PX286 - Methods of Mathematical Physics

Chung

April 7, 2024

Contents

0.1 Acknowledgements and Tips

These notes were authored by Chung Xu (and perhaps further edited by later PhysSoc exec) and should provide a comprehensive coverage of PX286 Methods of Mathematical Physics. These notes follow closely on content covered both in the 2022-23 and 2023-24 academic years, and is a mix of my own content plus old lecture notes created by Prof. Gareth Alexander and Prof. Rudolf Roemer during pandemic years.

0.1.1 Tips

The exam is only 1 hour long. The majority of the exam is *calculation* - usually working out an integral (e.g. for Fourier transforms), series (Parseval/Bessel's inequalities), numerical calculations (e.g. diffraction) or Lagrange multipliers and indices. Therefore

Learn how to do these *quickly* and *correctly*.

Chapter 1

Fourier Series

1.1 Introduction and Definitions

Definition 1.1.1. A **periodic function** is one which attains the same values after a certain interval. given a function $f(x)$, then $\exists P$ such that $f(x+P) = f(x)$

Functions need not be symmetric (in the sense that reflecting the function through an axis gives the same function) over an interval. For example, $f(x) = \sin\left(\frac{\pi x}{L}\right)$ over $x \in [0, 2\pi]$ is periodic but not symmetric at $x = \pi$.

In contrast, the function $f : [-\pi, \pi] \to \mathbb{R}, f(x) = \cos(x)$ is symmetric about $x = 0$ and periodic.

Definition 1.1.2. Let $f : [-L, L] \to \mathbb{R}$ be a periodic function. It can be expressed as a sum of complex functions so that

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{\frac{in\pi x}{L}} \tag{1.1}$$

where $c_n \in \mathbb{C}$. This is the **complex Fourier series** of a function $f(x)$ on the interval $[-L, L]$.

An alternative expression is derived by substituting the exponential for its complex trigonometric form:

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(k_n x) + \sum_{n=1}^{\infty} b_n \sin(k_n x) \tag{1.2}$$

where $k_n = \frac{n\pi}{L}, a_n, b_n \in \mathbb{R}$. This is called the **Trigonometric Fourier series** of the function

Fourier series consider the study of Fourier approximants, and how they converge.

Definition 1.1.3. Let $f_N : [-L, L] \to \mathbb{R}$ be a periodic function. It can be expressed as a sum of other functions so that

$$f_N(x) = \sum_{-N}^{N} c_n e^{\frac{in\pi x}{L}} \tag{1.3}$$

These f_N are called the **Fourier Approximants**

Definition 1.1.4. The quantities c_n are the **Fourier coefficients**.

Their expression is given as a proposition to be proved below. Finally:

Definition 1.1.5. The nth **Fourier mode** is

$$e^{\frac{in\pi x}{L}}$$

It is this complex exponential, related by Euler's theorem that $e^{ix} = \cos(x) + i\sin(x)$, which produces the sinusoidal behaviour we know and expect.

Theorem 1.1.1. *Dirichlet Conditions. For a function to be represented completely by a Fourier series, we must have*

- *The function must be periodic*
- *It must be single-valued and continuous, but must have a finite number of discontinuities*
- *It must only have a finite number of maxima and minima in one period*
- *The function must be absolutely integrable, i.e.*

$$\int_{-\infty}^{\infty} |f(x)|dx < \infty$$

- *At discontinuities, the sum converges to* $\frac{1}{2}\left[f(x^+) + f(x^-)\right]$

Proof. Not examinable. □

You can intuitively see why these conditions should hold. The function should be periodic, since we are approximating with periodic functions. The second condition is equivalent to requiring piecewise continuity with finite numbers of discontinuities. This is so the series can actually converge to f at each x. For the third condition, if you have an infinite number of extrema in one period, the series may not converge. For example, consider the function

$$f : [-1, 1] \to \mathbb{R} = f(x) = \cos\frac{1}{x}$$

which oscillates infinitely as $x \to 0$ from either side.

The fourth condition can actually be proved

1.2 Fourier Coefficients

Proposition 1.2.1. *The Fourier coefficients for a function $f : [-L, L] \to \mathbb{R}$ are given by*

$$c_n = \frac{1}{2L}\int_{-L}^{L} e^{\frac{-in\pi x}{L}} f(x)dx \tag{1.4}$$

Proof. We integrate the Fourier series over one time period, starting at $t = t_0$ and going to $t = t_0 + P$ where P is the period, as follows:

$$\int f(x)dx = \int \sum_{-\infty}^{\infty} c_n e^{\frac{in\pi x}{L}} dx$$

We know that the function f must be absolutely integrable, so using an integral test, the series must converge absolutely, so the series must converge. Hence, we can use Tonelli's theorem and swap the integral and summation:

$$\int f(x)dx = \sum_{-\infty}^{\infty} \int c_n e^{\frac{in\pi x}{L}} dx$$

This is a bit weird to do without some complex analysis. It is helpful to see how this converts to the series form, find a_0, a_n, b_n. We know

$$\cos(x) = \frac{e^{ix} + e^{-ix}}{2} \qquad \cos(x) = \frac{e^{ix} - e^{-ix}}{2i}$$

Using this, this transform the equation to:

$$\int f(x)dx = \sum_{-\infty}^{\infty} \int c_n e^{\frac{in\pi x}{L}} dx$$

□

The Proposition also establishes the following lemma:

Lemma 1.2.1. *A Fourier mode is defined by $e^{ik\pi x/L}$. Using the definition of orthogonality over $[-L, L]$:*

$$\int \overline{f(x)} g(x) = 0$$

We see then using $f(x) = e^{\frac{im\pi x}{L}}, g(x) = e^{\frac{in\pi x}{L}}$

$$\frac{1}{2L} \int_{-L}^{L} \overline{e^{\frac{im\pi x}{L}}} e^{\frac{in\pi x}{L}} dx \tag{1.5}$$

$$= \frac{1}{2L} \int_{-L}^{L} e^{-\frac{im\pi x}{L}} e^{\frac{in\pi x}{L}} dx \tag{1.6}$$

$$= \frac{1}{2L} \int_{-L}^{L} e^{\frac{i(n-m)\pi x}{L}} dx = 0 \tag{1.7}$$

$$\tag{1.8}$$

whenever $m \neq n$. Trivially, if $m = n$, the integral becomes 1.

This is referred to as the orthogonality of Fourier modes.

1.3 Fourier sine and cosine series, and orthogonality relationships

We can transform between the cosine/sine series form and the exponential form of the Fourier series using Euler's theorem, that $e^{ix} = \cos(x) + i\sin(x)$.

Indeed, the Fourier coefficients c_n transform as follows:

$$c_n = \frac{1}{2L}\int_{-L}^{L}\left(\cos\left(\frac{n\pi x}{L}\right) - i\sin\left(\frac{n\pi x}{L}\right)\right)f(x)dx = \frac{1}{2}(a_n - ib_n)$$

Corollary 1.3.1. *We see $a_{-n} = -a_n$ and $b_{-n} = -b_n$. This allows us to reduce the lower limit of the series from $-\infty$ to 1. If f is real, then $c_{-n} = \overline{c_n}$.*

Using Euler's theorem again, we arrive at

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \sin(k_n x) + \sum_{n=1}^{\infty} b_n \sin(k_n x), k_n = \frac{n\pi x}{L}$$

Corollary 1.3.2. *The orthogonality relationships are as follows:*

$$\frac{1}{L}\int_{-L}^{L}\sin(k_m x)\sin(k_n x)dx = \delta_{mn} \tag{1.9}$$

$$\frac{1}{L}\int_{-L}^{L}\cos(k_m x)\cos(k_n x)dx = \delta_{mn} \tag{1.10}$$

$$\frac{1}{L}\int_{-L}^{L}\sin(k_m x)\cos(k_n x)dx = 0 \tag{1.11}$$

It is helpful to note

- The averages of $\sin^2, \cos^2$ over a period is a half

1.4 Properties of Fourier series

1.4.1 Shift rule

Let f, g be functions with Fourier coefficients c_n, d_n and suppose $g(x) = f(x - x_0)$. Then

$$d_n = e^{-\frac{in\pi x_0}{L}} c_n$$

1.4.2 Derivatives

Derivatives come under quite a lot of conditions. These conditions are an **if and only if** statement:

- $f(x)$ is continuous over $[-L, L]$
- Endpoints match $f(L) = f(-L)$
- $f'(x)$ is piecewise continuous over $[-L, L]$

Then $\forall x \in [-L, L]$:

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(k_n x) + b_n \sin(k_n x) \tag{1.12}$$

$$\implies f'(x) = \sum_{n=1}^{\infty} -k_n a_n \sin(k_n x) + k_n b_n \cos(k_n x) \tag{1.13}$$

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{\frac{in\pi x}{L}} \tag{1.14}$$

$$\implies f'(x) = \sum_{n=-\infty}^{\infty} \frac{in\pi}{L} c_n e^{\frac{in\pi x}{L}} \tag{1.15}$$

1.4.3 Integration

Term-by-term is fine if $f(x)$ is **piecewise continuous over** $[-L, L]$.

$$\int_{-L}^{x} f(x')dx' = \left[\frac{1}{2}a_0 x' + \sum_{n=1}^{\infty} \frac{a_n}{k_n} \sin(k_n x) - \frac{b_n}{k_n} \cos(k_n x)\right]_{-L}^{x} \tag{1.16}$$

$$= \frac{1}{2}a_0(x + L) + \sum_{n=1}^{\infty} \frac{a_n}{k_n} \sin(k_n x) - \frac{b_n}{k_n}(\cos(k_n x) - (-1)^n) \tag{1.17}$$

For the complex Fourier series, this evaluates as, under appropriate convergence tests,

$$\int_{-L}^{x} \sum_{n=-\infty}^{\infty} c_n e^{\frac{in\pi x'}{L}} dx' = \left[\sum_{n=-\infty}^{\infty} c_n \frac{L}{in\pi} e^{\frac{in\pi x'}{L}}\right] \tag{1.18}$$

$$= -\sum_{n=-\infty}^{\infty} c_n \frac{iL}{n\pi} e^{\frac{in\pi(x+L)}{L}} \tag{1.19}$$

1.4.4 Discontinuities

Suppose f is discontinuous at x_0 but the two one-sided limits $\lim_{\delta \to 0} f(x_0 \pm \delta) = f_{\pm}(x_0)$ both exist and the Fourier series converges away from x_0 (i.e, the Fourier series does not converge to x_0). Then at x_0 the Fourier series converges to the **midpoint** of either side of x_0.

For example, consider the function defined by

$$f(x) = \begin{cases} 1 & 0 < x < L \\ -1 & -L < x < 0 \end{cases}$$

Then the Fourier series hits the origin, halfway between 1 and -1 as expected.

1.5 Truncated series

Definition 1.5.1. A truncated Fourier series is one where we iterate over a finite number of times, rather than taking sums over infinities:

$$f_N(x) = \sum_{n=-N}^{N} c_n e^{\frac{in\pi x}{L}}, -L \leq x \leq L \tag{1.20}$$

The goal is now to determine the c_n which would be the best fit.

Definition 1.5.2. The **cost function** is a root-squared deviation defined by

$$\Delta = \int_{-L}^{L} \left[f_N(x) - f(x)\right]^2 dx \tag{1.21}$$

We aim to minimise this function (Calculus of Variations) so we go to find the extrema:

$$\begin{aligned}
\frac{\partial \Delta}{\partial c_m} &= \frac{\partial}{\partial c_n} \int_{-L}^{L} \left[\sum_{n=-N}^{N} c_n e^{\frac{in\pi x}{L}} - f(x) \right]^2 dx \\
&= 2 \int_{-L}^{L} \left[\sum_{n=-N}^{N} c_n e^{\frac{in\pi x}{L}} - f(x) \right] e^{\frac{im\pi x}{L}} dx \\
\implies \int_{-L}^{L} f(x) e^{\frac{im\pi x}{L}} dx &= \int_{-L}^{L} \sum_{n=-N}^{N} c_n e^{\frac{i(n+m)\pi x}{L}} dx = 2Lc_{-m}
\end{aligned}$$

This is as derived, so the Fourier coefficient for an infinite series is just as good as a coefficient for a truncated series.

1.5.1 Gibbs Oscillations

In all Fourier series, when approximating infinite sums by finite values of N, there are deviations from the exact values of $f(x)$ in the vicinity of discontinuities in $f(x)$. This is quite a general problem and stems from our imprecision in defining the convergence of our Fourier series.

There are always undershoots and overshoots.

- As $N \to \infty$, the peak width decrease in the order of $\frac{1}{N}$. This can be visualised as the FS stretching out to match the curve, e.g. FS for a linear function $f(x) = x$, or a square wave.
- As $N \to \infty$, the peak height goes to a constant $\delta \approx 0.09\Delta$
- Gibbs phenomenon at jump discontinuities will remain even as $N \to \infty$.

1.6 Parseval's theorem

Theorem 1.6.1. *(Parseval's theorem). Let f, g be functions defined over $[-L, L]$ and c_n, d_n be their complex Fourier coefficients. Then*

$$\frac{1}{2L} \int_{-L}^{L} \overline{g(x)} f(x) dx = \sum_{n=-\infty}^{\infty} \overline{d}_n c_n \tag{1.22}$$

Proof. This proceeds via direct substitution and doing some algebra.

$$\begin{aligned}\frac{1}{2L}\int_{-L}^{L}\overline{g(x)}f(x)dx &= \frac{1}{2L}\int_{-L}^{L}\sum_{n=-\infty}^{\infty}\overline{d}_n e^{-ik_n x}c_n e^{ik_n x}dx \\ &= \sum_{n=-\infty}^{\infty}\overline{d}_n c_n \frac{1}{2L}\int_{-L}^{L}dx \\ &= \sum_{n=-\infty}^{\infty}\overline{d}_n c_n\end{aligned}$$

The last line is achieved by noting that the integral of 1 over a symmetric interval evaluates to the interval itself, which is $2L$, so they cancel out. □

This leads to some nice corollaries:

Corollary 1.6.1.

$$\frac{1}{2L}\int_{-L}^{L}|f(x)|^2dx = \sum_{n=-\infty}^{\infty}|c_n|^2$$

Proof. This follows directly from realising that the product of $\overline{f(x)}f(x) = |f(x)|^2$, and the Fourier coefficients $c_n^* c_n = |c_n|^2$, and applying Parseval's theorem. □

In trigonometric form, Parseval's theorem is realised as follows:

$$\begin{aligned}f(x)^2 &= \frac{a_0^2}{4} + \sum_{n=1}^{\infty}a_n^3\cos^2(k_n x) + \sum_{n=1}^{\infty}b_n^2\sin^2(k_n x) + \sum_n\sum_m ... \\ \overline{f(x)^2} &= \frac{1}{2L}\int_{-L}^{L}\frac{a_0^2}{4} + \sum_{n=1}^{\infty}a_n^2\overline{\cos^2(k_n x)} + \sum_{n=1}^{\infty}b_n^2\overline{\sin^2(k_n x)} \\ &= \frac{a_0^2}{4} + \frac{1}{2}\sum_{n=1}^{\infty}a_n^2 + b_n^2\end{aligned}$$

where we used the fact that averages of $\sin^2$ and $\cos^2$ is a half.

1.6.1 Riemann-Lebesgue lemma

Lemma 1.6.1. *If $f \in L^1(\mathbb{R}^n)$ be integrable, then its Fourier transform vanishes at infinity*

More precisely, the Fourier coefficients should vanish

Proof. Begin with the substitution $y = x + \frac{L}{n}$:

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

Shifting the interval $[-L + L/n, L + L/n] \to [-L, L]$ keeps this integral the same by periodicity. Taking an average:

$$\frac{c_n + c_n}{2} = \frac{1}{4L}\left|\int_{-L}^{L}e^{-in\pi x/L}(f(x) - f(x-\frac{L}{n}))dx\right| < \int_{-L}^{L}(f(x) - f(x-\frac{L}{n}))dx < \epsilon$$

Since ϵ is arbitrary, then the $|c_n| \to 0$ as we go to infinity (Analysis). □

1.6.2 Orthogonal basis

The orthogonality of Fourier modes can be viewed as exactly analogous to that for the Cartesian basis vectors, namely $\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$, in a finite-dimensional vector space.

