NST Mathematics Part IB Summary Notes

Tai Yingzhe, Tommy (ytt26)

Contents

Notation. *b.c.s stands for boundary conditions, i.c.s stands for initial conditions, s.t. stands for such that.*

Foreword. *A notes compilation for IB Mathematics for the Natural Science Tripos. Each section is taken from the respective section from the Mathematics Tripos. I would recommend readers to read the course notes from the Mathematics Tripos since concepts are discussed in greater detail and with more rigour. Do let me know by email (crsid: ytt26) if there are any mistakes in this compilation.*

1   Vector calculus

Suffix notation. Einstein summation convention. Contractions using δ_{ij} and ϵ_{ijk}. Reminder of vector products, grad, div, curl, ∇^2 and their representations using suffix notation. Divergence theorem and Stokes' theorem. Vector differential operators in orthogonal curvilinear coordinates, e.g. cylindrical and spherical polar coordinates. Jacobian **[6]**

1.1   Recap

We begin the topic on vector calculus with a quick review on important concepts introduced in Part IA NST Mathematics.

Differentials

If $f = f(u_1, \ldots, u_n)$, then the differential $df = \frac{\partial f}{\partial u_i} du_i$. The $\{du_i\}$ are called differential forms, defined to be linearly independent if $\{u_i\}$ are independent. Similarly, if $\mathbf{x} = \mathbf{x}(u_1, \ldots, u_n)$, then $d\mathbf{x} = \frac{\partial \mathbf{x}}{\partial u_i} du_i$. By ordinary calculus,

$$f(u_1 + \delta u_1, \ldots, u_n + \delta u_n) - f(u_1, \ldots, u_n) = \frac{\partial f}{\partial u_i} \delta u_i + o(\delta \mathbf{u})$$

where $\frac{o(\delta \mathbf{u})}{|\delta \mathbf{u}|} \to 0$ as $|\delta \mathbf{u}| \to 0$. Up to first order, $\delta f = \frac{\partial f}{\partial u_i} \delta u_i$, then similarly $\delta \mathbf{x} = \frac{\partial \mathbf{x}}{\partial u_i} \delta u_i$.

Proposition 1.1 (Multivariate chain rule)**:** Let $(u_1, \ldots, u_n)$ be a new set of coordinates that depend on the old coordinates $(x_1, \ldots, x_n)$. If $F(u_1, \ldots, u_n) = f(x_1, \ldots, x_n)$, where $x_i = x_i(u_1, \ldots, u_n)$, $i = 1, \ldots, n$. Then

$$\frac{\partial F}{\partial u_i} = \frac{\partial f}{\partial x_j} \frac{\partial x_j}{\partial u_i} \tag{1.1}$$

Gradient

Definition 1.1 (Gradient Operator)**:** The gradient operator is defined s.t. for a function $f : \mathbb{R}^3 \to \mathbb{R}$, it satisfies

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + \mathbf{h} \cdot \boldsymbol{\nabla} f(\mathbf{x}) + o(\mathbf{h})$$

The directional derivative $D_{\mathbf{v}}$ is s.t.

$$f(\mathbf{x} + t\mathbf{v}) = f(\mathbf{x}) + t D_{\mathbf{v}} f(\mathbf{x}) + o(t)$$

Recall that $\boldsymbol{\nabla} f$ points in a direction of gradient increase in f.

Proposition 1.2 (Coordinates independent chain rule)**:** Suppose we have a curve $t \mapsto \mathbf{x}(t)$, $F(t) = f(\mathbf{x}(t))$ changes like

$$\frac{dF}{dt} = \frac{d}{dt} f(\mathbf{x}(t)) = \frac{d\mathbf{x}}{dt} \cdot \boldsymbol{\nabla} f(\mathbf{x}(t)) \tag{1.2}$$

Corollary 1.1: $\boldsymbol{\nabla} f(\mathbf{x})$ is normal to any level surface $f(\mathbf{x}) = 0$ in $\mathbb{R}^3$ at the point $\mathbf{x}$.

Line integral and conservative fields

Definition 1.2 (Curve)**:** A curve C is the image of a map $t \in [a, b]$ s.t. $t \mapsto \mathbf{r}(t)$.

Definition 1.3 (Differentiable curve)**:** C is differentiable if each component $r_i(t)$ is differentiable.

Definition 1.4 (Regular curve)**:** C is regular if $|\mathbf{r}'(t)| \neq 0 \ \forall t$.

Definition 1.5 (Smooth curve)**:** C is smooth if C is differentiable and regular.

Remarks 1.1. *So for a differentiable curve C parametrised by $t \mapsto \mathbf{r}(t)$, where $a \leq t \leq b$, we have*

$$\ell(C) = \int_C ds = \int_a^b |\mathbf{r}'(t)| dt$$

Similarly we define

$$\int_C f(\mathbf{r}) ds = \int_a^b f(\mathbf{r}(t)) |\mathbf{r}'(t)| dt$$

We think of ds as the arc-length element. Informally we have $ds = |\mathbf{r}'(t)|dt = \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}dt \implies ds^2 = dx^2 + dy^2 + dz^2$. If the curve C is piecewise smooth, i.e. made up of M smooth curves $C = C_1 \cup C_2 \cup \cdots \cup C_M$, then

$$\int_C f(\mathbf{r})ds = \sum_{i=1}^{M} \int_{C_i} f(\mathbf{r})ds$$

Definition 1.6 (Arc-length function)**:** Define the arc-length function

$$s(t) = \int_a^t |\mathbf{r}'(\tau)|d\tau$$

with $s(a) = 0$, $s(b) = \ell(C)$. The arc-length function is an increasing function of t, $\frac{ds}{dt} = |\mathbf{r}'(t)| \geq 0$. If the parametrization is regular, $\frac{ds}{dt} > 0$, we may invert to get $t = t(s)$.

Definition 1.7 (Line integral)**:** For a vector field $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3$ and a piecewise smooth curve $C : t \mapsto \mathbf{x}(t)$, where $t \in [a, b]$, we define the line integral

$$\int_C \mathbf{F} \cdot d\mathbf{x} = \int_a^b \mathbf{F}(\mathbf{x}(t)) \cdot \mathbf{x}'(t)dt \tag{1.3}$$

Line integrals are independent of parametrization.

Definition 1.8 (Circulation)**:** Line integrals defined along a closed curve is written as

$$\oint_C \mathbf{F} \cdot d\mathbf{x}$$

This is sometimes referred to as the circulation of $\mathbf{F}$ around the loop C.

Definition 1.9 (Exact differential)**:** $\mathbf{F} \cdot d\mathbf{x}$ is exact if it can be written as the differential fo a scalar function, i.e. $\mathbf{F} \cdot d\mathbf{x} = df$ for some scalar function f.

Definition 1.10 (Conservative field)**:** A vector field $\mathbf{F}$ is conservative if it can be written in the form $\mathbf{F} = \boldsymbol{\nabla} f$.

Proposition 1.3: $\mathbf{F} \cdot d\mathbf{x}$ is exact iff $\mathbf{F}$ is conservative.

Corollary 1.2: If $\mathbf{F}$ is conservative, then the line integral between any two fixed points is path independent.

Double and triple integrals

Example 1.1 (Example of double integral)**.** *Let D be a triangle with vertices (0,0), (1,0), (0,1). The integral of $f(x, y) = xy^2$ over D can be found by either integrating over horizontal strips first:*

$$\int_D f dA = \int_0^1 \left(\int_0^{1-y} xy^2 dx \right) dy = \int_0^1 \frac{1}{2}y^2(1-y)^2 dy = \frac{1}{60}$$

or over vertical strips first:

$$\int_D f dA = \int_0^1 \left(\int_0^{1-x} xy^2 dy \right) dx = \int_0^1 \frac{1}{3}x(1-x)^3 dx = \frac{1}{60}$$

Example 1.2 (Example of triple integral)**.** *Consider the region V bound by the plane $x + y + z = 1$ and the three planes $x = 0$, $y = 0$ and $z = 0$. It has volume*

$$\int_V dV = \int_0^1 \int_0^{1-x} \int_0^{1-x-y} dz dy dx = \int_0^1 \int_0^{1-x} 1 - x - y dy dx = \int_0^1 \frac{1}{2}(1-x)^2 dx = \frac{1}{6}$$

Suppose V has constant density $\rho = 1$, then define the centre of mass of an object as

$$\mathbf{X_{CM}} = \frac{1}{M} \int_V \rho \mathbf{x} dV$$

The symmetry dictates that $\mathbf{X_{CM}} = (r, r, r)$ for some $r > 0$.

$$r = \frac{1}{V} \int_V x dV = 6 \int_0^1 \frac{1}{2}x(1-x)^2 dx = \frac{1}{4}$$

hence the centre of mass is at $(0.25, 0.25, 0.25)$.

1.2 Suffix notation

Notation 1.1 (Einstein's summation convention). *Consider a sum $\mathbf{x} \cdot \mathbf{y} = \sum x_i y_i$. The summation convention says that we can drop the $\sum$ symbol and simply write $\mathbf{x} \cdot \mathbf{y} = x_i y_i$. If suffixes are repeated once, summation is understood. The rules of this convention are:*

1. *Suffix appears once in a term: free suffix*

2. *Suffix appears twice in a term: dummy suffix and is summed over*

3. *Suffixes cannot appear three times or more.*

Example 1.3. $[(\mathbf{a} \cdot \mathbf{b})\mathbf{c} - (\mathbf{a} \cdot \mathbf{c})\mathbf{b}]_i = a_j b_j c_i - a_j c_j b_i$ summing over j understood.

Objects with more than one index are called tensors. Here we will define two important tensors:
Definition 1.11 (Kronecker delta)**:**

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{1.4}$$

Example 1.4.

1. $a_i \delta_{i1} = a_1$. In general, $a_i \delta_{ij} = a_j$ (i is dummy, j is free).

2. $\delta_{ij} \delta_{jk} = \delta_{ik}$

3. $\delta_{ii} = n$ if we are in $\mathbb{R}^n$.

4. $a_p \delta_{pq} b_q = a_p b_p$ with p, q both dummy suffices and summed over.

Definition 1.12 (Alternating symbol ε_{ijk})**:** Consider rearrangements of $1, 2, 3$ by swapping two or none of the elements. Even permutations include $(1, 2, 3)$, $(2, 3, 1)$ and $(3, 1, 2)$. The odd permutations are $(2, 1, 3)$, $(1, 3, 2)$ and $(3, 2, 1)$. Define

$$\varepsilon_{ijk} = \begin{cases} +1 & ijk \text{ is even permutation} \\ -1 & ijk \text{ is odd permutation} \\ 0 & \text{otherwise (i.e. repeated suffices)} \end{cases} \tag{1.5}$$

ε_{ijk} has 3 free suffices. We have $\varepsilon_{123} = \varepsilon_{231} = \varepsilon_{312} = +1$ and $\varepsilon_{213} = \varepsilon_{132} = \varepsilon_{321} = -1$. $\varepsilon_{112} = \varepsilon_{111} = \cdots = 0$.

1. $\varepsilon_{ijk} \delta_{jk} = \varepsilon_{ijj} = 0$

2. If $a_{jk} = a_{kj}$ (i.e. a_{ij} is symmetric), then $\varepsilon_{ijk} a_{jk} = \varepsilon_{ijk} a_{kj} = -\varepsilon_{ikj} a_{kj}$. Since $\varepsilon_{ijk} a_{jk} = \varepsilon_{ikj} a_{kj}$ (we simply renamed dummy suffices), we have $\varepsilon_{ijk} a_{jk} = 0$.

Proposition 1.4: $(\mathbf{a} \times \mathbf{b})_i = \varepsilon_{ijk} a_j b_k$

Theorem 1.1:

$$\varepsilon_{ijk} \varepsilon_{ipq} = \delta_{jp} \delta_{kq} - \delta_{jq} \delta_{kp} \tag{1.6}$$

Proof. Proof by exhaustion:

$$\text{RHS} = \begin{cases} +1 & \text{if } j = p \text{ and } k = q \\ -1 & \text{if } j = q \text{ and } k = p \\ 0 & \text{otherwise} \end{cases}$$

LHS: Summing over i, the only non-zero terms are when $j, k \neq i$ and $p, q \neq i$. If $j = p$ and $k = q$, LHS is $(-1)^2$ or $(+1)^2 = 1$. If $j = q$ and $k = p$, LHS is $(+1)(-1)$ or $(-1)(+1) = -1$. All other possibilities result in 0. $\qquad \square$

Proposition 1.5: $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a})$

Proof. In suffix notation, we have $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = a_i(\mathbf{b} \times \mathbf{c})_i = \varepsilon_{ijk}b_j c_k a_i = \varepsilon_{jki}b_j c_k a_i = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a})$.
□

Theorem 1.2 (Vector triple product)**:** $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}$

Proof. $[\mathbf{a} \times (\mathbf{b} \times \mathbf{c})]_i = \varepsilon_{ijk}\varepsilon_{kpq}a_j b_p c_q = (\delta_{ip}\delta_{jq} - \delta_{iq}\delta_{jp})a_j b_p c_q = a_j b_i c_j - a_j c_i b_j = (\mathbf{a} \cdot \mathbf{c})b_i - (\mathbf{a} \cdot \mathbf{b})c_i$
□

Definition 1.13 (Divergence)**:** The divergence of $\mathbf{F}$ is defined by $\boldsymbol{\nabla} \cdot \mathbf{F}$. In Cartesian coordinates,

$$\boldsymbol{\nabla} \cdot \mathbf{F} = \frac{\partial F_i}{\partial x_i} \tag{1.7}$$

Definition 1.14 (Curl)**:** The curl of $\mathbf{F}$ is defined by $\boldsymbol{\nabla} \times \mathbf{F}$. In Cartesian coordinates,

$$[\boldsymbol{\nabla} \times \mathbf{F}]_i = \epsilon_{ijk}\frac{\partial F_k}{\partial x_j} \tag{1.8}$$

Definition 1.15 (Laplacian)**:** For a scalar field $f : \mathbb{R}^3 \mathbb{R}$, we define the Laplacian of f to be $\boldsymbol{\nabla} \cdot \boldsymbol{\nabla} f$. In Cartesian coordinates,

$$\boldsymbol{\nabla} \cdot \boldsymbol{\nabla} f = \frac{\partial^2 f}{\partial x_i \partial x_i} \tag{1.9}$$

Notation 1.2. *The Laplacian of a scalar function is written as* $\nabla^2 f$. *Note that* $\frac{\partial^2 f}{\partial x_i \partial x_i} \neq \frac{\partial^2 f}{\partial x_i^2}$.

Proposition 1.6: The Laplacian of a vector field $\mathbf{F}$ is

$$\nabla^2 \mathbf{F} = \boldsymbol{\nabla}(\boldsymbol{\nabla} \cdot \mathbf{F}) - \boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \mathbf{F}) \tag{1.10}$$

Proof. Evaluate $\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \mathbf{F})$:

$$\epsilon_{ijk}\partial_i\epsilon_{pqj}\partial_p F_q = (\delta_{kp}\delta_{iq} - \delta_{kq}\delta_{ip})\partial_i\partial_p F_q = \partial_i\partial_k F_i - \partial_i\partial_i F_k$$

Hence, $\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \mathbf{F}) = \boldsymbol{\nabla}(\boldsymbol{\nabla} \cdot \mathbf{F}) - \nabla^2 \mathbf{F}$.
□

Proposition 1.7: For scalar functions f, g and vector fields $\mathbf{F}, \mathbf{G}$, we have

1. $\boldsymbol{\nabla}(fg) = (\boldsymbol{\nabla}f)g + (\boldsymbol{\nabla}g)f$;

2. $\boldsymbol{\nabla} \cdot (f\mathbf{F}) = (\boldsymbol{\nabla}f) \cdot \mathbf{F} + f(\boldsymbol{\nabla} \cdot \mathbf{F})$;

3. $\boldsymbol{\nabla} \times (f\mathbf{F}) = (\boldsymbol{\nabla}f) \times \mathbf{F} + f(\boldsymbol{\nabla} \times \mathbf{F})$;

4. $\boldsymbol{\nabla}(\mathbf{F} \cdot \mathbf{G}) = \mathbf{F} \times (\boldsymbol{\nabla} \times \mathbf{G}) + \mathbf{G} \times (\boldsymbol{\nabla} \times \mathbf{F}) + (\mathbf{F} \cdot \boldsymbol{\nabla})\mathbf{G} + (\mathbf{G} \cdot \boldsymbol{\nabla})\mathbf{F}$;

5. $\boldsymbol{\nabla} \times (\mathbf{F} \times \mathbf{G}) = \mathbf{F}(\boldsymbol{\nabla} \cdot \mathbf{G}) - \mathbf{G}(\boldsymbol{\nabla} \cdot \mathbf{F}) + (\mathbf{G} \cdot \boldsymbol{\nabla})\mathbf{F} - (\mathbf{F} \cdot \boldsymbol{\nabla})\mathbf{G}$;

6. $\boldsymbol{\nabla} \cdot (\mathbf{F} \times \mathbf{G}) = (\boldsymbol{\nabla} \times \mathbf{F}) \cdot \mathbf{G} - \mathbf{F} \cdot (\boldsymbol{\nabla} \times \mathbf{G})$.

Proof. Use suffix notation and the identity from Theorem 1.1.

1. $\partial_i(fg) = (\partial_i f)g + (\partial_i g)f$;

2. $\partial_i(fF_i) = (\partial_i f)F_i + f(\partial_i F_i)$;

3. $\epsilon_{ijk}\partial_i(fF_j) = \epsilon_{ijk}(\partial_i f)F_j + \epsilon_{ijk}f(\partial_i F_j)$;

4. Start from $\mathbf{G} \times (\boldsymbol{\nabla} \times \mathbf{F}) + \mathbf{F} \times (\boldsymbol{\nabla} \times \mathbf{G})$: $F_i\partial_k G_i + G_i\partial_k F_i - G_i\partial_i F_k - F_i\partial_i G_k$.

5. $\epsilon_{ijk}\partial_i\epsilon_{pqj}F_p G_q = (\delta_{kp}\delta_{iq} - \delta_{kq}\delta_{ip})(F_p\partial_i G_q + G_q\partial_i F_p) = F_k\partial_i G_i + G_i\partial_i F_k - F_i\partial_i G_k - G_k\partial_i F_i$;

6. $\partial_k\epsilon_{ijk}F_i G_j = \epsilon_{ijk}G_j\partial_k F_i + \epsilon_{ijk}F_i\partial_k G_j = (\epsilon_{kij}\partial_k F_i)G_j - (\epsilon_{kji}\partial_k G_j)F_i$

□

Definition 1.16 (Irrotational, solenoidal, conservative): Let $\mathbf{F}$ be a vector field, then it is

- conservative if $\exists$ a scalar function f s.t. $\mathbf{F} = \boldsymbol{\nabla} f$;

- solenoidal if $\boldsymbol{\nabla} \cdot \mathbf{F} = 0$;

- irrotational if $\boldsymbol{\nabla} \times \mathbf{F} = \mathbf{0}$.

Proposition 1.8: For scalar field f, vector field $\mathbf{F}$,

$$\boldsymbol{\nabla} \times \boldsymbol{\nabla} f = \mathbf{0}, \quad \boldsymbol{\nabla} \cdot (\boldsymbol{\nabla} \times \mathbf{F}) = 0 \tag{1.11}$$

Proof. We will show this only for Cartesian coordinates. $[\epsilon_{ijk}\partial_i\partial_j f]_k = 0$, $\epsilon_{ijk}\partial_i\partial_j F_k = 0$. They evaluate to zero since the partial derivatives are symmetric. $\qquad\square$

Theorem 1.3:

1. If $\mathbf{F}$ is conservative, then it is always irrotational.

2. If $\mathbf{F} = \boldsymbol{\nabla} \times \mathbf{A}$ (where $\mathbf{A}$ is the vector potential), then it is solenoidal.

Proof. For 1, from the definition of conservative field, $\mathbf{F} = \boldsymbol{\nabla} f$. f satisfies Proposition 1.8. For 2, $\mathbf{A}$ satisfies Proposition 1.8. Note that f and $\mathbf{A}$ are not unique. $\qquad\square$

1.3 Orthogonal curvilinear coordinates

Definition 1.17 (Orthogonal curvilinear coordinates and scale factors): (u, v, w) are a set of orthogonal curvilinear coordinates if the vectors

$$\mathbf{e_u} = \frac{\partial_u \mathbf{r}}{|\partial_u \mathbf{r}|}, \quad \mathbf{e_v} = \frac{\partial_v \mathbf{r}}{|\partial_u \mathbf{r}|}, \quad \mathbf{e_w} = \frac{\partial_w \mathbf{r}}{|\partial_w \mathbf{r}|} \tag{1.12}$$

form a right-handed, orthonormal basis, i.e. $\mathbf{e_u} \times \mathbf{e_v} = \mathbf{e_w}$. The scale factors would be say, $h_u = |\partial_u \mathbf{x}|$, then the line element is

$$d\mathbf{r} = \frac{\partial \mathbf{r}}{\partial u} du + \frac{\partial \mathbf{r}}{\partial v} dv + \frac{\partial \mathbf{r}}{\partial w} dw = h_u du\,\mathbf{e_u} + h_v dv\,\mathbf{e_v} + h_w dw\,\mathbf{e_w} \tag{1.13}$$

Example 1.5 (Cylindrical polar coordinates). *The cylindrical polar coordinates (ρ, ϕ, z) are defined by*

$$\mathbf{r}(\rho, \phi, z) = \rho\cos\phi\,\mathbf{e_x} + \rho\sin\phi\,\mathbf{e_y} + z\,\mathbf{e_z}, \quad 0 \le \rho < \infty, \quad 0 \le \phi < 2\pi, \quad -\infty < z < \infty$$

These are orthogonal curvilinear coordinates with

$$\boldsymbol{e_\rho} = \cos\phi\,\mathbf{e_x} + \sin\phi\,\mathbf{e_y}, \quad \boldsymbol{e_\phi} = -\sin\phi\,\mathbf{e_x} + \cos\phi\,\mathbf{e_y}, \quad \mathbf{e_z}$$

The scale factors are $h_\rho = 1$, $h_\phi = \rho$ and $h_z = 1$. The corresponding line element

$$d\mathbf{r} = d\rho\,\boldsymbol{e_\rho} + \rho d\phi\,\boldsymbol{e_\phi} + dz\,\mathbf{e_z}$$

Change in the coordinate $\phi \mapsto \phi + \delta\phi$ produces a first order change $\mathbf{r} \mapsto \mathbf{r} + \rho\delta\phi\,\boldsymbol{e_\phi}$.

Example 1.6 (Spherical polar coordinates). *The spherical polar coordinates (r, θ, ϕ) are defined by*

$$\mathbf{r}(r, \theta, \phi) = r\cos\phi\sin\theta\,\mathbf{e_x} + r\sin\phi\sin\theta\,\mathbf{e_y} + r\cos\theta\,\mathbf{e_z}, \quad 0 \le r < \infty, \quad 0 \le \theta \le \pi, \quad 0 \le \phi < 2\pi$$

These are orthogonal curvilinear coordinates with

$$\boldsymbol{e_r} = \cos\phi\sin\theta\,\mathbf{e_x} + \sin\phi\sin\theta\,\mathbf{e_y} + \cos\theta\,\mathbf{e_z}, \quad \boldsymbol{e_\theta} = \cos\phi\cos\theta\,\mathbf{e_x} + \sin\phi\cos\theta\,\mathbf{e_y} - \sin\theta\,\mathbf{e_z},$$

$$\boldsymbol{e_\phi} = -\sin\phi\,\mathbf{e_x} + \cos\phi\,\mathbf{e_y}$$

The scale factors are $h_r = 1$, $h_\theta = r$ and $h_\phi = r\sin\theta$. The corresponding line element

$$d\mathbf{r} = dr\,\boldsymbol{e_r} + rd\theta\,\boldsymbol{e_\theta} + r\sin\theta d\phi\,\mathbf{e_\phi}$$

Change in the coordinate $\phi \mapsto \phi + \delta\phi$ produces a first order change $\mathbf{r} \mapsto \mathbf{r} + r\sin\theta\delta\phi\,\boldsymbol{e_\phi}$.

Gradients

Recall in Cartesian coordinates in $\mathbb{R}^3$,

$$f(\mathbf{x} + \delta\mathbf{x}) = f(\mathbf{x}) + \frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial y}\delta y + \frac{\partial f}{\partial z}\delta z + o(\delta\mathbf{x})$$

Then with the definition of gradient in suffix notation, $\boldsymbol{\nabla} f = \mathbf{e_i}\frac{\partial f}{\partial x_i}$, hence the differential is $df = \frac{\partial f}{\partial x_i}dx_i = \boldsymbol{\nabla} f \cdot d\mathbf{x}$.

Proposition 1.9: If (u, v, w) is a set of orthogonal curvilinear coordinates and $f = f(u, v, w)$

$$\boldsymbol{\nabla} f = \frac{1}{h_u}\frac{\partial f}{\partial u}\mathbf{e_u} + \frac{1}{h_v}\frac{\partial f}{\partial v}\mathbf{e_v} + \frac{1}{h_w}\frac{\partial f}{\partial w}\mathbf{e_w} \tag{1.14}$$

Proof. If $f = f(u, v, w)$ and $\mathbf{x} = \mathbf{x}(u, v, w)$, then the differential and line element are respectively

$$df = \frac{\partial f}{\partial u}du + \frac{\partial f}{\partial v}dv + \frac{\partial f}{\partial w}dw, \quad d\mathbf{x} = h_u du\mathbf{e_u} + h_v dv\mathbf{e_v} + h_w dw\mathbf{e_w}$$

where $h_u = |\partial_u\mathbf{x}|$, etc. By writing $\boldsymbol{\nabla} f = (\boldsymbol{\nabla} f)_u\mathbf{e_u} + (\boldsymbol{\nabla} f)_v\mathbf{e_v} + (\boldsymbol{\nabla} f)_w\mathbf{e_w}$ and substituting everything in, using the fact that $\{\mathbf{e_u}, \mathbf{e_v}, \mathbf{e_w}\}$ are orthonormal,

$$\frac{\partial f}{\partial u}du + \frac{\partial f}{\partial v}dv + \frac{\partial f}{\partial w}dw = h_u(\boldsymbol{\nabla} f)_u du + h_v(\boldsymbol{\nabla} f)_v dv + h_w(\boldsymbol{\nabla} f)_w dw$$

Since the differentials $\{du, dv, dw\}$ are linearly independent, then we obtain our required result, e.g. $(\boldsymbol{\nabla} f)_u = \frac{1}{h_u}\frac{\partial f}{\partial u}$. □

Example 1.7. *In cylindrical polar coordinates: if $f = f(\rho, \phi, z)$,*

$$\boldsymbol{\nabla} f = \frac{\partial f}{\partial \rho}\boldsymbol{e_\rho} + \frac{1}{\rho}\frac{\partial f}{\partial \phi}\boldsymbol{e_\phi} + \frac{\partial f}{\partial z}\mathbf{e_z}$$

In spherical polar coordinates: if $f = f(r, \theta, \phi)$,

$$\boldsymbol{\nabla} f = \frac{\partial f}{\partial r}\mathbf{e_r} + \frac{1}{r}\frac{\partial f}{\partial \theta}\boldsymbol{e_\theta} + \frac{1}{r\sin\theta}\frac{\partial f}{\partial \phi}\boldsymbol{e_\phi}$$

Proposition 1.10: In orthogonal curvilinear coordinates, $\boldsymbol{\nabla} \cdot \mathbf{F}$, $\boldsymbol{\nabla} \times \mathbf{F}$ and ∇^2 are defined respectively as

$$\frac{1}{h_u h_v h_w}\left[\frac{\partial}{\partial u}(h_v h_w F_u) + \frac{\partial}{\partial v}(h_w h_u F_v) + \frac{\partial}{\partial w}(h_u h_v F_w)\right] \tag{1.15}$$

$$\frac{1}{h_v h_w}\left[\frac{\partial}{\partial v}(h_w F_w) - \frac{\partial}{\partial w}(h_v F_v)\right]\mathbf{e_u} + \frac{1}{h_u h_w}\left[\frac{\partial}{\partial w}(h_u F_u) - \frac{\partial}{\partial u}(h_w F_w)\right]\mathbf{e_v} + \frac{1}{h_u h_v}\left[\frac{\partial}{\partial u}(h_v F_v) - \frac{\partial}{\partial v}(h_u F_u)\right]\mathbf{e_w} \tag{1.16}$$

$$\nabla^2 = \frac{1}{h_u h_v h_w}\left[\frac{\partial}{\partial u}\left(\frac{h_v h_w}{h_u}\frac{\partial}{\partial u}\right) + \frac{\partial}{\partial v}\left(\frac{h_u h_w}{h_v}\frac{\partial}{\partial v}\right) + \frac{\partial}{\partial w}\left(\frac{h_v h_u}{h_w}\frac{\partial}{\partial w}\right)\right] \tag{1.17}$$

Jacobians

Proposition 1.11: Let $x = x(u, v)$ and $y = y(u, v)$ be a differentiable bijection with differentiable inverse that sends E in the (u, v)-plane to D in the (x, y)-plane in a one-to-one fashion. Then

$$\int\int_D f(x, y)dxdy = \int\int_E f(x(u, v), y(u, v))\left|\frac{\partial(x, y)}{\partial(u, v)}\right|dudv \tag{1.18a}$$

where

$$\frac{\partial(x, y)}{\partial(u, v)} = \det\begin{bmatrix}\frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v}\end{bmatrix} \tag{1.18b}$$

This is often denoted as the Jacobian, J, s.t. $dxdy = |J|dudv$.

Example 1.8. *Consider the transformation to polar coordinates (ρ, ϕ) defined by $x = \rho \cos \phi$ and $y = \rho \sin \phi$, then the Jacobian is*

$$|J| = \left| \det \begin{bmatrix} \cos \phi & -\rho \sin \phi \\ \sin \phi & \rho \cos \phi \end{bmatrix} \right| = \rho$$

Now let D be the region in the (x, y)-plane defined by

$$D = \{(x, y) : x > 0, y > 0, x^2 + y^2 < R^2\}$$

This region is mapped in a one-to-one fashion in the region E in the (ρ, ϕ)-plane defined by

$$E = \{(\rho, \phi) : 0 < \rho < R, 0 < \phi < \frac{\pi}{2}\}$$

By Proposition 3.5, $\int_D f(x, y) dx dy = \int_E f(\rho \cos \phi, \rho \sin \phi) \rho d\rho d\phi$. Let's consider the integral

$$I = \int_0^\infty e^{-x^2} dx$$

Take the previous result with the limit $R \to \infty$, then

$$I^2 = \int_0^\infty e^{-x^2} dx \int_0^\infty e^{-y^2} dy = \int_0^{\pi/2} \int_0^\infty e^{-\rho^2} \rho d\rho d\phi = \frac{\pi}{2} \int_0^\infty \frac{d}{d\rho} - 0.5 e^{-\rho^2} d\rho = \frac{\pi}{4}$$

Taking the positive square root, $I = \sqrt{\pi}/2$.

Proposition 1.12: Let $x = x(u, v, w)$ and $y = y(u, v, w)$ and $z = z(u, v, w)$ be a smooth bijection with smooth inverse that sends volume U in the (u, v, w)-plane to volume V in the (x, y, z)-plane in a one-to-one fashion. Then

$$\int \int \int_V f(x, y, z) dx dy dz = \int \int \int_U f(x(u, v, w), y(u, v, w), z(u, v, w)) \left| \frac{\partial(x, y, z)}{\partial(u, v, w)} \right| du dv dw \tag{1.19a}$$

where

$$\frac{\partial(x, y, z)}{\partial(u, v, w)} = \det \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{bmatrix} \tag{1.19b}$$

This is still denoted as the Jacobian, J, s.t. $dx dy dz = |J| du dv dw$.

Example 1.9. *Consider a sphere of radius a with a cylinder of radius $b < a$ removed. The resulting volume is*

$$V = \{(x, y, z) : 0 \le x^2 + y^2 + z^2 \le a^2, x^2 + y^2 \ge b^2\} = \{(\rho, \phi, z) : b \le \rho \le a, 0 \le z^2 + \rho^2 \le a^2, 0 \le \phi < 2\pi\}$$

The volume is

$$\int_V dV = \int_a^b \int_0^{2\pi} \int_{-\sqrt{a^2 - \rho^2}}^{\sqrt{a^2 - \rho^2}} \rho dz d\phi d\rho = 2\pi \int_b^a 2\rho \sqrt{a^2 - \rho^2} d\rho = \frac{4\pi}{3} (a^2 - b^2)^{3/2}$$

1.4 Surface integrals and integral theorems

Surface integrals

Definition 1.18 (Surface): A two-dimensional surface in $\mathbb{R}^3$ can be defined implicitly using a function $f : \mathbb{R}^3 \to \mathbb{R}$ via

$$S = \{\mathbf{x} : f(\mathbf{x}) = 0\}$$

Remarks 1.2.

- *From Corollary 1.1, any normal to the surface at $\mathbf{x}$ must be parallel to $\boldsymbol{\nabla} f(\mathbf{x})$;*

- *A surface S is regular if $\boldsymbol{\nabla} f \neq \mathbf{0}$ for $\mathbf{x} \in S$;*

- *It is useful to parametrize a surface using some coordinates, say (u, v):*

$$S = \{\mathbf{x} = \mathbf{x}(u, v), (u, v) \in D\}$$

for some region D in the (u, v)-plane. Again, this surface is regular if

$$\frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v} \neq 0$$

s.t. the normal is defined to be

$$\mathbf{n} = \frac{\frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v}}{|\frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v}|} \tag{1.20}$$

We fix the convention that normal vectors in your immediate vicinity should be on your left as you traverse (each part) of the boundary of the surface S.

