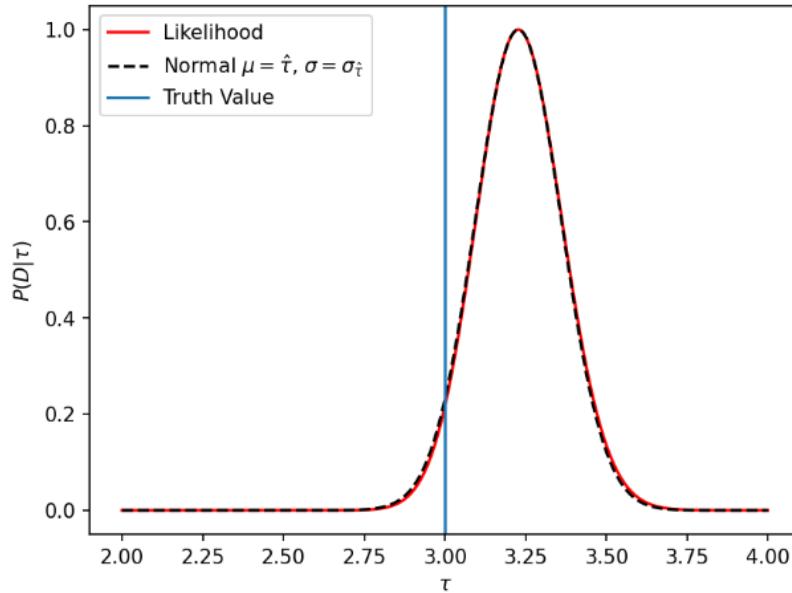


Figure 26.7

Using the log likelihood has:

- Been a nicer calculation, which is computationally easier, as we can strip out all the constant terms.
- Allowed us to calculate the uncertainties on our predicted values.
- Allowed us to get back to the standard likelihood (as the black and red curves above are approximately the same) anyways.

4.1 Multidimensional Generalisation

This works in 1D, but we can generalise to a higher number of parameters. We assume here however that $P(D | \theta)$ is going to be a normal distribution. If this is not the case, we cannot use the approximations for uncertainties, and the problem becomes too complex for first year stats. It is generally the case that it will approximate a normal as many datapoints are taken. If this isn't true, it becomes a problem for Y4 Bayesian Stats.

We find the maximum likelihood (given multiple parameters $\hat{\theta}$) with:

$$\frac{\partial \mathcal{L}}{\partial \theta_1} = \frac{\partial \mathcal{L}}{\partial \theta_2} = \dots = 0$$

And the Hessian is given by:

$$\mathbf{H}(\hat{\theta}) = \begin{bmatrix} \frac{\partial^2 \mathcal{L}}{\partial \theta_1^2} \Big|_{\theta=\hat{\theta}} & \frac{\partial^2 \mathcal{L}}{\partial \theta_1 \partial \theta_2} \Big|_{\theta=\hat{\theta}} & \dots & \frac{\partial^2 \mathcal{L}}{\partial \theta_1 \partial \theta_k} \Big|_{\theta=\hat{\theta}} \\ \frac{\partial^2 \mathcal{L}}{\partial \theta_2 \partial \theta_1} \Big|_{\theta=\hat{\theta}} & \frac{\partial^2 \mathcal{L}}{\partial \theta_2^2} \Big|_{\theta=\hat{\theta}} & \dots & \frac{\partial^2 \mathcal{L}}{\partial \theta_2 \partial \theta_k} \Big|_{\theta=\hat{\theta}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \mathcal{L}}{\partial \theta_k \partial \theta_1} \Big|_{\theta=\hat{\theta}} & \frac{\partial^2 \mathcal{L}}{\partial \theta_k \partial \theta_2} \Big|_{\theta=\hat{\theta}} & \dots & \frac{\partial^2 \mathcal{L}}{\partial \theta_k^2} \Big|_{\theta=\hat{\theta}} \end{bmatrix}.$$

Wed 22 Oct 2025 12:00

Lecture S7 - Fitting a Straight Line 1

We want to create a model for a straight line:

$$M(x, \theta) = mx + c$$

Where datapoints are given by this model and some additive noise:

$$D = M(x, \theta) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma_i)$$

The general recipe for line fitting is given by:

1. A generative model for the data, with knowledge of how the noise is distributed.
2. Likelihood function.
3. A method for finding the maximum likelihood.
4. Method for finding the uncertainties on best fit parameters.
5. A method for checking how good the fit is.

We can write down the likelihood function for this model as:

$$\begin{aligned} P(D_i | \theta) &= \frac{1}{\sigma_{D_i} \sqrt{2\pi}} \exp\left(\frac{-(D_i - M(x_i, \theta))^2}{2\sigma_{D_i}^2}\right) \\ P(D | \theta) &= \prod_{i=1}^n P(D_i | \theta) \end{aligned}$$

And again:

$$\mathcal{L} = \ln P(D | \theta) = \sum_{i=1}^n \ln P(D_i | \theta) \propto \sum_{i=1}^n \left(\frac{-(D_i - M(x_i, \theta))^2}{2\sigma_{D_i}^2} \right)$$

We want to find the parameters of distribution that maximise the (log)likelihood:

$$\hat{\theta} = \arg \max_{\theta} P(D | \theta)$$

Or:

$$\hat{\theta} = \arg \max_{\theta} \mathcal{L}$$

There are a number of different approaches to do this:

- Find where all first derivatives equal zero (as last lecture, various clever algorithms to do so).
- Brute force on a grid.
- Iterative or stochastic methods.
- Analytic maximisation for a simple linear model - see next lecture.

1 Finding Maximum Likelihood

The most crude way to do this is to build a grid of all values of m and c , and iterate through over all points (with some resolution) to find the maximum likelihood generated by them. Plotting M (as colour):

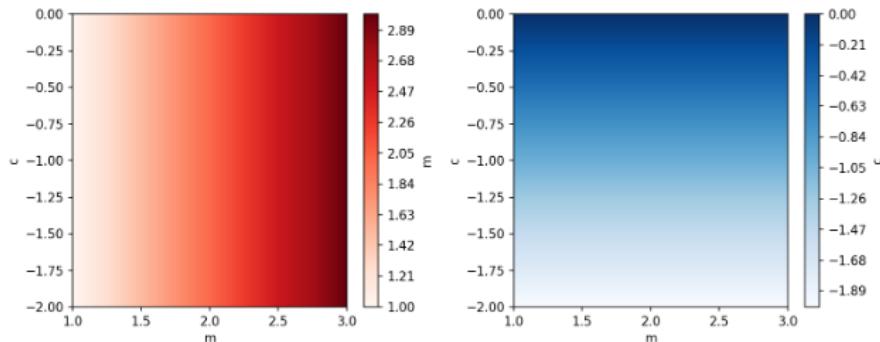


Figure 27.1

And the likelihood:

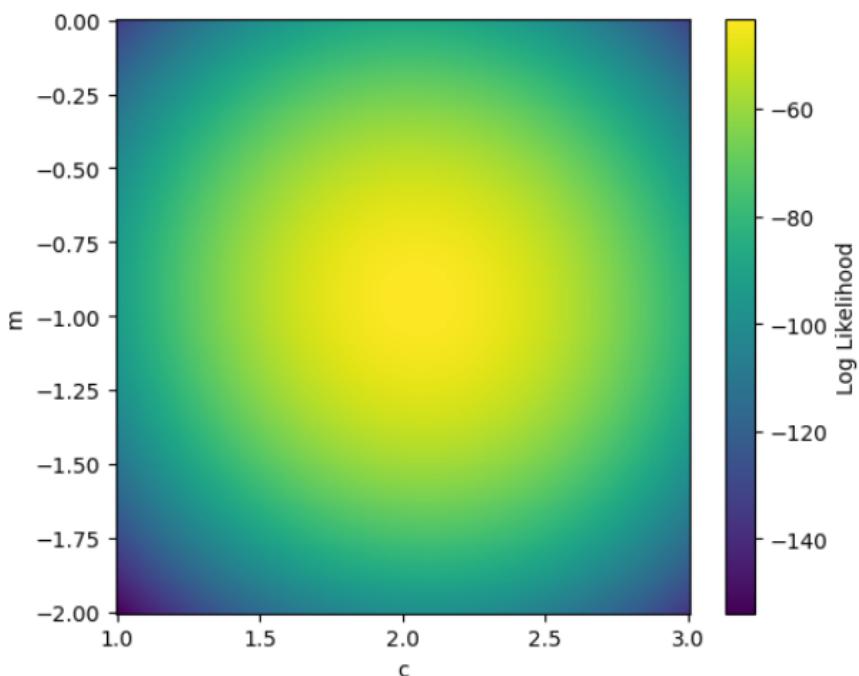


Figure 27.2

We assume that the grid point with the highest likelihood and the true point with the highest likelihood are the same. In this case, the grid resolution is small enough that this is true, but it may not always be. This gives:

$$m = 2.0552763819095476$$

$$c = -0.9447236180904524$$

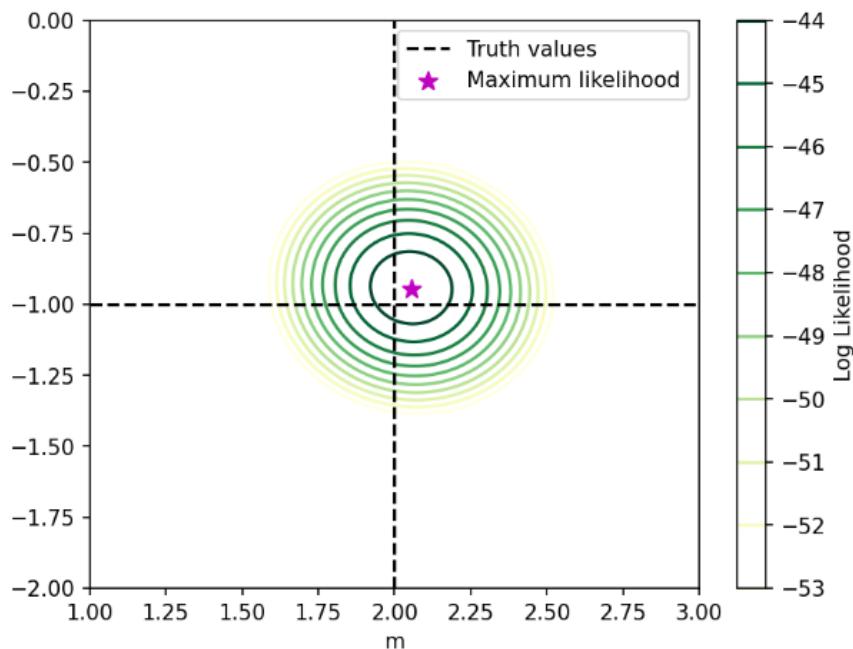


Figure 27.3

We see that this does a good, but not perfect job, of fitting the data. This is due to noise in the data, and is acceptable, provided it's within a reasonable uncertainty.

Thu 23 Oct 2025 09:00

Lecture S8 - Fitting a Straight Line 2

1 Uncertainties on Best Fit Parameters

1.1 What do these uncertainties actually mean?

Previously, we found that the likelihood of a single value could be described by a Normal distribution.

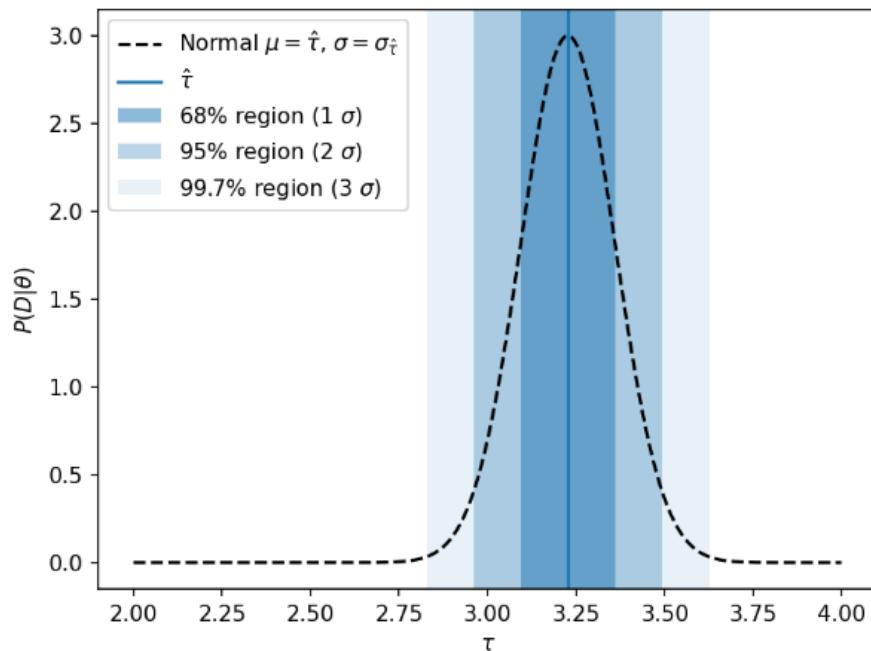


Figure 28.1

We can therefore quote the best fit value as $\hat{\tau} = 3.227 \pm 0.133$. What we're effectively saying is that our 1 sigma uncertainty encompasses 68% of the probability density such that:

$$\int_{\hat{\tau}-\sigma_{\hat{\tau}}}^{\hat{\tau}+\sigma_{\hat{\tau}}} P(D | \tau) d\tau \approx 0.68$$

And the same for 2 sigma uncertainty with 0.95, and 3 sigma uncertainty with 0.997. We can therefore say that while it may be common for the value to lie outside the 1 sigma uncertainty, it is rare for it to lie outside the 3 sigma uncertainty and if this happens (including error), something probably went wrong with our measurement.

Here, where we're attempting to determine an uncertainty, we instead increase the value of $\sigma_{\hat{\tau}}$ until this first integral is satisfied.

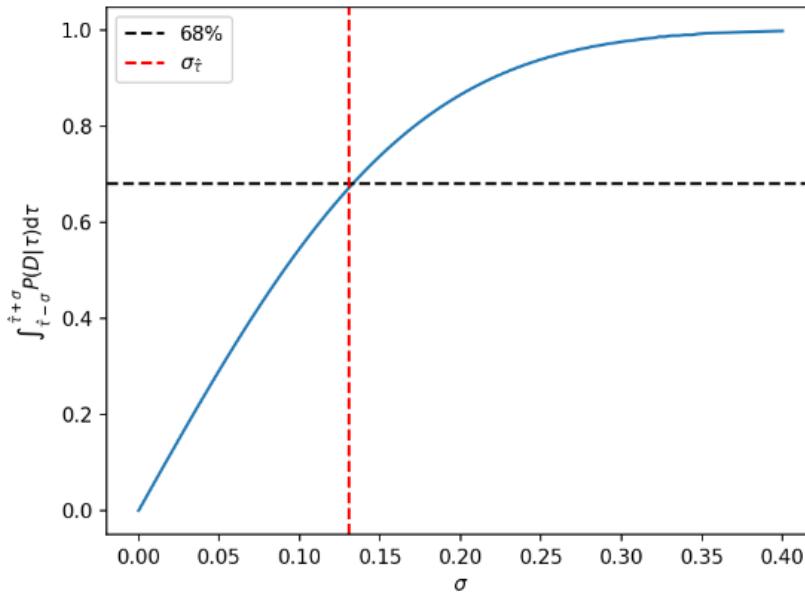


Figure 28.2

However, with a line of best fit, we want to consider uncertainties in two dimensions. This makes life a little bit more difficult, we want some boundary on the parameter space that is centred on the best fit parameters and encompasses 68% of the whole probability space.

Plotting the same distribution again, but including the 1D likelihood for each parameter individually:

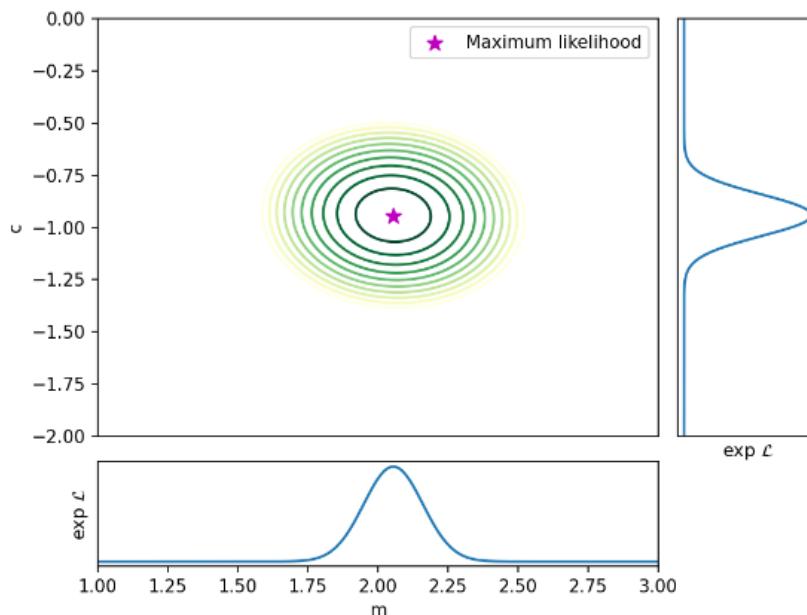


Figure 28.3

In general, we shouldn't just separately calculate the 1 dimensional approach per parameter and combine them (as the parameters may be correlated), but it's a useful starting point. Swapping to log likelihood:

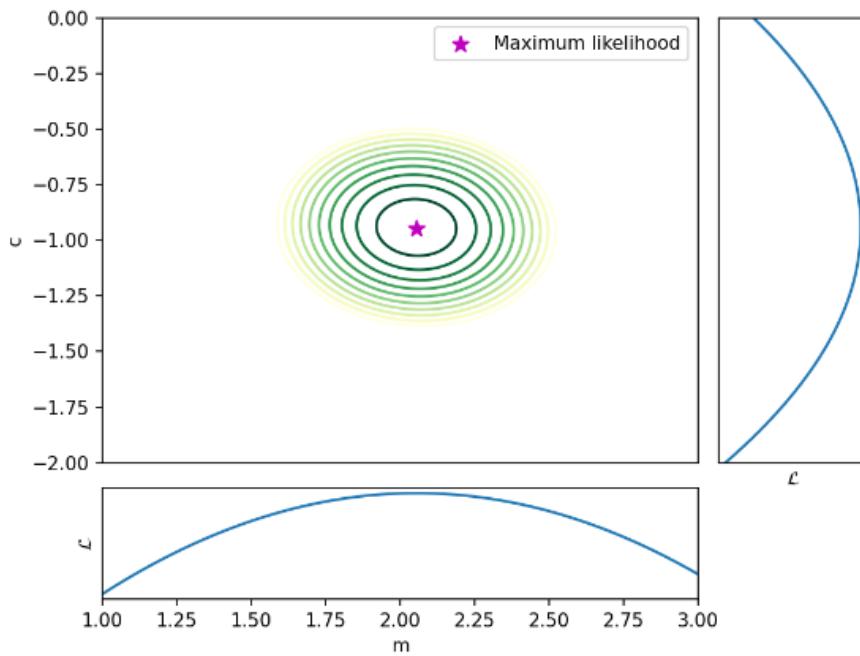


Figure 28.4

We want to estimate:

$$\frac{\partial^2 \mathcal{L}}{\partial m^2} \Big|_{\theta=\hat{\theta}}$$

$$\frac{\partial^2 \mathcal{L}}{\partial c^2} \Big|_{\theta=\hat{\theta}}$$

While we shouldn't use these 1D distributions to estimate the best fit parameters, this problem has deliberately been created to minimise the correlation between m and c and creating something with them that describes $P(D | \theta)$ is still instructive. We can however create a grid of the values of m and c , calculate the likelihood for all points with some resolution and pick the maximum. This is slow, but crucially does find the maximum point of the 3D distribution surface and is not the same as finding the max for each variable individually.

We assume that $P(D | \theta)$ can be described by a two dimensional normal distribution, and we can build this from two discrete normal distributions of two independent variables (with the caveats above). We say that the mean value of $P(D | m)$ is \hat{m} and the uncertainty is therefore given by:

$$\sigma_m^2 = \left(-\frac{\partial^2 \mathcal{L}}{\partial m^2} \Big|_{\theta=\hat{\theta}} \right)^{-1}$$

And likewise for c , \hat{c} , σ_c^2

$$\sigma_c^2 = \left(-\frac{\partial^2 \mathcal{L}}{\partial c^2} \Big|_{\theta=\hat{\theta}} \right)^{-1}$$

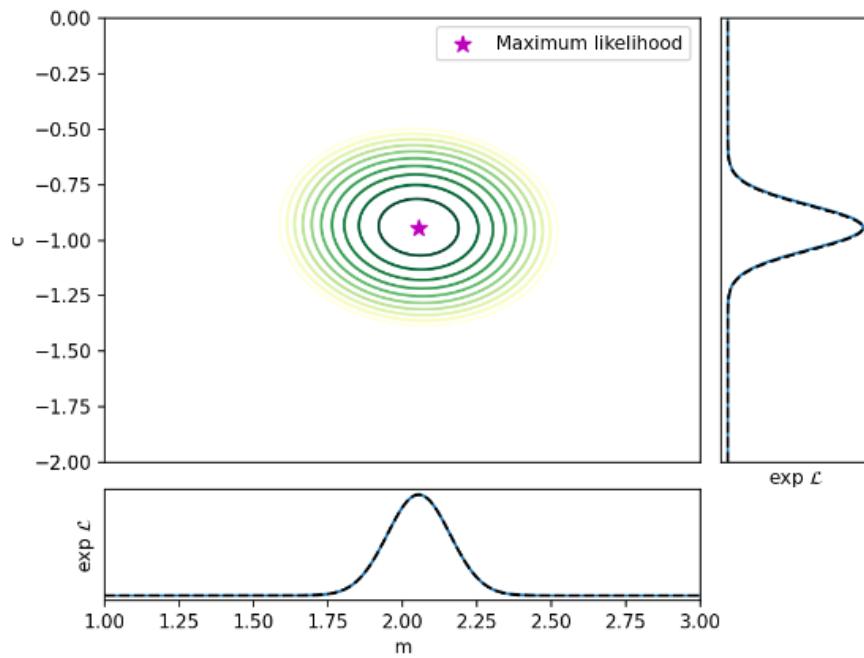


Figure 28.5

Calculating our summary stats and plotting them:

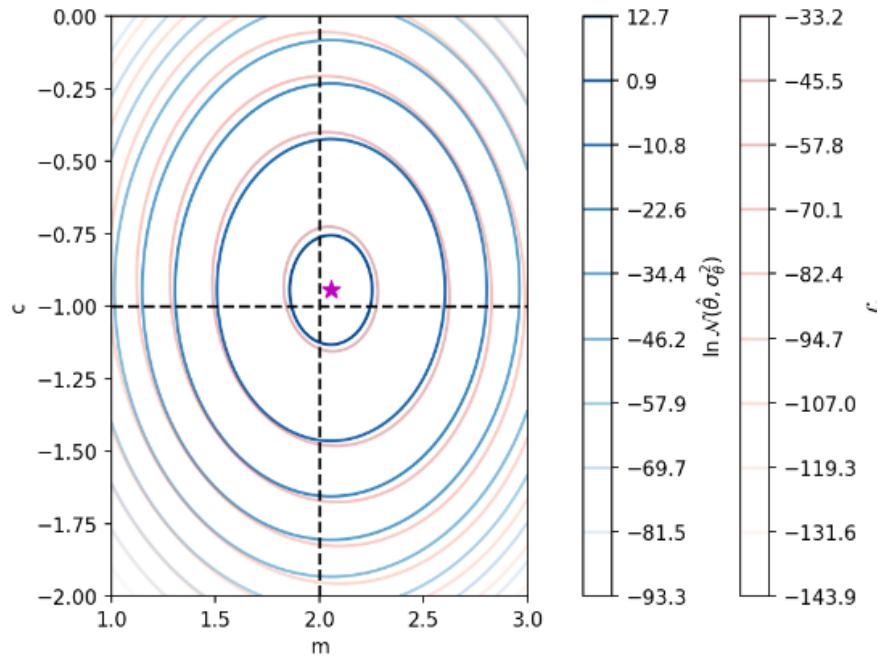


Figure 28.6

We can see that this is pretty good agreement between our estimate and the actual log likelihood function. We now have everything we need to quote the best fit parameters and (crucially) their uncertainty.

Wed 29 Oct 2025 12:00

Lecture S9 - Linear Regression

Previously, we've used numerical methods to determine the best fit parameters for a line of best fit to some data. This is good, because it easily generalises to more complex problems, however lines of best fit have a more specific (and easier to implement) algebraic method.

1 Method

And, we have to perform the same broad steps:

1. A generative model for the data, with knowledge of how the noise is distributed.
2. Likelihood function.
3. A method for finding the maximum likelihood.
4. Method for finding the uncertainties on best fit parameters.
5. A method for checking how good the fit is.

1.1 Generative Model

We use the same generative model as before, with a straight line fit with some additively generated noise (with a standard deviation that may differ from point to point, σ_{D_i}):

$$M(x, \theta) = mx + c$$

1.2 Likelihood Function

Using the same likelihood and log likelihood formulae has before, we can take this one step further by defining (as the extra factor of -2 does not matter when calculating the maxima/minima and ignoring it will make the algebra nicer):

$$\chi^2 = -2\mathcal{L} = \sum_{i=1}^n \left(\frac{D_i - M(x_i, \theta)}{\sigma_{D_i}} \right)^2$$

1.3 Finding Maximum Likelihood (Minimum χ^2)

Since our model is linear, there will only be one turning point for χ^2 , so we can be sure that the minimum of χ^2 will be at the point where the first derivatives (wrt m and c) are zero.

$$\frac{\partial(\chi^2)}{\partial m} = \frac{\partial(\chi^2)}{\partial c} = 0$$

Rather than doing this numerically, as last lecture, we can do it algebraically:

$$\chi^2 = \sum_{i=1}^n \left(\frac{D_i - M(x_i, \theta)}{\sigma_{D_i}} \right)^2 = \sum_{i=1}^n \left(\frac{D_i - mx_i - c}{\sigma_{D_i}} \right)^2$$

w.r.t m :

$$\begin{aligned}\frac{\partial(\chi^2)}{\partial m} &= \frac{\partial}{\partial m} \sum_{i=1}^n \left(\frac{D_i - mx_i - c}{\sigma_{D_i}} \right)^2 = 0 \\ &= \sum_{i=1}^n \frac{\partial}{\partial m} \left(\frac{D_i - mx_i - c}{\sigma_{D_i}} \right)^2\end{aligned}$$

Let $u_i = (D_i - mx_i - c)/\sigma_{D_i}$, so $\partial u_i / \partial m = -x_i / \sigma_{D_i}$:

$$\begin{aligned}&= \sum_{i=1}^n \frac{\partial u_i}{\partial m} \frac{\partial}{\partial u_i} u_i^2 \\ &= \sum_{i=1}^n \frac{\partial u_i}{\partial m} (2u_i) \\ &= \sum_{i=1}^n \left(\frac{-x_i}{\sigma_{D_i}} \right) (2u_i) \\ &= -2 \sum_{i=1}^n \frac{x_i u_i}{\sigma_{D_i}}\end{aligned}$$

So:

$$\begin{aligned}\frac{\partial(\chi^2)}{\partial m} &= -2 \sum_{i=1}^n \frac{x_i}{\sigma_{D_i}} \frac{D_i - mx_i - c}{\sigma_{D_i}} \\ \boxed{\frac{\partial(\chi^2)}{\partial m} = -2 \sum_{i=1}^n \left(\frac{x_i(D_i - mx_i - c)}{\sigma_{D_i}^2} \right) = 0}\end{aligned}$$

w.r.t c :

$$\begin{aligned}\frac{\partial(\chi^2)}{\partial c} &= \frac{\partial}{\partial c} \sum_{i=1}^n \left(\frac{D_i - mx_i - c}{\sigma_{D_i}} \right)^2 = 0 \\ &= \sum_{i=1}^n \frac{\partial}{\partial c} \left(\frac{D_i - mx_i - c}{\sigma_{D_i}} \right)^2\end{aligned}$$

Let $u_i = (D_i - mx_i - c)/\sigma_{D_i}$, so $\partial u_i / \partial c = -1 / \sigma_{D_i}$:

$$\begin{aligned}&= \sum_{i=1}^n \frac{\partial u_i}{\partial c} \frac{\partial}{\partial u_i} u_i^2 \\ &= \sum_{i=1}^n \frac{\partial u_i}{\partial c} (2u_i) \\ &= \sum_{i=1}^n \left(\frac{-1}{\sigma_{D_i}} \right) (2u_i) \\ &= -2 \sum_{i=1}^n \frac{u_i}{\sigma_{D_i}}\end{aligned}$$

So:

$$\boxed{\frac{\partial(\chi^2)}{\partial c} = -2 \sum_{i=1}^n \frac{(D_i - mx_i - c)}{\sigma_{D_i}^2} = 0}$$

2 Putting it All Together

We now have two simultaneous equations with two variables (m, c), so we can solve for the optimum values:

$$\begin{aligned}\frac{\partial(\chi^2)}{\partial c} &= -2 \sum_{i=1}^n \frac{(D_i - mx_i - c)}{\sigma_{D_i}^2} = 0 \\ \frac{\partial(\chi^2)}{\partial m} &= -2 \sum_{i=1}^n \frac{x_i(D_i - mx_i - c)}{\sigma_{D_i}^2} = 0\end{aligned}$$

We can make some substitutions for ease:

$$S = \sum_{i=1}^n \frac{1}{\sigma_{D_i}^2}$$

$$S_x = \sum_{i=1}^n \frac{x_i}{\sigma_{D_i}^2}$$

$$S_{xx} = \sum_{i=1}^n \frac{x_i^2}{\sigma_{D_i}^2}$$

$$S_D = \sum_{i=1}^n \frac{D_i}{\sigma_{D_i}^2}$$

$$S_{Dx} = \sum_{i=1}^n \frac{x_i D_i}{\sigma_{D_i}^2}$$

2.1 Subbing into $\partial(\chi^2)/\partial c$

$$\begin{aligned}-2 \sum_{i=1}^n \frac{(D_i - mx_i - c)}{\sigma_{D_i}^2} &= 0 \\ \sum_{i=1}^n \frac{(D_i - mx_i - c)}{\sigma_{D_i}^2} &= 0\end{aligned}$$

$$S_D - mS_x - cS = 0 \implies S_D = mS_x + cS$$

2.2 Subbing into $\partial(\chi^2)/\partial m$

$$\begin{aligned}-2 \sum_{i=1}^n \frac{x_i(D_i - mx_i - c)}{\sigma_{D_i}^2} &= 0 \\ \sum_{i=1}^n \frac{x_i(D_i - mx_i - c)}{\sigma_{D_i}^2} &= 0\end{aligned}$$

$$S_{Dx} - mS_{xx} - cS_x = 0 \implies S_{Dx} = mS_{xx} + cS_x$$

2.3 Combining

We now have two simultaneous equations to solve:

$$\begin{cases} S_D = mS_x + cS & (1) \\ S_{Dx} = mS_{xx} + cS_x & (2) \end{cases}$$

Rearranging (1) gives:

$$c = \frac{S_d - mS_x}{S} \quad (3)$$

And (3) into (2):

$$S_{Dx} = mS_{xx} + \frac{S_d - mS_x}{S} S_x$$

$$\begin{aligned}
 S_{Dx} &= mS_{xx} + \frac{S_D S_x}{S} - \frac{m(S_x)^2}{S} \\
 S_{Dx} - \frac{S_D S_x}{S} &= mS_{xx} - \frac{m(S_x)^2}{S} \\
 S_{Dx} - \frac{S_D S_x}{S} &= m \left(S_{xx} - \frac{(S_x)^2}{S} \right) \\
 m &= \left(S_{Dx} - \frac{S_D S_x}{S} \right) / \left(S_{xx} - \frac{(S_x)^2}{S} \right)
 \end{aligned}$$

Simplifying to:

$$m = \frac{SS_{Dx} - S_D S_x}{SS_{xx} - S_x^2}$$

And for c , rearranging (2) gives:

$$m = \frac{S_D - cS}{S_x} \quad (4)$$

(4) into (2)

$$\begin{aligned}
 S_{Dx} &= \frac{S_D - cS}{S_x} S_{xx} + cS_x \\
 S_{Dx} &= \frac{S_D S_{xx}}{S_x} - \frac{cSS_{xx}}{S_x} + cS_x \\
 S_{Dx} - \frac{S_D S_{xx}}{S_x} &= cS_x - \frac{cSS_{xx}}{S_x} \\
 S_{Dx} - \frac{S_D S_{xx}}{S_x} &= c \left(S_x - \frac{SS_{xx}}{S_x} \right) \\
 c &= \left(S_{Dx} - \frac{S_D S_{xx}}{S_x} \right) / \left(S_x - \frac{SS_{xx}}{S_x} \right)
 \end{aligned}$$

Simplifying to:

$$c = \frac{S_D S_{xx} - S_x S_{Dx}}{SS_{xx} - S_x^2}$$

2.4 And Finally...

To simplify, let $\Delta = SS_{xx} - S_x^2$:

$$\begin{aligned}
 \hat{c} &\equiv \langle c \rangle = \frac{S_D S_{xx} - S_x S_{Dx}}{\Delta} \\
 \hat{m} &\equiv \langle m \rangle = \frac{SS_{Dx} - S_D S_x}{\Delta}
 \end{aligned}$$

We have therefore managed to calculate the best fit parameters \hat{m} and \hat{c} in closed form without any numerical methods.

3 Uncertainties on Best Fit Parameters

Given we're now in 2D, the Hessian matrix is given as:

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 \mathcal{L}}{\partial \theta_1^2} \Big|_{\theta=\hat{\theta}} & \frac{\partial^2 \mathcal{L}}{\partial \theta_1 \partial \theta_2} \Big|_{\theta=\hat{\theta}} \\ \frac{\partial^2 \mathcal{L}}{\partial \theta_2 \partial \theta_1} \Big|_{\theta=\hat{\theta}} & \frac{\partial^2 \mathcal{L}}{\partial \theta_2^2} \Big|_{\theta=\hat{\theta}} \end{bmatrix}$$

And the covariance matrix Σ is given as:

$$\Sigma = -\mathbf{H}^{-1}$$

We therefore need to calculate the relevant second derivatives. Note that while we could ignore the -2 term before, as it did not matter for finding the location of the maximum, it does matter for errors and cannot be left off. We therefore go back to working in \mathcal{L} and not χ^2

We know the first derivatives of the log likelihood are:

$$\begin{aligned}\frac{\partial(\chi^2)}{\partial m} &= -2(S_{Dx} - mS_{xx} - cS_x) \implies \frac{\partial \mathcal{L}}{\partial m} = S_{Dx} - mS_{xx} - cS_x \\ \frac{\partial(\chi^2)}{\partial c} &= -2(S_D - mS_x - cS) \implies \frac{\partial \mathcal{L}}{\partial c} = S_D - mS_x - cS\end{aligned}$$

Taking second derivatives:

$$\begin{aligned}\frac{\partial^2 \mathcal{L}}{\partial m^2} &= \frac{\partial}{\partial m} (S_{Dx} - mS_{xx} - cS_x) = -S_{xx} \\ \frac{\partial^2 \mathcal{L}}{\partial c^2} &= \frac{\partial}{\partial c} (S_D - mS_x - cS) = -S\end{aligned}$$

And for the term wrt both variables:

$$\frac{\partial^2 \mathcal{L}}{\partial m \partial c} = \frac{\partial^2 \mathcal{L}}{\partial c \partial m} = \frac{\partial}{\partial c} (S_{Dx} - mS_{xx} - cS_x) - S_x$$

Hence (taking $\theta_1 = m$, $\theta_2 = c$):

$$\mathbf{H} = \begin{bmatrix} -S_{xx} & -S_x \\ -S_x & -S \end{bmatrix}$$

And finally:

$$\Sigma = -\mathbf{H}^{-1} = \frac{1}{SS_{xx} - S_x^2} \begin{bmatrix} S & -S_x \\ -S_x & S_{xx} \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} S & -S_x \\ -S_x & S_{xx} \end{bmatrix},$$

Therefore (from the definitions of the covariance matrix):

$$\begin{aligned}Var(m) &= \Sigma_{11} = \frac{S}{\Delta} \\ Var(c) &= \Sigma_{22} = \frac{S_{xx}}{\Delta} \\ Cov(m, c) &= \Sigma_{12} = \Sigma_{21} = -\frac{S_x}{\Delta}\end{aligned}$$

And using the definition of correlation:

$$Cor(m, c) = \frac{Cov(m, c)}{\sqrt{Var(m)Var(c)}} = \frac{-S_x}{\sqrt{SS_{xx}}}$$

Thu 30 Oct 2025 09:00

Lecture S10 - Goodness of Fit

Using likelihood, we can quantify how close our model is to the data, but how do we know if we've got the right model in the first place? We can use a *goodness of fit statistic* to quantify this probabilistically.

1 Straight Line Example

Lets start by taking some data and fitting a straight line to it, using last lectures content:

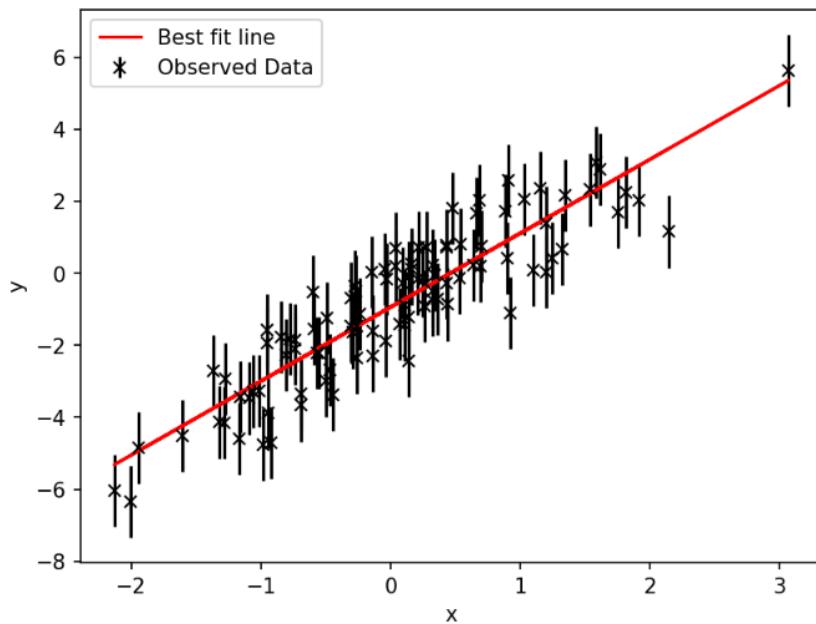


Figure 30.1

We can calculate the χ^2 value for this fit with:

$$\chi^2 = \sum_{i=1}^n \left(\frac{D_i - M(x_i, \theta)}{\sigma_{D_i}} \right)^2$$

Where $M(x_i, \theta) = \hat{m}x_i + \hat{c}$. This gives $\chi^2 \approx 86.37$. We can generate many different datasets, and create lines of best fit for them, and calculate each fit's χ^2 . If we plot these values as a histogram we get:

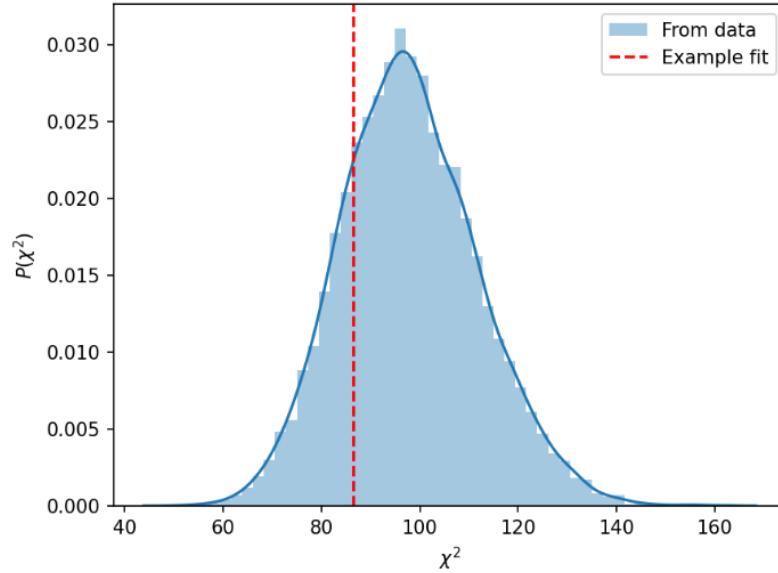


Figure 30.2

This is called a “ χ^2 distribution” and is generated by the frequencies of χ^2 values across many realisations of the data, if the model fitted to the data is the same model actually used to generate the data. A chi squared distribution can be described using the number of “degrees of freedom”, k , and we denote a chi squared distribution with k degrees as χ_k^2 . The PDF is given by:

$$P(x; k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}$$

Where $\Gamma(x)$ is the gamma function, an interpolation of the factorial function across all reals. The number of degrees of freedom is given by the number of data points minus the number of fitted parameters, i.e. $k = N - M$. Comparing our calculated distribution to the theoretical distribution given by the PDF gives:

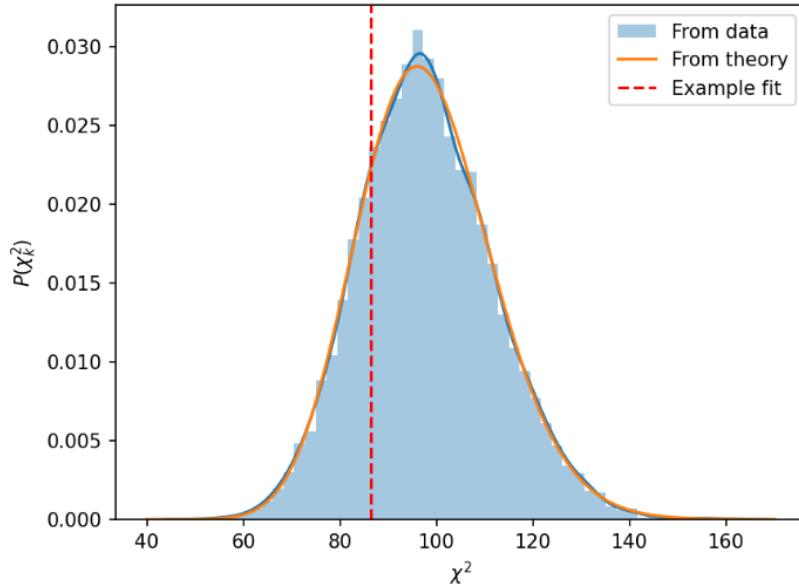


Figure 30.3

Which is a pretty good fit. We can also look at this distribution for different numbers of data points, where a larger number of data points gives a larger number of degrees of freedom:

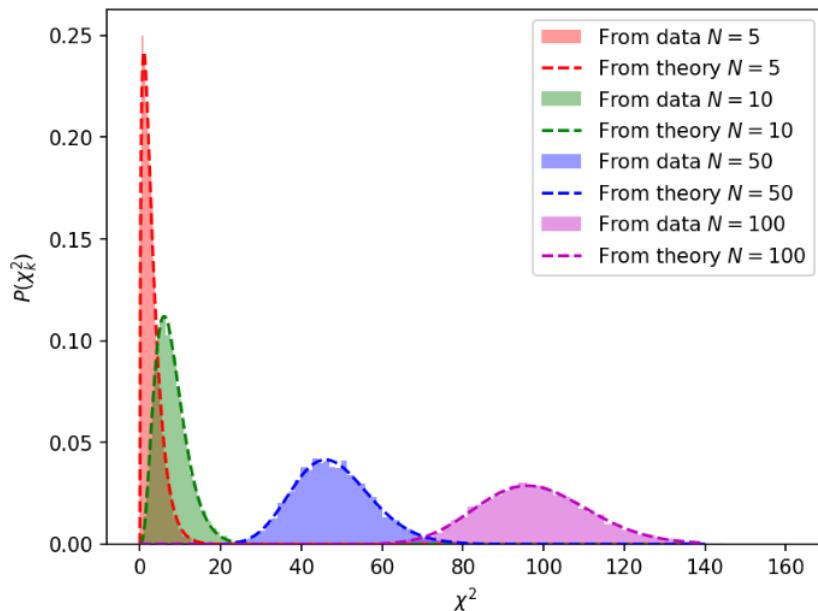


Figure 30.4

If we are fitting the correct model to our data, then we know the distribution of expected χ^2 values is (a χ_k^2 distribution). This provides a goodness of fit statistic, by checking to see if the χ^2 value from the best fit parameters is consistent with what we'd expect to see if our model was correctly chosen.

If assume our model is correct, we can ask how likely it is that we'd get the data χ^2 value from the χ_k^2 likelihood. If this is reasonably likely, we say the goodness of fit is acceptable. If not, then we have a problem. This problem could be many things, including:

- The incorrect model is being fitted - *larger χ^2 than expected*.
- The uncertainties on the data are too small to account for the observed noise - *larger χ^2 than expected*.
- The uncertainties on the data are too large so account for more than the actual noise - *smaller χ^2 than expected*.
- Something else has gone wrong...

The test itself cannot tell us exactly which of these is true, we have to use scientific judgement.

2 Quantifying the Likelihood

Since the χ_k^2 distribution is continuous, the likelihood of getting a specific χ^2 value is zero. Instead, we reframe and look at “what is the chance of getting this value of χ^2 or larger?”. This is done by:

$$P(\chi^2 \geq a) = \int_a^{\infty} P(\chi^2; k) d\chi^2$$

Crucially, the probability returned is only valid on the assumption that the model fitted to the data is correct. If this value is above a certain threshold, we say this is evidence of a sensible fit. This is **not the chance that the model is correct, as the model is just that, a model, and is almost certainly never entirely correct**. It is the chance of getting this χ^2 value or larger *if the model is correct*.