The Fourier modes serve as an orthogonal basis for the infinite-dimensional vector space of square-integrable periodic functions, $L^2([-L, L])$ (This is L-squared space, a Hilbert space with the 2-norm). The Fourier coefficients of any function are the components of the vector in the Fourier basis.

1.6.3 Average of a function

Definition 1.6.1.

$$\bar{h(x)} = \langle h(x) \rangle = \frac{1}{2L} \int_{-L}^{L} h(x) dx \tag{1.23}$$

1.7 Fourier in multiple dimensions

We extend everything into vectors over any vector space of dimension n. Let $\mathbf{x} = (x_1, \ldots x_n)$

$$f(\mathbf{x}) = \sum_{\mathbf{k}} c_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}} \qquad\qquad c_{\mathbf{k}} = \frac{1}{(2L)^3} \int_{[-L,L]^3} \mathrm{e}^{-i\mathbf{k}\cdot\mathbf{x}} f(\mathbf{x}) d^3x$$

Chapter 2

Fourier Transforms

Definition 2.0.1. The **wavevector**, denoted $\mathbf{k}_n = \frac{n\pi}{L}$

We want to keep the spacing uniform., and this spacing is $\delta k = \frac{\pi}{L} \implies 1 = \frac{L}{\pi}\delta k$. Then we see that the sum over the n Fourier modes is really just a sum over wavevectors

$$f(x) = \sum_{k_n} \frac{L}{\pi}\delta k c_n e^{ik_n x} = \frac{1}{2\pi}\sum_{k_n} 2c_n L e^{ik_n x}\delta k$$

Obviously, we simply move from an integer index n to a real, but still discrete index k_n. As $\delta k \to 0$, we start obtaining the structure of a Riemann integral.

Looking at the expression for the Fourier coefficients:

$$c_n = \frac{1}{2L}\int_{-L}^{L} e^{-in\pi x/L} \implies \frac{Lc_n}{\pi} = \frac{1}{2\pi}\int_{-L}^{L} e^{0ik_n x}$$

This leads us to the key definition

Definition 2.0.2. The **Fourier transform** (and its inverse) of a function x is defined by

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

It is important to note that the definition above is the **symmetrical** Fourier transform. There are other conventions (see notes, cba). I will denote the forward Fourier transform and inverse Fourier transform of functions $f(x), \tilde{f}(k)$ by $\mathfrak{F}[f(x)], \mathfrak{F}^{-1}[\tilde{f}(x)]$

2.1 Examples, Lorentzians and Gaussians

EXAMPLE 1.

The exponentially decaying function

$$f(x) = e^{-\frac{|x|}{\xi}}$$

has a Fourier transform

$$\begin{aligned}\tilde{f}(k) = \int_{-\infty}^{\infty} e^{-ikx} f(x)dx &= \int_{-\infty}^{0} e^{-ikx} e^{x/\xi} dx + \int_{0}^{\infty} e^{-ikx} e^{-x/\xi} dx \\ &= \frac{e^{-ikx+x/\xi}}{-ik+1/\xi}\Big|_{-\infty}^{0} + \frac{e^{-ikx-x/\xi}}{-ik-1/\xi}\Big|_{0}^{\infty} \\ &= \frac{2\xi^{-1}}{k^2+\xi^{-2}}\end{aligned}$$

This is called a **Lorentzian lineshape**

Lemma 2.1.1. *The Fourier transform of a Gaussian is a Gaussian*

2.2 Properties of Fourier Transforms

Lemma 2.2.1. *The Fourier transformed function* $\tilde{f}(-k) = \overline{\tilde{f}(k)}$

Proof.

$$\tilde{f}(-k) = \int_{-\infty}^{\infty} \overline{e^{-ikx}} f(x)dx = \overline{\int_{-\infty}^{\infty} e^{ikx} f(x)dx} = \overline{\tilde{f}(x)}$$

□

Lemma 2.2.2. *Let* f, g *be functions satisfying the Dirichlet conditions so that the Fourier-transformed functions* $\tilde{f}, \tilde{g}$ *exist. Suppose that* $g(x) = f(x - x_0)$*. Then* $\tilde{g}(k) = e^{-ikx_0}\tilde{f}(k)$

Proof.

$$\tilde{g}(k) = \int_{-\infty}^{\infty} e^{-ikx} g(x)dx = \int_{-\infty}^{\infty} e^{-ikx} f(x - x_0)dx = \int_{-\infty}^{\infty} e^{-ik(y+x_0)} f(y)dy = e^{-ikx_0}\tilde{f}(k)$$

□

Lemma 2.2.3. *If* $\mathfrak{F}(f(t)) = \tilde{f}(k)$ *then* $\mathfrak{F}(f^{(n)}(t) = (ik)^n \tilde{f}(k)$

Proof.

$$\begin{aligned}\frac{d^n f(x)}{dk^n} &= \frac{1}{\sqrt{2\pi}} \frac{d^n}{dk^n} \int_{-\infty}^{\infty} e^{ikx} \tilde{f}(k)dk \\ = \frac{1}{\sqrt{2\pi}}(ik)^n \int_{-\infty}^{\infty} e^{ikx} \tilde{f}(k) &= (ik)^n \tilde{f}(k) \\ \implies \mathfrak{F}(f^{(n)}(t) &= (ik)^n \tilde{f}(k)\end{aligned}$$

□

However there are conditions attached to derivatives which must all be satisfied:

- $f^{(n)}(t)$ must be piecewise continuous
- $\int_{-\infty}^{\infty} f^{(n-1)}(t)dt$ must converge and be finite
- $\lim_{t\to\pm\infty} f^{(m)}(t) = 0 \forall m \in \{1, \dots, n\}$

Lemma 2.2.4. *Suppose* $\mathfrak{F}(f(x)) = \tilde{f}(\omega)$. *Then* $\mathfrak{F}[tf(t)] = i\frac{d\tilde{f}}{d\omega}$

Proof.

$$\begin{aligned}\frac{d\tilde{f}}{d\omega} &= \frac{1}{\sqrt{2\pi}}\frac{d}{d\omega}\int_{-\infty}^{\infty} e^{-i\omega t}f(t)dt \\ &= \frac{1}{\sqrt{2\pi}}(-it)\int_{-\infty}^{\infty} e^{-i\omega t}f(t)dt \\ \implies i\frac{d\tilde{f}}{d\omega} &= \mathfrak{F}[tf(t)]\end{aligned}$$

□

These properties are very useful since they turn hard to compute operations into simple ones, i.e. differentiation becomes multiplication!

2.2.1 Damped oscillations

Consider the function $f(t) = Ae^{-|\gamma|t}\cos(\omega_0 t)$.

Expanding this with complex exponentials we get

$$\begin{aligned}f(t) &= \frac{A}{2}\left(e^{-|\gamma|t}e^{i\omega_0 t} + e^{-|\gamma|t}e^{-i\omega_0 t}\right) \\ \mathfrak{F}[e^{-|\gamma|t}] &= \frac{1}{\sqrt{2\pi}}\frac{2\gamma}{\gamma^2+\omega^2} \\ \implies \mathfrak{F}[f(t)] &= \frac{A}{2}\frac{1}{\sqrt{2\pi}}\left(\frac{2\gamma}{\gamma^2+(\omega-\omega^0)^2} + \frac{2\gamma}{\gamma^2+(\omega+\omega^0)^2}\right)\end{aligned}$$

This means we get 2 Lorentzians at $\omega = \pm\omega_0$.

2.2.2 Application to spectroscopy

In spectroscopy experiments, waves of various frequencies ω and typically give information about the Fourier transform of some signal $f(t)$.

- Spectroscopic lines are often Lorentzians
- Line precision (how close the lines are together) tells us information about the frequencies involved
- Line width tells us information about the damping parameter γ
- However with line width, various dynamical factors can affect it. The environment is often out of our control. Specific atoms/molecules we desire may be in a mixture with other compounds.

- The more lines that appear, the more dynamical factors that are (potentially) affecting the system
- Twice broadening may occur. This is due to the Heisenberg Uncertainty Principle

2.3 Parseval's Theorem and FT identities

2.3.1 Parseval's Theorem

Previously, Parseval's theorem was stated for Fourier series. We now evolve this into the transform.

Theorem 2.3.1. *(Parseval's theorem). The inner product of 2 functions in function space is proportional to the inner product of their Fourier transforms. More mathematically, let f, g be functions satisfying the Dirichlet conditions, so their Fourier transforms $\tilde{f}, \tilde{g}$ exist and are well-defined. Then*

$$\int_{-\infty}^{\infty} \overline{g(x)} f(x) dx = \frac{1}{2\pi} \int_{-\infty}^{[} \infty] \overline{\tilde{g}(k)} \tilde{f}(k) dk$$

Proof. This is direct calculation:

$$\begin{aligned}
\int_{-\infty}^{\infty} \overline{g(x)} f(x) dx &= \int_{-\infty}^{\infty} \left[\overline{\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ikx} \tilde{g}(k) dk} \right] f(x) dx \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \overline{\tilde{g}(k)} \left[\int_{-\infty}^{\infty} e^{-ikx} \tilde{f}(x) dx \right] dk \\
&= \frac{1}{2\pi} \int_{-\infty}^{[} \infty] \overline{\tilde{g}(k)} \tilde{f}(k) dk
\end{aligned}$$

□

Sometimes Parseval's theorem for FT is known as Plancherel's theorem. A more familiar form of this theorem is when $f = g$

$$\int_{-\infty}^{\infty} |f(x)|^2 dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\tilde{f}(k)|^2 dk$$

2.3.2 Riemann-Lebesgue lemma

We work a lot with trigonometric functions and must study their convergence (or lack of) under various operations, including Fourier transforms. In the Fourier transform, we consider $x = 0$:

$$f(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{f}(k) dk$$

By absolute convergence of Riemann integrals (Analysis 3), this will converge if the integral of the absolute value of $\tilde{f}$ does. This is known as the Riemann-Lebesgue lemma. We prove via substituting in $y = x + \frac{\pi}{k}$

Proof.

$$\tilde{f}(k) = \int_{-\infty}^{\infty} e^{-ikx} f(x)dx = \int_{-\infty}^{\infty} e^{-ik(y-\pi/k)} f(y-\pi/k)dy = -\int_{-\infty}^{\infty} e^{-iky} f(y-\pi/k)dy$$

$$\left|\frac{\tilde{f}(k)+\tilde{f}(k)}{2}\right| = \left|\frac{1}{2}\int_{-\infty}^{\infty} e^{-ikx}(f(x)-f(x-\pi/k))dx\right| \leq \frac{1}{2}\int_{-\infty}^{\infty} |f(x)-f(x-\pi/k)|\,dx$$

where the last step is justified by theorem from Analysis 3. □

So then the Fourier transform absolutely converges as $k \to \pm\infty$.

2.4 Convolutions

Definition 2.4.1. The **convolution** between 2 functions f, g is the function $f \star g$ defined by

$$(f \star g)(x) = \int_{-\infty}^{\infty} f(x-y)g(y)dy \tag{2.2}$$

Proposition 2.4.1. *The convolution is symmetric, i.e.* $f \star g = g \star f$

Proof.

$$(f \star g)(x) = \int_{-\infty}^{\infty} f(x-y)g(y)dy = \int_{\infty}^{-\infty} f(t)g(x-t)(-dt) = \int_{-\infty}^{\infty} g(x-t)f(t)dt = g \star f$$

□

Proposition 2.4.2. *(Convolution theorem). The convolution is the inverse Fourier transform of a product*

Proof.

$$\begin{aligned}(f \star g)(x) = \int_{-\infty}^{\infty} f(x-y)g(y)dy &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{ik(x-y)}\tilde{f}(k)g(y)dkdy \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx}\tilde{f}(k)\left(\int_{-\infty}^{\infty} e^{-iky}g(y)dy\right)dk \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx}\tilde{f}(k)\tilde{g}(k)dk = \mathfrak{F}^{-1}[\sqrt{2\pi}\tilde{f}(k)\tilde{g}(k)]\end{aligned}$$

□

Proposition 2.4.3. *The Inverse Convolution theorem: the Fourier transform of a product of two functions* $\mathfrak{F}[f(x)g(x)]$ *is proportional to the convolution*

$$\tilde{f}(kx) \star \tilde{g}(k) = \int_{-\infty}^{\infty} \tilde{f}(k-k')\tilde{g}(k')dk'$$

Proof. There are 2 ways to do this. One is by evaluating $\mathfrak{F}^{-1}[\tilde{f} \star \tilde{g}]$, and the other is by brute force substitution.

Method one: evaluating the inverse

$$\mathfrak{F}^{-1}[\tilde{f} \star \tilde{g}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ikx} \left(\int_{-\infty}^{\infty} \tilde{f}(k-k')\tilde{g}(k')dk' \right) dk$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ikx} \left(\int_{-\infty}^{\infty} \tilde{f}(k-k')\tilde{g}(k')dk' \right) dk = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ik'x} \tilde{g}(k') \left(\int_{-\infty}^{\infty} e^{i(k-k')x} \tilde{f}(k-k')dk \right) dk'$$

$$\int_{-\infty}^{\infty} e^{ik'x} \tilde{g}(k') \left(\int_{-\infty}^{\infty} e^{i(k'')x} \tilde{f}(k'')dk'' \right) dk'$$

$$= \frac{\sqrt{2\pi}}{2\pi} \int_{-\infty}^{\infty} e^{ik'x} \tilde{g}(k') f(x) dk'$$

$$= \sqrt{2\pi} f(x) g(x)$$

$$\implies \tilde{f} \star \tilde{g} = \mathfrak{F}[\sqrt{2\pi} f(x) g(x)]$$

□

Chapter 3

Dirac Delta Function

3.1 Elementary Properties of the Dirac Delta Function

Definition 3.1.1. The Dirac delta function (distribution), denoted $\delta(x)$ is a function localised at $x = 0$, which when integrated with a function, simply evaluates it at that point

Indeed we can see this with the Fourier transform:

$$\begin{aligned}
\tilde{f}(k) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ikx'} f(x') dx' \\
\implies f(x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ikx} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ikx'} f(x') dx' dk \\
&= \int_{-\infty}^{\infty} \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ik(x-x')} dk \right) f(x') dx' \\
\implies f(x) &= \int_{-\infty}^{\infty} \delta(x - x') f(x') dx' \\
\implies \delta(x - x') &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ik(x-x')} dk
\end{aligned}$$

So we can see the integral of a delta function with a function at a point $x = x'$, evaluates the function at x'.