Definition 1.19 (Boundary and closed surfaces): A surface S will either have no boundary ($\partial S = \varnothing$) or it have a boundary made of a collection of piecewise smooth curves ($\partial S = \cup_i C_i$). Surfaces with a boundary are called closed.

Example 1.10. *The hemisphere can be parametrized as*

$$
\begin{aligned}
S &= \{(x, y, z) : x^2 + y^2 + z^2 = 1, z \geq 0\} \\
&= \{\mathbf{x}(\theta, \phi) = \sin\theta\cos\phi\mathbf{e_x} + \sin\theta\sin\phi\mathbf{e_y} + \cos\theta\mathbf{e_z}, \quad 0 \leq \theta \leq \frac{\pi}{2}, \quad 0 \leq \phi < 2\pi\}
\end{aligned}
$$

with a boundary ∂S

$$\partial S = \{(x, y, z) : x^2 + y^2 + 1, \quad z = 0\}$$

Unfortunately, a small change $(u, v) \mapsto (u + \delta u, v + \delta v)$ does not necessarily produce an area of $\delta u \delta v$ on S, so the surface integral is not $\int_D du dv$!

Proposition 1.13: The surface area of an arbitrary surface S is

$$\int_S dS = \int \int_D \left| \frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v} \right| du dv \tag{1.21a}$$

Proof. To leading order, the change in $(u, v) \mapsto (u + \delta u, v + \delta v)$ produces a parallelogram generated by $\frac{\partial \mathbf{x}}{\partial u}\delta u$ and $\frac{\partial \mathbf{x}}{\partial v}\delta v$ which has area

$$\left| \frac{\partial \mathbf{x}}{\partial u}\delta u \times \frac{\partial \mathbf{x}}{\partial v}\delta v \right| = \left| \frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v} \right| \delta u \delta v$$

This leads us to define the scalar area element and vector area element by

$$dS = \left| \frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v} \right| du dv, \quad d\mathbf{S} = \frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v} du dv = \mathbf{n} dS \tag{1.21b}$$

The surface area of S is $\int_S dS$ as desired. $\qquad\square$

Example 1.11. *Consider the hemisphere S with radius R, then*

$$dS = \left|\frac{\partial \mathbf{x}}{\partial \theta} \times \frac{\partial \mathbf{x}}{\partial \phi}\right| d\theta d\phi = |R\boldsymbol{e_\theta} \times R\sin\theta\boldsymbol{e_\phi}| = R^2\sin\theta d\theta d\phi \implies \int_S dS = \int_0^{pi/2} d\theta \int_0^{2\pi} R^2\sin\theta d\phi = 2\pi R^2$$

Example 1.12 (Flux integrals)**.** *Suppose we model a fluid flow with the vector field $\mathbf{u} = \mathbf{u}(\mathbf{x})$. Given a surface S, how do we calculate how much fluid passes through S per unit time? On a small patch δS on S, the volume of fluid crossing it in time δt is $(\mathbf{u} \cdot \delta\mathbf{S})\delta t$. So total amount of fluid that passes over S in time δt is*

$$\delta t \int_S \mathbf{u} \cdot d\mathbf{S}$$

So the integral gives the rate at which fluid is passing through S.

Proposition 1.14: Surface integrals are independent of parametrization.

Integral theorems

Theorem 1.4 (Green's theorem)**:** If $P = P(x, y)$ and $Q = Q(x, y)$ are continuously differentiable functions on $A \subset \mathbb{R}^2$ and ∂A a collection of piecewise smooth curves, then

$$\oint_{\partial A} Pdx + Qdy = \int\int_A \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dxdy \tag{1.22}$$

The orientation of the boundary ∂_A is s.t. A lies to your left as you traverse it.

Example 1.13. *Let $P = -\frac{1}{2}y$, $Q = \frac{1}{2}x$, then the Green's theorem tells us*

$$\int\int_A dxdy = \frac{1}{2}\oint_{\partial A} xdy - ydx$$

Let A be the ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$, so that C has parametrization $t \mapsto (a\cos t, b\sin t)^T$ where $t \in [0, 2\pi]$. The area is

$$\frac{1}{2}\int_0^{2\pi} (ab\cos^2 t + ab\sin^2 t)dt = \pi ab$$

Theorem 1.5 (Stokes' Theorem)**:** Let $\mathbf{F} = \mathbf{F}(\mathbf{x})$ be a continuously differentiable vector field, and let surface S be orientable, piecewise regular with piecewise smooth boundary ∂S, then

$$\int_S (\boldsymbol{\nabla} \times \mathbf{F}) \cdot d\mathbf{S} = \oint_{\partial S} \mathbf{F} \cdot d\mathbf{x} \tag{1.23}$$

The orientable bit is important: it means that there's a consistent choice of unit normal $\mathbf{n}$ on S that varies continuously from point to point.

Example 1.14. *Let S be a section of the sphere defined in spherical coordinates by*

$$S = \{\mathbf{x}(\theta, \phi) = \mathbf{e_r} : 0 \leq \theta \leq \alpha, \quad 0 \leq \phi < 2\pi\}$$

Let $\mathbf{F} = (-x^2y, 0, 0)^T \implies \boldsymbol{\nabla} \times \mathbf{F} = (0, 0, x^2)^T$. Then

$$d\mathbf{S} = \frac{\partial \mathbf{x}}{\partial \theta} \times \frac{\partial \mathbf{x}}{\partial \phi} d\theta d\phi = \boldsymbol{e_\theta} \times (\sin\theta\boldsymbol{e_\phi})d\theta d\phi = \mathbf{e_r}\sin\theta d\theta d\phi$$

The surface integral is

$$(\boldsymbol{\nabla} \times \mathbf{F}) \cdot d\mathbf{S} = x^2(\mathbf{e_z} \cdot \mathbf{e_r})\sin\theta d\theta d\phi \implies \int_0^{2\pi}\cos^2\phi d\phi \int_0^\alpha \sin^2\theta\cos\theta\sin\theta d\theta = \frac{\pi}{4}\sin^4\alpha$$

The boundary is described by the curve

$$t \mapsto \sin\alpha\cos t\mathbf{e_x} + \sin\alpha\sin t\mathbf{e_y} + \cos\alpha\mathbf{e_z}$$

where $t \in [0, 2\pi]$, then

$$\oint_C \mathbf{F} \cdot d\mathbf{x} = -\int_0^{2\pi}\sin^2\alpha\cos^2 t\sin\alpha\sin t\sin\alpha(-\sin t)dt = \frac{\pi}{4}\sin^4\alpha$$

Proposition 1.15: If $\mathbf{F} = \mathbf{F}(\mathbf{x})$ is a continuously differentiable vector field and for every simple loop C

$$\oint_C \mathbf{F} \cdot d\mathbf{x} = 0$$

then $\boldsymbol{\nabla} \times \mathbf{F} = \mathbf{0}$.

Theorem 1.6 (Divergence theorem in $\mathbb{R}^3$)**:** If $\mathbf{F} = \mathbf{F}(\mathbf{x})$ be a continuously differentiable vector field and V is a volume with a piecewise regular boundary ∂V, then

$$\int_V \boldsymbol{\nabla} \cdot \mathbf{F} dV = \int_{\partial V} \mathbf{F} \cdot d\mathbf{S} \tag{1.24}$$

where the normal to ∂V points outwards from V.

Theorem 1.7 (Divergence theorem in $\mathbb{R}^2$)**:** If $\mathbf{F} = \mathbf{F}(\mathbf{x})$ be a continuously differentiable vector field and $D \subset \mathbb{R}^2$ a region with piecewise smooth boundary ∂D, then

$$\int_D \boldsymbol{\nabla} \cdot \mathbf{F} dA = \oint_{\partial D} \mathbf{F} \cdot d\mathbf{n} ds \tag{1.25}$$

where the normal to ∂D points outwards from D.

Example 1.15. *Let V be a cylinder, defined using cylindrical polars (ρ, ϕ, z) by*

$$V = \{(\rho, \phi, z) : 0 \le \rho \le R, \quad -h \le z \le h, \quad 0 \le \phi < 2\pi\}$$

Consider the vector field $\mathbf{F} = \mathbf{x}$, then $\int_V \boldsymbol{\nabla} \cdot \mathbf{F} dV = 3\pi R^2 h 2 = 6\pi h R^2$. Alternatively, we can use the divergence theorem. The boundary ∂V is made of three surfaces

$$S_R = \{(\rho, \phi, z) : \rho = R, \quad -h \le z \le h, \quad 0 \le \phi < 2\pi\}$$

$$S_\pm = \{(\rho, \phi, z) : 0 \le \rho \le R, \quad z = \pm h, \quad 0 \le \phi < 2\pi\}$$

On S_R, we have $d\mathbf{S} = \boldsymbol{e_\rho} R d\phi dz$ and $\mathbf{x} \cdot \boldsymbol{e_\rho} = R$, so

$$\int_{S_R} \mathbf{F} \cdot d\mathbf{S} = \int_{-h}^{h} \int_0^{2\pi} R^2 d\phi dz = 4\pi R^2 h$$

On $S_\pm$, we have $d\mathbf{S} = \pm \mathbf{e_z} \rho d\rho d\phi$ and $\mathbf{e_z} \cdot \mathbf{x} = \pm h$, so

$$\int_{S_\pm} \mathbf{F} \cdot d\mathbf{S} = \int_0^{2\pi} \int_0^{R} h\rho d\rho d\phi = \pi R^2 h$$

Then as expected,

$$\int_{\partial V} \mathbf{F} \cdot d\mathbf{S} = \int_S \mathbf{F} \cdot d\mathbf{S} + \int_{S_+} \mathbf{F} \cdot d\mathbf{S} + \int_{S_-} \mathbf{F} \cdot d\mathbf{S} = 6\pi R^2 h$$

Proposition 1.16: If $\mathbf{F}$ is continuously differentiable and for every closed surface S,

$$\int_S \mathbf{F} \cdot d\mathbf{S} = 0$$

then $\boldsymbol{\nabla} \cdot \mathbf{F} = 0$.

2 Partial differential equations

Linear second-order partial differential equations; physical examples of occurrence, the method of separation of variables (Cartesian coordinates only) **[2]**

2.1 Recap

Fourier series

Definition 2.1 (Periodic function)**:** A function $f(x)$ is periodic if $f(x + T) = f(x) \; \forall x$, where T is the period.

Definition 2.2 (Inner product)**:** The inner product $\langle f, g \rangle$ for two functions (w.r.t unit weight) in the interval $[\alpha, \beta]$ is

$$\langle f, g \rangle = \int_\alpha^\beta f^*(x) g(x) dx$$

Proposition 2.1: The set of functions

$$\left\{ g_n(x) = \cos \frac{n\pi x}{L}, \quad h_n(x) = \sin \frac{n\pi x}{L} \right\}_{0 \le n < \infty}$$

which are periodic on $0 \le x < 2L$ (period $T = 2L$), are mutually orthogonal, i.e.

$$\langle h_n, h_m \rangle = \int_0^{2L} \sin \frac{n\pi x}{L} \sin \frac{m\pi x}{L} dx = \left\{ \begin{array}{ll} L\delta_{nm} & m \ne 0 \\ 0 & m = 0 \end{array} \right. \tag{2.1}$$

$$\langle g_n, g_m \rangle = \int_0^{2L} \cos \frac{n\pi x}{L} \cos \frac{m\pi x}{L} dx = \left\{ \begin{array}{ll} L\delta_{nm} & m \ne 0 \\ 2L\delta_{nm} & m = 0 \end{array} \right. \tag{2.2}$$

$$\langle h_n, g_m \rangle = 0, \quad \forall n, m \tag{2.3}$$

Corollary 2.1: Integrate odd power of trigonometric functions over integer number of cycles gives zero.

Definition 2.3 (Fourier series)**:** We can express any 'well-behaved' periodic function $f(x)$ with period $2L$ as

$$f(x) = \frac{1}{2} a_0 + \sum_{n=1}^\infty a_n \cos \frac{n\pi x}{L} + b_n \sin \frac{n\pi x}{L} \tag{2.4}$$

where a_n, b_n are constants s.t. RHS is convergent $\forall x$ where $f(x)$ is continuous. At a discontinuity, the Fourier series (1.4) approaches the midpoint $\frac{1}{2}(f(x_+) + f(x_-))$.

Proposition 2.2: The Fourier coefficients are

$$a_n = \frac{1}{L} \int_0^{2L} f(x) \cos \frac{n\pi x}{L} dx \tag{2.5a}$$

$$b_n = \frac{1}{L} \int_0^{2L} f(x) \sin \frac{n\pi x}{L} dx \tag{2.5b}$$

Proof. Integrate (2.4) $\times \sin \frac{m\pi x}{L}$ for some m:

$$\int_0^{2L} \sin \frac{m\pi x}{L} f(x) dx = \sum_{n=1}^\infty L b_n \delta_{mn} = L b_m$$

by orthogonality relations (2.1,2.2,2.3). For a_n, we use $\cos \frac{m\pi x}{L}$ instead. $\square$

Proposition 2.3 (Complex Fourier series)**:** The complex representation of a Fourier series is

$$f(x) = \sum_{n=-\infty}^\infty c_n e^{in\pi x/L} \tag{2.6a}$$

where for $n > 0$, we have $c_0 = \frac{1}{2} a_0$, $c_{\pm n} = \frac{1}{2}(a_n \mp i b_n)$, or equivalently,

$$c_n = \frac{1}{2L} \int_{-L}^L f(x) e^{-in\pi x/L} dx \tag{2.6b}$$

Proposition 2.4 (Half-range series)**:** Consider $f(x)$ defined only on $0 \leq x < L$. We extend its range over $-L \leq x < L$ in two simple ways:

- if f is required to be odd, i.e. $f(-x) = -f(x)$ with period $2L$, then

$$a_n = 0, \quad b_n = \frac{2}{L} \int_0^L f(x) \sin \frac{n\pi x}{L} dx \tag{2.7a}$$

  and the Fourier sine series will be

$$f(x) = \sum_{n=1}^{\infty} b_n \sin \frac{n\pi x}{L} \tag{2.7b}$$

- if f is required to be even, i.e $f(-x) = f(x)$ with period $2L$, then

$$b_n = 0, \quad a_n = \frac{2}{L} \int_0^L f(x) \cos \frac{n\pi x}{L} dx \tag{2.8a}$$

  and the Fourier cosine series will be

$$f(x) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n \cos \frac{n\pi x}{L} \tag{2.8b}$$

Linear PDEs and superposition

Definition 2.4 (Homogeneous differential equation)**:** A differential equation is homogeneous if $y = 0$ is a solution. Inhomogeneous otherwise.

Definition 2.5 (Homogeneous boundary conditions)**:** The boundary conditions are homogeneous if they are zero. Inhomogeneous otherwise.

Proposition 2.5: If y_1 and y_2 are separately solutions for a linear PDE, then $c_1 y_1 + c_2 y_2$ is also a solution of the PDE for some constants c_1, c_2.

2.2 Separation of variables in Cartesian

Definition 2.6 (Separation of variables)**:** For a generic partial differential equation (PDE) of variables $x_1, \ldots x_n$, with homogeneous boundary conditions, the method of separation of variables asserts that the solution of the PDE $y(x_1, \ldots, x_n)$ has the form

$$y(x_1, \ldots, x_n) = \prod_{i=1}^{n} X_i(x_i) \tag{2.9}$$

The procedure steps:

1. Separate the variables in the ansatz.

2. Determine the admissible form for eigenvalues and associated eigenvectors from b.c.s.

3. Determine the form of the other separated function using the eigenvalues, and hence find a particular solution.

4. Sum across all possible particular solutions to form a general series representation.

5. Determine the coefficients in the series representation from the i.c.s.

Example 2.1. *Consider the one-dimensional wave equation:*

$$\frac{1}{c^2}\frac{\partial^2 y}{\partial t^2} = \frac{\partial^2 y}{\partial x^2}, \quad y(0,t) = y(L,t) = 0, \quad y(x,0) = p(x), \quad \frac{\partial y}{\partial t}(x,0) = q(x)$$

We use separation of variables to solve it, i.e. $y(x,t) = X(x)T(t)$, we obtain

$$\frac{1}{c^2}\frac{\ddot{T}}{T} = \frac{X''}{X}$$

But LHS is a function of t only, while RHS is a function of x only, so both sides must be equal to a constant, say $-\lambda$, then we obtain two independent ordinary differential equations

$$X'' + \lambda X = 0, \quad \ddot{T} + \lambda c^2 T = 0$$

There are three possibilities for λ in the spatial ODE, but bear in mind that we are restricted by the b.c.s.

- *$\lambda < 0$: Take $\chi^2 := -\lambda$, then $X(x) = Ae^{\chi x} + Be^{-\chi x} = \tilde{A}\cosh(\chi x) + \tilde{B}\sinh(\chi x)$. After imposing b.c.s which is also $X(0) = X(L) = 0$, then $\tilde{A} = \tilde{B} = 0$;*

- *$\lambda = 0$: Then $X(x) = Ax + B$, which is obvious that $A = B = 0$ by b.c.s;*

- *$\lambda > 0$, then $X(x) = A\cos\sqrt{\lambda}x + B\sin\sqrt{\lambda}x$ and imposing b.c.s, $A = 0$ and $B\sin\sqrt{\lambda}L = 0 \implies \sqrt{\lambda}L = n\pi$, hence*

$$X_n(x) = B_n \sin\frac{n\pi x}{L}, \quad \lambda_n = \frac{n^2\pi^2}{L^2}, \quad n > 0$$

  *are the eigenfunctions and eigenvalues of the system. These are **normal modes** because the spatial shapes in x does not change in time t. The $n = 1$ mode is called the fundamental mode, while the $n > 1$ modes are called overtones, with the first overtone being $n = 2$.*

Substitute eigenvalues $\lambda_n = \frac{n^2\pi^2}{L^2}$ into temporal ODE,

$$\ddot{T} + \frac{n^2\pi^2 c^2}{L^2}T = 0 \implies T_n(t) = C_n \cos\frac{n\pi c}{L}t + D_n \sin\frac{n\pi c}{L}t$$

Thus a specific solution to the wave equation satisfying b.c.s is $y_n(x,t) = T_n(t)X_n(x)$. Since the wave equation is linear and homogeneous (with homogeneous b.c.s), we can add the solution together to find the general solution. By construction, the general solution will satisfy b.c.s. Suppose we further impose our i.c.s,

$$p(x) = \sum_{n=1}^{\infty} C_n \sin\frac{n\pi x}{L}, \quad q(x) = \sum_{n=1}^{\infty} D_n \frac{n\pi c}{L}\cos\frac{n\pi x}{L}$$

so the coefficients are those for the Fourier sine series (2.7b)

$$C_n = \frac{2}{L}\int_0^L p(x)\sin\frac{n\pi x}{L}dx, \quad D_n = \frac{2}{n\pi c}\int_0^L q(x)\sin\frac{n\pi x}{L}dx$$

Thus the general solution is

$$y(x,t) = \sum_{n=1}^{\infty}\left(c_n \cos\frac{n\pi ct}{L} + D_n \sin\frac{n\pi ct}{L}\right)\sin\frac{n\pi x}{L}$$

3   Green's functions

Response to impulses, delta function (treated heuristically), Green's functions for initial and boundry value problems.   **[3]**

3.1   Delta functions

Technically, Dirac-delta functions are distributions and not ordinary functions. But after some abuse of notation, we can write

Definition 3.1 (Dirac delta function)**:** We have $\delta(x) = 0$ whenever $x \neq 0$ and $\delta(x)$ to be not finite at $x = 0$. Hence, we generalize $\delta(x - \xi)$ with the following properties

$$\delta(x - \xi) = 0 \ \forall x \neq \xi, \quad \int_{-\infty}^{\infty} \delta(x - \xi)dx = 1 \tag{3.1}$$

This acts as a linear operator $\int \delta(x - \xi)dx$ on an arbitrary function $f(x)$ to produce a number $f(\xi)$, i.e.

$$\int_{-\infty}^{\infty} \delta(x - \xi)f(x)dx = f(\xi) \tag{3.2}$$

provided $f(x)$ is 'well-behaved' at $x = \xi, \pm\infty$. In turn, we define the derivative of the Dirac delta:

$$\int_{-\infty}^{\infty} \delta'(x - \xi)f(x)dx = -f'(\xi) \tag{3.3}$$

Remarks 3.1. *$\delta(x)$ always appears inside an integral as a linear operator where it's well-defined. Physically, it represents a unit point source.*

Example 3.1 (Some limiting approximations)**.** *A discrete example for the dirac delta function is*

$$\delta_n(x) = \begin{cases} 0 & x > \frac{1}{n} \\ n/2 & |x| \leq \frac{1}{n} \\ 0 & x < -\frac{1}{n} \end{cases} \tag{3.4a}$$

where we take $n \to \infty$. A continuous example would be

$$\delta_\epsilon(x) = \frac{1}{\epsilon\sqrt{\pi}}e^{-x^2/\epsilon^2} \tag{3.4b}$$

where we take $\epsilon \to 0$. We can check (3.2) for $\xi = 0$ for all well-behaved functions $f(x)$ at $x = 0, \pm\infty$.

$$\lim_{\epsilon \to 0} \int_{-\infty}^{\infty} \frac{1}{\epsilon\sqrt{\pi}}e^{-x^2/\epsilon^2}f(x)dx = \lim_{\epsilon \to 0} \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}}e^{-y^2}f(\epsilon y)dy = \lim_{\epsilon \to 0} \int_{-\infty}^{\infty} \frac{e^{-y^2}}{\sqrt{\pi}}(f(0)+\epsilon y f'(0)+\dots)dy = f(0)$$

Other examples would be

$$\delta_n(x) = \frac{\sin nx}{\pi x} = \frac{1}{2\pi}\int_{-n}^{n} e^{ikx}dk \tag{3.5}$$

$$\delta_n(x) = \frac{n}{2}\operatorname{sech}^2(nx) \tag{3.6}$$

Definition 3.2 (Heaviside function)**:** The unit step function aka the Heaviside function

$$H(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{3.7}$$

is another instance of a generalized function on $\mathbb{R}$ by

$$H(\phi) = \int_{\mathbb{R}} H(x)\phi(x)dx = \int_{0}^{\infty} \phi(x)dx$$

Proposition 3.1:

$$H(x) = \int_{-\infty}^{x} \delta(x')dx' \tag{3.8}$$

Theorem 3.1 (Properties of Dirac delta function)**:**

1. Sampling property:

$$\int_a^b f(x)\delta(x-\xi)dx \begin{cases} f(\xi) & a < \xi < b \\ 0 & \text{otherwise} \end{cases} \tag{3.9}$$

2. Even property:

$$\int_{-\infty}^{\infty} f(x)\delta(-(x-\xi))dx = \int_{-\infty}^{\infty} f(x)\delta(x-\xi)dx \tag{3.10}$$

3. Scaling property:

$$\int_{-\infty}^{\infty} f(x)\delta(a(x-\xi))dx = \frac{1}{|a|}f(\xi) \tag{3.11}$$

4. Advanced scaling: Suppose $g(x)$ has n isolated zeros at $x_1, \ldots, x_n$, then (with $g'(x_i) \neq 0$)

$$\delta(g(x)) = \sum_{i=1}^{n} \frac{\delta(x-x_i)}{|g'(x_i)|} \tag{3.12}$$

5. Isolation property: If $g(x)$ is continuous at $x = 0$, then

$$g(x)\delta(x) = g(0)\delta(x) \tag{3.13}$$

Example 3.2. *Since the roots for $x^2 - 1$ are ± 1, then*

$$I = \int_{-\infty}^{\infty} f(x)\delta(x^2-1)dx = \int_{1-\epsilon}^{1+\epsilon} f(x)\frac{\delta(x-1)}{|2x|}dx + \int_{-1-\epsilon}^{-1+\epsilon} f(x)\frac{\delta(x+1)}{|2x|}dx = \frac{1}{2}(f(1) + f(-1))$$

Proposition 3.2: The eigenfunction expansion of $\delta(x)$ is

$$\delta(x) = \frac{1}{2L} \sum_{n=-\infty}^{\infty} e^{in\pi x/L} \tag{3.14}$$

Proof. The Fourier coefficient (2.6b) is $c_n = \frac{1}{2L}\int_{-L}^{L} \delta(x)e^{-in\pi x/L}dx = \frac{1}{2L}$.   □

This is consistent with the sampling property, where we take $f(x) = \sum_{n=-\infty}^{\infty} d_n e^{in\pi x/L}$,

$$\int_{-L}^{L} f(x)\delta(x)dx = \frac{1}{2L} \sum_{n=-\infty}^{\infty} d_n \int_{-L}^{L} e^{in\pi x/L}e^{-in\pi x/L}dx = \sum_n d_n = f(0)$$

Aside from expanding $\delta(x)$ as a Fourier series, any other basis of orthogonal function will do.

3.2 Solving ODEs with Green's functions

Proposition 3.3 (Boundary value problem)**:** The solution for the boundary value problem $\mathcal{L}y = f(x)$ with homogeneous b.c.s $y(a) = y(b) = 0$ is

$$y(x) = y_+(x) \int_\alpha^x \frac{y_-(\xi)f(\xi)d\xi}{\alpha(\xi)W(\xi)} + y_-(x) \int_x^b \frac{y_+(\xi)f(\xi)d\xi}{\alpha(\xi)W(\xi)} \tag{3.15}$$

where $y_\pm(x)$ are an appropriate linear combinations of the two linearly independent homogeneous solutions that solve $\mathcal{L}y = 0$.

Proof. The corresponding Green's function satisfy

$$\mathcal{L}G(x,\xi) = \delta(x - \xi), \quad G(a,\xi) = G(b,\xi) = 0$$

For $a \leq x < \xi$: $G_1(x,\xi) = Ay_1(x) + By_2(x)$ s.t. $Ay_1(a) + By_2(a) = 0$, where y_1 and y_2 are two linearly independent homogeneous solutions. This defines a complementary function (1.29) $y_-(x)$ s.t. $y_-(a) = 0$, so the general homogeneous solution with $G_1 = 0$ at $x = a$.

$$G_1 = Cy_-(x), \quad y_-(a) = 0$$

For $\xi < x \leq b$: similarly find

$$G_2 = Dy_+(x), \quad y_+(b) = 0$$

where y_+ is also a complementary function.

Suppose G is discontinuous, so $G \propto H(x-\xi)$ which means $G' \propto \delta(x-\xi)$ and hence $G'' \propto \delta'(x-\xi)$. But

$$\mathcal{L}G \propto \alpha(x)\delta'(x-\xi) + \beta(x)\delta(x-\xi) + \gamma(x)H(x-\xi)$$

LHS is $\delta(x-\xi)$ with no term containing $\delta'(x-\xi)$, hence a contradiction. We thus must have

$$[G]_{\xi_-}^{\xi_+} = 0 \implies Cy_-(\xi) = Dy_+(\xi) \tag{3.16}$$

Integrate $\mathcal{L}G = \delta(x-\xi)$ across $x = \xi$: LHS gives

$$1 = \int_{\xi_-}^{\xi_+} \mathcal{L}G dx = \int_{\xi_-}^{\xi_+} (\alpha G'' + \beta G' + \gamma G)dx = \alpha(\xi)[G']_{\xi_-}^{\xi_+} + (\beta - \alpha')[G]_{\xi_-}^{\xi_+} + \int_{\xi_-}^{\xi_+} (\gamma - \beta' + \alpha'')G dx$$

where $\int_{\xi_-}^{\xi_+} (\gamma - \beta' + \alpha'')G dx = 0$, $[G]_{\xi_-}^{\xi_+} = 0$ by continuity (3.16). Hence,

$$[G']_{\xi_-}^{\xi_+} = \frac{1}{\alpha(\xi)} \implies Dy_+'(\xi) - Cy_-'(\xi) = \frac{1}{\alpha(\xi)} \tag{3.17}$$

Solving them gives

$$C(\xi) = \frac{y_+(\xi)}{\alpha(\xi)W(\xi)}, \quad D(\xi) = \frac{y_-(\xi)}{\alpha(\xi)W(\xi)}$$

where the Wronskian is

$$W(\xi) = y_-(\xi)y_+'(\xi) - y_+(\xi)y_-'(\xi)$$

and $W(\xi) \neq 0$ if y_+, y_- are linearly independent.

$$G(x,\xi) = \frac{1}{\alpha(\xi)W(\xi)} \begin{cases} y_-(x)y_+(\xi) & a \leq x < \xi \\ y_+(x)y_-(\xi) & \xi < x \leq b \end{cases} = \frac{H(\xi-x)y_-(x)y_+(\xi) + H(x-\xi)y_+(x)y_-(\xi)}{\alpha(\xi)W(\xi)} \tag{3.18}$$

so the solution to (3.21) with $y(a) = y(b) = 0$ is $y(x) = \int_a^b G(x,\xi)f(\xi)d\xi = \int_a^x G_2(x,\xi)f(\xi)d\xi + \int_x^b G_1(x,\xi)f(\xi)d\xi$, and we recover (3.29). $\qquad \square$

Example 3.3. *Solve $y'' = y = f(x)$, $y(0) = y(1) = 0$ by constructing the $G(x,\xi)$.*

    1. homogeneous solution $y_1 = e^x$, $y_2 = e^{-x}$, so with homogeneous b.c.s (by inspection)

$$G(x,\xi) = \begin{cases} C(\xi)\sinh x & 0 \leq x < \xi \\ D(\xi)\sinh(1-x) & \xi < x \leq 1 \end{cases}$$

2. continuous at $x = \xi$ gives $C \sinh \xi = D \sinh(1 - \xi) \implies C = \frac{D \sinh(1-\xi)}{\sinh \xi}$

3. discontinuous at $x = \xi$: $-D \cosh(1 - \xi) - C \cosh \xi = 1 \implies -D \sinh(1) = \sinh \xi$, so the solution is

$$y(x) = -\frac{\sinh(1 - x)}{\sinh(1)} \int_0^x \sinh \xi f(\xi) d\xi - \frac{\sinh(x)}{\sinh(1)} \int_x^1 \sinh(1 - \xi) f(\xi) d\xi$$

Remarks 3.2 (Inhomogeneous b.c.s). *Consider the modified problem: find y_p solution to $\mathcal{L}y = 0$ satisfying b.c.s $y(a) = c$ and $y(b) = d$.*

Then we try $y_g(x) = y(x) - y_p(x)$ s.t. $\mathcal{L}y_g = 0$ with $y_g(a) = y_g(b) = 0$. Consider $y'' - y = f(x)$ but with b.c.s $y(0) = 0$ and $y(1) = 1$. The particular solution $y_p = A \sinh x + B \cosh x$ has the b.c.s $y_p(0) = 0 \implies B = 0$ and $y_p(1) = 1 \implies A = \frac{1}{\sinh(1)}$. Now solve for $y_g = y - y_p$ with $y_g(0) = 0$, $y_g(1) = 0$ with solution $y(x) = \frac{\sinh x}{\sinh 1} + y_g(x)$.

Remarks 3.3 (Higher order ODEs). *If $\mathcal{L}y = f(x)$ to nth order (coefficient $\alpha(x)$ for $y^{(n)}(x)$) with homogeneous b.c.s, then we generalize the Green's function to have the properties:*

1. G_1, G_2 are homogeneous solution satisfying homogeneous b.c.s;

2. continuity $G_1 = G_2$, $G_1' = G_2'$, ... $G_1^{(n-2)} = G_2^{(n-2)}$ at $x = \xi$;

3. jump on $(n-1)$th derivative

$$[G^{(n-1)}]_{\xi_-}^{\xi_+} = G_2^{(n-1)}|_{\xi_+} - G_1^{(n-1)}|_{\xi_-} = \frac{1}{\alpha(\xi)}$$

Proposition 3.4 (Initial value problem)**:** The solution for the initial value problem $\mathcal{L}y = f(t)$ for $t \geq a$, with $y(a) = 0 = y'(a)$ is

$$y(t) = \int_a^t \frac{y_+(t)f(\tau)}{\alpha(\tau)y_+'(\tau)} d\tau \tag{3.19}$$

Proof. The corresponding Green's function satisfy

$$\mathcal{L}G(t, \tau) = \delta(t - \tau), \quad G(a, \tau) = G'(a, \tau) = 0$$

For $t < \tau$: $G_1 = Ay_1(t) + By_2(t)$ with $Ay_1(a) + By_2(a) = 0$ and $Ay_1'(a) + By_2'(a) = 0$. But since $W(a) \neq 0$ (y_1, y_2 independent), $A = B = 0$, so $G(t, \tau) = 0$ for $a \leq t \leq \tau$, i.e. no change until impulse $t = \tau$. For $t > \tau$: by G continuity (3.27), $G_2(\tau, \tau) = 0$, so choose $G_2 = Dy_+(t)$ with $y_+(\tau) = Ay_1(\tau) + By_2(\tau) = 0$. But discontinuity in G' gives

$$[G']_{\tau_-}^{\tau_+} = G_2'(\tau, \tau) - G_1'(\tau, \tau) = Dy_+'(\tau) = \frac{1}{\alpha(\tau)}$$

Hence, we have

$$G(t, \tau) \begin{cases} 0 & a \leq t \leq \tau \\ \frac{y_+(t)}{\alpha(\tau)y_+'(\tau)} & a \leq \tau < t \end{cases} \tag{3.20}$$

Solution follows from $y(t) = \int G(t, \tau)f(\tau)d\tau$. $\qquad\qquad\square$

Remarks 3.4. *Causality is 'built-in' as only forces acting prior to t affect the solution at t.*

Example 3.4. *Solve $y'' - y = f(t)$ with $y(0) = y'(0) = 0$.*

1. homogeneous solution and i.c.s: $t < \tau$, $G_1 = 0$ while $t > \tau$, $G_2 = Ae^t + Be^{-t}$.