Taking the previous example of $\chi^2 = 86.37$, with $k = 100 - 2 = 98$. This gives:

$$P(\chi^2 \geq 86.37) = \int_{86.37}^{\infty} P(\chi^2; 86.36) d\chi^2$$

This needs to be evaluated numerically, as it becomes unpleasant for not-trivial values of k . This gives a probability of 0.793. This is a reasonable value, not worryingly high (i.e. better than 99%) or worryingly low (i.e. less than 1%), so we say the fit is adequate.

To counterexample, say we have a poor fit. The data is from $y = 0.85x^2 + 2x - 1$, and we try to fit a linear model to it:

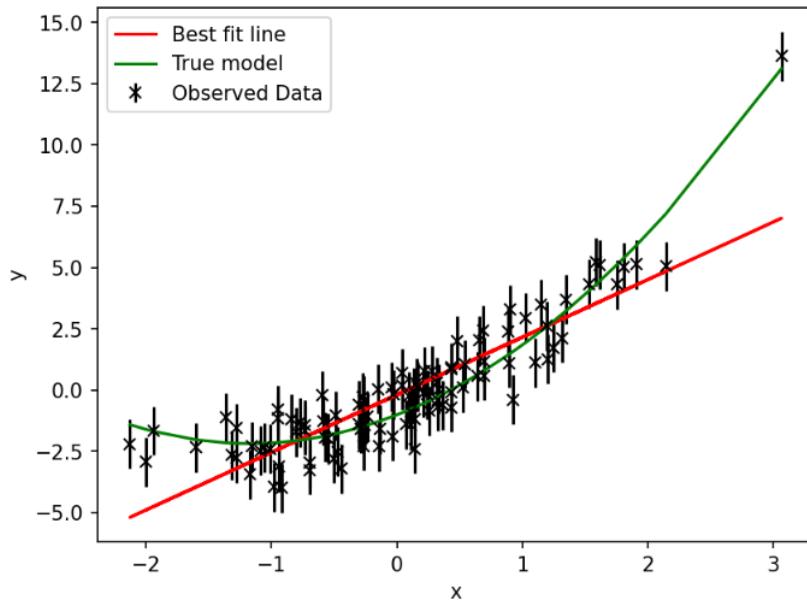


Figure 30.5

This gives a χ^2 value of 166.407. This is much larger than before, lets plot it on the distribution of the expected χ^2 :

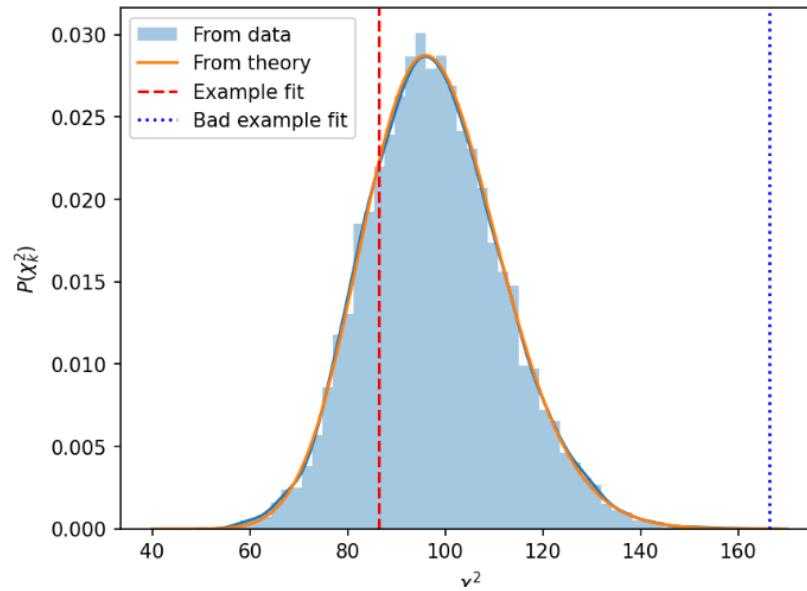


Figure 30.6: Probability is $1.98e - 05$

So we can confidently say that the fit is not an acceptable quality fit. This is despite it not being a terrible fit by eye.

Thu 06 Nov 2025 09:00

Lecture S11 - End of Stats: Revision

Revision lecture (focussed on sheet 6 of the non-assessed problems). No new content covered.

Fri 07 Nov 2025 11:00

Lecture P1 - Start of Probability: Introduction

What is probability? Probability is the pure mathematical description of randomness.

1 Empirical Probability

Say we want to group trees into four sets:

- Tall, or not.
- Variegated (has a lighter coloured leaf border) or not.

In a park of 142 trees, we observe:

| | Tall | Not Tall | Total |
|----------------|------|----------|-------|
| Variegated | 20 | 41 | 61 |
| Not Variegated | 72 | 9 | 81 |
| Total | 92 | 50 | 142 |

Figure 32.1

We denote Tall as T , Variegated as V and not Tall/not Variegated as \bar{T} and \bar{V} . We use n as the number of number of trees that satisfy the parameters, i.e $n(T, \bar{V}) = 72$. $N = 142$ is the total number of trees, so:

$$N = n(T, V) + n(T, \bar{V}) + n(\bar{T}, \bar{V}) + n(\bar{T}, V)$$

Similarly, we define the fraction of trees that meet the provided criteria as f , i.e:

$$f(T, V) \equiv \frac{n(T, V)}{N}$$

Hence:

$$1 = f(T, V) + f(T, \bar{V}) + f(\bar{T}, \bar{V}) + f(\bar{T}, V)$$

We define probability as the limit of this as $N \rightarrow \infty$:

$$P(T, V) = \lim_{N \rightarrow \infty} f(T, V) = \lim_{N \rightarrow \infty} \frac{n(T, V)}{N}$$

And discrete probability as the limiting fraction of the times an event will occur. Probability is bounded between 0 (the event **never** occurs) and 1 (the event **always** occurs). In a particular experiment, the total probability is one, therefore something must happen (we just may not know what exactly).

2 Set Theory

A set is a collection of elements, i.e.:

$$A = \{1, 2, 3\}$$

A deck of cards is a set, with 52 elements. Elements can be anything, including other sets.

2.1 Subsets

Consider some set A . Another set, B is a “subset” of A if it contains (solely) some of the elements of A . We denote this $B \subset A$.

For example:

$$A = \{1, 2, 3, 4, 5, 6\}$$

$$B = \{1, 2, 3\}$$

$$C = \{4, 5, 6\}$$

$$D = \{1, 5, 7\}$$

$$B \subset A$$

$$C \subset A$$

$$D \not\subset A$$

2.2 Sample Space

We observe a particular outcome, and the set of all possible outcomes we could observe is called the sample space Ω . Any event (that can occur) is a subset of Ω .

If drawing 5 cards from a deck, all possible combinations of 5 cards form the sample space. Drawing an ace is an event.

3 Probability Functions

Probability assigns a value to every event in Ω to quantify how likely it is to happen. This value is the probability of the event.

If something must happen, then we have:

$$P(\Omega) = 1$$

This is called ‘normalisation’.

Thu 13 Nov 2025 09:00

Lecture P2 - Combinatorics and Counting

1 Combinatorics

This is the way of counting the number of ways of something happening.

For example, we toss two dice. How many ways are there of getting a given total. There is one way to get a two (1, 1) but three ways of getting a four: (3, 1), (1, 3), (2, 2).

Combinatorial factors get very large extremely quickly, i.e. factorially or exponentially. For example, the number of Sudoku boards is huge at $\approx 6.7 \times 10^{21}$.

Again, we use Ω to represent the set of possible outcomes, and we use $|\Omega|$ to represent the number of outcomes.

2 Sampling

Sampling involves having some set of N objects, and selecting k of them. We can:

- **Sample with replacement and keep order:**
 - Pick 4 numbers for a PIN freely.
- **Sample without replacement and keep order:**
 - Ensure that no numbers in the PIN are repeated.
- **Sample with replacement and ignore order:**
 - An ice cream shop has vanilla, chocolate and strawberry. You have paid for two scoops in a bowl.
 - You may pick any combination of the three flavours, including repeats, i.e two scoops of chocolate, one vanilla one chocolate or two strawberry are all valid. The order the scooper puts them into the bowl doesn't matter, adding strawberry and then chocolate is the same bowl of ice cream as chocolate then strawberry.
- **Sample without replacement and ignore order.**
 - Making a fruit salad with two fruits from bananas, cherries, apples, mangoes.
 - You cannot choose the same fruit twice, but the order is irrelevant, so adding apples and then mangoes vs mangoes and then apples results in the same salad.

In summary:

- With replacement, once object has been selected, it goes back into the set of possible objects and can be chosen again.
- With non-replacement, once an object has been selected it is removed from the pool and cannot be chosen again.
- Order or not is whether or not the order selections happen in is something we keep track of or not, i.e. is 1234 an equivalent event to 4321.

2.1 Counting

Most simply, we have the counting rule. This states that if we have n ways of doing one thing, and m ways of doing a second, the number of possible combinations if we do both things is $n \times m$

In general: We sample k things, the first has n_1 choices, the second has n_2 , etc up to n_k . The total choices is $n_1 \times n_2 \times \cdots \times n_k$

2.2 Sampling with Replacement

We sample k things from N objects with replacement. The first has N choices, the second still has the same N choices etc. Therefore:

$$|\Omega| = \underbrace{N \times N \times \cdots \times N}_{k \text{ times}} = N^k$$

2.3 Sampling without Replacement

We sample k things from N objects without replacement. The first has N choices, the second time we cannot make the same choice again, so have $N - 1$ choices. The third time we cannot make the same choice as the first or the second, so have $N - 2$ etc. Therefore:

$$|\Omega| = N \times (N - 1) \times (N - 2) \times \cdots \times (N - k + 1) = \frac{N!}{(N - k)!}$$

We denote this ${}^N P_k$ or ${}^N V_k$ (where the v is for ‘variations’ and p for ‘permutations’).

2.4 Permutations

We consider order and sample N things from N objects without replacements, i.e. we draw every object from the set of possible objects, in some order.

We get:

$$|\Omega| = N \times (N - 1) \times (N - 2) \times \cdots \times 2 \times 1 = N!$$

This is commonly called a ‘full permutation’ of N , while the previous ${}^N P_k$ is called a ‘ k -permutation’ of N .

2.5 Unordered Samples

In a hand of cards the order the cards appear in the hand is irrelevant. We have no replacement as you cannot have the same card twice, but a hand with (for example) the queen of hearts and jack of diamonds vs the jack of diamonds and the queen of hearts is just the same deck.

If we consider order, we get $\frac{N!}{(N-k)!}$ but this ends up overcounting when we don’t want order. We therefore need to divide out the total ways of permuting $k!$ objects. For example, a three card hand would (if considering order) count each set of 3 cards $3!$ times, so we divide out to only count all of these once. This gives:

$$|\Omega| = \frac{N!}{k!(N - k)!} \equiv {}^N C_k \equiv \binom{N}{k}$$

Pronounced “n choose k”.

2.6 Interpretations of Binomial Coefficient

Notably, this n choose k notation pops up in the binomial expansion too:

$$(a + b)^N = \sum_{k=0}^N \binom{N}{k} a^k b^{N-k}$$

It is also the number of ways to split N objects into two groups, where one group has k objects and the other has $N - k$ objects.

3 Multinomial

Say we split a deck of 52 cards into 4 even hands, the order of each is irrelevant. The first person picks 13 cards from 52, with $\binom{51}{13}$ options, the second then picks from the remaining 39 and has $\binom{39}{13}$ etc.

This gives:

$$|\Omega| = \binom{52}{13} \times \binom{39}{13} \times \binom{26}{13} \times \binom{13}{13} = \frac{52!}{13! 39!} \frac{39!}{13! 26!} \frac{26!}{13! 13!} \frac{13!}{13! 0!} = \frac{51!}{13! 13! 13! 13!}$$

This is called the multinomial coefficient.

In general if we have N objects and partition into p containers, n_1, n_2, \dots, n_p , we have:

$$|\Omega| = \binom{N}{n_1} \times \binom{N-n_1}{n_2} \times \binom{N-n_1-n_2}{n_3} \times \cdots \times \binom{n_p}{n_p}$$

$$|\Omega| = \frac{N!}{n_1! n_2! \cdots n_p!} \equiv \binom{N}{n_1 \ n_2 \ \cdots \ n_p}$$

Hence:

$$(a+b+c)^N = \sum_{k_1+k_2+k_3=N} = \binom{N}{k_1 \ k_2 \ k_3} a^{k_1} b^{k_2} c^{k_3}$$

4 Uniform Probability

Assuming every outcome is equally likely to happen, i.e. probability is uniform, the probability of an event A is given by:

$$P(A) = \frac{|A|}{|\Omega|}$$

I.e. the probability of A is the fraction of the sample space it takes up.

Thu 14 Nov 2025 11:00

Lecture P3 - Combining Probabilities

Today we will arrive at:

- The formula for $P(A \cap B)$ (Probability of A and B).
- Summing mutually exclusive events.

1 More Set Theory

We have a sample space Ω , and a subset labelled A . We then have the remainder of Ω (the portion of Ω which is not in A), denoted "A Complement" - A^C or \bar{A} . We also have a subset labelled B .

We can define A using set builder notation, to slightly redundantly say "A is the set of all x'es which are in A":

$$A = \{x \mid x \in A\}$$

There is some overlap between A and B. We denote this intersection as $A \cap B$.

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}$$

Everything written in A or B (including the intersection) is called the union, $A \cup B$:

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}$$

Note that "or" in standard language excludes both, i.e. you may have x or you may have y. In mathematics, we refer to this as XOR (exclusive or). "Or" by itself does allow for this case of both, so an item in A or B may be in A alone, B alone, or both (i.e. in the intersection).

We also have the empty set $\emptyset = \{\}$. If two sets have no common elements, the intersection is this empty set. We say that the events are mutually exclusive (they cannot both happen) and the sets are pairwise disjoint. The empty set is the complement of Ω , $\emptyset = \Omega^C$.

2 De Morgan's Laws

De Morgan's Laws give us these relations:

$$(A \cup B)^C = A^C \cap B^C \quad (34.1)$$

$$(A \cap B)^C = A^C \cup B^C \quad (34.2)$$

This can be illustrated visually as follows:

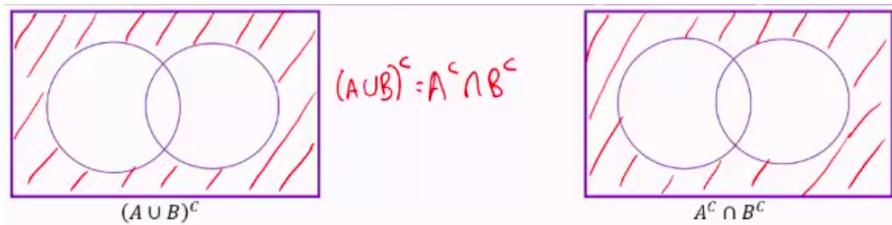
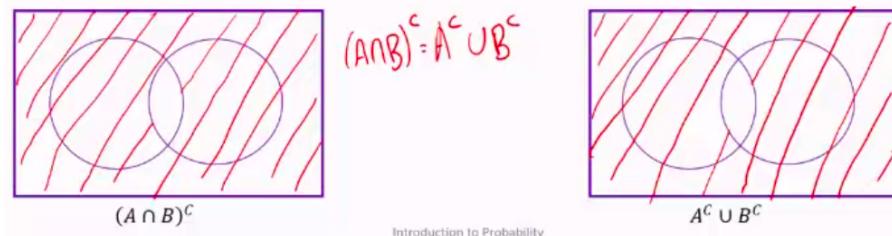


Figure 34.1



Introduction to Probability

Figure 34.2

3 The Inclusion-Exclusion Principle

The number of elements in A and B is given by:

$$|A \cup B| = |A| + |B| - |A \cap B|$$

The last term is required to account for the intersection of A and B being included in A, and included in B. Therefore it double-counts the intersection, and we subtract it away.

The same is true of probability functions:

$$P(A) + P(B) = P(A \cup B) + P(A \cap B)$$

In other words, the probability of A + the probability of B is the probability of A or B + the probability of A + B.

This has the following consequences:

$$P(\emptyset) = 0$$

As:

$$P(A) = \frac{|A|}{|\Omega|} \implies P(\emptyset) = \frac{0}{|\Omega|} = 0$$

And:

$$P(A) = p \implies P(A^c) = 1 - p$$

As:

$$\Omega = A \cup A^c$$

$$P(\Omega) = P(A) + P(A^c) - P(A \cup A^c) \quad \text{By inclusion-exclusion principle.}$$

$$P(A \cup A^c) = P(\emptyset) = 0$$

$$1 = P(A) + P(A^c) \quad \text{As: } P(\Omega) = 1$$

4 Multiple Events

Given some events e_n , the inclusion-exclusion principle says:

$$P(e_1 \cup e_2) = P(e_1) + P(e_2) - P(e_1 \cap e_2)$$

Hence for independent events ($e_1 \cap e_2 = \emptyset$), the probability of e_1 or e_2 occurring is:

$$P(e_1 \cup e_2) = P(e_1) + P(e_2)$$

4.1 What about 3 events?

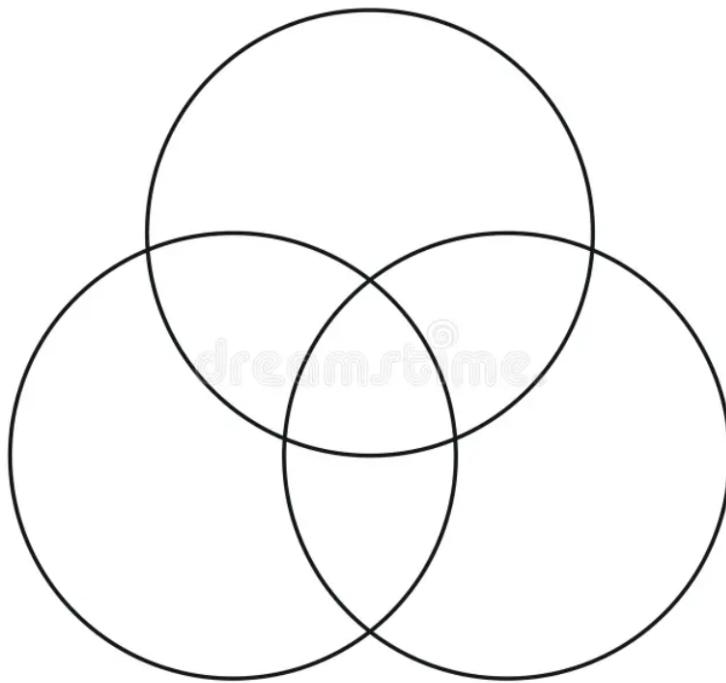


Figure 34.3

For three events, our venn diagram becomes more complex. We want to calculate $|e_1 \cup e_2 \cup e_3|$ (in this general example, which circle is which event is irrelevant).

This is given by:

$$|e_1 \cup e_2 \cup e_3| = |e_1| + |e_2| + |e_3| - |e_1 \cap e_2| - |e_1 \cap e_3| - |e_2 \cap e_3| + |e_1 \cap e_2 \cap e_3|$$

Note the final term, as the central portion is included three times when summing the whole event, but subtracted three times when removing intersections, so we must add it back.

If they are all pairwise disjoint, so $e_i \cup e_j = \emptyset$, then:

$$P(e_1 \cup e_2 \cup \dots \cup e_n) = P(e_1) + P(e_2) + \dots + P(e_n)$$

Or:

$$P\left(\bigcup_{n=1}^N e_n\right) = \sum_{n=1}^N P(e_n)$$

4.2 Normalisation

If the events are all mutually exclusive, and the sample space is “tiled” by the events (i.e. one of them must happen), then:

$$P(\Omega) = P(e_1) + P(e_2) + \dots + P(e_n) = 1$$

Thu 20 Nov 2025 09:00

Lecture P4 - Conditional Probability

1 Axioms of Probability

For a sample space Ω , a distribution $P(x)$ must satisfy:

1. $P(x) \geq 0$ for any $x \in \Omega$ (for discrete events)
2. $P(\Omega) = 1$
3. $P(e_1 \cup e_2 \cup \dots \cup e_n) = P(e_1) + P(e_2) + \dots + P(e_n)$
 - if the elements are pairwise disjoint ($e_i \cap e_j = \emptyset, i \neq j$).

2 Conditional Probability

2.1 Example

Throw two dice. What is the probability that we see a 4, given the total was 6?

This “given that” is the key. It provides us with an extra piece of information that the final probability depends on.

| | | | | | |
|------------|---------------------|---------------------|---------------------|---------------------|---------|
| $(1,1)$ | $(2,1)$ | $(3,1)$ | $(4,1)$ | $(\underline{5,1})$ | $(6,1)$ |
| $(1,2)$ | $(2,2)$ | $(3,2)$ | $(\underline{4,2})$ | $(5,2)$ | $(6,2)$ |
| $\Omega =$ | $(1,3)$ | $(2,3)$ | $(\underline{3,3})$ | $(4,3)$ | $(5,3)$ |
| | $(1,4)$ | $(\underline{2,4})$ | $(3,4)$ | $(4,4)$ | $(5,4)$ |
| | $(\underline{1,5})$ | $(2,5)$ | $(3,5)$ | $(4,5)$ | $(5,5)$ |
| | $(1,6)$ | $(2,6)$ | $(3,6)$ | $(4,6)$ | $(5,6)$ |
| | | | | | $(6,6)$ |

Figure 35.1: The sample space.

We can now see that our probability is $\frac{2}{5}$. This is different than if we had not considered the extra information, and would have drastically changed our answer.

2.2 Definition

The conditional probability of A given B is written $P(A | B)$ and is defined by:

$$P(A | B) \equiv \frac{P(A \cap B)}{P(B)} = \frac{\text{number of events in } A \text{ and } B}{\text{number of events in } B} \quad P(B) \neq 0$$

This is the fraction of events in B where both A and B happen.

2.3 Verifying This is Still a Probability

It may not be obvious that this is still a valid probability (i.e. that taking ratios of probabilities still yields a probability).

Proof. Assume that P is a valid probability function and let $Q(A | B) \equiv \frac{P(A \cap B)}{P(B)}$

Consider the first axiom of probability:

$Q(A | B) \geq 0$. This is satisfied, as $P(x) \geq 0$ for any $x \in \Omega$.

Consider the second:

$$Q(\Omega | B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1 \text{ so satisfied.}$$

Consider the third:

If $a_1 \cap a_2 = \emptyset$, is $Q(a_1 \cup a_2 | B) = Q(a_1 | B) + Q(a_2 | B)$?

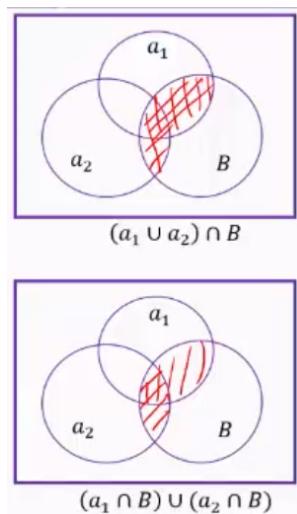


Figure 35.2

The definition of Q gives:

$$\begin{aligned} Q(a_1 \cup a_2 | B) &= \frac{P([a_1 \cup a_2] \cap B)}{P(B)} \\ &= \frac{P([a_1 \cap B] \cup [a_2 \cap B])}{P(B)} \\ &= \frac{P(a_1 \cap B) + P(a_2 \cap B)}{P(B)} \\ &= \frac{P(a_1 \cap B)}{P(B)} + \frac{P(a_2 \cap B)}{P(B)} \\ &= Q(a_1 | B) + Q(a_2 | B) \end{aligned}$$

So yes, the third axiom is satisfied.

Since all three axioms are satisfied, Q , hence P , is a valid probability distribution. □

3 Reconditioning

Reconditioning is using $P(A | B)$ to determine $P(B | A)$.

$$(1) \quad P(A | B) = \frac{P(A \cap B)}{P(B)}$$

$$(2) \quad P(B | A) = \frac{P(B \cap A)}{P(A)}$$

Using the fact that $A \cap B = B \cap A$:

$$\begin{aligned} P(B | A) &= \frac{P(A \cap B)}{P(A)} \\ \implies P(A \cap B) &= P(A | B)P(B) = P(B | A)P(A) \\ \implies P(B | A) &= \boxed{\frac{P(A | B)P(B)}{P(A)}} \end{aligned}$$

3.1 Example

A rare cold-like disease has symptoms with probability 0.95. The probability in the population of having the disease is 0.0001. The probability of having the cold-like symptoms in the population is 0.4, either from the disease or any other cold.

Given someone has symptoms, what is the probability they have the disease? Let d be having the disease, and let s be having symptoms. Then:

$$P(d) = 0.0001$$

$$P(s) = 0.4$$

$$P(s | d) = 0.95$$

Therefore:

$$\begin{aligned} P(d | s) &= \frac{P(s | d)P(d)}{P(s)} \\ &= \frac{0.95 \times 0.0001}{0.4} = 0.002 \end{aligned}$$

So, as expected, it is very unlikely for someone to have the disease even if they present with symptoms.

4 Statistical Independence

If knowing B happened has no impact on whether A happens or not, then:

$$P(A | B) = P(A)$$

In this case:

$$P(A \cap B) = P(A | B)P(B) = P(A)P(B)$$

$$P(x_1 \cap x_2 \cap \dots \cap x_n) = P(x_1)P(x_2) \dots P(x_n)$$

This (if it is true) is called *statistical independence*.

Fri 21 Nov 2025 11:00

Lecture P5 - Law of Total Probability and Bayes Theorem

1 Total Probability

Given some sets A and B that are not mutually exclusive, we can take an event in A and break it down into:

1. The piece of A which is also in B .
2. The piece of A which is not also in B .

This means that:

$$P(A) = P(A \cap B) + P(A \cap B^C)$$

1.1 The Law of Total Probability

This becomes more useful when we consider multiple events. Say we have a sequence of disjoint sets B_1, B_2, \dots, B_n . These events tile A . We then have:

$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) + \dots + P(A \cap B_n) \\ P(A) &= \sum_{n=1}^N P(A \cap B_n) \end{aligned}$$

This distribution $P(A)$ is called the marginal distribution.

2 Change of Notation

We now abstract ourselves away from sets, so change our notation:

$$P(A \cap B) \mapsto P(A, B)$$

This is called the “joint distribution”. Our previous equations now become:

$$\text{Marginalisation: } P(A) = \sum_{n=1}^N P(A, B_n)$$

$$\text{Conditional Probability: } P(A | B) = \frac{P(A, B)}{P(B)}$$

$$\text{Statistical Independence: } P(A, B) = P(A)P(B)$$

2.1 Example 1

If the joint distribution $P(x, y)$ is given by:

| $P(x, y)$ | $x = 0$ | $x = 1$ | $x = 2$ |
|-----------|---------|---------|---------|
| $y = 0$ | $2/9$ | $2/9$ | $2/9$ |
| $y = 1$ | $1/9$ | $1/9$ | $1/9$ |

Figure 36.1

What is $P(y)$? We marginalise over x so:

$$\begin{aligned} P(y = 1) &= \sum_{x=0}^2 P(x, y = 1) \\ &= 1/9 + 1/9 + 1/9 = 1/3 \end{aligned}$$

$$\begin{aligned} P(y = 0) &= \sum_{x=0}^2 P(x, y = 0) \\ &= 2/9 + 2/9 + 2/9 = 2/3 \end{aligned}$$

3 Conditional Probability with the Law of Total Probability

We have that:

$$P(A, B) = P(A | B)P(B)$$

So we can adapt our law of total probability to use this:

$$P(A) = \sum_{n=1}^N P(A, B_n) = \sum_{n=1}^N P(A | B_n)P(B_n)$$

3.1 Example

If:

$$P(A | B) = 0.2$$

$$P(A | \bar{B}) = 0.4$$

$$P(B) = 0.1$$

What is $P(A)$? We marginalise (taking the two possible outcomes as B happening or not):

$$\begin{aligned} P(A) &= P(A | B)P(B) + P(A | \bar{B})P(\bar{B}) \\ &= 0.2 \times 0.1 + 0.4 \times (1 - 0.1) = 0.38 \end{aligned}$$

4 Bayes' Theorem

We can now derive the general form of Bayes' Theorem by combining the definition of conditional probability with the Law of Total Probability.

The definition of conditional probability is

$$P(A | B) = \frac{P(A, B)}{P(B)}$$

We can recondition to get:

$$P(A, B) = P(B | A)P(A)$$

Substituting this into the definition gives:

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}$$

However, we often do not know $P(B)$ explicitly. We need to marginalise over all possible states of A to get this:

$$P(B) = \sum_A P(B, A) = \sum_A P(B | A)P(A)$$

$$P(A | B) = \boxed{\frac{P(B | A)P(A)}{\sum_A P(B | A)P(A)}}$$

5 Bayes' Examples

5.1 Example 1

Likely to come up on exam.

There are two boxes:

- Box A has 5 gold and 5 silver.
- Box B has 5 gold coins and 10 silver.

If a box is picked at random, and a coin picked at random from this box, what is the probability that the box chosen was A, given the coin was silver.

We want $P(A | s)$:

$$\begin{aligned} P(A | s) &= \frac{P(s | A)P(A)}{P(s | A)P(A) + P(s | B)P(B)} \\ &= \frac{1/2}{1/2 + 2/3} = 3/7 \end{aligned}$$

5.2 Example 2

Polygraphs are used to screen people. Either an individual tells the truth, or they lie. The (imperfect) polygraph returns 0 if it thinks the person lies, or 1 if being truthful, with the following probabilities:

$$P(0 | \text{lies}) = 0.88$$

$$P(1 | \text{truth}) = 0.86$$

People lie rarely, about 1% of the time. What is the probability that a person is actually lying if the polygraph gives a zero? I.e. we want $P(\text{lie} | 0)$

$$\begin{aligned} P(L | 0) &= \frac{P(0 | L)P(L)}{P(0 | L)P(L) + P(0 | T)P(T)} \\ &= \frac{0.88 \times 0.01}{0.88 \times 0.01 + (1 - 0.86) \times (1 - 0.01)} \\ &= 0.06 \end{aligned}$$

We therefore also know: $P(T | 0) = 0.94$. Even though the polygraph is “accurate”, so assume that so few people will lie which makes the results unreliable. This means it is inherently difficult to conduct large-scale testing for rare events.

Thu 27 Nov 2025 09:00

Lecture P6 - Ordered Events and Expectation Values

1 Ordering

So far we've just had fairly abstract events. We now want to add a structure into these events, by considering what happens when the events are ordered. This (rather than physical order) means what if each event is associated with an actual number.

For example, events being getting a head or getting a tail is abstract, while counting the number of heads when tossing a coin is numerical and ordered.

1.1 Probability Mass Functions (PMFs)

If we have a variable x , the probability of observing x is $P(x)$. Since we now consider cases where events are numerical values, we can graph $P(x)$.

We previously had a probability function that assigns probabilities to events. Now we are assigning probabilities to numbers, this is called a probability mass function (PMF).

1.2 Normalisation

The events in a PMF are disjoint from the rest. We can therefore normalise through summation to 1:

$$\sum_x P(x) = 1$$

1.3 Cumulative Distributions

This is the probability that a variable is less than or equal to a certain value.

$$C(x) \equiv P(X \leq x) = \sum_{X \leq x} P(x)$$

| x | 1 | 2 | 3 | 4 | 5 | 6 |
|--------|-------|-------|-------|-------|-------|-------|
| $P(x)$ | $1/6$ | $1/6$ | $1/6$ | $1/6$ | $1/6$ | $1/6$ |
| $C(x)$ | $1/6$ | $1/3$ | $1/2$ | $2/3$ | $5/6$ | 1 |

Figure 37.1

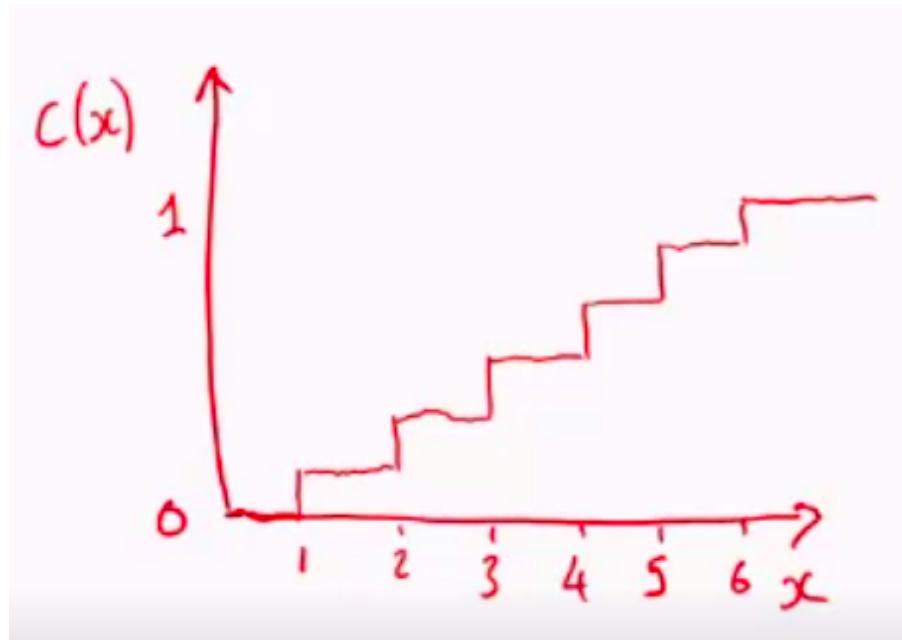


Figure 37.2

2 Expectation Values

The expectation value for a probability mass function is defined by:

$$\langle x \rangle = \sum_x x P(x)$$

It's the weighted sum of all the outcomes (weighted by their probability of occurring), and is a measure of location. Note that the expectation value may not be an actual value that our discrete distribution can obtain, it may be between two.

The best way to describe what it actually represents is the long-term average of a distribution (if you take many draws from the PMF and average them).

2.1 Expectation Value of a Function

$$\langle f \rangle = \sum_x f(x) P(x)$$

This is again a weighted sum over the values of x .

Note: For some constant c :

$$\begin{aligned} \langle c \rangle &= \sum_x x P(x) = c \sum_x P(x) = c \\ \langle c \rangle &= c \end{aligned}$$

This implies that:

$$\langle \langle x \rangle \rangle = \langle x \rangle$$

2.2 Linear Combinations

We stated these as fact earlier in the stats portion of the module. We can now derive them.

$\langle ax + b \rangle$:

$$\begin{aligned}\langle ax + b \rangle &= \sum_x (ax + b)P(x) \\ &= \sum_x axP(x) = \sum_x bP(x) \\ &= a \sum_x xP(x) + b \sum_x P(x) \\ &= a\langle x \rangle + b\end{aligned}$$

3 Measure of Dispersion

Now we know $\langle x \rangle$, we can ask the question how far, on average, does the value of x get away from this.

In other words, what is the expectation value of the difference between x and $\langle x \rangle$? Lets try:

$$\langle x - \langle x \rangle \rangle = \langle x \rangle - \langle \langle x \rangle \rangle = 0$$

Ah, this is always 0, so doesn't work. . . What else can we try?

$$\langle |x - \langle x \rangle| \rangle = \text{MAD}(x)$$

This is the Mean Absolute Deviation of x . This is perfectly valid, but not really very common anymore. More commonly, we use the variance:

$$\langle |x - \langle x \rangle|^2 \rangle = \sum_x (x - \langle x \rangle)^2 P(x) = \text{Var}(x)$$

3.1 Simplifying Variance

$$\begin{aligned}\text{Var}(x) &\equiv \langle (x - \langle x \rangle)^2 \rangle \\ &= \langle x^2 - 2\langle x \rangle x + \langle x \rangle^2 \rangle \\ &= \langle x^2 \rangle - 2\langle x \rangle \langle x \rangle + \langle \langle x \rangle^2 \rangle \\ &= \langle x^2 \rangle - 2\langle x \rangle^2 + \langle x \rangle^2 \\ \text{Var}(x) &= \langle x^2 \rangle - \langle x \rangle^2\end{aligned}$$

We are further going to define *standard deviation* as the square root of variance:

$$\text{std}(x) = \sqrt{\langle x^2 \rangle - \langle x \rangle^2}$$

3.2 Linear Variance Combinations

We now want to consider the same linear combinations but for variance this time:

$$\begin{aligned}\text{var}(ax + b) &= \langle (ax)^2 \rangle - \langle ax \rangle^2 \\ &= \langle a^2 x^2 \rangle - a^2 \langle x \rangle^2 \\ &= a^2 \{ \langle x^2 \rangle - \langle x \rangle^2 \} \\ \boxed{\text{var}(ax + b) = a^2 \text{var}(x)}\end{aligned}$$

Fri 28 Nov 2025 11:00

Lecture P7 - Discrete Distributions

1 Parametric Distributions

So far, we've treated $P(x)$ very arbitrarily. We now want to define it a bit more tightly.

So far, we've explicitly defined values for every combination (i.e. $P(x = 0) = a, P(x = 1) = b$) etc. Instead of doing this, we can define them *parametrically*. We add some number of parameters θ in:

$$P(x | \theta)$$

We can have more than one parameter:

$$P(x | \theta_1, \theta_2, \dots, \theta_n) = P(x | \theta)$$

1.1 Bernoulli Distribution

The Bernoulli Distribution is the simplest distribution we can define. It underpins all of the rest. We define two events, 0 and 1. 1 happens with probability p and 0 with probability $1-p$. We then run a **single** trial.

If:

$$P(x = 0 | p) = 1 - p \quad P(x = 1 | p) = p$$

Then:

$$P(x | p) = p^x(1-p)^{1-x} \quad x = 0, 1$$

Or:

$$P(x | p) = (1-x)(1-p) + xp \quad x = 0, 1$$

Properties

$$\begin{aligned} \langle x \rangle &\equiv \sum_x xP(x | p) = \sum_{x=0}^1 xp^x(1-p)^{1-x} = 0p^0(1-p)^1 + 1p^1(1-p)^0 = p \\ \langle x^2 \rangle &\equiv \sum_x x^2P(x | p) = p \\ \text{var}(x) &= p - p^2 = p(1-p) \end{aligned}$$

1.2 Binomial Distribution

The binomial distribution is the sum of Bernoulli distributions. If we toss N coins and count the number of heads k , what is the distribution of k ?

Let $P(H) = p$, so $P(T) = 1 - p$.

If $n = 3$:

$$\Omega \{ \text{HHH}, \text{HHT}, \text{HTH}, \text{HTT}, \text{THH}, \text{THT}, \text{TTH}, \text{TTT} \}$$

We assume that each throw of the coin is independent of the previous throw. Therefore $P(\text{HHH}) = P(H)P(H)P(H) = P(H)^3$

For k heads, we use the union. For $k = 1$, we have:

$$P(k = 1) = P(\text{TTT} \cup \text{TTH} \cup \text{THT} \cup \text{HTT})$$

Each one of these options has probability $p(1-p)^2$, so $P(k = 1) = 3p(1-p)^2 = {}^3C_1 p(1-p)^2$ (as there are 3 possible outcomes that give us that combination of H and T).

This gives us:

$$P(k = 0) = {}^3_0 (1-p)^3$$

$$P(k = 1) = {}^3_1 p(1-p)^2$$

$$P(k = 2) = {}^3_2 p^2(1-p)$$

$$P(k = 3) = {}^3_3 p^3$$

We can get the final probability mass function:

$$P(k | N, p) = {}^N_k p^k (1-p)^{N-k}$$

And therefore the expectation value:

$$\begin{aligned} \langle k \rangle &\equiv \sum_{k=0}^N k P(k | N, p) \\ &= \sum_{k=0}^N k \frac{N!}{k!(N-k)!} p^k (1-p)^{N-k} \\ &= 0 + \sum_{k=1}^N k \frac{N!}{k!(N-k)!} p^k (1-p)^{N-k} \\ &= \sum_{k=1}^N \frac{N!}{(k-1)!(N-k)!} p^k (1-p)^{N-k} \\ &= Np \sum_{k=1}^N \frac{(N-1)!}{(k-1)!(N-k)!} (p^{k-1})(1-p)^{N-k} \end{aligned}$$

Let: $t = k - 1 \implies k = t + 1$:

$$\begin{aligned} &= Np \sum_{t=0}^{N-1} \frac{N!}{(t)!(N-1-t)!} (p^t)(1-p)^{N-1-t} \\ &= Np \sum_{t=0}^{N-1} P(t | N-1, p) \\ &= Np \times 1 \\ \langle k \rangle &= Np \end{aligned}$$

We can do the same thing for $\langle k^2 \rangle$ to get:

$$\langle k^2 \rangle = \sum_{k=0}^N k^2 P(k | N-1, p)$$

$$\langle k^2 \rangle = N^2 p^2 - Np^2 + Np$$

Hence:

$$\text{var}(k) = \langle k^2 \rangle - \langle k \rangle^2 = N^2 p^2 - Np^2 + Np - N^2 p^2 = Np - Np^2 = \boxed{Np(1-p)}$$

1.3 Poisson Distribution

What if p was very small, but N was very large? Consider N people living in a town, each of whom (rarely) goes to a shop, independently with probability p . We use $N \rightarrow \infty$ and $p \rightarrow 0$, fixed by $Np = \lambda$.

$$P(k | N, p) = \frac{N!}{k!(N-k)!} p^k (1-p)^{N-k}$$

If $k \ll N$, then:

$$\frac{N!}{(N-k)!} = \frac{N \times (N-1) \times \cdots \times (N-k+1) \times \overbrace{(N-k) \times \cdots \times 1}^{(N-k)!}}{\underbrace{(N-k) \times \cdots \times 1}_{(N-k)!}}$$

$$N(N-1)(N-2)\dots(N-k+1) \approx \underbrace{N \cdot N \cdot \dots \cdot N}_{k \text{ times}} = N^k$$

Hence $N!(N-k)! \approx N^k$ (better and better approximation as N gets larger and larger):

$$P(k | N, p) = \frac{N^k}{k!} p^k (1-p)^{N-k}$$

And using $\lambda = np$:

$$P(k) = \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{N}\right)^{N-k}$$

Taking limits as $N \rightarrow \infty$:

$$\lim_{N \rightarrow \infty} \left(1 - \frac{x^N}{N}\right) = e^{-x} \quad \lim_{N \rightarrow \infty} \left(1 - \frac{x^a}{N}\right) = 1$$

Hence:

$$P(k | \lambda) \equiv \frac{\lambda^k}{k!} e^{-\lambda}$$

This gives us the probability mass function for the Poisson Distribution. It is an approximation for the binomial distribution for very large N . It has expectation value $\langle k \rangle = \lambda$, as the underlying binomial has expectation Np and $\lambda = Np$.

By the same trick, for the underlying binomial $\text{var}(k) = Np(1-p)$, so for very large N and very small p , $\text{var}(k) = \lambda$ too.

Thu 04 Dec 2025 09:00

Lecture P8 - Multivariate Distributions

For one variable, we can get $P(x)$:

$$\langle x \rangle = \sum_x xP(x)$$

For multiple variables, we get the same thing a slightly different way by marginalising to find $P(x)$ and using that:

$$\langle x \rangle \equiv \sum_x \sum_y xP(x, y) = \sum_x xP(x)$$

The two methods (direct sum including y or marginalising) are equivalent, they're just conceptually different.

1 Sums of Random Variables

We say x was drawn from some distribution $P(x)$ by writing:

$$x \sim P(x)$$

This tells us that x is a random variable, and follows the distribution given after the \sim .

Consider N random variables, all drawn from some distribution $P(x)$

$$x_1 \sim P(x) \quad x_2 \sim P(x) \dots \quad x_n \sim P(x)$$

We define the total:

$$t = x_1 + x_2 + x_3 + \dots + x_n$$

Working out $P(t)$ directly at this point is too difficult for us. We can however ask about the expected total $\langle t \rangle$ or $\text{var}(t)$. Recall the sample mean is:

$$\bar{x} = \frac{1}{N}(x_1 + \dots + x_n) = \frac{t}{N}$$

We already know expectation is linear, so $\langle x_1 + x_2 + \dots + x_n \rangle = \langle x_1 \rangle + \langle x_2 \rangle + \dots + \langle x_n \rangle$. The expectation value is given easily by:

$$\langle t \rangle = \langle x_1 + x_2 + \dots + x_n \rangle$$

For the variance, let's consider $N = 2$:

$$\begin{aligned} \text{var}(x_1 + x_2) &= \langle (x_1 + x_2)^2 \rangle + \langle x_1 + x_2 \rangle^2 \\ &= \langle x_1^2 + x_2^2 + 2x_1x_2 \rangle - (\langle x_1 + x_2 \rangle)^2 \\ &= \langle x_1^2 \rangle + \langle x_2^2 \rangle + 2\langle x_1x_2 \rangle - \langle x_1 \rangle^2 - \langle x_2 \rangle^2 - 2\langle x_1 \rangle \langle x_2 \rangle \\ &= \langle x_1^2 \rangle - \langle x_1 \rangle^2 + \langle x_2^2 \rangle - \langle x_2 \rangle^2 + 2(\langle x_1x_2 \rangle - \langle x_1 \rangle - \langle x_2 \rangle) \\ &= \text{var}(x_1) + \text{var}(x_2) + 2\text{cov}(x_1, x_2) \end{aligned}$$

Where $\text{cov}(x, y)$ is the covariance.