Dirac Delta function

$$\delta(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ik(x)} dk \tag{3.1}$$

$$\delta(x - x') = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ik(x-x')} dk \tag{3.2}$$

Lemma 3.1.1. *$\delta(x)$ is an even function, i.e. $\delta(-x) = \delta(x)$*

Proof.

$$\delta(-(x-x_0)) = \delta(x_0-x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ik(x_0-x)}dk = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ik(x-x_0)}dk$$
$$= -\frac{1}{2\pi}\int_{\infty}^{-\infty} e^{-ik(x-x_0)}dk = \frac{1}{2\pi}\int_{\infty}^{-\infty} e^{ik(x_0-x)}(-dk)$$
$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ik(x-x_0)}dk = \delta(x-x_0)$$

□

The Dirac delta function also operates over finite intervals, namely

$$\int_a^b f(x)\delta(x-x_0)dx = \begin{cases} f(x_0) & x_0 \in (a,b) \\ 0 & otherwise \end{cases} \tag{3.3}$$

Definition 3.1.2. Dirac delta function as a limit.

$$\delta(x) = \lim_{Q\to\infty}\delta_Q(x) = \lim_{Q\to\infty}\int_{-Q}^{Q} e^{ikx}dk$$
$$= \lim_{Q\to\infty}\frac{1}{2\pi}\left[\frac{e^{ikx}}{ix}\right]_{-Q}^{Q} = \frac{1}{\pi x}\left(\frac{e^{iQx}-e^{-iQx}}{2i}\right) = \frac{\sin(Qx)}{\pi x}$$

- As $Q\to\infty, \delta_Q(x)$ becomes narrower and higher (larger amplitude)
- $\delta(x)$ is infinite high and infinitesimally wide.
- The width and height of $\delta(x)$ are inversely proportional to each other. Namely, the height of $\delta_Q(x)$ is Q/π (measured from the x-axis), but its width (measured from the y-axis, so this is more of a half-width) is π/Q
- $\lim_{Q\to\infty}\delta_Q(x-x_0) = \int_{-Q}^{Q}\delta(x-x_0)f(x)dk = \lim_{x\to x_0} f(x) = f(x_0)$
- This means we can say: $\int_{-\infty}^{\infty}\delta(x-x')f(x)dk = f(x)$
- $\int_{-\infty}^{\infty}\delta(x)dx = 1$

3.2 Convolutions

Lemma 3.2.1. *Let f be a function and $g(x) = \delta(x-x_0)$. Then*

$$f\star g(x) = \int_{-\infty}^{\infty} f(x-y)\delta(y-x_0)dy = f(x-x_0) \tag{3.4}$$

Proof. This follows directly from the definition of the delta function. Namely, the delta function is non-zero $\iff y = x_0 \in (-\infty,\infty)$ which means we return $f(x-x_0)$ directly. □

This actually generalises to any sum of delta functions, $g(x) = \sum_i \delta(y-x_i)$

3.3 Shifting

Related to convolutions, we can shift in the integration bounds to get a function out, namely:

$$\int_a^b \delta(x-x_0)f(x) = \int_{a-x_0}^{b-x_0} \delta(x')f(x'+x_0)dx' = \begin{cases} f(x_0) & a-x_0 < 0 < b-x_0 \\ 0 & \text{otherwise} \end{cases}$$

3.4 Dirac comb

Definition 3.4.1. The **Dirac comb** is the infinite set of regularly spaced delta functions localised at the points $x_n = nP \forall n \in \mathbb{Z}$:

$$g(x) = \sum_{n=-\infty}^{\infty} \delta(x-nP)$$

One use is we can create a periodic pattern by copying some feature (e.g., peaks in a signal) represented by the function at f at points of the comb. To do this, we convolve: $f \star g$:

Lemma 3.4.1. *Show that the Fourier transform $\tilde{h}(k) = \mathfrak{F}[f \star g]$ consists of a set of delta functions of the same spacing as before, but with strengths given by $\tilde{f}(k_n)$*

Proposition 3.4.1. *The Fourier transform of a Dirac comb is a Fourier series:*

Proof. Define

$$g(x) = \sum_{n=-\infty}^{\infty} \delta(x-nP)$$

$$\implies \tilde{g}(k) = \int_{-\infty}^{\infty} \left(\sum_{n=-\infty}^{\infty} \delta(x-nP) \right) dx = \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-ikx}\delta(x-nP)dx = \sum_{n=-\infty}^{\infty} e^{-iknP}$$

□

- This is a Fourier series of a function of k
- This has period $2L = \frac{2\pi}{P}$
- Fourier coefficients $c_n = 1 \forall n$

Corollary 3.4.1. *The original function $g(x)$ can be written as*

$$g(x) = \sum_{n=-\infty}^{\infty} \frac{1}{P} e^{\frac{2in\pi x}{P}}$$

This leads to

$$\sum_{n=-\infty}^{\infty} \delta(x-nP) = \frac{1}{P} \sum_{n=-\infty}^{\infty} e^{2in\pi x/P}$$

Corollary 3.4.2. *We Fourier transform the Dirac comb, and we get*

$$\tilde{g}(k) = \sum_{n=-\infty}^{\infty} e^{-iknP} = \frac{2\pi}{P} \sum_{n=-\infty}^{\infty} \delta(k - \frac{2n\pi}{P})$$

In other words, the Fourier transform of a Dirac comb is another Dirac comb. This result manifests itself in the appearance of (crystal) diffraction patterns.

3.5 Delta function with function argument

When the argument of the delta function is itself a function, $\delta(f(x))$, it behaves like a sum of (ordinary) delta functions localised at each of the roots of the function f. This allows us to derive properties such as scaling.

Dirac delta of a function

Theorem 3.5.1. *Let $f : U_r \to \mathbb{R}, f \in C^1(\mathbb{R})$ be a one-time differentiable function with continuous derivative, such that $f(x_r) = 0 \forall r \in \mathbb{N}, f'(x_r) \neq 0 \forall r$. U_r is an open set containing the root x_r and nothing else. Let $g(x)$ be a function at least defined on the x_r. Let f, g both be integrable. Also require:*

$$\bigcup_r U_r \subset [a, b]; a, b \in \mathbb{R}$$

Then f, g satisfy

$$\int_a^b g(x)\delta(f(x))dx = \sum_{roots,i} \frac{g(x_i)}{|f'(x_i)|}$$

where the $x_i \in (a, b) \forall i$

Proof. On each U_r the map f is restricted to a bijection $f|_{U_r}$ between U_r and an open set V centred on the origin, so that by the change of variables $y = f(x)$, with inverse $x = \phi(y)$, we get

$$\int_{U_r} g(x)\delta(f(x))dx = \int_V g(\phi(y))\delta(y)\frac{1}{|f'(\phi(y))|}dy = \frac{g(\phi(0))}{|f'(\phi(9))|} = \frac{g(x_r)}{|f'(x_r)|}$$

□

Corollary 3.5.1. *Scaling of a delta function.*

$$\delta(af(x)) = \frac{1}{|a|}\delta(x)$$

Proof. Indeed, we set $f(x) \to yf(x)$ above. This yields us:

$$\int_a^n \delta(yx)g(x) = \frac{b/y}{a/y}\delta(x')g(x'/y)dx' = \frac{f(0)}{|a|}$$

□

3.6 Heaviside Step Function

Definition 3.6.1. The **Heaviside Step Function** is defined as

$$H(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

Lemma 3.6.1. *The Dirac delta is the derivative of the Heaviside step function. Similarly, the Heaviside step function is the antiderivative of the Dirac delta:*

$$\delta(x) = \frac{dH(x)}{dx} \qquad \int_a^b \delta(x)dx = H(b) - H(a)$$

3.7 Applications of Fourier Transforms and Delta function

3.7.1 Convolution of Lorentzians

Recall the Lorentzians:

$$\tilde{f}(\omega) \propto \frac{1}{\gamma^2 + \omega^2} \iff f(t) \propto e^{-\gamma|t|} \quad \tilde{F}(\omega) \propto \frac{1}{\Gamma^2 + (\omega - \Omega)^2} \iff F(t) \propto e^{-\Gamma|t|} e^{i\Omega t}$$

We will convolve the signal $\tilde{F}(\omega)$ with the resolution function $\tilde{f}(\omega)$. Using the convolution theorem, we will Fourier transform the product of the functions $F(t)f(t)$:

$$(\tilde{F} \star \tilde{f})(\omega) \propto \mathfrak{F}[F(t)f(t)] \propto \mathfrak{F}[e^{-\Gamma|t|} e^{i\Omega t} e^{-\gamma|t|}] \propto \frac{1}{(\gamma + \Gamma)^2 + (\omega - \Omega)^2}$$

As a result, the signal is **broadened**. The widths $\Gamma, \gamma \to \Gamma + \gamma$

Chapter 4

Fourier Analysis in Multiple Dimensions

We now aim to generalise everything we have done previously to n dimensions. Space is three-dimensional, so $n = 3$, or $n = 2$ as will be considered in a later chapter on diffraction. We work on the vector space $\mathbb{R}^n$ with a vector $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and the wavevector $\mathbf{k} = (k_1, \ldots, k_n)$.

4.1 Conventions and Fourier Transform

Depending on what resource you use, people will use different factors of 2π outside the integrals. We will be using the **symmetric** convention:

Definition 4.1.1. The (symmetric) **spatial** Fourier transform in n-dimensions is

$$\tilde{f}(\mathbf{k}) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{-i\mathbf{k}\cdot\mathbf{x}} \tilde{f}(\mathbf{x}) d^n x \qquad f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{i\mathbf{k}\cdot\mathbf{x}} \tilde{f}(\mathbf{k}) d^n k \tag{4.1}$$

We can also transform in time *and* space, so time $t \leftrightarrow \omega$, the (angular) frequency. So if we do both, the convention is

Complete Fourier Transform

Definition 4.1.2. Th **complete** Fourier transform of $f(\mathbf{x}, t)$ in n spatial dimensions and 1 time dimension is

$$\tilde{f}(\mathbf{k}, \omega) = \int_{\infty}^{\infty} \int_{\mathbb{R}^n} e^{-i(\mathbf{k}\cdot\mathbf{x} - \omega t)} f(\mathbf{x}) d^n x dt \tag{4.2}$$

$$f(\mathbf{x}, t) = \frac{1}{(2\pi)^{n+1}} \int_{-\infty}^{\infty} \int_{\mathbb{R}^n} e^{i(\mathbf{k}\cdot\mathbf{x} - \omega t)} \tilde{f}(\mathbf{k}, \omega) d^n k d\omega \tag{4.3}$$

with the motivation being that the standard expression for a plane wave is $\exp(i(\mathbf{k} \cdot \mathbf{x}\omega t))$ rather than $\exp(i(\mathbf{k} \cdot \mathbf{x} + \omega t))$.

4.2 Convolutions in 3D

Convolutions work exactly the same way.

Definition 4.2.1. The convolution of 2 functions $f, g : \mathbb{R}^n \to \mathbb{R}$ is $f * g(\mathbf{x})$ defined by

$$f * g(\mathbf{x}) = \int_{\mathbb{R}^n} f(\mathbf{x} - \mathbf{y}) g(\mathbf{y}) d^n y \tag{4.4}$$

In particular,

Theorem 4.2.1. *The convolution theorem (and its inverse) still hold, namely:*

$$\mathfrak{F}[f * g](\mathbf{k}) = \tilde{f}(\mathbf{k})\tilde{g}(\mathbf{k}) \tag{4.5}$$

Proof. Follow the proof as in for 1 dimension. It is identical in that regard. □

4.3 Delta functions

Definition 4.3.1. The 3D Dirac delta function is

$$\delta(\mathbf{x} - \mathbf{y}) = \delta(x_1 - y_1)\,\delta(x_2 - y_2) \cdots \delta(x_n - y_n) = \prod_{i=1}^{n} \delta(x_i - y_i) \tag{4.6}$$

If this definition isn't clear to you, remember the point of a Dirac delta is to select points which are equal to each other. 2 vectors $\mathbf{x}, \mathbf{y}$ are the same if $x_i = y_i \forall i$. Therefore their Dirac delta is the same as testing if the Dirac delta of each component is 1.

Given that, the following property still holds, which is analogous to the version in 1D:

Lemma 4.3.1.

$$\int_{\mathbb{R}^n} f(\mathbf{x}) \delta(\mathbf{x} - \mathbf{y}) d^n x = f(\mathbf{y}) \tag{4.7}$$

and it still has a relationship to the Fourier transform, namely now we are in n-dimensions

Lemma 4.3.2.

$$\delta(\mathbf{x} - \mathbf{y}) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \mathrm{e}^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})} d^n k \tag{4.8}$$

4.4 Parseval's theorem

It is the exact same thing

Theorem 4.4.1. *Parseval's theorem in n-dimensions*

$$\int_{\mathbb{R}^n} \overline{g(\mathbf{x})} f(\mathbf{x}) d^n x = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \overline{\tilde{g}(\mathbf{k})} \tilde{f}(\mathbf{k}) d^n k. \tag{4.9}$$

and from this, you get the corollary if $f = g$

Corollary 4.4.1.

$$\int_{\mathbb{R}^n} |f(\mathbf{x})|^2 d^n x = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} |\tilde{f}(\mathbf{k})|^2 d^n k \tag{4.10}$$

4.5 Derivatives

Definition 4.5.1. Gradient of a function $f : \mathbb{R}^n \to \mathbb{R}$ is an operator ∇ such that

$$\nabla f = \mathbf{e}_1 \frac{\partial f}{\partial x_1} + \mathbf{e}_2 \frac{\partial f}{\partial x_2} + \cdots + \mathbf{e}_n \frac{\partial f}{\partial x_n}, \tag{4.11}$$

where $\mathbf{e}_i$ are the normal basis vectors of $\mathbb{R}^n$. In different geometries such as spherical or cylindrical, the form will be more complicated.

Then the Fourier transform of the gradient satisfies:

Theorem 4.5.1. *The Fourier transform of a gradient of a function is*

$$\mathfrak{F}[\nabla f](\mathbf{k}) = i\mathbf{k}\tilde{f}(\mathbf{k}) \tag{4.12}$$

and this exists provided f vanishes at infinity, namely $f \to 0$ as $|\mathbf{x}| \to \infty$ so the integrals in Definition 4.1.1 converge.

Proof. Use the fact integrals are linear operators:

$$\begin{aligned}
\mathfrak{F}[\nabla f](\mathbf{k}) &= \int_{\mathbb{R}^n} \mathrm{e}^{-i\mathbf{k}\cdot\mathbf{x}} \left(\mathbf{e}_1 \frac{\partial f}{\partial x_1} + \mathbf{e}_2 \frac{\partial f}{\partial x_2} + \cdots + \mathbf{e}_n \frac{\partial f}{\partial x_n} \right) d^n x, \\
&= \mathbf{e}_1 \int_{\mathbb{R}^n} \mathrm{e}^{-i\mathbf{k}\cdot\mathbf{x}} \frac{\partial f}{\partial x_1} d^n x + \mathbf{e}_2 \int_{\mathbb{R}^n} \mathrm{e}^{-i\mathbf{k}\cdot\mathbf{x}} \frac{\partial f}{\partial x_2} d^n x + \cdots + \mathbf{e}_n \int_{\mathbb{R}^n} \mathrm{e}^{-i\mathbf{k}\cdot\mathbf{x}} \frac{\partial f}{\partial x_n} d^n x, \\
&= \mathbf{e}_1 i k_1 \tilde{f}(\mathbf{k}) + \mathbf{e}_2 i k_2 \tilde{f}(\mathbf{k}) + \cdots + \mathbf{e}_n i k_n \tilde{f}(\mathbf{k}), \\
&= i\mathbf{k}\tilde{f}(\mathbf{k}).
\end{aligned}$$

□

In general, taking a derivative will add on a prefactor of $i\mathbf{k}$ to $\tilde{f}(\mathbf{k})$ for every derivative taken. Observe that every even derivative is the divergence of another gradient, so $\nabla^2 f = \nabla \cdot \nabla f$. However every odd derivative is a gradient, so $\nabla^3 f = \nabla(\nabla \cdot \nabla f)$ and so

$$\mathfrak{F}[\nabla^{2n-1} f](\mathbf{k}) = i^{2n-1} k^{2n-2} \mathbf{k} \tilde{f}(\mathbf{k}) \qquad \mathfrak{F}[\nabla^{2n} f](\mathbf{k}) = (ik)^{2n} \mathbf{k} \tilde{f}(\mathbf{k}) \tag{4.13}$$

where $k = |\mathbf{k}|$.

4.5.1 Time and temporal FT

Note that the spatial FT of a time derivative, leaves the time derivative *unchanged*:

$$\mathfrak{F}\left[\frac{\partial f(\mathbf{x}, t)}{\partial t} \right] = \frac{\partial \tilde{f}(\mathbf{k}, t)}{\partial t} \tag{4.14}$$

if we take the time FT of a time derivative we get

$$\mathfrak{F}\left[\frac{\partial f(\mathbf{x}, t)}{\partial t} \right] = -i\omega \tilde{f}(\mathbf{x}, \omega) \tag{4.15}$$

But instead if we also take the spatial *and* time FT (yes the notation is a bit poor here)

$$\mathfrak{F}\left[\frac{\partial f(\mathbf{x}, t)}{\partial t} \right] = \omega \mathbf{k} \tilde{f}(\mathbf{k}, \omega) \tag{4.16}$$

Note both space and time each produce a factor of i which multiply to give $i^2 = -1$ hence the minus sign. However we are also taking the convention of the exponential being $\exp[i(\mathbf{k} \cdot \mathbf{x} - \omega t)]$ so we have a 3rd minus sign which cancels it out.