2. continuity: $G(\tau, \tau) = 0 \implies G_2 = D \sinh(t - \tau)$ for $t > \tau$.

3. jump: $[G'] = \frac{1}{\alpha} = 1 \implies G_2'(t, \tau) = D \cosh 0 = D = 1$.

The solution is $y(t) = \int_0^t f(\tau) \sinh(t - \tau) d\tau$

4   Fourier transform

Fourier transforms; relation to Fourier series, simple properties and examples, convolution theorem, correlation functions, Parseval's theorem and power spectra   **[2]**

4.1   Motivation and properties

Definition 4.1 (Fourier transform)**:** The Fourier transform of a function $f(x)$ is

$$\tilde{f}(k) := \mathcal{F}[f(x)] = \int_{-\infty}^{\infty} f(x)e^{-ikx}dx \tag{4.1}$$

and the inverse Fourier transform is

$$f(x) = \mathcal{F}^{-1}[\tilde{f}(k)] = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{f}(k)e^{ikx}dk \tag{4.2}$$

Beware there are several conventions.

Theorem 4.1 (Fourier inversion theorem)**:** For the following to be true

$$\mathcal{F}^{-1}[\mathcal{F}[f(x)]] = f(x) \tag{4.3}$$

it is sufficient to require f and $\tilde{f}$ to be absolutely integrable, i.e. $\int_{-\infty}^{\infty} |f(x)|dx = M < \infty$, so $f \to 0$ as $x \to \pm\infty$.

From here onwards, we assume all functions mentioned will have a Fourier transform.

Example 4.1 (Gaussian)**.** *The Fourier transform of the Gaussian function*

$$f(x) = \frac{1}{\sigma\sqrt{\pi}}e^{-x^2/\delta^2} \tag{4.4}$$

Discard the $\sin(kx)$ term since it is odd, while the Gaussian is even, the result would be

$$\tilde{f}(k) = \frac{1}{\sigma\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-x^2/\delta^2}e^{-ikx}dx = \frac{1}{\sigma\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-x^2/\delta^2}\cos(kx)dk = e^{-k^2\sigma^2/4} \tag{4.5}$$

A simpler way to obtain the result is by considering $\tilde{f}'(k)$:

$$
\begin{aligned}
\frac{d\tilde{f}}{dk} = \tilde{f}'(k) &= -\frac{1}{\sigma\sqrt{\pi}}\int_{-\infty}^{\infty} xe^{-x^2/\delta^2}\sin(kx)dx \\
&= \frac{1}{\sigma\sqrt{\pi}}[0.5\sigma^2 e^{-x^2/\delta^2}\sin kx]_{-\infty}^{\infty} - \frac{1}{\sigma\sqrt{\pi}}\int_{-\infty}^{\infty}(0.5k\sigma)^2 e^{-x^2/\sigma^2}\cos kxdx \\
&= -\frac{k\sigma^2}{2}\tilde{f}(k)
\end{aligned}
$$

Example 4.2. *For $f(x) = e^{-a|x|}$ where $a > 0$, the Fourier transform is*

$$\tilde{f}(k) = \frac{2a}{a^2 + k^2} \tag{4.6a}$$

This can be done in two ways:

- *integrate $2\int_0^{\infty} e^{-ax}\cos kxdx$ by parts twice;*

- *integrate $\int_0^{\infty} e^{-(a-ik)x}dx + \int_{-\infty}^0 e^{(a+ik)x}dx$ directly.*

Note if $f(x) = e^{-ax}$ only for $x > 0$ instead, with $f(x) = 0$ for $x \leq 0$, then

$$\tilde{f}(k) = \frac{1}{ik + a} \tag{4.6b}$$

Remarks 4.1 (Relation to Fourier series)**.** *Write the Fourier series as $f(x) = \sum_{n=-\infty}^{\infty} c_n e^{ik_n x}$ where $k_n = \frac{n\pi}{L}$, so write $k_n = n\Delta k$ with $\Delta k = \frac{\pi}{L}$, then*

$$c_n = \frac{1}{2L}\int_{-L}^{L} f(x)e^{-ik_n x}dx = \frac{\Delta k}{2\pi}\int_{-L}^{L} f(x)e^{-ik_n x}dx \implies f(x) = \sum_{n=-\infty}^{\infty}\frac{\Delta k}{2\pi}e^{ik_n x}\int_{-L}^{L} f(x')e^{-ik_n x'}dx'$$

But recall the Riemann sum definition of the integral of a function:

$$\lim_{\Delta k \to 0}\sum_{n\in\mathbb{Z}}\Delta k \; g(n\Delta k) = \int_{\mathbb{R}} g(k)dk$$

where $g(k_n = n\Delta k) = \frac{e^{ik_n x}}{2\pi}\int_{-L}^{L} f(x')e^{-ikx'}dx'$, so take limit $L \to \infty$ (extend to non-periodic functions) to get

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx}\int_{-\infty}^{\infty} f(x')e^{-ikx'}dx'dk = \mathcal{F}^{-1}[\mathcal{F}[f(x)]]$$

i.e. recover (4.3). Note when $f(x)$ is discontinuous at x, the Fourier transform gives

$$\mathcal{F}^{-1}[\mathcal{F}[f(x)]] = \frac{1}{2}(f(x_-) + f(x_+)) \tag{4.7}$$

Theorem 4.2 (Properties of Fourier transform)**:**

1. linearity:
$$h(x) = \lambda f(x) + \mu g(x) \iff \tilde{h}(k) = \lambda\tilde{f}(k) + \mu\tilde{g}(k) \tag{4.8}$$

2. translation and re-phasing:
$$h(x) = f(x-\lambda) \iff \tilde{h}(k) = e^{-i\lambda k}\tilde{f}(k) \tag{4.9}$$
$$h(x) = e^{i\lambda x}f(x) \iff \tilde{h}(k) = \tilde{f}(k-\lambda) \tag{4.10}$$

3. scaling:
$$h(x) = f(\lambda x) \iff \tilde{h}(k) = \frac{1}{|\lambda|}\tilde{f}(k/\lambda) \tag{4.11}$$

4. derivative:
$$h(x) = xf(x) \iff \tilde{h}(k) = i\tilde{f}'(k) \tag{4.12}$$
$$h(x) = f'(x) \iff \tilde{h}(k) = ik\tilde{f}(k) \tag{4.13}$$

5. duality:
$$g(x) = \tilde{f}(x) \iff \tilde{g}(k) = 2\pi f(-k) \tag{4.14}$$

Proof.

1. Trivial since the integration is linear.

2. Using the substitution $x = y + \lambda$, $\int_{-\infty}^{\infty} f(x-\lambda)e^{-ikx}dx = \int_{-\infty}^{\infty} f(y)e^{-ik(y+\lambda)}dy = e^{-i\lambda k}\tilde{f}(k)$.

3. Using the substitution $y = \lambda x$ for $\lambda > 0$.
$$\int_{-\infty}^{\infty} f(\lambda x)e^{-ikx}dx = \frac{1}{\lambda}\int_{-\infty}^{\infty} f(y)e^{-iky/\lambda}dy = \frac{1}{\lambda}\tilde{f}(k/\lambda)$$
   Take care of the case $\lambda < 0$, i.e. do modulus for λ.

4. $\int_{-\infty}^{\infty} xf(x)e^{-ikx}dk = \frac{-1}{i}\frac{d}{dk}\int_{-\infty}^{\infty} f(x)e^{-ikx}dx$ and $\tilde{h}(k) = [f(x)e^{-ikx}]_{-\infty}^{\infty} + ik\int_{-\infty}^{\infty} f(x)e^{-ikx}dx = ik\tilde{f}(k)$.

5. Consider $f(-x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\tilde{f}(k)e^{-ikx}dk$, so $k \leftrightarrow x \implies f(-k) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\tilde{f}(x)e^{-ikx}dx$.

$\square$

Corollary 4.1: $f(-x) = \frac{1}{2\pi}\mathcal{F}[\tilde{f}(k)] = \frac{1}{2\pi}\mathcal{F}^2[f(x)]$.

Example 4.3 (Top-hat)**.** *Consider top-hat function $f(x) = 1$ for $|x| \leq a$ (where $a > 0$) and $f(x) = 0$ for $|x| > a$. The Fourier transform is*

$$\tilde{f}(k) = \int_{-\infty}^{\infty} f(x)e^{ikx}dx = \int_{-a}^{a} \cos(kx)dx = \frac{2\sin(ka)}{k} \tag{4.15}$$

Fourier inversion theorem (4.3) gives

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx}\frac{2\sin ka}{k}dk = \left\{ \begin{array}{ll} 1 & |x| \leq a \\ 0 & |x| > a \end{array} \right.$$

Set $x = 0$, then take $k \to x$ to get the Dirichlet's discontinuity formula:

$$\int_{-\infty}^{\infty} \frac{\sin(ax)}{x}dx = \left\{ \begin{array}{ll} \pi/2 & a > 0 \\ 0 & a = 0 \\ -\pi/2 & a < 0 \end{array} \right. = \frac{\pi}{2}\operatorname{sgn}(a) \tag{4.16}$$

4.2 Convolution and Parseval's theorem

Theorem 4.3 (Convolution theorem)**:**

$$\tilde{h}(k) = \tilde{f}(k)\tilde{g}(k) \iff h(x) = (f * g)(x) := \int_{-\infty}^{\infty} f(y)g(x-y)dy \tag{4.17}$$

$$h(x) = f(x)g(x) \iff \tilde{h}(k) = \frac{1}{2\pi}(\tilde{f} * \tilde{g})(k) := \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{f}(p)\tilde{g}(k-p)dp \tag{4.18}$$

Proof. By duality (4.14), (4.18) follows from (4.17).

$$\begin{aligned} h(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{f}(k)\tilde{g}(k)e^{ikx}dk \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(y)e^{-iky}dy \right)\tilde{g}(k)e^{ikx}dk \\ &= \int_{-\infty}^{\infty} f(y)\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{g}(k)e^{ik(x-y)}dk \right)dy \\ &= \int_{-\infty}^{\infty} f(y)g(x-y)dy \\ &= (f * g)(x) \end{aligned}$$

$\square$

Proposition 4.1: Fourier transform preserves the inner product between two functions, i.e.

$$\langle g, f \rangle = \int_{-\infty}^{\infty} f(x)g^*(x)dx = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{f}(k)\tilde{g}^*(k)dk = \frac{1}{2\pi}\langle \tilde{g}, \tilde{f} \rangle \tag{4.19}$$

Proof. Consider the Fourier transform of $g^*(-x)$:

$$\int_{-\infty}^{\infty} g^*(-x)e^{-ikx}dx = \left(\int_{-\infty}^{\infty} g(-x)e^{ikx}dx \right)^* = \left(\int_{-\infty}^{\infty} g(y)e^{-iky}dy \right)^* = \tilde{g}^*(k)$$

where we substitute $y = -x$. Substitute into (4.17) by replacing $g(x-y) \to g^*(y-x)$, then

$$h(x) = \int_{-\infty}^{\infty} f(y)g^*(y-x)dy = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{f}(k)\tilde{g}^*(k)e^{ikx}dk$$

Take $x = 0$ and switch the dummy variable $y \to x$ on LHS to obtain (4.19). $\square$

Corollary 4.2 (Parseval's theorem)**:**

$$\int_{-\infty}^{\infty} |f(x)|^2 dx = \frac{1}{2\pi}\int_{-\infty}^{\infty} |\tilde{f}(k)|^2 dk \tag{4.20}$$

Proof. $g^* = f^*$ for (4.19). $\square$

4.3 Fourier transforms of generalized functions

Proposition 4.2:

$$\delta(x - u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ik(x-u)} dk \tag{4.21}$$

Proof. Consider the inversion

$$f(x) = \mathcal{F}^{-1}[\mathcal{F}[f(x)]] = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(u)e^{-iku} du \right) e^{ikx} dk = \int_{-\infty}^{\infty} f(u) \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ik(x-u)} dk \right) du$$

Result follows from the definition of $\delta(x - u)$. $\qquad\qquad\square$

Proposition 4.3:

   1.

$$\mathcal{F}[\delta(x)] = 1 \iff \mathcal{F}[1] = 2\pi\delta(k) \tag{4.22}$$

   2.

$$\mathcal{F}[\delta(x - a)] = e^{-ika} \tag{4.23}$$

Proof.

   1.

$$\mathcal{F}\delta[\phi] = 2\pi\delta[\mathcal{F}^{-1}\phi] = 2\pi\delta \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ikx}\phi(k)dk \right] = 2\pi \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi(k)dk$$

      Second part follows from the Fourier inversion theorem (Theorem 4.1).

   2. Follow from the translation and re-phasing property (Theorem 4.2.2).

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$$

Corollary 4.3:

$$f(x) = \cos\omega x \iff \tilde{f}(k) = \pi(\delta(k + \omega) + \delta(k - \omega))$$
$$f(x) = \sin\omega x \iff \tilde{f}(k) = \pi(\delta(k + \omega) - \delta(k - \omega))$$

Trivial. One can then find $\mathcal{F}^{-1}$ for $\sin(\omega k)$ and $\cos(\omega k)$ using duality (4.14).

Proposition 4.4: The Fourier transform of the Heaviside function is p.v.$(ik)^{-1} + \pi\delta(k)$, where p.v. refers to Cauchy principal value (exclude the point $k = 0$ from any integral containing this term).

4.4 Applications

Power spectrum

Definition 4.2 (Correlation): The correlation of two functions, $h = f \otimes g$, is defined by

$$h(x) = \int_{-\infty}^{\infty} [f(y)]^* g(x + y) \; dy$$

Note the argument of h is the shift between the arguments of f and g. The correlation quantifies the relationship between two oscillatory functions.

Definition 4.3 (Autocorrelation): Correspondingly, the autocorrelation of a function $f(x)$ is $f \otimes f$.

Proposition 4.5: The Fourier transform of a correlation $f \otimes g$ is $[\tilde{f}(k)]^* \tilde{g}(k)$.

Corollary 4.4 (Wiener-Khinchin theorem):

$$\mathcal{F}[f \otimes f] = |\tilde{f}|^2$$

Remarks 4.2. *The power spectra $\Phi(k) := |\tilde{f}(k)|^2$ quantifies the spectral content of a signal $f(t)$. In particular,*

- *a perfectly periodic signal has a spectrum consisting of a series of delta functions at the principal frequency and its harmonics. Its autocorrelation does not decay as $t \to \infty$.*

- *white noise is an ideal random signal with autocorrelation function proportional to $\delta(t)$, i.e. signal is perfectly correlated, and has a flat spectrum, i.e. constant Φ.*

Motivation: ODE for BVP

Consider

$$y'' - y = f(x)$$

with homogeneous b.c.s $y \to 0$ as $x \to \pm\infty$. Take Fourier transform

$$(-k^2 - 1)\tilde{y} = \tilde{f}$$

by (4.13), so solution is

$$\tilde{y}(k) = \frac{-\tilde{f}(k)}{1 + k^2} := \tilde{f}(k)\tilde{g}(k)$$

where $\tilde{g}(k) = -\frac{1}{1+k^2}$, but $\tilde{g}$ is the Fourier transform of $g(x) = -\frac{1}{2}e^{-|x|}$ (see (4.6)). Thus by Convolution theorem (4.17)

$$y(x) = \int_{-\infty}^{\infty} f(u)g(x-u)du = -\frac{1}{2}\int_{-\infty}^{\infty} f(u)e^{-|x-u|}du = -\frac{1}{2}\int_{-\infty}^{x} f(u)e^{u-x}du - \frac{1}{2}\int_{x}^{\infty} f(u)e^{x-u}du$$

which is in the form of a BVP Green's function. Verify by constructing $G(x, u)$ directly.

Motivation: Signal processing (IVP)

Suppose given input $I(t)$ acted upon by linear operator $\mathcal{L}_{in}$ to yield output $O(t)$. The Fourier transform of the input $\tilde{I}(\omega)$ is denoted the resolution

$$\tilde{I}(\omega) = \int_{-\infty}^{\infty} I(t)e^{-i\omega t}dt \tag{4.24}$$

In frequency domain, $\mathcal{L}_{in}I(t)$ means $\tilde{I}(\omega)$ is multiplied by a transfer function $\tilde{R}(\omega)$ to yield output

$$O(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{R}(\omega)\tilde{I}(\omega)e^{i\omega t}d\omega \tag{4.25}$$

with response function given by

$$R(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{R}(\omega)e^{i\omega t}d\omega \tag{4.26}$$

By the Convolution theorem (4.17), the output is $O(t) = \int_{-\infty}^{\infty} I(u)R(t-u)du$. If we assume no input $I(t) = 0$ for $t < 0$ and zero output $O(t) = 0$ for $t < 0$, i.e. $R(t-u)$ has source $\delta(t-u)$, so we require by causality $0 < u < t$,

$$O(t) = \int_{0}^{t} I(u)R(t-u)du \tag{4.27}$$

This has the same form as the IVP problem, where the transfer function is the Green's function.

General transfer functions for ODEs

Suppose input/output relation given by a linear ODE of nth order

$$\mathcal{L}O(t) := \left(\sum_{i=1}^{n} a_i \frac{d^i}{dt^i} \right) O(t) = I(t) \tag{4.28}$$

Set $\mathcal{L}_{in} = 1$. Take Fourier transform, $\sum_{j=1}^{n} a_j (i\omega)^j \tilde{O}(\omega) = \tilde{I}(\omega)$. So the transfer function is

$$\tilde{R}(\omega) = (a_0 + a_1(\omega) + \cdots + a_n(i\omega)^n)^{-1} \tag{4.29}$$

Factorize the nth degree polynomial into products of n roots, $(i\omega - c_j)^{k_j}$ where k_j is the multiplicity, with $j = 1, 2, \ldots J$ (with repeated roots) s.t. $\sum_{j=1}^{J} k_j = n$. Using partial fractions, this can be written as

$$\tilde{R}(\omega) = \frac{1}{(i\omega - c_1)^{k_1} \ldots (i\omega - c_J)^{k_J}} = \sum_{j=1}^{J} \sum_{m=1}^{k_j} \frac{\Gamma_{jm}}{(i\omega - c_j)^m} \tag{4.30}$$

for some constant Γ_{jm}. For repeated root,

$$\frac{1}{(i\omega - c_j)^{k_j}} \rightarrow \frac{\Gamma_{j1}}{i\omega - c_j} + \frac{\Gamma_{j2}}{(i\omega - c_j)^2} + \cdots + \frac{\Gamma_{jk_j}}{(i\omega - c_j)^{k_j}}$$

To solve, we must invert $\frac{1}{(i\omega - a)^m}$ for $m \geq 1$. We know (4.6a) that $F^{-1} \frac{1}{i\omega - a} = e^{at}$ for $t > 0$ and $\text{Re}[a] < 0$ and 0 for $t < 0$. So assume $\text{Re}[c_j] < 0 \; \forall j$ to eliminate the growing exponential solution. For $m = 2$, note $i \frac{d}{d\omega} \frac{1}{i\omega - a} = \frac{1}{(i\omega - a)^2}$ and recall (4.12) $\mathcal{F}^{-1}(tf(t)) = if'(\omega)$, so $\mathcal{F}^{-1}(\frac{1}{(i\omega - a)^2}) = te^{at}$ for $t > 0$ and 0 otherwise. By induction, for $t > 0$, $\mathcal{F}^{-1}(\frac{1}{(i\omega - a)^m}) = \frac{t^{m-1}e^{at}}{(m-1)!}$. The response function takes the form

$$R(t) = \sum_{j} \sum_{m} \Gamma_{jm} \frac{t^{m-1}}{(m-!)!} e^{c_j t}, \quad t > 0 \tag{4.31}$$

We can solve (4.28) in Green's function form (4.27) or directly invert $R(\omega)I(\omega)$ for the given polynomial $I(\omega)$.

Example 4.4 (Damped oscillator)**.** *Solve $\mathcal{L}y = y'' + 2py' + (p^2 + q^2)y = f(t)$ with damping $p > 0$ and homogeneous b.c.s $y(0) = y'(0) = 0$. The Fourier transform is*

$$(i\omega)^2 \tilde{y} + 2ip\omega \tilde{y} + (p^2 + q^2)\tilde{y} = \tilde{f} \implies \tilde{y} = \frac{\tilde{f}}{-\omega^2 + 2ip\omega + p^2 + q^2} := \tilde{R}(\omega)\tilde{f}(\omega)$$

Inverting with Convolution theorem (4.17),

$$y(t) = \int_0^t R(t - \tau)f(\tau)dt, \quad R(t - \tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{i\omega(t-\tau)}d\omega}{p^2 + q^2 + 2ip\omega - \omega^2}$$

One can show $\mathcal{L}R(t - \tau) = \delta(t - \tau)$ using (4.23), i.e. the response function $R(t - \tau)$ is the Green's function.

5 Matrices

N-dimensional vector spaces, matrices, scalar product, transformation of basis vectors. Eigenvalues and eigenvectors of a matrix; degenerate case, stationary property of eigenvalues. Orthogonal and unitary transformations. Quadratic and Hermitian forms, quadric surfaces. **[5**

5.1 Recap

The recap from IA Maths contains far too much content to be included in this summary notes. The concepts needed: Linear equations. Notion of a vector space; linear mappings. Matrix addition and multiplication. Determinant of a matrix. Statement of the main properties of determinants. Inverse matrix. Equations $Ax = 0$ with non-zero solutions. Symmetric, antisymmetric and orthogonal matrices. Eigenvalues and eigenvectors for symmetric matrices. hessian matrix as an example.

5.2 Vector space and span

Definition 5.1 (Vector space)**:** A vector space over a field $\mathbb{F}$ (sometimes known as $\mathbb{F}$-space) is an abelian group $(V, +)$ equipped with scalar multiplication:

$$\mathbb{F} \times V \to V, \quad (\lambda, v) \mapsto \lambda v$$

s.t. it satisfies the following axioms:

1. $\lambda(v_1 + v_2) = \lambda v_1 + \lambda v_2$

2. $(\lambda_1 + \lambda_2)v = \lambda_1 v + \lambda_2 v$

3. $\lambda_1(\lambda_2 v) = (\lambda_1 \lambda_2)v$

4. $1v = v$ and $-1v = -v$

5. $0v = 0$ where 0 is the additive unit

$\forall \lambda, \lambda_i \in \mathbb{F}, v, v_i \in V$.
We will only consider finite-dimensional vector spaces.

Definition 5.2 (Subspace)**:** Let V be a $\mathbb{F}$-space. The subset $U \subseteq V$ is a subspace written $U \leq V$, if it satisfies

1. $0 \in U$,

2. $\lambda_1 u_1 + \lambda_2 u_2 \in U \ \forall u_i \in U, \ \forall \lambda_i \in \mathbb{F}$.

If $U \leq V$, then U itself is a $\mathbb{F}$-space.

Definition 5.3 (Spanning set)**:** Let V be a $\mathbb{F}$-space, $S \subseteq V$ (S could be infinite). Define the span of S to be

$$\langle S \rangle := \left\{ \sum_{s \in S} \lambda_s s \mid \lambda_s \in \mathbb{F} \right\} \tag{5.1}$$

where the linear combination is finite.

Definition 5.4 (Span)**:** Let V is a $\mathbb{F}$-space, $S \subseteq V$, then S spans V if $\langle S \rangle = V$.

Definition 5.5 (Finite dimensional)**:** V is finite dimensional over $\mathbb{F}$ if it is spanned by a finite set.

Definition 5.6 (Linear independent)**:** The set of vectors $\{v_1, \dots, v_n\} \in V$ is a linearly independent set over $\mathbb{F}$ if

$$\sum_{i=1}^{n} \lambda_i v_i = 0 \implies \lambda_i = 0 \ \forall i \tag{5.2}$$

$S \subseteq V$ is linearly independent if every finite subset of it is. Otherwise, it is linearly dependent. A linearly independent set thus cannot contain 0.

Definition 5.7 (Basis)**:** S is a basis for V if S is linearly independent and spans V.

Definition 5.8 (Matrix)**:** An $m \times n$ matrix over $\mathbb{F}$ is an array with m rows and n columns, entries in $\mathbb{F}$. We write the matrix A as its entries, i.e.

$$A = (a_{ij}), \quad 1 \le i \le m, \ 1 \le j \le n$$

We define $M_{m,n}(\mathbb{F})$ to be the set of all such matrices.

Definition 5.9 (Change of basis)**:** Consider two sets of bases for V:

$$B = (v_1, \ldots, v_n), \quad B' = (v'_1, \ldots, v'_n)$$

The change of basis matrix from B' to B is $P = (P_{ij})$ given by $v'_j = \sum_i P_{ij} v_i$, i.e.

$$P = ([v'_1]_B | \ldots | [v'_n]_B) = [\mathrm{Id}]_{B',B}$$

Definition 5.10 (Space of linear maps)**:** Let the space of linear maps between the $\mathbb{F}$-spaces V, W be $\mathcal{L}(V, W)$.

Definition 5.11 (Endomorphisms)**:** Let V be $\mathbb{F}$-space of dimension $n < \infty$, with basis $B = \{v_1, \ldots, v_n\}$. Then, $\alpha \in \mathcal{L}(V) := \mathcal{L}(V, V)$ is defined to be an endomorphism of V, i.e. linear map from V to itself.

5.3   Inner product and adjoint map

Definition 5.12 (Symmetric bilinear form)**:** Let U, V be $\mathbb{F}$-spaces. $\phi : U \times V \to \mathbb{F}$ is a bilinear form if it's linear in both arguments.

$$\phi(u, -) : \ V \to \mathbb{F} \in V^*, \ \forall u \in U$$

$$\phi(-, v) : \ U \to \mathbb{F} \in U^*, \ \forall v \in V$$

A bilinear form ϕ on V is symmetric if $\phi(u, v) = \phi(v, u) \ \forall u, v \in V$.

Definition 5.13 (Hermitian form)**:** Let V and W be $\mathbb{C}$-spaces. A Hermitian form is a function $\phi : V \times W \to \mathbb{C}$ s.t.

$$\phi(\lambda_1 v_1 + \lambda_2 v_2, w) = \overline{\lambda}_1 \phi(v_1, w) + \overline{\lambda}_2 \phi(v_2, w) \ \forall \lambda_i, \mu_i \in \mathbb{C}$$

$$\phi(v, \mu_1 w_1 + \mu_2 w_2) = \mu_1 \phi(v, w_1) + \mu_2 \phi(v, w_2) \ \forall v_i, w_i \in \mathbb{C}^n$$

where $\phi(u, v) = \phi(v, u) \ \forall u, v \in V$. Essentially, linear in second argument, and anti-linear in first argument.

Definition 5.14 (Definiteness)**:** ϕ is symmetric bilinear form on the real vector space V.

- ϕ is positive definite if $\phi(u, u) > 0 \ \forall u \neq 0$.

- ϕ is positive semi-definite if $\phi(u, u) \ge 0 \ \forall u$.

- ϕ is negative definite if $\phi(u, u) < 0 \ \forall u \neq 0$.

- ϕ is negative semi-definite if $\phi(u, u) \le 0$.

- else indefinite.

Similar terminology for Hermitian forms.

Definition 5.15 (Inner product)**:** Let V be a vector space over either $\mathbb{R}$ or $\mathbb{C}$. An inner product (IP) on V is a positive-definite symmetric (Hermitian) form ϕ on V.

Lemma 5.1 (Cauchy-Schwarz inequality)**:**

$$|\langle u, v \rangle| \le ||u|| \, ||v|| \quad \forall u, v \tag{5.3}$$

Proof. $\mathbb{F}$ is $\mathbb{R}$ or $\mathbb{C}$. W.l.o.g. $u \neq 0$. $\forall t \in \mathbb{F}$,

$$0 \le ||tu - v||^2 = |t|^2 ||u||^2 - t \langle u, v \rangle - \overline{t} \overline{\langle u, v \rangle} + ||u||^2$$

Now set $t = \frac{\overline{\langle u, v \rangle}}{||u||^2}$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Definition 5.16 (Orthogonal, orthonormal)**:** Vectors $\{e_1, \ldots, e_k\} \in V$ are orthogonal if $\langle e_i, e_j \rangle = 0 \ \forall i \neq j$, and orthonormal if $\langle e_i, e_j \rangle = \delta_{ij}$.

Lemma 5.2: If $e_1, \ldots, e_k$ are orthogonal and non-zero, then they are linearly independent More-over, if $v = \sum_{i=1}^{k} \lambda_i e_i$, then $\lambda_i = \frac{\langle v, e_i \rangle}{\langle e_i, e_i \rangle}$.

Proof. $\langle v, e_j \rangle = \langle \sum_i \lambda_i e_i, e_j \rangle = \lambda_j \langle e_j, e_j \rangle$ by orthogonality. $\qquad \square$

Theorem 5.1 (Gram-Schmidt)**:** Let V be a IP space and $\{v_1, v_2, \ldots\}$ is a countable set of linearly independent vectors, then there's a sequence $e_1, e_2, \ldots$ of orthonormal eigenvectors s.t. $\text{span}\{v_1, \ldots, v_k\} = \text{span}\{e_1, \ldots, e_k\} \ \forall k$ (be careful of conflicting notation).

Proof. Induct on k. If $k = 1$, we are trivially done. Say we've found $e_1, \ldots, e_k$.

$$e'_{k+1} = v_{k+1} - \sum_{i=1}^{k} \langle v_{k+1}, e_i \rangle e_i$$

Then, $\langle e'_{k+1}, e_i \rangle = 0$ for $i = 1, \ldots, k$ by construction. Clearly,

$$\text{span}\{v_1, \ldots, v_{k+1}\} = \text{span}\{e_1, \ldots, e_k, e'_{k+1}\}$$

We then have $e'_{k+1} \neq 0$ by the linear independence of v_i's. Set $e_{k+1} = \frac{e'_{k+1}}{||e'_{k+1}||}$ and we will obtain an orthonormal set of $k + 1$ vectors from the original $k + 1$ v_i's. $\qquad \square$

Corollary 5.1: Let V be a finite-dimensional IP space. Any orthonormal set of vectors can be extended to an orthonormal basis.

Proof. Suppose $e_1, \ldots, e_k$ is an orthonormal set. They are linearly independent, so can extend to a basis $e_1, \ldots, e_k, v_{k+1}, \ldots, v_n$. Apply Gram-Schmidt. (no effect on the first k vectors). $\qquad \square$

Definition 5.17 (Orthonormal columns)**:** $A \in M_{m,n}(\mathbb{R})$ has orthonormal columns if $A^T A = I$. $A \in M_{m,n}(\mathbb{C})$ has orthonormal columns if $A^T \overline{A} = I$.

Definition 5.18 (Orthogonal, unitary)**:** $A \in M_n(\mathbb{R})$ is orthogonal if $A^T A \iff A^{-1} = A^T$. $A \in M_n(\mathbb{C})$ is unitary if $A^T \overline{A} = I \iff A^{-1} = \overline{A}^T$.

Proposition 5.1: Let V and W be finite-dimensional IP spaces, $\alpha \in \mathcal{L}(V, W)$. There is a unique linear map $\alpha^* : W \to V$ s.t.

$$\langle \alpha(v), w \rangle = \langle v, \alpha^*(w) \rangle \quad \forall v \in V, w \in W$$

If B and C are orthonormal bases for V and W respectively, then $[\alpha^*]_{C,B} = ([\alpha]_{B,C})^T$, where α^* is the adjoint of α.

Definition 5.19 (Hermitian conjugate)**:** $A^\dagger = \overline{A}^T$ is called the Hermitian conjugate.

Definition 5.20 (Self-adjoint map and isometries)**:** Let V be IP space, $\alpha \in \mathcal{L}(V)$ with adjoint α^*.

1. If $\langle \alpha v, w \rangle = \langle v, \alpha w \rangle \ \forall v, w$, then α is self-adjoint and it is called symmetric (Hermitian) if it is over $\mathbb{R}$ ($\mathbb{C}$).

2. If $\langle \alpha v, \alpha w \rangle = \langle v, w \rangle \ \forall v, w$, then α is an isometry and it is called orthogonal (unitary) if it is over $\mathbb{R}$ ($\mathbb{C}$).

Theorem 5.2: α is self-adjoint iff $\alpha^* = \alpha$ while α is an isometry iff $\alpha^* = \alpha^{-1}$.

Proof. For the former, we have $||\alpha v||^2 = ||v||^2$ and α injective, so can define α^{-1}.

$$\forall v, w \quad \langle v, \alpha^*(w) \rangle = \langle \alpha v, w \rangle = \langle \alpha v, \alpha(\alpha^{-1}(w)) \rangle = \langle v, \alpha^{-1}(w) \rangle \implies \alpha^{-1} = \alpha^*$$

For the latter, we have $\langle \alpha v, \alpha w \rangle = \langle v, \alpha^*(\alpha(w)) \rangle = \langle v, w \rangle \ \forall v, w$. Hence, $\alpha^* = \alpha^{-1}$. $\qquad \square$

Lemma 5.3: Let V be a finite-dimensional real (complex) IP space,

1. $\alpha \in \mathcal{L}(V)$ is self-adjoint iff for any orthonormal basis B, $[\alpha]_B$ is symmetric (Hermitian).