2 Covariance

Recall that:

$$\text{var}(x) = \sum_x (x - \langle x \rangle)^2 P(x)$$

We now define:

$$\begin{aligned} \text{cov}(x, y) &= \sum_{xy} (x - \langle x \rangle)(y - \langle y \rangle) P(x, y) \\ &= \langle (x - \langle x \rangle)(y - \langle y \rangle) \rangle = \langle xy \rangle - x\langle y \rangle - y\langle x \rangle + \langle x \rangle \langle y \rangle \\ &= \langle xy \rangle - \langle y \rangle \langle x \rangle - \langle x \rangle \langle y \rangle + \langle x \rangle \langle y \rangle \\ &= \langle xy \rangle - \langle x \rangle \langle y \rangle \end{aligned}$$

This is a generalisation for variance for multiple variables, so note that:

$$\text{var}(x) = \text{cov}(x, x)$$

Covariance measures linear association between two variables:

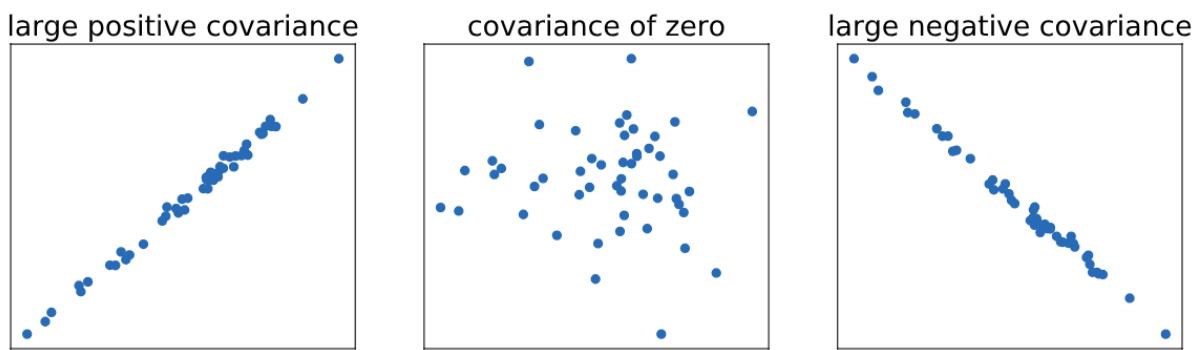


Figure 39.1

The caveat it's only linear is important - there may well be an association between two variables, but if it's not linear covariance won't test for it. For example, a true n -wavelengths of a sine wave will have zero covariance due to symmetry, but there is obviously an association (sinusoidal).

We define an additional measure called correlation using covariance (again, this is linear):

$$\text{corr}(x, y) = \frac{\text{cov}(x, y)}{\text{std}(x)\text{std}(y)}$$

This is useful, because it standardises, i.e. the maximum value for perfect linear association from the correlation function is 1. This helps us to standardise regardless of units (i.e. if we were testing the correlation of two measurements with two different units, changing one of the units from km to mm would change the covariance. This is not helpful, as the actual underlying correlation stays the same. The corr function avoids this)

3 Covariance and Independence

If x and y are independent, then we know:

$$P(x, y) = P(x, y)$$

How does this impact the covariance however?

$$\text{cov}(x, y) = \langle xy \rangle - \langle x \rangle \langle y \rangle$$

Assume they are independent, then:

$$\begin{aligned}\langle xy \rangle &= \sum_x \sum_y xy P(x)P(y) \\ &= \sum_x xP(x) \sum_y yP(y) \\ &= \langle x \rangle \langle y \rangle\end{aligned}$$

Hence if they are independent, then the covariance is zero (this does not go backwards).

3.1 Variance of Sum

If the variables are independent, the variance of the sum is the sum of the variances.

If they are not independent, we must include the covariance:

$$\text{var}(x_1 + x_2 + \dots + x_n) = \sum_n + 2 \sum_{m>n} \text{cov}(x_n, x_m)$$

As an example, what is the expectation value and the variance if we sum N independent Bernoulli variables, each with the same parameter p .

$$x_1 \sim \text{Bern}(p) \quad x_2 \sim \text{Bern}(p) \quad \dots \quad x_n \sim \text{Bern}(p)$$

For expectation values:

$$\langle x_1 + x_2 + \dots + x_n \rangle = \sum_{i=1}^N \langle x_i \rangle = Np$$

And for variance:

$$\text{var}(x_1 + \dots + x_n) = \sum_{i=1}^N \text{var}(x_i) = Np(1-p)$$

This is our results for the binomial distribution, as expected. Since we assume independence here, this is why the binomial distribution requires independent trials.

4 Change of Variables (Discrete)

We have some $P_x(x)$, in a sample space Ω_x . We make a transformation of x , $y = f(x)$.

For example, we have a particle moving with random velocity, and we want to the random variable that models kinetic energy. Kinetic energy is the new post-transformation random variable - the “induced distribution”.

What then is $P_y(y)$ or Ω_y ?

Consider a fair six-sided die:

$$P_x(x) = 1/6, \quad x = 1, 2, 3, 4, 5, 6$$

Example 1: $y = x - 2$ This changes the sample space but not the distribution shape itself. This is simple, as it's a bijection:

$$P_y(y) = \frac{1}{6} \quad y = -1, 0, 1, 2, 3, 4$$

Example 2: $z = |x - 2|$ This is a bit trickier, because it is not a bijection. Two different values of x ($x = 1, 3$) both map to $z = 1$. We therefore need to consider both cases to ensure the new PMF is still normalised:

$$P_z(z) = \begin{cases} 2/6 & z = 1 (x = 1, 3) \\ 1/6 & z = 0, 2, 3, 4 (x = 2, 4, 5, 6) \end{cases}$$

The general formula for a particular case of $P_y(y)$ is:

$$P_y(y) = \sum_{x:y=f(x)} P_x(x)$$

The $x : y = f(x)$ notation means that we are summing over values of x such that $y = f(x)$. In English, the formula is saying “to find the probability of the new variable having value y , sum the probabilities for all the x s that transform into that value of y ”.

Ω_y is the unique set of values that arise from $y = f(x)$, $\forall x \in \Omega_x$.

Fri 05 Dec 2025 11:00

Lecture P9 - Continuous Probability

So far, we have only discussed discrete probability. In this, $P(x)$ means the probability of x happening and we've ignored the possibility of an infinite number of possible events, because it never caused a problem.

1 Continuous Distributions

1.1 Continuous Random Walk

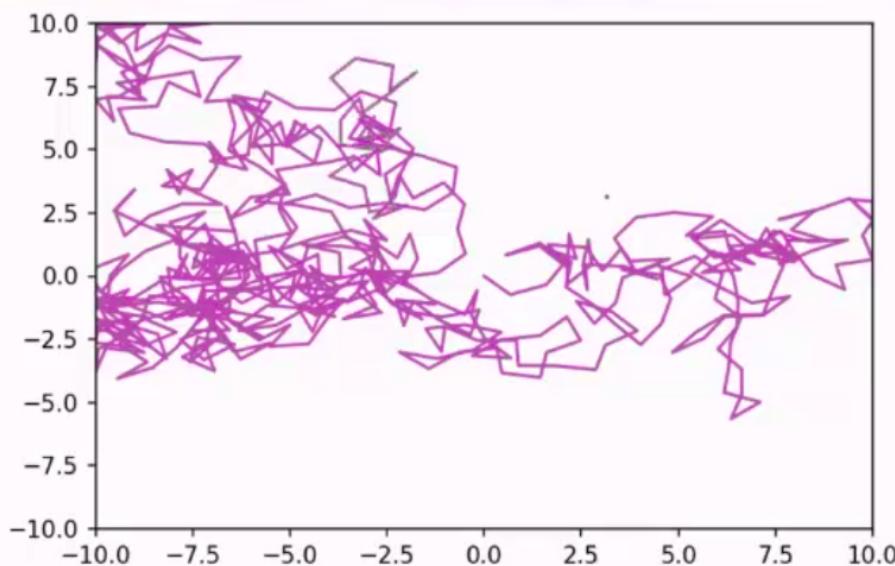


Figure 40.1: The path taken by the particle in a random walk.

We have a particle conducting a random walk. What is the probability of the particle landing exactly on (π, π) ? This is zero, and indeed it will be zero for every specific point we can name.

We have to think about probability a bit differently, and we have to ask the question about the particle being *near* (π, π) . We cannot consider exact values.

Continuous probability allows possible values to be real numbers (uncountably infinite) whereas before we were limited to discrete integers (countably infinite). Ω is therefore \mathbb{R} or a subset thereof.

We have to set our question to be “What is the probability that x lies in some interval”. Effectively, sums become integrals. $P(x)$ is now a probability density function, and the area under $P(x)$ is what gives us probability, rather than values of $P(x)$ alone.

1.2 Properties of a PDF

- $P(x) \geq 0$:
- $P(x) = 0, \forall x \notin \Omega$

- $P(x) > 0, \forall x \in \Omega$
 - It is normalised, hence:
- $$\int_{\Omega} P(x) dx = 1$$
- Since we care about the area under $P(x)$ to get probabilities, $P(x)$ itself may be bigger than 1, provided the integral is never bigger than 1.

All the old formulae still hold, but with integration instead of summation:

- Expectation Values:

$$\langle x \rangle = \int_{\Omega} x P(x) dx$$

- Expectation of a Function:

$$\langle f \rangle = \int_{\Omega} f(x) P(x) dx$$

- Variance:

$$\begin{aligned} \text{var}(x) &= \langle (x - \langle x \rangle)^2 \rangle \\ &= \int_{\Omega} (x - \langle x \rangle)^2 P(x) dx \\ &= \langle x^2 \rangle - \langle x \rangle^2 \end{aligned}$$

2 Cumulative Distributions

The definition of cumulative probability is the same:

$$C(x) \equiv \text{Probability}(X \leq x)$$

However we calculate it a bit differently as we can no longer simply sum the options below, we must integrate:

$$\begin{aligned} C(x) &= \int_{-\infty}^x P(x) dx \\ \frac{dC(x)}{dx} &= P(x) \end{aligned}$$

We can also use this to determine the probability of x being between two values, with:

$$P(a \leq x \leq b) = P(x \leq b) - P(x \leq a) = C(b) - C(a)$$

2.1 Example

A PDF is given by:

$$P(x) = \frac{3}{2}(1 - x^2) \quad 0 \leq x \leq 1$$

Find:

$$P\left(\frac{1}{4} \leq x \leq \frac{1}{2}\right)$$

$$\begin{aligned} P\left(\frac{1}{4} \leq x \leq \frac{1}{2}\right) &= C\left(\frac{1}{2}\right) - C\left(\frac{1}{4}\right) \\ C(x) &\equiv \int_0^x \frac{3}{2}(1 - x^2) dx \\ &= \frac{3x}{2} - \frac{x^3}{2} \end{aligned}$$

Hence $P\left(\frac{1}{4} \leq x \leq \frac{1}{2}\right) = C\left(\frac{1}{2}\right) - C\left(\frac{1}{4}\right)$ is easy (albeit faffy) to evaluate.

3 Change of Variables

As we did for discrete, let $x \sim P_x(x)$ and we set $y = f(x)$. What is $P_y(y)$?

This isn't needed for CoV to work, but for ease lets assume that $\Omega_x = [a, b]$, i.e. is some finite region. We need to conserve the normalisation, so:

$$1 = \int_a^b P_x(x) dx$$

We use the fact that $y = f(x) \implies x = f^{-1}(y)$. This is true only for monotonic functions. For this to work, we need $f(x)$ to have a unique inverse and is either always increasing or always decreasing. We can treat this as an integration subsitution, so:

$$dx = \frac{df^{-1}y}{dy} dy \implies 1 = \int_{f(a)}^{f(b)} \frac{df^{-1}}{dy} P_x(f^{-1}(y)) dy \equiv \int_{\Omega_y} P_y(y) dy$$

Where we call this new distribution the induced distribution. We have the general formula (again assuming monotonicity):

$$P_y(y) = \left| \frac{d}{dy} f^{-1}(y) \right| P_x(f^{-1}(y))$$

Note that the $|...|$ appear to ensure we handle the cases (where $f(x)$ is strictly increasing or strictly decreasing) as the strictly decreasing case will add an extra negative sign in the derivative. When integrating, we would swap the limits to cancel this extra negative, but we can't here as there's no integral, so taking the absolute value ensures that $P_y(y)$ is always positive.

Thu 10 Dec 2025 11:00

Lecture P10 - Example Distributions and Central Limit Theorem

1 Uniform Distribution

The continuous uniform distribution is a flat distribution between two points, a and b . The PDF is given by:

$$P(x | a, b) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

1.1 Expectation Value

$$\begin{aligned} \langle x \rangle &= \int_{\Omega} x P(x) dx \\ &= \frac{1}{b-a} \int_a^b x dx \\ &= \frac{1}{2} \frac{b^2 - a^2}{b-a} = \frac{1}{2} \frac{(b-a)(b+a)}{(b-a)} \\ \langle x \rangle &= \frac{b+a}{2} \end{aligned}$$

2 Exponential Distribution

The exponential distribution models exponential growth or decay. There are two different forms, depending on what units we give the parameter μ or λ :

$$P(x | \mu) = \begin{cases} \mu e^{-\mu x} & \text{if } 0 \leq x < \infty \\ 0 & \text{otherwise.} \end{cases}$$

$$P(x | \lambda) = \begin{cases} \frac{1}{\lambda} e^{-x/\lambda} & \text{if } 0 \leq x < \infty \\ 0 & \text{otherwise.} \end{cases}$$

2.1 Expectation Value

Considering the first form:

$$\langle x \rangle = \int_0^\infty \mu e^{-\mu x} dx$$

Letting $t = \mu x \implies dt = \mu dx$:

$$= \frac{1}{\mu} \int_0^\infty t e^{-t} dt$$

Solving by parts:

$$\begin{aligned} &= \frac{1}{\mu} \int_0^\infty \underbrace{t}_u \underbrace{e^{-t}}_{dv} dt \\ &= \frac{1}{\mu} [-te^{-t}]_0^\infty + \frac{1}{\mu} \int_0^\infty e^{-t} dt \end{aligned}$$

And, as boundary conditions (can do more formally via L'Hopital's rule) cause the first term to equal zero:

$$\begin{aligned} &= \int_0^\infty e^{-t} dt \\ \langle x \rangle &= \left[\frac{1}{\mu} - e^{-t} \right]_0^\infty = \frac{1}{\mu} \end{aligned}$$

3 Normal Distribution and the Central Limit Theorem

The Normal distribution (a.k.a Gaussian Distribution)

$$P(x | \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Properties:

- Expectation: $\langle x \rangle = \mu$,
- Variance: $\text{var}(x) = \sigma^2$.

To say that the random variable x follows a normal distribution we write:

$$x \sim \mathcal{N}(\mu, \sigma^2)$$

Or:

$$x \sim \mathcal{N}(x | \mu, \sigma^2)$$

3.1 Standard Normal

If $\mu = 0$ and $\sigma = \sigma^2 = 1$, we call the distribution the standard normal distribution.

$$P(x | 0, 1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

We can rescale any normal distribution onto the standard normal.

$$\begin{aligned} \text{let: } z &= \frac{x-\mu}{\sigma} \implies x = \sigma z + \mu \\ P_z(z) &= \left| \frac{d}{dz} f^{-1}(z) \right| P_x(f^{-1}(z)) \\ &= \sigma \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\sigma z + \mu - \mu}{2\sigma^2}\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \end{aligned}$$

3.2 CLT

The Central Limit Theorem states that rescaled sums of standard variables appear to be Normally distributed. We will not prove it formally.

Let:

$$x_1 \sim P(x) \quad x_2 \sim P(x) \quad \dots \quad x_n \sim P(x)$$

We define:

$$X = x_1 + x_2 + \dots + x_n$$

We now form the **rescaled** variable:

$$z = \frac{X - \langle X \rangle}{\text{std}(X)}$$

Assuming independence, we have already shown that:

$$\langle X \rangle = N \langle x \rangle$$

$$\text{var}(X) = N \text{var}(x)$$

Hence:

$$z = \frac{X - N \langle x \rangle}{\sqrt{N} \text{std}(x)}$$

The CTL states that:

$$\lim_{N \rightarrow \infty} z = \mathcal{N}(0, 1) \quad \text{if } \text{std}(x) < \infty$$

This is regardless of what $P(x)$ is, even if it is not normal.

It also says that averages appear Normal. Consider the sample mean:

$$\bar{x} = \frac{X}{N} = \frac{1}{N}(x_1 + x_2 + \dots + x_n)$$

$$\langle \bar{x} \rangle = \left\langle \frac{X}{n} \right\rangle = \frac{1}{N} \langle X \rangle = \frac{1}{N} N \langle x \rangle = \langle x \rangle$$

And:

$$\text{var}(\bar{x}) = \frac{1}{N^2} \text{var}(X) = \frac{1}{N} \text{var}(x)$$

Again lets define a standard variable:

$$z = \frac{\bar{x} - \langle x \rangle}{\text{std}(\bar{x})/\sqrt{N}} \rightarrow \mathcal{N}(0, 1)$$

We can invert this to get:

$$\begin{aligned} \bar{x} &= \frac{\text{std}(x)}{\sqrt{N}} z + \langle x \rangle \\ \bar{x} &\sim \mathcal{N}\left(\langle x \rangle, \frac{\text{var}(x)}{N}\right) \\ \bar{x} &\sim \mathcal{N}\left(\mu, \frac{\sigma^2}{N}\right) \end{aligned}$$

This tells us that when we construct a sample mean from something, we can treat it as a Normal distribution (regardless of the source distribution). The variance is a decreasing function of N , so taking more data creates a distribution better centred on the sample mean.

3.3 Example

The Erlang Distribution has PDF:

$$P(x | \lambda, k) = \frac{\lambda^x x^{k-1} e^{-\lambda x}}{(k-1)!}$$

With:

$$\langle x \rangle = \frac{k}{\lambda}$$

$$\text{var}(x) = \frac{k}{\lambda^2}$$

What is the limiting distribution of the sum of N Erlang distributed random variables, if N is large?

Let:

$$X = x_1 + x_2 + \cdots + x_n \quad \text{where all are Erlang distributed random variables}$$

We have:

$$\langle x \rangle = N \frac{k}{\lambda} \quad \text{var}(x) = N \frac{k}{\lambda^2}$$

Creating our rescaled variable:

$$z = \frac{X - \langle X \rangle}{\text{std}(x)} = \frac{X - (Nk/\lambda)}{\sqrt{\frac{Nk}{\lambda^2}}} \xrightarrow{\text{C.L.T.}} \mathcal{N}(0, 1)$$

And inverting and solving for X :

$$X = \sqrt{\frac{Nk}{\lambda^2}} z + \frac{Nk}{\lambda}$$

So:

$$X \sim \mathcal{N}\left(\frac{Nk}{\lambda}, \frac{Nk}{\lambda^2}\right)$$

Fri 12 Dec 2025 11:00

Lecture P11 - End of Probability: Variance Propagation

1 Variance Propagation

Say we have $x \sim P_x(x)$, but we don't know what the actual distribution $P_x(x)$ is. We have however been able to determine (or estimate) values for $\langle x \rangle$ and $\text{var}(x)$. We set $y = f(x)$, and what is $P_y(y)$? This is a common scenario in e.g. labs, where we can measure a sample mean and estimate a standard deviation, but we cannot necessarily actually determine the underlying distribution.

We would formally use:

$$P_y(y) = \left| \frac{d}{dy} f^{-1}(y) \right| P_x(x)(f^{-1}(y))$$

But that is impossible here, as we do not know $P_x(x)$. We can however calculate values for $\langle y \rangle$ and $\text{var}(y)$.

1.1 Example

Imagine an experiment where we want to determine kinetic energy. This is difficult (as we would have to transfer this energy into another easier to measure form first). Instead, we measure velocity and use:

$$E = \frac{1}{2}mv^2 \quad \text{where } v \text{ is a random variable.}$$

What is $\langle E \rangle$ and $\text{var}(E)$?

We use Taylor Series:

$$f(x) = f(x_0) + (x - x_0)f'(x_0) + \frac{(x - x_0)^2}{2!}f''(x_0) + \dots$$

This generates a polynomial approximation for $f(x)$ about the point x_0 . What if x is a random variable? Lets assume $\langle x \rangle = \mu$ and $\text{var}(x) = \sigma^2$.

$$f(x) = f(\mu) + (x - \mu)f'(\mu) + \frac{(x - \mu)^2}{2}f''(\mu) + \dots$$

We have effectively decoupled the f and the x , allowing us to work with it more easily. We take expectations:

$$\langle f(x) \rangle \approx \langle f(\mu) \rangle + \langle (x - \mu)f'(\mu) \rangle + \frac{1}{2}\langle (x - \mu)^2f''(\mu) \rangle + \dots$$

$$\langle f(\mu) \rangle = f(\mu)$$

$$\langle (x - \mu)f'(\mu) \rangle = 0$$

$$\frac{1}{2}\langle (x - \mu)^2f''(\mu) \rangle = \frac{\sigma^2}{2}f''(\mu)$$

We know the second term is zero, and if we assume the third term and onwards are 'small':

$$\langle f \rangle \approx f(\mu)$$

We therefore use $f(\langle x \rangle)$ as an approximation for $\langle f \rangle$.

Now, for variance:

$$\text{var}(f) = \langle f^2 \rangle - \langle f \rangle^2 \approx \langle f^2 \rangle - (f(\mu))^2$$

Approximating to the first order with $f(x) \approx f(\mu) + (x - \mu)f'(\mu)$:

$$\begin{aligned} \langle f^2 \rangle &\approx \langle (f(\mu) + (x - \mu)f'(\mu))^2 \rangle - \langle f(\mu)^2 \rangle + 2f(\mu)f'(\mu)\langle (x - \mu) \rangle + f'(\mu)^2\langle (x - \mu)^2 \rangle \\ &\approx f(\mu)^2 + \sigma^2 f'(\mu)^2 \\ \implies \text{var}(f) &\approx \sigma^2 f'(\mu)^2 \end{aligned}$$

Going back to the example:

Say we've measured $\langle v \rangle = v_0$ and $\text{var}(v) = \sigma^2$. We want to estimate $\langle E \rangle$ and $\text{var}(E)$.

$$\begin{aligned} E &= \frac{1}{2}mv^2 \\ \langle E \rangle &\approx \frac{1}{2}m\langle v \rangle^2 \\ \text{var}(E) &\approx \left(\frac{dE}{dv} \right)^2 \sigma^2 \\ &= m^2v^2\sigma^2 \end{aligned}$$

Here, v is a ‘smooth’ function with a constant second derivative, so the third term of the Taylor Expansion before is indeed small and this is a decent approximation, but it may not always be.

2 Example II

If x is drawn according to a Poisson distribution with parameter λ , estimate the value of:

$$\left\langle \frac{1}{\sqrt{1+x}} \right\rangle$$

And:

$$\text{var}\left(\frac{1}{\sqrt{1+x}}\right)$$

Note that:

$$\begin{aligned} \langle x \rangle &= \text{var}(x) = \lambda \\ \frac{d}{dx} \frac{1}{\sqrt{1+x}} &= -\frac{1}{2(1+x)^{3/2}} \end{aligned}$$

From the definition of discrete expectation values:

$$\left\langle \frac{1}{\sqrt{1+x}} \right\rangle = \sum_{x=0}^{\infty} \frac{1}{\sqrt{1+x}} \frac{\lambda^x}{x!} e^{-\lambda}$$

This is...unpleasant, so we need the approximation to reasonably determine it.

$$\begin{aligned} \langle f(x) \rangle &\approx f(\langle x \rangle) = \frac{1}{\sqrt{1+\langle x \rangle}} = \frac{1}{\sqrt{1+\lambda}} \\ \text{var}\left(\frac{1}{\sqrt{1+x}}\right) &= \left(\frac{d}{dx} \frac{1}{\sqrt{1+x}} \right)^2 \text{var}(x) \\ &= \frac{1}{4(1+\lambda)^3} \lambda \end{aligned}$$

3 General Formulae

If we have N statistically independent random variables, $x_1, x_2, x_3, \dots, x_n$ with:

$$\langle x_i \rangle = \mu_i \quad \text{var}(x_i) = \sigma_i^2$$

For some function $f(x_1, x_2, \dots, x_n)$:

$$\langle f \rangle \approx f(\mu_1, \mu_2, \dots, \mu_n)$$

$$\text{var}(f) \approx \left(\frac{\partial f}{\partial x_1} \right)^2 \sigma_1^2 + \left(\frac{\partial f}{\partial x_2} \right)^2 \sigma_2^2 + \dots + \left(\frac{\partial f}{\partial x_n} \right)^2 \sigma_n^2$$

We estimate the values of μ_i, σ_i^2 using sample values, i.e. in labs.

3.1 Multivariate Example

Again we want to determine kinetic energy, but this time we measure both m and v with:

$$\langle v \rangle = v_0 \quad \langle m \rangle = m_0$$

$$\text{var}(v) = \sigma_v^2 \quad \text{var}(m) = \sigma_m^2$$

$$E = \frac{1}{2}mv^2$$

We again want to estimate $\langle E \rangle$ and $\text{var}(E)$.

$$\begin{aligned} \langle E \rangle &\approx \frac{1}{2}\langle m \rangle \langle v \rangle^2 = \frac{1}{2}m_0v_0^2 \\ \text{var}(E) &\approx \left(\frac{dE}{dm} \right)^2 \text{var}(m) + \left(\frac{dE}{dv} \right)^2 \text{var}(v) \\ &= \frac{v_0^4}{4}\sigma_m^2 + m_0^2v_0^2\sigma_v^2 \end{aligned}$$

End of Module.

LC Mathematics for Physicists 1A

Mon 29 Sep 2025 12:00

Lecture 1 - Course Welcome and Introduction to Vectors

Recommended Course Books

- *Mathematical Techniques...* by DW Jordan and Smith, 3rd Edition
 - Closest to the course.
- *Engineering Mathematics* by Stroud.
 - Lots of examples and extremely clear.
- *Mathematical Methods for the Physical Sciences* by Mary Boas (Wiley).
 - Succinct and faster.
- *Elementary Vector Analysis* by Weatherburn.
 - Vector specific book.

1 Intro To Vectors

Some physical quantities can be represented by numbers, i.e. charge, speed, time, etc. Some other quantities have an associated direction as well as a magnitude, for example velocity, acceleration, position, electric fields, etc.

We call the single-number quantities scalars, and the directional quantities vectors. Some quantities (such as moment of inertia, covered in CMR2) depend on more than one direction, we call these *tensors*.

We will initially deal with vectors geometrically, in terms of points and the vectors which arise from them. Consider two points (in 2D for now). We label one as the origin O and one as point A , we define the vector \vec{OA} as being the distance and direction from O to A .

The vector has a “sense” from O to A , we could also have the vector \vec{AO} which would point in the opposite direction:

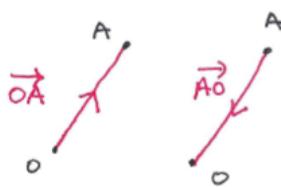


Figure 43.1

We denote the magnitude (length) of a vector as $|\vec{OA}|$ which is the distance from O to A .

Once we define a vector, we can move it anywhere we want, and it is not constrained as having to actually start at O and end at A . We can ‘liberate’ the vector from its initial points and translate it anywhere we want in the plane, and it is still the same vector provided the magnitude and direction do not change. Once we’ve used O and A to define \vec{OA} , we can copy the vector anywhere.

2 Vector Operations

We now want to define standard mathematical operations for vectors, starting with addition.

2.1 Vector Addition

Consider three points, O, A, B . We can naturally define $\overrightarrow{OA}, \overrightarrow{OB}, \overrightarrow{AB}$

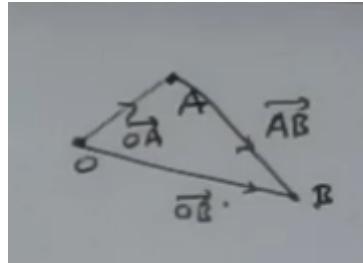


Figure 43.2

We define the following:

$$\overrightarrow{OA} + \overrightarrow{AB} = \overrightarrow{OB}$$

2.2 Vector Subtraction

We note that \overrightarrow{OA} and \overrightarrow{AO} are the same vector with opposite directions, therefore:

$$\begin{aligned}\overrightarrow{OA} + \overrightarrow{AO} &= \overrightarrow{O} \\ \overrightarrow{OA} &= -\overrightarrow{AO}\end{aligned}$$

Hence \overrightarrow{OA} and \overrightarrow{AO} are inverses of each other, just like 2 and -2 are for numbers. We consider our equation from addition and add \overrightarrow{BO} to both sides:

$$\begin{aligned}\overrightarrow{OA} + \overrightarrow{AB} + \overrightarrow{BO} &= \overrightarrow{OB} + \overrightarrow{BO} \\ \overrightarrow{OA} + \overrightarrow{AB} - \overrightarrow{OB} &= 0\end{aligned}$$

Which gives us an example of subtraction.

3 Better Notation

We use the following improved and more compact notation, which is especially useful if we have some number of several points relative to an origin.

Given the points A, B, C we have so far written $\overrightarrow{OA}, \overrightarrow{OB}, \overrightarrow{OC}$. Instead, we now write:

$$\begin{aligned}\overrightarrow{OA} &= \mathbf{a} \\ \overrightarrow{OB} &= \mathbf{b} \\ \overrightarrow{OC} &= \mathbf{c}\end{aligned}$$

Note that vectors are handwritten with an underline, so \underline{a} , but may instead be typed in bold font, so \mathbf{a} .

Say O is the centre of mass of n particles, we can also use the notation $r_1, r_2, r_3, \dots, r_n$ to denote the position vectors of these particles wrt the origin O .

3.1 Physical Example

Recall the definition of equilibrium:

- $\mathbf{F} = m\mathbf{a}$, for equilibrium $\mathbf{F} = 0$.
- Zero net moment.

Say we have a particle of mass m hanging by a single spring. The spring exerts some tension force \mathbf{T} , while the mass has a weight force $\mathbf{F}_g = m\mathbf{g}$.

For this particle to be in equilibrium, we must have:

$$\mathbf{T} + m\mathbf{g} = 0$$

$$\mathbf{T} = -m\mathbf{g}$$

Simple enough! Consider an example where two springs connect to the particle and connect to some ceiling. We now have:

$$\mathbf{T}_1 + \mathbf{T}_2 + m\mathbf{g} = 0$$

Again, simple.

4 Midpoints

Consider a triangle with points (clockwise), O , A , and B . Say we want the position vector of the midpoint of AB , denoted M .

Wed 01 Oct 2025 09:00

Lecture 2

Thu 02 Oct 2025 16:00

Lecture 3

Mon 06 Oct 2025 12:00

Lecture 4

Wed 08 Oct 2025 09:00

Lecture 5

Thu 09 Oct 2025 16:00

Lecture 6

Mon 13 Oct 2025 12:00

Lecture 7

Wed 15 Oct 2025 09:00

Lecture 8

Thu 16 Oct 2025 16:00

Lecture 9

Mon 20 Oct 2025 12:00

Lecture 10

Wed 22 Oct 2025 09:00

Lecture 11

Thu 23 Oct 2025 16:00

Lecture 12

Mon 27 Oct 2025 12:00

Lecture 13

Wed 29 Oct 2025 09:00

Lecture 14

Thu 30 Oct 2025 16:00

Lecture 15

Mon 03 Nov 2025 12:00

Lecture 16

Wed 05 Nov 2025 09:00

Lecture 17

Thu 06 Nov 2025 16:00

Lecture 18

Mon 10 Nov 2025 12:00

Lecture 19

Wed 12 Nov 2025 09:00

Lecture 20

Thu 13 Nov 2025 16:00

Lecture 21

Mon 17 Nov 2025 12:00

Lecture 22

Wed 19 Nov 2025 09:00

Lecture 23

Thu 20 Nov 2025 16:00

Lecture 24

Mon 24 Nov 2025 12:00

Lecture 25

Wed 26 Nov 2025 09:00

Lecture 26

Thu 27 Nov 2025 16:00

Lecture 27

Mon 01 Dec 2025 12:00

Lecture 28

Wed 03 Dec 2025 09:00

Lecture 29

Thu 04 Dec 2025 16:00

Lecture 30

Mon 08 Dec 2025 12:00

Lecture 31

Wed 10 Dec 2025 09:00

Lecture 32

Thu 11 Dec 2025 16:00

Lecture 33

LC Optics and Waves

Wed 01 Oct 2025 11:00

Lecture 1 - Intro to Waves and SHM Recap

0.1 Course Objectives

- Have a sound understanding of basic wave properties
- Have a basic understanding of interference effects, inc diffraction
- Be able to use simple geometric optics and understand the fundamentals of optical instruments.

0.2 Recommended Textbooks

1. University Physics, Young and Freedman (Ch 15, 16 for Waves, Ch 33-36 for Optics)
2. Physics for Scientists and Engineers (Ch 20, 21 for Waves, Ch 22-24 for Optics)
3. 5e, Tipler and Mosca, (Ch 15, 16 for Waves, 31-33 for Optics)
4. Fundamentals of Optics, Jenkins and White
5. Optics, Hecht and Zajac

What is a wave? Waves occur when a system is disturbed from equilibrium and the disturbance can travel from one region to another region. Waves carry energy, but do not move mass. The course aim is to derive basic equations for describing waves, and learn their physical properties.

0.3 Periodic Motion

Waves are very linked to periodic motion. Therefore we recap periodic motion first.

It has these characteristics:

- A period, T (the time for one cycle)
- A frequency, f , the number of cycles per unit time ($f = \frac{1}{T}$)
- An amplitude, A , the maximum displacement from equilibrium.

Periodic motion continues due to the restoring force. When an object is displaced from equilibrium, the restoring force acts back towards the equi point. The object reaches equi with a non-zero speed, so the motion continues past the equi point and continues forever.

0.4 Energy

Periodic motion is an exchange between potential and kinetic energy, with no energy loss. Energy is conserved.

0.5 Simple Harmonic Motion

If the restoring force is directly proportional to the displacement $F = -kx$, then the periodic motion becomes Simple Harmonic Motion and the object is called a harmonic oscillator.

In a single dimension, displacement is given by:

$$x = A \cos(\omega t + \phi)$$

Where $\omega = 2\pi f$ is the angular velocity, and ϕ is the phase angle. In cases like this, where the phase angle is 90 deg we can simplify to $x = -A\sin(\omega t)$

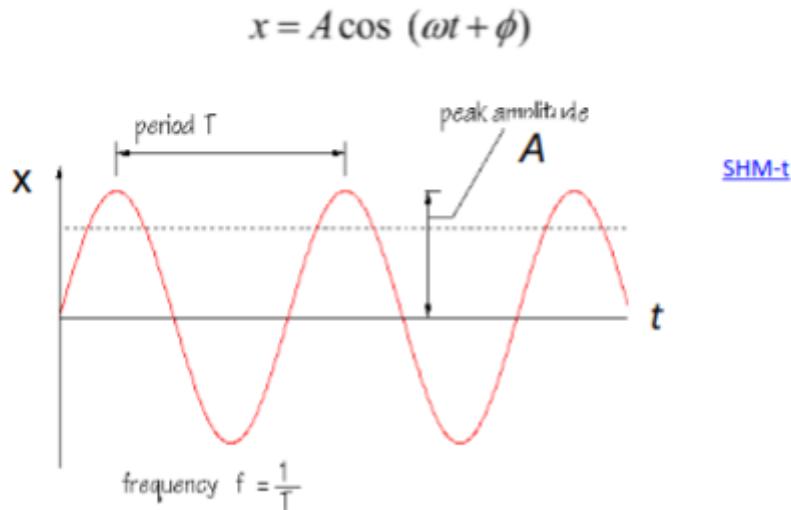


Figure 76.1: A Phase Angle of 90

0.6 More SHM Equations

Velocity

$$v_x = \frac{dx}{dt} = -\omega A \sin(\omega t + \phi)$$

Acceleration

$$a_x = \frac{dv_x}{dt} = \frac{d^2x}{dt^2} = -\omega^2 A \cos(\omega t + \phi)$$

Both properties are signed to indicate direction, as they are both vectors.

Thu 02 Oct 2025 13:00

Lecture 2 - Wave Functions

0.1 Sine Waves

Mechanical Waves A mechanical wave is a disturbance through a medium. It's formed of a single wave pulse or a periodic wave.

Mechanical Waves have the following properties:

1. Transverse: Where displacement of the medium is perpendicular to the direction of propagation.
2. Longitudinal, displacement of the medium is in the same direction as propagation.
3. Propagation depends on the medium the wave moves through (i.e. density, rigidity)
4. The medium does not travel with the wave.
5. Waves have a magnitude and a direction.
6. The disturbance travels with a known exact speed.
7. Waves transport energy but not matter throughout the medium.

0.2 Wave Functions

We want to define a wave function in terms of two variables, x and t . In any given moment, if we consider a single point on the wave (i.e. $t = 0$), and wait a short while, the wave will have travelled to some $t = t_1 > 0$.

In order to quantify displacement, we therefore want to specify both the time, and the displacement. This will let us find the wave speed, acceleration and the (new) wave number.

We are also able to talk about the velocity and acceleration of individual particles on the wave.

0.3 Wave Function for a Sine Wave

Consider a sine wave. We want to find a wave function in the form $y(x, t)$. Consider the particle at $x = 0$.

We can express the wave function at this point as $y(x = 0, t) = A \cos \omega t$. However we want to expand this to any general point. Now consider a point (2) which is one wavelength away. We know the behaviour of particle 1 is mirrored by particle 2 (with a time lag).

Since the string is initially at rest, it takes on period (T) for the propagation of the wave to reach point 2, therefore point 2 is lagging behind the motion of point 1. The wave equation is therefore (if particle two has $x = \lambda$) $y(x = \lambda, t) = A \cos(\omega t - 2\pi)$.

For arbitrary x , $y(x, t) = A \cos(\omega t - \frac{x}{\lambda} \cdot 2\pi)$ to account for this delay. This quantity is called the wave number:

$$\text{Wave Number: } k = \frac{2\pi}{\lambda}$$

So:

$$\begin{aligned} y(x, t) &= A \cos(\omega t - kx) \\ &= A \cos(kx - \omega t) \end{aligned}$$

Note the second step is possible as \cos is an even function. k can also be signed to indicate direction: if $k > 0$, the wave travels in the positive x . If $k < 0$, the wave travels in the negative x direction. Again, $\omega = 2\pi f$

0.4 Displacement Stuff

Considering a point (starting at equi), the time taken for the particle on the sin wave to reach maximum displacement, minimum displacement and back takes the time period T . The speed of the wave is distance travelled over the time taken. We take the distance to be the wavelength λ , as we know the time by definition this takes is one time period T . Therefore wave speed v is:

$$v = \frac{\lambda}{T} = \lambda f$$

Since $\lambda = \frac{2\pi}{k}$ and $f = \frac{\omega}{2\pi}$ (as ω is defined as $\frac{2\pi}{T}$), we can also write:

$$v = \frac{2\pi}{k} \cdot \frac{\omega}{2\pi} = \frac{\omega}{k}$$

0.5 Particle Velocity

We can also determine the velocity of individual particles in the medium. We can use this to determine the acceleration.

We know that

$$y(x, t) = A \cos(kx - \omega t)$$

The vertical velocity v_y is therefore given by:

$$v_y = \frac{dy(x, t)}{dt}$$

Which is unhelpful (as we can't differentiate two variables at once), we can slightly cheat this by looking at purely a certain value of x , and therefore treating x as constant (to get a single variable derivative).

$$v_y = \left. \frac{dy(x, t)}{dt} \right|_{x=\text{const.}}$$

However this is notationally yucky, so we therefore use the notation:

$$\frac{\partial y(x, t)}{\partial t}$$

To represent the same thing. Finally (carrying out the partial derivative):

$$v_y(x, t) = \frac{\partial y(x, t)}{\partial t} = \omega A \sin(kx - \omega t)$$

0.6 Particle Acceleration

We can work out particle acceleration (transverse acceleration) by differentiating in the same manner again:

$$a_y(x, t) = \frac{\partial^2 y(x, t)}{\partial t^2} = \frac{\partial}{\partial t} \left(\frac{\partial y(x, t)}{\partial t} \right) = -\omega^2 A \cos(kx - \omega t) = -\omega^2 y(x, t).$$

Wed 08 Oct 2025 11:00

Lecture 3 - Generalised Wavefunctions

Recap: For a sine wave, the wavefunction is:

$$y(x, t) = A \cos(kx - \omega t)$$

Where k is the wave number, $k = \frac{2\pi}{\lambda}$ and omega is $\frac{2\pi}{T} = 2\pi f$

1 More Wavefunctions

What is the general form of a triangular shaped wave? What about a square stepped wave? TODO: Diagram

We know that each particle (i.e. along a string) copies the motion of its immediate left-hand neighbour (for a particle moving in the positive x), with a time delay proportional to their distance. Every wave is describable $y(x, t)$ wave function, and every mechanical wave relies on a medium (i.e. a piece of string or water, concrete etc) to travel.

Ideally, we want to be able to find a general form of a wave function.

In a single fixed instant of time, the moving wave pulse is stationary, so purely a function of $y = f(x)$. We want a wave function where we can input any value of t , so we need a moving frame of reference. We define this frame of reference as O' (for the origin) and x', y' axes. This frame of reference moves with the wave pulse and at the same speed, therefore y' is a function of x' only, independant of speed.

1.1 New System

x is the distance from the origin O to the relevant point, while x' is the distance from O' .

$$x = x' + vt$$

$$x' = x - vt$$

$$y' = f(x') = f(x - vt)$$

However, as the wave is moving purely in one direction (along x), $y = y'$, so:

$$y = f(x - vt)$$

1.2 Back to Basics

Going back to:

$$\begin{aligned} y(x, t) &= A \cos(kx - \omega t) \\ &= A \cos\left[k\left(x - \frac{\omega}{k}t\right)\right] \\ &= A \cos[k(x - vt)] \end{aligned}$$

(Note, this is true for a wave in the positive x, for a wave moving in the negative x this would be $x + vt$)

1.3 Equivalent Representations

There are some equivalent representations for a sine wave:

$$y(x, t) = A \cos\left(2\pi f \frac{x}{v} - \omega t\right) = A \cos\left(2\pi \left[\frac{x}{\lambda} - \frac{t}{T}\right]\right) = \text{TODO}$$

2 More Differentiation

We've already looked at differentiating with respect to t , but what about x ? This would give us the slope of the string at that point:

$$\frac{\partial y(x, t)}{\partial x}$$

And the curvature of the string:

$$\frac{\partial^2 y(x, t)}{\partial x^2} = -k^2 A \cos(kx - \omega t) = -k^2 y(x, t)$$

Note the similarities here with the equation for transverse acceleration:

$$(1) \frac{\partial^2 y(x, t)}{\partial t^2} = -\omega^2 y(x, t)$$

$$(2) \frac{\partial^2 y(x, t)}{\partial x^2} = -k^2 y(x, t)$$

Dividing (1) by (2):

$$\frac{\partial^2 y(x, t)/\partial t^2}{\partial^2 y(x, t)/\partial x^2} = \frac{\omega^2}{k^2} = v^2$$

Therefore:

$$\frac{\partial^2 y(x, t)}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 y(x, t)}{\partial t^2}$$

We call this the 'wave equation' as every wave function $y(x, t)$ must satisfy it, regardless of whether or not it is periodic or its direction of travel. If $y(x, t)$ does not satisfy this, it is not a wave function.

2.1 An Example

$$y(x, t) = \frac{x^3 - vt^2}{e^t}$$

TODO, the example in our own time

3 Wave Equation for a String

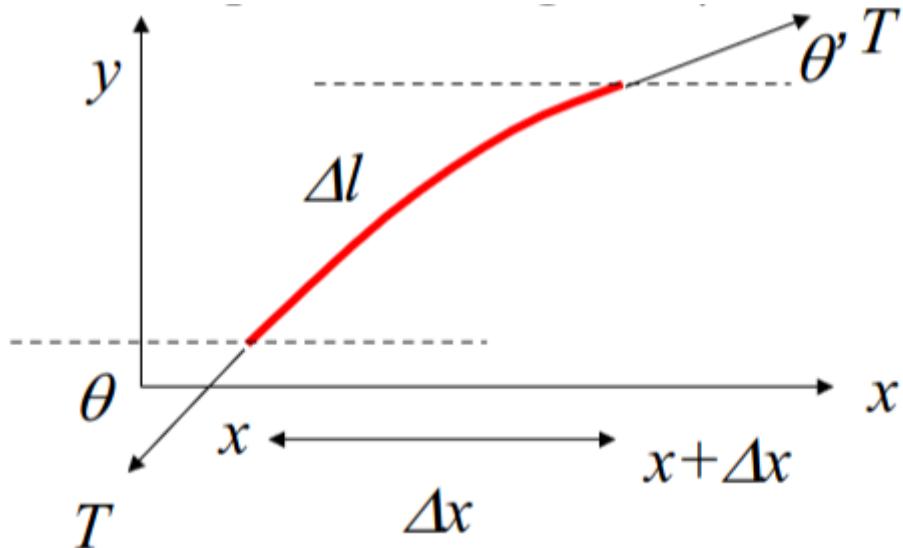


Figure 78.1: A snapshot of the wave.

Lets say we have some string, suspended horizontally under tension. We generate a single wave pulse and allow it to propagate down the string.

We assume the string is 1D, under tension T (constant throughout) and has mass per unit length μ .

Consider a small segment of string of length ΔL from x to $x + \Delta x$. This string makes angle θ with the horizontal at the bottom of the string, and angle θ' with the horizontal at the top of the string.