4.6 Using FT to solve PDEs

This is one of the most important uses for Fourier transforms, Taking FT's using the statements above can simplify PDEs to a huge extent. Of course the real solutions will depend on whether the functions are integrable in the first place.

4.6.1 Wave equation

The wave equation in 3D is

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = \nabla^2 u \tag{4.17}$$

and suppose we have an initial value problem $u(\mathbf{x}, 0) = u_0(\mathbf{x})$, $\dfrac{\partial u}{\partial t}(\mathbf{x}, 0) = v_0(\mathbf{x})$.

Taking FTs of both sides in space gives

$$\frac{1}{c^2}\frac{\partial^2 \tilde{u}}{\partial t^2} = -k^2\tilde{u}, \quad \Rightarrow \quad \tilde{u}(\mathbf{k}, t) = \tilde{f}(\mathbf{k})\mathrm{e}^{-ikct} + \tilde{g}(\mathbf{k})\mathrm{e}^{ikct}. \tag{4.18}$$

We must also FT the initial values so $\tilde{u}(\mathbf{k}, 0) = \tilde{u}_0(\mathbf{k})$, $\dfrac{\partial \tilde{u}}{\partial t}(\mathbf{k}, 0) = \tilde{v}_0(\mathbf{k})$.

We get the Fourier-transformed solution as

$$\tilde{u}(\mathbf{k}, t) = \frac{1}{2}\left[\mathrm{e}^{-ikct} + \mathrm{e}^{ikct}\right]\tilde{u}_0(\mathbf{k}) + \frac{1}{2cik}\left[\mathrm{e}^{ikct} - \mathrm{e}^{-ikct}\right]\tilde{v}_0(\mathbf{k}). \tag{4.19}$$

It is now time to inverse Fourier transform this. We suppose that all the integrals converge so we are solving

$$u(\mathbf{x}, t) = \frac{1}{(2\pi)^{3/2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)}\tilde{u}(\mathbf{k}, t)dk_1dk_2dk_3 \tag{4.20}$$

The integrand becomes

$$\begin{aligned}&\frac{1}{2}\left[\exp\left(i(\mathbf{k}\cdot\mathbf{x} - t(\omega - kc)\right) + \exp\left(-i(\mathbf{k}\cdot\mathbf{x} - t(\omega - kc)\right)\right]\tilde{u}_0(\mathbf{k}) \\ +&\frac{1}{2ick}\left[\exp\left(i(\mathbf{k}\cdot\mathbf{x} - t(\omega - kc)\right) - \exp\left(-i(\mathbf{k}\cdot\mathbf{x} - t(\omega - kc)\right)\right]\tilde{v}_0(\mathbf{k})\end{aligned} \tag{4.21}$$

We set $\omega^1(k) \coloneqq \omega - kc$ and use the formulae that

$$\cos(mx) = \frac{e^{imx} + e^{-imx}}{2} \qquad \sin(mx) = \frac{e^{imx} - e^{-imx}}{2i} \tag{4.22}$$

we get the integrand as

$$\cos(\mathbf{k}\cdot\mathbf{x} - \omega^1(k)t)\tilde{u}_0(\mathbf{k}) + \frac{1}{ck}\sin(\mathbf{k}\cdot\mathbf{x} - \omega^1(k)t)\tilde{v}_0(\mathbf{k}) \tag{4.23}$$

We do not know what $\tilde{v}_0, \tilde{u}_0$ are but let's assume we can also inverse FT them back. Doing this, will get you the solution.

4.6.2 Diffusion Equation

MathsPhys students will meet this again in PX449 Kinetic Theory in 3rd year if on the 4 year course.

In 3D, the diffusion equation is

$$\frac{\partial u}{\partial t} - D\nabla^2 u = f(\mathbf{x}, t)$$

where $f(\mathbf{x}, t)$ acts as a 'source' for the 'solute' u. There are 2 ways we can solve this, with either spatial or time FT.

Spatial FT : Taking the spatial Fourier transform we find

$$\frac{\partial \tilde{u}}{\partial t} + Dk^2 \tilde{u} = \tilde{f}(\mathbf{k}, t)$$

This is a one-dimensional ordinary differential equation in t and you can apply ODE methods you have learnt previously to do so.

The other method (and the one you'll encounter in later Physics courses) is instead taking a

Complete FT to get

$$\left(-i\omega + Dk^2\right)\tilde{u}(\mathbf{k}, \omega) = \tilde{f}(\mathbf{k}, \omega) \quad \Rightarrow \quad \tilde{u}(\mathbf{k}, \omega) = \frac{1}{-i\omega + Dk^2}\tilde{f}(\mathbf{k}, \omega) \coloneqq \tilde{G}(\mathbf{k}, \omega)\tilde{f}(\mathbf{k}, \omega) \tag{4.24}$$

Taking the inverse FT by the convolution theorem is

$$u(\mathbf{x}, t) = \int_{-\infty}^{\infty} \int_{\mathbb{R}^n} G(\mathbf{x} - \mathbf{y}, t - s) f(\mathbf{y}, s) d^n y ds \tag{4.25}$$

where G is the **Green function** for the PDE, where Green functions are the response of the differential equation to a delta function source in time and space: $f(\mathbf{x}, t) = \delta(\mathbf{x})\delta(t)$.

We can usually find Green functions explicitly, first by inverting the temporal FT:

$$\tilde{G}(\mathbf{k}, \omega) = \frac{1}{-i\omega + Dk^2}, \quad \Rightarrow \quad \tilde{G}(\mathbf{k}, t) = \begin{cases} \mathrm{e}^{-Dk^2 t} & t > 0, \\ 0 & t < 0. \end{cases} \tag{4.26}$$

Then invert the spatial FT

$$G(\mathbf{x}, t) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \mathrm{e}^{i\mathbf{k}\cdot\mathbf{x} - Dk^2 t} d^n k = \frac{1}{(2\pi)^{n/2}(2Dt)^{n/2}} \mathrm{e}^{-\frac{|\mathbf{x}|^2}{4Dt}}, \tag{4.27}$$

which is valid for $t > 0$.

Remark. The Green function method is applied here to the diffusion equation, but applies generally to linear PDEs in $\mathbb{R}^n$ with constant coefficients.

Chapter 5

Wave optics

5.1 Waves and the 1D Wave Equation

The 1D wave equation is a second-order, autonomous, homogeneous partial differential equation

$$\frac{\partial^2 \psi}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 \psi}{\partial t^2}$$

Possible solutions $\psi(x,t)$ include:

$$Ae^{i(kx-\omega t+\varphi)} \qquad A\sin(kx-\omega t+\varphi) \qquad A\cos(kx-\omega t+\varphi)$$

Definition 5.1.1. The **wavelength** is the distance one cycle of a wave takes in space. It is denoted λ

Definition 5.1.2. The **wavenumber** is defined as $k = \frac{2\pi}{\lambda}$. This can be intuitively thought of as the number of wave cycles in 1 meter of space

Definition 5.1.3. The **phase velocity** is $\frac{\omega}{k}$

The initial phase is φ

Definition 5.1.4. A wave is **monochromatic** if it only has a single frequency $\omega = ck$

Lemma 5.1.1. *The Principle of Superposition. If $\psi_1(x,t).\psi_2(x,t)$ are (linearly independent) solutions, then the solution $\psi(x,t) = \psi_1(x,t) + \psi_2(x,t)$ is also a solution to the wave equation*

Physically, the addition of the 2 waves could represent the superposition of the 2 waves at a point, where the waves come from different sources.

Definition 5.1.5. A wave with a non-zero phase velocity, i.e. not stationary, are called **travelling waves**

$$v = \frac{x' - x}{t' - t}$$

where $(x', t'), (x, t)$ are different points of the wave cycle

Definition 5.1.6. Points of constant phase are called **wavefronts**.

5.2 3D wave equation

Some definitions must be modified a bit.

Definition 5.2.1. The **wavevector** is defined as $|\mathbf{k}| = \frac{2\pi}{\lambda}$. This can be intuitively thought of as the number of wave cycles in 1 meter of space

Definition 5.2.2. A wave is **monochromatic** if it only has a single frequency $\omega = \mathbf{c} \cdot \mathbf{k}$

Our wave equation and solution now generalises to

$$\nabla^2 = \frac{1}{v^2}\frac{\partial^2 \psi}{\partial t^2} \qquad \psi(\mathbf{r}, t) = Ae^{i(\mathbf{k}\cdot\mathbf{r} - \omega t + \varphi)}$$

- The phase of the wave is now given by $\mathbf{k} \cdot \mathbf{x} - \omega t + \varphi$
- At any time $t, \psi(\mathbf{r}, t)$ is the same at all points where $\mathbf{k} \cdot \mathbf{r}$ is a constant
- Wavefronts are now *surfaces* of equal phase
- The wavevector points **normal** to the surface

5.3 Spherical Waves

If we have spherically symmetric waves $\psi(\mathbf{r}, t) = \psi(r, t)$ then the Laplacian takes the form

$$\nabla^2 \psi = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial \psi}{\partial r}\right) + \ldots = \frac{1}{r}\frac{\partial^2}{\partial r^2}(r\psi) + \ldots \tag{5.1}$$

where the azimuth ϕ and polar θ terms are dropped because the wave radiates out equally in all directions.

- The wavefronts are now spherically symmetric surfaces defined by kr being constant

5.4 Electromagnetic Waves

We will now look at electromagnetic properties and things they that do (diffraction and interference for now). The next chapter will then be entirely dedicated to Fraunhofer diffraction and a have more rigorous treatment of this property of light (and waves in general).

5.4.1 Maxwell's equations

$$\nabla^2 \mathbf{E} = \frac{1}{c^2}\frac{\partial^2 \mathbf{E}}{\partial t^2} \qquad \nabla^2 \mathbf{B} = \frac{1}{c^2}\frac{\partial^2 \mathbf{B}}{\partial t^2}$$

These are Maxwell's equations for the propagation of light *in vacuum*. These are simply vector form of the wave equation. Important properties include

- They are **transverse waves**, meaning their direction of oscillation is perpendicular to the propagation direction. Mathematically, the vector fields $\mathbf{E}, \mathbf{B}$ satisfy $\mathbf{E} \times \mathbf{B} \propto \mathbf{k}$ where $\mathbf{k}$ points in the direction of propagation.

- The vector fields $\mathbf{E} \perp \mathbf{B}$
- The **energy density/intensity/irradiance** $\propto |\mathbf{E}|^2, |\mathbf{B}|^2$ or $\propto c|\mathbf{E}|^2$
- We focus on electric field due to stronger matter coupling
- For a spherical wave, the amplitude varies as $\propto \frac{1}{r}$, so the intensity varies as $\propto \frac{1}{r^2}$. This is the **inverse square law**
- The rate of energy flow through a spherical surface around a source is independent of the sphere radius (Gauss's Law)

From now on, we consider the electromagnetic waves being monochromatic and having the same direction of propagation. This allows us to reduce our problem to the scalar wave equation for convenience.

5.4.2 Interference

Suppose we have this scenario:

- 2 point sources, placed vertically above each other (they lie in the same plane) sending out spherically symmetric, monochromatic waves at fixed initial phases φ_1, φ_2. They have position vectors $\mathbf{r}_1, \mathbf{r}_2$ respectively
- The waves have the same frequency ω, $\implies k = |\mathbf{k}_1| = |\mathbf{k}_2| = \frac{\omega}{c}$
- The waves have amplitudes E_1, E_2
- Assume electric field vectors are parallel to each other, so we can use the scalar field approach
- An observer is at position vector $\mathbf{R}$ away. Then the distances from each source are $|\mathbf{R} - \mathbf{r_1}|, |\mathbf{R} - \mathbf{r_2}|$ respectively.

Then:

$$E_1(\mathbf{R}, t) = \frac{E_1}{|\mathbf{R} - \mathbf{r_1}|} \cos(\mathbf{k} \cdot (\mathbf{R} - \mathbf{r}_1) - \omega t + \varphi_1) \quad E_2(\mathbf{R}, t) = \frac{E_2}{|\mathbf{R} - \mathbf{r_2}|} \cos(\mathbf{k} \cdot (\mathbf{R} - \mathbf{r}_2) - \omega t + \varphi_2)$$

Definition 5.4.1. The **reduced electric field** is the electric field amplitude divided by the distance

In particular, we will use this and separate out the time-independent parts of the phase. For further convenience, we define $R_1 = |\mathbf{R} - \mathbf{r}_1|, R_2 = |\mathbf{R} - \mathbf{r}_2|$

$$F_1 = \frac{E_1}{R_1} \qquad F_2 = \frac{E_2}{R_2} \qquad \phi_1 = kR_1 + \varphi_1 \quad \phi_2 = kR_2 + \varphi_2$$
$$E_1(\mathbf{R}, t) = F_1 \cos(\phi_1 - \omega t) \quad E_2(\mathbf{R}, t) = F_2 \cos(\phi_2 - \omega t)$$

The superposition is then adding these waves up. We do this by harmonic addition formulae:

$$\begin{aligned} E = (E_1 + E_2)(\mathbf{R}, t) &= F_1(\cos(\phi_1)\cos(\omega t) + \sin(\phi_1)\sin(\omega t)) + F_2(\cos(\phi_2)\cos(\omega t) + \sin(\phi_2)\sin(\omega t)) \\ &= (F_1\cos(\phi_1) + F_2\cos(\phi_3))\cos(\omega t) + (F_1\sin(\phi_1) + F_2\sin(\phi_2)) \\ E(\mathbf{R}, t) &= F\cos(\phi - \omega t) = F(\cos(\phi)\cos(\omega t) + \sin(\phi)\sin(\omega t)) \end{aligned}$$

We can clearly see the equivalence of the 2 forms.

This derivation can also be quickly done with complex exponentials by setting

$$E(\mathbf{R},t) = F\cos(\phi - \omega t) = \Re F e^{i\phi} e^{-i\omega t} = \Re A e^{-i\omega t}$$

5.4.3 Phasors

We derived above that

$$\begin{aligned} F\cos(\phi) = (F_1\cos(\phi_1) + F_2\cos(\phi_3)) \qquad & F\sin(\phi) = (F_1\sin(\phi_1) + F_2\sin(\phi_2)) \\ \implies \qquad & F^2 = F_1^2 + F_2^2 - 2F_1F_2\cos(\phi_1 - \phi_2) \end{aligned}$$

A phasor diagram is a geometric interpretation. Imagine F_1, F_2, F as vectors and forming a vector triangle for $F = F_1 + F_2$. The angles ϕ_1, ϕ_2 are the angles $F_[1], F_2$ make to the horizontal respectively.

The angle ϕ is the angle between the tail end of F and F_1.

Equivalently then, this is just the cosine law!

Furthermore, as $I \propto F^2$, then

$$I = I_1^2 + I_2^2 - 2\sqrt{I_1 I_2}\cos(\phi_1 - \phi_2) \tag{5.2}$$

If the 2 sources are coherent, the total intensity is **NOT** the sum of the intensities of both.

5.4.4 Interference and Intensity

The interference term, the $2F_1F_2\cos(\phi_1 - \phi_2)$ depends on the phase difference $\phi_1 - \phi_2$ between the 2 waves. This depends on 2 things:

- The *initial phase difference* $\varphi_1 - \varphi_2$. Imagine shifting the 2 waves initially. Then you may get different amounts of constructive and destructive interference.
- The *path difference*: $k(R_1 - R_2) = \frac{2\pi}{\lambda}(R_1 - R_2)$

The intensity is proportional to $|F|^2$, and there are 3 cases in comparison to $I_1 + I_2$:

- $I > I_1 + I_2$ suggests (complete) constructive interference
- $I < I_1 + I_2$ suggests (complete) destructive interference
- $I = I_1 + I_2$ means the polarisations of the waves are perpendicular

Furthermore, the extrema are

- $I_{\max} = I_1 + I_2 + 2\sqrt{I_1 I_2} = 4I_0$ if $I_1 = I_2 = I_0$
- $I_{\min} = I_1 + I_2 - 2\sqrt{I_1 I_2} = 0$ if $I_1 = I_2 = I_0$

These correspond to in phase and anti-phase respectively.