2. $\alpha \in \mathcal{L}(V)$ is an isometry iff for any orthonormal basis B, $[\alpha]_B$ is orthogonal (unitary).

Proof. For orthonormal basis B, $[\alpha^*]_B = [\alpha]_B^T$. We have for self-adjoint: $[\alpha]_B = [\alpha]_B^T$, and for isometry: $[\alpha]_B^{-1} = [\alpha]_B^T$. $\qquad\square$

5.4 Diagonalisation

Definition 5.21 (Equivalent matrices)**:** $A, A' \in M_{m,n}(\mathbb{F})$ are equivalent if $A' = Q^{-1}AP$ for some invertible matrices $P \in M_{n,n}(\mathbb{F})$ and $Q \in M_{m,m}(\mathbb{F})$.

Definition 5.22 (Similar/conjugate matrix)**:** For $A, B \in M_n(\mathbb{F})$ to be similar (conjugate), $\exists$ an invertible $P \in M_n(\mathbb{F})$ s.t.

$$A = P^{-1}BP$$

Definition 5.23 (Diagonalisable and triangulable)**:** $\alpha \in \mathcal{L}(V)$ is said to be diagonalisable if $\exists B$ with $[\alpha]_B$ diagonal. α is said to be triangulable if $\exists B$ s.t. $[\alpha]_B$ is upper-triangular. $A \in M_n(\mathbb{F})$ is diagonalizable if conjugate to a diagonal matrix.

Definition 5.24 (Eigenvalue, eigenvector, eigenspace)**:** λ is an eigenvalue of $\alpha \in \mathcal{L}(V)$ if $\exists v \in V \setminus \{0\}$ s.t. $\alpha v = \lambda v$. $v \in V$ is an eigenvector if $\alpha v = \lambda v$ for some $\lambda \in \mathbb{F}$.

$$V_\lambda = \{\, v \in V \mid \alpha v = \lambda v \,\} \le V$$

is called the λ-eigenspace of α.

Definition 5.25 (Characteristic polynomial)**:** The characteristic polynomial of an endomorphism $\alpha \in \mathcal{L}(V)$ is

$$\chi_\alpha(t) = \det(\alpha - t\,\mathrm{Id})$$

but that of a square matrix $A \in M_n(\mathbb{F})$ is

$$\chi_A(t) = \det(A - tI) \tag{5.4}$$

Conjugate matrices have the same characteristic polynomial.

Theorem 5.3: λ is an eigenvalue of A iff $\det(A - \lambda I) = 0$.

Proof. Suppose that λ is an eigenvalue and $\mathbf{x}$ is the associated eigenvector.

$$(A - \lambda I)\mathbf{x} = \mathbf{0} \implies \mathbf{x} \in \mathrm{Ker}(A - \lambda I)$$

But $\mathbf{x} \ne \mathbf{0}$. So $\mathrm{Ker}(A - \lambda I)$ is non-trivial and $\det(A - \lambda I) = 0$. The argument for the converse is similar. $\qquad\square$

Theorem 5.4: Suppose $n \times n$ matrix A has distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then the corresponding eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are linearly independent.

Proof. Suppose $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are linearly dependent, then $\exists d_i \ne 0 \; \forall i = 1, 2 \ldots, r$ so that

$$d_1\mathbf{x}_1 + d_2\mathbf{x}_2 + \cdots + d_r\mathbf{x}_r = \mathbf{0}.$$

Suppose that this is the shortest non-trivial linear combination that gives $\mathbf{0}$ (we may need to re-order $\mathbf{x}_i$). Now apply $(A - \lambda_1 I)$ to the whole equation to obtain

$$d_1(\lambda_1 - \lambda_1)\mathbf{x}_1 + d_2(\lambda_2 - \lambda_1)\mathbf{x}_2 + \cdots + d_r(\lambda_r - \lambda_1)\mathbf{x}_r = \mathbf{0}.$$

Assuming distinct eigenvalues, all terms except the first are not zero, so

$$d_2(\lambda_2 - \lambda_1)\mathbf{x}_2 + \cdots + d_r(\lambda_r - \lambda_1)\mathbf{x}_r = \mathbf{0},$$

and we have found a shorter linear combination that gives $\mathbf{0}$. A contradiction. $\qquad\square$

Theorem 5.5: Let V be $\mathbb{F}$-space with $\dim V < \infty$. Let $\alpha \in \mathcal{L}(V)$, then α is diagonalisable iff $p(\alpha) = 0$ for some polynomial p which is a product of distinct linear factors.

Theorem 5.6 (Simultaneous diagonalisation)**:** Let $\alpha, \beta \in \mathcal{L}(V)$ be diagonalisable. Then α, β are simultaneously diagonalisable iff they commute.

Proof. Assume the eigenbasis of A is also the eigenbasis of B. Let this be $\{\mathbf{e_n}\}$ s.t. $A\mathbf{e_n} = a_n \mathbf{e_n}$ and $B\mathbf{e_n} = b_n \mathbf{e_n}$. Since each eigenvalue is distinct, the eigenvectors are linearly independent, i.e. the only solution to $\mathbf{q} = \sum_{p=1}^{n} \alpha_p \mathbf{e_p} = \mathbf{0}$ is $\alpha_p = 0 \ \forall p$. Consider

$$\mathbf{0} = \prod_{p=1, p\neq s}^{n} (A - a_p I)\mathbf{q} = \prod_{p=1, p\neq s}^{n} (a_s - a_p)\alpha_s \mathbf{e_s}$$

which implies $\alpha_s = 0 \ \forall \mathbf{e_s}$, so the eigenvectors form a basis. Apply $AB - BA$ to an arbitrary vector $\mathbf{v} = \sum_{p=1}^{n} \beta_p \mathbf{e_p}$ written as a linear combination of the basis vectors:

$$(AB - BA)\mathbf{v} = \sum_{p=1}^{n} (a_p b_p - b_p a_p)\beta_p \mathbf{e_p} = \mathbf{0} \implies AB = BA$$

Conversely, if $AB = BA$, then $a_n B\mathbf{e_n} = AB\mathbf{e_n} = BA\mathbf{e_n}$, and so $B\mathbf{e_n}$ is an eigenvector of A with eigenvalue a_n. But all the eigenvalues of A are given to be distinct, so $B\mathbf{e_n}$ is parallel to $\mathbf{e_n}$, $\exists b_n$ s.t. $B e_n = b_n e_n$. Hence, the eigenvectors of A are also the eigenvectors of B. $\qquad\square$

Definition 5.26 (Multiplicity)**:** Given λ to be an eigenvalue of the endomorphism $\alpha \in \mathcal{L}(V)$,

- a_λ, the algebraic multiplicity of λ, is the multiplicity of λ as a root of the characteristic polynomial χ_α;

- $g_\lambda = n(\alpha - \lambda \operatorname{Id})$, the geometric multiplicity of λ, where n is the nullity.

Proposition 5.2: For $\mathbb{F} = \mathbb{C}$, $\alpha \in \mathcal{L}(V)$. The following are equivalent:

1. α diagonalisable.

2. $a_\lambda = g_\lambda \ \forall$ eigenvalues λ of α.

3. $c_\lambda = 1 \ \forall$ eigenvalues λ of α.

5.5 Spectral theory

Definition 5.27 (Spectrum)**:** For finite linear maps, the spectrum is the set of eigenvalues.

Lemma 5.4: Let V be an IP space (with no assumption on its dimension), then if $\alpha \in \mathcal{L}(V)$ is self-adjoint, then

1. α has real eigenvalues;

2. the eigenvectors of α with distinct eigenvalues, are orthogonal.

Proof. Let $\alpha v_i = \lambda_i v_i$ for $v_i \neq 0$ and $\lambda_i \in \mathbb{C}$, then

$$0 = \langle \alpha v_1, v_2 \rangle - \langle v_1, \alpha v_2 \rangle = \lambda_1 \langle v_1, v_2 \rangle - \overline{\lambda}_2 \langle v_1, v_2 \rangle$$

For $v_2 = v_1$, $\langle v_1, v_2 \rangle \neq 0$, and so $0 = \lambda_1 - \overline{\lambda}_2 = \lambda_1 - \overline{\lambda}_1$, and thus $\lambda_i \in \mathbb{R} \ \forall i$. For $v_2 \neq v_1$, then $\lambda_1 - \overline{\lambda}_2 = \lambda_1 - \lambda_2 \neq 0$, and so $\langle v_2, v_1 \rangle = 0$. $\qquad\square$

Theorem 5.7: Let V be a finite-dimensional IP space with $\alpha \in \mathcal{L}(V)$ being self-adjoint, then V has an orthonormal basis of eigenvectors of α.

Proof. We have $\mathbb{F}$ being $\mathbb{R}$ or $\mathbb{C}$. Induct on $n = \dim V$. Trivial when $n = 0$ or $n = 1$. Say $[\alpha]_B = A$ w.r.t. the standard basis. By the fundamental theorem of algebra, $\chi_A(t)$ has a root in $\mathbb{C}$, so A, viewed as a complex matrix, has an eigenvalue. By Lemma 8.7, this eigenvalue is real. For either field, α has an eigenvalue λ with corresponding eigenvector $v_1 \neq 0$.

Let $U = \langle v_1 \rangle^{\perp} \leq V$. We want $\alpha(U) \leq U$. If $u \in U$,

$$\langle \alpha u, v_1 \rangle = \langle u, \alpha v_1 \rangle = \langle u, \lambda v_1 \rangle = \lambda \langle u, v_1 \rangle = 0$$

so $\alpha u \in \langle v_1 \rangle^{\perp}$. $\alpha|_U$ is self-adjoint since α is self-adjoint. Then by induction on $\dim V$, there is an orthonormal basis $\{v_2, \ldots, v_n\}$ of U which are eigenvectors of $\alpha|_U$. Then $\{\frac{v_1}{||v_1||}, v_2, \ldots, v_n\}$ is an orthonormal basis of eigenvectors of α. $\qquad\square$

Corollary 5.2: V is a finite-dimensional IP space. If $\alpha \in \mathcal{L}(V)$ is self-adjoint, then V is the orthogonal direct sum of the eigenspaces of α, i.e. $V = \perp_i V_{\lambda_i}$.

Lemma 5.5: Let V be a complex IP space (with no assumption on its dimension), then if $\alpha \in \mathcal{L}(V)$ is self-adjoint, then

1. α has eigenvalues with modulus 1;

2. the eigenvectors of α with distinct eigenvalues, are orthogonal.

Proof. Let $\alpha v_i = \lambda_i v_i$ for $v_i \neq 0$ and $\lambda_i \in \mathbb{C}$, then

$$0 = \langle \alpha v_1, v_2 \rangle - \langle v_1, \alpha v_2 \rangle = \lambda_1 \langle v_1, v_2 \rangle - \overline{\lambda}_2^{-1} \langle v_1, v_2 \rangle$$

For $v_2 = v_1$, $\langle v_1, v_2 \rangle \neq 0$, and so $0 = \lambda_1 - \overline{\lambda}_2^{-1} = \lambda_1 - \overline{\lambda}_1^{-1} \implies |\lambda_i| = 1$, $\forall i$. For $v_2 \neq v_1$, then $\lambda_1 - \overline{\lambda}_2^{-1} = \lambda_1 - \lambda_2 \neq 0$, and so $\langle v_2, v_1 \rangle = 0$. $\qquad\square$

Theorem 5.8: Let V be a finite-dimensional complex IP space with $\alpha \in \mathcal{L}(V)$ being unitary, then V has an orthonormal basis of eigenvectors of α.

Proof. Similar to Theorem 8.3. By the fundamental theorem of algebra, $\chi_A(t)$ has a root in $\mathbb{C}$, so A, viewed as a complex matrix, has an eigenvalue. Fix $v_1 \in V \backslash \{0\}$ s.t. $\alpha v_1 = \lambda v_1$ and $||v_1|| = 1$.

Let $U = \langle v_1 \rangle^{\perp} \leq V$. We want $\alpha(U) \leq U$. If $u \in U$,

$$\langle \alpha u, v_1 \rangle = \langle u, \alpha^{-1} v_1 \rangle = \langle u, \overline{\lambda}^{-1} v_1 \rangle = \overline{\lambda}^{-1} \langle u, v_1 \rangle = 0$$

so $\alpha u \in \langle v_1 \rangle^{\perp}$. $\alpha|_U$ is self-adjoint since α is self-adjoint. Then by induction on $\dim V$, there is an orthonormal basis $\{v_2, \ldots, v_n\}$ of U which are eigenvectors of $\alpha|_U$. Then $\{v_1, v_2, \ldots, v_n\}$ is an orthonormal basis of eigenvectors of α. $\qquad\square$

Theorem 5.9: An $n \times n$ Hermitian matrix has precisely n orthogonal eigenvectors.

Proof. Let $\lambda_1, \lambda_2, \ldots, \lambda_r$ be the distinct eigenvalues of H ($r \leq n$), with a set of corresponding orthonormal eigenvectors $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$. Extend to a basis of the whole of $\mathbb{C}^n$ and use Gram-Schmidt to create an orthonormal basis

$$\tilde{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r, \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{n-r}\}.$$

Now write it in a form of an unitary matrix P.

$$P = \begin{pmatrix} \uparrow & \uparrow & & \uparrow & \uparrow & & \uparrow \\ \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_r & \mathbf{u}_1 & \ldots & \mathbf{u}_{n-r} \\ \downarrow & \downarrow & & \downarrow & \downarrow & & \downarrow \end{pmatrix}$$

If we change basis, we have

$$P^{-1}HP = P^{\dagger}HP = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \ldots & \lambda_r & 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & c_{11} & c_{12} & \ldots & c_{1,n-r} \\ 0 & 0 & \ldots & 0 & c_{21} & c_{22} & \ldots & c_{2,n-r} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & c_{n-r,1} & c_{n-r,2} & \ldots & c_{n-r,n-r} \end{pmatrix}$$

Here C is an $(n-r) \times (n-r)$ Hermitian matrix. The eigenvalues of C are also eigenvalues of H because $\det(H - \lambda I) = \det(P^{\dagger}HP - \lambda I) = (\lambda_1 - \lambda) \ldots (\lambda_r - \lambda) \det(C - \lambda I)$. We can keep repeating the process on C until we finish all rows. If the eigenvalues of C are all distinct, there are $n - r$ orthonormal eigenvectors $\mathbf{w}_j$ (for $j = r + 1, \ldots, n$) of C. Otherwise, use Gram-Schmidt. Let

$$Q = \begin{pmatrix} 1 & & & & & & & \\ & 1 & & & & & & \\ & & \ddots & & & & & \\ & & & 1 & & & & \\ & & & & \uparrow & \uparrow & & \uparrow \\ & & & & \mathbf{w}_{r+1} & \mathbf{w}_{r+2} & \ldots & \mathbf{w}_n \\ & & & & \downarrow & \downarrow & & \downarrow \end{pmatrix}$$

with other entries 0 (where we have a $r \times r$ identity matrix block on the top left corner and a $(n - r) \times (n - r)$ with columns formed by $\mathbf{w}_j$). Since the columns of Q are orthonormal, Q is unitary. So $Q^{\dagger}P^{\dagger}HPQ = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r, \lambda_{r+1}, \ldots, \lambda_n)$, where the first r λs are distinct and the remaining ones are copies of previous ones. The n linearly-independent eigenvectors are the columns of PQ. So it now follows that H is diagonalizable via a transformation PQ.   □

5.6   Quadrics

Definition 5.28 (Quadratic form)**:** A map $Q: V \to \mathbb{F}$ is a quadratic form if there is a bilinear form $\phi: V \times V \to \mathbb{F}$ s.t. $Q(u) = \phi(u, u) \; \forall u \in V$.

Example 5.1. *Let $V = \mathbb{R}^2$, then*

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} := \mathbf{x}^T A \mathbf{x} = ax^2 + (b+c)xy + dy^2$$

Wouldn't change if we replace A with $\frac{1}{2}(A + A^T)$.

Theorem 5.10: Hermitian forms are real.

Proof. $(\mathbf{x}^\dagger H \mathbf{x})^* = (\mathbf{x}^\dagger H \mathbf{x})^\dagger = \mathbf{x}^\dagger H^\dagger \mathbf{x} = \mathbf{x}^\dagger H \mathbf{x} \in \mathbb{R}$. By Theorem 4.10, any Hermitian matrix can be diagonalized with a unitary transformation. So $F(\mathbf{x}) = \mathbf{x}^\dagger H \mathbf{x} = \mathbf{x}^\dagger U D U^\dagger \mathbf{x}$. Write $\mathbf{x}' = U^\dagger \mathbf{x}$. So $F = (\mathbf{x}')^\dagger D \mathbf{x}'$, where $D = \text{diag}(\lambda_1, \ldots, \lambda_n)$. We know that $\mathbf{x}'$ is the vector $\mathbf{x}$ relative to the eigenvector basis. So $F(\mathbf{x}) = \sum_{i=1}^n \lambda_i |x_i'|^2 \in \mathbb{R}$. $\qquad\square$

Definition 5.29 (Principal axes)**:** The eigenvectors of A are known as the principal axes.

Example 5.2. *Take $F = 2x^2 - 4xy + 5y^2 = \mathbf{x}^T S \mathbf{x}$, where $\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}$ and $S = \begin{pmatrix} 2 & -2 \\ -2 & 5 \end{pmatrix}$. We can always choose the matrix to be symmetric since for any antisymmetric A, we have $\mathbf{x}^\dagger A \mathbf{x} = 0$. The eigenvalues are $1, 6$ with corresponding eigenvectors $\frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ -2 \end{pmatrix}$. Now change basis with*

$$Q = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 1 \\ 1 & -2 \end{pmatrix}$$

Then $\mathbf{x}' = Q^T \mathbf{x} = \frac{1}{\sqrt{5}} \begin{pmatrix} 2x + y \\ x - 2y \end{pmatrix}$. Then $F = (x')^2 + 6(y')^2$. So $F = c$ is an ellipse.

Definition 5.30 (Quadric)**:** A quadric is an n-dimensional surface defined by the zero of a real quadratic polynomial, i.e.
$$\mathbf{x}^T A \mathbf{x} + \mathbf{b}^T \mathbf{x} + c = 0,$$
where A is a real $n \times n$ matrix, $\mathbf{x}, \mathbf{b}$ are n-dimensional column vectors and c is a constant scalar.

Proposition 5.3: We can transform a quadric form to a quadratic form.

Proof. Retaining the symmetric part $S = (A + A^T)/2$ gives

$$\mathbf{x}^T S \mathbf{x} + \mathbf{b}^T \mathbf{x} + c = 0$$

By Proposition 4.9, since S is real and symmetric, we can diagonalize it using $S = QDQ^T$ with D diagonal. We write $\mathbf{x}' = Q^T \mathbf{x}$ and $\mathbf{b}' = Q^T \mathbf{b}$. So we have

$$(\mathbf{x}')^T D \mathbf{x}' + (\mathbf{b}')^T \mathbf{x}' + c = 0.$$

If S is invertible, i.e. with no zero eigenvalues, then write $\mathbf{x}'' = \mathbf{x}' + \frac{1}{2} D^{-1} \mathbf{b}'$ which shifts the origin to eliminate the linear term $(\mathbf{b}')^T \mathbf{x}'$ and finally have the required qudratic form after dropping the prime superscripts. $\qquad\square$

Example 5.3. *Consider $n = 2$, then $\lambda_1 x_1^2 + \lambda_2 x_2^2 = k$. If*

1. $\lambda_1 \lambda_2 > 0$: *we have ellipses with axes coinciding with eigenvectors of S. (We require $\text{sgn}(k) = \text{sgn}(\lambda_1, \lambda_2)$, or else we would have no solutions at all)*

2. $\lambda_1 \lambda_2 < 0$: *say $\lambda_1 = k/a^2 > 0$, $\lambda_2 = -k/b^2 < 0$. So we obtain $\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = 1$, which is a hyperbola.*

3. $\lambda_1 \lambda_2 = 0$: *Say $\lambda_2 = 0$, $\lambda_1 \neq 0$. Note that in this case, our symmetric matrix S is not invertible and we cannot shift our origin using as above. From our initial equation,*

$$\lambda_1 (x_1')^2 + b_1' x_1' + b_2' x_2' + c = 0.$$

We perform the coordinate transform (which is simply completing the square!)

$$x_1'' = x_1' + \frac{b_1'}{2\lambda_1}, \quad x_2'' = x_2' + \frac{c}{b_2'} - \frac{(b_1')^2}{4\lambda_1 b_2'}$$

to remove the x_1' and constant term. Dropping the primes, we have $\lambda_1 x_1^2 + b_2 x_2 = 0$, which is a parabola. Note that we assumed $b_2' \neq 0$. If $b_2' = 0$, we have $\lambda_1 (x_1')^2 + b_1' x_1' + c = 0$. If we solve this quadratic for x_1', we obtain 0, 1 or 2 solutions for x_1 (straight line(s)).

Definition 5.31 (Focus-directrix property): Let the foci of a conic section be $(\pm ae, 0)$ and the directrices be $x = \pm a/e$, where a and e are the scale and eccentricity of the conic section.

This conic section is the set of points whose distance from focus is $e\times$ distance from directrix which is closer to that of focus (unless $e = 1$, where we take the distance to the other directrix).

Now consider the different cases of e:

1. $e < 1$. By definition,

$$\sqrt{(x - ae)^2 + y^2} = e\left(\frac{a}{e} - x\right) \implies \frac{x^2}{a^2} + \frac{y^2}{a^2(1 - e^2)} = 1$$

which is an ellipse with semi-major axis a and semi-minor axis $a\sqrt{1 - e^2}$. (if $e = 0$, then we have a circle).

2. $e > 1$: we have hyperbola

$$\sqrt{(x - ae)^2 + y^2} = e\left(x - \frac{a}{e}\right) \implies \frac{x^2}{a^2} - \frac{y^2}{a^2(e^2 - 1)} = 1$$

and we have a hyperbola.

3. $e = 1$: we have parabola

$$\sqrt{(x - a)^2 + y^2} = (x + 1) \implies y^2 = 4ax$$

Conics also work in polar coordinates (r, θ) centred on one of the foci. Let l be s.t. l/e is the distance from the focus to the directrix, i.e. $l = a|1 - e^2|$. The focus-directrix property gives

$$r = e\left(\frac{l}{e} - r\cos\theta\right) \implies r = \frac{l}{1 + e\cos\theta}$$

We see that $r \to \infty$ if $\theta \to \cos^{-1}(-1/e)$, which is only possible if $e \geq 1$, i.e. hyperbola or parabola. But ellipses have $e < 1$. So r is bounded, as expected.

Example 5.4. *Consider $n = 3$, then $\lambda_1 x^2 + \lambda_2 y^2 + \lambda_3 z^2 = k$. The possible cases are:*

- $\lambda_i > 0 \; \forall i$ *and $k > 0$: ellipsoid;*

- $\lambda_1 = \lambda_2 > 0$, $\lambda_3 > 0$ *and $k > 0$: spheroid with z being rotational symmetry axis;*

- $\lambda_1 = \lambda_2 = \lambda_3 > 0$ *and $k > 0$: sphere;*

- $\lambda_1, \lambda_2 > 0$, $\lambda_3 = 0$ *and $k > 0$: elliptic cylinder about z axis;*

- $\lambda_1, \lambda_2 > 0$, $\lambda_3 < 0$ *and $k > 0$: hyperboloid of one sheet;*

- $\lambda_1 > 0$, $\lambda_2, \lambda_3 < 0$ *and $k > 0$: hyperboloid of two sheets;*

- $\lambda_1 > 0$, $\lambda_2 = \lambda_3 = 0$ *and $k \geq 0$: plane.*

5.7   Stationary property of eigenvalues

Definition 5.32 (Rayleigh quotient)**:** The Rayleigh quotient associated with a Hermitian matrix A is the normalized Hermitian form

$$\lambda(x) = \frac{x^T A x}{x^T x} \in \mathbb{R} \tag{5.5}$$

with $\lambda(\alpha x) = \lambda(x)$ s.t. if x is the eigenvector of A then $\lambda(x)$ is the corresponding eigenvalue.

Proposition 5.4: The eigenvalues of A are the stationary values of the function $\lambda(x)$.

Proof.

$$\begin{aligned}
\delta\lambda &= \lambda(x + \delta x) - \lambda(x) \\
&= \frac{(x + \delta x)^\dagger A(x + \delta x)}{(x + \delta x)^\dagger (x + \delta x)} - \lambda(x) \\
&= \frac{x^\dagger A x + (\delta x)^\dagger A x + x^\dagger A(\delta x) + O(\delta x^2)}{x^\dagger x + (\delta x)^\dagger x + x^\dagger(\delta x) + O(\delta x^2)} - \lambda(x) \\
&= \frac{(\delta x)^\dagger A x + x^\dagger A(\delta x) - \lambda(x)[(\delta x)^\dagger x + x^\dagger(\delta x)] + O(\delta x^2)}{x^\dagger x + (\delta x)^\dagger x + x^\dagger(\delta x) + O(\delta x^2)} \\
&= \frac{(\delta x)^\dagger (A x - \lambda(x)x) + (x^\dagger A - \lambda(x)x^\dagger)(\delta x)}{x^\dagger x} + O(\delta x^2)
\end{aligned}$$

The first-order variation of $\lambda(x)$ vanishes $\forall$ perturbations $\delta(x)$ iff $Ax = \lambda(x)x \iff x^\dagger A = \lambda(x)x^\dagger$. $\qquad\square$

See Rayleigh-Ritz variational principle later.

6   Elementary analysis

Idea of convergence and limits. O notation. Statement of Taylor's theorem with discussion of remainder. Convergence of series; comparison and ratio tests. Power series of a complex variable; circle of convergence. Analytic functions: Cauchy-Riemann equations, rational functions and $\exp(z)$. Zeros, poles and essential singularities.   **[3]**

6.1   Recap

Proposition 6.1 (Cauchy's principle of convergence)**:** A necessary and sufficient condition for the series $\sum_n u_n$ to converge is that, for any positive number ε,

$$|S_{N+m} - S_N| = |u_{N+1} + \cdots + u_{N+m}| < \varepsilon$$

$\forall m \in \mathbb{Z}^+$ and for sufficiently large N, where $S_N = \sum_{n=n_0}^{N} u_n$ is the partial sum.

Definition 6.1 (Absolute and conditional convergence)**:** If $\sum_n |u_n|$ converges, then $\sum_n u_n$ also converges. In this case, $\sum_n u_n$ is said to converge absolutely. If $\sum_n |u_n|$ diverges, then $\sum_n u_n$ may or may not converge. If it does, it is said to converge conditionally.

Proposition 6.2: A necessary condition for $\sum_n u_n$ to converge is that $u_n \to 0$ as $n \to \infty$.

Example 6.1. *Harmonic series satisfy this condition but still diverges. This is a necessary but not sufficient condition.*

Proposition 6.3 (Ratio test)**:** Consider the ratio of successive terms,

$$r_n = \left| \frac{u_{n+1}}{u_n} \right| \tag{6.1}$$

Suppose $r_n \to L$ as $n \to \infty$, then if

- $L < 1$: $\sum_n u_n$ converges absolutely;

- $L > 1$: $\sum_n u_n$ diverges (u_n does not tend to 0)

- $L = 1$: a different test is required.

Definition 6.2 (Limits)**:** $f(z)$ tends to the limit L as $z \to z_0$ if for any $\varepsilon > 0$, $\exists \delta > 0$ s.t. $|f(z) - L| < \varepsilon \; \forall z$ s.t. $|z - z_0| < \delta$. Write as $\lim_{z \to z_0} f(z) = L$.

Definition 6.3 (Continuity)**:** $f(z)$ is continuous at $z = z_0$ if $f(z) \to f(z_0)$ as $z \to z_0$.

Definition 6.4 (Order parameters)**:**

1. "$f(x) = o(g(x))$ as $x \to x_0$" if $\lim\limits_{x \to x_0} \frac{f(x)}{g(x)} = 0$. Intuitively, $f(x)$ is much smaller than $g(x)$.

2. "$f(x) = O(g(x))$ as $x \to x_0$" if $\frac{f(x)}{g(x)}$ is bounded as $x \to x_0$. Intuitively, $f(x)$ is about as big as $g(x)$.

   Note that for $f(x) = O(g(x))$ to be true, $\lim\limits_{x \to x_0} \dfrac{f(x)}{g(x)}$ need not exist.

Usually, x_0 is either 0 or infinity. Clearly, we have $f(x) = o(g(x))$ implies $f(x) = O(g(x))$.

Note $o(g(x))$ itself is not a function but represents a class of functions (namely functions that satisfy the property above), and we are saying that f is in this class.

Example 6.2.

- $x = o(\sqrt{x})$ *as* $x \to 0$ *and* $\sqrt{x} = o(x)$ *as* $x \to \infty$.

- $\sin 2x = O(x)$ *as* $x \to 0$ *as* $\sin \theta \approx \theta$ *for small* θ.

- $\sin 2x = O(1)$ *as* $x \to \infty$ *even though the limit does not exist.*

Theorem 6.1 (Taylor's series)**:** For n-times differentiable f, we have

$$f(x + h) = f(x) + hf'(x) + \frac{h^2}{2!}f''(x) + \cdots + \frac{h^n}{n!}f^{(n)}(x) + E_n \tag{6.2a}$$

where $E_n = o(h^n)$ as $h \to 0$. If $f^{(n+1)}$ exists, then $E_n = O(h^{n+1})$. Equivalently, we can write it as

$$f(x) = f(x_0) + (x - x_0)f'(x_0) + \cdots + \frac{(x - x_0)^n}{n!}f^{(n)}(x_0) + E_n \tag{6.2b}$$

When the limit as $n \to \infty$ is taken, the Taylor series of $f(x)$ about the point $x = x_0$ is obtained.

**Theorem 6.2** (L'Hôpital's rule)**:** Let $f(x)$ and $g(x)$ be differentiable at x_0, and $\lim\limits_{x \to x_0} f(x) = \lim\limits_{x \to x_0} g(x) = 0$. Then

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = \lim_{x \to x_0} \frac{f'(x)}{g'(x)} \tag{6.3}$$

6.2 Analytic complex functions

Definition 6.5 (Extended complex plane)**:** The extended complex plane is defined as $\mathbb{C}^* := \mathbf{C} \cup \{\infty\}$.

Definition 6.6 (Riemann sphere)**:** We can reach the point at ∞ by going off in any direction, and all are equivalent. Conceptually, this is a sphere on $\mathbb{C}$ with south pole S at $z = 0$. For any point $z \in \mathbb{C}$, draw a line through north pole N to z and note where the line intersects the sphere. This specifies an equivalent point P on the sphere. ∞ is equivalent to N.
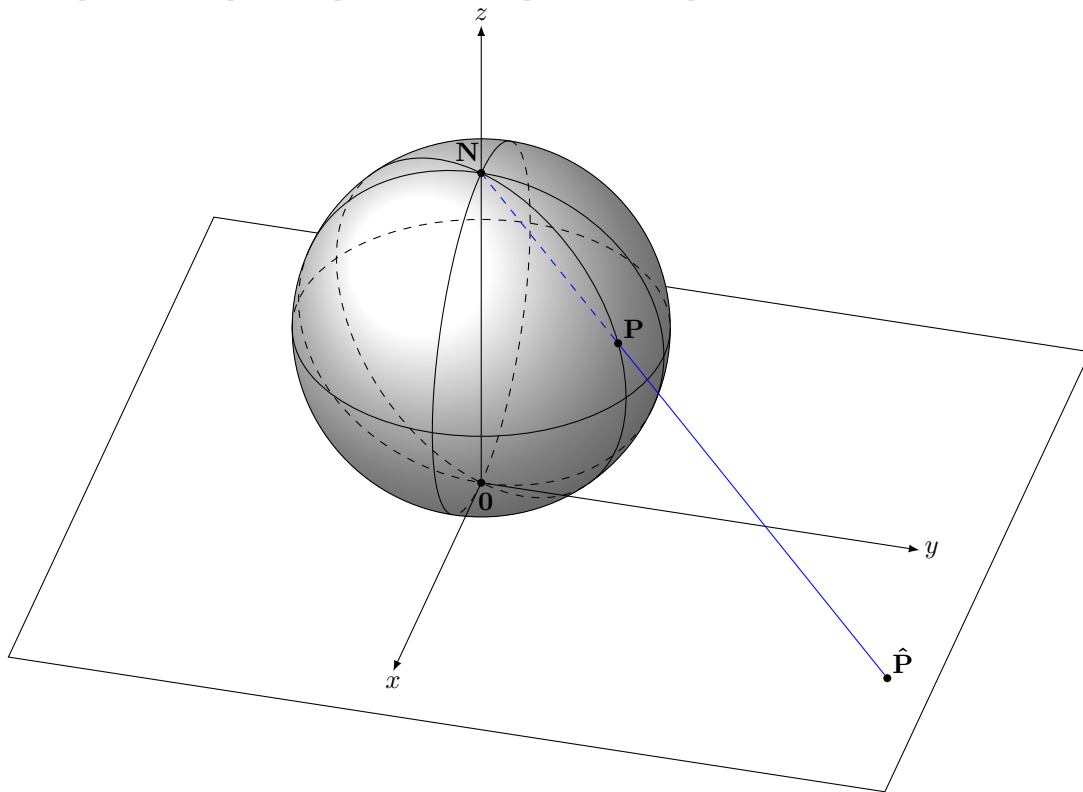
Remarks 6.1. *To investigate the properties of ∞, use the substitution $\xi = \frac{1}{z}$. A function $f(z)$ has a particular property at $z = \infty$ if $f(1/\xi)$ has that property at $\xi = 0$.*

6.2.1 Complex differentiation

Definition 6.7 (Complex differentiability)**:** A complex function $f(z)$ is complex differentiable if the limit

$$f'(z) = \lim_{\delta z \to 0} \frac{f(z + \delta z) - f(z)}{\delta z} \tag{6.4}$$

exists and is independent of the direction of approach. $f'(z)$ is its complex derivative.