Net force (transverse in y) is:

$$F_y = T \sin \theta' - T \sin \theta$$

And using the small angle approx $\sin \theta \approx \tan \theta = \frac{dy}{dx}$:

$$F_y = T \left(\frac{dy}{dx} \Big|_{x+\Delta x} - \frac{dy}{dx} \Big|_x \right)$$

Using differentiation by first principles:

$$\frac{dy}{dx} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

We can say:

$$\frac{d^2y}{dx^2} = \frac{d}{dx} \left(\frac{dy}{dx} \right) = \lim_{\Delta x \rightarrow 0} \frac{\frac{dy}{dx} \Big|_{x+\Delta x} - \frac{dy}{dx} \Big|_x}{\Delta x}$$

Therefore:

$$F_y = T \frac{d^2y}{dx^2} \Delta x$$

The mass of this section of string is $\mu \Delta x$, and considering the acceleration in the y direction we can plug into $F = ma$ to get:

$$F = ma$$

$$T \frac{d^2y}{dx^2} \Delta x = \mu \Delta x \frac{d^2y}{dt^2}$$

$$\frac{d^2y}{dx^2} = \frac{\mu}{T} \frac{d^2y}{dt^2}$$

The LHS is the rate of change of the string's gradient, which as mentioned is the curvature of the string. The RHS includes the transverse acceleration of the string, therefore acceleration is proportional to curvature. To evaluate this at a fixed time/position we should write:

$$\frac{\partial^2 y}{\partial x^2} = \frac{\mu}{T} \frac{\partial^2 y}{\partial t^2}$$

And comparing to the wave equation we get:

$$\frac{1}{v^2} \frac{\partial^2 y}{\partial t^2} = \frac{\mu}{T} \frac{\partial^2 y}{\partial t^2}$$

$$\frac{1}{v^2} = \frac{\mu}{T}$$

$$v = \sqrt{\frac{T}{\mu}}$$

So a wave travels faster under a higher tension with a lower mass per unit length.

Thu 09 Oct 2025 13:00

Lecture 4 - Waves at Boundaries

1 Recap

We previously derived:

$$v = \sqrt{\frac{T}{\mu}}$$

And we also have:

$$v = \frac{\omega}{k}$$

The former is useful explicitly for a wave travelling over a string, while the latter is applicable to the movement of any wave. On a string, a higher tension yields a higher restoring force and therefore a higher speed, while a higher mass per unit length gives a higher mass for some arbitrary length of string, therefore a lower acceleration and lower speed.

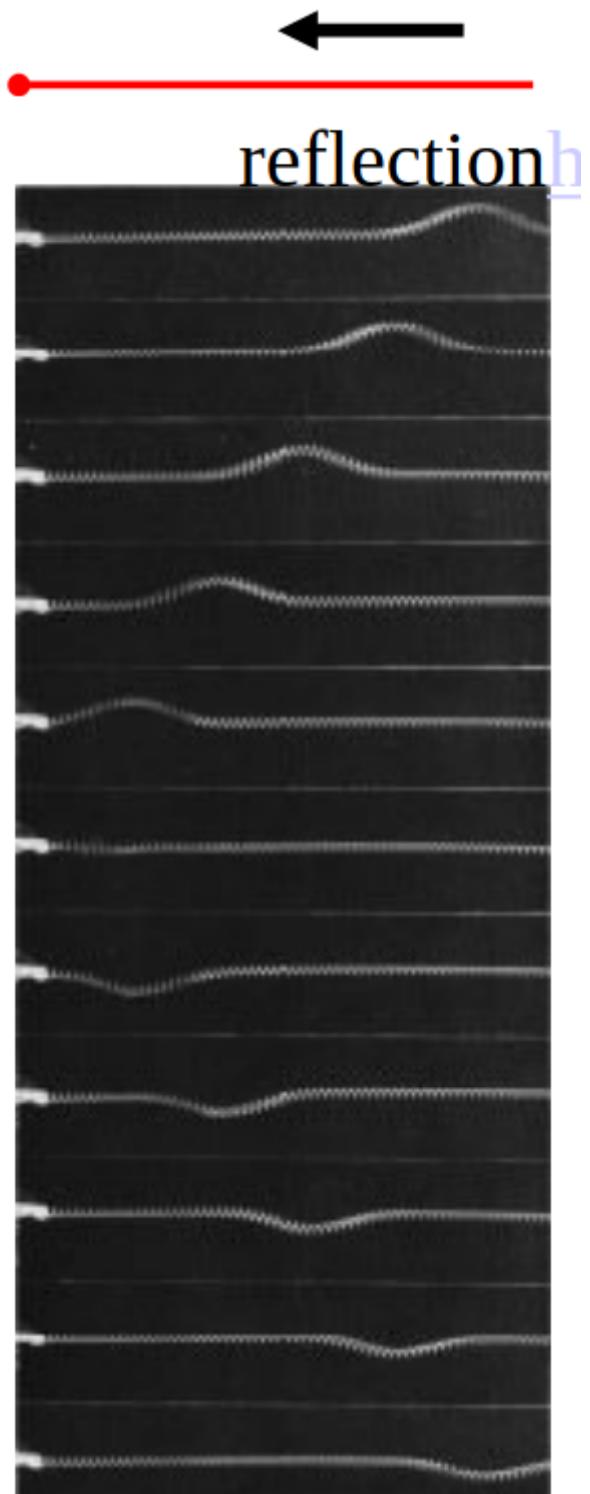
2 A Quick Interlude

The speed of a mechanical wave has the general form:

$$v = \sqrt{\frac{\text{Restoring force returning to equilibrium}}{\text{Inertia resisting return to equilibrium}}}$$

3 Reflection

When a wave hits a fixed boundary, it is reflected and inverted. Lets consider a case where a string is fixed on the LHS and is reflected back:



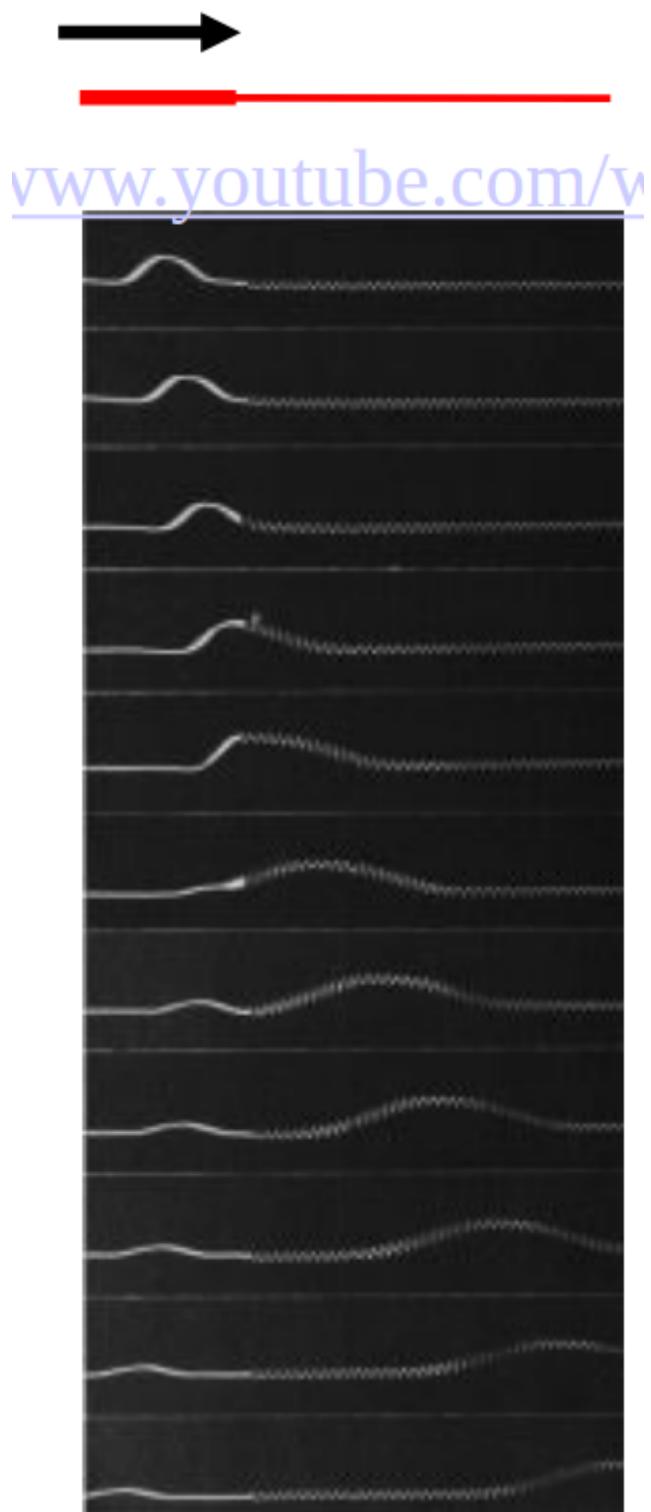


Figure 79.1: Or with two strings, going from thick to thin:

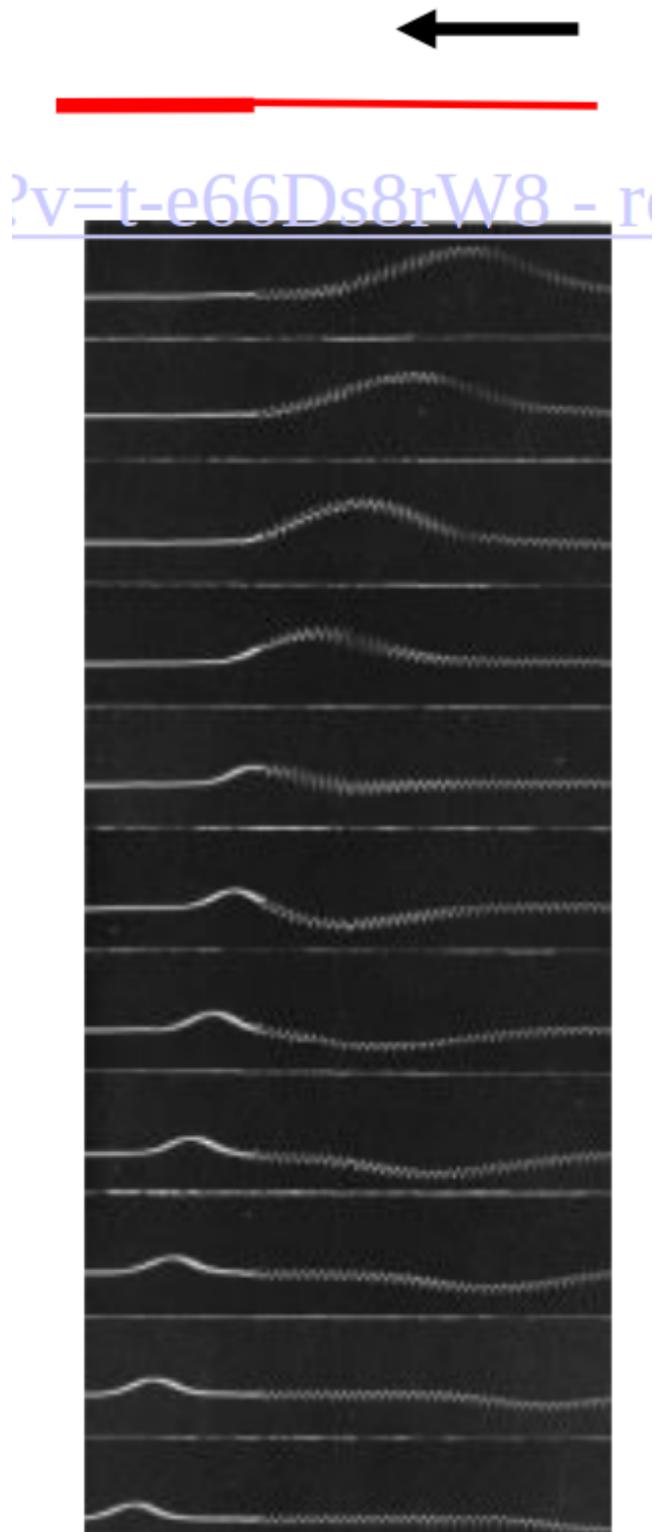


Figure 79.2: Or from thin to thick.

4 Waves Interacting at Boundaries

Say we have two pieces of string connected to each other, one thin string with mass per unit length μ_1 and a thicker string with μ_2 (both under the same tension, T). If a wave pulse is passed along from the thin string to the thicker string, what happens at the point of connection, P ?

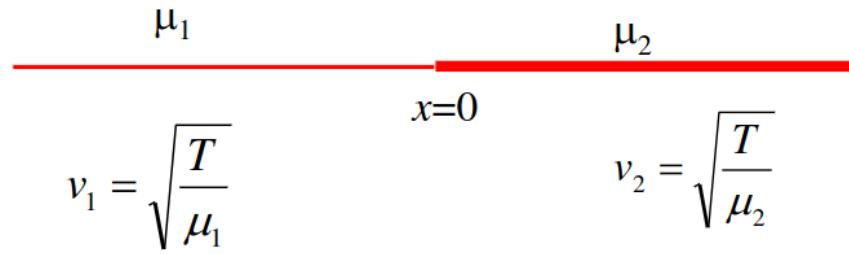


Figure 79.3: The two connected strings.

Consider a travelling wave coming from the left:

$$y_1 = A \cos(k_1 x - \omega_1 t)$$

$$y_2 = B \cos(k_2 x - \omega_2 t)$$

At the point of connection, we'll define this as $x = 0$. Here (as the string is not broken, so it must be connected):

$$y_1 = y_2 \quad (79.1)$$

Also, force is finite, therefore curvature must be finite (as force is proportional to curvature). Therefore we cannot have any discontinuities in curvature.

$$\frac{\partial y_1}{\partial x_1} = \frac{\partial y_2}{\partial x_2} \quad (79.2)$$

Disregarding non-linear effects, so assuming that the frequency with which the waves travel down the string is the same for both parts of the string, $\omega_1 = \omega_2 = \omega$.

From (1) and noting $x = 0$:

$$A \cos(-\omega t) = B \cos(\omega t) \quad (79.3)$$

And from (2) with the same note:

$$-k_1 A \sin(-\omega t) = -k_2 B \sin(-\omega t) \quad (79.4)$$

However this suggests that $A = B = 0$. This is technically a solution, but not really - it doesn't represent an actual wave (just two flat lines with no amplitude ever)

4.1 Including Reflection

The previous case did not work as we disregarded reflection at the wave boundary, P . Lets add an extra wave (the C term) to represent the reflection back into the light string:

$$y_1 = A \cos(k_1 x - \omega t) + C \cos(k_1 x + \omega t) = B \cos(k_2 x - \omega t)$$

From $y_1 = y_2$ at $x = 0$ we get:

$$y_1 = A \cos(-\omega t) + C \cos(-\omega t) = B \cos(-\omega t)$$

So: $A + C = B$

From (2) we get:

$$-k_1 A \sin(-\omega t) - k_1 C \sin(\omega t) = -k_2 B \sin(-\omega t)$$

Which has solutions:

$$B = \frac{2k_1}{k_1 + k_2} A$$

$$C = \frac{k_1 - k_2}{k_1 + k_2} A$$

We have the incident wave:

$$y_1 = A \cos(k_1 x - \omega t)$$

The transmitted wave:

$$y_2 = B \cos(k_2 x - \omega t) = \frac{2k_1}{k_1 + k_2} A \cos(k_2 x - \omega t)$$

And lastly the reflected wave:

$$y_3 = C \cos(k_1 x + \omega t) = \frac{k_1 - k_2}{k_1 + k_2} A \cos(k_1 x - \omega t)$$

We note:

k_1 is the wave number in the medium where the incident wave comes from.

k_2 is the wave number in the medium where the transmitted wave goes into.

4.2 Example

If the wave comes in from the left, then $\mu_2 > \mu_1$ per example diagram, then $k_2 > k_1$, and C is negative.

Note that if the single pulse has a positive y amplitude, then the transmitted pulse in the heavier string will also have a positive y amplitude, but will have a smaller magnitude. The reflected wave will have a negative y -amplitude as the reflection inverts it.

If the wave comes from the right (from thick to thin), then $\mu_1 > \mu_2$, $k_1 > k_2$ and C is positive.

Wed 15 Oct 2025 11:00

Lecture 5 - Wave Applications and Introduction to Standing Waves

1 Wave Applications

Reflection and transmission have some very important real word applications:

- Fibre Optics.
- Vision and Photography.
- X-Ray Imaging.
- Sonar Imaging.
- Ultrasound Imaging
- Police Speed Checks.

We will look at a few in a bit more detail.

1.1 Electron Microscopes

We have two types:

- Scanning Electron Microscope (SEM)
 - We take some sample, and generate a very narrow beam of electrons. Some of these electrons will scatter off the sample back up. We take a detector, “scan” it around many angles and detect angles of scattering, to build a picture of the sample’s surface.
- Transmission Electron Microscope (TEM)
 - We take a very thin sample, fire the electron beam at it. Some of the electron beam will pass through the sample, where we can detect it on the other side. Areas of the sample which are thinner act as being more transparent, with a greater rate of transmission. We can measure this difference.

1.2 Solar Cells

Consider a solar cell with light incident on it at some angle. We want to maximise absorption of light and minimise reflection, to maximise the energy efficiency of the cell.

1.3 Ultrasound Imaging

So far, we have considered a wave to only be able to transmit or reflect at a boundary. What we have not considered yet is attenuation. This is where a wave passing through a medium loses some energy to the medium. As the wave propagates, it loses energy and hence decreases in amplitude.

For example, a sound wave propagating though a medium causes particles in the medium to oscillate. This causes a heating effect on the medium, where some of the wave energy is lost thermally resulting in a decrease in amplitude (volume).

Ultrasound is sound with very high frequencies, compared to audible sound:

$$f_{\text{ultrasound}} : 2 - 20 \text{MHz}$$

$$f_{\text{sound}} : 20 \text{Hz} - 20 \text{kHz}$$

Wave speed is given by the below, where ρ is density and B is bulk modulus (resistance to compression):

$$v = \sqrt{\frac{B}{\rho}}$$

This means:

- Ultrasound's wavelength in air (10MHz) is $33\mu\text{m}$.
- Sound's wavelength in air (at 300Hz) is $\approx 1\text{m}$

Ultrasound has a longer wavelength in fat (human tissue) of about $150\mu\text{m}$ as while fat does have a higher density, it has a much greater bulk modulus, and a greater B/ρ ratio.

Lets consider transmission of ultrasound at an air-fat boundary. Since the wavelength in air is smaller, and the density of air is much smaller, $C \approx A$, so $B \approx 0$ and there is (almost) zero transmission from air to fat.

This is why ultrasound gel is used, to ensure that there is transmission from air to gel and gel to fat, effectively bridging the gap.

2 Introduction to Standing Waves

Consider a string (of indefinite length) attached to the wall on the left. We send series of wave pulses (y_1) into the string from the right, which travel left and reflect. The reflected wave (y_2) travels back to the right.

If we continue sending a wave signal from $y_1(x, t)$ we would have a continuous wave signal back in the form of $y_2(x, t)$.

This causes the two waves to overlap and interfere with each other. We can say that the resultant amplitude for any point on the string is given by:

$$y = y_1(x, t) + y_2(x, t)$$

This works because the wave equation is a linear partial differential equation. If two functions y_1, y_2 are solutions to the wave equation, then this means their sum will be two (hence y is a valid final wave).

The incident wave is:

$$y_1 = -A \cos(kx + \omega t)$$

And the reflected wave is (assuming the two waves have equal amplitude, period and wavelength, with the second travelling in the opposite direction and being inverted due to the reflection):

$$y_2 = A \cos(kx - \omega t)$$

We add these together:

$$y_{\text{total}} = y_1 + y_2 = A \left[\underbrace{\cos(kx)}_A - \underbrace{\cos(kx + \omega t)}_B + \underbrace{\cos(kx - \omega t)}_A \right]$$

$$\text{using: } \cos(A \pm B) = \cos A \cos B \mp \sin A \sin B$$

$$y_{\text{total}} = A[-\cos(kx) \cos(\omega t) + \sin(kx) \sin(\omega t)] + A[\cos(kx) \cos(\omega t) + \sin(kx) \sin(\omega t)]$$

$$y_{\text{total}} = 2A \sin(kx) \sin(\omega t)$$

This gives us the general form of a sinusoidal standing wave function.

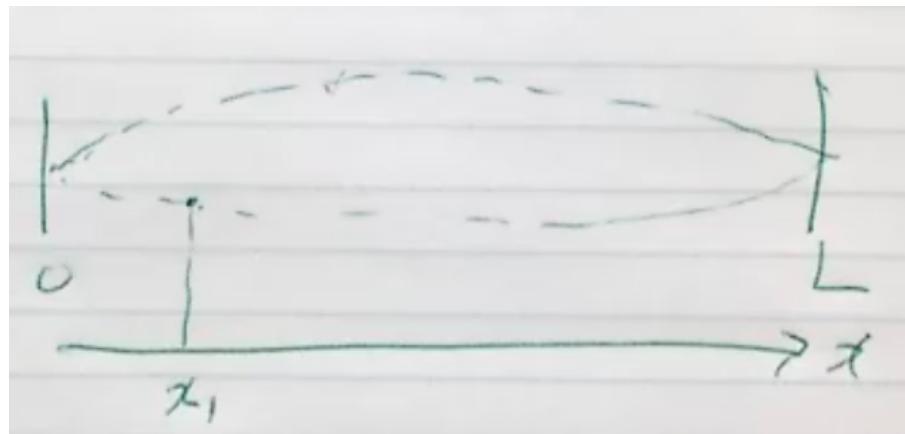


Figure 80.1

At some point $x = x_1$:

$$y_t(x = x_1, t) = \boxed{2A\sin(kx_1)} \sin(\omega t)$$

As x_1 is a constant, the boxed term is now another constant too. Let this constant be E .

$$y_t(x = x_1, t) = E \sin \omega t$$

This is the equation for a harmonic oscillator. This point of the string oscillates in simple harmonic motion with a maximum amplitude dependant on position.

If we consider a higher frequency wave, we may encounter points where $\sin kx = 0$. These are called nodes.

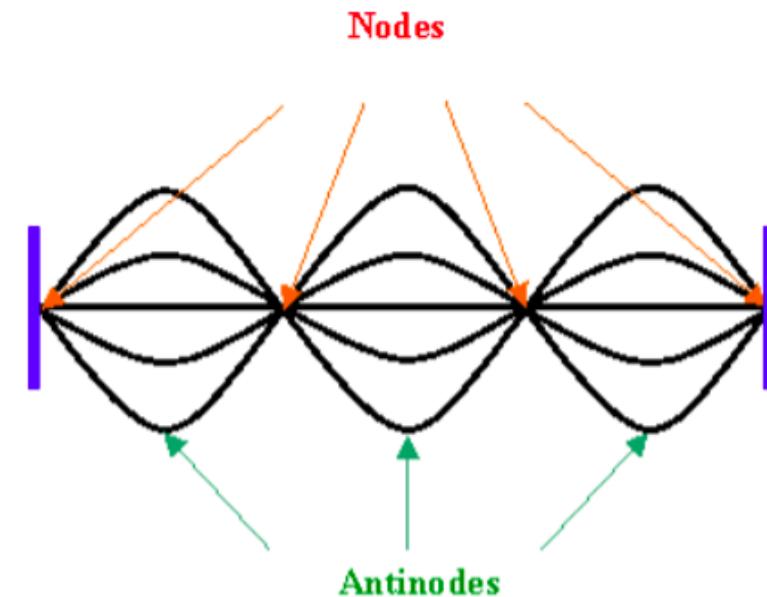


Figure 80.2

The amplitude at a node is always $y = 0$, regardless of time, as the sine term is zero regardless of time. This wave does not propagate, it is “stationary”.

3 Standing Waves and the Wave Equation

It's important to note that standing waves also satisfy the wave equation. Carrying out derivatives, we can get:

$$\frac{\partial y}{\partial x} = 2Ak \cos kx \sin \omega t$$

$$\frac{\partial^2 y}{\partial x^2} = 2A(-k^2) \sin kx \sin \omega t$$

$$\frac{\partial y}{\partial t} = 2A\omega \sin kx \cos \omega t$$

$$\frac{\partial^2 y}{\partial t^2} = 2A(-\omega^2) \sin kx \sin \omega t$$

And substituting into the wave equation:

$$\begin{aligned} \frac{\left(\frac{\partial^2 y}{\partial t^2}\right)}{\left(\frac{\partial^2 y}{\partial x^2}\right)} &= \frac{2A(-\omega^2) \sin kx \sin \omega t}{2A(-k^2) \sin kx \sin \omega t} = \frac{\omega^2}{k^2} = v^2 \\ \therefore \frac{\partial^2 y}{\partial x^2} &= \frac{1}{v^2} \frac{\partial^2 y}{\partial t^2} \end{aligned}$$

So the wave equation is satisfied.

Thu 16 Oct 2025 13:00

Lecture 6 - Standing Waves 2 Electric Boogaloo

1 Recap

$$\frac{\partial y}{\partial x} = 2Ak \cos(kx) \sin(\omega t)$$

$$\frac{\partial^2 y}{\partial x^2} = 2A(-k^2) \sin(kx) \sin(\omega t)$$

$$\frac{\partial y}{\partial t} = 2A\omega \sin(kx) \cos(\omega t)$$

$$\frac{\partial^2 y}{\partial t^2} = 2A(-\omega^2) \sin(kx) \sin(\omega t)$$

Therefore:

$$\frac{\frac{\partial^2 y}{\partial t^2}}{\frac{\partial^2 y}{\partial x^2}} = \frac{2A(-\omega^2) \sin(kx) \sin(\omega t)}{2A(-k^2) \sin(kx) \sin(\omega t)} = \frac{\omega^2}{k^2} = v^2$$

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 y}{\partial t^2}$$

Therefore a standing wave still obeys the wave equation, as it must.

2 Standing Wave Properties

2.1 Wavelength

Consider a horizontal string from $x = 0$ to $x = L$, with both ends fixed. We generate a sinusoidal wave pulse, which must satisfy:

$$y(x, t) = 2A \sin(kx) \sin(\omega t)$$

We know at $x = L$ and $x = 0$, $y = 0$ at all times as fixed at this point. Therefore:

$$kL = n\pi, (n \in \mathbb{N})$$

For $n = 1$, we have half a wavelength on the string:

$$\lambda = \frac{2L}{1} = 2L$$

And this has a general form: $\lambda = \frac{2L}{n}$.

2.2 Frequency

$$f_n = \frac{v}{\lambda_n} = \frac{v}{\left(\frac{2L}{n}\right)} = \frac{nv}{2L}$$

Crucially:

$$f_1 = \frac{v}{2L}$$

Is the first harmonic (or fundamental). $f_2 = 2f_1$ is the second harmonic, or first overtone, etc. All of these, f_n where $n \in \mathbb{N}$ are called “normal modes”. For each normal mode, the corresponding frequency is called the resonant frequency (natural frequency of the system).

What happens if we try to create a standing wave where $\lambda \neq \frac{2L}{n}$? In short, we cannot. The system will reject any attempts to do so.

2.3 Energy

Energy is proportional to ω^2 . Energy can only take certain discrete values (corresponding to f_1, f_2, \dots, f_n), we find that the system has quantised possible values for energy.

To generate a wave with a higher frequency we either have to use a shorter L, or a higher v. A higher v is achieved by using a lighter string or placing the system under higher tension.

3 Sound Waves

3.1 Notation

Displacement of a sound wave is denoted:

$$s(x, t) = S_m \cos(kx - \omega t)$$

And pressure is given by:

$$\Delta P(x, t) = \Delta P_m \sin(kx - \omega t)$$

4 Different Boundary Conditions

The equations for standing waves given is only true for the boundary conditions of both ends fixed. If we vary these (for example left end fixed, right end not, wave initially travelling left) we get a different solution. For example, the first harmonic:

$$L = \frac{1}{4\lambda_1}$$

$$f_1 = \frac{v}{4L}$$

Where the left end forms a node (as required by boundary conditions) and the right end forms an antinode, as it is free to move. For the third harmonic:

$$L = \frac{3}{4\lambda_3}$$

$$f_3 = \frac{3v}{4L} = 3f_1$$

And fifth:

$$L = \frac{5}{4\lambda_5}$$

$$f_5 = \frac{5v}{4L} = 5f_1$$

Notably, this system cannot support even harmonics.

Thu 22 Oct 2025 11:00

Lecture 7 - Energy and Power of Waves

1 Kinetic Energy for a Sine Wave

Assuming a sine wave created by a harmonic oscillator. We can consider every particle on this string as acting as its own harmonic oscillator. We can determine the kinetic energy for one single harmonic oscillator (of an infinitesimally small length of string)

$$\begin{aligned} KE &= \frac{1}{2}mv^2 \\ (KE)_{\max} &= \frac{1}{2}mv_{\max}^2 \end{aligned}$$

Mass is given by the density of the string per unit length multiplied by the length of the section of string: μdx , and the maximum velocity by ωA :

$$(KE)_{\max} = \frac{1}{2}(\mu dx)(\omega A)^2$$

This gives us the kinetic energy of a single oscillator. We now want to get the total energy across a single wavelength, by integrating:

$$\begin{aligned} E &= \int_0^\lambda \frac{1}{2}\omega^2 A^2 \mu dx \\ E &= \frac{1}{2}\omega^2 A^2 \mu \int_0^\lambda dx \\ E &= \frac{1}{2}\omega^2 A^2 \mu \lambda \end{aligned}$$

Note that strictly speaking, at any point half the wave's energy is kinetic and half is potential. Instead of calculating potential energy, we consider a singular oscillator. In this single oscillator, the two quantities oscillate in opposition to each other, so max KE means zero PE. We can say $E_{\text{total}} = KE_{\max}$ therefore, for a singular oscillator. This works out when we integrate across all oscillators.

And for power:

$$\begin{aligned} \text{Power} &= \frac{\text{Energy}}{\text{Time}} \\ P &= \frac{\frac{1}{2}(\omega A)^2 \mu \lambda}{T} \\ &= \frac{1}{2}(\omega A)^2 \mu v \end{aligned}$$

2 Standing Waves

Considering a standing wave with $\lambda = 2L$ (hence a single loop with two nodes at each boundary and one antinode in the middle):

$$y_t = y_1 + y_2$$

$$\implies (KE)_t = (KE)_1 + (KE)_2$$

Where y_1 and y_2 are identical except for their direction, but they carry the same kinetic energy. The KE of a standing wave is the sum of the KE of two waves that make it up.

Note: Normally, A is the amplitude of the travelling wave (hence 2A is the amplitude of the standing wave created by them), however it is sometimes ambiguous what is being referred to by A - some questions may give the standing wave amplitude and also denote this as A.

3 Interference

Superposition: When waves overlap in the same region, the resulting wave is the algebraic sum of waves (they interfere)

Consider two waves:

$$y_1 = A \cos(kx - \omega t) \quad y_2 = A \cos(kx - \omega t + \delta)$$

Where both waves are travelling in the same direction with the same amplitude, but the second wave has some phase shift δ .

We can sum the waves as with standing waves:

$$y = A \cos(\underbrace{kx - \omega t}_{\alpha}) + A \cos(\underbrace{kx - \omega t + \delta}_{\beta})$$

$$\text{Using: } \cos \alpha + \cos \beta = 2 \cos \frac{\alpha + \beta}{2} \cos \frac{\alpha - \beta}{2}$$

$$y = 2A \cos\left(\frac{\delta}{2}\right) \cos\left(kx - \omega t + \frac{\delta}{2}\right)$$

So the amplitude of the new travelling (not standing) wave is dependant on delta, where amplitude is given by $2A \cos(\delta/2)$.

Here are three examples of phase difference where y_1 is the blue line, y_2 is green and y_{total} is red:

$$y = 2A \cos\left(\frac{\delta}{2}\right) \cos(kx - \omega t + \frac{\delta}{2})$$

$$2 \cos\left(\frac{\delta}{2}\right) = 2 \quad \delta = 0$$

$$2 \cos\left(\frac{\delta}{2}\right) = 1.85 \quad \delta = \pi/4$$

$$2 \cos\left(\frac{\delta}{2}\right) = 1.41 \quad \delta = \pi/2$$

$$2 \cos\left(\frac{\delta}{2}\right) = 0 \quad \delta = \pi$$

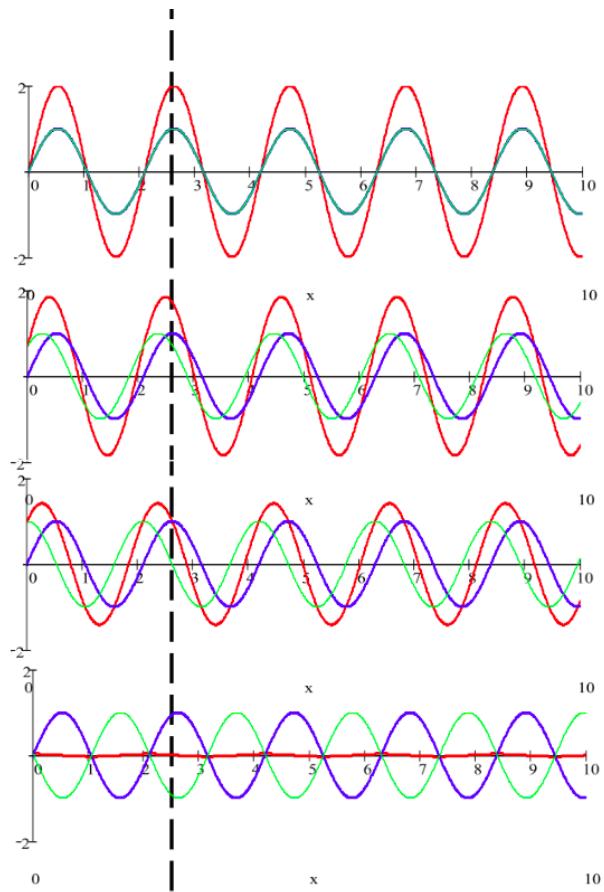


Figure 82.1

Given two arbitrary waves, differences in phase can arise from either a difference in angular frequency, and/or in x .

3.1 Example

Suppose we have two sound sources, S_1, S_2 . They oscillate in phase and emit harmonic waves of equal frequency and wavelength. Let's consider some point P where the waves interact where the path differences are different (i.e. length from S_1 to $P \neq$ length from S_2 to P).

If there is an integer number of wavelengths in path difference, the interference is totally constructive, if path difference is equal to a half number of wavelengths (i.e. 1.5, 2.5, 3.5) then the interference is totally destructive.

From S_1 :

$$y_1(l_1, t) = A \cos(kl_1 - \omega t)$$

From S_2 :

$$y_2(l_2, t) = A \cos(kl_2 - \omega t)$$

Giving us a phase difference of:

$$\delta = (kl_1 - \omega t) - (kl_2 - \omega t) = k(l_1 - l_2)$$

$$\delta = \frac{2\pi}{\lambda} \Delta l$$

Hence the signal strength at point P is dependant on the path difference.

4 Beats

Lets consider interference from waves with slightly different frequencies (but the same amplitude). We again have two sources S_1 and S_2 . We observe the resultant amplitude after these have interfered at some value of x (say $x = 0$ so the kx term is 0), where both sources are the same distance from this point, P .

$$p_1 = A \cos \omega_1 t$$

$$p_2 = A \cos \omega_2 t$$

$$p = p_1 + p_2 = A(\cos \omega_1 t + \cos \omega_2 t)$$

$$p = 2A \cos\left(\frac{\omega_1 - \omega_2}{2}t\right) \cos\left(\frac{\omega_1 + \omega_2}{2}t\right)$$

We also note the average value of omega as $\omega_{avg} = (\omega_1 + \omega_2)/2$ and the variation from this value as $\Delta\omega = (\omega_1 - \omega_2)/2$

$$p = 2A \cos(\Delta\omega t) \cos(\omega_{avg} t)$$

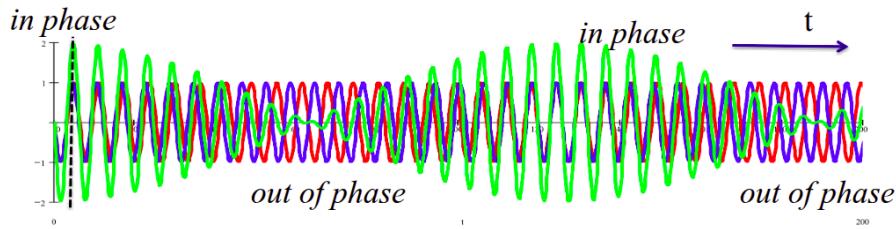


Figure 82.2: The resultant wave p in green, with p_1 and p_2 in red and blue.

Note that (as the difference in frequency is small) the two waves start mostly in phase, leading to a high amplitude resultant wave. As they drift out of phase, the resultant amplitude drops (becoming zero when in antiphase). Phase difference is periodic, so they return in phase and this repeats, creating a cosine wave with oscillating amplitude.

We can consider the 'envelope' that contains the resultant wave (given in purple):

The amplitude oscillates with angular frequency:

$$\omega = \left(\frac{\omega_1 - \omega_2}{2}\right) = 2\pi f \implies f = \left(\frac{\omega_1 - \omega_2}{4\pi}\right)$$

The human ear cannot really tell the difference between the green signal and the purple envelope signal, and we have a perceived beat frequency of:

$$f_{beat} = 2\left(\frac{\omega_1 - \omega_2}{4\pi}\right) = \frac{\omega_1}{2\pi} - \frac{\omega_2}{2\pi} = |f_1 - f_2|$$

The 2 arises as our ears don't care about the difference between a positive and a negative amplitude. The cosine wave goes positive, zero, negative, zero (one oscillation) while the human ear perceives loud, quiet, loud, quiet (two oscillations).

Note: This holds up when the amplitude of the two waves is the same. If the amplitudes are different, we would have to re-derive to take this into account.

Thu 23 Oct 2025 13:00

Lecture 8 - EM Standing Waves and Lasers

1 Lasers

A traditional laser has two mirrors on either side of a cavity. One is (ideally) 100% reflecting, while another is almost perfectly reflecting (say 99% reflective), but allows some transmission. This causes EM waves to reflect back and forth (with some transmitted to actually cause the visible laser).

These reflecting waves cause a standing wave inside the laser cavity. This standing wave oscillates with the equation given in Lec 07.

2 A Quick Audio Interlude

Listening to various sounds from a frequency generator, we notice two things:

- The human ear is very sensitive to changes in frequency, even in the single Hz range.
- A square wave has a higher perceived pitch than a sine wave of the same frequency.
 - This is because a square wave can be decomposed into the sum of many sine waves, and some of these sine wave components have a higher frequency compared to the square wave itself, which our ears can detect.

3 And Back to Superposition

Given two travelling waves:

$$y_1 = A \cos(k_1 x - \omega_1 t)$$

$$y_2 = A \cos(k_2 x - \omega_2 t)$$

$$y = y_1 + y_2$$

Hence:

$$y = A \cos(k_1 x - \omega_1 t) + A \cos(k_2 x - \omega_2 t)$$

And using a double angle formula:

$$y = 2A \cos\left(\frac{k_1 - k_2}{2}x - \frac{\omega_1 - \omega_2}{2}t\right) \cos\left(\frac{k_1 + k_2}{2}x - \frac{\omega_1 + \omega_2}{2}t\right)$$

Or briefly:

$$y = 2A \underbrace{\cos(\Delta k x - \Delta \omega t)}_{\text{group}} \underbrace{\cos(k_{avg} x - \omega_{avg} t)}_{\text{carrier}}$$

We define the phase velocity as:

$$v_{\text{phase}} = \frac{\omega_{avg}}{k_{avg}}$$

And the group velocity as:

$$v_{\text{group}} = \frac{\Delta \omega}{\Delta k}$$

4 Carriers and Groups

What actually is a carrier and a group? Lets take radio as an example. Radio waves travel for potentially hundreds of kilometers, and it's not possible to send sound waves remotely close to that distance because of high attenuation.

The idea therefore arose of using a radio frequency wave to carry sound. Lets consider an example EM RF wave at 500kHz. We then approximate our sound wave as a sine wave of pressure variations, say at 1000Hz.

The first thing we can do is to modulate the amplitude of the RF wave by the audio wave. The RF wave keeps the same frequency, but changes amplitude depending on the amplitude of the audio wave.

Upon receipt of the wave, we filter out the high frequency oscillations of the radio wave itself, but we extract a lower frequency wave defined by a wave that passes through the maximum points of oscillation. This effectively restores the sound wave.

This is known as AM radio. The carrier is the high frequency wave, and the group is the lower frequency constructed wave.

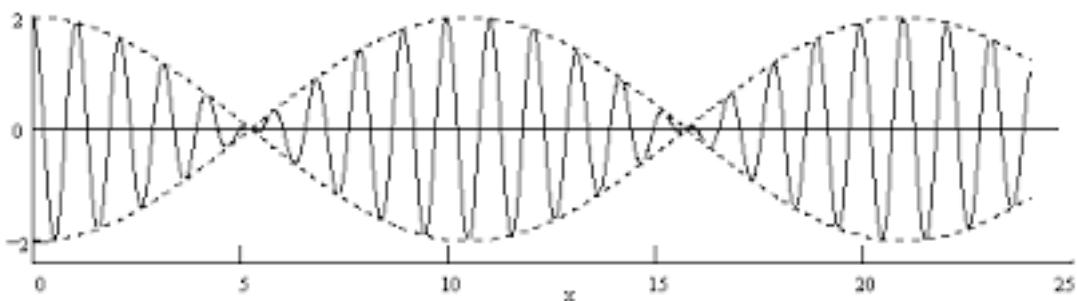


Figure 83.1: Carrier and Group waves. The lower freq. dotted line is the extracted wave (group), the higher frequency wave is the carrier.

5 Velocities

Which of the two waves have a higher speed?

Assuming wave speed and frequency are independent, we have:

$$V = \omega_1/k_1 \quad V = \omega_2/k_2$$

$$v_{\text{phase}} = \frac{\omega_{\text{avg}}}{k_{\text{avg}}} = \frac{(\omega_1 + \omega_2)/2}{(k_1 + k_2)/2} = V$$

And for group:

$$v_g = \frac{(\omega_1 - \omega_2)/2}{(k_1 - k_2)/2} = V$$

Hence they move at the same speed. This is only true on the assumption that wave speed does not depend on frequency, which is not generally true. Depending on the circumstances, phase velocity can be greater or smaller than group velocity.

In a medium where the wave speed depends on frequency, we call this medium dispersive.

Air is non-dispersive (i.e. sound), and in a vacuum the same applies (i.e. speed of light).

What about the speed of light in glass? Does this depend on frequency? We'll investigate later in the optics section.

6 A Few More Definitions

Coherence: Two (or more) sources that are in phase, or have a constant phase difference (Δ) are said to be coherent.

6.1 Adding Coherent Waves

Constant phase difference means we can nicely add amplitudes, so total intensity:

$$I = |A|^2 = (A_1 + A_2 + A_3 + \dots + A_n)^2$$

6.2 Adding Incoherent Waves

We directly add intensities instead:

$$I = I_1 + I_2 + \dots = A_1^2 + A_2^2 + \dots$$

6.3 The Decibel Scale

$$\beta = 10 \log\left(\frac{I}{I_0}\right)$$

Where I is the intensity of the source, and I_0 is a reference level approximately at threshold of hearing equal to $10^{-12} W/m^2$

At the threshold of hearing:

$$\beta = 10 \log\left(\frac{10^{-12}}{10^{-12}}\right) = 10 \log(1) = 0 dB$$

Crucially, 0dB is not zero sound intensity.

At the upper pain threshold:

$$\beta = 10 \log\left(\frac{1}{10^{-12}}\right) = 120 dB$$

But, how do we calculate I here?

$$I = \frac{\text{Power}}{\text{Area}}$$

Power is related to energy, so:

$$I = \frac{\text{energy}}{\text{time} \times \text{area}}$$

And to get intensity for an actual volume:

$$I = \frac{\text{energy} \times \text{length}}{\text{time} \times \text{volume}}$$

$$I = \frac{\text{energy}}{\text{volume}} \times \text{wavespeed}$$

What displacement on the eardrum do we actually have? For 120dB, $10^{-5} m$, or 10^{-11} for 0dB. The latter measurement is smaller than an atom. The ear is a very sensitive instrument.

7 Human Senses

Interestingly, we see, hear and smell all on a log scale. Physical feeling is not very scientific and can't easily be quantified. Emotional feeling is also not quantifiable...

Tue 29 Oct 2025 12:00

Lecture 9 - Doppler Shift, Shockwaves and Optics I

Doppler Shift: If a wave emitter and observer are moving relative to each other, the observed frequency will be different to the emitted frequency. As wave emitting objects moves towards an observer, the perceived frequency of the wave increases. As the object moves away, the frequency decreases.

We consider two scenarios: (1) a moving source, and (2) a moving receiver, where the other is stationary. We will consider this for **sound waves**.

1 Case 1: Moving Source, Stationary Receiver

The source has frequency f_s and period $T_s = 1/f_s$. It has velocity u_s , and the velocity of waves in the medium is v .

At $t = 0$: The wave source creates a sound wave pulse, which begins to propagate out in all directions.

At $t = T_s$: The initial sound wave has moved by $d = vt$, so distance travelled by the wave is vT_s distance units. The source has moved to the right by $u_s T_s$ distance units.