Additionally, if the electric field vectors are not parallel, then we must use the vector cosine rule

5.4.5 Coherence, Interference and Diffraction

Definition 5.4.2. 2 or more waves are **coherent** if they have the same frequency (or frequencies close to each other), and a constant initial phase difference. This may be zero or non-zero.

2 beams that interfere will produce a stable interference pattern if the beams are coherent

- **Optical interference** corresponds to the interaction of 2 or more light waves giving a resultant intensity which deviates from the resultant sum of the individual intensities
- **Optical diffraction** is the term describing the deviation from rectilinear propagation. This means instead of $\mathbf{k}$ being constant for a given position at a fixed time, but varying the position (e.g., along a wavefront) gives different $\mathbf{k}$

There are similar properties applied to other waves, but the physical parameters (such as wavelength) differ

5.5 Young's Double Slits

(insert figure) This is a strong piece of experimental evidence for the wave behaviour of light.

- Consider 2 coherent sources of light, with equal intensities $I_1 = I_2$, and equal initial phase $\varphi_1 = \varphi_2$
- They are incident on a screen with 2 small slits at positions $\mathbf{r}_1, \mathbf{r}_2$. The slits are separated by distance a
- There is some observer at position $\mathbf{R}$. The phase difference at this point depends only on the path difference $R_1 - R_2$ since the initial phase is the same and constant
- If the observer $\mathbf{R}$ is very far away, then $\mathbf{R} - \mathbf{r}_1, \mathbf{R} - \mathbf{r}_2$ are nearly parallel to each other and so the direction of propagation, we can approximate $\mathbf{k} = \mathbf{k}_1 = \mathbf{k}_2$
- The apertures can be taken to be infinitely small or point sources, if the vectors $\mathbf{R}, \mathbf{r}_i, \mathbf{k}$ are in a plane perpendicularto the screen (i.e., parallel to the normal vector of the screen)
- Now we use elementary geometry and we see $R_2 - R_1 = a\sin(\theta)$ where θ is the angle between the screen and the normal between the 2 waves.

If $I_1 = I_2 = I_0$, the intensity is given by (we see derivation later in Fraunhofer diffraction)

$$I \propto 2I_0 + 2I_0\cos(ka\sin(\theta))$$

$$I = \propto 2I_0\left[1 + \cos\left(\frac{2\pi a}{\lambda}\sin(\theta)\right)\right] \propto 4I_0\cos^2\left(\frac{\pi a}{\lambda}\sin(\theta)\right)$$

We use $\propto$ because we ignore the $\frac{1}{R^2}$ variance. Notice we have an oscillatory term in our proportionality relationship for intensity. This means we will get maxima and minima of intensity. These are the **interference fringes**.

To find the **fringe spacing**, we must consider the change in $\sin(\theta)$ since the other numbers are fixed. Using the small angle approximation, $\Delta(\sin(\theta)) \approx \Delta\theta = \frac{\lambda}{a}$.

Lemma 5.5.1. *Suppose the screen is a distance L away from the slits and the fringe position is denoted X, then the **fringe spacing** denoted ΔX is:*

$$\Delta X = \frac{L\lambda}{a}$$

5.5.1 Many slits

Suppose we have many slits instead. Choose a slit to have the position index 0, so it is centred on the origin. Then the slits are at positions x_j.

We again suppose amplitudes and initial phases are the same(same construction as we did for 2 slits). Again, let $\mathbf{R}$ be far away, and let $|\mathbf{R} - \mathbf{r}_j| = R_j$. Superimposing everything:

$$A(\mathbf{R}) \propto A_0 \sum_k e^{ikx\sin(\theta)}$$

Furthermore, if for each $j, R_[j] - R_0 = x_j \sin(\theta)$, this implies the equation directly above.

5.6 Huygens-Fresnel Principle

This principle was devised way before our understanding of light was that good. Hence, it has a lot of shortcomings, but is still useful. This is stated as a theorem, but isn't actually a true theorem

Theorem 5.6.1. *Huygens-Fresnel Theorem. (Incorrect, but useful tool). Every unobstructed point of a wavefront, at a given instant, serves as a source of secondary spherical wavelets, with the same frequency as the primary wave. The amplitude at any point beyond this specific point, is the superposition of all these secondary waves, taking into account their amplitudes and relative phases*

The keeping of the frequency allows us to derive laws of reflection and refraction. However, there are issues:

- Doesn't really explain how light propagates
- There should also be backwards propagating wavelets as the waves are spherical. This should cause extra interference, but this is not observed.

Chapter 6

Fraunhofer Diffraction

6.1 Fraunhofer vs Fresnel

The aim is to describe light propagation through an aperture. We use the HF principle to combine spherical wavelets originating from points **r** inside the width, and to determine the effects at position **R**.

- Fraunhofer says φ varies linearly with **r** over the aperture
- Fresnel says φ varies quadratically with **r** over the aperture

Fraunhofer diffraction works in the **far-field limit**. This is when $R_0 >> \frac{a^2}{\lambda}$ where R_0 is distance to observation, a is slit dimension and λ is wavelength.

6.2 Spherical Waves

As we know, all waves should satisfy the 3D wave equation. In this section, we consider everything emitting from point sources, and in this case, if they spread out equally in all directions, they are spherical waves, as stated before.

We further assume they are monochromatic (so we don't have to deal with a wavevector, just a wavenumber).The solution to the wave equation takes the form: $U(\mathbf{x}, t) = u(\mathbf{x})e^{-ikct}$. Substituting this into the wave equation:

$$\begin{aligned}\nabla^2 U(\mathbf{x},t) = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}U(\mathbf{x},t) &\implies \nabla^2 u(\mathbf{x})e^{-ikct} = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}u(\mathbf{x})e^{-ikct} \\ &\implies \nabla^2 u(\mathbf{x})e^{-ikct} = \frac{1}{c^2}(-ikc)^2 u(\mathbf{x})e^{-ikct} \\ &\implies (\nabla^2 + k^2)u = 0\end{aligned}$$

Definition 6.2.1. The equation $(\nabla^2 + k^2)u = 0$ is the **Helmholtz equation**. It is an eigenvalue equation with eigenvalue k^2, u is the eigenfunction.

It is a linear, second-order, homogeneous, autonomous, partial differential equation and is a time-independent form of the wave equation.

Since we are considering spherical geometry, it is ideal to use spherical coordinates to ease the calculation. In this case, the spherical Laplacian transforms the Helmholtz equation into:

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{du}{dr}\right) + k^2 u = 0$$

However recall from Equation 5.1 that the Laplacian is equal to a much simpler expression, namely, the Helmholtz equation then reduces to

$$\frac{1}{r}\frac{\partial^2}{\partial r^2}(ru) + k^2 u = 0$$

Lemma 6.2.1. *Solve the Helmholtz equation to get the fundamental solution*

Proof.

$$\frac{1}{r}\frac{d}{dr}\frac{d}{dr}ru + k^2 u = 0 \implies \frac{1}{r}\frac{d}{dr}\left(r\frac{du}{dr} + u\right) + k^2 u = 0$$
$$\frac{1}{r}\left(\frac{du}{dr} + r\frac{d^2 r}{du^2}\right) + k^2 u = 0$$
$$\implies \frac{d^2 u}{dr^2} + \frac{1}{r}\frac{du}{dr} + k^2 u = 0$$

See this resource for the rest of the details: □

We then observe $ru \propto e^{\pm ikr}$ and the fundamental solution is:

$$u = -\frac{1}{4\pi r}e^{ikr} := G(\mathbf{x}) \tag{6.1}$$

- We choose the positive exponent because the full waveform has the factor $e^{ik(r-ct)}$ and the wavefronts propagate radially outwards as time increases, i.e. they are **outgoing spherical** waves which is what one would expect from a point source. Taking the minus sign is physically interpreted as waves incident at a point, e.g. a point *sink*.

- The constant of proportionality is $-\frac{1}{4\pi}$. To see why, integrate the Helmholtz equation over a ball of radius R centred on the origin and apply the divergence theorem to the Laplacian term. This will be done in the following lemma (NON-EXAMINABLE):

Lemma 6.2.2. *NON-EXAMINABLE.*
Let $\mathbb{B}(\mathbf{0}, r)$ be the open ball of radius R centred at the origin. The term $\partial\mathbb{B}$ is the boundary of this ball. We include the boundary (divergence theorem). Let $\mathbf{n}$ be the positive outward unit normal to the boundary.

$$\int_{B(\mathbf{0},R)} (\nabla^2 + k^2) G dV = 1$$

Proof. Warning: heavy mathematics incoming:

$$\int_{B(\mathbf{0},R)} (\nabla^2 + k^2) G dV = \int_{\partial B(\mathbf{0},R)} \mathbf{n}\cdot\nabla G dA + \int_{B(\mathbf{0},R)} k^2 G dV$$
$$\int_{B(\mathbf{0},R)} k^2 G dV = \int_0^{2\pi}\int_0^{\pi}\int_0^{R} -\frac{k^2}{4\pi r}e^{ikr} r^2 \sin(\theta) dr d\theta d\phi$$

Now the integrand is continuous w.r.t. r, it is also continuous w.r.t θ. These functions are as a result, Riemann-integrable over the given domains. We can therefore use Fubini's theorem and swap the limits of integration as so:

$$\begin{aligned}
-\frac{k^2}{4\pi}\int_0^R\int_0^\pi\int_0^{2\pi} re^{ikr}\sin(\theta)d\phi d\theta dr \\
\frac{-k^2}{2}\int_0^R\int_0^\pi re^{ikr}\sin(\theta)d\theta dr \\
\frac{-k^2}{2}\int_0^R re^{ikr}\left[-\cos(\theta)\right]_0^\pi dr \\
= -k^2\int_0^R re^{ikr}dr = e^{ikR}(ikR-1)+1 \\
\int_{\partial\mathbb{B}(\mathbf{0},R)}\mathbf{n}\cdot\nabla G dA = \int_0^{2\pi}\int_0^\pi \frac{\partial}{\partial r}\left(\frac{-1}{4\pi r}e^{ikr}\right)_{r=R} R^2\sin(\theta)d\theta d\phi \\
\implies \int_{\mathbb{B}(0,R)}(\nabla^2+k^2)G dV = 1
\end{aligned}$$

□

This makes the integral independent of R, in particular, $(\nabla^2+k^2)G=\delta(\mathbf{x})$, our fundamental solution is the (outgoing) Green function for the Helmholtz equation. This makes sense: we put a source at the origin

6.3 Fraunhofer Diffraction

6.3.1 Diffraction from a slit

In this scenario, we align the coordinate system such that the origin is halfway through the slit. The $\hat{x}$ direction is aligned with the slit upwards, and the $\hat{z}$ direction lies perpendicular to x, normal to the slit. The y-axis points out of the plane.

- We have an opaque screen normal to $\hat{z}$
- The light is incident on this screen, parallel to the $\hat{z}$ direction
- The slit ranges in $\left[-\frac{a}{2},\frac{a}{2}\right]\hat{x}$
- Both the screen and slit are infinitely long in the $\hat{y}$ direction
- As a result, the physics takes place in the xz direction

We consider an arbitrary point along the slit, at $\mathbf{r}=(x,0)$ and an observer at the opaque screen at $\mathbf{R}=(X,L)$. Furthermore, we define

- R_0 is the distance between the centre of the slit (the origin) to the observer, so $|\mathbf{R}|=R_0$
- R is the distance from the position along the slit to the observer, namely $R=|\mathbf{R}-\mathbf{r}|$
- L is the perpendicular distance between the slit origin and opaque screen. We assume they are parallel, so this distance is fixed everywhere

- X is the perpendicular distance from the observer to the z-axis
- θ is the angle between the slit and normal to the rays

Clearly then by Pythagoras' theorem, $R_0 = \sqrt{X^2 + L^2}$ and $R \approx R_0 - x\sin(\theta)$.

From now on, make the assumption that the electric field amplitude E_0 and initial phase φ are independent of x.

Now:

Proposition 6.3.1. *The superposition of spherical waves at position* $\mathbf{R}$ *is given by*

$$E(\mathbf{R}) \propto \Re \int_{-\frac{a}{2}}^{\frac{a}{2}} \frac{E}{R(x)} e^{i(kR(x)-\omega t+\varphi)} dx$$

and the relative intensity as a function of θ *is*

$$\frac{I(\theta)}{I(0)} = \frac{\sin^2\left(\frac{k_x a}{2}\right)}{\frac{k_x a}{2}}$$

where $k_x = k\sin(\theta)$ *(x-component of wave-vector* $\mathbf{k}$*)*

6.4 Fraunhofer Diffraction Integral (FDI)

You **do not** need to know this integral for the exam, but it will explain why we get the following results in this chapter

The FDI can be worded as so:

Fraunhofer Diffraction Integral

The diffracted wave amplitude observed on a screen a large distance from an aperture is proportional to the Fourier transform of the aperture.

Now, consider some aperture (screen with holes) and our observation screen. We parameterise the aperture coordinates by (y_1, y_2) and the observation plane by (x_1, x_2). The total diffracted wave amplitude is then given by adding up these sources from each point of the aperture a distance D away.

$$u(\mathbf{x}) = \int_{\mathbb{R}^2} \frac{-1}{4\pi|\mathbf{x}-\mathbf{y}|} \mathrm{e}^{ik|\mathbf{x}-\mathbf{y}|} A\left(y_1, y_2\right) dy_1 dy_2 \tag{6.2}$$

where the aperture itself will be given by a function $A(y_1, y_2)$ that takes the value 1 at points where light can get through the screen and 0 at points where it cannot. This is **Huygen's principle**.

using the Fraunhofer limit, $D \gg x_1, x_2, y_1, y_2$. In the denominator, we approximate $|\mathbf{x}-\mathbf{y}| \sim D$ but not in the exponential. For the exponential term, we see

$$\begin{aligned}
|\mathbf{x}-\mathbf{y}| &= [(\mathbf{x}-\mathbf{y})\cdot(\mathbf{x}-\mathbf{y})]^{1/2} \\
&= \left[|\mathbf{x}|^2 - 2\mathbf{x}\cdot\mathbf{y} + |\mathbf{y}|^2\right]^{1/2} \\
&= |\mathbf{x}| - \frac{\mathbf{x}\cdot\mathbf{y}}{|\mathbf{x}|} + \cdots \\
&\approx D - \frac{x_1 y_1 + x_2 y_2}{D}
\end{aligned}$$

Substituting this back in to u gives the FDI as

$$u(\mathbf{x}) = \frac{-\mathrm{e}^{ikD}}{4\pi D} \int_{\mathbb{R}^2} \mathrm{e}^{-ik(x_1 y_1 + x_2 y_2)/D} A(y_1, y_2)\, dy_1 dy_2 \tag{6.3}$$

We can now analyse various 2D scenarios. Note that you should be able to attribute the diffraction pattern to the type of aperture (e.g. circular, rectangular etc.). This was a question asked in a previous exam (when this module was called PX276). You do not need to be able to do all this analysis in the exam - it is too short for that.

6.5 2D Fraunhofer Diffraction

In this section, we will be using our multi-dimensional Fourier transform and the FDI seen previously to analyse diffraction patterns through 2D slits.