Definition 6.8 (Analytic)**:** We say f is analytic at z if $\exists$ a neighbourhood of z throughout which f' exists.

Definition 6.9 (Entire)**:** A function is entire if it is analytic throughout $\mathbb{C}$.

Definition 6.10 (Singularity)**:** A singularity of f is a point at which it is not analytic, or not even defined.

Remarks 6.2. *Analyticity is a strong property with two consequences:*

1. *If a function is analytic, then it is differentiable infinitely many number of times (not true for real functions).*

2. *A bounded entire function is constant.*

Proposition 6.4: Suppose $f(z) = u(x, y) + iv(x, y)$ is differentiable at $z = x + iy$, where $x, y \in \mathbb{R}$ and $u(x, y), v(x, y)$ are real-valued functions, then f must satisfy the Cauchy-Riemann equations.

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \tag{6.5}$$

Proof. Take δz to be δx and $i\delta y$ separately, and show $f'(z)$ has the same value regardless of approach.

$$\lim_{\delta x \to 0} \frac{f(z + \delta x) - f(z)}{\delta x} = \lim_{\delta x \to 0} \frac{u(x + \delta x, y) + iv(x + \delta x, y) - ((u(x, y) + iv(x, y)))}{\delta x} = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}$$

$$\lim_{\delta y \to 0} \frac{f(z + i\delta y) - f(z)}{i\delta y} = \lim_{\delta y \to 0} \frac{u(x, y + i\delta y) + iv(x, y + i\delta y) - ((u(x, y) + iv(x, y)))}{i\delta y} = -i\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}$$

Comparing them gives the Cauchy-Riemann equations. Denote it as CRE. $\qquad\square$

Remarks 6.3. *For the converse to be true, i.e. function satisfying CRE must be differentiable, we require the partial derivatives u_x, u_y, v_x, v_y to be differentiable in both x and y together.*

Example 6.3.

1. $f(z) = e^z = e^x(\cos y + i\sin y)$ *is entire.* $\frac{\partial u}{\partial x} = e^x \cos y = \frac{\partial v}{\partial y}$, $\frac{\partial u}{\partial y} = -e^x \sin y = -\frac{\partial v}{\partial x}$ *and hence* $f'(z) = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = e^x \cos y + ie^x \sin y = e^z$ *as expected.*

2. *If $f(z)$ is analytic and $f(z_0) \neq 0$, then $\frac{1}{f(z)}$ is analytic at z_0, so $\frac{1}{z}$ is analytic except at $z = 0$.*

3. *Any rational function $f(z) = \frac{P(z)}{Q(z)}$, where $P(z)$ and $Q(z)$ are polynomials, is analytic except at points where $Q(z) = 0$. For instance, $\frac{z}{z^2+1}$ is analytic except at $z = \pm i$.*

4. *Many standard real functions can be extended naturally to complex functions and obey the usual rules for their derivatives. They include polynomials, trigonometric functions, logarithmic functions and hyperbolic functions.*

5. *Product/composition of two analytic functions is also analytic. May use product/quotient/chain rules.*

Remarks 6.4. *It is sometimes useful to write the complex trigonometric functions as*

$$\sin z = \sin(x + iy) = \sin x \cos(iy) + \cos x \sin(iy) = \sin x \cosh y + i \cos x \sinh y$$

Example 6.4. *Examples of non-analytic functions are*

1. $f(z) = |z|$ *has $u = \sqrt{x^2 + y^2}$, $v = 0$ and do not satisfy CRE, so $|z|$ is nowhere analytic. But, let $f(z) = |z|^2 = x^2 + y^2$, then the CRE is satisfied only at the origin $(2x = 0, 2y = 0)$, so the function is only differentiable at 0. $|z|^2$ is still nowhere analytic since no neighbourhood of 0 throughout which f is differentiable.*

2. $f(z) = \bar{z} = x - iy$ *has $u = x$, $v = -y$ and again do not satisfy CRE. So $\bar{z}$ is nowhere analytic even if z is analytic.*

6.2.2 Harmonic functions

Definition 6.11 (Harmonic function)**:** A function satisfying Laplace's equation in some two-dimensional domain is said to be harmonic there.

Proposition 6.5: If $f(z) = u + iv$ is analytic, then it is harmonic.

Proof. Since f is analytic, it satisfies CRE.

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial}{\partial x}\frac{\partial v}{\partial y} = \frac{\partial}{\partial y}\frac{\partial v}{\partial x} = \frac{\partial}{\partial y}\left(-\frac{\partial u}{\partial y}\right) = -\frac{\partial^2 u}{\partial y^2}$$

And thus u is a solution to Laplace's equation and by definition, it is harmonic. Similar for v. $\square$

Definition 6.12 (Harmonic conjugate)**:** u, v are said to be harmonic conjugates if they satisfy CRE, i.e. if we know one function then we know the other, up to a constant.

Example 6.5. *Consider* $u(x,y) = x^2 - y^2$ *(harmonic),* v *satisfies* $\frac{\partial v}{\partial y} = \frac{\partial u}{\partial x} = 2x \implies v = 2xy + g(x)$, *so* $-2y = \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} = -2y - g'(x)$. *So* $g(x)$ *is a constant, say* α. *Then the corresponding analytic function whose real part is* u, *is thus*

$$f(z) = x^2 - y^2 + 2ixy + i\alpha = (x + iy)^2 + i\alpha = z^2 + i\alpha$$

Remarks 6.5. *If the domain in which the functions are harmonic is not simply connected, then this method might give a solution that is multi-valued. For instances, if* $u = \frac{1}{2}\ln|x^2 + y^2|$, *which is harmonic on* $|z| \neq 0$, *the corresponding* $f(z) = \ln z$, *which is multi-valued. See Section 1.4.*

Proposition 6.6: Contours of harmonic conjugate functions are perpendicular to each other.

Proof. The gradient $\boldsymbol{\nabla} u = (\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y})^T$ is perpendicular to contours of u (curves of constant u). Similar for v, then

$$\boldsymbol{\nabla} u \cdot \boldsymbol{\nabla} v = \frac{\partial u}{\partial x}\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\frac{\partial v}{\partial y} = \frac{\partial u}{\partial x}\left(-\frac{\partial u}{\partial y}\right) + \frac{\partial u}{\partial y}\frac{\partial u}{\partial x} = 0$$

where we used the CRE, and so the result follows. $\square$

Proposition 6.7:

1. Every stationary point of u is a stationary point of v and vice-versa.

2. Stationary points for which $\begin{vmatrix} \partial_{xx}u & \partial_{xy}u \\ \partial_{yx}u & \partial_{yy}u \end{vmatrix} \neq 0$ must be saddle points.

Proof.

1. If we have a stationary point of u, then $\frac{\partial u}{\partial x} = \frac{\partial u}{\partial y} = 0$. By CRE, then $\frac{\partial v}{\partial x} = \frac{\partial v}{\partial y} = 0$, and hence a stationary point of v.

2. Evaluating the Hessian, we have

$$\frac{\partial}{\partial x}\frac{\partial u}{\partial x} \times \frac{\partial}{\partial y}\frac{\partial u}{\partial y} - \left(\frac{\partial^2 u}{\partial x \partial y}\right)^2 = -\left(\frac{\partial^2 v}{\partial x \partial y}\right)^2 - \left(\frac{\partial^2 u}{\partial x \partial y}\right)^2 < 0$$

   For a function of 2 variables $u(x, y)$, if the determinant of the Hessian is negative at this point, this point is a saddle point.

   $\square$

Corollary 6.1: If $f(z) = u(x,y) + iv(x,y)$ and $g(z) = s(x,y) + it(x,y)$ are analytic functions, then so is $g(f(z))$, and hence deduce that $s(u(x,y), v(x,y))$ satisfies Laplace's equation.

Proof. We have $f(x+iy) = u(x,y) + iv(x,y)$ to be analytic, then f satisfies the CRE, i.e. $(\frac{\partial u}{\partial x})_y = (\frac{\partial v}{\partial y})_x$ and $(\frac{\partial u}{\partial y})_x = -(\frac{\partial v}{\partial x})_y$. Similarly, $g(u+iv) = s(u,v) + it(u,v)$ is analytic, this gives $(\frac{\partial s}{\partial v})_u = (\frac{\partial t}{\partial v})_u$ and $(\frac{\partial s}{\partial v})_u = -(\frac{\partial t}{\partial u})_v$. Consider the composite function

$$g(f(z)) = s(u(x,y), v(x,y)) + it(u(x,y), v(x,y))$$

then we have by chain rule,

$$\left(\frac{\partial s}{\partial x}\right)_y = \left(\frac{\partial u}{\partial x}\right)_y\left(\frac{\partial s}{\partial u}\right)_v + \left(\frac{\partial v}{\partial x}\right)_y\left(\frac{\partial s}{\partial v}\right)_u = \left(\frac{\partial v}{\partial y}\right)_x\left(\frac{\partial t}{\partial v}\right)_u + \left(\frac{\partial t}{\partial u}\right)_v\left(\frac{\partial u}{\partial y}\right)_x = \left(\frac{\partial t}{\partial y}\right)_x$$

$$\left(\frac{\partial s}{\partial y}\right)_x = \left(\frac{\partial u}{\partial y}\right)_x\left(\frac{\partial s}{\partial u}\right)_v + \left(\frac{\partial v}{\partial y}\right)_x\left(\frac{\partial s}{\partial v}\right)_u = -\left(\frac{\partial v}{\partial x}\right)_y\left(\frac{\partial t}{\partial v}\right)_u - \left(\frac{\partial t}{\partial u}\right)_v\left(\frac{\partial u}{\partial x}\right)_y = \left(\frac{\partial t}{\partial x}\right)_y$$

Hence composite functions obey the CRE and $s(u,v)$ is thus harmonic and also satisfy the Laplace's equation. $\square$

6.2.3 Zeros

Definition 6.13 (Zero): The zeros of an analytic function f are the points z_0 where $f(z_0) = 0$. A zero is of order N if, in its Taylor series $\sum_n a_n(z - z_0)^n$, the first non-zero coefficient is a_N. Alternatively, it is of order N if $0 = f(z_0) = f'(z_0) = \dots f^{(N-1)}(z_0)$ but $f^{(N)}(z_0) \neq 0$. In particular, a zero of order 1 is called a simple zero.

Example 6.6.

1. $z^3 + iz^2 + z + i = (z - i)(z + i)^2$ *has a simple zero at $z = i$ and a zero of order 2 at $z = -i$.*

2. $\sinh z$ *has zeroes where* $\frac{e^z - e^{-z}}{2} = 0 \iff e^{2z} = 1 \iff z = n\pi i, n \in \mathbb{Z}$. *The zeros are all simple since its derivative* $\cosh(n\pi i) = \cos(n\pi) \neq 0$.

3. *Since $\sinh z$ has a simple zero at $z = \pi i$, $\sinh^3 z$ has a zero of order 3 there. If needed, we can find its Taylor series about πi by visiting $\xi = z - \pi i$:*

$$\sinh^3 z = \sinh^3(\xi + \pi i) = (-\sinh\xi)^3 = -(\xi + \frac{1}{3!}\xi^3 + \dots)^3 = -(z - \pi i)^3 - \frac{1}{2}(z - \pi i)^5 - \dots$$

   where we used the standard Taylor series for $\sinh z$.

6.2.4 Laurent series

If f might have a singularity at z_0, then we cannot necessarily expect it to have a Taylor series there.

Definition 6.14 (Laurent series): If f is analytic in an annulus $R_1 < |z - z_0| < R_2$, then it has a Laurent series about z_0, of the form

$$f(z) = \sum_{n=-\infty}^{\infty} a_n(z - z_0)^n \tag{6.6}$$

which is convergent within the annulus.

Remarks 6.6.

1. *We can show that Laurent series for f about a particular z_0 is unique with any given annulus*

2. *Taylor series is just a special case of Laurent series, with $a_n = 0 \;\forall n < 0$ and $R_1 = 0$. The radius of convergence of a Taylor series is always the distance to the nearest singularity.*

Example 6.7.

1. $\frac{e^z}{z^3}$ *has a Laurent series about $z_0 = 0$ given by*

$$\frac{e^z}{z^3} = \sum_{m=0}^{\infty} \frac{z^{m-3}}{m!} = \sum_{n=3}^{\infty} \frac{z^n}{(n+3)!}$$

   so $a_n = \frac{1}{(n+3)!}$ for $n \geq -3$.

2. $e^{1/z}$ *has a Laurent series about* $z_0 = 0$ *given by*

$$e^{1/z} = 1 + \frac{1}{z} + \frac{1}{2!z^2} + \frac{1}{3!z^3}$$

so $a_n = \frac{1}{(-n)!}$ *for* $n \leq 0$.

3. $f(z) = \frac{e^z}{z^2-1}$ *has a singularity at* $z_0 = 1$ *but is analytic in the annulus* $0 < |z - z_0| < 2$ *(since the only singularities are at* ± 1*). Write everything in* $\xi = z - z_0$, *so* $f(z)$ *is*

$$\frac{e^\xi e^{z_0}}{\xi(\xi+2)} = \frac{e^{z_0}}{2\xi} e^\xi (1+0.5\xi)^{-1} = \frac{e^{z_0}}{2\xi}(1+\xi+\frac{1}{2!}\xi^2+\dots)(1-0.5\xi+0.25\xi^2+\dots) = \frac{e^{z_0}}{2\xi}(1+0.5\xi+0.25\xi^2+\dots)$$

which gives $\frac{1}{2}e^{z_0}(\frac{1}{z-z_0} + \frac{1}{2} + \frac{z-z_0}{4} + \dots)$. *Hence* $a_{-1} = \frac{1}{2}e$, $a_0 = \frac{1}{4}e$, *etc. This series is valid in the whole annulus (our expansion of* $(1+0.5\xi)^{-1}$ *was valid for* $|0.5\xi| < 1$, *i.e.* $|z - z_0| < 2$*).*

4. *If* $f(z) = \frac{1}{z-a}$ *for* $a \in \mathbb{C}$, *then it is analytic in* $|z| < |a|$, *so it has a Taylor series about* $z_0 = 0$ *given by*

$$\frac{1}{z-a} = -\frac{1}{a}\left(1 - \frac{z}{a}\right)^{-1} = -\sum_{n=0}^{\infty} a^{-n-1} z^n$$

where the binomial expansion is valid since $|\frac{z}{a}| < 1$. *In* $|z| > |a|$, *it has a Laurent series in the annulus* $|a| < |z| < \infty$, *given by*

$$\frac{1}{z-a} = \frac{1}{z}\left(1 - \frac{a}{z}\right)^{-1} = \sum_{m=0}^{\infty} \frac{a^m}{z^{m+1}} = \sum_{n=-\infty}^{-1} a^{-n-1} z^n$$

where the binomial expansion is valid since $|\frac{a}{z}| < 1$.

5. *This doesn't seem to work for* $f(z) = z^{-1/2}$. *We cannot find a Laurent series about* $z_0 = 0$. *The reason is the required branch cut would pass through any annulus about 0, so cannot find such an annulus in which* f *is analytic. (Of course,* $z^{-1/2}$ *has a Taylor series about other points* $z_0 \neq 0$, *except those on the branch cut).*

Remarks 6.7 (Radius of convergence). *Suppose we have a Laurent series that we know to be valid in annulus* $R_1 < |z - z_0| < R_2$, *but there are no singularities on* $|z - z_0| = R_2$. *Then the outer radius of convergence can actually be pushed outwards until the circle touches a singularity, say* z_2.

Similarly, if no singularity in $|z - z_0| = R_1$, *the inner radius can be pulled in until we encounter a singularity, say* z_1. *Then our Laurent series in fact converges in the enlarged annulus*

$$\{z : \ |z_1| = R_1' < |z - z_0| < R_2' = |z_2|\}$$

In other words, the annulus of convergence of a Laurent series is always maximally large, with a singularity on each of the bounding circles, even if we originally derived it in a smaller region. By Remarks 3.1.1, the new Laurent series must be the same as the old one, noting that they have an overlap region where they are both valid.

Example 6.8. $\operatorname{cosec} z$ *has a Laurent series*

$$\left(z - \frac{z^3}{3!} + \dots\right)^{-1} = z^{-1}\left(1 - \frac{z^2}{6} + \dots\right)^{-1} = z^{-1} + \frac{z}{6} + \dots$$

for sufficiently small $z \neq 0$, *but it is not immediately obvious when the binomial expansion will be valid. The singularities of* $\operatorname{cosec} z$ *closest to 0 are at* $z = \pm\pi$, *so the radius of convergence is in fact* $0 < |z| < \pi$.

6.2.5 Classification of singularities

Definition 6.15 (Isolated singularity)**:** Suppose f has a singularity at $z = z_0$. If there is a neighbourhood of z_0 in which f is analytic (except at z_0) then f has an isolated singularity at z_0. Otherwise, we call it a non-isolated singularity.

Example 6.9.

1. $\operatorname{cosech} z$ *has an isolated singularity at* $z = n\pi i$, $n \in \mathbb{Z}$.

2. $\operatorname{cosech} \frac{1}{z}$ *has an isolated singularity at* $z = \frac{1}{n\pi i}$, $n \neq 0$ *and a non-isolated singularity at* $z = 0$ *(since there are other arbitrarily close singularities, i.e. n large).*

3. $\operatorname{cosech} z$ *has a non-isolated singularity at* ∞.

4. $z^{-1/2}$ *has a branch point singularity at* $z = 0$. *This is a type of non-isolated singularity since* $z^{-1/2}$ *is not analytic at any point on the branch cut. We usually treat this as a separate case.*

Definition 6.16 (Removable singularity)**:** If f has a removable singularity at z_0, then $\lim_{z \to z_0} f(z)$ is finite, i.e. $f(z) = a_0 + a_1(z - z_0) + \ldots$ for $0 < |z - z_0| < r_0$ s.t. $\lim_{z \to z_0} f(z) = a_0$. Then, we can 'remove the singularity' by redefining $f(z_0)$ as a_0. Then f will become analytic at z_0.

Definition 6.17 (Pole)**:** If f has a pole at z_0, then $\lim_{z \to z_0} |f(z)| = \infty$.

Definition 6.18 (Essential isolated singularity)**:** If f has an essential isolated singularity, f does not tend to any finite or infinite limit at this point.

Remarks 6.8. *If f has an isolated singularity at z_0, we can find an annulus $0 < |z - z_0| < r_0$ say within which f is analytic and therefore has a Laurent series. This gives a way to classify singularities.*

1. *Check for a branch point singularity.*

2. *Check for a non-isolated singularity.*

3. *Otherwise, consider the coefficients of Laurent series $\sum_{n=-\infty}^{\infty} a_n(z - z_0)^n$:*

   - *If $a_n = 0$ $\forall n < 0$, then f has a removable singularity at z_0.*
   - *If $\exists N > 0$ s.t. $a_n = 0$ $\forall n < -N$, but $a_{-N} \neq 0$, then f has a pole of order N at z_0 (simple pole at $N = 1$, double pole at $N = 2$, etc).*
   - *If there is no such N, then f has an essential isolated singularity,*

In fact, we can show that f takes all possible complex values (except at most one) in any neighbourhood of z_0, however small. For instance, $e^{1/z}$ takes all values except 0.

Proposition 6.8:

1. If $g(z)$ has a zero of order N at z_0, then $\frac{1}{g(z)}$ has a pole of order N there.

2. A rational function $f(z) = \frac{P(z)}{Q(z)}$ where P and Q are polynomials in z, has a singularity at any z_0, where Q has a zero. But if $P(z_0) = 0$ as well and assuming Q has a simple zero, then the singularity is removable by redefining $f(z_0) = \frac{P'(z_0)}{Q'(z_0)}$.

Example 6.10.

1. $\frac{\cos z}{z}$ *has Laurent series* $z^{-1} - \frac{1}{2}z + \frac{1}{24}z^3 + \dots$ *about* $z = 0$ *so has a simple pole at 0.*

2. $\frac{1}{z-i}$ *has a simple pole at* $z = i$, *since it is its own Laurent series.*

3. $\frac{z^2}{(z-1)^2(z-i)^3}$ *has a double pole at* $z = 1$ *and triple pole at* $z = i$. *To show formally that, for instance, there is a double pole at* $z = 1$, *note* $\frac{z^2}{(z-i)^3}$ *is analytic there, so has a Taylor series* $b_0 + b_1(z - 1) + b_2(z - 1)^2 + \dots$ *for some* b_n. *Hence,*

$$\frac{z^2}{(z - 1)^2(z - i)^3} = \frac{b_0}{(z - 1)^2} + \frac{b_1}{z - 1} + b_2 + \dots$$

4. $\cot z$ *has a simple pole at 0, since* $\tan z$ *has a simple zero there.*

5. z^2 *has a double pole at* ∞.

6. $e^{1/z}$ *has an essential isolated singularity at 0 because* $a_n \neq 0$ *for* $n \leq 0$.

7. $\sin\frac{1}{z}$ *also has an essential isolated singularity at* $z = 0$ *because using the standard Taylor series for* $\sin$, *there are non-zero* a_n *for infinitely many negative* n.

8. $f(z) = \frac{e^z - 1}{z}$ *has a removable singularity at 0 since* $f(z) = 1 + \frac{1}{2!}z + \frac{1}{3!}z^2 + \dots$. *By defining* $f(0) = 1$, *we would remove the singularity and obtain an entire function.*

9. $f(z) = \frac{\sin z}{z}$ *is not defined at 0, but has a removable singularity there. Remove it by defining* $f(0) = 1$.

7   Series solutions to ODEs

Homogeneous equations; solution by series (without full discussion of logarithmic singularities), exemplified by Legendre's equation. Classification of singular points. Indicial equation and local behaviour of solutions near singular points.   **[3]**

We will consider series solutions to the equations of the form

$$p(x)y'' + q(x)y' + r(x)y = 0 \tag{7.1}$$

Definition 7.1 (Ordinary and singular points)**:** The point $x = x_0$ is an ordinary point of the differential equation if $\frac{q}{p}$ and $\frac{r}{p}$ have Taylor series about x_0. Otherwise, x_0 is a singular point.

If x_0 is a singular point but the equation can be written as

$$P(x)(x - x_0)^2 y'' + Q(x)(x - x_0)y' + R(x)y = 0 \tag{7.2}$$

where $\frac{Q}{P}$ and $\frac{R}{P}$ have Taylor series about x_0, then x_0 is a regular singular point.

Example 7.1.

1. $(1 - x^2)y'' - 2cy' + 2y = 0$. $x = 0$ *is an ordinary point. However,* $x = \pm 1$ *are (regular) singular points since* $p(\pm 1) = 0$.

2. $\sin x y'' + \cos x y' + 2y = 0$. $x = n\pi$ *are regular singular points while all others are ordinary.*

3. $(1 + \sqrt{x})y'' - 2xy' + 2y = 0$. $x = 0$ *is an irregular singular point because* $\sqrt{x}$ *is not differentiable at* $x = 0$.

Proposition 7.1: If x_0 is

- an ordinary point, then the equation is guaranteed to have two linearly independent solutions of the form $y = \sum_{n=0}^{\infty} a_n(x - x_0)^n$, i.e. Taylor series about x_0. The solution must be convergent in some neighbourhood of x_0;

- a regular singular point, then there is at least one solution of the form $y = \sum_{n=0}^{\infty} a_n(x - x_0)^{n+\sigma}$, i.e. Frobenius series about x_0 where $a_0 \neq 0$ to ensure the index $\sigma \in \mathbb{C}$ is unique.

7.1   Ordinary points

Example 7.2 (Ordinary point)**.** *Consider* $(1 - x^2)y'' - 2xy' + 2y = 0$. *Find a series solution about* $x = 0$ *(which is an ordinary point).*
We try $y = \sum_{n=0}^{\infty} a_n x^n$ *and compare the coefficient of* x^n *and obtain the following general recurrence relation:*

$$0 = n(n-1)a_n + [-(n-2)^2 - (n-2) + 2]a_{n-2} = n(n-1)a_n - (n^2 - 3n)a_{n-2}$$

Here we do not divide by anything since they might be zero.

First consider the case $n = 0$. *The left hand side gives* $0 \cdot a_0 = 0$ *(the right hand side is* 0 *since* $a_{n-2} = 0$*). So any value of* a_0 *satisfies the recurrence relationship, and it can take any arbitrary value. This corresponds to a constant of integration. Similarly, by considering* $n = 1$, a_1 *is arbitrary. For* $n > 1$, n *and* $n - 1$ *are non-zero. So we have*

$$a_n = \frac{n-3}{n-1}a_{n-2}$$

In this case (but generally not), we can further simplify it to obtain:

$$a_{2k} = \frac{-1}{2k-1}a_0,$$

$$a_{2k+1} = 0.$$

So we obtain

$$y = a_0\left[1 - \frac{x^2}{1} - \frac{x^4}{3} - \frac{x^6}{5} - \cdots\right] + a_1 x$$

$$= a_0\left[1 - \frac{x}{2}\ln\left(\frac{1+x}{1-x}\right)\right] + a_1 x$$

Notice the logarithmic behaviour near $x = \pm 1$ *which are regular singular points.*

7.2   Regular singular points

Example 7.3 (Regular singular point)**.** *Consider* $4xy'' + 2(1 - x^2)y' - xy = 0$. *Note that* $x = 0$
is a singular point. But since $\frac{Q}{P} = \frac{1-x^2}{2}$ *and* $\frac{R}{P} = -\frac{x^2}{4}$ *both have Taylor series,* $x = 0$ *is a regular*
singular point. Try $y = \sum_{n=0}^{\infty} a_n x^{n+\sigma}$ *with* $a_0 \neq 0$ *and comparing the coefficient of* $x^{n+\sigma}$*, we*
obtain the general recurrence relation

$$[4(n+\sigma)(n+\sigma-1)+2(n+\sigma)]a_n - [2(n-2+\sigma)+1]a_{n-2} = 0 \implies 2(n+\sigma)(2n+2\sigma-1)a_n = (2n+2\sigma-3)a_{n-2}.$$

The $n = 0$ *case gives the indicial equation for the index* σ*:*

$$2\sigma(2\sigma - 1)a_0 = 0.$$

Since $a_0 \neq 0$*, we must have* $\sigma = 0$ *or* $\frac{1}{2}$*. The* $\sigma = 0$ *solution corresponds to an analytic ("Taylor*
series") solution, while $\sigma = \frac{1}{2}$ *corresponds to a non-analytic one.*

When $\sigma = 0$*, the recurrence relation becomes*

$$2n(2n - 1)a_n = (2n - 3)a_{n-2}.$$

When $n = 0$*, this gives* $0 \cdot a_0 = 0$*. So* a_0 *is arbitrary. For* $n > 0$*, we can divide and obtain*

$$a_n = \frac{2n - 3}{2n(2n - 1)}a_{n-2}.$$

We can see that $a_1 = 0$ *and so are subsequent odd terms. If* $n = 2k$*, i.e.* n *is even, then*

$$a_{2k} = \frac{4k - 3}{4k(4k - 1)}a_{2k-2}$$

$$y = a_0 \left[1 + \frac{1}{4 \cdot 3}x^2 + \frac{5}{8 \cdot 7 \cdot 4 \cdot 3}x^4 + \cdots\right]$$

Note that we have only found one solution in this case. Now when $\sigma = \frac{1}{2}$*, we obtain*

$$(2n + 1)(2n)a_n = (2n - 2)a_{n-2}$$

When $n = 0$*, we obtain* $0 \cdot a_0 = 0$*, so* a_0 *is arbitrary. To avoid confusion with the* a_0 *above, call it*
b_0 *instead. When* $n = 1$*, we obtain* $6a_1 = 0$ *and* $a_1 = 0$ *and so are subsequent odd terms. For even*
n*,*

$$a_n = \frac{n - 1}{n(2n + 1)}a_{n-2}$$

So

$$y = b_0 x^{1/2}\left[1 + \frac{1}{2 \cdot 5}x^2 + \frac{3}{2 \cdot 5 \cdot 4 \cdot 9}x^4 + \cdots\right]$$

7.3   Resonance of solutions

Note that the indicial equation has two roots σ_1, σ_2. Consider the two different cases:

1. If $\sigma_2 - \sigma_1 \notin \mathbb{Z}$, then there are two linearly independent Frobenius solutions

$$y = \left[(x - x_0)^{\sigma_1} \sum_{n=0}^{\infty} a_n(x - x_0)^n\right] + \left[(x - x_0)^{\sigma_2} \sum_{n=0}^{\infty} b_n(x - x_0)^n\right].$$

   As $x \to x_0$, $y \sim (x - x_0)^{\sigma_1}$, where $\Re(\sigma_1) \leq \Re(\sigma_2)$

2. If $\sigma_2 - \sigma_1 \in \mathbb{Z}$ (including when they are equal), there is one solution of the form

$$y_1 = (x - x_0)^{\sigma_2} \sum_{n=0}^{\infty} a_n(x - x_0)^n$$

with $\sigma_2 \geq \sigma_1$. In this case, $\sigma = \sigma_1$ will not give a valid solution, as we will later see. Instead, the other solution is (usually) in the form

$$y_2 = \ln(x - x_0)y_1 + \sum_{n=0}^{\infty} b_n (x - x_0)^{n+\sigma_1}.$$

This form arises from resonance between the two solutions. But if the resonance somehow avoids itself, we can possibly end up with two regular Frobenius series solutions.

We can substitute this form of solution into the differential equation to determine b_n.

Example 7.4 (Regular singular point with logarithmic solution). *Consider $x^2 y'' - xy = 0$. $x = 0$ is a regular singular point. It is already in equidimensional form $(x^2 y'') - x(y) = 0$. Try*

$$y = \sum_{n=0}^{\infty} a_n x^{n+\sigma}$$

with $a_0 \neq 0$. We obtain

$$\sum a_n x^{n+\sigma}[(n+\sigma)(n+\sigma-1) - x] = 0.$$

The general recurrence relation is

$$(n+\sigma)(n+\sigma-1)a_n = a_{n-1}.$$

$n = 0$ gives the indicial equation

$$\sigma(\sigma - 1) = 0.$$

Then $\sigma = 0, 1$. We are guaranteed to have a solution in the form $\sigma = 1$. When $\sigma = 1$, the recurrence relation becomes

$$(n+1)na_n = a_{n-1}.$$

When $n = 0$, $0 \cdot a_0 = 0$ so a_0 is arbitrary. When $n > 0$, we obtain

$$a_n = \frac{1}{n(n+1)}a_{n-1} = \frac{1}{(n+1)(n!)^2}a_0.$$

So

$$y_1 = a_0 x \left(1 + \frac{x}{2} + \frac{x^2}{12} + \frac{x^3}{144} + \cdots\right).$$

When $\sigma = 0$, we obtain

$$n(n-1)a_n = a_{n-1}.$$

When $n = 0$, $0 \cdot a_0 = 0$ and a_0 is arbitrary. When $n = 1$, $0 \cdot a_1 = a_0$. However, $a_0 \neq 0$ by our initial constraint. Contradiction. So there is no solution in this form (If we ignore the constraint that $a_0 \neq 0$, we know that a_0 is arbitrary. But this gives exactly the same solution we found previously with $\sigma = 1$). The other solution is thus in the form

$$y_2 = y_1 \ln x + \sum_{n=0}^{\infty} b_n x^n.$$

8   Sturm-Liouville theory

Self-adjoint operators, eigenfunctions and eigenvalues, reality of eigenvalues and orthogonality of eigenfunctions. Eigenfunction expansions and determinantion of coefficients. Legendre polynomials; orthogonality

[3]

8.1   Review of 2nd order linear ODEs

We wish to solve the general inhomogeneous ODE

$$\mathcal{L}y = \alpha(x)y'' + \beta(x)y' + \gamma(x)y = f(x) \tag{8.1}$$

The homogeneous equation

$$\mathcal{L}y = 0 \tag{8.2}$$

has two independent solutions $y_1(x)$, $y_2(x)$ (besides the trivial solution $y = 0$), the complementary function $y_c(x)$ is a general solution of (1.28):

$$y_c(x) = Ay_1(x) + By_2(x) \tag{8.3}$$

where A and B are arbitrary constants. The inhomogeneous equation

$$\mathcal{L}y = f(x) \tag{8.4}$$

is solved by the particular integral $y_p(x)$. So the general solution of (8.4) is then

$$y(x) = y_p(x) + Ay_1(x) + By_2(x) \tag{8.5}$$

The boundary or initial data are required to determine A, B in (8.5).

- b.c.s: solve (8.4) on $a < x < b$ given

  - Dirichlet: y at $x = a, b$
  - Neumann: y' at $x = a, b$
  - Cauchy (mixed): $y + ky' = 0$

  Homogeneous b.c.s are often assumed, $y(a) = y(b) = 0$. This can always be achieved by adding complementary functions (8.3), i.e. $\tilde{y}(x) = y(x) + Ay_1(x) + By_2(x)$, s.t. $\tilde{y}(a) = \tilde{y}(b) = 0$.