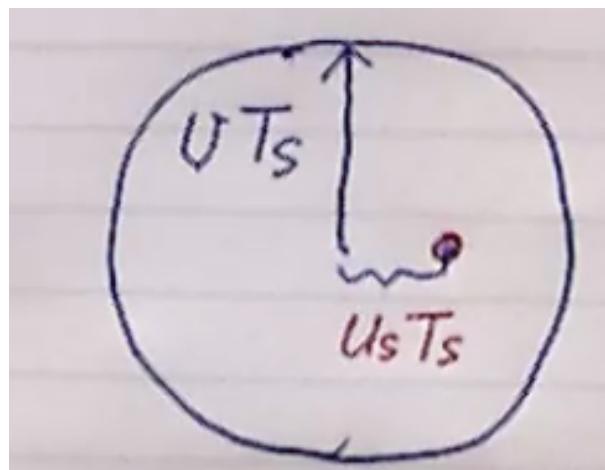


Figure 84.1

As this continues, we get something which looks like this:

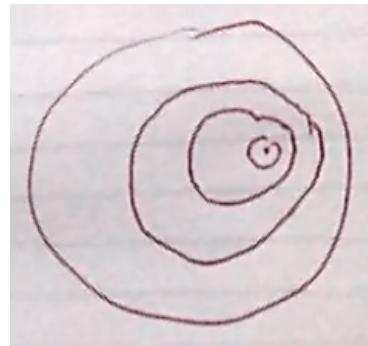


Figure 84.2

This effectively gives us a wave which has two different frequencies and two different wavelengths, one in the forward direction and one each in the backwards.

We denote the distance between the source's new location and the wavelet in the forward direction as λ_f , and in the reverse direction as λ_b .

$$\lambda_f = vT_s - u_s T_s$$

And the new frequency in the forward direction:

$$\begin{aligned} f_f &= \frac{v}{\lambda_f} \\ f_f &= \frac{v}{(v-u_s)T_s} \\ f_f &= \frac{v}{(v-u_s)(1/f_s)} \\ f_f &= f_s \frac{v}{v-u_s} \end{aligned}$$

We can do the same analysis considering the wavelength in the backwards direction:

$$f_b = f_s \frac{v}{v+u_s}$$

2 Case 2: Stationary Source, Moving Receiver

Consider a periodic wave with wavelength λ and speed v . Lets look at the case where the receiver is moving towards the source at a speed u_r .

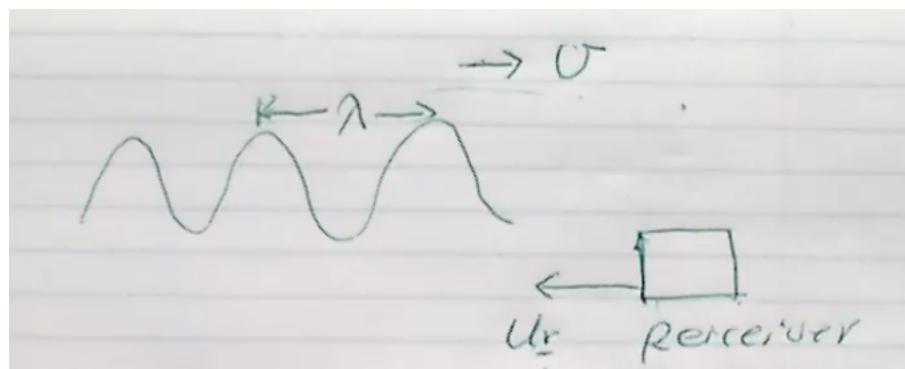


Figure 84.3

The receiver perceives the wave as moving at speed $v + u_r$ ¹. It therefore observes the frequency:

$$f_f = \frac{v + u_r}{\lambda}$$

And if the receiver is moving in the opposite direction:

$$f_b = \frac{v - u_r}{\lambda}$$

Note that as the source is not moving, the wavelength remains constant.

3 Case 3: Both Source & Receiver Moving

In summary, for a moving source:

$$f_r = \frac{v}{v \pm u_s} f_s \quad (- \text{ towards, } + \text{ away.})$$

And for a moving receiver:

$$f_s = \frac{v \pm u_r}{v} f_r \quad (+ \text{ towards, } - \text{ away.})$$

We can combine these two equations to get the general form:

$$f_r = \frac{v \pm u_r}{v \pm u_s} f_s$$

Note that the two \pm s are independent and the numerator/denominator values for the sign should be chosen according to the sign conventions for the numerator/denominator stated in the two individual equations. If moving towards, frequency increases. If moving away, frequency decreases.

3.1 What about light waves?

The doppler shift in frequency depends on the source/receiver speed relative to the medium. For light, we do not consider a traditional medium. We therefore cannot talk about absolute motion, we can only consider relative motion between the source and the receiver.

We have:

$$f_r = f_s \sqrt{\frac{c \pm u}{c \mp u}}$$

If we expand this and ignore terms in u^2/c^2 (as they'll be tiny). We get:

$$\frac{\Delta f}{f_s} = \pm \frac{u}{c} \quad \text{where: } \Delta f = f_r - f_s$$

This only holds for $u \ll c$, as if this isn't true the u^2/c^2 terms will be non-negligible. The square root appears due to time dilation.

- **Blueshift (+):** Source approaching (frequency increases).
- **Redshift (-):** Source receding (frequency decreases).

4 Shock Waves

Shock waves arise when the velocity of the source is faster than the velocity of waves in the medium. We have previously implicitly assumed $u < v$, so this did not happen. We end up with a large number of wavefronts building up in front of the object, forming a cone of high amplitude:

¹This does not take special relativity into account, so no ultrafast receivers here, and since it's a sound wave, $v \ll c$.

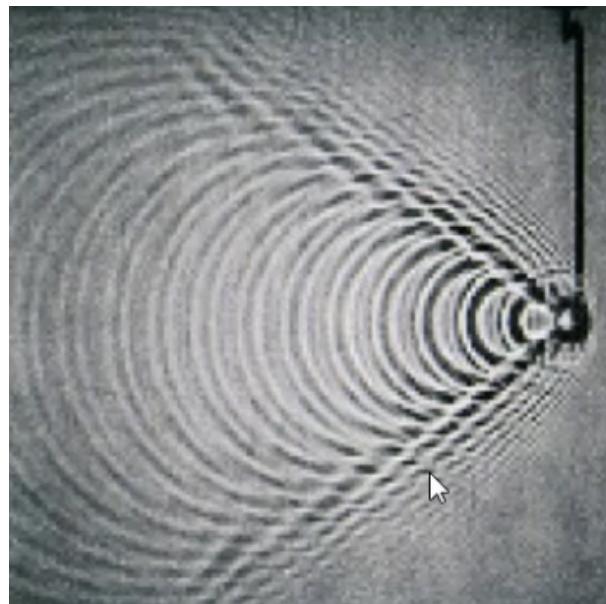


Figure 84.4: A shock wave cone.

If $u = v$, this creates a shock wave which is a straight line tangent to the motion.

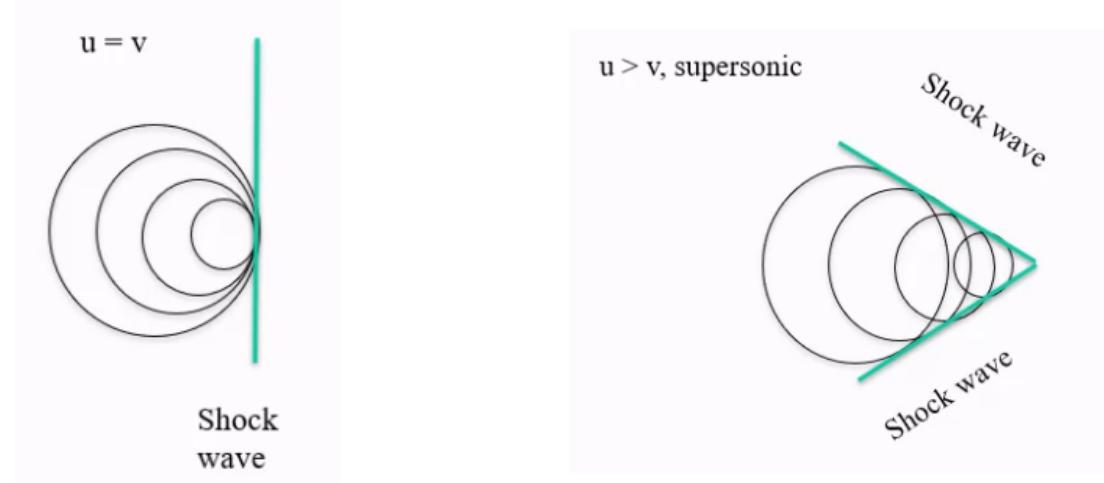


Figure 84.6

Figure 84.5

The angle created from the horizontal to either the top or bottom of this cone is known as the Mach Cone Angle θ , and is given by the Mach Number u/v (where v is velocity of the sound, u is the velocity of the source):

$$\sin \theta = \frac{v}{u}$$

When travelling along the path of the moving source, the sound waves produced later are heard first. This is the reverse of what we'd expect for a slower source. This is because, as the source moves faster than sound, the source gets closer to the receiver faster than the old sound waves can travel. The sound waves emitted at (for example) $t = 10\text{s}$ are therefore emitted much closer to the receiver and arrive before the sound emitted at (for example) $t = 5\text{s}$.

At the back of the supersonic source, the frequency is significantly shifted:

$$f_b = \frac{v}{v+u} f_s < \frac{1}{2} f_s$$

This equation suggests that in front of the source, the frequency becomes negative. This is because this analysis does not apply in front of the source, as waves do not pass the cone. The exception to this is a sonic boom, which is the extremely high amplitude of the cone itself being heard, as passes a receiver.

5 Optics Part One

Now we've studied waves, we can formally apply this to light. Electromagnetic waves are oscillations of the electric (E) and magnetic (B) fields.

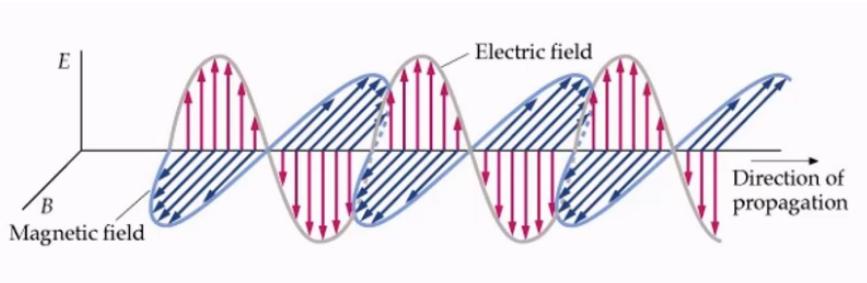


Figure 84.7

With the following properties:

- Light needs no medium, it can travel in a vacuum.
- Light does not involve the oscillation of particles (ignoring weird QM stuff).
- Light is a transverse wave (with f , ω , k , standing waves, doppler shift etc).

For electromagnetic waves, we have the wave equation:

$$\frac{\partial^2 E}{\partial x^2} = \epsilon_0 \mu_0 \frac{\partial^2 E}{\partial t^2}$$

Where ϵ and μ are two constants which determine how well waves propagate through electric and magnetic fields respectively. The zero subscript indicates that we are talking about free space (i.e. in a vacuum).

We have the following for the speed of light:

$$c = \sqrt{\frac{1}{\epsilon_0 \mu_0}}$$

Hence:

$$c^2 = \frac{1}{\epsilon_0 \mu_0} \implies 1/c^2 = \epsilon_0 \mu_0$$

Therefore the wave equation derived previously (with $1/v^2$ in place of $\epsilon_0 \mu_0$) is still satisfied.

For materials with *relative permittivity* ϵ_r and *relative permeability* μ_r (permittivity, ϵ and permeability μ for a specific material relative to the free-space values), the speed of light is:

$$v = \frac{c}{\sqrt{\epsilon_r \mu_r}} = \frac{c}{n}$$

Where $n = \sqrt{\epsilon_r \mu_r} > 1$. Light therefore travels slower in matter.

Thu 30 Oct 2025 13:00

Lecture 10 - Optics II

1 Key Principles

There are two main principles that we'll use:

1. Huygen's Principle

- Each point on a wavefront serves as a source of spherical secondary *wavelets* that advance with speed and frequency identical to the primary wave.
- If we consider each point on a wave as a spherical wavelet source, the wavefront is given by a line tangent to all of these.

2. Fermat's Principle

- The actual path taken by a beam of light is the one which takes the least time to traverse.
- i.e. $dt/dl = 0$
- Light could theoretically take an infinite number of possible routes between any two points, but in practice (in the same material) it will travel in a straight line as this is the fastest route.
- When light is travelling through an interface (i.e. an air to glass boundary), the fastest point is no longer a straight line, as refraction appears.

2 Reflection

Lets consider some mirror as a reflecting surface, with light travelling from point A to point B. There are multiple different theoretical paths that the light could take, so Fermat's principle requires the shortest possible time taken.

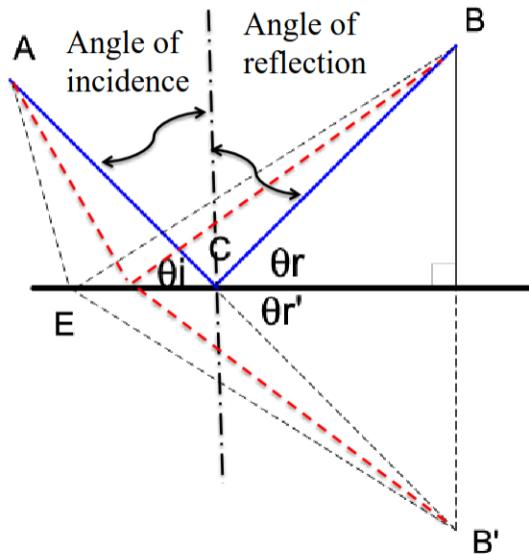


Figure 85.1

We require, by Fermat's Principle (as same material, so shortest time is simply shortest path):

$$AC + CB = \min$$

By considering congruent triangles (let O be the midpoint of B and B'), therefore:

$$AC + CB = AC + CB'$$

Provided ACB' is a straight line, $\theta_i = \theta_r (= \theta_{r'})$. Therefore the angles of incidence and reflection must be equal.

3 Refraction

When light is travelling through a material, the speed of light in each is different. Therefore, a straight line is no longer the fastest path.

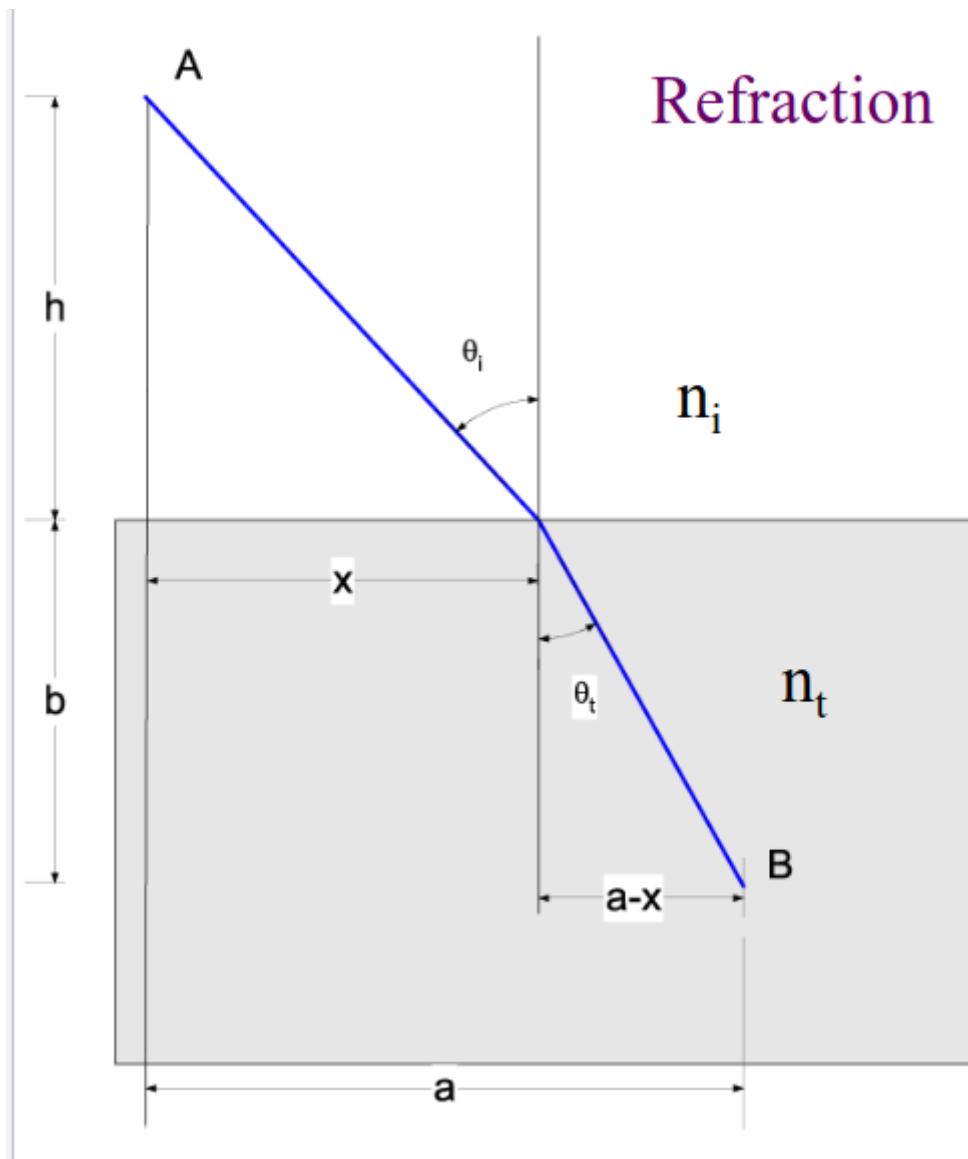


Figure 85.2

Where:

- n_i is the incident index.
- n_t is the post-refraction index.
- θ_i is the angle of incidence.
- θ_t is the angle of refraction.
- Light is travelling from point A to point B.
- Distances are given per diagram.

The aim is to be able to find the true trajectory (per Fermat) that takes the shortest amount of time. By $s = d/t$, for given distance values:

$$t = \frac{\sqrt{h^2 + x^2}}{v_i} + \frac{\sqrt{b^2 + (a-x)^2}}{v_t}$$

Where v_i, v_t are speeds of light in the respective media. t is a function of x , with constants, so to find a min value we differentiate wrt x and solve for minima:

$$\frac{dt}{dx} = \frac{x}{v_i \sqrt{h^2 + x^2}} + \frac{-(a-x)}{v_t \sqrt{b^2 + (a-x)^2}} = 0$$

Substituting values for distances, we get:

$$\frac{\sin \theta_i}{v_i} = \frac{\sin \theta_t}{v_t}$$

And since $n = c/v$, we get:

$$n_i \sin \theta_i = n_t \sin \theta_t \quad (85.1)$$

The shortest (wrt time) path must have angles which satisfies this. This is called Snell's Law.

4 Connecting Fermat and Huygen

Speed of light in a medium is less than in a vacuum, this is categorised by the index of refraction, n :

$$n = \frac{c}{v}$$

Where c is speed of light in a vacuum, and v is speed of light in the medium. i.e. for water, $n = 1.333$, for air, $n = 1.0003$.

Suppose we have a wavefront travelling as a series of wavelets. As soon as each wavelet hits the glass, they slow down. This leads to an uneven distribution of speeds across the wavelength, as some have slowed down and some have not. Therefore, the wave bends around and we get a new wavefront at a different angle to the previous wavefront.

Note the condition where the light travels perpendicular to the interface, i.e. $\theta_i = \theta_t = 0$. In this case, all wavelets hit the medium and change speed at exactly the same time, therefore there is no difference in velocity across the wavefront, so no direction change.

5 But how?

How does light actually “choose” which path to take? Apparently it just does and we have to wait to find out...

6 Combining Reflection and Refraction

When light strikes a boundary surface, there is two components - both reflected and transmitted (just like on a string). To determine this, we have to use the wave equation:

$$E = E_0 \cos(kx - \omega t)$$

Noting that instead of traditional mechanical amplitude we use the amplitude of the electromagnetic field at this point, E_0 . For now, we ignore polarisation as a potential scenario.

7 More Index

We note that since $v = f\lambda$, the index of refraction will change depending on the frequency of light:

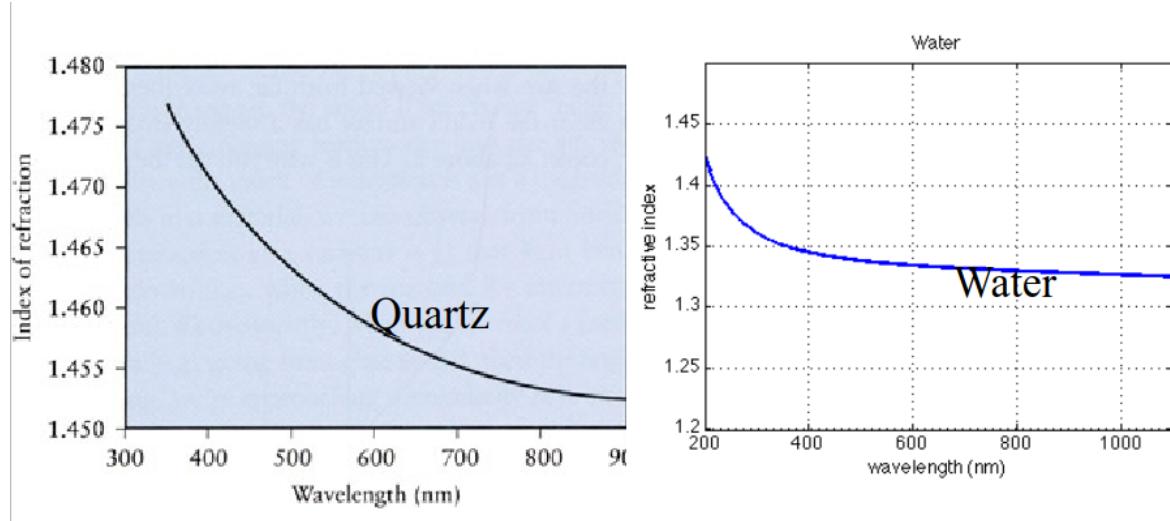


Figure 85.3

When we shine light of multiple frequencies (i.e. white light), the higher frequency light bends further than the lower frequency light. This causes dispersion of light, i.e. the formation of a rainbow out of a prism. This arises because each frequency has a different speed and therefore a different change in θ .

If we shine pulses of light of multiple frequencies down a fibre optic line, the higher frequencies will have a higher n , hence a lower speed and will arrive later.

This allows us to explain the formation of sunsets and rainbows etc. In the case of a sunset, when we see the sun just above the horizon, the sun has actually set just below the horizon. We cannot therefore see the sun via direct line of sight, and yet we can still see it as if we could?

This is because the light from the sun is refracted and bends towards us. The atmosphere is higher density at the bottom, and lower density as altitude increases. The amount of refraction depends on density, hence changing the index of refraction. We can say that n is a function of y , where y is height.

We imagine it as being actually present precisely where we see it, because our brain does not account for this refraction and extrapolates the light as a straight line. We therefore see the sun has just about to set, when in reality the sun physically has set below the horizon.

Wed 05 Nov 2025 11:00

Lecture 11 - Reflection and Refraction Examples I

1 Simple Mirror

A man is standing in front of a simple flat mirror. He is height hm . What is the minimum height of the mirror if he wishes to see the full image of himself?

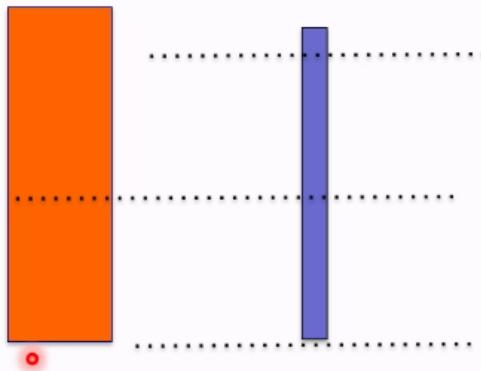


Figure 86.1: The man, modelled as a box (orange) and the mirror of indeterminate height (blue)

Consider a ray from the top of the man's head to his eye level:

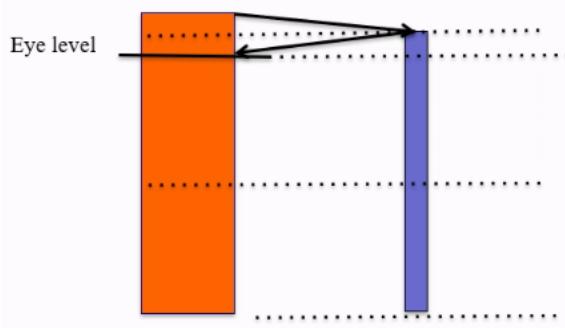


Figure 86.2

And a ray from feet level again to the man's eye (we also label some dimensions and add two extra red rays):

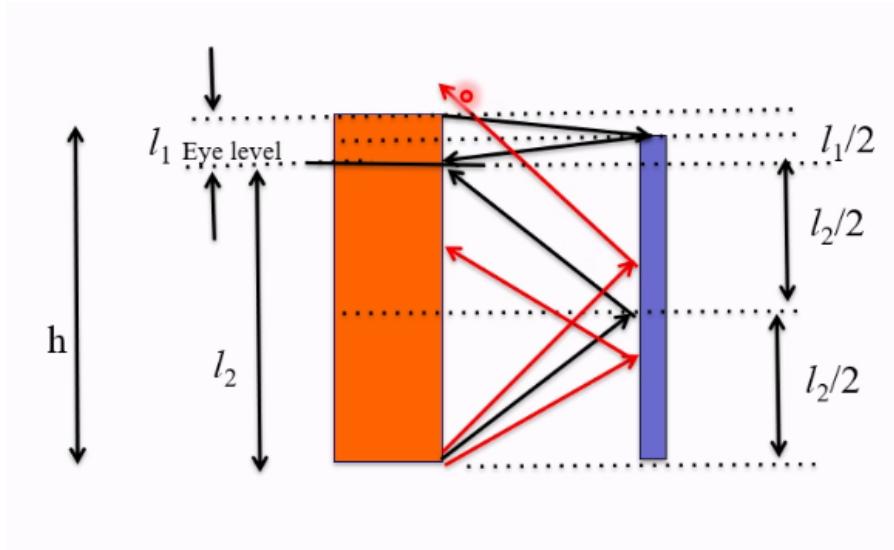


Figure 86.3

Note that the red rays are not observed by the man. The lower red ray strikes the man at too low an angle so is below eye level, while the opposite is true for the upper red ray. Neither are therefore seen.

This means that the lower half of the mirror is redundant, as any rays which strike the lower section will be reflected at too low an angle to meet the man's eyes. The upper portion of the mirror is therefore the only required portion, giving a final height of:

$$l_1/2 + l_2/2 = (l_1 + l_2)/2 = h/2$$

2 Corner Cube Reflector

Note: Pg 1088 Y&F A corner cube reflector is three mutually orthogonal reflecting surfaces:

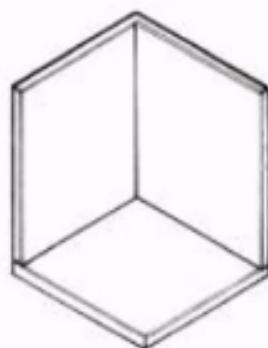


Figure 86.4: Corner Cube Reflector

We'll consider a simplified setup, with two orthogonal mirrors in a 2D scenario:

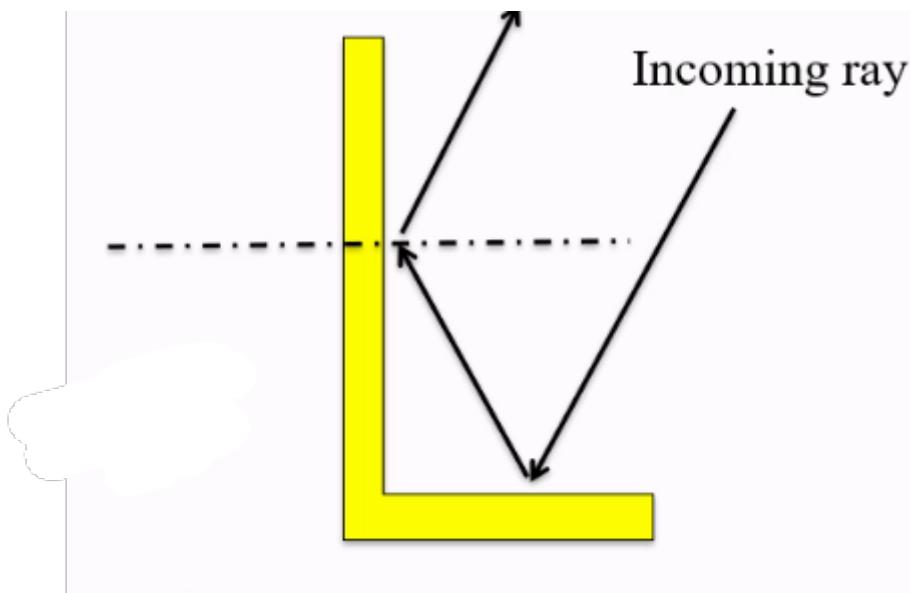


Figure 86.5

Here, the incoming ray is reflected twice (assuming it does not arrive parallel to a mirror), once on each mirror. We can easily analyse the problem as we know that the angle of incidence and reflection for the initial incoming ray are equal, denoted θ_1 .

The angle of incidence and reflection for the second reflection are also equal, and are given by $90 - \theta_1$. Note that this is equal to the angle made by the incoming wave with the horizontal.

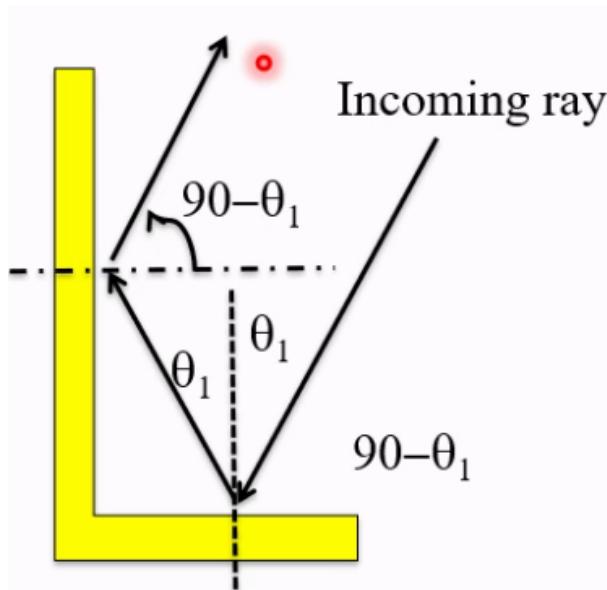


Figure 86.6

This means that the reflected ray travels back towards the emitter, at the same angle as it was initially.

3 Light Travelling Through a Piece of Glass

Consider a 1cm thick slab of glass, where a light ray is fired into it at angle of incidence θ_1 .

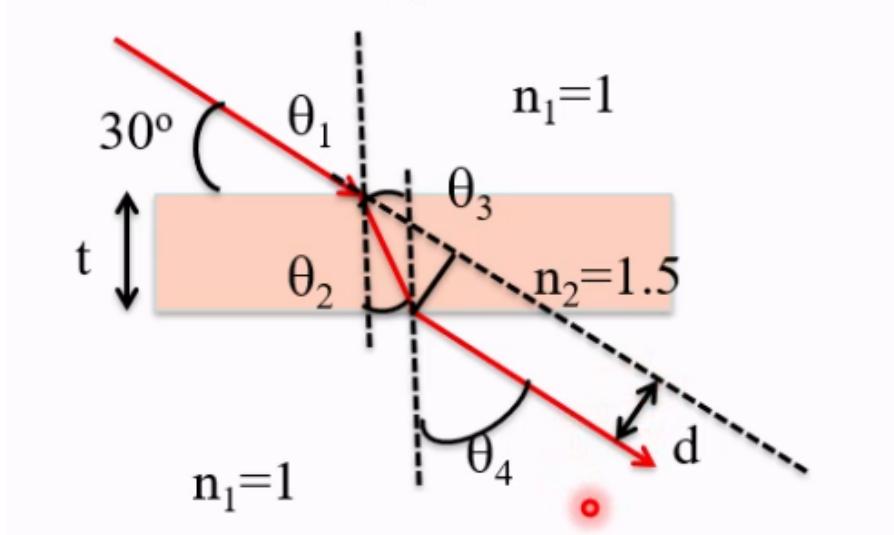


Figure 86.7

We use Snell's Law to find theta two:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2$$

$$\theta_2 = \arcsin\left(\frac{n_1 \sin \theta_1}{n_2}\right) = 35.3^\circ$$

And again:

$$n_2 \sin \theta_3 = n_1 \sin \theta_4$$

$$\theta_2 = \theta_3$$

$$\Rightarrow \theta_4 = \arcsin\left(\frac{n_2 \sin \theta_3}{n_1}\right) = 60^\circ$$

Therefore as $\theta_1 = \theta_4$ the light ray exits the material at the same angle it entered, just with some displacement. We denote this displacement d . We can find it again using some trig:

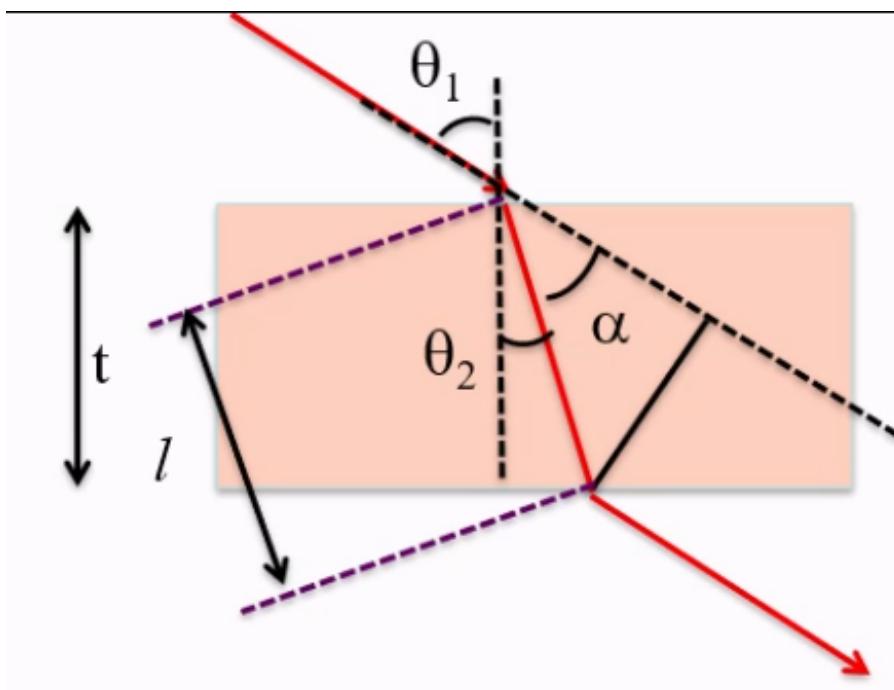


Figure 86.8

The distance is given by:

$$\begin{aligned}d &= l \sin \alpha = l \sin(\theta_1 - \theta_2) \\&= \frac{t}{\cos \theta_2} \sin(\theta_1 - \theta_2) = 0.51\text{cm}\end{aligned}$$

4 Apparent Depth of an Object

Consider an example with a fish in the pond. We want to estimate how

Thu 06 Nov 2025 13:00

Lecture 12 - Reflection and Refraction Examples II

Wed 12 Nov 2025 11:00

Lecture 13 - Image Formation

1 Sign Rules

We have the mirror equation:

$$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f}$$

The numerical values of s and s' give us the distance, but the position of the formed image is determined by their signs.

2 Mirrors

The image formed by a mirror is:

- *Virtual*, as we effectively see it appears as being behind the mirror, where there is no real light.
- *Erect*, as there is no inversion (if you stand in front of a mirror, you see your body in the same top-to-bottom orientation as real life).
- *Reversed*, as if you look in a mirror your body faces you, and you see your face, not your back.

2.1 Angled Mirrors

If we have a pair of hinged mirrors, changing the angle between them changes the way that the image is produced. Crucially, it changes how many images of the object we see:



Figure 88.1

LC Quantum Mechanics 1

Fri 03 Oct 2025 12:00

Lecture 1 - Atomic Structure

1 Atomic Structure

1.1 What is the course?

- Quantum mech is weird and unintuitive, we will build up a case in the course for why this weird theory was necessary and why we're confident it works.
- Each week will be a self-contained concept and/or historical experiment, working up to the Schrödinger Equation and wave-particle duality.
- Names and dates do not need to be memorised.
- Recommended text: University Physics (Young and Freedman).
- Office hours: 13:00 – 13:50 Fridays (immediately post-lecture), Physics East Rm 207.

1.2 Atomic Structure

What actually is an atom? What does it actually look like inside?

Early Clues

- Periodic Table (Mendeleev, 1869), periodic patterns in elements properties.
- Radioactivity (Becquerel, 1896, Curie 1898)
- Atoms emit and absorb specific discrete wavelengths, (Balmer, 1884)
- Discovery of the Electron (Thomson 1897). Cathode rays - heating metal in a vacuum with an electric field above it, to strip away electrons from the metal.
 - This showed electrons were negatively charged and extremely light (1/2000th of the atomic mass).

Atoms emit/absorb light at discrete wavelengths

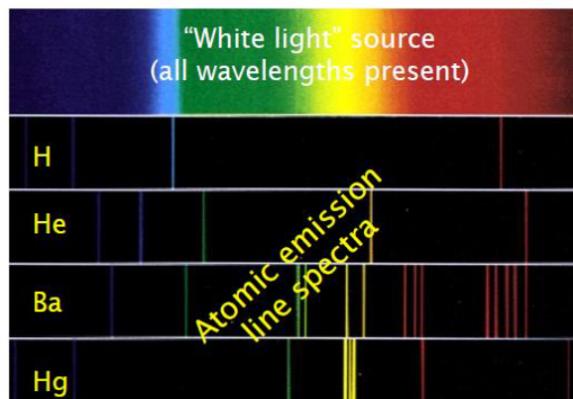


Figure 89.1: Absorption Spectra

1.3 Plum Pudding Model

A solid, uniform lump of positively charged matter, approximately 10^{-10} m across. This had evenly distributed negative charges (electrons) scattered throughout.

1.4 Discovery of the Nucleus

Geiger and Marsden (1908-1913), in an experiment designed by Rutherford fired alpha particles (He nuclei, mass of $4u$, charge of $+2e$) at thin gold foil and measured the deflection / scattering. The accelerating voltage gave these alpha particles an energy of ≈ 5 MeV.

They found that most α were scattered only by small angles, but (surprisingly) a small number were scattered right back towards to emitter (through $\theta > 90^\circ$). The distribution of the angles is approximately Normally distributed, with a mean of 0. Only approximately 1 in 8,000 fired α s were scattered by $\theta > 90^\circ$ back towards the emitter ("back-scattering").

Can back-scattering be explained with the Plum Pudding Model? No, it cannot. A plum pudding is too large, and has an insufficient charge density to produce the repulsion force required at the distances required.

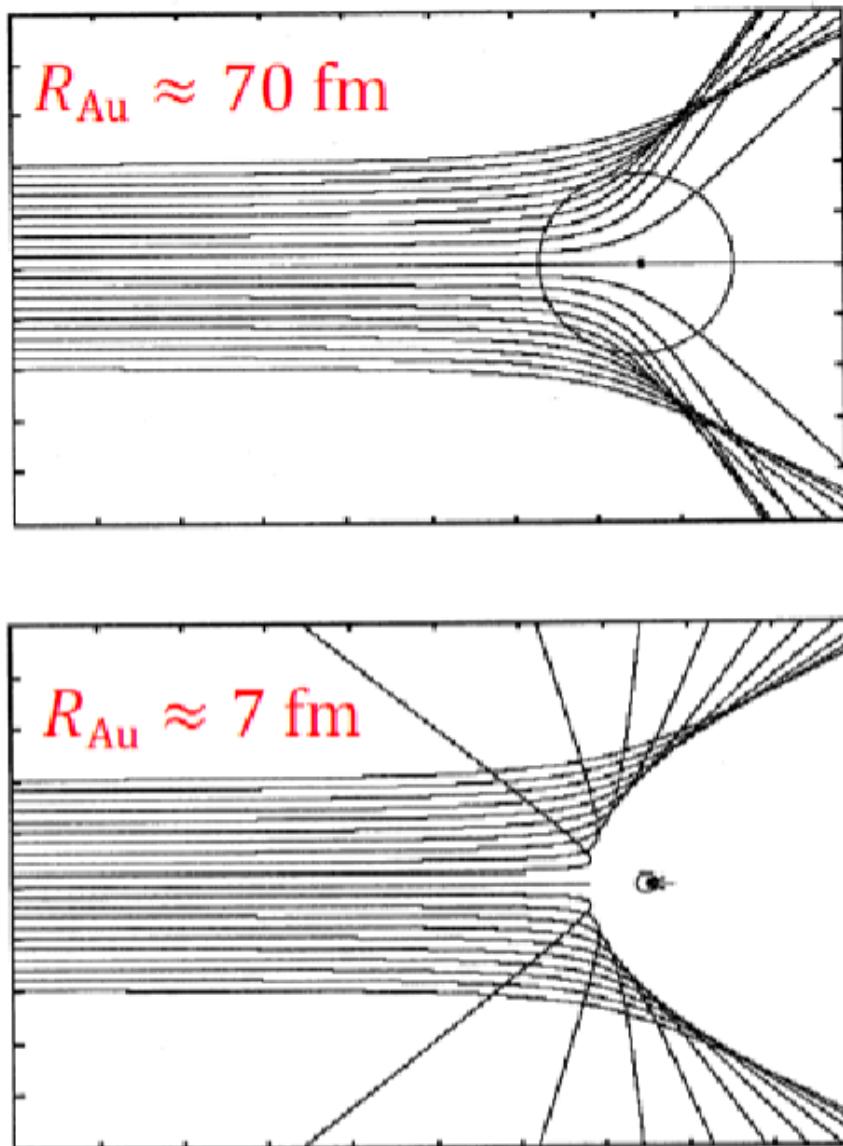


Figure 89.2: Rutherford Scattering Experiment Data

The first image is the theoretical data, from an atomic radius of 70 fm, consistent with the plum pudding model. The traces represent paths of incoming, then deflected alpha particles. Notably, the scattering is of small angles, less than 90°. The second image is data from a much smaller volume of the same charge, not consistent with the plum pudding model. Here, backscattering occurs, which (as this was experimentally observed) shows that the plum pudding model is not accurate.

2 Demonstrating by Calculation

Lets work out the work done to take an α from infinity to the pudding centre. If the electrostatic repulsion is not enough to overcome this, we cannot stop the α and cannot back scatter.

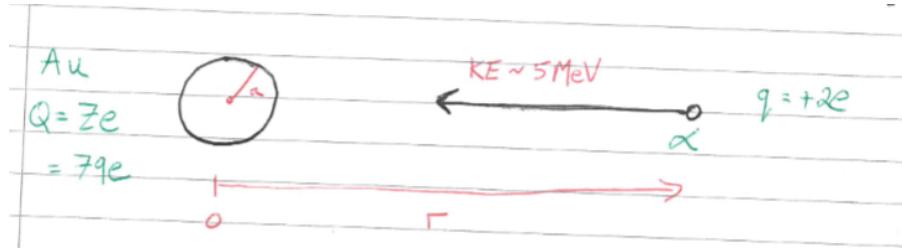


Figure 89.3: The experiment

2.1 Assumptions

- The atom stays still.
- Ignore the gold electrons (this is fine, as they would cancel some positive charge and make repulsion weaker. If we can't do it without them, it would be equally impossible to do it with).

2.2 Calculating Repulsive Force

Coulomb Potential Energy is:

$$u(r) = \frac{qQ}{4\pi\epsilon_0 r}$$

Force is:

$$F(r) = -\frac{du}{dr} = \frac{qQ}{4\pi\epsilon_0 r^2}$$

Change in potential energy ($u_2 - u_1$) is work done:

$$\int_{u_1}^{u_2} du = - \int_{r_1}^{r_2} F(r) dr$$

From outside the atomic radius, we treat the atomic pudding as a point charge of charge Q . From inside the atomic radius, we treat it as a smaller point charge $Q'(r)$, where we only consider the charge inside the portion of the pudding where $r < a$, where r is the current position inside the sphere and a is the atomic radius. We totally disregard any of the charge which sits at a greater radius than the current position.