6.5.1 Recap of 1D slits

Recall some important properties of fringes and slits, e.g. from Young's slits:

- The fringe spacing ΔX is **inversely proportional** to the slit spacing a, so we will observe a larger fringe for a narrower slit. This applies to *each* dimension of the slit

6.5.2 Rectangular Slit

Consider a rectangular slit of width $2w$ and height $2h$ centred on the origin, so the aperture function is

$$A(y_1, y_2) = \begin{cases} 1 & (y_1, y_2) \in [-w, w] \times [-h, h] \\ 0 & \text{otherwise.} \end{cases} \tag{6.4}$$

Using the FDI to find the observed amplitude:

$$\tilde{A}(x_1, x_2) := \int_{\mathbb{R}^2} \mathrm{e}^{-ik(x_1 y_1 + x_2 y_2)/D} A(y_1, y_2)\, dy_1 dy_2 = \left[2w \frac{\sin(kwx_1/D)}{kwx_1/D}\right] \left[2h \frac{\sin(khx_2/D)}{khx_2/D}\right] \tag{6.5}$$

This has the intensity

$$I(x_1, x_2) \propto \left|\tilde{A}(x_1, x_2)\right|^2 = I_0 \left[\frac{\sin(kwx_1/D)}{kwx_1/D}\right]^2 \left[\frac{\sin(khx_2/D)}{khx_2/D}\right]^2 \tag{6.6}$$

The distance of the zero in intensity scales **inversely** with the physical size of the aperture. Thus to see any appreciable spread in a beam the linear dimensions of the aperture should be comparable to the wavelength of the light.

There is an example in Fig 6.1.

We see that the longer side $2h = 20\lambda$ has a **shorter** diffraction pattern (on the y-axis) whereas the shorter side $2h = 10\lambda$ has a larger fringe (seen on the x-axis).

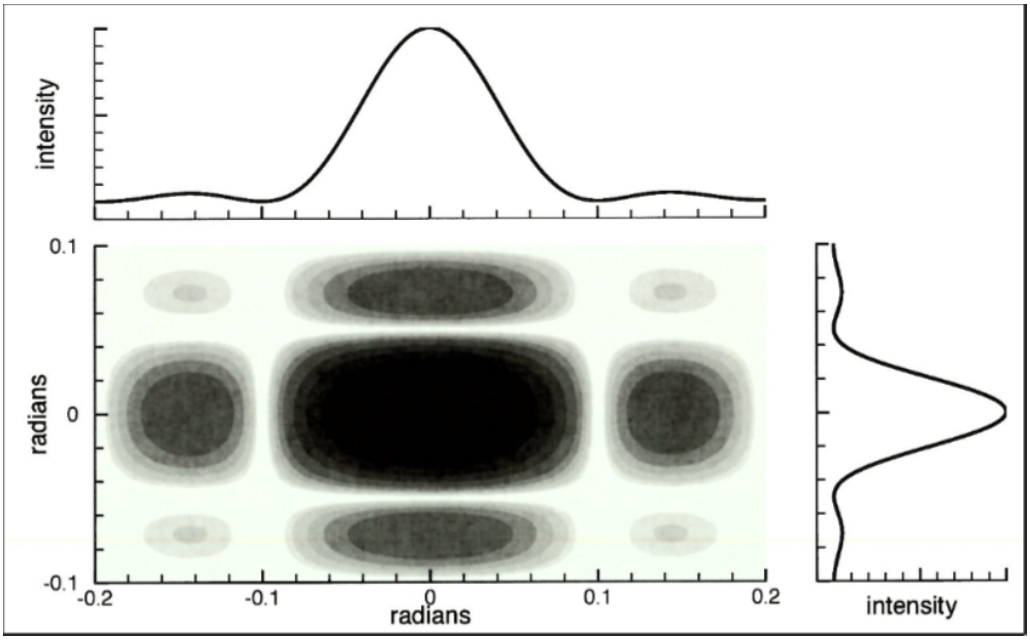

Figure 6.1: Diffraction pattern and intensity curve of a rectangular slit with dimensions $2w = 10\lambda, 2h = 20\lambda$ with λ wavelength of incident monochromatic light.

6.5.3 Non-normal incidence

Suppose the incident light makes an angle θ to the normal direction of the aperture, i.e. its wavevector is $(k\sin\theta, 0, k\cos\theta)$. This modulates the light at the aperture by a phase factor $e^{iy_1 k\sin\theta}$ so that the aperture function is now

$$A_\theta(y_1, y_2) = e^{iy_1 k\sin\theta} A_0(y_1, y_2) \tag{6.7}$$

where $A_0(y_1, y_2)$ is the aperture function for normally incident light. By the convolution theorem

$$\tilde{A}_\theta(x_1, x_2) = \tilde{A}_0(x_1 - D\sin\theta, x_2) \tag{6.8}$$

In other words, the **diffraction pattern is the same** but its **position is shifted from the centre** of the screen to a point corresponding to the straight line between the source and the screen (making an angle θ with the normal direction).

6.5.4 Young's double slits

Consider two identical rectangular slits of width $2w$ and height $2h$, with centres separated by a distance 2ℓ along the horizontal direction (y_1). The aperture function may be written in terms of that of a single rectangular slit as

$$A\left(y_1, y_2\right) = A_{\text{slit}}\ \left(y_1 + \ell, y_2\right) + A_{\text{slit}}\ \left(y_1 - \ell, y_2\right)$$

The diffracted wave amplitude then follows from the shift property of Fourier transforms

$$\tilde{A}\left(x_1, x_2\right) = \mathrm{e}^{ik\ell x_1/D}\tilde{A}_{\text{slit}}\ \left(x_1, x_2\right) + \mathrm{e}^{-ik\ell x_1/D}\tilde{A}_{\text{slit}}\ \left(x_1, x_2\right) = 2\cos\left(k\ell x_1/D\right)\tilde{A}_{\text{slit}}\ \left(x_1, x_2\right)$$

Using $I \propto |\tilde{A}|^2$, it then follows that the intensity observed on a distant screen is

$$I\left(x_1, x_2\right) = I_0\cos^2\left(k\ell x_1/D\right)\left[\frac{\sin\left(kwx_1/D\right)}{kwx_1/D}\right]^2\left[\frac{\sin\left(khx_2/D\right)}{khx_2/D}\right]^2.$$

6.5.5 Circular Apertures

Consider a circular aperture of radius a so the aperture function is

$$A(y_1, y_2) = \begin{cases} 1 & \sqrt{y_1^2 + y_2^2} < a \\ 0 & \text{otherwise} \end{cases} \tag{6.9}$$

We switch to polar coordinates with the transformation

$$r = \sqrt{x_1^2 + x_2^2} \qquad \rho = \sqrt{y_1^2 + y_1^2} \tag{6.10}$$

and ϕ for the angle in the aperture plane. Performing the FDI gives

$$\tilde{A}(r) = \int_{\phi=-\pi}^{\pi} \int_{\rho=0}^{a} \mathrm{e}^{-ikr\rho\cos\phi/D} \rho d\rho d\phi = \int_{\rho=0}^{a} \left(\int_{\phi=-\pi}^{\pi} \mathrm{e}^{-ikr\rho\cos\phi/D} d\phi \right) \rho d\rho. \tag{6.11}$$

This integral cannot be solved using elementary functions. We define its solution to be a special kind of function called a **Bessel function of the first kind** with

$$J_0(\rho) = \frac{1}{2\pi} \int_0^{2\pi} \mathrm{d}\phi \mathrm{e}^{\mathrm{i}\rho\cos\phi} \quad \text{and} \quad \int_0^{\alpha} \rho \mathrm{d}\rho J_0(\rho) = \alpha J_1(\alpha) \tag{6.12}$$

of order 0 and 1 respectively (denoted by subscripts). You don't really need to know anything about these functions, rather that they allow us to directly produce the solutions for the diffraction pattern and intensity:

$$\tilde{A}(r) = \pi a^2 \frac{2J_1(kra/D)}{kra/D} \qquad I(r) = I_0 \left(\frac{2J_1(kra/D)}{kra/D} \right)^2 \tag{6.13}$$

We call the resulting diffraction pattern (a bunch of concentric discs) the **Airy pattern** or Airy discs.

The intensity at the centre of the bright Airy disc ($r = 0$) is I_0 and it first vanishes when

$$J_1(kra/D) = 0 \quad \text{for} \quad r \neq 0$$

The first non-trivial zero of $J_1(x)$ occurs when $x \approx 1.22\pi$, which recovers the angular size (r/D) of the Airy disc as

$$\delta\theta \approx 1.22 \frac{\lambda}{2a} \tag{6.14}$$

You may recall this formula from other modules or reading. It is the Rayleigh criterion for a circular aperture. An Airy disc can be seen in Fig 6.2.

6.6 Rayleigh Criterion

Suppose we have 2 distant point sources whose light travels through some lens or mirror (like our eye, or telescope). Assume the sources are independent, i.e. their light is incoherent with each other. Then we will observe 2 diffraction patterns, one formed by each source, which may or may not overlap.

Just remember:

$$\delta\theta \approx 1.22 \frac{\lambda}{2a} = \frac{\text{wavelength of incident light}}{\text{diameter of aperture}} \tag{6.15}$$

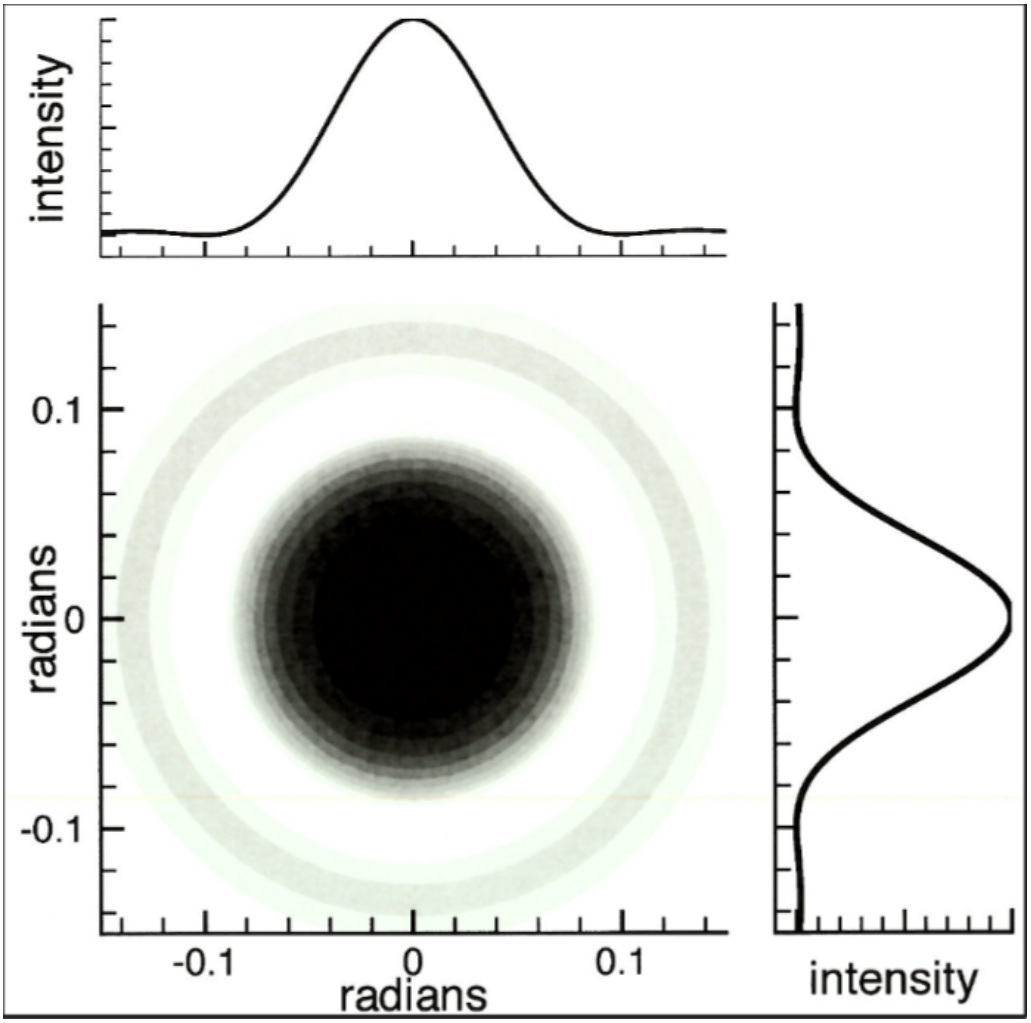

Figure 6.2: The diffraction pattern and intensity for a circular aperture, also called an Airy disc.

Definition 6.6.1. We call the minimum resolvable angle $\delta\theta_{\text{min}}$ the **critical angle** - the angle at which you can just about resolve the 2 objects, i.e. see each of them individually.

Fig. 6.3 depicts 2 objects which are barely resolvable and Fig. 6.4 depicts objects which are clearly resolvable.

6.7 Diffraction grating

Definition 6.7.1. A **diffraction grating** is an optical element (aperture) consisting of a large number of identical, equally spaced, narrow slits.

We will suppose as usual that the slits are rectangular of width $2w$ and height $2h$, and that the spacing between them is 2ℓ.

$$y_1 = -(N-1)\ell + n2\ell, \quad n = 0, 1, \ldots, N-1,$$

and $y_2 = 0$. The aperture function is

$$A(y_1, y_2) = \sum_{n=0}^{N-1} A_{\text{slit}} \ (y_1 + (N-1)\ell - 2n\ell, y_2),$$

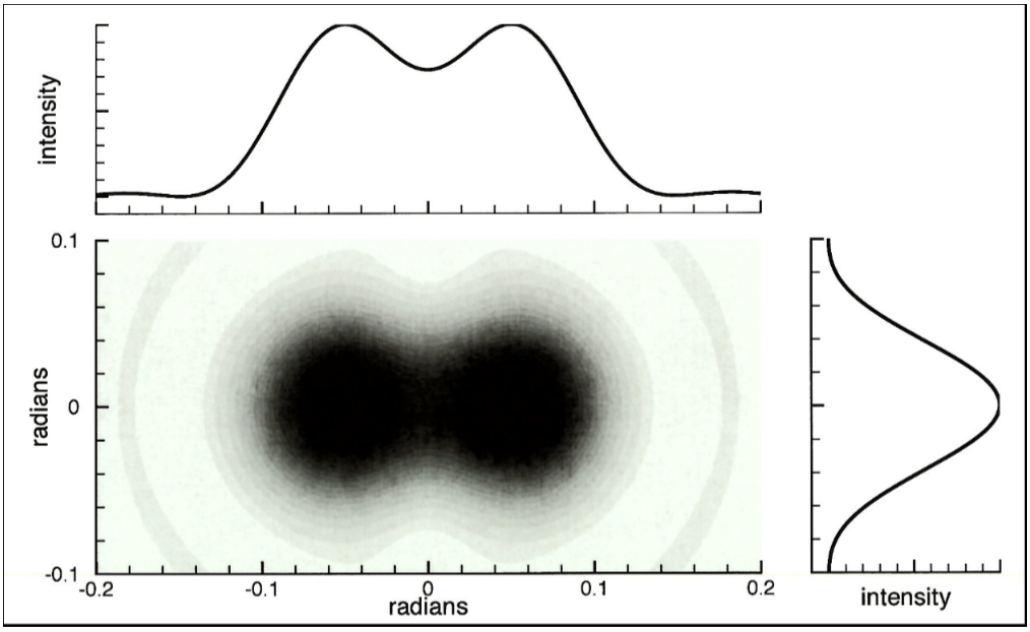

Figure 6.3: Diffraction and intensity of minimally resolvable objects, when $\delta\theta = \delta\theta_{\min}$.

and hence the diffracted wave amplitude is

$$\begin{aligned}
\tilde{A}(x_1, x_2) &= \sum_{n=0}^{N-1} \mathrm{e}^{ik(N-1)\ell x_1/D - i2nk\ell x_1/D} \tilde{A}_{\text{slit}}(x_1, x_2), \\
&= \mathrm{e}^{ik(N-1)\ell x_1/D} \left(\sum_{n=0}^{N-1} \mathrm{e}^{-i2nk\ell x_1/D} \right) \tilde{A}_{\text{slit}}(x_1, x_2), \\
&= \mathrm{e}^{ik(N-1)\ell x_1/D} \left(\frac{1 - \mathrm{e}^{-i2Nk\ell x_1/D}}{1 - \mathrm{e}^{-i2k\ell x_1/D}} \right) \tilde{A}_{\text{slit}}(x_1, x_2), \\
&= \frac{\sin(Nk\ell x_1/D)}{\sin(k\ell x_1/D)} \tilde{A}_{\text{slit}}(x_1, x_2).
\end{aligned}$$

An example of repeating rectangular slits is seen in Fig. 6.5

We observe the same behaviour as Fig 6.1 but now we have small fringes between the larger, blocky fringes. These small fringes are the interference patterns generated by the diffraction of the light through the slits.