- i.c.s: solve (8.4) for $x \geq a$, given y, y' at $x = a$.

To solve (8.1) using eigenfunction expansions, we must first solve the related eigenvalue problem

$$\alpha(x)y'' + \beta y'(x) + \gamma(x)y = -\lambda\rho(x)y \tag{8.6}$$

with specified b.c.s. This form occurs after invoking separation of variables in several dimensions.

8.2   Self-adjoint operators

Definition 8.1 (Weighted inner product)**:** Define inner product between two continuous complex-valued functions f, g w.r.t the weight function $w(x)$ on $a \leq x \leq b$ to be

$$\langle f, g \rangle_w = \int_a^b w(x)f^*(x)g(x)dx = \langle wf, g \rangle = \langle f, wg \rangle \tag{8.7}$$

The weight function $w(x)$ is a real-valued, non-negative function that has at most finitely many zeroes on the domain $[a, b]$.

Definition 8.2 (Norm)**:** The norm of a complex-valued function w.r.t some weight function $w(x)$ is

$$||f||_w := \sqrt{\langle f, f \rangle_w}$$

Since f is continuous and w has finitely many zeroes on $[a, b]$, $\langle f, f \rangle_w = 0$ iff $f(x) = 0$ identically on $[a, b]$.

Definition 8.3 (Self-adjoint): $\mathcal{L}$ is said to be self-adjoint on $a \leq x \leq b \; \forall$ functions y_1, y_2 satisfying appropriate b.c.s, if

$$\langle y_1, \mathcal{L}y_2 \rangle = \langle \mathcal{L}y_1, y_2 \rangle, \text{ i.e. } \int_a^b y_1^*(x)\mathcal{L}y_2(x)dx = \int_a^b (\mathcal{L}y_1(x))^* y_2(x)dx \tag{8.8}$$

Definition 8.4 (Sturm-Liouville equation): When $\mathcal{L}$ is self-adjoint, then the eigenvalue problem (8.6) can be expressed in Sturm-Liouville (SL) form

$$\mathcal{L}y := -(py')' + qy = \lambda wy \tag{8.9}$$

where the weight function $w(x) \geq 0 \; \forall x$.

Proposition 8.1: The eigenvalue problem (8.6) can be recasted to SL form as long as suitable b.c.s are satisfied.

Proof. Multiply (8.6) by an integration factor $\mu(x)$ to find

$$-\lambda\mu\rho y = \mu\alpha y'' + \mu\beta y' + \mu\gamma y = (\mu\alpha y')' + y'(-\mu'\alpha - \mu\alpha' + \mu\beta)y' + \mu\gamma y$$

Eliminate the y' term in order to cast it to SL form,

$$\frac{1}{\mu(x)}\frac{d\mu}{dx} = \frac{\beta - \alpha'}{\alpha} \implies \mu(x) = \exp\left(\int^x \frac{\beta - \alpha'}{\alpha}dx\right) \tag{8.10}$$

In that case, $p(x) := \mu(x)\alpha(x)$, $q(x) := -\mu(x)\gamma(x)$, $w(x) = \mu(x)\rho(x)$. Note, $\mu(x) \geq 0$. In order to ensure $\mathcal{L}$ is self-adjoint on $a \leq x \leq b \; \forall$ eigenfunctions y_1, y_2, we require it to satisfy (8.8), so

$$\langle y_1, \mathcal{L}y_2 \rangle - \langle \mathcal{L}y_1, y_2 \rangle = \int_a^b (-y_1(py_2')' + y_1qy_2 + y_2(py_1')' - y_2qy_1)dx$$

$$= \int_a^b (-(py_1y_2')' + (py_1'y_2)')dx$$

$$= [-py_1y_2' + py_1'y_2]_a^b = 0 \tag{8.11}$$

The self-adjoint compatible b.c.s include

- homogneneous b.c.s $y(a) = y(b) = 0$ or $y'(a) = y'(b) = 0$, or mixed $y + ky' = 0$, etc;

- periodic $y(a) = y(b)$;

- singular points of ODE $p(a) = p(b) = 0$;

- combinations of the above.

$\square$

Theorem 8.1: For self-adjoint operators $\mathcal{L}$,

1. the eigenvalues λ_n are real,

2. the eigenfunctions y_n are orthogonal w.r.t the weight function $w(x)$;

3. the eigenfunctions y_n form a complete set.

Proof. We shall not prove the third part. For the first two, consider two eigenfunctions y_n, y_m with eigenvalues λ_n, λ_m respectively, i.e.

$$\mathcal{L}y_{n,m} = \lambda_{n,m}wy_{n,m} \tag{8.12}$$

Consider

$$\int_a^b (y_m^*\mathcal{L}y_n - y_n\mathcal{L}y_m)dx = (\lambda_n - \lambda_m^*)\int_a^b wy_ny_mdx$$

LHS is zero by (8.12). When $n = m$, RHS is $(\lambda_n - \lambda_n^*)\int_a^b w|y_n|^2dx$, but $\int w|y_n|^2dx \geq 0$ for $y_n \neq 0$, so $\lambda_n = \lambda_n^* \in \mathbb{R}$, i.e. real eigenvalues. In particular, if λ_n are non-degenerate, then $y_n^* = y_n$ are

real eigenfunctions.

If $m \neq n$ s.t. the eigenvalues are distinct, i.e. $\lambda_m \neq \lambda_n$, then

$$\int_a^b wy_ny_m dx = 0 \quad \forall n \neq m \tag{8.13}$$

i.e. y_m and y_n are orthogonal w.r.t the weight function $w(x)$ on $a \leq x \leq b$. $\qquad\square$

Remarks 8.1.

- *One can show that the regular SL problem (homogeneous b.c.s) has simple eigenvalues λ_n, i.e. unique eigenfunctions. This is done by considering two eigenfunctions u, v with the same eigenvalue λ and use*

$$u\mathcal{L}v - (\mathcal{L}u)v = [-p(uv' - u'v)]'$$

- *We can eliminate the weight function $w(x)$ by redefining $\tilde{y} = \sqrt{w}y$ and replacing $\mathcal{L}y$ by $\frac{1}{\sqrt{w}}\mathcal{L}\frac{\tilde{y}}{\sqrt{w}}$ but it is generally simpler to keep it.*

Example 8.1. *Consider the Hermite equation for simple harmonic oscillator:*

$$y'' - 2xy' + 2ny = 0$$

We wish to cast it to SL form (8.9). Compare with (8.6), $\alpha = 1$, $\beta = -2x$, $\gamma = 0$, $\lambda\rho = 2n$. By (8.10), the integration factor is $\exp(\int^x -2xdx) = e^{-x^2}$, hence

$$\mathcal{L}y = -(e^{-x^2}y')' = 2ne^{-x^2}y \tag{8.13}$$

We can eliminate the weight function $w(x)$ with $\tilde{y} = e^{-x^2/2}y$ to find

$$\tilde{\mathcal{L}}\tilde{y} = -\tilde{y}'' + (x^2 - 1)\tilde{y} = 2n\tilde{y}$$

8.3 Completeness and Parseval's identity

One can show that the eigenvalues λ_n form a countably infinite sequence $\lambda_1, \lambda_2, \ldots$ with $|\lambda_n| \to \infty$ as $n \to \infty$, and that the corresponding set of orthogonal eigenfunctions $y_1(x), y_2(x), \ldots$ form a complete basis for functions on $[a, b]$ satisfying these b.c.s. Completeness implies we can expand any well-behaved function $f(x)$ on $a \leq x \leq b$ as a discrete sum. This discreteness arises because the domain $[a, b]$ is compact, and because of our b.c.s. Write

$$f(x) = \sum_{n=1}^{\infty} a_ny_n(x) \tag{8.14}$$

To find the expansion coefficients, take

$$\int_a^b w(x)y_m(x)f(x)dx = \sum_{n=1}^{\infty} a_n \int_a^b wy_ny_m dx = a_n \int_a^b wy_m^2 dx \implies a_n = \frac{\int_a^b w(x)y_n(x)f(x)dx}{\int_a^b w(x)y_n^2(x)dx} \tag{8.15}$$

where we exploited orthogonality (8.13). The eigenfunctions are usually normalized for convenience. Eigenfunctions with unit norm take the form

$$Y_n(x) = \frac{y_n(x)}{\sqrt{\int_a^b wy_n^2 dx}}, \quad \langle Y_n, Y_m \rangle = \delta_{mn} \tag{8.16}$$

with $f(x)$ now written as $f(x) = \sum_{n=1}^{\infty} A_nY_n(x)$.

Proposition 8.2 (Bessel's inequality)**:** For an arbitrary expansion in terms of the set of normalized eigenfunctions, $f(x) = \sum_{n=1}^{\infty} A_nY_n(x)$ satisfy the inequality

$$\int_a^b wf^2 dx \geq \sum_{n=1}^{\infty} A_n^2$$

Again, it is an inequality if the eigenbasis is incomplete.

Proof. Consider

$$\int_a^b (f(x) - \sum_{n=1}^\infty a_n y_n)^2 w dx = \int_a^b (f^2 - 2f \sum_n a_n y_n + \sum_n a_n^2 y_n^2) w dx = \int_a^b w f^2 dx - \sum_{n=1}^\infty a_n^2 \int_a^b w y_n^2 dx$$

where we used orthogonality (8.13) and the relation (8.16). If the eigenfunctions are complete, then the series converges (inequality becomes an equality).

$$\int_a^b w f^2 dx = \sum_{n=1}^\infty a_n^2 \int_a^b w y_n^2 dx = \sum_{n=1}^\infty A_n^2 \tag{8.17}$$

$\square$

Definition 8.5 (Partial sum): We define the partial sum of the normalized eigenfunctions to be $S_N(x) = \sum_{n=1}^N a_n Y_n$, s.t.

$$f(x) = \lim_{N \to \infty} S_N(x) \tag{8.18}$$

Definition 8.6 (Mean square error): The mean square error $\epsilon_N = \int_a^b w (f(x) - S_N(x))^2 dx$ is s.t. $\lim_{N \to \infty} \epsilon_N = 0$.

This is a global definition of convergence in the mean (unlike the pointwise convergence of Fourier series).

Proposition 8.3 (Least squares approximation): We want to approximate an arbitrary function $f(x)$ as a finite sum of the eigenfunctions. The error in the partial sum (8.18) is minimized by a_n (8.15) for the $N = \infty$ expansion.

Proof. Extremize ϵ_N in terms of a_n:

$$\frac{\partial \epsilon_N}{\partial a_n} = -2 \int_a^b y_n w \left(f - \sum_{m=1}^N a_m y_m \right) dx = -2 \int_a^b (w f y_n - a_n y_n^2) dx = 0$$

where we used (8.15). This is a minimum since $\frac{\partial^2 \epsilon_N}{\partial a_n^2} = \int_a^b w y_n^2 dx \geq 0$. $\square$

Example 8.2 (Legendre polynomials). *Consider the Legendre's equation*

$$(1 - x^2) y'' - 2x y' + \lambda y = 0 \tag{8.19}$$

on the interval $-1 \leq x \leq 1$ with y finite at $x = \pm 1$ (regular singular point of ODE, hence give rise to self-adjoint compatible b.c.s). (8.19) casted in SL form (1.35) will have $p = 1 - x^2$, $q = 0$, $w = 1$. How do we solve (8.9)? Seek a power series about $x = 0$ (ordinary point): $y = \sum_n c_n x^n$. Substitute and we obtain

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

Comparing coefficients of x^n to obtain the recurrence relation:

$$(n+2)(n+1) c_{n+2} - n(n-1) c_n - 2n c_n + \lambda_n c_n = 0 \implies c_{n+2} = \frac{n(n+1) - \lambda}{(n+1)(n+2)} c_n \tag{8.20}$$

So specifying c_0 and c_1 will give two linearly independent solution (near $x = 0$):

$$y_{even} = c_0 \left(1 + \frac{\lambda}{2!} x^2 + \frac{(6-\lambda)(-\lambda)}{4!} x^4 + \dots \right), \quad y_{odd} = c_1 \left(x + \frac{2-\lambda}{3!} x^3 + \dots \right)$$

*But as $n \to \infty$, $\frac{c_{n+2}}{c_n} \to 1$ so these are geometric series with radius of convergence $|x| < 1$, i.e. divergent at $x = \pm 1$. What can we do to ensure finiteness? Take $\lambda = \ell(\ell+1)$ with $\ell \in \mathbb{Z}$, then one of the series terminates, i.e. $c_n = 0 \; \forall n \geq \ell + 2$. These are the **Legendre polynomials**, eigenfunctions of (8.19) on $-1 \leq x \leq 1$ with normalization convention $P_\ell(1) = 1 \; \forall \ell$.*

$$P_0(x) = 1, \quad P_1(x) = x, \quad P_2(x) = \frac{1}{2}(3x^2 - 1), \quad P_3(x) = \frac{1}{2}(5x^3 - 3x) \tag{8.21a}$$

with eigenvalues $0, 2, 6, 12$ respectively. $P_\ell(x)$ has ℓ number of zeros, with $P_\ell(x)$ being even (odd) if ℓ is even (odd). There is an alternative definition using the generating function:

$$\frac{1}{\sqrt{1 - 2xt + t^2}} \approx 1 + \frac{1}{2}(2xt - t^2) + \frac{3}{8}(2xt - t^2)^2 + \cdots = 1 + xt + \frac{1}{2}(3x^2 - 1)t^2 + \cdots = \sum_{\ell=0}^{\infty} P_\ell(x)t^\ell$$
(8.21b)

In fact, the Legendre polynomials have a generic form described by the Rodrigue's formula

$$P_n(x) = \frac{1}{2^n n!}\frac{d^n}{dx^n}(x^2 - 1)^n$$

Using this formula, one can show that the Legendre polynomials are orthonormal.

$$\int_{-1}^{1} P_n P_m dx = \frac{2}{2n + 1}\delta_{nm}$$
(8.22)

These nice properties allow us to expand any $f(x)$ on $-1 \leq x \leq 1$ as

$$f(x) = \sum_{\ell=0}^{\infty} a_\ell P_\ell(x)$$
(8.23)

where

$$a_\ell = \frac{2\ell + 1}{2}\int_{-1}^{1} f(x)P_\ell(x)dx$$
(8.24)

8.4 Sturm-Liouville theory and inhomogeneous ODE

Consider the inhomogeneous problem with homogeneous b.c.s on $a \leq x \leq b$

$$\mathcal{L}y = f(x) = wF(x)$$
(8.25)

Given the eigenfunctions $y_n(x)$ satisfying $\mathcal{L}y_n = \lambda_n w y_n$. Recall Section 1.1.6 where we expand $y(x)$ and $F(x)$ in a complete set of eigenfunctions of $\mathcal{L}$, i.e. $F(x) = \sum_n a_n y_n(x)$, $y(x) = \sum_n c_n y_n(x)$, with a_n's known (8.15) and c_n's unknown. Substitute into (8.25) and by orthogonality (8.31), the particular solution is

$$\mathcal{L}y = \mathcal{L}\sum_n c_n y_n = \sum_n c_n \lambda_n w y_n = w\sum_n a_n y_n \implies y(x) = \sum_{n=1}^{\infty} \frac{a_n}{\lambda_n}y(x)$$
(8.26)

assuming $\lambda_n \neq 0 \ \forall n$. We can generalize this method. For instance, driving forces often induce a linear response term $\tilde{\lambda}wy$, where $\tilde{\lambda}$ is fixed. The solution (8.26) becomes

$$\mathcal{L}y - \tilde{\lambda}wy = f(x) \implies y(x) = \sum_{n=1}^{\infty} \frac{a_n}{\lambda_n - \tilde{\lambda}}y_n(x)$$
(8.27)

again $\tilde{\lambda}_n \neq \lambda_n \ \forall n$. In fact, the solution (8.26) can be represented as an integral using Green's functions (more later, for now treat it as a formal inverse to the differential operator $\mathcal{L}$):

$$y(x) = \sum_{n=1}^{\infty} \frac{a_n}{\lambda_n}y_n(x) = \sum_n \frac{y_n(x)}{\lambda_n||y_n||_w^2}\int_a^b w(\xi)F(\xi)y_n(\xi)d\xi$$

$$= \int_a^b \sum_n \frac{y_n(x)y_n(\xi)}{\lambda_n||y_n||_w^2}w(\xi)F(\xi)d\xi$$

$$:= \int_a^b G(x,\xi)f(\xi)d\xi$$
(8.28)

where the Green's function is $G(x,\xi) = \sum_{n=1}^{\infty} \frac{y_n(x)y_n(\xi)}{\lambda_n||y_n||_w^2}$, where $(x,\xi) \in [a,b] \times [a,b]$. G depends on $\mathcal{L}$ both through its eigenfunctions and the self-adjoint compatible b.c.s.

9   Variational principles

Lagrange multipliers, examples with two or three variables. Euler-Lagrange equations and examples.
Variational principles; Fermat's principle; Hamilton's principle and deduction of Lagrange's equation.
Variational principle for the lowest eigenvalue and for higher eigenvalues (Rayliegh-Ritz)   **[6]**

9.1   Extremizing functionals

For this topic, we will focus on finding stationary points and not on their classification (maxima, minima, etc). Also, we will restrict ourselves to first-order variation.

Definition 9.1 (Functional)**:** A functional is a function that takes in another real-valued function as an argument. We usually write them as $F[x]$ (square brackets), where $x = x(t) : \mathbb{R} \to \mathbb{R}$ is a real function. We say that $F[x]$ is a functional of the function $x(t)$.

We wish to extremize the functional of the form

$$F[y] = \int_\alpha^\beta f(y, y', x) dx \tag{9.1}$$

with a given smooth function $f(y, y', x)$. F depends on y but not x. The boundary conditions (b.c.s) is s.t. $y(\alpha)$ and $y(\beta)$ are fixed. Consider small perturbations $y \to y + \varepsilon\eta(x)$ in (9.1). Compute $F[y + \varepsilon\eta]$ where $\eta(\alpha) = \eta(\beta) = 0$.

Lemma 9.1: If $g : [\alpha, \beta] \to \mathbb{R}$ is continuous on $[\alpha, \beta]$, and

$$\int_\alpha^\beta g(x)\eta(x) dx = 0$$

$\forall \eta(x)$ continuous on $[\alpha, \beta]$ and s.t. $\eta(\alpha) = \eta(\beta) = 0$, then $g(x) = 0 \ \forall x \in [\alpha, \beta]$.

Back to (9.1), we look at its first order variation

$$F[y + \varepsilon\eta] = \int_\alpha^\beta f(x, y + \varepsilon\eta, y' + \varepsilon\eta') dx = F[y] + \int_\alpha^\beta \left(\frac{\partial f}{\partial y}\eta + \frac{\partial f}{\partial y'}\eta' \right) dx + O(\varepsilon^2)$$

Proposition 9.1 (Euler-Lagrange Equations)**:** With first order variation, $\mathcal{F}$ will satisfy the Euler-Lagrange (EL) equations

$$\frac{d}{dx}\frac{\partial f}{\partial y'} = \frac{\partial f}{\partial y} \tag{9.2}$$

when F is extremized/stationary.

Proof. At the extrema, we have $F[y + \varepsilon\eta] = F[y] + O(\varepsilon^2)$, i.e. $\frac{\partial F}{\partial \varepsilon}|_{\varepsilon=0} = 0$. Integrate the ε term by parts:

$$0 = \int_\alpha^\beta \left\{ \frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'} \right) \right\}\eta(x) dx + \left[\frac{\partial f}{\partial y'}\eta \right]_\alpha^\beta$$

The second term has to be zero since $\eta(\alpha) = 0 = \eta(\beta) \ \forall \eta$. Apply Lemma 9.1 where $g = \frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'} = 0 \ \forall x \in [\alpha, \beta]$. This gives us the Euler-Lagrange equations as required. $\square$

Remarks 9.1.

- *The Euler-Lagrange equations are necessary conditions for an extremum.*

- *Equation (9.2) is a second order ODE for $y(x)$ with $y(\alpha) = y_1$, $y(\beta) = y_2$.*

- *LHS of (9.2) is denoted as $\frac{\delta F[y]}{\delta y(x)}$, and is called the **functional derivative**.*

- *Other b.c.s are possible. For instance, $\frac{\partial f}{\partial y'}|_{\alpha,\beta} = 0$, $y'(\alpha) = y'(\beta) = 0$ and mixed b.c.s (one end fixed, the other end free).*

In some cases, (9.2) can be integrated once to obtain a first order ODE, called the first integral.

Corollary 9.1: If f does not explicitly depend on y, then

$$\frac{\partial f}{\partial y'} = K \tag{9.3}$$

for some constant K.

Proof. Follows from Equation (9.2) after imposing $\frac{\partial f}{\partial y} = 0$. $\qquad\square$

Example 9.1 (Geodesics on the Euclidean plane)**.**

$$F[y] = \int_\alpha^\beta \sqrt{dx^2 + dy^2} = \int_\alpha^\beta \sqrt{1 + y'^2}dx$$

Denote $f(y') := \sqrt{1 + y'^2}$. *We observe that* $\frac{\partial f}{\partial y} = 0 \implies \frac{y'}{\sqrt{1+y'^2}} = $ *constant, which means* $y(x) = mx + c$. *The shortest path is a straight line, as expected. The constants m and c are obtained from the fixed values of $y(\alpha)$ and $y(\beta)$.*

Example 9.2 (Geodesics on a sphere)**.** *Consider a sphere embedded in $\mathbb{R}^3$, i.e. $\mathcal{S}^2 \subset \mathbb{R}^3$, where*

$$\mathcal{S}^2 = \{x = \sin\theta\sin\phi, \ y = \sin\theta\cos\phi, \ z = \cos\theta, \quad 0 \le \phi < 2\pi, \ 0 \le \theta < \pi\}$$

Parametrize curves on $\mathcal{S}^2$ by θ, i.e. $\phi(\theta)$. Change $x \to \theta$, $y \to \phi$ in the Euler-Lagrange equations. The functional is

$$F[\phi] = \int_{\theta_1}^{\theta_2} \sqrt{1 + \sin^2\theta(\phi')^2}d\theta$$

where $\phi' = \frac{d\phi}{d\theta}$. Let $f(\theta, \phi') := \sqrt{1 + \sin^2\theta(\phi')^2}$, then we see that $\frac{\partial f}{\partial \phi} = 0$. Then by Corollary 9.1,

$$K = \frac{\partial f}{\partial \phi'} = \frac{\sin^2\theta\phi'}{\sqrt{1 + \phi'^2\sin^2\theta}} \implies \phi - \phi_0 = \pm K \int \frac{d\theta}{\sin\theta\sqrt{\sin^2\theta - K^2}}$$

where K is some constant. We solve this integral by using $\cot\theta = u$, then we get

$$\phi - \phi_0 = \pm K \int \frac{-du}{\sqrt{1 - K^2(1 + u^2)}} = \pm K \int \frac{K}{\sqrt{1 - K^2}} \frac{-du}{\sqrt{1 - \frac{K^2}{1-K^2}u}} = \pm \cos^{-1}\left(\frac{K}{\sqrt{1 - K^2}}\cot\theta\right)$$

which gives $\cot\theta = \pm\frac{\sqrt{1-K^2}}{K}\cos(\phi - \phi_0)$. The constant ϕ_0 surface is a great circle.

Corollary 9.2: If f does not explicitly depend on x,

$$f - y'\frac{\partial f}{\partial y'} = K \tag{9.4}$$

for some constant K.

Proof. For a general function $f = f(x, y, y')$,

$$
\begin{aligned}
\frac{d}{dx}\left(f - y'\frac{\partial f}{\partial y'}\right) &= \frac{\partial f}{\partial x} + y'\frac{\partial f}{\partial y} + y''\frac{\partial f}{\partial y'} - \frac{d}{dx}\left(y'\frac{\partial f}{\partial y'}\right) \\
&= \frac{\partial f}{\partial x} + y'\frac{\partial f}{\partial y} + y''\frac{\partial f}{\partial y'} - y'\frac{d}{dx}\frac{\partial f}{\partial y'} - y''\frac{\partial f}{\partial y'} \\
&= \frac{\partial f}{\partial x} + y'\left(\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right)\right) \\
&= \frac{\partial f}{\partial x} + y'0
\end{aligned}
$$

where we used the chain rule and (9.2). Result follow by setting $\frac{\partial f}{\partial x} = 0$. $\qquad\square$

Proposition 9.2 (Fermat's principle)**:** Consider light traveling between two fixed points in an arbitrary medium of refractive index $\mu(\mathbf{x})$, then it travels along paths that require the least time.

Proposition 9.3 (Snell's law): For a ray with velocity $c = c(x)$, then by Fermat's Principle, the quantity $\frac{\sin \theta}{c(x)}$ is a constant.

Proof. The ray is described by a path $y = y(x)$ and the speed of light $c(x, y)$ depends on the position. The time taken is a functional

$$T[y] = \int \frac{dl}{c} = \int_\alpha^\beta \frac{\sqrt{1 + y'^2}}{c(x, y)} dx$$

where the integrand is $f(x, y, y')$. Assume $c = c(x)$, then $\frac{\partial f}{\partial y} = 0$, by (9.4),

$$K = \frac{\partial f}{\partial y'} = \frac{y'}{\sqrt{1 + y'^2}} \frac{1}{c(x)}$$

Since the ray is launched at θ_1, where $y' = \tan \theta$, then we obtain:

$$\frac{\sin \theta_1}{c(x_1)} = \frac{\sin \theta}{c(x)}, \ \forall x(t > 0) \tag{9.5}$$

$\square$

Definition 9.2 (Lagrangan and action): For a particle in $\mathbb{R}^3$, its kinetic energy and potential energy are respectively T and V. The Lagrangian is defined to be the result after subtracting potential energy from the kinetic energy (We will see why we do this later).

$$\mathcal{L}(\mathbf{x}, \dot{\mathbf{x}}, t) = T - V \tag{9.6}$$

The action for a path in configuration space between two points at times t_1 and t_2 respectively, is

$$S[\mathbf{x}] = \int_{t_1}^{t_2} \mathcal{L}(\mathbf{x}, \dot{\mathbf{x}}, t) dt \tag{9.7}$$

9.2   Constrained variation

Proposition 9.4: The corresponding EL equations with constraints have the form

$$\frac{d}{dx} \left(\frac{\partial}{\partial y'} (f - \lambda g) \right) = \frac{\partial}{\partial y} (f - \lambda g) \tag{9.8}$$

Proof. Extremize $F[y] = \int_\alpha^\beta f(x, y, y') dx$, subject to the constraint $G[y] = \int_\alpha^\beta g(x, y, y') dx = k$ for some constant k. Use Lagrange multiplier method, i.e. extremize $\Phi[y; \lambda] = F[y] - \lambda G[y]$. We replace f in (9.2) by $f(x, y, y') - \lambda g(x, y, y')$ to obtain (9.8). $\square$

Example 9.3 (Dido's isoperimetric problem). *What simple closed plane curve of fixed length L maximizes the enclosed area? This curve is convex - any two points inside these regions can be connected by a straight line that is contained in the region. Let x along the curve monotonically increase from α to β, and decreases from β to α.*
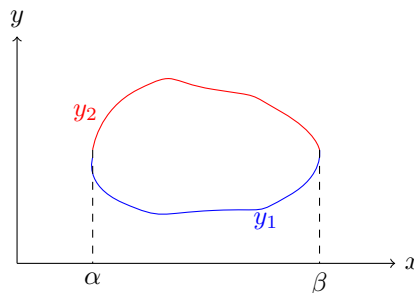
The area functional is

$$A[y] = \int_\alpha^\beta y_2(x) - y_1(x)dx = \oint_C y(x)dx$$

subjected to the constraint

$$L[y] = \oint_C dl = \oint_C \sqrt{1 + y'^2}dx$$

Set $h(y, y') = y - \lambda\sqrt{1 + y'^2}$. So $\frac{\partial h}{\partial x} = 0$ then by (2.5),

$$K = h - y'\frac{\partial h}{\partial y'} = y - \lambda\sqrt{1 + y'^2} + y'\lambda\frac{y'}{\sqrt{1 + y'^2}} = y - \frac{\lambda}{\sqrt{1 + y'^2}} \implies y'^2 = \frac{\lambda^2}{(y - K)^2} - 1$$

which upon solving gives $\int_{y_0}^y \frac{y - k}{\sqrt{\lambda^2 - (y - k)^2}}dy = x - x_0 \implies (x - x_0)^2 + (y - y_0)^2 = \lambda^2$, a circle of radius λ. But given our constraint, we have $2\pi\lambda = L$.

Example 9.4 (Sturm-Liouville problem). *Consider $\rho(x) > 0$ for $x \in [\alpha, \beta]$ and*

$$F[y] = \int_\alpha^\beta \rho(x)(y')^2 + \sigma(x)y^2 dx \tag{9.9a}$$

subjected to the constraint

$$G[y] = \int_\alpha^\beta w(x)y^2 dx \tag{9.9b}$$

where $\rho(x)$, $\sigma(x)$ and $w(x)$ are real functions of x, defined for $x \in [\alpha, \beta]$. Minimize F subject to $G = 1$ (fixed ends), the functional will be

$$\Phi[y; \lambda] = F[y] - \lambda(G[y] - 1)$$

By (9.8), the integrand would be $h = \rho(y')^2 + \sigma y^2 - \lambda(wy^2 - \frac{1}{\beta - \alpha})$ and so

$$0 = \frac{d}{dx}(2\rho y' - 2\sigma y + 2\lambda wy) = 2((\rho y)' - (\sigma - \lambda w)y)$$

which gives

$$-\frac{d}{dx}(\rho y') + \sigma y = \lambda wy \implies \mathcal{L}y := \lambda wy \tag{9.9c}$$

where $\mathcal{L}$ is the Sturm-Liouville operator. (9.9c) is also known as the eigenvalue-eigenvector problem. As long as we can cast any differential equation into (9.9), then (9.9) is known as the Sturm-Liouville (SL) form, as we seen earlier in Chapter 8.

Remarks 9.2.

1. *The time-independent Schrödinger's equation is of SL form $\rho = w = 1$, $\sigma(x) = V(x)$.*

2. *If $\sigma(x) > 0$, then $F[y] > 0$.*

Proposition 9.5: As long as $F[y] > 0$, then there exists a lowest positive eigenvalue.

Proof. Multiply (9.9) with y and integrate from α to β,

$$[-yy'\rho]_\beta^\alpha + F[y] = \lambda G[y]$$

First term is 0 due to b.c.s, so the lowest eigenvalue equals the minimum of $\frac{F[y]}{G[y]}$.   □

Remarks 9.3. *The original problem is equivalent to a problem of minimizing F/G because the scale is fixed by the normalization constant drops out of this ratio.*

Proposition 9.6: For a problem with several dependent variables, we will obtain n Euler-Lagrange equations, each of the form

$$\frac{d}{dx}\frac{\partial f}{\partial y_i'} = \frac{\partial f}{\partial y}, \ \forall i = 1, \ldots, n \tag{9.10}$$

This is a system of n second order ODEs.

Proof. The functional is now $F[\mathbf{y}] = \int_\alpha^\beta f(x, \mathbf{y}, \mathbf{y}')dx$. Consider small perturbations $y_i \to y_i + \varepsilon\eta_i(x)$ where $i = 1, \ldots, n$, ε small, and $\eta_i(\alpha) = \eta_i(\beta) = 0$. Following the derivation of the Euler-Lagrange equations,

$$F[\mathbf{y} + \varepsilon\boldsymbol{\eta}] = \int_\alpha^\beta \sum_{i=1}^n \eta_i \left\{ \frac{d}{dx}\left(\frac{\partial f}{\partial y_i'}\right) - \frac{\partial f}{\partial y_i} \right\} dx$$

where the boundary terms vanish. (9.10) follows from Lemma 9.1. $\qquad\square$

Corollary 9.3: The corresponding first integrals are

- If $\frac{\partial f}{\partial y_j} = 0$ for some j, then $\frac{\partial f}{\partial y_j'}$ is a constant.

- If $\frac{\partial f}{\partial x} = 0$, then $f - \sum_{i=1}^n y_i' \frac{\partial f}{\partial y_i'}$ is a constant.

Example 9.5 (Geodesics on surfaces)**.** *We want to find the shortest path on the surface Σ embedded in $\mathbb{R}^3$, given by $g(x, y, z) = 0$, between two points $A, B \in \Sigma$ (assuming one exists). Parameterize the curve with t s.t. $\mathbf{x}(0)$ is at A while $\mathbf{x}(1)$ is at B. The functional is*

$$\Phi[\mathbf{x}, \lambda] = \int_0^1 \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2} - \lambda(t)g(x, y, z)dt$$

The Lagrange multiplier is now a function of t, as we want the entire curve to lie on Σ. Varying w.r.t. λ trivially recovers the constraint $g(x, y, z) = 0$, whereas varying w.r.t. x_i (where $x_1 = x$, $x_2 = y$, $x_3 = z$) gives a system of ODEs.

$$\frac{d}{dt} \frac{\dot{x}_i}{\sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}} + \lambda\frac{\partial g}{\partial x_i} = 0$$

Alternatively, we could also have solved the constraints directly.