If charge is spread uniformly, the total charge is proportional to the volume of the sphere. So:

$$\frac{Q'}{Q} = \frac{\frac{4}{3}\pi r^3}{\frac{4}{3}\pi a^3}$$

$$Q' = Q \frac{r^3}{a^3}$$

Inside the Pudding

$$F = \frac{qQ'}{4\pi\epsilon_0 r^2}$$

$$F = \frac{qQr^3}{4\pi\epsilon_0 r^2 a^3}$$

$$F = \frac{qQr}{4\pi\epsilon_0 a^3}$$

$$F = \frac{qQ}{4\pi\epsilon_0 a^3} \times r$$

Hence inside, $F \propto r$

Outside the Pudding

$$F = \frac{Qq}{4\pi\epsilon_0 r^2}$$

$$F = \frac{Qq}{4\pi\epsilon_0} \times \frac{1}{r^2}$$

Hence outside, $F \propto \frac{1}{r^2}$. We are therefore integrating the area under this (almost) triangle:

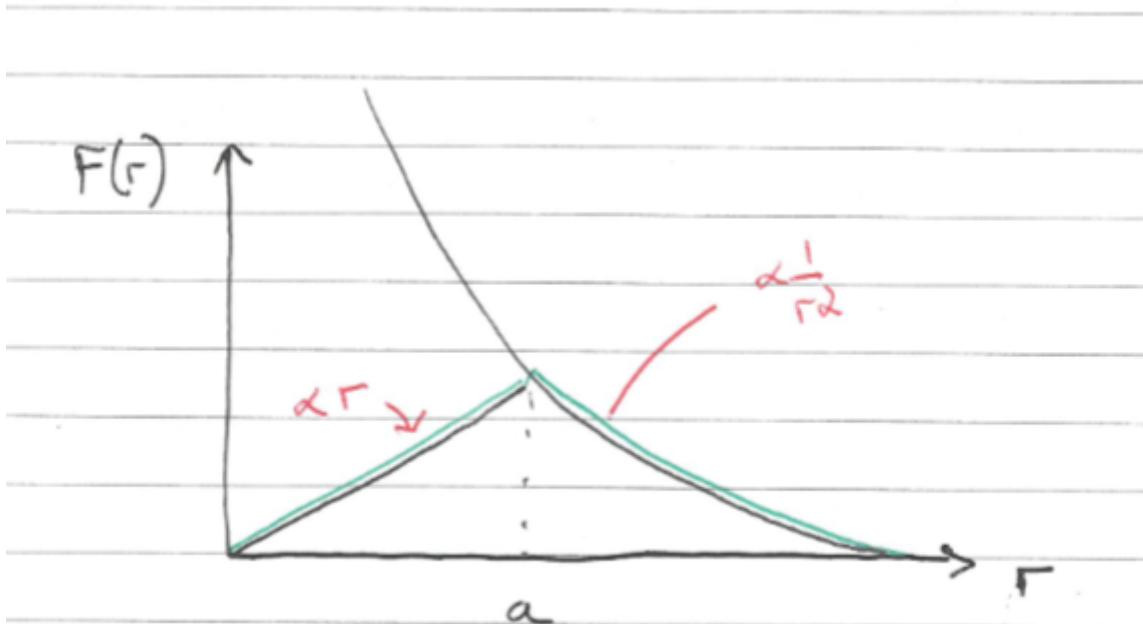


Figure 89.4: Radius vs electrostatic repulsion force

$$\Delta u = - \int_{r_1}^{r_2} F(r) dr$$

Splitting into two sections (when the alpha particle is outside vs inside the atomic radius), and integrating across all r (as we are attempting to work out the work done to bring an α from infinity to the charge, at which point distance is 0):

$$\begin{aligned} &= - \int_{\infty}^a \frac{qQ}{4\pi\epsilon_0 r^2} dr - \int_a^0 \frac{qQr}{4\pi\epsilon_0 a^3} dr \\ &= - \frac{qQ}{4\pi\epsilon_0} \int_{\infty}^a \frac{1}{r^2} dr - \frac{qQ}{4\pi\epsilon_0 a^3} \int_a^0 r dr \end{aligned}$$

$$\begin{aligned}
&= -\frac{qQ}{4\pi\epsilon_0} \lim_{x \rightarrow \infty} \int_x^a \frac{1}{r^2} dr - \frac{qQ}{4\pi\epsilon_0 a^3} \int_a^0 r dr \\
&= -\frac{qQ}{4\pi\epsilon_0} \lim_{x \rightarrow \infty} \left[-\frac{1}{r} \right]_x^a - \frac{qQ}{4\pi\epsilon_0 a^3} \left[\frac{1}{2} r^2 \right]_a^0 \\
&= -\frac{qQ}{4\pi\epsilon_0} \left(-\frac{1}{a} - 0 \right) - \frac{qQ}{4\pi\epsilon_0 a^3} \left(\frac{1}{2} 0^2 - \frac{1}{2} a^2 \right) \\
&= -\frac{qQ}{4\pi\epsilon_0} \left(-\frac{1}{a} \right) - \frac{qQ}{4\pi\epsilon_0 a^3} \left(-\frac{1}{2} a^2 \right) \\
&= \frac{qQ}{4\pi\epsilon_0 a} + \frac{qQa^2}{8\pi\epsilon_0 a^3} \\
&= \frac{qQ}{4\pi\epsilon_0 a} + \frac{qQ}{8\pi\epsilon_0 a} \\
&= \frac{qQ}{4\pi\epsilon_0 a} + \frac{1}{2} \frac{qQ}{4\pi\epsilon_0 a} \\
&= \frac{3}{2} \frac{qQ}{4\pi\epsilon_0 a}
\end{aligned}$$

As required! Plugging in values gives us:

$$\begin{aligned}
\Delta u &= \frac{3}{2} \frac{(2e)(79e)}{4\pi(8.854 \times 10^{-12}) \times 10^{-10}} \\
&= 5.45 \times 10^{-16} \text{ J} = 3.41 \text{ keV}
\end{aligned}$$

This is much less than the kinetic energy of the 5 MeV alpha particle, therefore (as this value is maximum work done against the repulsive force) a plum pudding could not stop, and therefore could not backscatter a 5 MeV alpha particle. However, since $\Delta u \propto 1/a$, a smaller volume of charge could. How small, however?

$$\begin{aligned}
\Delta u = 5 \text{ MeV} &= - \int_{\infty}^{r_{max}} \frac{qQ}{4\pi\epsilon_0 r^2} dr \\
-5 \text{ MeV} &= \frac{qQ}{4\pi\epsilon_0} \int_{\infty}^{r_{max}} \frac{1}{r^2} dr \\
-5 \text{ MeV} &= \frac{qQ}{4\pi\epsilon_0} \lim_{x \rightarrow \infty} \left[-\frac{1}{r} \right]_x^{r_{max}} \\
-5 \text{ MeV} &= \frac{qQ}{4\pi\epsilon_0} \left[-\frac{1}{r_{max}} - \lim_{x \rightarrow \infty} \frac{1}{x} \right] \\
-5 \text{ MeV} &= \frac{qQ}{4\pi\epsilon_0} \left[-\frac{1}{r_{max}} \right] \\
5 \text{ MeV} &= \frac{qQ}{4\pi\epsilon_0 r_{max}} \\
r_{max} &= \frac{qQ}{4\pi\epsilon_0 (5 \text{ MeV})}
\end{aligned}$$

Substitution and rearrangement gives $r_{max} = 4.5 \times 10^{-14} \text{ m} = 45 \text{ fm}$. This gives us the “distance of closest approach”. The nucleus cannot be any smaller than this, or an incoming alpha particle would collide with it. This is not the true size of the nucleus (a gold nucleus is smaller at $\approx 7 \text{ fm}$), but an alpha particle is not energetic enough to get this close. It instead gives the maximum size.

2.3 Next Idea: The Solar System Model

Therefore, the next idea was an orbiting solar system model, where electrons orbit in fixed paths around a central nucleus. However, accelerating charges (i.e. a charge in circular motion) radiate energy, so this orbiting electron would be on a decaying path to crash into the nucleus. We can observe this does not happen, so need another idea...

Bohr made two postulates:

- The electron in hydrogen moves in a set non-radiating circular orbit.
- Radiation is only emitted or absorbed when an electron moves from one orbit to another.

This works (at least for hydrogen) and explains the absorption spectra, but for now lacks a physical grounding.

Fri 10 Oct 2025 12:00

Lecture 2 - The Ultraviolet Catastrophe

In this lecture:

- How classical theories fail to explain black body radiation (“The Ultraviolet Catastrophe”).
- How quantising light into photons gives predictions that fit this observation.

1 Black Body Radiation

A ‘black body’ is an idealised perfect object, that does not reflect, and absorbs internally all light (regardless of wavelength) incident upon it. No light is transmitted, so nothing shines out the other side. The object is perfectly black.

All bodies emit electromagnetic energy, usually outside the visible portion of the spectrum. For example, Paul Hollywood (and other humans) emit at about 300K, which is infrared (at the temperature which night vision goggles are tuned to).

For the black body, emission spectrum is **only** from this thermal emission (no reflection, no fluorescence, etc). Hotter objects are brighter and bluer (hotter means higher energy, and therefore a shorter wavelength)

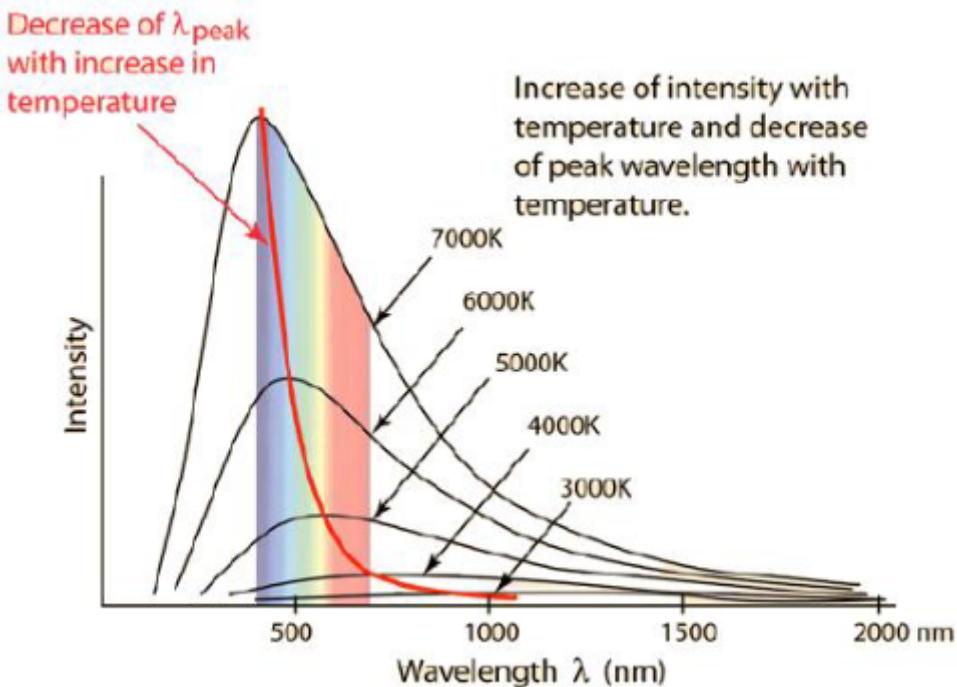


Figure 90.1: Observed Emission Spectra

However, we run into a problem. If we plot the spectra predicted by classical thermodynamics, vs the observed spectra for a given temperature object, the classical prediction gets it totally wrong, especially at shorter wavelengths.

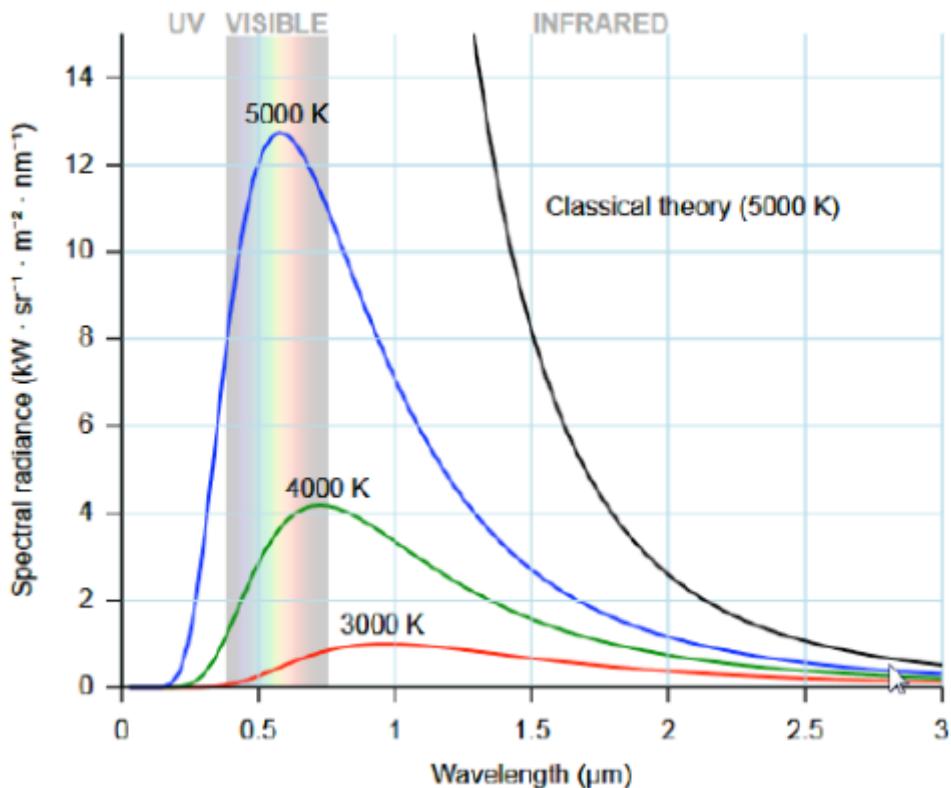


Figure 90.2: Predicted and Observed Spectra for 5000K (and Observed for 4000K and 3000K)

1.1 Notation

$I(\lambda)$ is the intensity for an emitted wavelength λ . I is the total intensity across all wavelengths per unit time (in W/m^2 , power per unit area).

$$I = \int_0^\infty I(\lambda) d\lambda$$

I is the total area under the $I(\lambda)$ curve, i.e. the sum of intensity per wavelength, across every wavelength.

2 The Ultraviolet Catastrophe

2.1 Empirical Results

The Stefan-Boltzmann Law gives $I = \sigma T^4$, where σ is the Stefan-Boltzmann constant, $\sigma = 5.67 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$.

Wien's Displacement Law gives $\lambda_{\text{peak}} = \frac{b}{T}$, where $b = 2.898 \times 10^{-3} \text{ K} \cdot \text{m}$.

2.2 Why does classical mechanics break?

Lets model the $I(\lambda)$ spectrum by slotting standing waves into a cavity. Inside the blackbody, EM waves form standing waves (in a limited number of possible configurations).

We can simplify by considering a 1D cavity of length L . We can consider 'cavity modes' as the possible standing waves that can exist in this cavity. As we know the wave is bound at each end, the displacement at each end of the cavity must be 0. Therefore, the only possible waves must obey this, and these are cavity modes.

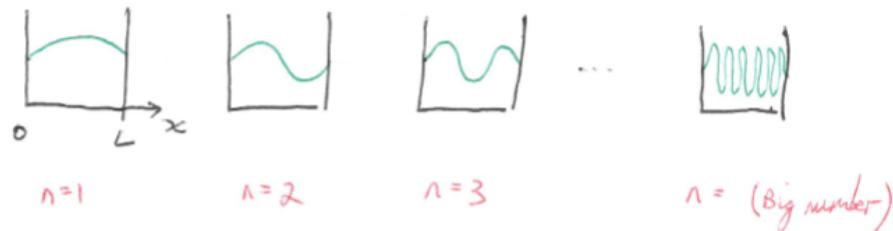


Figure 90.3: Possible cavity modes

The amplitude $a(x)$ can be given by this:

$$a(x) = \sin\left(\frac{n\pi x}{L}\right), \quad n = 1, 2, 3, \dots \quad (\text{mode number})$$

And by inspection from the figures:

$$\lambda = \frac{2L}{n}$$

And therefore the number of nodes per wavelength is:

$$n(\lambda) = \frac{2L}{\lambda}$$

Moving to a 3D cavity, the number of modes allowed in a volume scales with volume, meaning the density of modes becomes proportional to $1/\lambda^4$.

So, classically (in 3D):

$$I(\lambda) \propto \frac{n(\lambda)}{\lambda} \times k_B T \propto \frac{1}{\lambda^4}$$

Where the first term is the density of modes at lambda, and the second is the average energy of nodes. As we head to UV and $\lambda \rightarrow 0$, $I(\lambda) \rightarrow \infty \dots$ which is not accurate. This is the UV Catastrophe!

2.3 Where did it go wrong?

The issue was assuming that all cavity modes have average energy $k_B T$ - the “Equipartition Theorem” (which we’ll meet in later courses).

In brief: the probability distribution of energies is a “Boltzmann Distribution”:

$$p(E) = \frac{\exp\left(-\frac{E}{k_B T}\right)}{k_B T}$$

Average energy:

$$\bar{E} = \int_0^\infty E p(E) dE = k_B T$$

We assume that all cavity modes have the same average energy, and yet as lambda becomes very small, the number of possible cavity modes becomes very large. This causes the energy to tend to infinity, which the UV Catastrophe says is incorrect...

2.4 Planck’s Hypothesis

A rather desperate Planck hypothesised that energy was quantised, i.e. it comes in discrete packets, called quanta. The energy of these quanta is proportional to frequency. This was radical at the time, even though we accept it now.

$$\Delta E = hf = \frac{hc}{\lambda}$$

Where $h = 6.626 \times 10^{-34} \text{ J} \cdot \text{s}$ is Planck's constant.

Sticking this into the Partition Function from statistical mechanics (which we will properly encounter later on, for now don't worry!), we get an average energy:

$$\bar{E}(\lambda) = \frac{hc/\lambda}{\exp(hc/\lambda k_B T) - 1}$$

Looking at limits:

$$\text{For } \bar{E}(\lambda \rightarrow \infty) : \quad \frac{hc}{\lambda k_B T} \ll 1$$

And the Taylor Series of e^x :

$$\exp\left(\frac{hc}{\lambda k_B T}\right) \approx 1 + \frac{hc}{\lambda k_B T} + \dots$$

Yields:

$$\bar{E}(\lambda) \approx \frac{hc/\lambda}{1 + (hc/\lambda k_B T) - 1} = k_B T$$

This is a good sign, because it means that Planck's Hypothesis holds the correct classically predicted and empirically observed behaviour for higher wavelengths.

Now what about lower wavelengths, where the classical behaviour broke?

$$\text{For } \bar{E}(\lambda \rightarrow 0) : \quad \exp\left(\frac{hc}{\lambda k_B T}\right) \rightarrow \infty \quad (\text{very quickly})$$

$$\text{For } \bar{E}(\lambda \rightarrow 0) : \quad \frac{hc}{\lambda} \rightarrow \infty \quad (\text{slower})$$

Therefore, in the expression:

$$\bar{E}(\lambda) = \frac{hc/\lambda}{\exp(hc/\lambda k_B T) - 1}$$

The numerator and denominator both tend to infinity, but the denominator does so much faster. Therefore (and this can be done in a less handwavy manner via L'Hopital):

$$\bar{E}(\lambda \rightarrow 0) \rightarrow \frac{1}{\infty} \rightarrow 0$$

Which recovers the behaviour at UV wavelengths, so no Catastrophe!

3 Conclusion

- This strange quantisation hypothesis actually fits the data.
- Quantising energy means that the average energy of each cavity mode is wavelength dependant, and not fixed $k_B T$ as seen at larger wavelengths.
- This solves the UV Catastrophe!

Thu 17 Oct 2025 12:00

Lecture 3 - Particle Nature of Light

In this lecture:

- The photoelectric effect.
- Compton scattering.

Which are two examples where classical theory (light as a wave) break down.

1 The Photoelectric Effect

When shining ultraviolet light on a metal surface, electrons are emitted. This is the photoelectric effect.

Why are we not bombarded by electrons in daily life? For the electron to fly off, we must be in a vacuum. Otherwise, it'll immediately strike an air molecule and be absorbed.

Photoelectric Effect Background

- Discovered by Hertz, 1887
- Thomson (1889) went further, so did Lenard (1902) and others.
- Einstein won his Nobel Prize for explaining this, not from relativity.

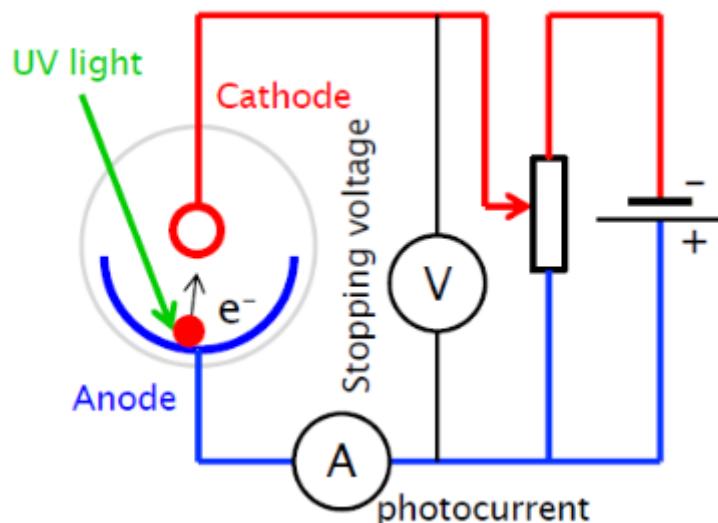


Figure 91.1: A circuit diagram for measuring the photoelectric effect.

The above setup would be encased in a glass ball (containing a vacuum), with a setup like this:

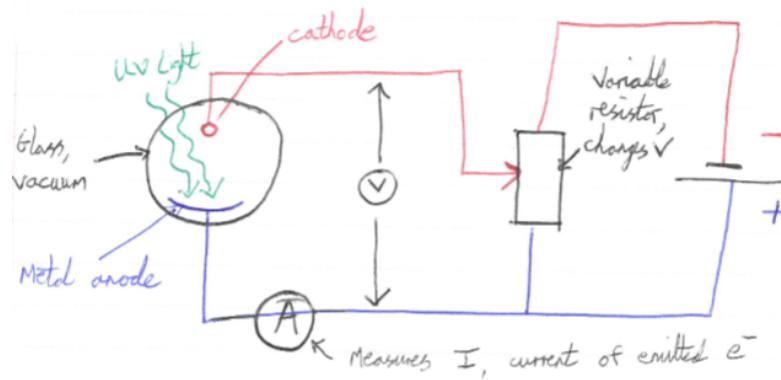


Figure 91.2: Experimental Setup

1.1 Results

Result One - Changing Intensity

For fixed UV wavelength, increasing the intensity of light increases the measured photocurrent:

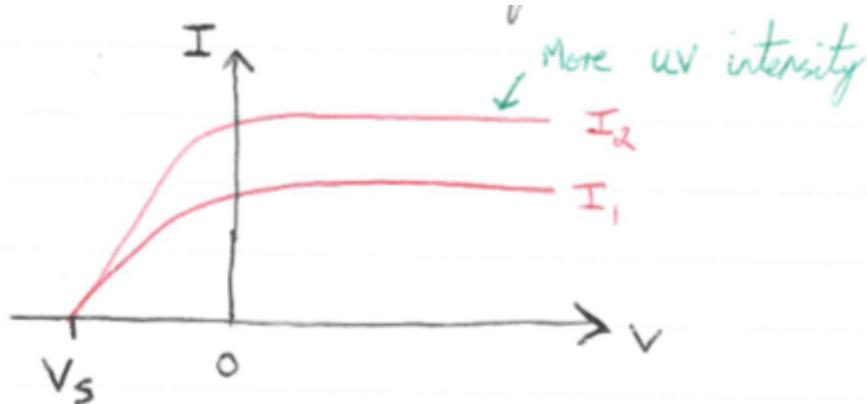


Figure 91.3

- Increasingly negative potential the cathode decreases photocurrent. At some potential v_s applied to the circuit (the “stopping potential”) this current drops to zero.
- Potential does not affect electron emission, however adding potential causes an electric field which effectively blows electrons back towards the anode. The stopping potential is when this electric field is perfectly strong to prevent electrons from reaching the cathode and causing a current.
- The fact this can happen consistently (i.e. no current means no electrons made it through) implies that there must be some maximum kinetic energy these electrons can have ($KE_{max} = eV_s$).
- The stopping potential is independent of UV intensity. More UV makes current increase, but does not change stopping potential (i.e. it does not give more energy to each electron, they each have the same energy). This does not make sense classically. Classically we would expect adding more energy to cause emitted electrons to have more energy, therefore changing the stopping potential.

Result Two - Changing Wavelength

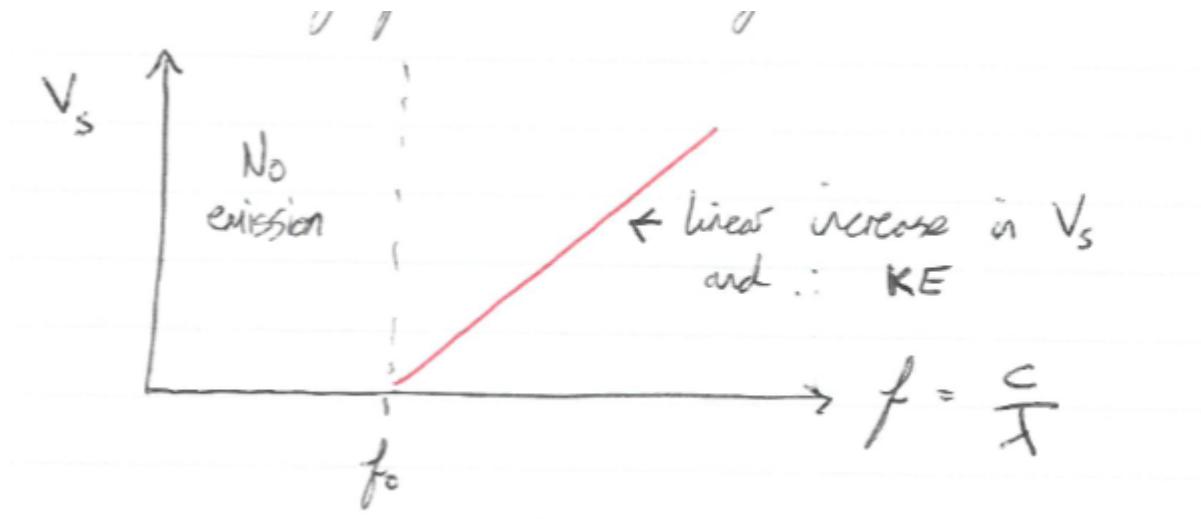


Figure 91.4

We have to reach some baseline threshold frequency f_0 before we see any photocurrent. After this, increasing wavelength increases photocurrent (and hence KE of emitted electrons) linearly.

- For a given metal, we find the threshold frequency f_0 , below which there is no emission of electrons (no current). If below the frequency f_0 , intensity is irrelevant. This contradicts classical mechanics which would suggest that turning up the light intensity would supply more (and potentially sufficient) energy.
- Above the threshold, the energy of individual emitted photons depends on UV frequency and not intensity (by result one).

1.2 Conclusions

Classically

Classically, we expect energy to be proportional to intensity, $\therefore v_s$ should increase with greater intensity. We also expect there to be no link between frequency and energy, hence no threshold frequency. We'd expect no threshold frequency, instead being a time delay as electrons "soak up" energy to reach the required threshold.

In theory, great, in practice *this is not observed*.

Einstein's Proposal

Energy in light comes from photons with energy $E = hf$. There is a minimum energy required for an electron to be able to escape from the metal. This minimum energy is called the work function ϕ .

$$KE_{\max} = hf - \phi = eV_s$$

Now:

- Higher intensity means more of the same particles (more photons), but the energy of each is unchanged.
- $E = hf$ so frequency changes energy (as observed).
- The Bohr model says that an electron can only have certain electron energy transitions when the correct energy is supplied (an electron cannot gradually soak up energy). This explains why there is a cutoff below the work function, and no observed time delay (as the "soaking up" that causes the delay does not happen). Either an incoming photon has sufficient energy, or it does not. Having more photons does not help.
- The first incoming photons immediately releases an electron (assuming the incoming light has sufficient energy), therefore there's no time delay.

1.3 In Practice

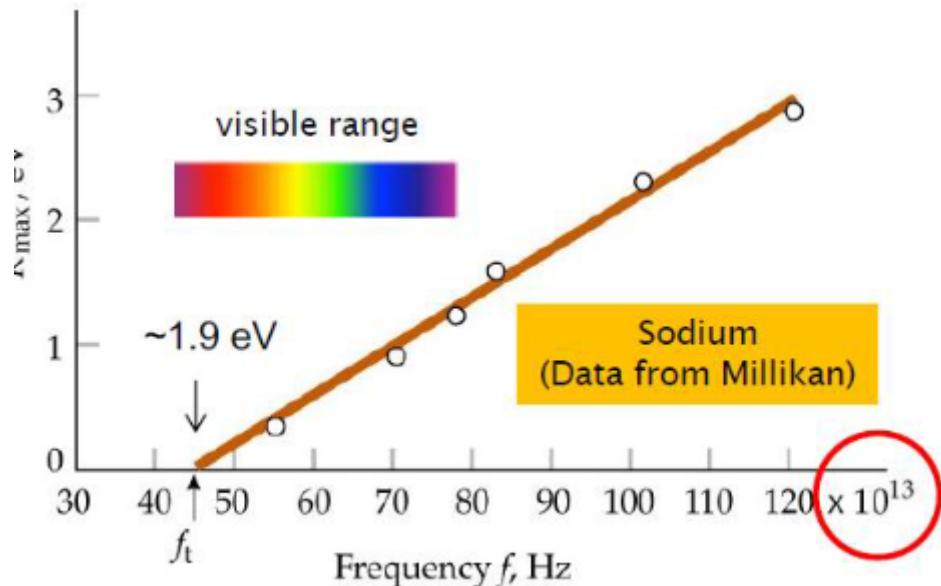


Figure 91.5: Sodium photocurrent measurements by Robert Millikan

2 Compton Scattering

Compton Scattering is the scattering of x-rays off carbon atoms.

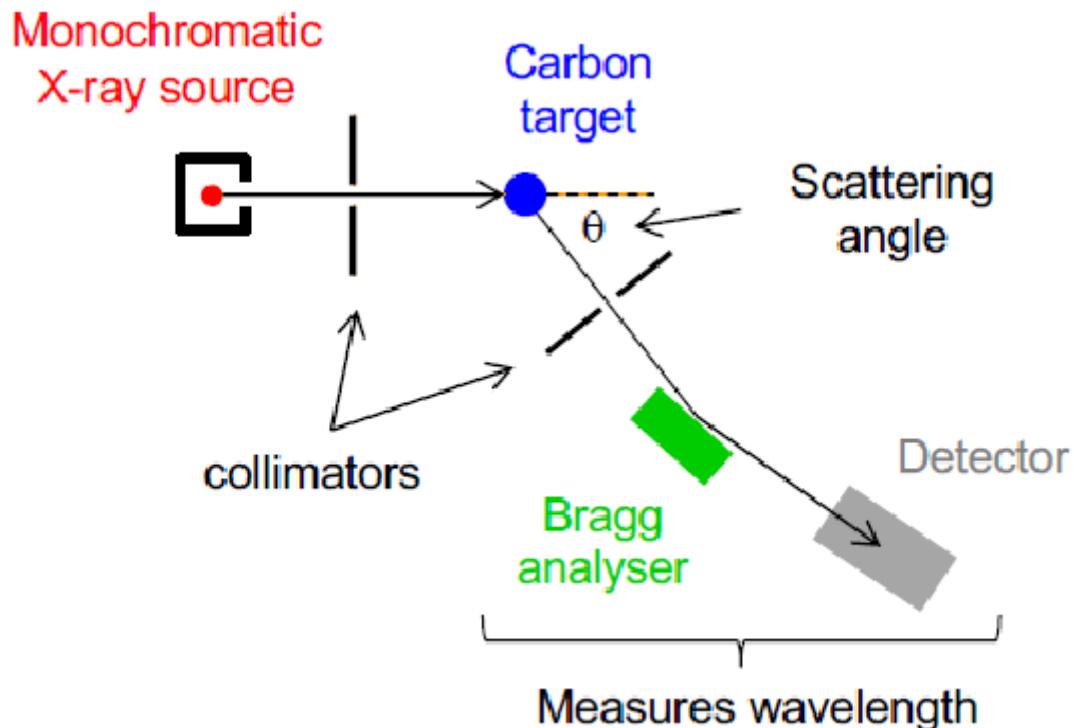


Figure 91.6: A Compton Scattering experimental setup.

The surprising result is that two wavelengths were observed (not just the original) - λ_1, λ_2 , where λ_1 is the original and λ_2 is different. Classically this is hard to explain and λ should not change.

The difference between these two wavelengths increases with scattering angle θ . This can be explained if the x-ray beam is a stream of photons, but not classically.

Possible options when a photon collides:

- Scattering off the whole atom (Rayleigh Scattering). The photon hits a tightly bound electron that cannot move independently. The photon effectively scatters off the entire atom. Since $M_{\text{atom}} \gg m_e$, the recoil energy of the atom is minimal and the photon loses almost no energy, hence the wavelength is effectively unchanged. This gives us the existing λ_1 peak.
- Scattering off a free electron (Compton Scattering): The photon hits a loosely bound electron which has a much lower binding energy, so the electron is knocked loose. The photon transfers momentum and energy to this electron, leading to a loss of energy and therefore change of wavelength. This gives the new λ' peak.

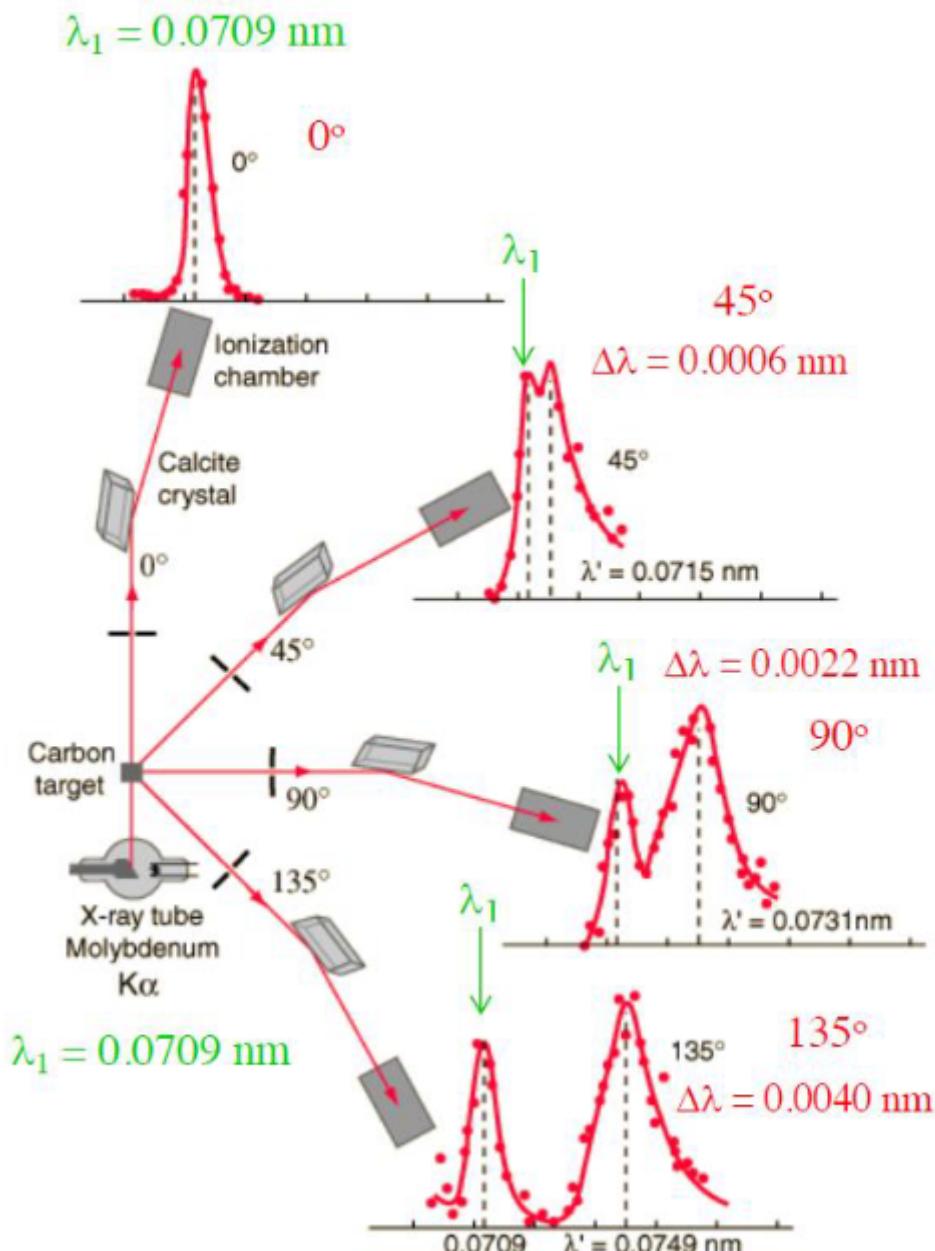


Figure 91.7: The observed results

3 Deriving Compton's Equation

Given an incoming photon with energy E_1 , wavelength λ_1 and momentum p_1 . This strikes an electron and is deflected by angle θ . The electron is deflected by some angle ϕ such that momentum is conserved. The new deflected photon has E_2 , λ_2 , p_2 .

We must consider relativistic effects here given the high speed ($E^2 = p^2c^2 + m^2c^4$)

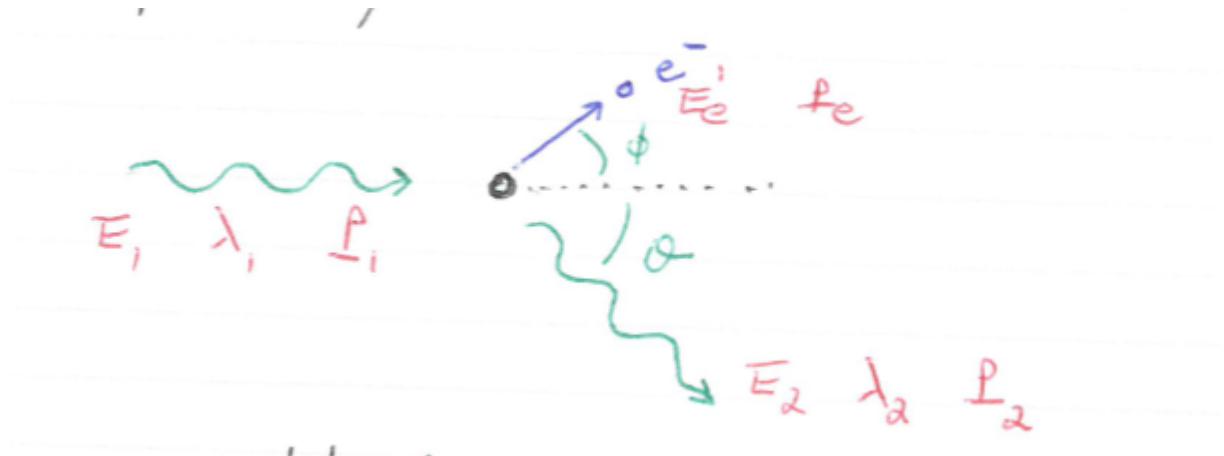


Figure 91.8

3.1 Setup

For a massless photon ($m = 0$):

$$E = pc$$

And:

$$E = \frac{hc}{\lambda}$$

So:

$$p = \frac{h}{\lambda} \quad (91.1)$$

3.2 Conservation and Relativity

Conserving momentum (underlines omitted for speed):

$$p_1 = p_e + p_2$$

$$p_e = p_1 - p_2$$

Squaring both sides:

$$\begin{aligned} p_e^2 &= p_1^2 + p_2^2 - 2p_1 \cdot p_2 \\ p_e^2 &= p_1^2 + p_2^2 - 2p_1 p_2 \cos \theta \end{aligned} \quad (91.2)$$

And then by conservation of energy:

$$E_1 + E_e = E_2 + E'_e$$

Where E_1 is the incoming photon energy, E_e is the energy of the electron at rest in atom before collision, E_2 is the deflected photon energy and finally E'_e is the deflected electron's energy.

Using this:

$$\begin{aligned} p_1 c + m_e c^2 &= p_2 c + \sqrt{p_e^2 c^2 + m_e^2 c^4} \\ \implies p_1 - p_2 + m_e c &= \sqrt{p_e^2 + m_e^2 c^2} \end{aligned}$$

And squaring both sides:

$$(p_1 - p_2)^2 + \cancel{m_e^2 c^2} + 2m_e c(p_1 - p_2) = p_e^2 + \cancel{m_e^2 c^2}$$

Substituting in Eqn 91.2 for p_e^2

$$\begin{aligned}(p_1 - p_2)^2 + 2m_e c(p_1 - p_2) &= p_e^2 \\ (p_1 - p_2)^2 + 2m_e c(p_1 - p_2) &= p_1^2 + p_2^2 - 2p_1 p_2 \cos \theta\end{aligned}$$

And rearranging:

$$\begin{aligned}(p_1 - p_2)^2 + 2m_e c(p_1 - p_2) &= p_1^2 + p_2^2 - 2p_1 p_2 \cos \theta \\ p_1^2 + p_2^2 - 2p_1 p_2 + 2m_e c(p_1 - p_2) &= p_1^2 + p_2^2 - 2p_1 p_2 \cos \theta \\ -2p_1 p_2 + 2m_e c(p_1 - p_2) &= -2p_1 p_2 \cos \theta \\ -p_1 p_2 + m_e c(p_1 - p_2) &= -p_1 p_2 \cos \theta \\ m_e c(p_1 - p_2) &= -p_1 p_2 \cos \theta + p_1 p_2 \\ m_e c(p_1 - p_2) &= p_1 p_2(1 - \cos \theta)\end{aligned}$$

Substituting Eqn 91.1:

$$\begin{aligned}m_e c(p_1 - p_2) &= p_1 p_2(1 - \cos \theta) \\ m_e c\left(\frac{h}{\lambda_1} - \frac{h}{\lambda_2}\right) &= \frac{h}{\lambda_1} \frac{h}{\lambda_2} (1 - \cos \theta) \\ m_e c\left(\frac{h}{\lambda_1} - \frac{h}{\lambda_2}\right) &= \frac{h^2}{\lambda_1 \lambda_2} (1 - \cos \theta) \\ m_e c\left(\frac{1}{\lambda_1} - \frac{1}{\lambda_2}\right) &= \frac{h}{\lambda_1 \lambda_2} (1 - \cos \theta) \\ m_e c\left(\frac{\lambda_1 \lambda_2}{\lambda_1} - \frac{\lambda_1 \lambda_2}{\lambda_2}\right) &= h(1 - \cos \theta) \\ m_e c(\lambda_2 - \lambda_1) &= h(1 - \cos \theta) \\ (\lambda_2 - \lambda_1) &= \frac{h}{m_e c}(1 - \cos \theta)\end{aligned}$$

Which is the Compton Equation. This shows that the change in wavelength is proportional to $1 - \cos \theta$.

4 Conclusions

The photoelectric effect and Compton scattering are two more physical phenomena that cannot be explained using traditional classical mechanics with EM waves alone. They both require assuming photons of energy $E = hf$ to be adequately explained.

Fri 24 Oct 2025 12:00

Lecture 4 - Atomic Energy Levels and Spectra

In this lecture:

- The spectra of light emitted and absorbed by electrons in:
 - Hydrogen (simple)
 - Larger atoms (not simple...)
- Electronic shells and orbits.

1 Spectra

The Bohr Model says that electrons in an atom can hold discrete values “levels” of energy. As electrons go up or down these levels they must absorb or will emit a photon. This emission causes a discrete spectra of emitted frequencies, unique to the element causing it.

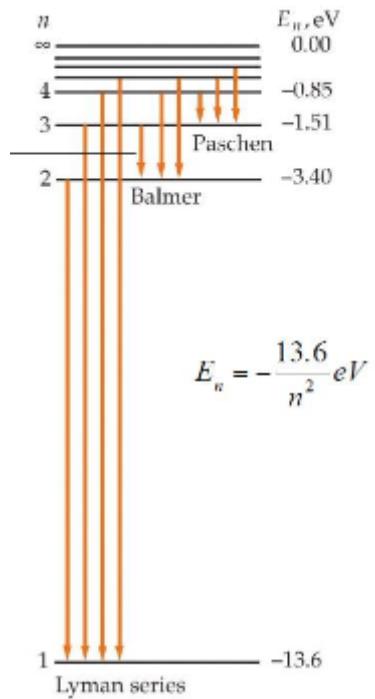


Figure 92.1: Example emission spectra.

Each of the transitions from one energy level to another have a discrete change in energy (measured in eV), therefore each transition will have a discrete wavelength of produced photons.

From the Bohr model, he postulated (from experimental observations) that light can *only* be absorbed or emitted when an electron goes up or down a discrete energy level (excitation or relaxation). The lowest energy level is known as the “Ground State” ($n = 1$).

2 Absorption and Emissions

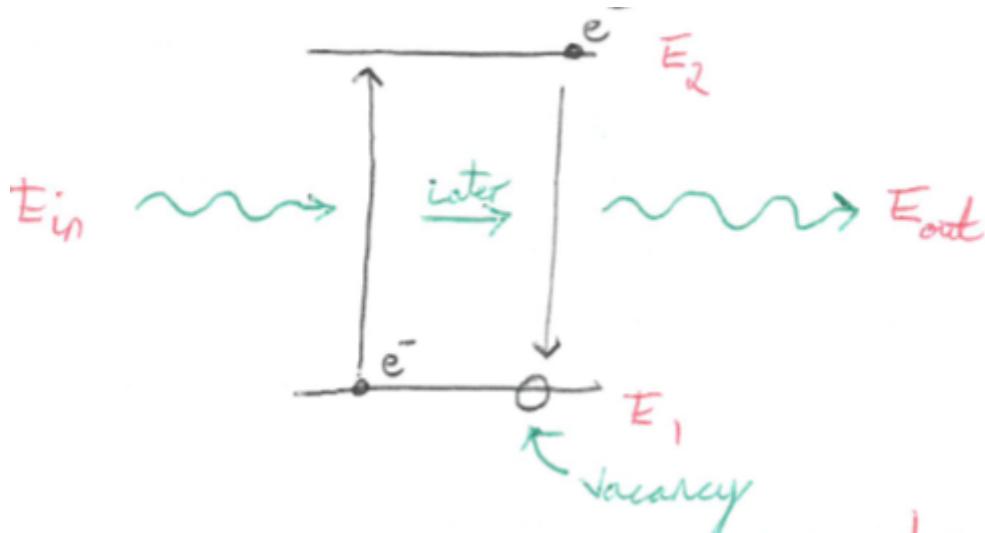


Figure 92.2

An electron is excited up an energy level, leaving a vacancy behind. Some time after, the electron drops back down into the ground state, emitting the energy in terms of a photon. By conservation of energy:

$$E_{\text{in}} = E_{\text{out}} = E_2 - E_1 = hf = \frac{hc}{\lambda}$$

Notably, the energy of the photon that initially triggered the excitation is given by $E_\gamma = E_2 - E_1$

3 Atomic Hydrogen

Atomic Hydrogen is by far the simplest example we can deal with (given the single proton and single electron). This single electron is ‘orbiting¹’ the nucleus at some distance r .