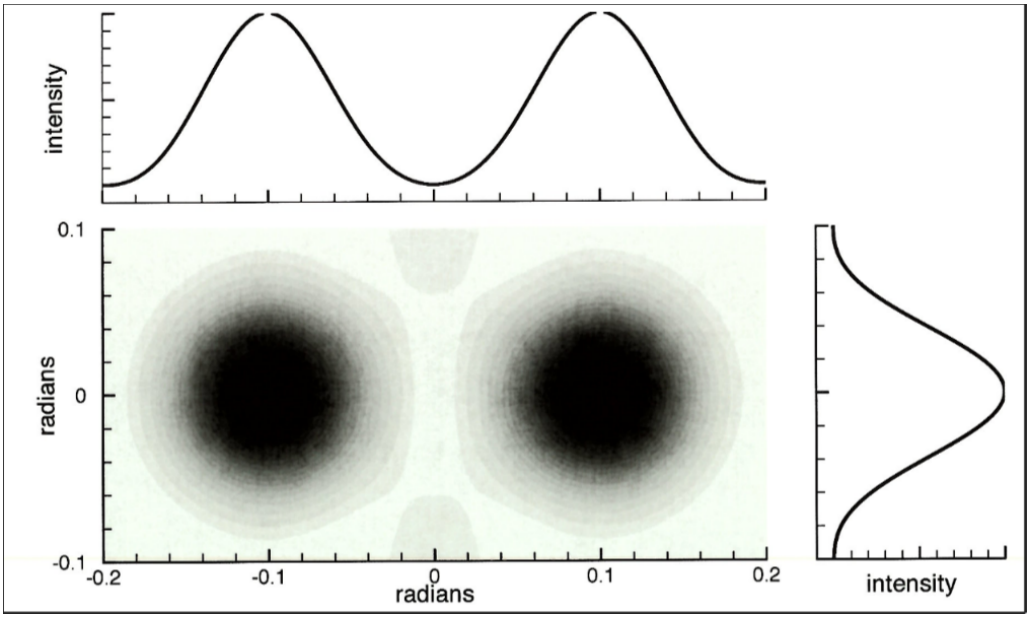

Figure 6.4: Diffraction and intensity of clearly resolvable objects, when $\delta\theta \gg \delta\theta_{\min}$.

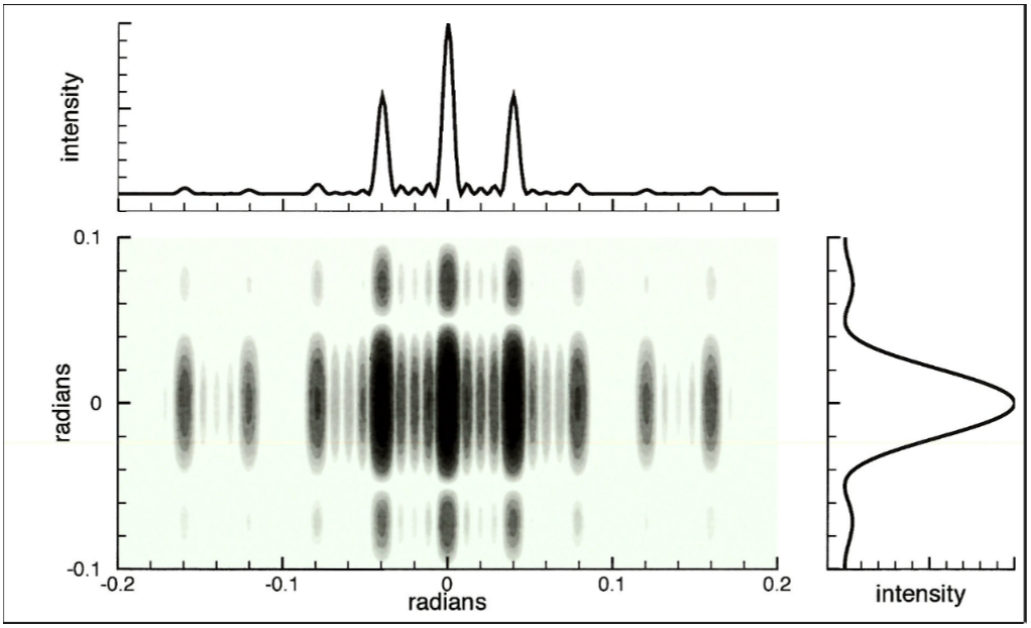

Figure 6.5: Diffraction pattern and intensity currve of a rectangular slit with dimensions $2w = 10\lambda, 2h = 20\lambda$ with λ wavelength of incident monochromatic light and $l = 25\lambda$

Chapter 7

Lagrange Multipliers

7.1 Definitions and Multivariable Calculus

The method of Lagrange multipliers is used to find critical points of functions subject to some constraint. We begin with some important definitions and mathematics.

Definition 7.1.1. For $1 \leq i \leq n, \partial_{x_i} f(x)$ is called the **ith-partial derivative** of $f : U \to \mathbb{R}^k \forall x \in U$. It is more simply denoted by $\partial_i(x) = \frac{\partial f}{\partial x_i}$

Definition 7.1.2. Let $f : \mathbb{R}^n \to \mathbb{R}^k$ be a differentiable function. Then the **grad** operator $\text{grad}(f) = \nabla f$ is a column vector of all partial derivatives of f. Namely, if $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}_n$, then

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

The definitions from now up until Definition 7.1.10 is **non-examinable** but are presented here for a bit more completeness.

Definition 7.1.3. Let $f : U \to \mathbb{R}, U \subset \mathbb{R}^2$ be a C^1 (continuously differentiable) function, and let $\mathbf{x} \in U$. Then a point $\mathbf{p}_c \in U : \nabla f(\mathbf{p}_c) = 0$ is called a **critical point** of f in U

Definition 7.1.4. Let f be a map $\mathbb{R}^n \to \mathbb{R}^k$. Then its **kernel**, $\ker(f) = \{\mathbf{x} \in \mathbb{R}^n | f(\mathbf{x}) = \mathbf{0} \in \mathbb{R}^k\}$

Definition 7.1.5. The **Jacobian matrix** at $x, \partial f(x)$ of $f : U \to \mathbb{R}^k$, where

$$f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_k(x) \end{pmatrix}$$

is a column vector, is defined as

$$\partial f(x) = \begin{pmatrix} \partial_1 f_1(x) & \ldots & \partial_n f_1(x) \\ \vdots & & \vdots \\ \partial_1 f_k(x) & \ldots & \partial_n f_k(x) \end{pmatrix}$$

For the Jacobian to exist, we have assumed that $f \in C^1(U, \mathbb{R}^k)$

Definition 7.1.6. The **Hessian matrix** at $x, \partial^2 f(x)$ of $f : U \to \mathbb{R}^k$, where

$$f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_k(x) \end{pmatrix}$$

is a column vector, is defined as

$$\partial^2 f(x) = \begin{pmatrix} \partial_{11} f_1(x) & \dots & \partial_{1n} f_1(x) \\ \vdots & & \vdots \\ \partial_{n1} f_k(x) & \dots & \partial_{nn} f_k(x) \end{pmatrix}$$

For the Jacobian to exist, we have assumed that $f \in C^2(U, \mathbb{R}^k)$

Definition 7.1.7. A critical point of f is called **non-degenerate/morse** if $\det(\partial^2 f) \neq 0$

Definition 7.1.8. A point $\mathbf{x} \in U$ is **isolated** if there exists a neighbourhood of $\mathbf{x}, N_{\mathbf{x}}$, where, $\forall \mathbf{x}_\delta \in N_{\mathbf{x}}, f(\mathbf{x}_\delta) \neq 0$.

This means a critical point is isolated, if any points really close to it, are not critical. A more topological definition is:

The functions:

$$x^2 + y^2 + z^2 \qquad x^2 + y^2 - z^2 \qquad x^2 - y^2 - z^2 \qquad -x^2 - y^2 - z^2$$

Have gradients:

$$2x + 2y + 2z \qquad 2x + 2y - 2z \qquad 2x - 2y - 2z \qquad -2x - 2y - 2z$$

which are zero at $\mathbf{x} = (x, y, z) = (0, 0, 0) = \mathbf{0}$. In particular, this is an isolated critical point, since if we move by a vector $\delta\mathbf{x} = (\delta x, \delta y, \delta z) \neq \mathbf{0}$, then the gradients become non-zero instantly

Definition 7.1.9. The number of minus signs appearing in each function (0, 1, 2 or 3) is the dimension of the negative eigenspace of the Hessian at the critical point, which is known as the **Morse index**

With this Morse index, we can use the following statement (without proof, non-examinable)

Corollary 7.1.1. *All non-degenerate critical points are equivalent to one of these canonical forms - the quadratic equations displayed above*

We will not be using this lemma directly, but it's important to realise for constraint problems, that there is often a non-trivial solution(s), i.e. the set of all critical points is not just the origin.

Definition 7.1.10. A **constraint function** is a function $g : \mathbb{R}^n \to \mathbb{R}$ which needs to vanish (equal 0)

This is very important for the method of Lagrange multipliers. Since we need to satisfy the constraint function, we need to find all points $\mathbf{c} \in \mathbb{R}^n : g(\mathbf{c}) = 0$. This is the **kernel** of the map.

This means, we must restrict our target function f to have domain of $\ker(g)$, i.e. instead of considering $f : \mathbb{R}^n \to \mathbb{R}$, we will be forced to restrict the domain of f such that we consider $f : \ker(g) \to \mathbb{R}$

7.2 Lagrange Multipliers

Lagrange Multipliers is mathematical method to solve (a system of) equations f subject to constraint functions $g_1, \ldots, g_k$. We will mathematically formulate a generic Lagrange Multipliers problem :

Theorem 7.2.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^1 function and let there be k constraint functions $g_1, \ldots, g_k : \mathbb{R}^n \rightarrowtail \mathbb{R}$ that are also C^1. Let $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$.*

Then the critical points of f subject to the constraints $g_i = 0$ are given by the unconstrained critical points of the function

$$f - \sum_{i=1}^{k} \lambda_i g_i$$

Namely, the following set of partial differential equations are satisfied:

$$\frac{\partial f}{\partial x_j} - \sum_{i=1}^{k} \lambda_i \frac{\partial g_i}{\partial x_j} = 0 \forall 1 \leq j \leq n \tag{7.1}$$

$$\implies \nabla f - \sum_{i=1}^{k} \lambda_i \nabla g_i = 0 \tag{7.2}$$

*where ∇f is the **grad** of f, and the λ_i are the **Lagrange Multipliers***

Remark. Why do we require $g_i = 0$? Well it's more of a convenience. Constraint functions could be in the form of something like $x^2 = 3y^2$. But since we are in real Hilbert space, we can make the rearrangement $x^2 - 3y^2 = 0$ and then cast this to $g(x, y) = 0$!

This means, when using this method, we will have:

- n partial differential equations for functions of n variables. This leads to n equations and n unknowns: the $x_1, \ldots, x_n$
- The k Lagrange multipliers, taking us up to $n + k$ unknowns with n equations
- But we have the k constraint equations, so we have $n + k$ equations for $n + k$ unknowns
- By the rank-nullity theorem we can find non-trivial critical points.

Remark. We need not use Cartesian coordinates x, y, z. We could have a problem in spherical, hyperbolic, conical coordinates... whatever we want really. We just need to make sure we differentiate with respect to the appropriate variables.

7.3 Stationary Phase and Fourier

Wave disturbances in a field $u(x, t)$ can be written as a Fourier transform:

$$u(x, t) = \frac{1}{(2\pi)^2} \int \int e^{i(kx - \omega t)} \tilde{u}(k, \omega) dk d\omega$$

where the integral is taken over all 2-tuples $k.\omega$ that satisfy the **dispersion relation** $D(k, \omega) = 0$.

In such an integral, the oscillations in the complex exponential interfere destructively except near those points (k, ω) where the phase is stationary, subject to the constraint that the dispersion relation is satisfied.

Thus, the wave disturbance can be approximated by the condition of stationary phase.

Using Lagrange Multipliers, the constrained critical points of the phase satisfy

$$x - \lambda \frac{\partial D}{\partial k} = 0 \qquad\qquad -t - \lambda \frac{\partial D}{\partial \omega} = 0$$

So then, rearranging and performing x/t:

$$\frac{x}{t} = \frac{-\partial D/\partial k}{\partial D/\partial \omega} = \frac{\partial \omega}{\partial k}$$

This is the **group velocity**

7.4 Shannon Entropy

Claude Shannon introduced a method for choosing probabilities based on only partial information using Lagrange Multipliers. Suppose X is a discrete random variable taking one of N values $x_i, i = 1, \ldots, N$, with probabilities p_i. If all we are told is that the expectation value of X is $\langle X \rangle$, then what values should we take for the probabilities? The solution is to maximise the entropy function:

$$S = -\sum_{i=1}^{N} p_i \ln(p_i)$$

subject to the constraints which are the *normalisation condition* (sum of probabilities add to 1) and that the expectation value is equal to $\langle X \rangle$:

$$g_1 = \sum_{i=1}^{N} p_i - 1 = 07 g_2 = \sum_{i=1}^{N} p_i x_i - \langle X \rangle = 0$$

As per the method, we have N equations of the form

$$\frac{\partial}{\partial p_i}(S = \lambda_1 g_1 - \lambda_2 g_2) = -(\ln(p_i + 1))\lambda_1 - \lambda_2 x_i = 0 \forall i$$

alongside the 2 constraint equations, for a total of $N + 2$ equations for $N + 2$ unknowns.

Rearranging for the probabilities:

$$p_i = e^{-(1+\lambda_1)-\lambda_2 x_i} = e^{-(1+\lambda_1)} e^{-\lambda_2 x_i} \tag{7.3}$$

Using the constraint function g_1:

$$\sum_{i=1}^{N} e^{-(1+\lambda_1)} e^{-\lambda_2 x_i} - 1 = 0$$

$$\sum_{i=1}^{N} e^{-\lambda_2 x_i} = e^{1+\lambda_1} := Z$$

We have just derived the **partition function** (see Statistical Mechanics).

The Lagrange multiplier λ_2 is determined from the constraint function g_2 from substituting in the partition function:

$$\sum_{i=1}^{N} p_i x_i = \langle X \rangle \implies \sum_{i=1}^{N} e^{-(1+\lambda_1)} e^{-\lambda_2 x_i} x_i = \langle X \rangle$$
$$= \frac{1}{e^{-(1+\lambda_1)}} \sum_{i=1}^{N} e^{-\lambda_2 x_i} x_i = \langle X \rangle$$
$$\implies \langle X \rangle = \frac{1}{Z} \sum_{i=1}^{N} e^{-\lambda_2 x_i} x_i = -\frac{\partial \ln(Z)}{\partial \lambda_2}$$

We have derived the expectation value!

7.5 Summary of Lagrange Multipliers

This is the general method to approach a Lagrange Multipliers question. It is almost guaranteed to appear and should be one of those questions that is a free 6-8 marks. So, here we go:

1. Identify the function you are trying to extremize (call it something like F if it isn't labelled). Additionally, identify all your constraint functions. Rearrange them so that they each equal 0, and label each function something memorable like g_1, g_2 etc.
2. Formulate the PDEs as in Eq. (7.1).
3. Solve them to either get a relationship between the Lagrange multipliers and known quantities, and eliminate them. Make sure whatever expressions you get are consistent with the other PDEs
4. Substitute the minimising parameter(s) back into F to get the quantity desired (aka., read the question again and confirm what it wants as a final answer).