Proposition 9.7: Suppose the integrand is a function containing up to $y^{(n)}$ (the nth derivative), then the corresponding Euler-Lagrange equation is

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'} + \frac{d^2}{dx^2}\left(\frac{\partial f}{\partial y''}\right) + \cdots + (-1)^n \frac{d^n}{dx^n}\left(\frac{\partial f}{\partial y^{(n)}}\right) = 0 \qquad (9.11)$$

Proof. Assume minimum y exists, then perturb it, $y \to y + \varepsilon\eta$ s.t. $\eta = \eta' = \cdots = \eta^{(n-1)} = 0$ at α, β. We have

$$F[y + \varepsilon\eta] - F[y] = \varepsilon \int_\alpha^\beta \left(\frac{\partial f}{\partial y}\eta + \cdots + \frac{\partial f}{\partial y^{(n)}}\eta^{(n)}\right)dx + O(\varepsilon^2)$$

Apply Lemma 9.1, and use $\frac{\partial f}{\partial y^{(n)}} = (-1)^n \frac{d^n}{dx^n}\frac{\partial f}{\partial y^{(n)}}$, then we obtain (9.11). $\qquad\square$

Example 9.6. *The corresponding first integral for $f(y, y', y'')$ is $\frac{\partial f}{\partial y'} - \frac{d}{dx}\frac{\partial f}{\partial y''}$ equals to a constant.*

Example 9.7. *Extremize $F[y] = \int_0^1 (y'')^2 dx$, where $y(0) = y'(0) = y(1) = 0$ and $y'(1) = 1$. Then the first integral gives*

$$K = \frac{d}{dx}(2y'') \implies y = x^3 - x^2$$

This is an absolute minimum. Let $y_0 = x^3 - x^2$, and $\eta(0) = \eta'(0) = \eta(1) = \eta'(1) = 0$ (do not assume η small), then

$$F[y_0 + \eta] - F[y_0] = \int_0^1 \eta''^2 dx + 2\int_0^1 y_0''\eta'' dx$$

$$> 4\int_0^1 (3x - 1)\eta'' dx$$

$$= 4\left([-\eta']_0^1 + \int_0^1 \frac{d}{dx}(3x\eta')) - \eta' dx\right)$$

$$= 4([3x\eta']_0^1 - [3\eta]_0^1)$$

$$= 4$$

Hence y_0 is the absolute minimum of f.

9.3 Rayleigh-Ritz method

By Proposition 9.5, there exists a minimal extremal value, say λ_0. Let y_0 be the eigenfunction of (9.9c) with eigenvalue λ_0, then

$$\lambda_0 = \Lambda[y_0] \geq 0 \tag{9.12}$$

Then for all other y,

$$\Lambda[y] \geq \lambda_0 \tag{9.13}$$

with equality iff $y = y_0$. This gives a way to find the upper bound on λ_0.

Proposition 9.8 (Rayleigh-Ritz method)**:** The Rayleigh-Ritz method uses a trial function constructed from a linearly independent set of basis functions s.t. they all satisfy the b.c.s, i.e. $y_{trial} = \sum_i c_i y_i$ s.t. y_i satisfy the b.c.s.

$\Lambda[\psi]$ is minimized with respect to c_i and the minimum value is an overestimate for the lowest true eigenvalue λ_0.

$$\lambda_0 \leq \min_{c_i} \Lambda(c_i) \tag{9.14}$$

By superposing a linearly independent set of functions that satisfy the b.c.s, and minimizing the resulting quotient $\frac{F}{G}$ with respect to these parameters, we will get an upper bound on the lowest eigenvalue.

Proof. The first order variation of $\Lambda = \frac{F}{G}$ is $\delta\Lambda = \frac{1}{G}(\delta F - \Lambda \delta G)$. Hence, the extremal values of Λ are thus $\Lambda = \frac{F}{G}$. This necessarily is an upper bound unless the trial function is the eigenfunction with the lowest eigenvalue. $\qquad\square$

Example 9.8. *Consider the Bessel's equation*

$$y'' + \frac{1}{r}y' + \lambda y = 0$$

subject to finite $y(0)$ and $y(1) = 0$. The eigenvalue scale with the squared of eigenfrequency. Cast the equation into SL form, we have

$$(ry')' + \lambda r y = 0, \quad w = r$$

Then, let $F[y] = \int_0^1 ry'^2 dr$ and $G[y] = \int_0^1 ry^2 dr$. Try $y_{trial} = a + br^2 + cr^4$ where $y(1) = 0 \implies a + b + c = 0$. We only include even powers since the equation is unchanged when $r \to -r$. Next, compute

$$F[y_{trial}] = b^2 + \frac{8}{3}bc + 2c^2 := f(b,c)$$

$$G[y_{trial}] = \frac{1}{6}b^2 + \frac{5}{12}bc + \frac{4}{15}c^2 := g(b,c)$$

Extremize $\Lambda(b,c) := \frac{f(b,c)}{g(b,c)}$ w.r.t b and c to get

$$4(\Lambda - 6)b = (32 - 5\Lambda)c, \quad 5(32 - 5\Lambda)b = 16(2\Lambda - 15)c \implies \Lambda = \frac{8}{3}(8 \pm \sqrt{34})$$

choosing the minus sign to obtain the lower root which is $\lambda \approx 5.284$.

Remarks 9.4. *We can find higher eigenvalues using this method since all the eigenfunctions are pairwise orthogonal. Suppose we want to guess $y_{trial}^{(r)}$ where $\lambda_0 < \lambda_1 < \lambda_2 < \ldots$, then let $y_{trial}^{(1)} = \sum_{n=1}^{\infty} a_n y_n$ (omit y_0 to ensure orthogonal to y_0).*

$$\Lambda[y_{trial}^{(1)}] = \frac{\langle y | \mathcal{L}y \rangle}{\langle y | y \rangle_w} = \frac{\sum_{n=1}^{\infty} \lambda_n |a_n|^2}{\sum_{n=1}^{\infty} |a_n|^2} \tag{9.15}$$

Since $|a_n|^2 \geq 0$ and $\lambda_n \geq \lambda_1$ for $n \geq 2$, then

$$\sum_{n=1}^{\infty} \lambda_n |a_n|^2 \geq \lambda_1 \sum_{n=1}^{\infty} |a_n|^2 \implies \Lambda[y_{trial}^{(1)}] \geq \lambda_1$$

This gives an upper bound on λ_1 from any trial function orthogonal to y_0. A reasonably good guess will give a good estimate for λ_1. It is generally hard to guess $y_{trial}^{(1)}$. But for symmetric potentials, we necessarily (see Quantum Mechanics) must have symmetric ground states. Hence, we can guess an antisymemtric wavefunction for the first excited state.

10   Laplace and Poisson's equations

Solution by separation of variables of Laplace's equation in plane polar coordinates, and spherical polar coordinates (axisymmetric case); Legendre polynomials again. Solution of Poisson's equation as an integral. Uniqueness for Poisson's equation with Dirichlet boundary conditions. Green's identity. Green's function for Laplace's eqution with simple boundary conditions using the method of images. Applications to electrostatic fields and steady heat flow.   **[5]**

Definition 10.1 (Poisson's and Laplace's equation)**:** When $F(\mathbf{x}) \neq 0$,

$$\nabla^2 \phi(\mathbf{x}) = F(\mathbf{x}) \tag{10.1}$$

$\phi(\mathbf{x})$ satisfies the Poisson's equation. If $F = 0$, it is called Laplace's equation.

Remarks 10.1. *Poisson's equation is defined over a domain $\Omega \in \mathbb{R}^n$. We need to prescribe some boundary conditions for ϕ on the boundary $\partial\Omega$. If $\Omega = \mathbb{R}^n$, the boundary condition is for $|\mathbf{x}| \to \infty$.*

Definition 10.2 (Dirichlet and Neumann problem)**:** For $\nabla^2 \phi = F$ in Ω, the Dirichlet problem is

$$\phi(\mathbf{x}) = f(\mathbf{x}), \text{ on } \partial\Omega$$

In comparison, the Neumann problem is defined for

$$\frac{\partial \phi(\mathbf{x})}{\partial \mathbf{n}} = \mathbf{n} \cdot \boldsymbol{\nabla} \phi(\mathbf{x}) = g(\mathbf{x}), \text{ on } \partial\Omega$$

Remarks 10.2.

- *We assume the solution continuously approaches the boundary data as $\mathbf{x}$ approaches the boundary $\partial\Omega$ from within Ω, i.e. we assume that ϕ and $\boldsymbol{\nabla}\phi$ are continuous on the set $\Omega \cup \partial\Omega$;*

- *ϕ must be well-defined on all of Ω in order to satisfy $\nabla^2 \phi = 0$ in the domain Ω. For instance, $\nabla^2 \frac{1}{|\mathbf{x}|} = 0$ is not well-defined for $\mathbf{x} = \mathbf{0}$. $\frac{1}{|\mathbf{x}|}$ is singular at $\mathbf{x} = \mathbf{0}$.*

10.1   Solving for various geometry

10.1.1   3D Cartesian

Proposition 10.1: The general solution to Laplace's equation in 3D Cartesian coordinates is

$$\phi(x, y, z) = \sum_{l,m,n} a_{lmn} X_l(x) Y_m(y) Z_n(z) \tag{10.2}$$

Proof. The Laplace's equation (10.1) becomes

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial z^2} = 0 \tag{10.3}$$

We use separation of variables $\phi(x, y, z) = X(x)Y(y)Z(z)$ and choose separation constants to be

$$\frac{X''}{X} = -\lambda_l, \quad \frac{Y''}{Y} = -\lambda_m, \quad \frac{Z''}{Z} = -\lambda_n = \lambda_l + \lambda_m \tag{10.4}$$

The result follows trivially. $\square$

Example 10.1 (Steady heat conduction)**.** *Consider (10.1) but with $\frac{\partial \phi}{\partial t} = 0$ for a semi-infinite rectangular bar with b.c.s*

$$\phi(0, y, z) = \phi(a, y, z) = \phi(x, 0, z) = \phi(x, b, z) = 0, \quad \phi(x, y, 0) = 1, \quad \lim_{z \to \infty} \phi = 0$$

Based on the b.c.s, solve for eigenmodes successively

$$X_l = \sin \frac{l\pi x}{a}, \quad \lambda_l = \frac{l^2 \pi^2}{a^2}, \quad l \in \mathbb{Z}$$

$$Y_m = \sin\frac{m\pi y}{b}, \quad \lambda_m = \frac{m^2\pi^2}{b^2}, \quad m \in \mathbb{Z}$$

$$Z_{lm} = e^{-\sqrt{(l/a)^2+(m/b)^2}\,\pi z}, \quad l, m \in \mathbb{Z}$$

The general solution would be

$$\phi(x,y,z) = \sum_{l,m} a_{lm}\sin\frac{l\pi x}{a}\sin\frac{m\pi y}{b}e^{-\sqrt{(l/a)^2+(m/b)^2}\,\pi z}$$

Now fix a_{lm} using $\phi(x,y,0)=1$, and using Fourier series,

$$a_{lm} = \frac{2}{b}\int_0^b \frac{2}{a}\int_0^a \sin\frac{l\pi x}{a}\sin\frac{m\pi y}{b}dxdy = \frac{4a}{al\pi}\frac{4b}{m\pi}$$

with l, m odd. Otherwise $a_{lm} = 0$. The heat flow solution would be

$$\phi(x,y,z) = \sum_{l,m\ odd}\frac{16}{\pi^2 lm}\sin\frac{l\pi x}{a}\sin\frac{m\pi y}{b}e^{-\sqrt{(l/a)^2+(m/b)^2}\,\pi z}$$

10.1.2 2D plane polar

Proposition 10.2: The general solution to Laplace's equation in 2D plane polar coordinates is

$$\phi(r,\theta) = \frac{1}{2}a_0 + c_0\ln r + \sum_{m=1}^{\infty}((a_m\cos m\theta + b_m\sin m\theta)r^m + (c_m\cos m\theta + d_m\sin m\theta)r^{-m}) \quad (10.5)$$

Proof. The Laplace's equation (10.1) becomes

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\phi}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2\phi}{\partial\theta^2} = 0 \quad (10.6)$$

Use separation of variables $\phi = R(r)\Theta(\theta)$ to find the radial equation $r(rR')' - \lambda R = 0$ and $\Theta'' + \lambda\Theta = 0$ where λ is the separation constant. Again, impose periodic b.c.s, we get $\lambda = m^2$. As for the radial equation, try $R = \alpha r^\beta$, to get $\beta^2 - m^2 = 0 \implies \beta = \pm m$. If $m = 0$, then $(rR')' = 0 \implies R = \ln r$. All these combines to give (10.5). $\square$

Example 10.2 (Soap film on a unit disk). *Solve (10.6) with a distorted wire of radius $r = 1$ with given b.c.s $\phi(1,\theta) = f(\theta)$ to find $\phi(r,\theta)$ for $r < 1$. Impose regularity at $r = 0 \implies c_m = d_m = 0$ $\forall m$. At $r = 1$, $\phi(1,\theta) = f(\theta) = \frac{1}{2}a_0 + \sum_{m=1}^{\infty}a_m\cos(m\theta) + b_m\sin(m\theta)$. So the Fourier series coefficients are:*

$$a_m = \frac{1}{\pi}\int_0^{2\pi}f(\theta)\cos m\theta d\theta, \quad b_m = \frac{1}{\pi}\int_0^{2\pi}f(\theta)\sin m\theta d\theta$$

We see the high harmonics r^m, $m \gg 1$ are confined near the edges.

10.1.3 3D axisymmetric case in spherical polars

Proposition 10.3: The general solution to Laplace's equation in 3D spherical polar coordinates under axisymmetric condition (no dependence on the azimuthal angle) is

$$\phi(r,\theta) = \sum_{l=0}^{\infty}(a_l r^l + b_l r^{-l-1})P_l(\cos\theta) \quad (10.7)$$

where $P_l(x)$'s are the Legendre polynomials in Example 8.2.

Proof. The Laplace's equation (10.1) becomes

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\phi}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\phi}{\partial\theta}\right) = 0 \quad (10.8)$$

Use separation of variables $\phi(r, \theta) = R(r)\Theta(\theta)$ to get

$$(\sin\theta\,\Theta')' + \lambda\sin\theta\,\Theta = 0, \quad (r^2 R')' - \lambda R = 0 \tag{10.9}$$

After identifying $x = \cos\theta$, the polar equation has the form of the Legendre's equation (8.19). The eigenvalues are thus $\lambda = l(l+1)$ with corresponding eigenfunctions $\Theta_l(\theta) = P_l(\cos\theta)$. As for the radial equation, we seek solution $R = \alpha r^\beta$ to obtain

$$\beta(\beta + 1) - l(l+1) = 0 \implies \beta = l, -l - 1$$

giving us the general solution (10.8). The constants a_l, b_l are determined by b.c.s, usually at fixed $r = a$, together with the orthogonality properties of $P_l(x)$ (8.22). $\qquad\square$

Example 10.3. *Solve $\nabla^2\phi = 0$ with axisymmetric b.c.s at $r = 1$ on a unit sphere, s.t. $\phi(1, \theta) = f(\theta)$. If $f(\theta)$ has such an expansion form $f(\theta) = \sum_{l=0}^\infty a_l P_l \cos\theta$, then*

$$a_l = \frac{2l+1}{2}\int_{-1}^{1} f(\theta)P_l(\cos\theta)d(\cos\theta)$$

Remarks 10.3. *The generating function (8.21b) is obtained for the potential $\Phi(\mathbf{r})$ in the following situation: a charge placed at point $\mathbf{r_0} = (0, 0, 1)$.*

$$\Phi(\mathbf{r}) = \frac{1}{|\mathbf{r} - \mathbf{r_0}|} = \frac{1}{\sqrt{x^2 + y^2 + (z-1)^2}} = \frac{1}{\sqrt{r^2 - 2r\cos\theta + 1}} = \sum_{l=0}^\infty P_l(\cos\theta)r^l$$

10.2 Uniqueness

Theorem 10.1 (Uniqueness theorem)**:** The solution to the Dirichlet problem is entirely unique, but that of the Neumann problem is unique up to an additive constant.

Proof. Let ϕ be the solution to the homogeneous

$$\nabla^2\phi = 0, \text{ in } \Omega, \quad B\phi = 0, \text{ on } \partial\Omega$$

where B is a linear operator s.t. $B\phi = \phi$ for Dirichlet and $B\phi = \frac{\partial\phi}{\partial\mathbf{n}}$ for Neumann. Consider the non-negative functional

$$I[\phi] = \int_\Omega |\boldsymbol{\nabla}\phi|^2 dV \geq 0$$

Clearly, $I[\phi] = 0$ iff $\boldsymbol{\nabla}\phi = \mathbf{0}$ (norm of a vector is zero only if the vector is the zero vector) in Ω. Note that

$$I[\phi] = \int_\Omega \boldsymbol{\nabla}\phi\cdot\boldsymbol{\nabla}\phi\,dV = \int_\Omega (\boldsymbol{\nabla}\cdot(\phi\boldsymbol{\nabla}\phi) - \phi\nabla^2\phi)dV$$

The second term vanishes since $\nabla^2\phi = 0$ in Ω. Using the divergence theorem on the first term, we find $I[\phi] = \int_{\partial\Omega} \phi\frac{\partial\phi}{\partial\mathbf{n}}dS$ which is zero since $B\phi = 0$ on $\partial\Omega$. Hence, $\boldsymbol{\nabla}\phi = 0$ throughout Ω, i.e. ϕ is constant. For the Dirichlet case, since $\phi = 0$ on $\partial\Omega$. By continuity of ϕ on $\Omega \cup \partial\Omega$ we conclude $\phi = 0$ throughout Ω, hence the solution to the Dirichlet problem is entirely unique. But for the Neumann case, we only know $\frac{\partial\phi}{\partial\mathbf{n}} = \mathbf{0}$ on $\partial\Omega$, and the solution is unique up to an additive constant. $\qquad\square$

Example 10.4. *Consider a charge density of the form $\rho(\mathbf{x}) = F(r)$ for $r \geq a$ and 0 otherwise. We will show that there is no electric field in $r < a$. The potential ϕ satisfies $\nabla^2\phi = -\frac{\rho(\mathbf{x})}{\epsilon_0} = 0$.*

In addition, the problem is spherically symmetric, so $\phi = \phi(r)$. In particular, ϕ is constant on the boundary $r = a$, i.e. $\phi(a) = \phi_0$. ϕ_0 still solves Laplace's equation inside the ball $r < a$ and has the correct boundary behaviour, so by uniqueness, we have found the solution. Since ϕ is constant throughout $r < a$, $\mathbf{E} = -\boldsymbol{\nabla}\phi = \mathbf{0}$ there.

10.3   Fundamental solution

Definition 10.3 (Poisson's equation and fundamental solution)**:** Consider the Poisson's equation

$$\nabla^2 \phi = -\rho(\mathbf{r}), \quad \phi : \mathbb{R}^n \to \mathbb{C} \tag{10.10}$$

The fundamental solution (satisfy same b.c.s as ϕ) to the Poisson's equation would satisfy

$$\nabla^2 G_{FS}(\mathbf{r}, \mathbf{r}') = \delta^{(n)}(\mathbf{r} - \mathbf{r}') \tag{10.11}$$

Remarks 10.4.

1. G_{FS} *inherits spherical symmetry from (10.10), i.e.* $G_{FS}(\mathbf{r}, \mathbf{r}') = G_{FS}(|\mathbf{r} - \mathbf{r}'|)$. *W.l.o.g.* $\mathbf{r}' = \mathbf{0}$, *so* $G_{FS} = G_{FS}(r)$.

2. *The Dirac delta function* $\delta^{(n)}(\mathbf{r})$ *on* $\mathbb{R}^n$ *has the following properties:*

   - $\delta(\mathbf{r} - \mathbf{r}') = 0 \; \forall \mathbf{r} \neq \mathbf{r}'$
   -
   $$\int_D \delta(\mathbf{r} - \mathbf{r}') d^3\mathbf{r} = \begin{cases} 1 & \mathbf{r}' \in D \\ 0 & otherwise \end{cases} \tag{10.12}$$

   - $\int_D f(\mathbf{r})\delta(\mathbf{r} - \mathbf{r}') d^3\mathbf{r} = f(\mathbf{r}').$

Proposition 10.4: The fundamental solution to the Poisson's equation in $\mathbb{R}^n$ have the form

$$G_{FS}(\mathbf{r} - \mathbf{r}') = \begin{cases} |\mathbf{r} - \mathbf{r}'| + C_{n=1} & n = 1 \\ \frac{1}{2\pi} \ln|\mathbf{r} - \mathbf{r}'| + C_{n=2} & n = 2 \\ \frac{-1}{A_n(n-2)} \frac{1}{|\mathbf{r}-\mathbf{r}'|^{n-2}} + C_{n\geq 3} & n \geq 3 \end{cases} \tag{10.13}$$

where A_n is the surface area of a unit S^{n-1} sphere.

Proof. Integrate over a n-sphere of radius r around $\mathbf{r}' = \mathbf{0}$, i.e. $B_r = \{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x} - \mathbf{y}| \leq r\}$, so using Divergence theorem for an outward pointing radial normal,

$$1 = \int_B \delta(\mathbf{r}) d^3\mathbf{r} = \int_B \nabla^2 G_{FS} d^3\mathbf{r} = \int_S \boldsymbol{\nabla} G_{FS} \cdot \mathbf{n} dS = \int_\Omega \frac{\partial G_{FS}}{\partial r} r^{n-1} d\Omega_n = r^{n-1} \int_{S^{n-1}} d\Omega_n \frac{\partial G_{FS}}{\partial r}$$

where $d\Omega_n$ is a measure on a unit S^{n-1} sphere, i.e. $d\Omega_2 = d\theta$, $d\Omega_3 = \sin\theta d\theta d\phi$. The remaining angular integral gives the total surface area A_n of our S^{n-1} sphere. The result is obtained by integrating $\frac{dG_{FS}}{dr} = \frac{1}{r^{n-1}A_n}$ and shift our origin back to arbitrary $\mathbf{r}'$. $\qquad\square$

Remarks 10.5.

1. *For* $n \geq 3$, *we can set* $\lim_{r \to 0} G_{FS}(r) = 0$ *and so the integration constant is 0. So for arbitrary* $\mathbf{r}'$, *we have the fundamental solution in three-dimensions to be*

   $$G(\mathbf{r}, \mathbf{r}') = \frac{-1}{4\pi|\mathbf{r} - \mathbf{r}'|} \tag{10.14}$$

2. *But for* $n = 1, 2$, *some other condition (specific to the problem) must instead be imposed to obtain the integration constant (which in general cannot be zero).*

3. *When* $n > 1$, *these fundamental solutions satisfy the Laplace's equation* $\forall \mathbf{r} \in \mathbb{R}^n - \{\mathbf{r}'\}$, *but that they become singular at* $\mathbf{r} = \mathbf{r}'$.

Proposition 10.5 (Green's identities)**:** Suppose $D \subset \mathbb{R}^n$ is a compact set with boundary ∂D, and let $\phi, \psi : D \to \mathbb{R}$ be a pair of functions on D that are regular throughout D, then they satisfy

1.
   $$\int_D \boldsymbol{\nabla} \cdot (\phi \boldsymbol{\nabla} \psi) d^3\mathbf{r} = \int_{\partial D} \phi \boldsymbol{\nabla} \psi \cdot \mathbf{n} dS \tag{10.15}$$

2.

$$\int_{\partial D} \left(\phi\frac{\partial\psi}{\partial n} - \psi\frac{\partial\phi}{\partial n}\right)dS = \int_D (\phi\nabla^2\psi - \psi\nabla^2\phi)d^3\mathbf{r} \tag{10.16}$$

3. Take ϕ and ψ to satisfy $\nabla^2\phi = -\rho$ and $\psi = G_{FS}(\mathbf{r}, \mathbf{r}')$.

$$\phi(\mathbf{r}') = \int_D G_{FS}(\mathbf{r}, \mathbf{r}')(-\rho(\mathbf{r}))dV + \int_{\partial D} \left(\phi(\mathbf{r})\frac{\partial G_{FS}}{\partial n} - G_{FS}(\mathbf{r}, \mathbf{r}')\frac{\partial\phi}{\partial n}\right)dS \tag{10.17}$$

Proof.

1. Follows from Divergence Theorem with outward pointing normal to $\partial\Omega$,

2. Interchange ϕ and ψ in (10.15) and subtract one from another,

3. Take ϕ in (10.16) s.t. $\nabla^2\phi = -\rho$ and $\psi = G_{FS}(\mathbf{r}, \mathbf{r}')$. To invoke Divergence theorem, we have to ensure functions are regular throughout Ω, so eliminating a neighbourhood of $\mathbf{r}'$ (where $G_{FS}(\mathbf{r}, \mathbf{r}')$ is singular at). Take $D = B_r = B_\epsilon = \{\mathbf{x} \in \mathbb{R}^n : \epsilon \le |\mathbf{x} - \mathbf{y}| \le r\}$, then apply (10.16), LHS is $\int_D \phi 0 - G_{FS}\nabla^2\phi dV = \int_D \rho G_{FS}dV$ while the RHS is

$$\int_{\partial D} \phi(\mathbf{n}\cdot\boldsymbol{\nabla}G_{FS}) - G_{FS}(\mathbf{n}\cdot\boldsymbol{\nabla}\phi)dS = \int_{S_r^{n-1}\cup S_\epsilon^{n-1}} \phi(\mathbf{n}\cdot\boldsymbol{\nabla}G_{FS}) - G_{FS}(\mathbf{n}\cdot\boldsymbol{\nabla}\phi)dS$$

On the inner boundary - a sphere of radius ϵ - the outward pointing normal is $\mathbf{n} = -\hat{\mathbf{r}}$, so

$$\mathbf{n}\cdot\boldsymbol{\nabla}G_{FS}|_{\text{inner}} = -\frac{1}{A_n}\frac{1}{\epsilon^{n-1}}, \quad dS = \epsilon^{n-1}d\Omega_n$$

Hence,

$$\int_{S_\epsilon^{n-1}} \phi(\mathbf{n}\cdot\boldsymbol{\nabla}G_{FS}) - G_{FS}(\mathbf{n}\cdot\boldsymbol{\nabla}\phi)dS = -\frac{1}{A_n}\int\phi d\Omega_n + \epsilon\int(\mathbf{n}\cdot\boldsymbol{\nabla}\phi)d\Omega_n \to -\phi(\mathbf{r}')$$

where since ϕ is regular at $\mathbf{r}'$ by assumption, the value of the integral $\int(\mathbf{n}\cdot\boldsymbol{\nabla}\phi)d\Omega_n$ is bounded. Furthermore, we have $\frac{1}{A_n}\int\phi d\Omega_n \to \phi(\mathbf{r}')$, i.e. the average value of ϕ on a small sphere is equal to that at the centre of the sphere. The result (10.17) follows.

$\square$

Example 10.5 (Dirichlet's problem). *Solve $\nabla^2\phi = -\rho$ on D with inhomogeneous Dirichlet b.c.s $\phi(\mathbf{r}) = h(\mathbf{r})$ on ∂D. The Dirichlet's Green's function will satisfy*

1. $\nabla^2 G(\mathbf{r}, \mathbf{r}') = 0 \ \forall\mathbf{r} \neq \mathbf{r}'$

2. $G(\mathbf{r}, \mathbf{r}') = 0$ *on the boundary ∂D*

3. $G(\mathbf{r}, \mathbf{r}') = G_{FS}(\mathbf{r}, \mathbf{r}') + H(\mathbf{r}, \mathbf{r}')$ *with $\nabla^2 H = 0$ and H finite $\forall\mathbf{r} \in D$.*

Green's second identity (10.16) with $\nabla^2\phi = -\rho$ and $\nabla^2 H = 0$ gives

$$\int_{\partial D} \left(\phi\frac{\partial H}{\partial n} - H\frac{\partial\phi}{\partial n}\right)dS = \int_D H\rho d^3\mathbf{r}$$

Now $G_{FS} = G - H$ in Green's third identity (10.17):

$$\phi(\mathbf{r}') = \int_D (G - H)(-\rho)d^3\mathbf{r} + \int_{\partial D} \left(\phi\frac{\partial(G - H)}{\partial n} - (G - H)\frac{\partial\phi}{\partial n}\right)dS$$

Subtract H terms above ($G = 0$, $\phi = h$ on ∂D), then the solution to the Dirichlet problem is

$$\phi(\mathbf{r}') = \int_D G(\mathbf{r}, \mathbf{r}')(-\rho(\mathbf{r}))d^3\mathbf{r} + \int_{\partial D} h(\mathbf{r})\frac{\partial G(\mathbf{r}, \mathbf{r}')}{\partial n}dS \tag{10.18}$$

Can further show Green's function is symmetric, i.e. $G(\mathbf{r}, \mathbf{r}') = G(\mathbf{r}', \mathbf{r}) \ \forall\mathbf{r} \neq \mathbf{r}'$.

Example 10.6 (Neumann's problem). *For the Neumann problem, we consider Neumann b.c.s $\mathbf{n}\cdot\boldsymbol{\nabla}\phi = g(\mathbf{r})$ on ∂D, and seek a harmonic function H' to make $\mathbf{n}\cdot\boldsymbol{\nabla}G = \frac{1}{A}$ on ∂D, where $A = \oint_{\partial D} dS$. Note that H' is unique up to an additive constant, then the solution to the Neumann problem is*

$$\phi(\mathbf{r}') = -\int_{\partial\Omega} g(\mathbf{r})G(\mathbf{r}, \mathbf{r}')dS - \int_D \rho(\mathbf{r})G(\mathbf{r}, \mathbf{r}')dV + \frac{1}{A}\oint_{\partial D} \phi(\mathbf{r})dS \tag{10.19}$$

10.4   Method of images

For symmetric domain D, we can construct Green's function with $G = 0$ on ∂D by cancelling the boundary signal with an opposite mirror or image Green's function placed outside of D.

Example 10.7 (Laplace's equation on half-plane in $\mathbb{R}^3$). *Solve $\nabla^2 \phi = 0$ on $D = \{(x, y, z) : z > 0\}$ with $\phi(x, y, 0) = \{h(x, y)$, $\phi \to 0$ as $|\mathbf{r}| \to \infty$. Now $G_{FS}(\mathbf{r}, \mathbf{r}') \to 0$ as $|\mathbf{r}| \to \infty$, but $G_{FS} \neq 0$ at $z = 0$. So for $G_{FS}(\mathbf{r}, \mathbf{r}')$ at $\mathbf{r}' = (x', y', z')$, subtract image $G_{FS}(\mathbf{r}, \mathbf{r}'')$ with $\mathbf{r}'' = (x', y', -z')$, to get $G(\mathbf{r}, \mathbf{r}') = 0$:*

$$-\frac{1}{4\pi|\mathbf{r} - \mathbf{r}'|} - \frac{-1}{4\pi|\mathbf{r} - \mathbf{r}'|} = \frac{1}{4\pi}\left(\frac{-1}{\sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2}} + \frac{1}{\sqrt{(x - x')^2 + (y - y')^2 + (z + z')^2}}\right) = 0$$

when $z = 0$, i.e. Dirichlet's b.c.s on all ∂D. We also have

$$\left.\frac{\partial G}{\partial n}\right|_{z=0} = \left.\frac{\partial G}{\partial z}\right|_{z=0} = -\frac{1}{4\pi}\left(\frac{z - z'}{|\mathbf{r} - \mathbf{r}'|^3} - \frac{z + z'}{|\mathbf{r} - \mathbf{r}''|^3}\right) = \frac{z'}{2\pi}\left[(x - x')^2 + (y - y')^2 + z'^2\right]^{-3/2}$$

at $z = 0$. The solution is then from (10.18) to be

$$\phi(x', y', z') = \frac{z'}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{h(x, y)}{((x - x')^2 + (y - y')^2 + z'^2)^{3/2}}dxdy$$

To apply the method of images to other simple geometry (with high symmetry) such as sphere, quarter-plane, etc.

11 Cartesian tensors

Transformation laws, addition, multiplication, contraction. Isotropic tensors, symmetric and anti-symmetric tensors. Principal axes and diagonalisation. Tensor fields, e.g. conductivity, polarizability, elasticity **[4]**

11.1 Transformation laws

Proposition 11.1 (Vectors)**:** The components of a vector with respect to two distinct orthonormal bases $\{\mathbf{e_j}\}$ and $\{\mathbf{e_j'}\}$ are related by an orthogonal matrix, i.e.

$$v_i' = R_{ij}v_j$$

Proof. The vector itself does not change under a change of bases, but its components do.

$$\mathbf{v} = v_i\mathbf{e_i} = v_i'\mathbf{e_i'}$$

Dotting both sides with $\mathbf{e_j}$ and $\mathbf{e_j'}$ separately:

$$v_i\delta_{ij} = v_i'\mathbf{e_i'}\cdot\mathbf{e_j} \implies v_j = v_i'\mathbf{e_i'}\cdot\mathbf{e_j}$$

$$v_i\mathbf{e_i}\cdot\mathbf{e_j'} = v_i'\delta_{ij} \implies v_j' = v_i\mathbf{e_i}\cdot\mathbf{e_j'}$$

If we define $R_{ji} := \mathbf{e_j'}\cdot\mathbf{e_i}$, then we see that $v_j' = R_{ji}v_i$ and $v_i = R_{ji}v_j'$. Next we need to show R is an orthogonal matrix.

$$v_i = R_{ij}v_j' = R_{ij}R_{kj}v_k \implies R_{ij}R_{kj} = \delta_{ik}$$

Hence, $RR^T = I$, i.e., R is orthogonal. $\qquad\square$

Moreover, since both systems are right-handed, i.e.

$$1 = \mathbf{e_1}\cdot(\mathbf{e_2}\times\mathbf{e_3}) = R_{i1}R_{j2}R_{k3}\mathbf{e_i'}\cdot(\mathbf{e_j'}\times\mathbf{e_k'}) = R_{i1}R_{j2}R_{k3}\epsilon_{ijk} = \det(R)$$

Hence, R is a rotation.