The potential energy of this electron is:

$$v(r) = \frac{-e^2}{4\pi\epsilon_0 r}$$

This P.E. is negative, so the electron is trapped in a potential well and must be supplied with energy to escape, assuming no K.E. We also know (if we take into account kinetic energy too):

$$E = T + V$$

Where E is total energy, T is K.E, V is P.E. If the total is negative, the electron is bound to the atom.

We can see from the equation for potential energy that $v(r) \propto -1/r$, which gives us the green line. Of the points on this green line, only some of them are actually discrete allowed orbits. The region at the top is called the continuum, which is where an electron has positive energy and has therefore left the nucleus.

¹for the sake of argument, even if in practice it does not really

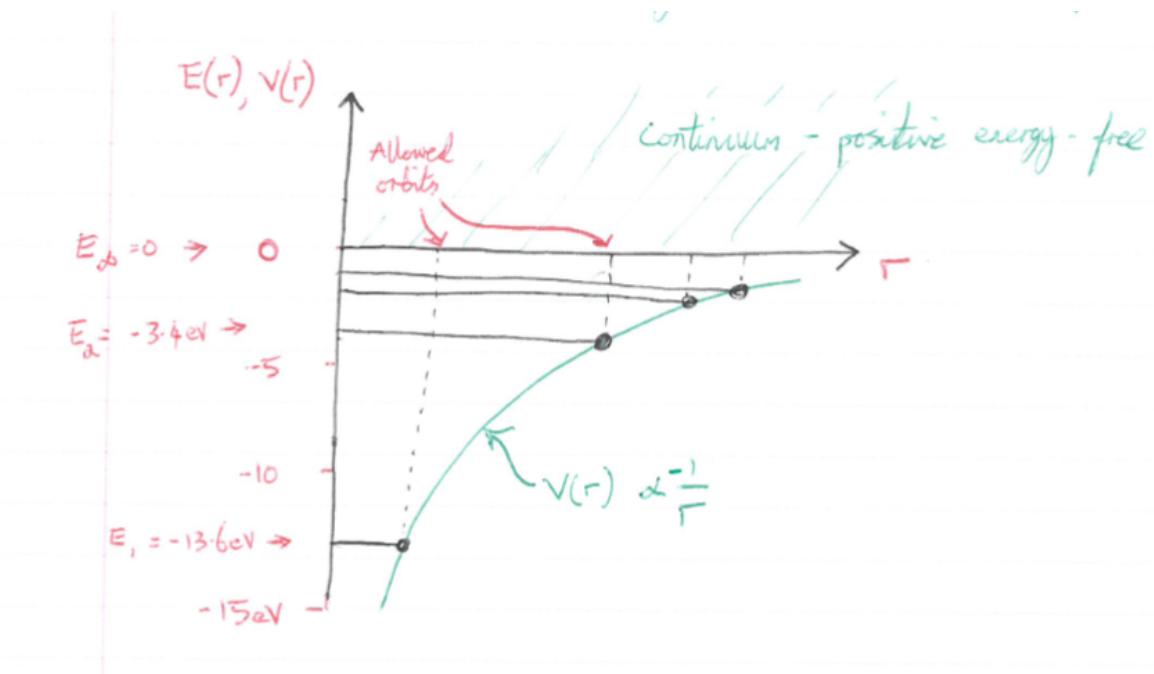


Figure 92.3

In Lec 08, we will properly derive from the Bohr model that the energy of each energy level is given by this:

$$E_n = \frac{-13.6\text{eV}}{n^2} \quad n = 1, 2, 3, \dots, \infty$$

Note that if $n \rightarrow \infty$, $E_n \rightarrow 0$. Emission and absorption happens when an electron moves from two different energy levels ($m \rightarrow n$).

$$E_\gamma = E_n - E_m = 13.6(1/n^2 - 1/m^2)\text{eV}$$

$$\frac{hc}{\lambda} = E_n - E_m = 13.6(1/n^2 - 1/m^2)\text{eV}$$

Or, finally:

$$\frac{1}{\lambda} = \frac{13.6\text{eV}}{hc} \left(\frac{1}{n^2} - \frac{1}{m^2} \right)$$

The $\frac{13.6\text{eV}}{hc}$ term is known as the Rydberg constant.

3.1 Balmer Series

The 'Balmer Series' is a portion of the hydrogen emission spectra which happens to take place at visible wavelengths. It is specifically transitions which take place to $n = 2$. To find these wavelengths, we set $n = 2$:

$$\frac{1}{\lambda} = R \left(\frac{1}{2^2} - \frac{1}{n^2} \right)$$

Moving from $3 \rightarrow 2$ gives us the Balmer Alpha line, for example.

4 Ionisation Energy of Hydrogen

The ionisation energy is the minimum energy required to kick an electron out of the ground state and into the continuum. This is just enough energy to barely make it free (i.e. when the continuum is reached the electron has 0 KE). Setting $n = 1, m = \infty$:

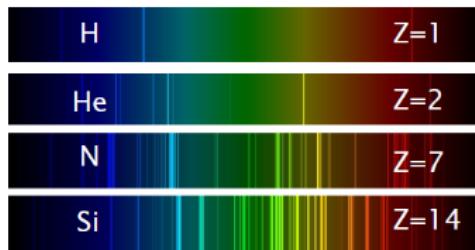
$$\frac{1}{\lambda} = R \left(\frac{1}{1} - \frac{1}{\infty} \right) = R$$

Where R is the Rydberg constant. Note: the energy levels can be very sharp, and they do not change (unless external stimuli like a strong external magnetic field are applied). We can therefore use this in e.g. atomic clocks.

5 Bigger Atoms

Bigger atoms gets more complex than we can reasonably consider now.

- Increasing complexity



- Sodium ($Z = 11$) energy levels

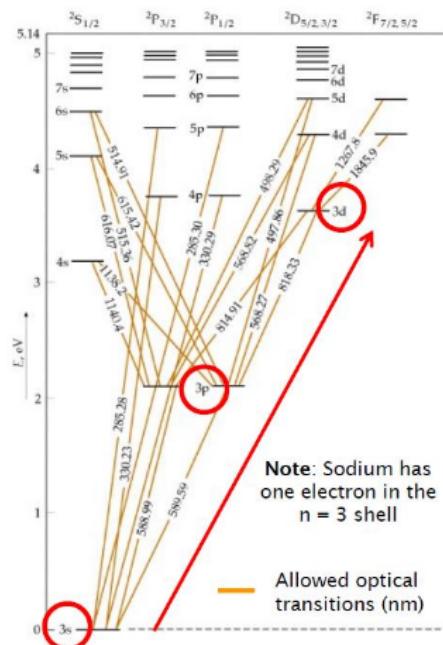


Figure 92.4

Electrons begin interacting with each other and the nucleus in weird ways, and it becomes too messy to calculate easily. If there is only a single electron, we can still use the Bohr model, just with a different charge of the nucleus, and we multiply by Z^2 when calculating. If we have multiple electrons, there is no general formula and it becomes unpleasant. The energy levels become split into multiple possible orbitals, i.e for $n = 1$, $1s$. For $n = 2$, $2s$, $2p$. For $n = 3$, $3s$, $3p$, $3d$ (where s , p , d are different electronic orbitals).

5.1 Spin

Electrons also have an intrinsic quantum property called ‘spin’, with a value $\frac{1}{2}$. They are fermions with two spin states (up, down). Note that spin is quantum, and has no relation to a physical geometric spin. The maximum occupancy of each level is $2n^2$ (i.e. the $1s$ state has space for one electron of each spin, i.e. one spin up and one spin down).

6 More Complexity...

- **Fine Splitting:** Electrons are moving, so have non-zero angular momentum. A moving charge creates a magnetic field, which changes the energy of the other electrons.
- **Hyperfine Splitting:** The nucleus also has a spin, they can be parallel (i.e. both have the same spin), or antiparallel, where the electron and the nucleus have opposing spins. These two configurations have slightly different energies (difference of $5.9\mu eV$). The parallel state is metastable, with a half life of 10 million years. There is so much hydrogen in the galaxy that radioastronomers can use this decay, and by detecting it they can image the location of hydrogen in the universe.

Fri 31 Oct 2025 12:00

Lecture 5 - X-Ray Production and Diffraction

In this lecture:

- Production of x-rays
- How to measure their wavelength
- Bragg scattering of x-rays by crystals and Bragg conditions.

1 Production of X-Rays

From the blackbody spectra lecture, we saw that something becoming hotter affects the wavelength of emission. Surely we, therefore, can just heat something hot enough for the peak of the emission spectra to be in the x-ray range?

Unfortunately not, this works, but would need the black body to be 1MK.

The more practical alternative is to fire high energy electrons (keV range) into a metal target in a vacuum.

2 What are they?

Electromagnetic radiation (light) outside of the visible spectrum. It sits beyond the UV portion.

Approximately 0.1keV ($\lambda = 10\text{nm}$) to 100keV ($\lambda = 0.01\text{nm}$), these are called soft and hard x-rays respectively.

3 Measuring Wavelength

How can we measure or select x-ray wavelengths? With visible light, we can use a prism to turn wavelength into an angle and measure the angle. However, a prism works for visible light because of its high refractive index. Unfortunately, for x-rays, $n \approx 1$ for all materials.

We therefore need some other way to split up wavelengths into angles, for which we can use a diffraction grating. Different wavelengths will be diffracted at different angles.

Another problem arises however... We need the slit separation d to be $d \approx \lambda$ for a diffraction grating to work. For x-rays, $\lambda \approx 10^{-10}\text{m}$, which is approximately the width of an atom. Good luck making that grating...

Luckily, nature has made these gratings for us — crystalline materials! Strontium titanate (SrTiO_3) has an approx 0.389nm spacing between strontium atoms. We can therefore use this crystal lattice as a diffraction grating.

4 X-Ray Diffraction from Crystals

We treat the x-rays as EM waves (and ignore weird photon stuff for now) and we assume that each atom scatters independently. We also assume that the atom absorbs the x-ray and later re-emits it, in all directions uniformly (at photon level, the emission is in a random direction, therefore we probabilistically treat it as in all).

We want to find the angles where *constructive interference* occurs. Zero intensity everywhere else due to very very high number of slots.

The Bragg Conditions determine where this interference happens. Consider a single plane of equally spaced atoms.

4.1 First Bragg Condition

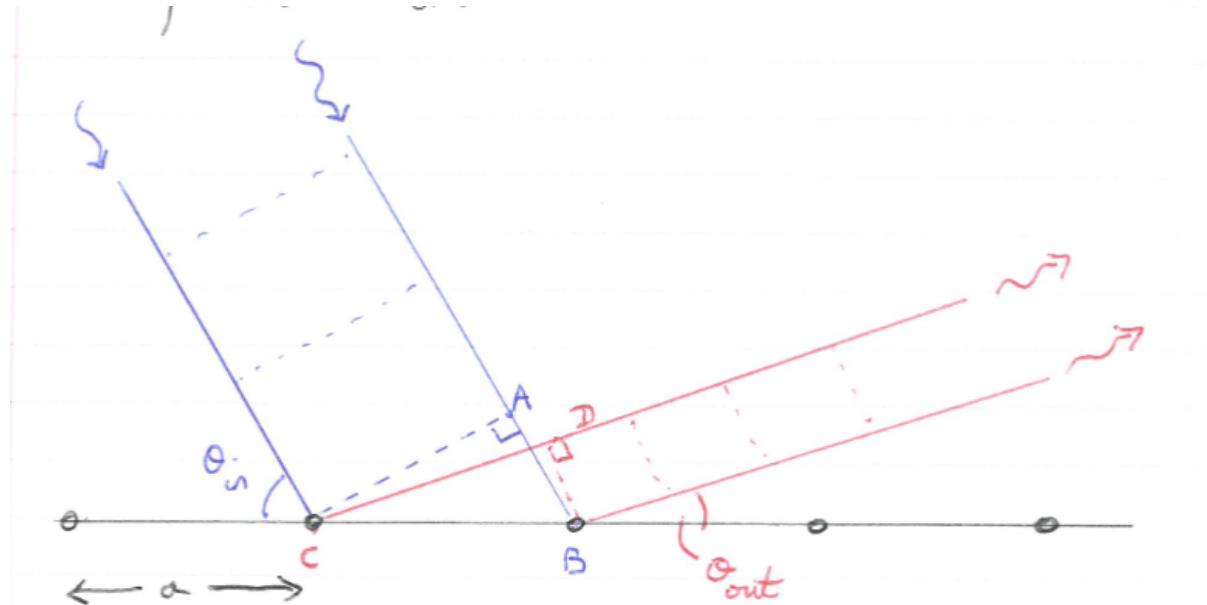


Figure 93.1

The outgoing waves (red) are all in phase when the wavefronts line up perpendicularly. This happens when path lengths **AB** and **CD** are the same. So:

$$a \cos \theta_{in} = a \cos \theta_{out}$$

$$\theta_{out} = \theta_{in}$$

We disregard higher orders where:

$$\cos(\theta_{out} + 2n\pi) = \cos(\theta_{in} + 2k\pi), \quad n, k \in \mathbb{Z}^+$$

And we only consider the case where they are explicitly equal. Note that, so far, this is independent of wavelength (i.e. there is 'no dispersion'). This means we're not acting like a diffraction grating yet, more of a mirror.

4.2 Second Bragg Condition

Now we consider the inclusion of a second plane of atoms (we have some much larger number of n planes).

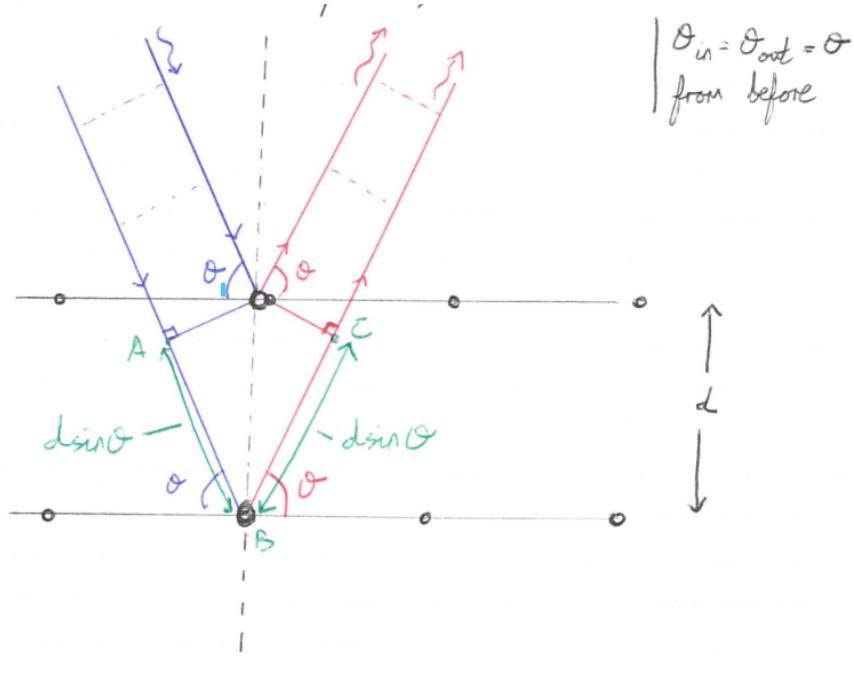


Figure 93.2

The ray which hits the lower plane travels further than the ray which hits the upper plane. For constructive interference to take place, this extra path difference must be an integer multiple of the wavelength. So:

$$AB + BC = 2d \sin \theta = n\lambda \quad n = 1, 2, 3, \dots$$

Where d is the separation of crystal planes. Note: This θ is not the beam deflection angle. It is the angle between the **horizontal** and the incoming beam.

This equation gives us different angles for different wavelengths, therefore we can select light (inc X-rays) of a certain wavelength by splitting multi-wavelength light into different angles, and physically selecting the one that we wish. This throws back to Lec 03, where a Bragg Analyser was used with a detector to measure the wavelength of the produced X-rays.

5 Uses

We can use this in two ways:

We can determine an unknown wavelength (or select a wavelength out of multi-frequency light) by varying θ and identifying the detection peak. Note that θ is here twice, the detector must also be moved, and therefore the total deflection angle is 2θ

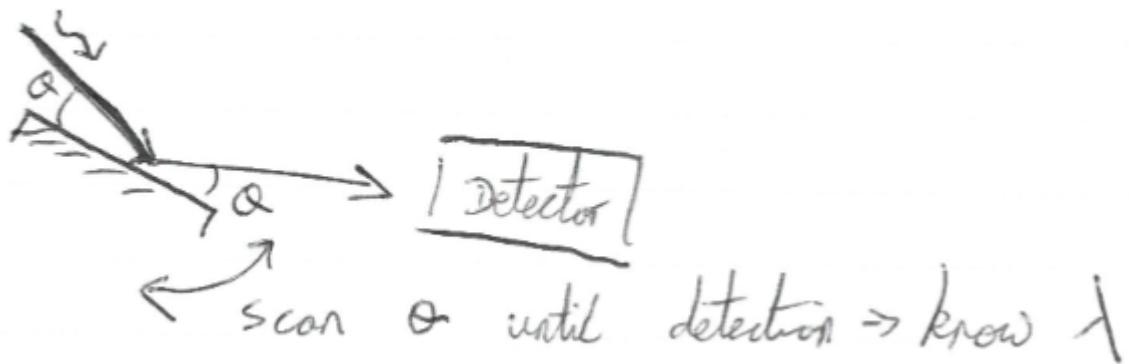


Figure 93.3

Alternatively, we can know lambda and fire at an unknown crystal to determine the atomic spacing. We can scan through 2θ and identify the angles where $n\lambda = 2d_i \sin \theta$ for various d_i . This leads us to x-ray crystallography, where we use x-rays to work out the structure of materials. Using crystals to measure or select an x-ray wavelength is called “wavelength dispersive x-ray spectroscopy”.

Fri 07 Nov 2025 12:00

Lecture 6 - X-Ray Spectra

In this lecture:

- The production of X-rays - the spectrum created.
- Attenuation of X-rays in matter.
- Absorption of X-rays in materials, and 'absorption edges'.

1 Typical Tube Spectra

As a reminder, the source is metal bombarded with high-energy electrons. This causes electron energy level changes in deep electron shells in the metal. This spectrum is characteristic of the anode (the thing being hit by the electrons) material.

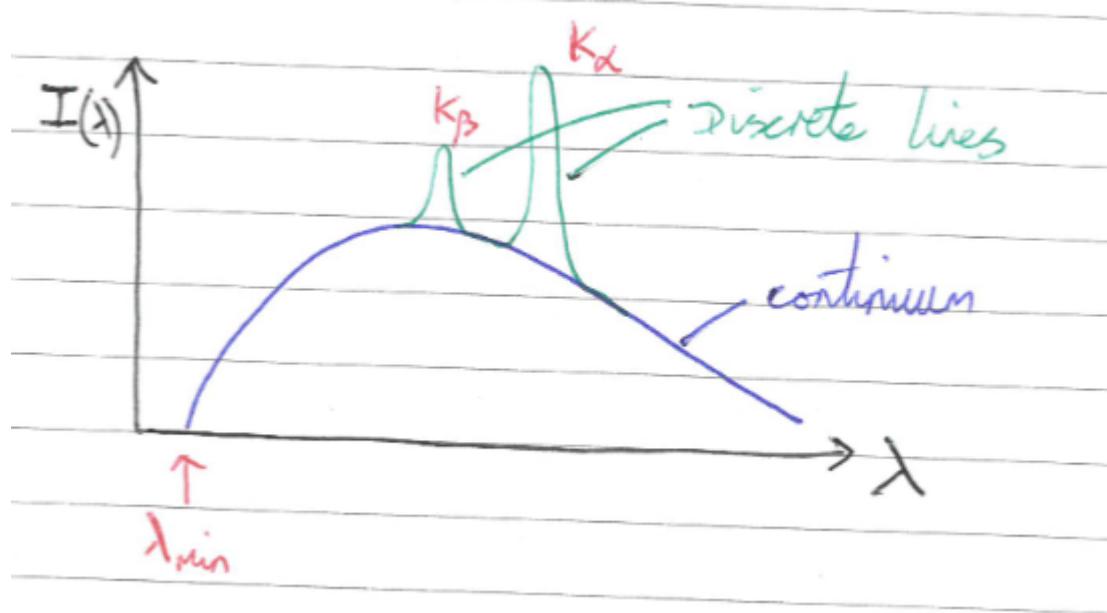


Figure 94.1: A typical x-ray spectra, with a background continuum and unique discrete peaks. Note there is some min λ

The continuous background curve is due to 'Bremsstrahlung', or 'braking radiation'.

2 Braking Radiation

Braking radiation is independent of anode materials. As an electron comes in at a high KE, it interacts with matter (charged nuclei etc) it is deflected and slows down. This lost KE is emitted as X-rays.

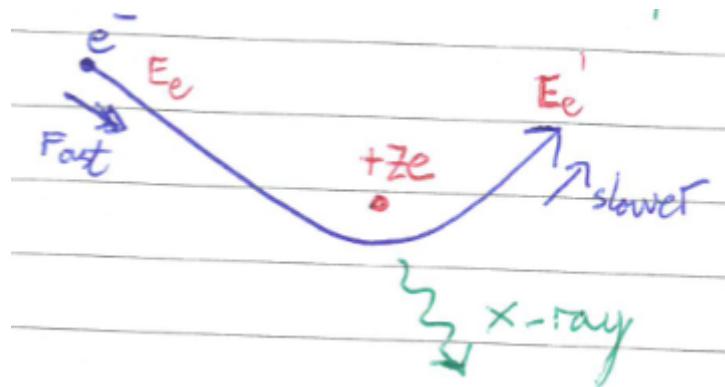


Figure 94.2

2.1 Lambda Min

By C.O.E:

$$E_{\text{x-ray}} = E_e - E'_e = \frac{hc}{\lambda}$$

If the electron came to a stop entirely, E'_e would be minimised (braking radiation maximised). The minimum wavelength λ_{\min} occurs at the highest x-ray energy, i.e. when E'_e is zero.

$$\begin{aligned} E_e - 0 &= \frac{hc}{\lambda_{\min}} = eV \\ \lambda_{\min} &= \frac{hc}{eV} \end{aligned}$$

Where V is the accelerating voltage.

3 Discrete Lines

These arise from inner shell electron transitions. These, unlike Bremsstrahlung, are characteristic of the anode material. We can use this to investigate the anode material, or we can use different materials to build different x-ray tubes with different emission spectra.

4 Jargon

Note the following stupid jargon.

- k means the $n = 1$ energy level.
- l means the $n = 2$ energy level.
- et cetera.

Lets again consider the atomic energy levels (and their associated maximum occupancy):

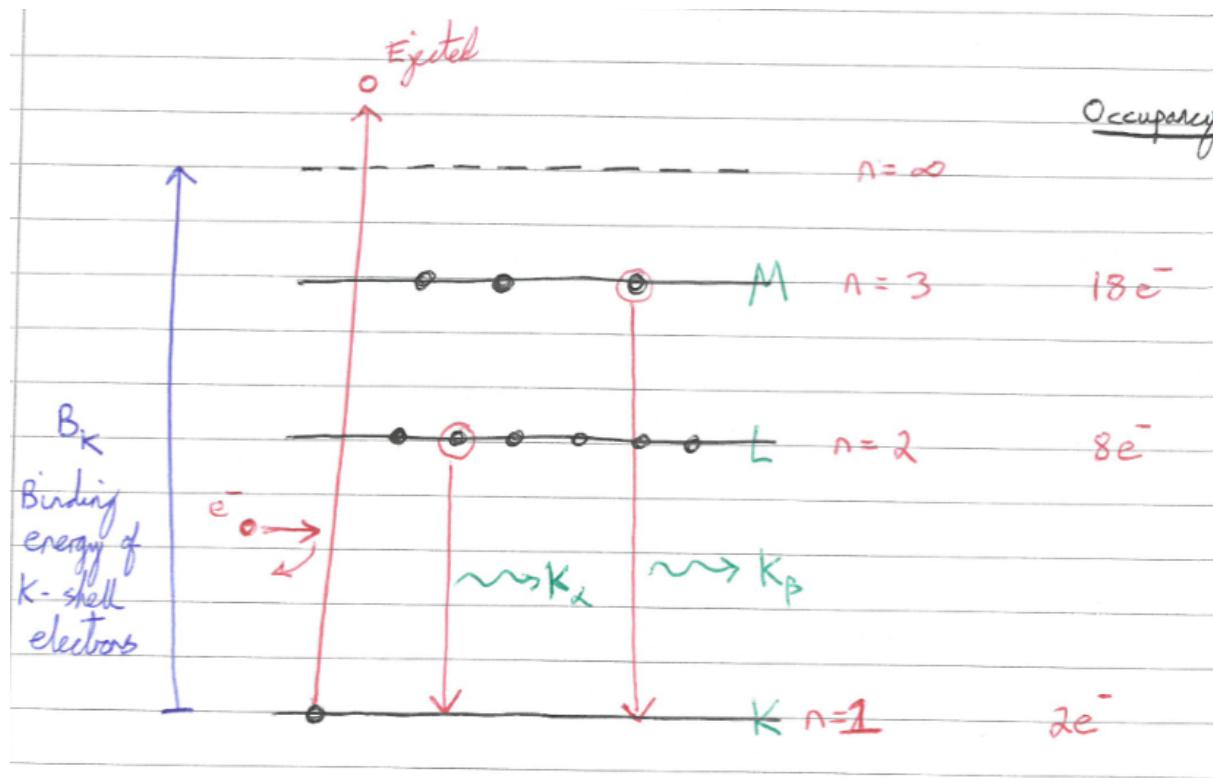


Figure 94.3

If the incoming energy eV is bigger than the binding energy B_k of the k -shell electrons, they can be ejected. This is called 'collisional ionisation'. The intention is to eject an electron from the $n=1$ shell. This leads to an open space in the $n=1$ shell (a 'hole'), and the atom will automatically settle into the lowest energy state, i.e. a higher level electron will drop down to fill the lower energy hole. This change in energy requires the emission of a photon.

This specific $n = 2$ to $n = 1$ state transition is called k_α (the k denoting 'to the 1st level'), while $n = 3$ to $n = 1$ would be k_β etc. Transitions from the third to the second would be L_α , etc etc.

k_α etc only appear if bombardment energy is sufficiently high to cause collisional ionisation. Lower energy lines will therefore appear sooner.

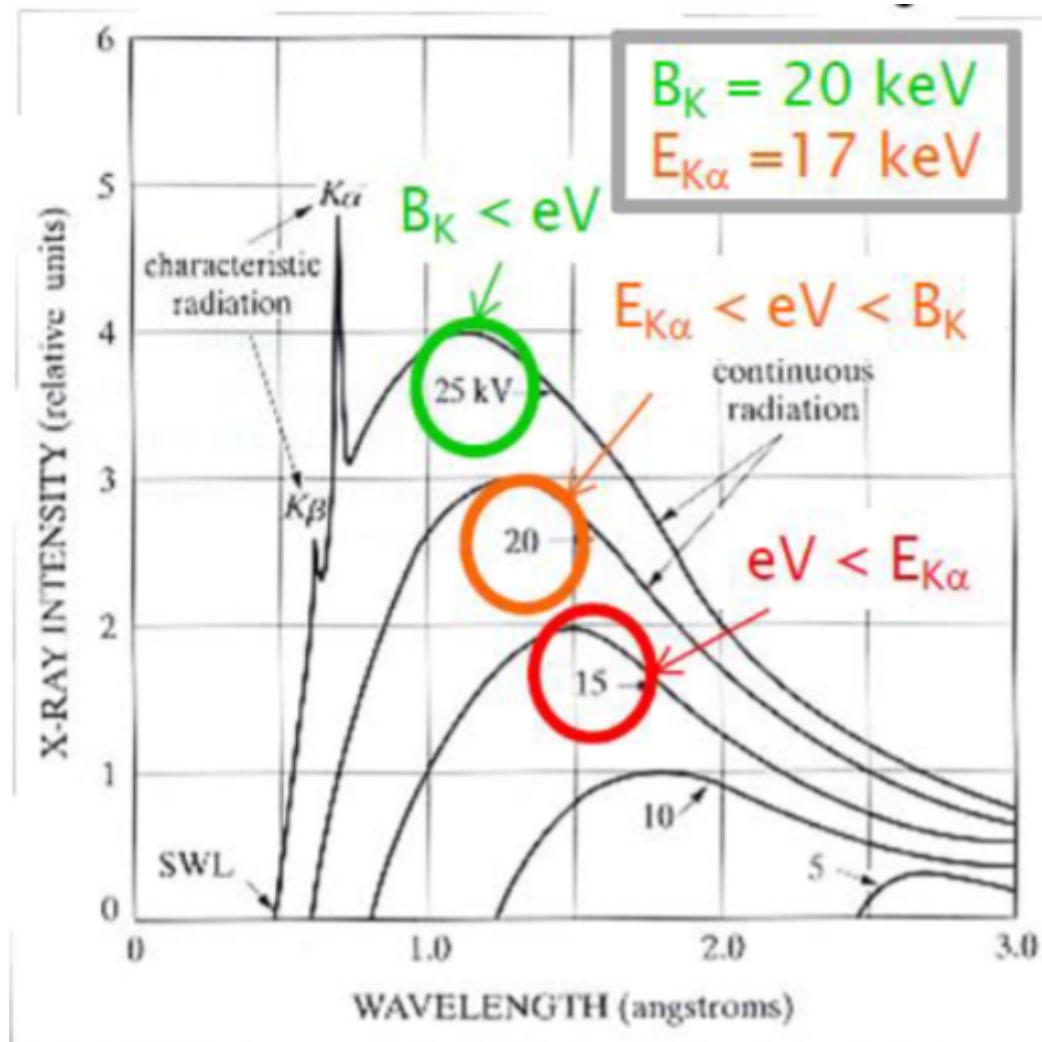


Figure 94.4: Molybdenum X-Ray Spectra (1 Angstrom = 1Å = 0.1nm)

Note that, while Bremsstrahlung is present consistently (albeit with a changing min wavelength where higher accelerating voltage leads to a higher energy x-ray and a lower min wavelength), the sharp spectra lines only appear at higher accelerating voltages where a sufficiently high threshold energy to trigger collisional ionisation is met.

5 X-Ray Absorption

As x-rays pass through a material, x-ray intensity will fall exponentially with distance travelled. Suppose we hit a material with x-rays of intensity I_0 . These x-rays travel x distance units. The final intensity is:

$$I = I_0 e^{-\mu x}$$

Where μ , the attenuation coefficient, depends on the material and incoming energy.

$$\mu \propto \frac{1}{E_{\text{x-ray}}^3}$$

This hearkens back to the photoelectric effect.

Absorption drops as energy of the x-ray falls below the binding energy of a given shell.

6 Absorption Edges

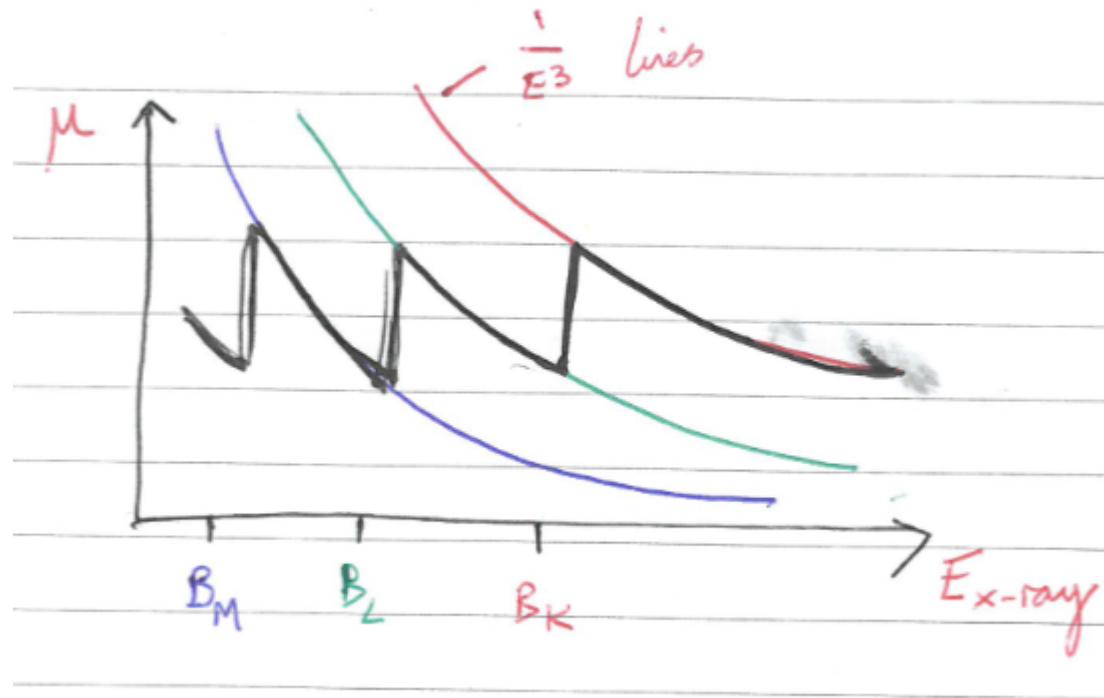


Figure 94.5: Absorption varying with x-ray energy.

Each line represents the $1/E^3$ value for a given energy shell. If we decrease energy and follow the red line along, we eventually hit the point where there is no longer sufficient energy for ejection from the k-shell. We then drop, and can now eject from everything except from the k-shell until we reach the point where we can no longer eject from the k-shell or the l-shell. We drop again, (blue line) and only have emission from all shells except for k and l.

Note:

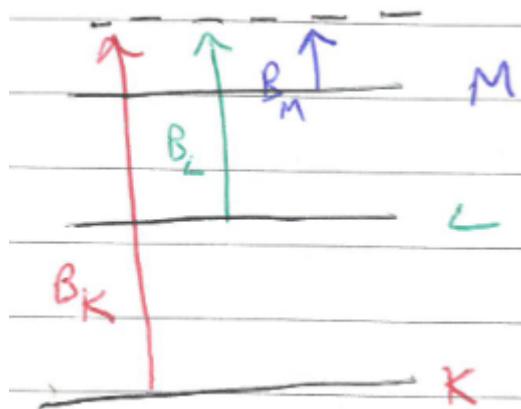


Figure 94.6

- Measuring these ‘absorption edges’ lets us determine what atom something is made of, and the associated electronic energy levels.
- Describing these as hard lines is inaccurate, there is (in reality) much more fine detail.
- This process is called x-ray absorption spectroscopy.

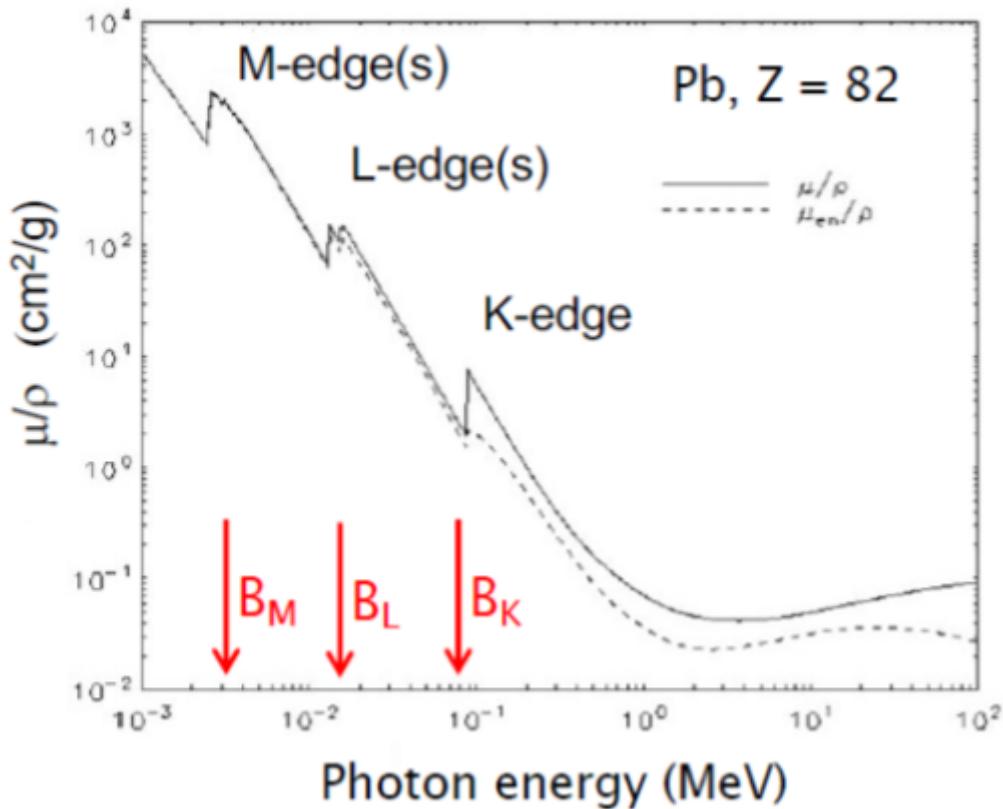


Figure 94.7: Fine detail visible in absorption edges (especially visible on L-edge).

7 Moseley's Law

If you measure the frequency of k alpha for some material, it is equal to a constant times $(Z-1)$ squared, where Z is the atomic number. This lets us identify elements:

$$f_{k\alpha} = (2.48 \times 10^{15})(Z - 1)^2$$

Note: the ' $- 1$ ' appears because there will be a second electron 'left behind' in the k shell, which reduces effective nuclear charge by one. This is called screening or shielding.

Fri 14 Nov 2025 12:00

Lecture 7 - "Matter Waves"

In this lecture:

- Particles acting as waves: de Broglie wavelength.
- Proof: Davisson and Germer experiment.

1 de Broglie Wavelength

We've seen light waves behaving like particles - photons have an energy and momentum related to their wave frequency and wavelength. We can see however that this goes both ways - particles can also behave as waves.

De Broglie suggested that matter/particles should have wave-like properties, and indeed everything behaves as matter, and as waves, and is simultaneously both and neither.

He proposed the De Broglie wavelength:

$$\lambda = \frac{h}{p} = \frac{h}{mv} \quad (95.1)$$

A bigger momentum mv means a smaller wavelength. Taking Paul Hollywood as an example, if he has a 250kg mass and can run at 43mph ($=20\text{m/s}$):

$$p = 5,000 \text{ kg ms}^{-1}$$

$$\lambda = \frac{6.6 \times 10^{-34}}{5000} = 10^{-37}$$

Which is so small it's practically irrelevant. Macroscopic objects have such small wavelengths we can effectively ignore them, hence why classical mechanics still holds on a larger scale.

1.1 For an Electron

For an electron accelerated by 54V, we have a KE of $54eV$. First we check that this is non-relativistic:

$$m_e c^2 = (9 \times 10^{-31})(3 \times 10^8)^2 = 8 \times 10^{-14} \text{ J} = 500,000 eV$$

Therefore KE is much less than the relativistic mass-energy, so classical mechanics are fine. Now:

$$E = \frac{1}{2}mv^2 = \frac{p^2}{2m} \implies p = \sqrt{2mE}$$

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{2mE}} = \frac{h}{\sqrt{2meV}}$$

$$\lambda = 1.67 \times 10^{-10} \text{ m} = 1.67 \text{ \AA}$$

This is no longer small enough that we can get away with totally ignoring it.

2 Experimental Verification - Davisson and Germer

They fired electrons into the surface of metal (with an accelerating voltage of 54V). They measured electron intensity in a scattered beam vs angle. They predicted that electrons should be diffracted, as if they were waves. They observed a constructive interference pattern, equivalent to x-rays.

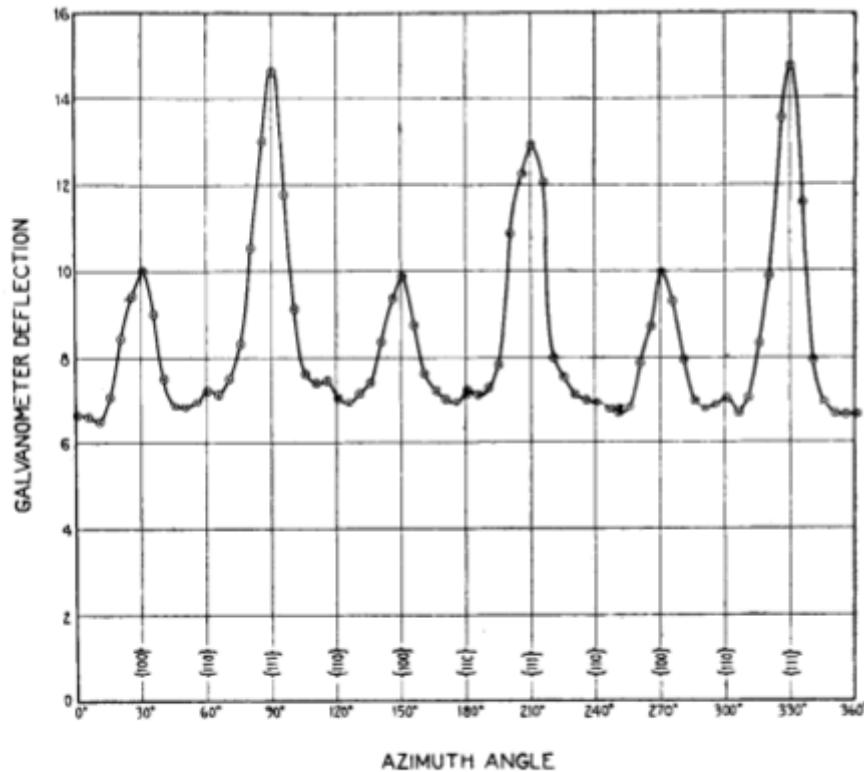


FIG. 2.—Intensity of electron scattering vs. azimuth angle—54 volts, co-latitude 50°.

Figure 95.1

3 Their Result

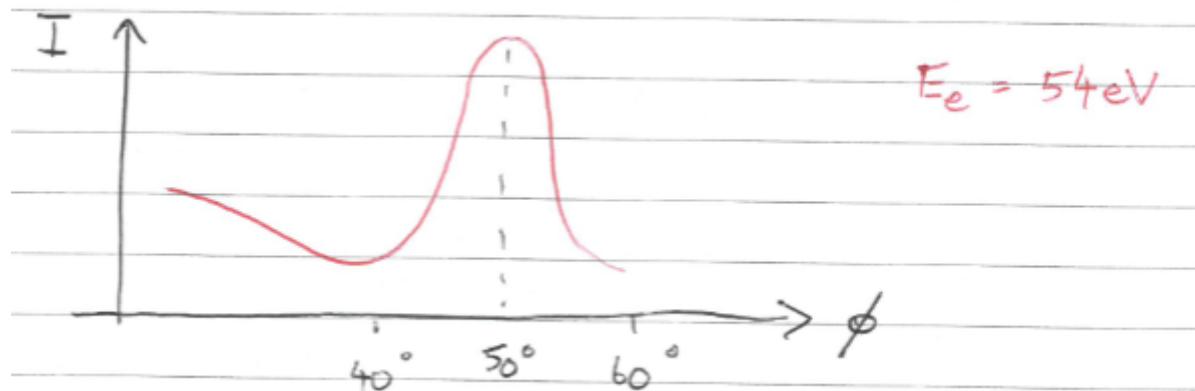


Figure 95.2

We interpret the peak at $\phi = 50$ as due to electron diffraction. Note that the 'scattering angle', ϕ measured is not the same as the Bragg angle θ .

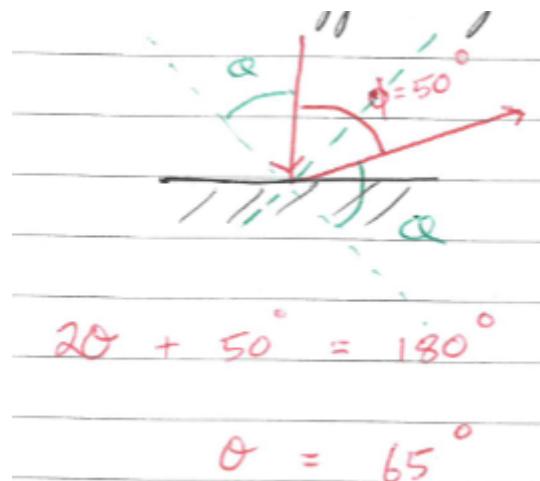


Figure 95.3

3.1 Explaining the Result

Given the Bragg condition $2d \sin \theta = n\lambda$, and $d = 0.91\text{\AA}$, we take $n = 1$ for the central peak, and substitute to get $\lambda = 1.65 \times 10^{-10}\text{m}$. This is (approximately) what we predicted earlier. Note that this is now the Bragg angle θ , not ϕ .

3.2 GP Thomson

At the same time, G.P. Thomson conducted 'powder diffraction'. Grinding a crystal into a powder creates a mess of smaller sub-crystals. This replaces the need for scanning and matching in/out angles, as there will be some correct orientation crystal for any input angle. Since the many small crystallites will be arranged at many angles, we can effectively consider getting all possible angles at once. This means that there will be some crystallites which always satisfy the Bragg conditions, if λ makes this possible.

There is some rotational symmetry, so we get rings produced, and if we use electrons and x-rays on the same target, we can see we get the same pattern.