Chapter 8

Indices and Notation

8.1 Einstein Summation Notation

Einstein Summation Notation can be summarised as following:

Do not write summation signs; they are implied by repeated (twice only) indices.

With the following remarks:

- Writing an index *once* implies *no* summation
- Writing an index *three times or more* implies you did an oopsie somewhere

Regardless, one should always check that the expression makes sense:

- The indices must span over the correct range of numbers. This is usually clear from context but it's better to explicitly state it.

Definition 8.1.1. A **dummy index** is a index that is summed over.

It's called dummy because even if you change the symbol, it still represents an index being summed over some range. There is the equivalent

Definition 8.1.2. A **free index** is an index which represents an independent dimension of a quantity. It is not used to represent summations.

Let's go through some examples to make this concrete. We first clear up some notation:

A **vector** in $\mathbb{R}^n$ is defined by $\mathbf{x} = (x_1, \ldots, x_n)$. A vector space like $\mathbb{R}^n$ has a **basis**, which is a subset containing vectors which are linearly independent and span the entire space. We denote these vectors by $\mathbf{e}_i$ for the ith basis vector. Then

$$\mathbf{x} = \sum_{i=1}^{n} x_i \mathbf{e}_i = x_i \mathbf{e}_i \tag{8.1}$$

Notice: we have removed the sum but kept the i. We *implicitly* sum over $i = 1, \ldots, n$. We could turn the i into a δ and it would mean the same thing. This is a **dummy index**.

Definition 8.1.3. Let V be a Euclidean vector space with an inner product (i.e. the regular dot product). Let $\mathbf{a}, \mathbf{b} \in V$, where $\mathbf{a} = (a_1, a_2, a_3); \mathbf{b} = (b_1, b_2, b_3)$. Then the dot product $(\cdot, \cdot)$ is defined as

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^{3} a_i b_i = a_i b_i$$

So the idea of index notation can greatly simplify writing stuff down as long as summation convention is implied.

We will denote an $m \times n$ matrix by (M_{ij}) where i represents the row index and j represents the column index. Note that M_{ij} (without the brackets) represents the (i, j)th element of the matrix M. Observe that whilst we can change i, j to be different letters, e.g. $M_{ij} = M_{\alpha\beta}$ if we set $i = \alpha, j = \beta$, we can't swap their *positions* without possibly changing the meaning! Namely, $M_{ij} \neq M_{ji}$ unless the matrix is symmetric! This means they are **free indices** - they are represent one 'dimension', are independent of each other and are not always used for summation.

So now we can extend this notation to various things.

8.1.1 Matrix multiplication and vectors

Consider 2 matrices A, B. Then their product $AB = (AB)_{ij}$ is defined as

$$(AB)_{ij} = \sum_{k=1} A_{ik} B_{kl} = A_{ik} B_{kj} \tag{8.2}$$

Remember to get the (i, j)th element of a product of matrices, you take the ith row of matrix A, then you *add* up the products of each element in that row (i.e., you index the columns by k) and multiply each element by the corresponding element in the jth column of B (i.e. you index the rows of B by k). To help you see it more clearly, an explicit calculation of a 3x3 multiplication is done:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \tag{8.3}$$

Consider the element c_{12}. To find this, you would 'multiply' row 1 of A by column 2 of B, so $c_{12} = a_{11}b_{12} + a_{12}b_{22} + a_{31}b_{32}$. Notice only the inner 2 indices are changing - this is what we are summing over!

If you were to multiply 3 matrices A, B, C together, the same logic applies: $(ABC)_{ij} = A_{ik}B_{kl}C_{lj}$. Here, we have 2 summations: one over k and one over l - they are the dummy indices. i, j are free (unlike us students :( ).

Next, consider the equation

$$A\mathbf{x} = \mathbf{b}$$

We can convert this into index notation. $\mathbf{x}, \mathbf{b}$ are vectors - so they only have 1 dimension we can write them with one index, e.g. x_j and b_i. $A = (A)_{ij}$, so

$$A\mathbf{x} = \mathbf{b} \equiv A_{ij} x_j = b_i$$

Namely, to get the ith element of $\mathbf{b}$, we take the ith row of A and multiply it by the elements of vector $\mathbf{x}$.

8.2 Special tensors

We will not cover what tensors are, but these 2 symbols are quite special and we will make use of them a lot.

Definition 8.2.1. The **Kronecker Delta** δ_{ij} is defined as

$$\delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases} \tag{8.4}$$

Observe that the identity matrix I_{ij} has ones when the row and column index match, i.e. $i = j$ and 0 when they don't. This is exactly what the Kronecker delta represents, so $I_{ij} = \delta_{ij}$.

Definition 8.2.2. The **Levi-Civita** or fully antisymmetric tensor $\epsilon_{ijk\ldots}$ exists in all dimensions and its elements are one of $0, \pm 1$. We will need only up to 2D (ϵ_{ij}) and 3D (ϵ_{ijk}) in this module, but its definition is quite confusing.

In 3D, i, j, k can each be any of $1, 2, 3$. First point:

$$\text{If any time } i = j \text{ or } j = k \text{ or } i = k, \text{ then it is 0. For example,}$$
$$\epsilon_{111} = \epsilon_{112} = \epsilon_{121} = \ldots = 0.$$

Second, consider the element $\epsilon_{123} = 1$, and then any element ϵ_{i_1,i_2,i_3}. We can pairwise swap 2 indices. Define p - the **parity** - as the number of pairwise swaps you need to do so that the order of i_1, i_2, i_3 is the same as 123. Then:

$$\epsilon_{i_1,i_2,i_3} = (-1)^p \epsilon_{123} \tag{8.5}$$

8.2.1 Indices that aren't being summed

So we have free indices and dummy indices. Can we still have indices which aren't either? Namely indices which are literally just labels? Yes! Consider the eigenvalue equation

$$A\mathbf{x} = \lambda \mathbf{x} \equiv A_{ij} x_j^\alpha = \lambda_\alpha x_i^\alpha \qquad\qquad \alpha = 1 \ldots n$$

Here, although α is an index, it only labels what number eigenvalue-eigenvector is being considered. It does *not* imply a sum!

8.3 Matrices

We are now going to state a whole bunch of definitions for matrices which you may be familiar with, but recast it all in index notation with summation convention. Some definitions will need multiple read throughs and it is a good idea to try and derive these yourself. Let A, B, C be matrices.

Definition 8.3.1. The **transpose** of a matrix $A = (A)_{ij}$ is when every row becomes a column and every column becomes a row. It is denoted $A^T = (A)_{ji}$

Definition 8.3.2. A matrix A is **symmetric** if $A_{ij} = A_{ji}$ for all i, j. This can only be true if the matrix is square.

A matrix A is **antisymmetric** if $A_{ij} = -A_{ji}$ for all i, j. This can only be true if the matrix is square.

Definition 8.3.3. A matrix is **orthogonal** if $A^T A = I$. Equivalently, $A_{ki}A_{kj} = \delta_{ij}$.

An example of an orthogonal matrix is the 2D rotation matrix.

Definition 8.3.4. The **trace** of a matrix A is the sum of all elements of its leading diagonal, so $\operatorname{tr} A = A_{ii}$.

Lemma 8.3.1. *The determinant of a* 3×3 *matrix is*

$$\det A = \epsilon_{ijk} A_{1i} A_{2j} A_{3k} \tag{8.6}$$

In 2D, it is

$$\det A = \epsilon_{ij} A_{1i} A_{2j} \tag{8.7}$$

Proof. This is non-examinable and requires some linear algebra/group theory from first or second year modules. However:

You must be able to use these definitions to prove determinant properties, namely you should be able to prove

$$\det AB = \det A \det B \tag{8.8}$$

□

8.3.1 Gotchas

Matrices by constructions are **not commutative**, i.e $AB \neq BA$ in general, except in some cases. However, scalars do commute[1].

Suppose we have two vectors $\mathbf{a}, \mathbf{b}$ and a matrix C. Then

$$C_{ij}a_j = C_{ij}b_j \not\Longrightarrow a_j = b_j \tag{8.9}$$

Moreover, there is no reason based off index notation that C would have an inverse.

Be careful when you swap the orders of terms in index notation. The only reason you can do this is because they are *numbers*. Just because you can swap them around in index form doesn't eman you can swap them in vector form!

8.4 Vector Calculus

Consider the divergence of a vector field $\mathbf{v} = (v_1(x_1, \ldots, x_n), \ldots, v_n(x_1, \ldots, x_n))$

$$\nabla \cdot \mathbf{v} = \frac{\partial v_1}{\partial x_1} + \ldots + \frac{\partial v_n}{\partial x_n} = \sum_{i=1}^{n} \frac{\partial v_i}{\partial x_i}$$

In particular, observe we again have 2 occurrence of the index i, so using Einstein notation we have:

$$\nabla \cdot \mathbf{v} = \frac{\partial v_i}{\partial x_i}$$

An alternative way to write this, is to define:

$$\partial_i = \frac{\partial}{\partial x_i} \implies \nabla \cdot \mathbf{a} = \partial_i v_i$$

This is a more concise way of writing the divergence.

[1] Matrices are vector spaces over a field of scalars

Lemma 8.4.1. *Consider the vector product in 3D of 2 vectors:* $\mathbf{u} = (u_1, u_2, u_3, \mathbf{v} = (v_1, v_2, v_3)$. *The vector product is:*

$$\mathbf{u} \times \mathbf{v} = (u_2v_3 - v_2u_3)\mathbf{e}_1 + (u_3v_1 - v_3u_1)\mathbf{e}_2 + (u_1v_2 - v_1u_2)\mathbf{e}_3$$

Then we can write the ith component of $\mathbf{u} \times \mathbf{v}$ *as* $(\mathbf{u} \times \mathbf{v})_i = \epsilon_{ijk}u_jv_k$

Proof. Consider the component of $\mathbf{e}_1$, which is $u_2v_3 - v_2u_3$. We observe that only 2s and 3s appear in this expression. In particular, we have $+u_2v_3\mathbf{e}_1$ which corr □

This allows us to define the curl of a vector (field):

Corollary 8.4.1.

$$(\nabla \times \mathbf{u})_i = \epsilon_{ijk}\partial_j v_k \tag{8.10}$$

Proposition 8.4.1. *Maxwell's equations in free space can be written in index notation as*

$$\begin{array}{lll} \nabla \cdot \mathbf{E} = 0 & \longrightarrow & \partial_i E_i = 0, \\ \nabla \cdot \mathbf{B} = 0 & \longrightarrow & \partial_i B_i = 0, \\ \nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0 & \longrightarrow & \epsilon_{ijk}\partial_j E_k + \partial_t B_i = 0, \\ \nabla \times \mathbf{B} - \mu_0\epsilon_0\frac{\partial \mathbf{E}}{\partial t} = 0 & \longrightarrow & \epsilon_{ijk}\partial_j B_k - \mu_0\epsilon_0\partial_t E_i = 0. \end{array}$$

Proof. This can be easily derived by using the definitions and statements proven above. □

Remark. A fairly nasty index question in the 2023 exam involved Maxwell's equations and proving vector identities!

You will be expected to use this to prove general vector calculus identities in the exam.

EXAMPLE 2.

We calculate $\nabla \times \nabla\phi$ using index notation

$$(\nabla \times \nabla\phi)_i \overset{1}{=} \epsilon_{ijk}\partial_j\partial_k\phi \overset{2}{=} \epsilon_{ijk}\partial_k\partial_j\phi \overset{3}{=} -\epsilon_{ikj}\partial_k\partial_j\phi \overset{4}{=} -(\nabla \times \nabla\phi)_i.$$

The first step converts to index notation. The second uses symmetry of mixed partial derivatives $\partial_j\partial_k\phi = \partial_k\partial_j\phi$. The third uses antisymmetry of the Levi-Civita symbol, $\epsilon_{ijk} = \epsilon_{ikj}$. The fourth converts back from index notation; the conclusion is clearly that $\nabla\times\ \nabla\phi = 0$.

Lemma 8.4.2. *Prove the identity* $\nabla \times (\nabla \times \mathbf{E}) = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2\mathbf{E}$.

Proof. The left-handside is

$$[\nabla \times (\nabla \times \mathbf{E})]_i \overset{1}{=} \epsilon_{ijk}\partial_j\left(\epsilon_{klm}\partial_l E_m\right) \overset{2}{=} \left(\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}\right)\partial_j\partial_l E_m \overset{3}{=} \partial_i\partial_j E_j - \partial_j\partial_j E_i$$

□

8.5 Tensors

Vectors talk about quantities with a magnitude and one direction, e.g. velocity is a vector where each element corresponds to a speed in one direction. A tensor generalises this to talk about quantities in *multiple directions*. They also obey specific transformation laws.

8.5.1 Orthogonal Matrices and Coordinate transforms

Under a general change of coordinates, the new coordinates x_i' will be related to the old ones by a transformation

$$x_i' = R_{ij}x_j + t_i,$$

where t_i is a translation and R_{ij} an orthogonal transformation - rotation or reflection. We will restrict to transformations that preserve the origin and have $t_i = 0$. These are known as **homogeneous** transformations.

Any such transformation does not change the length of the vector $\mathbf{x}$, which implies

$$x_i x_i = x_i' x_i' = R_{ij}x_j R_{ik}x_k, \quad \Rightarrow \quad R_{ij}R_{ik} = \delta_{jk}.$$

There are 2 types of orthogonal transformations: the transformations that preserve orientation (i.e. the rotations) and those that reverse it (i.e. the reflections).

Since the coordinates x_i are equally the components of the position vector $\mathbf{x}$, the relationship $x_i' = R_{ij}x_j$ also describes how the components of a vector change under the coordinate transformation. Thus, for any other vector like the force f_i or electric field E_i we have the same relationship

$$f_i' = R_{ij}f_j, \quad E_i' = R_{ij}E_j.$$

Let us see what this implies for other tensors. We use as an example the conductivity, $J_i = \sigma_{ij}E_j$. We have, trivially, that $J_i' = \sigma_{ij}'E_j'$ and then by the transformation for vector components

$$R_{ik}J_k = \sigma_{ij}'R_{jl}E_l, \quad \Rightarrow \quad R_{ik}\sigma_{kl} = R_{jl}\sigma_{ij}'.$$

Finally, using that R_{ij} is an orthogonal transformation we can write this as

$$\sigma_{ij}' = R_{ik}R_{jl}\sigma_{kl}.$$

For a tensor of **rank** n, one finds the natural extension of this relationship: the components transform according to

$$T'_{i_1 i_2 \cdots i_n} = R_{i_1 j_1}R_{i_2 j_2}\cdots R_{i_n j_n}T_{j_1 j_2 \cdots j_n}$$

8.5.2 Isotropic Tensors

Definition 8.5.1. An **isotropic tensor** is one whose components are independent of the coordinate system.

This means if you try to transform an isotropic tensor by any orthogonal transformation R_{ij}, you just end up with the same tensor.

Lemma 8.5.1. *The Kronecker Delta δ_{ij} is isotropic.*

Proof. The kronecker delta is a rank 2 tensor, so we try to transform with two orthogonal transformations, we see that

$$R_{ik}R_{jl}\delta_{kl} = \delta_{ij} \tag{8.11}$$

□

Proposition 8.5.1. *The only linearly independent rank 3 isotropic tensor is the Levi-Civita symbol ϵ_{ijk}*

Proof. See this useful document. □

Proposition 8.5.2. *There are 3 linearly independent 4th rank isotropic tensors:*

$$\delta_{ij}\delta_{kl} \qquad \delta_{ik}\delta_{jl} \qquad \delta_{il}\delta_{jl} \tag{8.12}$$

Remember - these are all **different** because the positions of the free indices are different. At this point it's not worth trying to think of these as matrices, they would be '4D' matrices at this point.

Look at that, you're at the end of the module now, good luck with revision!