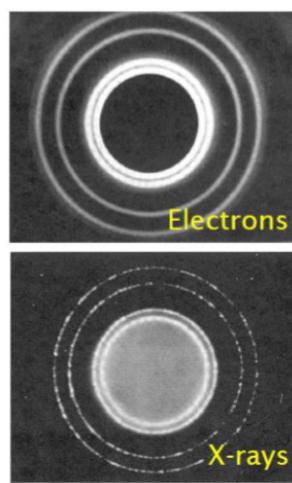


Figure 95.4

This is a little bit difficult to think about, as what does the 'amplitude' of a matter wave actually mean? This is easy to define with EM waves, in terms of the EM field, but not for matter...

4 Back to Bohr

Looking back to the Bohr model, we want to get:

$$E_n = \frac{-13.6eV}{n^2} \quad (95.2)$$

As a starting point, the electron must 'fit' into a circular orbit. Electrons are in a potential well and are bound to a positively charged nucleus:

$$V(r) = \frac{-e}{4\pi\epsilon_0 r}$$

$$u(r) = \frac{-e^2}{4\pi\epsilon_0 r}$$

The electron is in the n th orbital, with a radius r_n , velocity v_n . Plotting wave function amplitude against circumference:

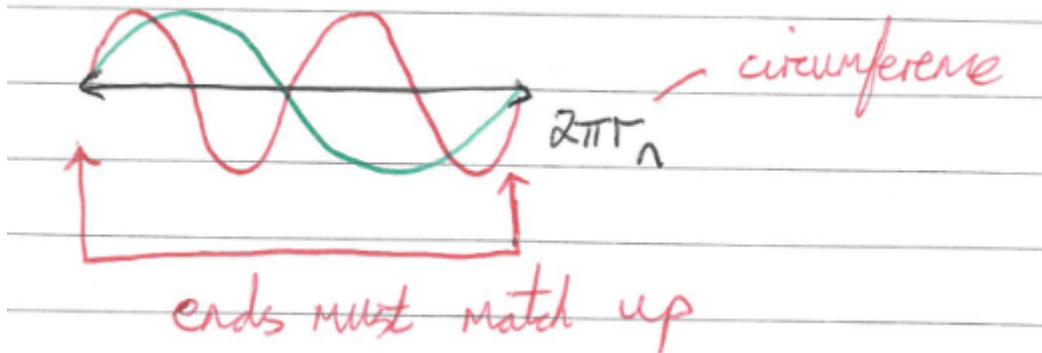


Figure 95.5

We note that this must be continuous, so the ends must join up:

$$2\pi r_n = n\lambda_n = n\frac{h}{p_n} = n\frac{h}{mv_n} \quad (95.3)$$

Also (this is non-relativistic so we can use circular motion)

$$F = ma = \frac{mv^2}{r} \quad (95.4)$$

$$\frac{e^2}{4\pi\epsilon_0 r_n} = mv_n^2 \quad (95.5)$$

We can use 95.3 and 95.5 to eliminate v_n and show:

$$r_n = n^2 a_0, \quad a_0 = \frac{\epsilon_0 h^2}{\pi m e^2} = 5.2918 \times 10^{-11} \text{ ("Bohr Radius")}$$

4.1 Energies

Using 95.3:

$$2\pi r_n = n\lambda_n = n\frac{h}{p_n} = n\frac{h}{mv_n}$$

$$p_n = \frac{nh}{2\pi r_n}$$

And using our new result:

$$p_n = \frac{nh}{2\pi r_n} = \frac{h}{2\pi n a_0}$$

Kinetic Energy

$$T_n = \frac{p_n^2}{2m} = \frac{h^2}{8\pi^2 mn^2 a_0^2} = \frac{me^4}{8\epsilon_0^2 h^2 n^2} = +\frac{13.6}{n^2}$$

Potential Energy

$$U_n = \frac{-e^2}{4\pi\epsilon_0 r_n} = \frac{-e^2}{4\pi\epsilon_0 n^2 a_0} = \frac{-me^4}{4\pi\epsilon_0^2 h^2 n^2} = -\frac{27.2}{n^2} eV$$

Total Energy

$$E_n = T_n + U_n = \frac{-13.6}{n^2} eV$$

This is the Bohr Model result, as required.

5 Conclusions

- Matter is wave-like, with a wavelength given by de Broglie wavelength.
- Electrons can therefore interfere, which we see with electron diffraction.
- A wave-like electron gives us the observed Bohr model result.

Fri 21 Nov 2025 12:00

Lecture 8 - Wave-Particle Duality

In this lecture:

- Particles acting as waves and vice-versa - depends on how we look at/measure them.
- Infinite potential wells - quantisation of energy in bound states, zero-point energy.
- The Heisenberg Uncertainty Principle.

1 Single Slit (Fraunhofer) Diffraction

We have seen particles behaving like waves (i.e. electron diffraction) and waves behaving like particles (i.e. the photoelectric effect). This seems to depend on how we look at them - different experiments give different results.

Broadly, they seem to be wave-y when in motion (i.e. interference), and particulate when detected. In a single-slit experiment, we can see both at the same time. We fire electrons through a single slit, and measure the diffraction pattern with a screen or movable detector:

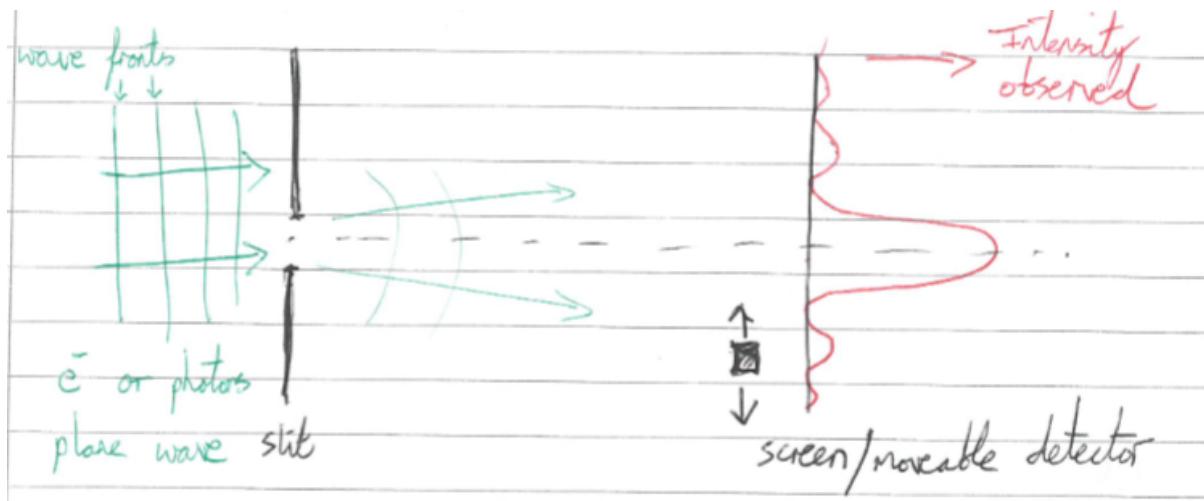


Figure 96.1

We can do this experiment with either photons, or with electrons and we observe the same result: both yield the same interference pattern due to superposition of waves cancelling each other out.

What if we turned the source way down and fired a single photon/electron at a time. We would therefore detect individual particles on our detector as they arrive one-by-one. Surely, therefore, we wouldn't get a diffraction pattern? We do, however, get a pattern... What are they interfering then, if there's no other particles to interfere with? It turns out the particle interferes with itself.

We view the intensity profile as a probability distribution of any one particle impacting here. The areas of the profile with a higher probability have a higher measured intensity, to be expanded on next lecture. The amplitude of a particle wave at a point represents the probability of the particle being found at that point.

2 Infinite Potential Wells

Given a particle in an infinite potential well, how is this impacted by wave-like behaviour of the particle?

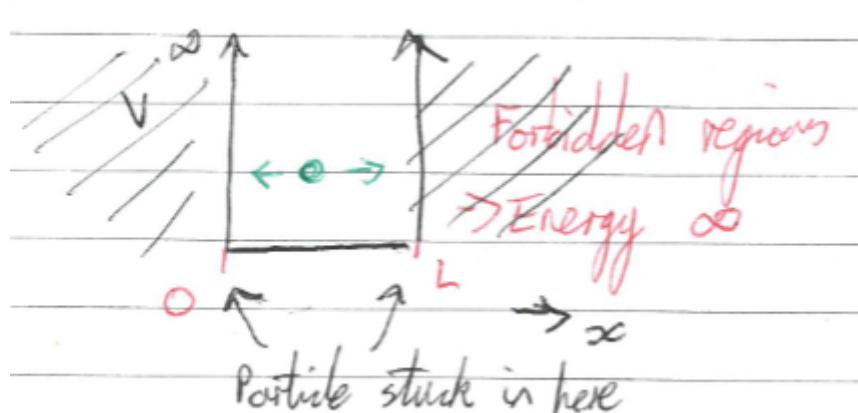


Figure 96.2

Classically, we can think of this as a squash court with a particle bouncing back and forth between the two left and right boundaries. The particle is stuck between zero and L , with potential energy measured on the y-axis. Note that since the y-axis is energy - it is not spatial. The only possible spatial movement of the particle is left-to-right.

Quantum mechanically, we have the following rules:

- The particle has finite energy. It therefore only has amplitude (probability of finding the particle) at this one spot.
- The wave function of a particle has to be continuous throughout. Discontinuities break things.

As a result of these two rules, the wavefunction $\psi(x)$ must be zero at the boundary points 0 and L . We also assume that we can define any wavefunction as a sine wave:

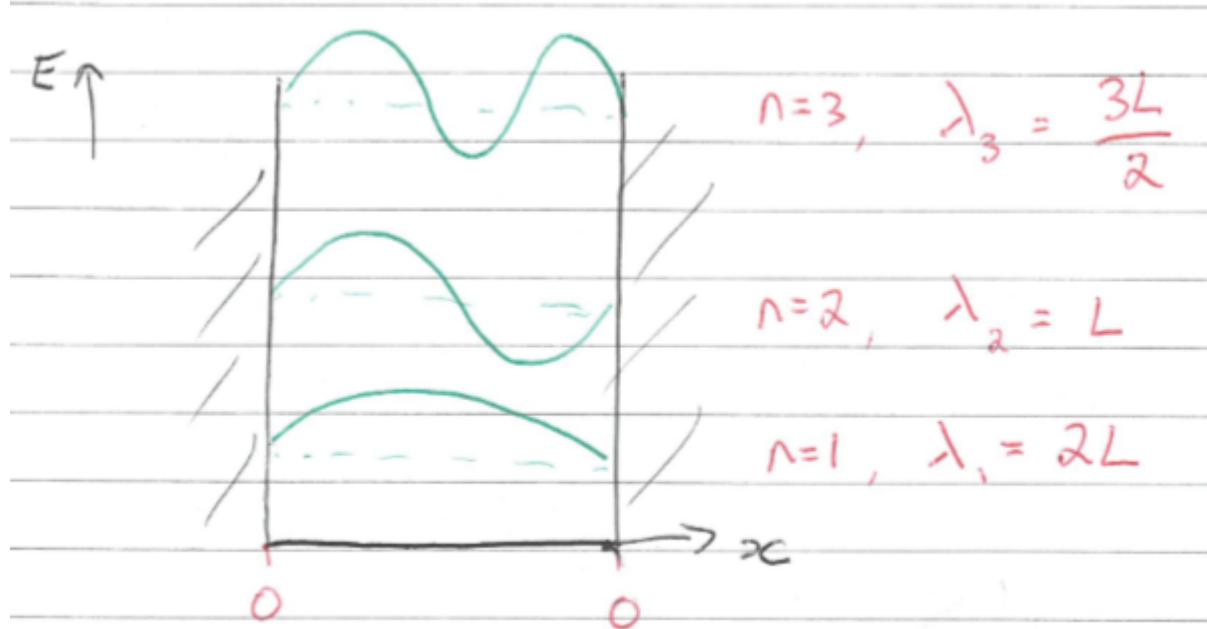
$$\psi(x) = A \sin \frac{n\pi x}{L}, \quad n = 1, 2, 3, \dots$$

Therefore:

$$\lambda_n = \frac{2L}{n}$$

We can generate more complicated shapes using Fourier Decomposition, where we define a more complex wave function as the sum of (potentially infinitely) many sine waves.

Considering this wave function at multiple different energies (infinitely many as we're in an infinite well):

Figure 96.3: Note: Bigger n is higher energy.

We know from last lecture that we can relate wavelength and momentum, so:

$$p = \frac{h}{\lambda} \implies p_n = \frac{n\hbar}{2L}$$

$$E = \frac{p^2}{2m} = \frac{n^2\hbar^2}{8mL^2}$$

Note that $E \propto n^2$ in an infinitely deep 1D potential well. Compare this to $E \propto \frac{1}{n^2}$ in a H atom.

Energy is uniquely determined for each quantum state $\psi_n(x)$ for the particle in the well. If we consider momentum: classically we say the particle is moving either to the left or or the right, with a known magnitude but an unknown direction (if we take a single observation as a snapshot in time).

For $n=1$:

$$p_x = \pm \frac{\hbar}{2L}$$

Which is effectively zero momentum but with an uncertainty of $\Delta p_x = \frac{\hbar}{2L}$. The expectation value (average value):

$$\langle p_x \rangle = 0$$

And for x (somewhere in the well):

$$\langle x \rangle = \frac{1}{2}L$$

With an uncertainty of:

$$\Delta x = \pm \frac{1}{2}L$$

3 Shrinking the Well

What happens if we shrink the well? Both momentum and energy are inversely proportional to cavity length:

$$p \propto \frac{1}{L}$$

$$E \propto \frac{1}{L^2}$$

Therefore they both go up. This is weird in classical mechanics. We now know less about momentum (higher uncertainty) but more about the particle's location (smaller set of possible values, lower uncertainty than before). This isn't a proof, but a hint towards the Heisenberg Uncertainty Principle.

4 Heisenberg Uncertainty Principle

$$\Delta x \Delta p_x \geq \frac{\hbar}{2} \quad (96.1)$$

Where: $\hbar = \frac{h}{2\pi}$. In QM, there is a limit to how well we can know two conjugate (effectively paired, formal word to mean a pair of variables where knowing one of them perfectly precludes knowing the other) observables simultaneously. This also applies to energy w.r.t. time:

$$\Delta E \Delta t \geq \frac{\hbar}{2}$$

Note that in a 2D system, $\Delta p_x \Delta y$ is unrestricted as they are independent and non-conjugate observables. In our example:

$$\Delta x \Delta p_x = \frac{L}{2} \frac{h}{2L} = \frac{h}{4} \geq \frac{h}{4\pi}$$

The single slit diffraction is a good example of this. If we measure the x of a photon by passing it through a slit, we lose information on its momentum and it spreads out along x .

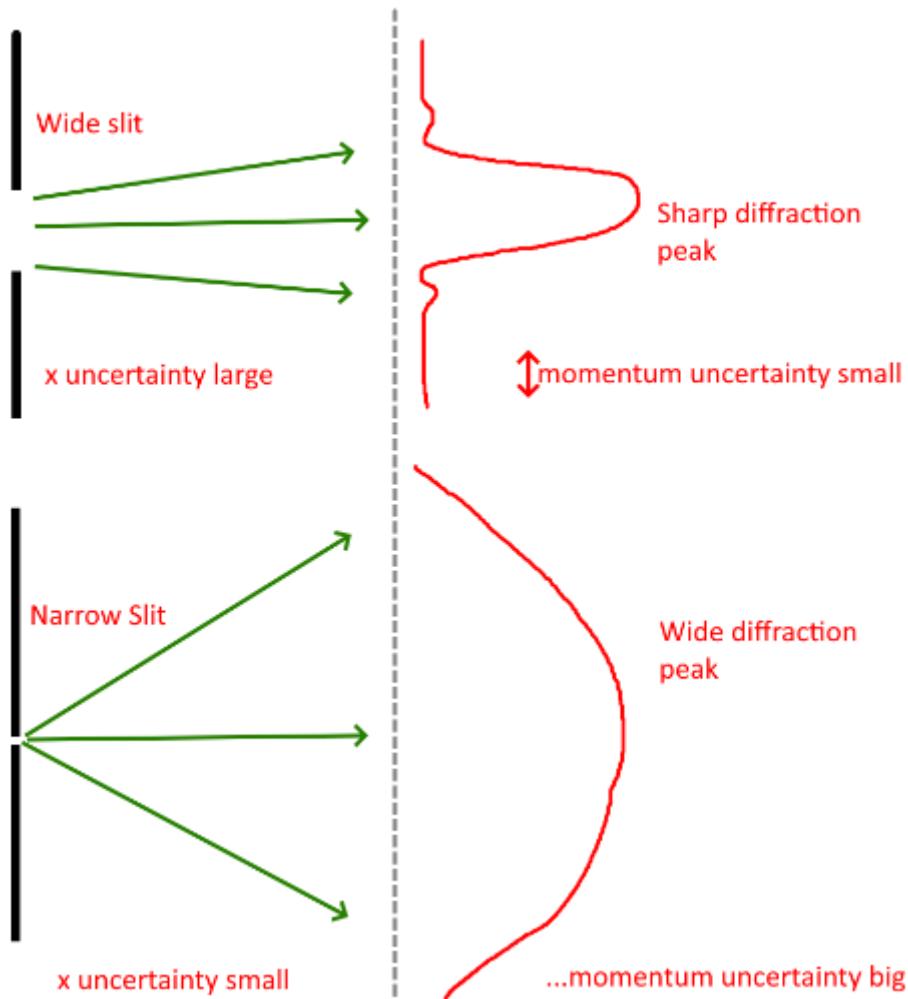


Figure 96.4

5 Conclusions

- Wave-particle duality allows particles to both diffract as if waves and to be quantised like particles.
- Wave-like particles explains why the energy in a bound system has to be quantised (as shown before) due to standing waves in a potential well.
- Energy levels depend on the shape of the potential well.
 - $E \propto n^2$ in an infinitely deep 1D well.
 - $\frac{1}{n^2}$ in a H atom.
- We cannot know a particle momentum and position simultaneously to perfect precision, we are limited that if we know one very very accurately, we cannot know the other very well.

Fri 28 Nov 2025 12:00

Lecture 9 - Wavefunctions for Quantum Particles

In this lecture:

- Reminder and recap of the classical wave equation, building to the complex quantum mechanical version.
- The idea of probability density and probability amplitude.
- The idea of standing waves and the “particle in a box” being modelled as one.

1 Recap of Classical Waves

The classical wave equation is:

$$\frac{\partial^2 E}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 E}{\partial t^2} \quad (97.1)$$

Solutions to this have the form:

$$E(x, t) = E \sin(kx - \omega t)$$

$$E(x, t) = E \cos(kx - \omega t)$$

As the wave equation is linear, linear sums of individual solutions form a solution. I.e. the summation of two different wave functions in this form is also a wave function. Recall that $k = 2\pi/\lambda$ “the wave number” and $\omega = 2\pi f$ “angular frequency”.

1.1 Direction

Wavefunctions with $kx - \omega t$ move to the right (in the positive x direction).

Wavefunctions with $kx + \omega t$ move to the left (in the negative x direction).

1.2 Phase

Consider some point on a wave. Each point has a phase ϕ , given by $\sin \phi = \sin kx - \omega t$. The phase of a point is constant, regardless of propagation.

$$\phi = kx - \omega t \quad (97.2)$$

For a constant point, i.e with phase $\pi/2$, the phase must be constant. Hence, as time increases, x must also increase, and hence the wave must be moving rightwards along the positive x-axis, as required.

To verify this, we can differentiate wrt time, considering a point where phi does not change, so the derivative is zero:

$$\frac{\partial \phi}{\partial t} = k \frac{\partial x}{\partial t} - \omega \frac{\partial t}{\partial t} = 0$$

$$k \frac{\partial x}{\partial t} = \omega$$

$$\frac{\partial x}{\partial t} = \frac{\omega}{k} > 0 \quad \text{so travelling right.}$$

This also happens to prove the wave speed equation again.

1.3 Things Get Complex

It is important to note that we are often dealing with complex representations, that we'll need when things become quantum-y. We can also write a solution to the wave equation as:

$$E(x, t) = \Re \{ E e^{i(kx - \omega t)} \}$$

This doesn't change anything, it's just a convenient repacking of the trigonometric terms into a slightly nicer form - it's just a maths trick. Even if not stated, there is always an implicit "take the real part" at the end, to get the actual cos() etc terms and real numbers we care about.

2 Quantum Mechanical Wave Function

In QM, the biggest change is that the particle wavefunction **truly is complex**. We do not take the real part at the end, and represents something which is fundamentally a complex number. Therefore, the wave function is:

$$\psi(x, t) = A e^{i(kx - \omega t)}$$

Again, note that both $\psi(x, t)$ and A are complex numbers. We can rewrite A as $A = A' e^i$, where A' is some real number.

We can now split up the terms:

$$\psi = A e^{ikx} e^{-i\omega t}$$

Where the first term is the spatial function $\psi(x)$ and the second term is the time dependence (or phase) $\phi(t)$. This is the wavefunction for a particle moving in the positive x, while something moving in the negative x would be:

$$\psi = A e^{-ikx} e^{-i\omega t}$$

Note that this is just conventions, and direction could instead be bundled into the A term, but we ignore this here.

3 What Does This Really Mean?

The amplitude of the quantum wavefunction for a particle represents the probability *amplitude*. This is a complex number, so has a magnitude and a phase. This phase is what gives us interference.

We physically cannot measure the phase of a particle such as an electron or a photon - it is not an observable. We can compare the phase difference between two particles, and one phase relative to another, but we cannot measure absolute phase in QM.

3.1 Observables

Something which we physically can measure is called an observable. They must be real numbers only. Something which is not an observable physically cannot be measured, regardless of how good the equipment is etc. In QM, this means we take complex conjugates to get to an observable.

Recall:

$$z = a + i b \quad z^* = a - i b$$

Then:

$$z * z = a^2 + b^2 = |z|^2$$

Which is real. This gives us *probability density*, which is something we can actually measure.

$$P(x) = \psi^*(x) \psi(x) = |\psi|^2$$

This gives us the probability density function for finding this particle at some position x . Note that we have to integrate over a range of x values to get an actual probability from this probability density.

The probability of finding the particle at some exact precise value $x = a$, where a is some real number is precisely zero. We must integrate over a range, and get a probability for this range - as x can take any continuous value, the probability of finding the particle at any one (of infinitely many) is infinitely small. This is also forbidden by the uncertainty principle (Lec 08).

4 Particle in a Box

Standing waves become a key concept for a particle constrained in a box. A standing wave is made up of two parts:

- A wave propagating in one direction: e^{ikx} .
- A reflected wave propagating in the opposite direction: e^{-ikx}

We know that we can add the superposition of two different solutions to the wave equation, to give another solution, so:

$$\psi = A_1 e^{ikx} e^{-i\omega t} + A_2 e^{-ikx} e^{-i\omega t}$$

Since the box is symmetrical, $A_1 = -A_2$ (a positive equality works too, we just get a cosine rather than a sine).

$$\begin{aligned}\psi &= A_1 (e^{ikx} - e^{-ikx}) e^{-i\omega t} \\ &= A_1 ([\cos(kx) + i \sin(kx)] - [\cos(-kx) + i \sin(-kx)]) e^{-i\omega t} \\ &= A_1 ([\cos(kx) + i \sin(kx)] - [\cos(kx) - i \sin(kx)]) e^{-i\omega t} \\ &= A_1 ([i \sin(kx)] - [-i \sin(kx)]) e^{-i\omega t} \\ &= A_1 2i \sin(kx) e^{-i\omega t}\end{aligned}$$

$\psi = A \sin(kx) e^{-i\omega t}$ where: $A = A_1 2i$

And using this to find probability density:

$$\begin{aligned}P(x) &= \psi^* \psi = A^* A \sin(kx) \sin(kx) e^{-i\omega t} e^{i\omega t} \\ &= |A|^2 \sin^2(kx)\end{aligned}$$

In the last lecture, we related k to the length of the box, and the energy state of the particle trapped in it. Using this, and considering the $n = 1$ energy level:

$$P(x) = |A|^2 \sin^2 \frac{\pi x}{L}$$

Note that in many situations, albeit not this one, $P(x)$ may be time dependent. $P(x)$ is also not uniform. Classically, we would expect a uniform probability for the particle's location equal at all points. However, using this QM wavefunction, we have a much greater probability of finding the particle at the centre of the box than the outer regions.

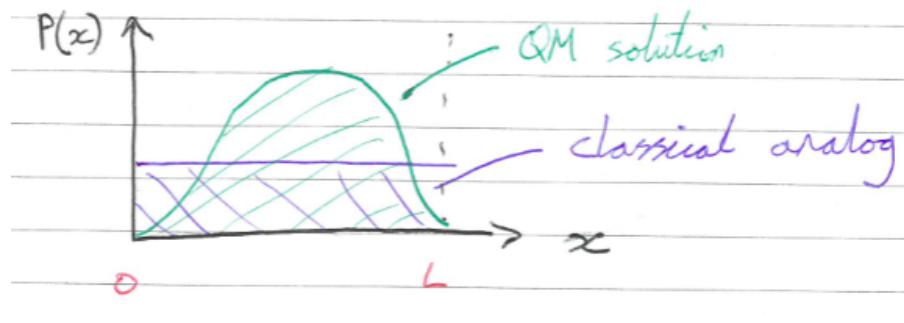


Figure 97.1

4.1 Normalisation

Previously, we've mostly ignored the value of A . However, in order to get numerical values out, we need to find a value for $|A|^2$. Since $P(x)$ is a probability distribution, the integral of the p.d.f. from negative to positive infinity must be one (i.e. the particle is trapped in the box, and must be *somewhere*).

$$\int_{-\infty}^{\infty} P(x) dx = \int_0^L |A|^2 \sin^2 \frac{\pi x}{L} dx = 1$$

We can change the bounds as we know the particle must be in the box somewhere, so the probability of finding it outside of the box is zero.

$$= |A|^2 \int_0^L \frac{1 - \cos\left(\frac{2\pi x}{L}\right)}{2} dx$$

$$= \frac{|A|^2}{2} \int_0^L \left(1 - \cos\left(\frac{2\pi x}{L}\right)\right) dx$$

$$\frac{|A|^2}{2} \left[x - \frac{L}{2\pi} \sin\left(\frac{2\pi x}{L}\right)\right]_0^L = 1$$

$$\frac{|A|^2}{2} (L - 0) = 1$$

$$|A|^2 \frac{L}{2} = 1$$

$$|A|^2 = \frac{2}{L}$$

Therefore:

$$P(x) = \frac{2}{L} \sin^2 kx \quad \text{for } n = 1$$

Note that this value of the prefactor A is true only for this geometry of problem.

4.2 Using This!

Now we have no unknown prefactors, we can find actual probabilities. For example, the probability of the particle being in the right half of the box is given by:

$$\begin{aligned} \int_{L/2}^{\infty} P(x) dx &= \int_{L/2}^L \frac{2}{L} \sin^2 kx dx \\ &= \frac{2}{2L} \int_{L/2}^L 1 - \cos(2kx) dx \\ &= \frac{1}{L} \left[x - \frac{\sin 2kx}{2k} \right]_{L/2}^L \end{aligned}$$

And using $k = 2\pi/\lambda = 2\pi/2L = \pi/L$ (for the $n = 1$) energy level, per Lec 08:

$$\begin{aligned} &= \frac{1}{L} \left[x - \frac{\sin \frac{2\pi x}{L}}{\frac{2\pi}{L}} \right]_{L/2}^L \\ &= \frac{1}{L} \left(\left[L - \frac{\sin \frac{2\pi L}{L}}{2\pi/L} \right] - \left[\frac{L}{2} - \frac{\sin \frac{2\pi L}{2L}}{2\pi/L} \right] \right) \end{aligned}$$

$$\begin{aligned} &= \frac{1}{L} \left(\left[L - \frac{\sin 2\pi}{2\pi/L} \right] - \left[\frac{L}{2} - \frac{\sin \pi}{2\pi/L} \right] \right) \\ &= \frac{1}{L} \left(L - \frac{L}{2} \right) \\ &= \frac{1}{2} \end{aligned}$$

This makes physical sense, as we're considering half the box, and the box (and the p.d.f. for the particle) are symmetrical.

Fri 05 Dec 2025 12:00

Lecture 10 - A Quantum Mechanical Wave Equation

In this lecture:

- The Schrödinger Equation.
- Expectation values of observables.

1 Recap

For a free particle, moving in the +ve x direction, the wavefunction is given by:

$$\psi = Ae^{ikx} e^{-i\omega t}$$

Where the first e-term is the position dependence, and the second is the time dependence. We can separate these two terms.

2 The Schrodinger Equation

Note: This is not derivable from basic physics. It is a postulate, like the Bohr model, which we are confident is true because it's been empirically demonstrated.

We want to build a wave equation for a quantum mechanical free particle in a potential $V(x, t)$. This is a 1D potential with free movement in the x direction only. This wave equation is the Schrodinger Equation.

We assume the following:

- Conservation of Energy applies, hence total energy is kinetic + potential energy. $T + V = E$.
- The equation is a linear differential equation. We need this, otherwise superposition would break, as this relies on two solutions to the wave equation being added together also forming a linear solution.
- This must not break the existing rules we've worked out, so $P = \frac{\hbar}{\lambda} = \hbar k$ and $E = hf = \hbar\omega$

2.1 Constructing the TISE

Now we play a game of “guess the terms” to try to determine values for kinetic, potential and total energy. We can then substitute these into $E = T + V$. This is **not** a formal derivation, but is a motivated construction.

We start with our wave function:

$$\psi = Ae^{ikx} e^{-i\omega t}$$

We note that this has k and ω present, so we have sufficient information to determine momentum and energy. We start by differentiating wrt x :

$$\frac{\partial}{\partial x} \psi(x, t) = ikAe^{ikx} e^{-i\omega t} = ik\psi$$

This lets us extract an ik , but we want a $\hbar k$. Since $(ik)(-\hbar) = \hbar k$, we can multiply by $-\hbar$ to get $\hbar k$:

$$-\hbar \frac{\partial}{\partial x} \psi = (-i\hbar)(ik)(\psi) = (\hbar k)\psi$$

Crucially, this seems to let us extract $\hbar k$, and therefore find momentum. We're going to define the momentum operator in the x direction as:

$$\hat{p}_x = -i\hbar \frac{\partial}{\partial x}$$

We can apply this operator to our wavefunction to determine x-directional momentum. Every observable will have an operator that we can apply to determine it.

We can write this as an eigenvalue equation (where the L.H.S is application of an operator, not multiplication):

$$\hat{p}_x[\psi] = p_x \psi$$

This is defined as doing some operator onto a function, and getting back some multiple of the wavefunction, where that multiple is a useful quantity. This p_x is the x-direction momentum and is a real number - we call this the eigenvalue. Note that, like matrices, operators are not commutative, so $\hat{p}_x \psi \neq \psi \hat{p}_x$.

It is important to note that the fact this works and we can get an eigenvalue tells us that momentum is a well defined quantity in this system, which is not always going to be true.

2.2 Finding Kinetic Energy

Now we have momentum, we can use the following to determine kinetic energy:

$$T = \frac{p^2}{2m}$$

What if we therefore had:

$$\hat{T} = \frac{1}{2m} \hat{p} \hat{p} = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2}$$

Where the two \hat{p} s represent two subsequent applications of the momentum operator.

2.3 Total Energy

For total energy, we want to pull out ω as $E = \hbar\omega$. We therefore get:

$$\hat{E} = i\hbar \frac{\partial}{\partial t}$$

2.4 Potential Energy

Potential Energy V is totally general and can be ugly. We only consider simple constant potentials in QM1. This gives us a neat:

$$\hat{V}\psi = V\psi$$

Where V is a known and constant value for potential.

2.5 Putting It All Together

Now we have these operators, we can substitute them into $T + V = E$ to get the Schrodinger Equation.

$$\boxed{-\frac{\hbar^2}{2m} \frac{\partial^2 \psi(x, t)}{\partial x^2} + V(x, t)\psi(x, t) = i\hbar \frac{\partial \psi(x, t)}{\partial t}}$$

When the potential is independent of time, we can factorise out the t dependence to get to the Time Independent Schrodinger Equation, TISE, (given in formula sheets). We begin by separating the time terms:

$$-\frac{\hbar^2}{2m} \frac{\partial^2 \psi(x)}{\partial x^2} \phi(t) + V(x)\psi(x)\phi(t) = i\hbar \frac{\partial \phi(t)}{\partial t} \psi(x)$$

Dividing both sides by $\phi(t)\psi(x)$:

$$\underbrace{\frac{1}{\psi(x)} \left[-\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + V(x)\psi(x) \right]}_{\text{spatially dependent}} = \underbrace{\frac{i\hbar}{\phi(t)} \frac{d\phi(t)}{dt}}_{\text{time dependent}}$$

Since the L.H.S is purely spatial, and the R.H.S purely time, they are entirely independent. It is only true that these are equal for all points in space and time if these are equal to a constant. We call this constant E .

$$\begin{aligned} \frac{1}{\psi(x)} \left[-\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + V(x)\psi(x) \right] &= E \\ -\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + V(x)\psi(x) &= E\psi(x) \end{aligned}$$

Or we can rewrite this as:

$$\left[-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x) \right] \psi = E\psi$$

The contents of the brackets form an operator. We call this the “Hamiltonian”:

$$\hat{H}[\psi] = E\psi$$

3 Expectation Values

Expectation values represent the average of an observable. It is denoted with angle brackets, for example:

$$\langle p_x \rangle$$

Is the average value for x-momentum if measured many times. In order to get this, we sandwich the operator between ψ^* and ψ and we integrate. For example, the operator for \hat{x} is just $= x$. Therefore:

$$\langle x \rangle = \int_{-\infty}^{\infty} \psi^* x \psi \, dx$$

Going back to our infinite well example:

$$\begin{aligned} \psi_1 &= \sqrt{\frac{2}{L}} \sin kx, \quad k = \pi/L \\ \langle x \rangle &= \frac{2}{L} \int_0^L x \sin^2\left(\frac{\pi}{L}x\right) dx \\ &= \frac{2}{2L} \int_0^L x \left(1 - \cos\left(\frac{2\pi x}{L}\right)\right) dx \end{aligned}$$

Carrying out the integral by parts, we get:

$$\langle x \rangle = \frac{L}{2}$$

Which is physically sensible, we expect the particle to lie, on average, in the middle. Note that this is an average. If we had a particle at $n = 2$, the standing wave created would have a node at $L/2$. This means it would be impossible to find the particle in the middle, but we can still have this as our expectation value.

4 In Conclusion

- Operators pull out observables from the wavefunction.
- We build operators for kinetic (T), potential (V) and total energy (E) and substitute into $T + V = E$ to get the Schrodinger Equation.
- Expectation values give us average values of an operator. This lets us find average values for things without well defined eigenvalues.

Fri 12 Dec 2025 12:00

Lecture 11 - Applications of The Schrodinger Equation

Recap:

- Operators bring out observables from wavefunctions.
- We can build operators for T, E, V and substitute them into energy conservation $T = E + V$ to get the Schrodinger Equation.
- Expectation values let us find ‘average’ values for observables that are not well defined eigenvalues of a wavefunction.

In this lecture:

- Using the Time Independent Schrodinger Equation to solve free-particle at a step problems.
- Reflection and quantum tunnelling.

1 Potential Step Where $E \geq V_0$

Consider a particle of mass m incident on a ‘potential step’ at $x = 0$. Note that the energy of the incoming particle has greater energy than the step potential, so $E > V_0$. This means the particle has sufficient energy to exist in the region, otherwise the particle would be trapped on the left (effectively half a potential well). This particle has enough energy to get past the step with excess K.E.

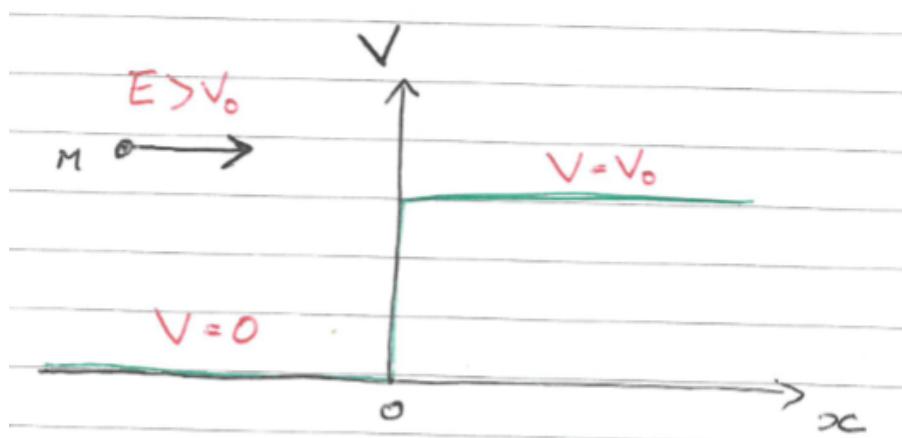


Figure 99.1

In each of the two regions of constant potential, there is no force on the particle:

$$F = -\frac{dV}{dx} = 0$$

The most general form of the T.I.S.E is given by the wave function in the two regions. We have to do this piecewise, as the solution in each area of constant potential will be different:

For $x < 0$, consider ψ_1 :

$$\psi_1 = Ae^{ik_1 x} + Be^{-ik_1 x}$$

Where the A term is the incident wave and the B term is the reflected wave.

For $x \geq 0$, consider ψ_2 :

$$\psi_2 = Ce^{ik_2 x}$$

Note that there is no reflection term, as the region of $x > 0$ has no boundary, hence no reflected wave from the right.

1.1 Verifying Solutions

Lets show that ψ_2 is a solution of the TISE:

$$\begin{aligned} -\frac{\hbar^2}{2m} \frac{d^2\psi_2}{dx^2} + V_0\psi_2 &= E\psi_2 \\ -\frac{\hbar^2}{2m} [(ik_2)^2 Ce^{ik_2 x}] + V_0\psi_2 &= E\psi_2 \\ -\frac{\hbar^2}{2m} (ik_2)^2 \psi_2 + V_0\psi_2 &= E\psi_2 \\ \implies k_2 &= \frac{\sqrt{2m(E - V_0)}}{\hbar} \end{aligned}$$

If this equality is true, then ψ_2 is a solution of the TISE. In other words, ψ_2 is a solution of the TISE for this value of k_2 (only).

Doing the same for ψ_1 gives us:

$$k_1 = \frac{\sqrt{2mE}}{\hbar}$$

So ψ_1 is a solution of the TISE for this value of k_1 .

1.2 Boundary Conditions

We still, however, do not know the values of A, B, C (where $A, B, C \in \mathbb{C}$). We can do this by “matching up” the two wavefunctions by applying boundary conditions for the boundary between the area of zero potential and the potential step (at $x = 0$)

(1) The wavefunction must be continuous. Since $p \propto \frac{\partial \psi}{\partial x}$, momentum will tend to infinity if there is a discontinuity in the wavefunction. Therefore, there cannot be any jump, and:

$$\psi_1(0) = \psi_2(0)$$

(2) The gradient of the wavefunction must be continuous. Since $E \propto \frac{\partial^2 \psi}{\partial x^2}$ the same logic applies. We cannot have a particle of infinite energy, so the gradient of the wavefunction must also be continuous. Note: this does not apply in the infinite potential well case, as we do allow the concept of infinite energy in that model.

$$\left. \frac{\partial \psi_1}{\partial x} \right|_{x=0} = \left. \frac{\partial \psi_2}{\partial x} \right|_{x=0}$$

From (1):

$$Ae^{ik_1 0} + Be^{-ik_1 0} + Ce^{ik_2 0}$$

$$\boxed{A + B = C}, \quad \text{noting: } A, B, C \in \mathbb{C}$$

From (2):

$$ik_1 (Ae^{ik_1 0} - Be^{-ik_1 0}) = ik_2 Ce^{ik_2 0}$$

$$A - B = \frac{k_2}{k_1} C$$

Lets use this to find the probability of a particle reflecting, given by R . From now on we specify $E \geq V_0$ not just $E > V_0$:

$$R = \left| \frac{\text{reflected amplitude}}{\text{incident amplitude}} \right|^2$$

$$R = \left| \frac{B}{A} \right|^2$$

Eliminating C shows that:

$$R = \left| \frac{k_1 - k_2}{k_1 + k_2} \right|^2$$

And we know $k_2 \propto \sqrt{E - V_0}$ and $k_1 \propto \sqrt{E}$. We can check some values:

- When $E = V_0$: $k_2 = 0 \implies R = 1$.
- When $E > V_0$: As $E \rightarrow \infty$, the $\sqrt{E - V_0}$ is dominated by the E term, and we end up with a square root decay curve tending towards zero.

This makes physical sense. We start with a high probability of reflection, which becomes smaller and smaller as we turn up the energy (or decrease the step). However, there is an asymptote at $R = 0$, and the curve is never equal to zero. There is always *some* probability of reflection.

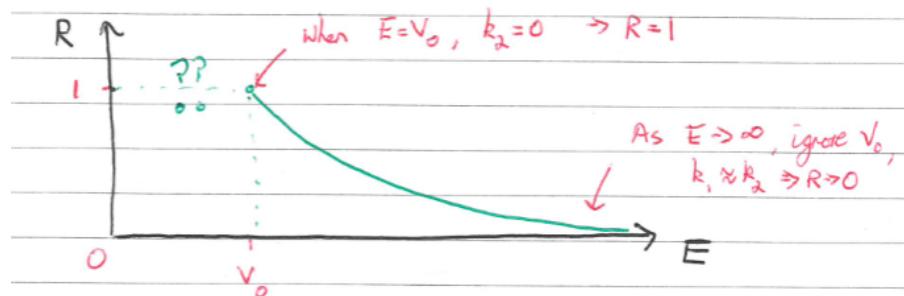


Figure 99.2

2 What if $E < V_0$?

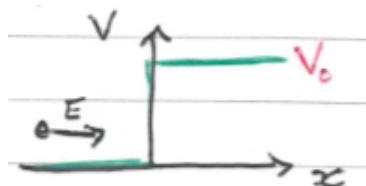


Figure 99.3

Classically, $x > 0$ is completely impossible. The particle must bounce off and cannot exist in this region. Let's approach from a QM perspective and reconsider k_2 :

$$k_2 = \frac{\sqrt{2m(E - V_0)}}{\hbar}$$

The contents of the square root are now negative. We can rewrite this (to simplify a little) as:

$$k_2 = \frac{i\sqrt{2m(V_0 - E)}}{\hbar}$$

This changes our wave function, and so:

$$\psi_2 = C \exp \left(ii \frac{\sqrt{2m(V_0 - E)}}{\hbar} x \right)$$

Letting $\alpha = \sqrt{2m(V_0 - E)/\hbar}$ which is a constant.

$$\psi_2 = Ce^{-\alpha x}$$

This is no longer a wavefunction, and is simply exponential decay. This makes sense, as the particle classically cannot truly exist in the region of too high energy. Sketching probability density against x :

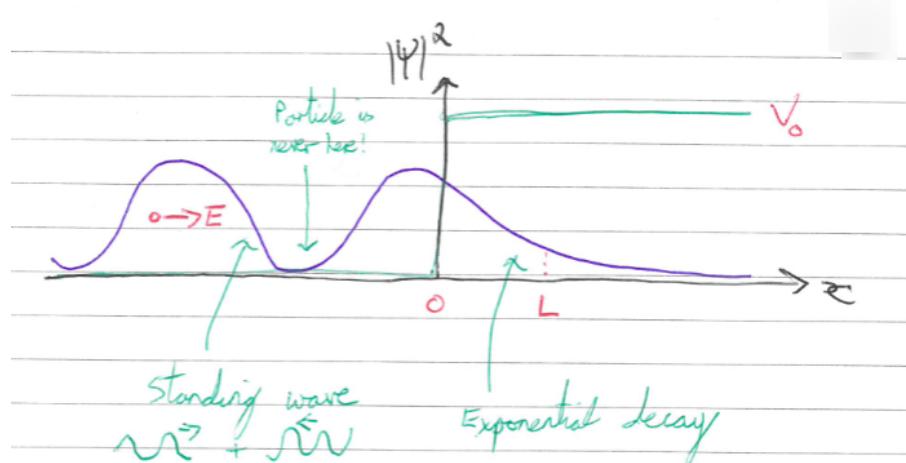


Figure 99.4

The “decay characteristic length” (depth of penetration into the barrier) is given by $1/\alpha$. This means that the particle is allowed to penetrate into the “forbidden zone” of the potential step. What if the forbidden zone ends at some $X = L$. We can see that the potential goes back to zero, and the particle goes back from exponential decay to being a wave. In effect, the particle has passed straight through the wall, despite having insufficient potential so being classically forbidden. This is called “Quantum Tunnelling”.

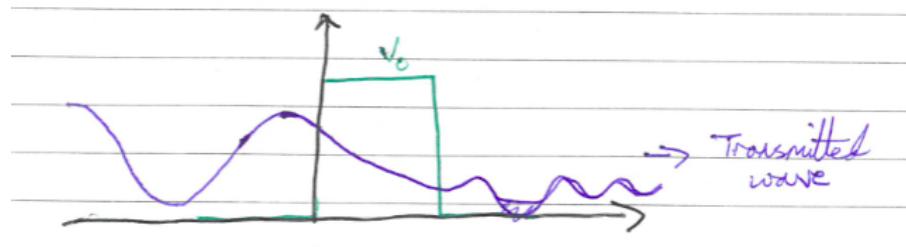


Figure 99.5

3 Conclusions

- For a finite step (or well) substitute in the general solutions to the TISE, piecewise.
- Use boundary conditions to match up the value and the slope of ψ at the boundary.
- There will always be some reflection, and therefore some interference.
- If energy is lower than the step potential, the wavefunction turns into exponential decay.
- If the step is of finite depth, since this exponential decay never reaches zero - there is some probability of the particle passing entirely through the step.

End of Module.