Remarks 11.1. *An orthogonal matrix R has $\det R = \pm1$. If $\det R = +1$, R is a rotation matrix, whereas if $\det R = -1$, R is an improper rotation (composition of rotation with a reflection). Improper rotations thus do not form a group. See Group theory later.*

Definition 11.1 (Axial vector)**:** A Cartesian axial-vector $\mathbf{a}$ is defined with respect to a set of orthonormal basis vectors $\mathbf{e_i}$ such that the coefficients a_i' w.r.t. another orthonormal basis $\mathbf{e_i'}$ is given by

$$a_i' = \det R\, R_{ij}a_j$$

In particular, when $\det R = -1$, the handedness of the coordinate system changes.

Example 11.1. *Angular momentum $\mathbf{L} = \mathbf{r}\times\mathbf{p} = (r,0,0)\times(0,p,0) = (0,0,rp)$ is an axial vector. Consider a passive transformation in the form of a reflectional transformation of the bases in the xy plane where $\mathbf{e_{1,2}'} = \mathbf{e_{1,2}}$ and $\mathbf{e_3'} = -\mathbf{e_3}$, then $\mathbf{L}$ is unchanged. Alternatively, consider an active transformation (a physical reflection of the vectors in the origin but keeping the same right-handed basis), i.e. $\mathbf{r}' = -\mathbf{r}$ and $\mathbf{p}' = -\mathbf{p}$. $\mathbf{L}$ is unchanged.*

Proposition 11.2 (Scalars)**:** Scalars are invariant under change of bases.

Proof. In terms of basis vectors $\{\mathbf{e_i}\}$, the scalar product $\sigma = \mathbf{a}\cdot\mathbf{b}$ is

$$\sigma = a_ib_j(\mathbf{e_i}\cdot\mathbf{e_j}) = a_ib_j\delta_{ij} = a_ib_i$$

If instead we used $\{\mathbf{e_i'}\}$, $\sigma' = a_i'b_i'$, then

$$\sigma' = R_{ip}R_{iq}a_pb_q = \delta_{pq}a_pb_q = a_pb_p = \sigma$$

where $a_i' = R_{ip}a_p$ and $b_i' = R_{iq}b_q$. $\qquad\square$

In fact, scalars are rank zero tensors and vectors are rank one tensors. Rank two tensors transform like

$$T_{ij}' = R_{ip}R_{jq}T_{pq}$$

11.2   Cartesian tensors of rank n

We generalize the transformation law to a rank n Cartesian tensor.

Definition 11.2 (Rank n Cartesian tensor)**:** A tensor of rank n has components $T_{ij\ldots k}$ (with n indices) with respect to each Cartesian basis $\{\mathbf{e_i}\}$ transform as

$$T'_{ij\ldots k} = R_{ip}R_{jq}\ldots R_{kr}T_{pq\ldots r}$$

upon changing from one right-handed Cartesian basis to another. Here $R_{ip}R_{jp} = \delta_{ij}$, corresponding to a rotation.

Notation 11.1. *We sometimes refer to tensors as their components, so $T_{ij\ldots k}$ is understood to be a tensor, rather than the components of a tensor.*

Example 11.2. *If $u_i, v_j, \ldots, w_k$ are n vectors, then $T_{ij\ldots k} = u_i v_j \ldots w_k$ define the components of an nth rank tensor since it transforms like*

$$T'_{ij\ldots k} = u'_i v'_j \ldots w'_k = R_{ip}R_{jq}\ldots R_{kr}u_p v_q \ldots w_r = R_{ip}R_{jq}\ldots R_{kr}T_{pq\ldots r}$$

Example 11.3. *The Kronecker delta δ_{ij} is a rank two tensor and invariant under a change of bases.*

$$\delta'_{ij} = R_{ip}R_{jq}\delta_{pq} = R_{ip}R_{jp} = \delta_{ij}$$

But the Levi-Civita symbol ϵ_{ijk} is a rank three pseudo-tensor. Pseudo-tensor since

$$\epsilon'_{123} = L_{1p}L_{2q}L_{3r}\epsilon_{pqr} = \det L$$

where we invoked the definition of determinant. In this case, ϵ'_{ijk} transforms like

$$\epsilon'_{ijk} = \det R R_{ip}R_{jq}R_{kr}\epsilon_{pqr} = \det(R)^2 \epsilon_{ijk} = \epsilon_{ijk}$$

We see that the Levi-Civita symbol transforms as a pseudo tensor.

Example 11.4. *The gradient is a rank one tensor. Since $x_j = R_{ij}x'_i$, we have $\frac{\partial x_j}{\partial x'_k} = R_{ij}\frac{\partial x'_i}{\partial x'_k} = R_{ij}\delta_{ik} = R_{kj}$, hence the chain rule gives*

$$\frac{\partial}{\partial x'_i} = \frac{\partial x_j}{\partial x'_i}\frac{\partial}{\partial x_j} = R_{ij}\frac{\partial}{\partial x_j}$$

Remarks 11.2. *Not all matrices are tensors. If it is a tensor of rank n, it must satisfy the corresponding transformation rule.*

Definition 11.3 (Tensor addition)**:** Two tensors of the same rank can be added in any coordinate system.

$$(T+S)_{ij\ldots k} = T_{ij\ldots k} + S_{ij\ldots k}$$

in any coordinate system. Both sides have the same rank.

Definition 11.4 (Scalar multiplication)**:** A tensor T of rank n can be multiplied by a scalar α.

$$(\alpha A)_{ij\ldots k} = \alpha A_{ij\ldots k}$$

It is trivial to check for the above two, their ranks will still be n.

Definition 11.5 (Tensor product)**:** Let T be a tensor of rank n and S be a tensor of rank m. The tensor product $T \otimes S$ is a tensor of rank $n + m$ defined by

$$(T \otimes S)_{x_1 x_2 \cdots x_n y_1 y_2 \cdots y_m} = T_{x_1 x_2 \cdots x_n}S_{y_1 y_2 \cdots y_n}$$

Once again, we can check that the resulting tensor has rank $n + m$.

Definition 11.6 (Tensor contraction)**:** For a tensor T of rank n with components $T_{ijp\cdots q}$, we can contract on the indices i, j (can contract on any pair) to obtain a new tensor of rank $n - 2$:

$$S_{p\cdots q} = \delta_{ij}T_{ijp\cdots q} = T_{iip\cdots q}$$

$$T'_{iik...l} = R_{ip}R_{iq}R_{kr}\ldots R_{ls}T_{pqr...s} = \delta_{pq}R_{kr}\ldots R_{ls}T_{pqr...s} = R_{kr}\ldots R_{ls}T_{ppr...s}$$

Definition 11.7 (Symmetric)**:** Tensor $T_{ij...k}$ is symmetric in indices (i, j) if

$$T_{ij...k} = T_{ji...k}$$

Proposition 11.3: Symmetric tensors remain symmetric after change of bases.

Proof.

$$T'_{ij...k} = R_{ip}R_{jq}\ldots R_{kr}T_{pq...r} = R_{ip}R_{jq}\ldots R_{kr}T_{pq...r} = R_{iq}R_{jp}\ldots R_{kr}T_{pq...r} = T'_{ji...k}$$

$\square$

Similarly, one can show the same for anti-symmetry in (i, j). A tensor that is totally symmetric (anti-symmetric) if it is symmetric (anti-symmetric) in each pair of indices.

Example 11.5. *Tensors δ_{ij} and $a_i a_j a_k$ are totally symmetric tensors of rank two and three respectively. ϵ_{ijk} is totally anti-symmetric.*

11.3 Tensor calculus

Definition 11.8 (Tensor field)**:** A tensor field of rank n $T_{ij...k}(\mathbf{x})$ is s.t. for each $\mathbf{x_0} \in \mathbb{R}^3$, $T_{ij...k}(\mathbf{x_0})$ define the components of an nth rank Cartesian tensor.

Proposition 11.4: If $T_{ij...k}(\mathbf{x})$ is a tensor field of rank n, then

$$\frac{\partial}{\partial x_{m_1}}\frac{\partial}{\partial x_{m_2}}\cdots\frac{\partial}{\partial x_{m_r}}T_{ij...s}(\mathbf{x})$$

is a tensor field of rank $r + s$.

Proof. Denote this quantity by a rank $r + s$ tensor $A_{m_1 m_2...m_r ij...s}$, then it transforms like

$$A'_{m_1 m_2...m_r ij...s} = R_{m_1 n_1}R_{m_2 n_2}\ldots R_{m_r n_r}\frac{\partial}{\partial x_{n_1}}\frac{\partial}{\partial x_{n_2}}\cdots\frac{\partial}{\partial x_{n_r}}R_{id}R_{je}\ldots R_{kf}T_{de...f}(\mathbf{x})$$

which is basically $R_{m_1 n_1}\ldots R_{m_r n_r}R_{id}\ldots R_{kf}A_{n_1...n_r a...f}$. $\square$

Example 11.6. *As mentioned, the gradient of a scalar field is a rank one tensor field. The divergence of a vector field transform like a scalar field*

$$\frac{\partial v'_i}{\partial x'_i} = R_{ip}\frac{\partial}{\partial x_p}R_{iq}v_q = R_{ip}R_{iq}\frac{\partial v_q}{\partial x_p} = \delta_{pq}\frac{\partial v_q}{\partial x_p} = \frac{\partial v_p}{\partial x_p}$$

The curl transforms as a vector field (rank one tensor field).

$$\epsilon'_{ijk}\frac{\partial v'_k}{\partial x'_j} = R_{ia}R_{jb}R_{kc}\epsilon_{abc}R_{jp}\frac{\partial}{\partial x_p}R_{kq}v_q = R_{ia}\epsilon_{abc}R_{jb}R_{jp}R_{kc}R_{kq}\frac{\partial v_q}{\partial x_p} = R_{ia}\epsilon_{abc}\delta_{bp}\delta_{cq}\frac{\partial v_q}{\partial x_p} = R_{ia}\epsilon_{abc}\frac{\partial v_c}{\partial x_b}$$

Proposition 11.5 (Generalized Divergence theorem)**:** For a smooth tensor field $T_{ij...k...l}(\mathbf{x})$ we have

$$\int_V \frac{\partial T_{ij...k...l}}{\partial x_k}dV = \int_{\partial V} T_{ij...k...l}n_k dS$$

Proof. Apply the usual Divergence theorem to the vector field $v_k := a_i b_j \ldots c_l T_{ij...k...l}$ where $a_i, b_j, \ldots, c_k$ are components of constant vectors $\mathbf{a}, \mathbf{b}, \ldots \mathbf{c}$. On the RHS the only free index is k since all of the others have been contracted. This gives

$$\int_V \frac{\partial v_k}{\partial x_k}dV = a_i b_j \ldots c_l \int_V \frac{\partial T_{ij...k...l}}{\partial x_k}dV,$$

$$\int_{\partial V} v_k n_k dS = a_i b_j \ldots c_l \int_{\partial V} T_{ij...k...l}n_k dS$$

The result follows since the vectors were arbitrary. $\square$

11.4   Rank 2 tensors

We may decompose an arbitrary rank two tensor into its symmetric and anti-symmetric part. The original tensor has $3^2 = 9$ independent components, but the components of a symmetric tensor S_{ij} are completely determined by the entries with $i \geq j$, of which there are $1 + 2 + 3 = 6$ components. Whereas an anti-symmetric tensor A_{ij} is completely determined by the components with $i > j$ (since the diagonal components are necessarily zero), of which there are $1 + 2 = 3$ components. Also, recall that A_{ij} is equivalent to $\epsilon_{ijk}\omega_k$.

Proposition 11.6: Every rank two tensor can be decomposed uniquely as

$$T_{ij} = S_{ij} + \epsilon_{ijk}\omega_k$$

where the vector $\boldsymbol{\omega}$ is defined by $\omega_i = \frac{1}{2}\epsilon_{ijk}T_{jk}$ and S_{ij} is symmetric.

Proof. Identify $S_{ij} = \frac{1}{2}(T_{ij} + T_{ji})$. Next show $\epsilon_{ijk}\omega_k = \frac{1}{2}(T_{ij} - T_{ji})$.

$$\epsilon_{ijk}\omega_k = \frac{1}{2}\epsilon_{ijk}\epsilon_{klm}T_{lm} = \frac{1}{2}(\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl})T_{lm} = \frac{1}{2}(T_{ij} - T_{ji})$$

Now suppose there were two such decompositions

$$S_{ij} + \epsilon_{ijk}\omega_k = \tilde{S}_{ij} + \epsilon_{ijk}\tilde{\omega}_k$$

then taking symmetric parts of both sides yeilds $S_{ij} = \tilde{S}_{ij}$, and so $\epsilon_{ijk}\omega_k = \epsilon_{ijk}\tilde{\omega}_k$. The latter implies $\omega_k = \tilde{\omega}_k$, hence the decomposition is unique.   □

Example 11.7 (Elastic Body). *Suppose that each point $\mathbf{x}$ in an elastic body undergoes a small displacement of $\mathbf{u}(\mathbf{x})$, then two nearby points that were initially separated by $\delta\mathbf{x}$ become separated by*

$$(\mathbf{x} + \delta\mathbf{x} + \mathbf{u}(\mathbf{x} + \delta\mathbf{x})) - \mathbf{x} - \mathbf{u}(\mathbf{x}) = \delta\mathbf{x} + \mathbf{u}(\mathbf{x} + \delta\mathbf{x}) - \mathbf{u}(\mathbf{x})$$

so the change in displacement between teo points in the body is $\mathbf{u}(\mathbf{x} + \delta\mathbf{x}) - \mathbf{u}(\mathbf{x})$. This gives a measure of how much deformation happens to the body. Using Taylor's theorem with Cartesian coordinates and suffix notation

$$u_i(\mathbf{x} + \delta\mathbf{x}) - u_i(\mathbf{x}) = \frac{\partial u_i}{\partial x_j}\delta x_j + o(\delta\mathbf{x})$$

By Proposition 11.6, decompose the second rank tensor

$$\frac{\partial u_i}{\partial x_j} = e_{ij} + \epsilon_{ijk}\omega_k$$

where $e_{ij} = \frac{1}{2}(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i})$ is called the linear strain tensor and $\omega_i = \frac{1}{2}\epsilon_{ijk}\frac{\partial u_j}{\partial x_k} = -\frac{1}{2}(\boldsymbol{\nabla} \times \mathbf{u})_i$. This tells us that

$$u_i(\mathbf{x} + \delta\mathbf{x}) - u_i(\mathbf{x}) = e_{ij}\delta x_j + [\boldsymbol{\omega} \times \delta\mathbf{x}]_i + o(\delta\mathbf{x})$$

the first term tells us about how the material stretches and/or compresses internally, while the second term is a solid body rotation, a change in displacement owing to a small rotation by an angle θ is

$$R\delta\mathbf{x} - \delta\mathbf{x} = \theta(\mathbf{n} \times \delta\mathbf{x}) + o(\theta)$$

Proposition 11.7: Suppose a body with uniform density $\rho(\mathbf{x})$ occupies a volume V. The moment of inertia is a rank-two tensor

$$I_{ij} = \int_{\mathcal{V}} \rho(\mathbf{x})(x_k x_k \delta_{ij} - x_i x_j)dV$$

where the integral is taken over $\mathcal{V} = \{(x_1, x_2, x_3) : \mathbf{x} = x_i\mathbf{e_i} \in V\}$. This is the inertia tensor.

Proof. Suppose each point in the body is rotating with constant angular velocity $\boldsymbol{\omega}$ about the origin, so the velocity of a point at $\mathbf{x} \in V$ is $\mathbf{v} = \boldsymbol{\omega} \times \mathbf{x}$. The total angular momentum about the origin is

$$L_i = \int_{\mathcal{V}} \rho(\mathbf{x})[\mathbf{x} \times (\boldsymbol{\omega} \times \mathbf{x})]_i dV = \int_{\mathcal{V}} \rho(\mathbf{x})(x_k x_k \omega_i - x_i x_j \omega_j) dV$$

where $\mathcal{V}' = \{(x_1', x_2', x_3') : \mathbf{x} = x_i' \mathbf{e_i'} \in V\}$. Next, we show it is a symmetric tensor of rank 2:

$$I_{ij}' = \int_{\mathcal{V}'} \rho(\mathbf{x})(x_k' x_k' \delta_{ij} - x_i' x_j') dV = R_{ip} R_{jq} \int_{\mathcal{V}} \rho(\mathbf{x})(x_k x_k \delta_{ij} - x_i x_j) dV = R_{ip} R_{jq} I_{pq}$$

Hence a rank two tensor.

$$I_{ji} = \int_{\mathcal{V}} \rho(\mathbf{x})(x_k x_k \delta_{ji} - x_j x_i) dV$$

Since δ_{ij} and $x_i x_j$ are symmetric in (i, j), I_{ij} is a symmetric tensor. $\qquad\square$

Example 11.8. *Consider an ellipsoid given in a particular Cartesian coordinate system by*

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} + \frac{x_3^2}{c^2} \leq 1$$

with uniform density ρ_0, so has mass $M = \frac{4\pi}{3} abc\rho_0$. Use spherical polar coordinates to compute the inertia tensor.

$$x_1 = ar\cos\phi\sin\theta, \quad x_2 = br\sin\phi\sin\theta, \quad x_3 = cr\cos\theta$$

with $0 \leq r \leq 1$. Note that if $i \neq j$ then by symmetry,

$$\int_V \rho_0 x_i x_j dV = 0$$

$$I_{11} = \rho_0 \int_V (x_2^2 + x_3^2) dV = \rho_0 abc \int_0^{2\pi} \int_0^\pi \int_0^1 r^2 (b^2 \sin^2\phi \sin^2\theta + c^2 \cos^2\theta) r^2 \sin\theta \, dr \, d\theta \, d\phi = \frac{M}{5}(b^2 + c^2)$$

and by symmetry

$$I_{22} = \frac{M}{5}(a^2 + c^2), \quad I_{33} = \frac{M}{5}(a^2 + b^2)$$

So in the coordinate system aligned with the axes of symmetry of the ellipsoid, $I = \mathrm{diag}(b^2 + c^2, a^2 + c^2, a^2 + b^2)\frac{M}{5}$. For a sphere $a = b = c$, $I_{ij} = \frac{2}{5} M \delta_{ij}$.

Proposition 11.8: If T_{ij} is a symmetric, rank-two tensor, then $\exists$ a choice of right-handed Cartesian coordinates axes in which $(T_{ij}) = \mathrm{diag}(\alpha, \beta, \gamma)$. The corresponding Cartesian axes are called the principal axes for T.

Proof. A real symmetric matrix (or more generally, normal matrix) can be diagonalized using orthogonal transformation R, for which we can assume $\det(R) = 1$ w.l.o.g. $\qquad\square$

The principal axes for the inertia tensor usually correspond to some obvious lines of symmetry on the body.

11.5   Invariant and isotropic tensors

Definition 11.9 (Isotropic tensors)**:** A tensor is isotropic if it is invariant under changes in Cartesian coordinates, i.e.

$$T'_{ij\ldots k} = R_{ip}R_{jq}\ldots R_{kr}T_{pq\ldots r} = T_{ij\ldots k}$$

for any rotation matrix R.

Proposition 11.9 (Classification of isotropic tensors)**:** Isotropic tensors on $\mathbb{R}^3$ are classified as

1. all tensors of rank 0 are isotropic;

2. there are no non-zero isotropic tensors of rank 1;

3. the most general rank two isotropic tensor is $\alpha\delta_{ij}$ where α is a scalar;

4. the most general rank two isotropic tensor is $\beta\epsilon_{ijk}$ where β is a scalar;

5. the most general rank four isotropic tensor is

$$\alpha\delta_{ij}\delta_{kl} + \beta\delta_{ik}\delta_{jl} + \gamma\delta_{il}\delta_{jk}$$

   with α, β, γ scalars;

6. the most general rank $n \geq 4$ isotropic tensor is a linear combination of products of Kronecker deltas and Levi-Civita tensors.

Proof. **Very sketchy proof!**

1. Rank 0 tensors (scalars) are isotropic by definition;

2. If v_i are the components of an isotropic vector then $v_i = R_{ij}v_j$ for any rotation. We will use rotations by π about each axis. Take about x_3-axis first

$$(R_{ij}) = \mathrm{diag}(-1, -1, 1)$$

   Then $v_1 = R_{1j}v_j = -v_1 \implies v_1 = 0$, $v_2 = R_{2j}v_j = -v_2 \implies v_2 = 0$. Similarly, using

$$(R_{ij}) = \mathrm{diag}(1, -1, -1)$$

   we find $v_3 = R_{3j}v_j = -v_3 \implies v_3 = 0$, i.e. $v_i = 0$ and this holds in all frames;

3. If T_{ij} is isotropic then $T_{ij} = R_{ip}R_{jq}T_{pq}$ for any R. We will use rotations by $\frac{\pi}{2}$ about each axis. First, we take about x_3-axis

$$(R_{ij}) = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

   Then

$$T_{13} = R_{1p}R_{3q}T_{pq} = R_{12}R_{33}T_{23} = T_{23}$$

$$T_{23} = R_{2p}R_{3q}T_{pq} = R_{21}R_{33}T_{13} = -T_{13}$$

   so $T_{13} = T_{23} = 0$. We also have

$$T_{11} = R_{1p}R_{1q}T_{pq} = R_{12}R_{12}T_{22} = T_{22}$$

   Next we take about x_1-axis

$$(R_{ij}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}$$

   Then

$$T_{32} = R_{3p}R_{2q}T_{pq} = R_{32}R_{23}T_{23} = -T_{23} = 0$$

$$T_{12} = R_{1p}R_{2q}T_{pq} = R_{11}R_{23}T_{13} = -T_{13} = 0$$

$$T_{31} = R_{3p}R_{1q}T_{pq} = R_{32}R_{11}T_{21} = -T_{21}$$

$$T_{21} = R_{2p}R_{1q}T_{pq} = R_{23}R_{11}T_{31} = T_{31}$$

so $T_{31} = T_{21} = 0$. We also have

$$T_{22} = R_{2p}R_{2q}T_{pq} = R_{23}R_{23}T_{33} = T_{33}$$

Combining the result from the two rotations, we conclude that all the off-diagonal elements of T are zero, i.e. $T_{ij} = 0$ if $i \neq j$ and all diagonal elements are equal, i.e. $T_{11} = T_{22} = T_{33}$. Hence $T_{ij} = \alpha \delta_{ij}$ for scalar α.

4. Similar, just more indices. Say take $T'_{ijk} = R_{ip}R_{jq}R_{kr}T_{pqr}$ for any R. Take rotation about x_1-axis like before, then

$$T_{111} = T_{222} = T_{333}$$

$$T_{112} = T_{223} = T_{331}, \quad T_{113} = T_{221} = T_{332}$$

$$T_{122} = T_{233} = T_{311}, \quad T_{133} = T_{211} = T_{322}$$

$$T_{121} = T_{232} = T_{313}, \quad T_{131} = T_{212} = T_{322}$$

$$T_{123} = T_{231} = T_{312}, \quad T_{132} = T_{213} = T_{321}$$

Then, rotate about x_3-axis like before,

$$T_{111} = T_{222} = -T_{111}$$

$$T_{112} = -T_{221} = -T_{112}$$

$$T_{211} = -T_{122} = -T_{211}$$

$$T_{121} = -T_{212} = -T_{121}$$

$$T_{123} = -T_{213}$$

We conclude the diagonal elements must be zero and

$$T_{112} = T_{223} = T_{331} = T_{113} = T_{221} = T_{332} = 0$$

$$T_{122} = T_{233} = T_{311} = T_{133} = T_{211} = T_{322} = 0$$

$$T_{121} = T_{232} = T_{313} = T_{131} = T_{212} = T_{323} = 0$$

$$T_{123} = T_{231} = T_{312} = -T_{132} = -T_{213} = -T_{321}$$

Hence, an isotropic third order tensor must be proportional to ϵ_{ijk}.

5. Not going to prove the remaining.

$\square$

Example 11.9. *Consider the array*

$$T_{ij} = \int_{\mathbb{R}^3} e^{-r^5} x_i x_j dV$$

T_{ij} is isotropic, so by Proposition 11.9, $T_{ij} = \alpha \delta_{ij}$ for some constant α. Contracting on i and j

$$3\alpha = \int_{\mathbb{R}^3} e^{-r^5} r^2 dV = 4\pi \int_0^\infty r^4 e^{-r^5} dr = \frac{4\pi}{5}$$

Hence $T_{ij} = \frac{4\pi}{15}\delta_{ij}$.

Example 11.10. *The inertia tensor is isotropic, so $I_{ij} = \alpha \delta_{ij}$ for some constant α. Contracting on i and j*

$$3\alpha = \int_{|\mathbf{x}|<R} 2\rho_0 r^2 dV = 2\rho_0 4\pi \int_0^R r^4 dr = \frac{6}{5}MR^2$$

where $M = \frac{4\pi}{3}R^3 \rho_0$. Hence, $I_{ij} = \frac{2}{5}MR^2\delta_{ij}$.

Example 11.11 (Strain tensor again). *We have seen that the linear strain tensor e_{ij} measures the infinitesimal deformation of a body in which each point $\mathbf{x}$ undergoes a displacement of $\mathbf{u}(\mathbf{x})$. Experiment suggests that the stresses experienced by a body depends linearly on the strain at each point.*

$$\sigma_{ij} = c_{ijkl}e_{kl}$$

where c_{ijkl} is a collection of $3^4 = 81$ numbers. We cannot immediately apply the quotient theorem since e_{kl} is not arbitrary but necessarily symmetric. However, if we assume $c_{ijkl} = c_{ijlk}$ then we can apply the quotient theorem. On this assumption, the array c_{ijkl} must correspond to the components of a rank 4 tensor, called the stiffness tensor. If the material is isotropic, then by Proposition 11.9,

$$c_{ijkl} = \lambda\delta_{ij}\delta_{kl} + \beta\delta_{ik}\delta_{jl} + \gamma\delta_{il}\delta_{kj}$$

for some scalars λ, β, γ so that

$$\sigma_{ij} = \lambda\delta_{ij}e_{kk} + \beta e_{ij} + \gamma e_{ji} = \lambda\delta_{ij}e_{kk} + 2\mu e_{ij}$$

where $e_{ij} = e_{ji}$ and $2\mu := \beta + \gamma$. This is Hooke's law for an isotropic material. Contract on i and j to get

$$\sigma_{ii} = (3\lambda + 2\mu)e_{kk}$$

assuming $3\lambda \neq -2\mu$, then $2\mu e_{ij} = \sigma_{ij} - \frac{\lambda}{3\lambda+2\mu}\sigma_{kk}\delta_{ij}$.

12   Contour integration

Integration along a path; elementary properties. Cauchy's theorem; proof by Cauchy-Riemann equations and divergence theorem in 2-D. Integral of $f'(z)$; Cauchy's formula for $f(z)$. Calculus of residues; examples of contour integration; point at infinity; multi-valued functions, branch points, $\log(z)$ **[4]**

12.1   Multi-valued functions

For $z = re^{i\theta}$, define $\ln z = \ln r + i\theta$. There are thus infinitely many values of $\ln z$, for θ may take an infinite number of values, e.g. $\ln i = \frac{\pi i}{2}, \frac{5\pi i}{2}, -\frac{3\pi}{2}, \ldots$, depending on which choice of θ we make.

12.1.1   Branch point

Definition 12.1 (Branch point)**:** A branch point of a multi-valued function is a point which is impossible to encircle with an arbitrarily small curve upon which the function is both continuous and single-valued.

Example 12.1.
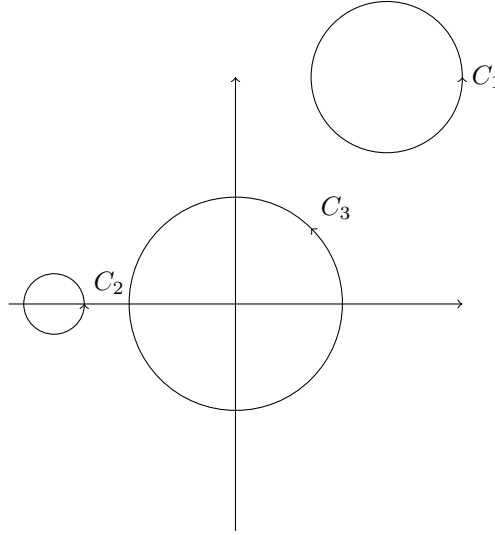
1. *Consider three closed curves (does not necessarily need to be circles) for $\ln z$:*

    - *$\mathcal{C}_1$: in region s.t. $\theta \in (0, \frac{\pi}{2})$, $\ln z$ will always be continuous and single-valued going around $\mathcal{C}_1$;*
    - *$\mathcal{C}_2$: in region s.t. $\theta \in (\frac{\pi}{2}, \frac{3\pi}{2})$, again $\ln z$ is continuous and single-valued;*
    - *$\mathcal{C}_3$: in such a way that encircles the origin, $\ln z$ can never be continuous and single-valued around $\mathcal{C}_3$. It must either 'jump' somewhere or be multi-valued.*

2. *$\ln(z-a)$ has a branch point at $z = a$, while $\ln \frac{z-1}{z+1}$ has two branch points at $z = \pm 1$.*

3. *$\ln z$ also has a branch point at ∞ because if $\xi = \frac{1}{z}$, $\ln z = -\ln \xi$ which has a branch point at $\xi = 0$.*

4. *$z^\alpha = r^\alpha e^{i\alpha\theta}$ has a branch point at 0 for $\alpha \notin \mathbb{Z}$. Consider a circle of radius r_0 centred at 0 and suppose w.l.o.g. we start at $\theta = 0$ and go round anti-clockwise. θ must vary continuously to ensure the continuity of $e^{i\alpha\theta}$, so as we get back almost to where we started, θ will approach 2π. But then there will be a jump in θ back to zero (in order to satisfy the single-valued requirement), and hence a jump in z^α from $r_0^\alpha e^{2\pi i\alpha}$ to r_0^α. We cannot, therefore, make z^α both continuous and single-valued on the circle. Note if $\alpha \in \mathbb{Z}$, then there is no jump, since $e^{2\pi i\alpha} = 1$.*

5. *Similarly, z^α has a branch point at ∞ for $\alpha \notin \mathbb{Z}$.*

6. *$\ln \frac{z-1}{z+1}$ does not have a branch point at ∞, because if $\xi = \frac{1}{z}$, then $\ln \frac{z-1}{z+1} = \ln \frac{1-\xi}{1+\xi}$. For ξ close to 0, $\frac{1-\xi}{1+\xi}$ remains close to 1, and therefore well away from the branch point of $\ln z$ at 0. So we can encircle $\xi = 0$ without $\ln \frac{1-\xi}{1+\xi}$ being discontinuous.*
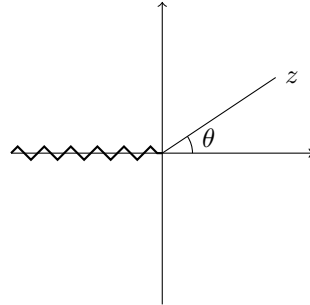
12.1.2 Branch cut

Definition 12.2 (Branch cut)**:** A branch cut is a curve (with ends possibly open, closed, or half-open) in the complex plane across which an analytic multivalued function is discontinuous.

Whenever a problem requires the use of a branch, it is important to specify it clearly. This can be done in two ways:

1. Define the function and parameter range exactly, e.g. $\ln z = \ln |z| + i \arg(z)$, where $\arg(z) \in (-\pi, \pi]$.

2. Specify the location of the branch cut and give the value of the required branch at a single point not on the cut. The values everywhere else (except on the cut) are then defined uniquely by continuity. e.g. $\ln z$ with a branch cut along $\mathbb{R}^-$ and $\ln(1) = 0$.

Remarks 12.1. *Note that a branch cut alone does not specify a branch (can have different branches for the same branch cut), nor is a single value of function sufficient by itself.*

Example 12.2. *If we wish to ensure $\ln z$ is continuous and single-valued on any curve $\mathcal{C}$, we must stop $\mathcal{C}$ from encircling zero, which is the branch point. There are various ways of doing this:*

- *Introduce a branch cut from $-\infty$ on the real axis to 0. Fix $\theta \in (-\pi, \pi]$ only s.t. we have defined a branch of $\ln z$ which is continuous and single-valued on any curve $\mathcal{C}$ that doesn't cross the branch cut. Here, we can choose any branch for the same cut. Take, $\ln(1) = 2\pi i$. This branch is analytic everywhere except on the negative real axis, where it is not even continuous, and at the branch point itself.*

  This is actually the canonical (i.e. standard) branch cut for $\ln z$, and the branch is called the principal value of the logarithm.

  It is usual to include $+\pi$ in the range of θ and not $-\pi$; but in fact this is of no consequence since we shall never evaluate the function precisely on the branch cut.

  What are the values of $\ln z$ just above and below the branch cut? Consider a point on the negative axis $z = x < 0$. Just above the branch cut, at $z = x + i0^+$, $\theta = \pi$ (in the limit) so $\ln z = \ln |x| + i\pi$. Just below it, at $z = x + i0^-$, $\ln z = \ln |x| - i\pi$.

- *Place branch cut along the negative imaginary axis and choose $\theta \in (-\frac{\pi}{2}, \frac{3\pi}{2}]$ (again, measured from the horizontal). Here, $\ln(1) = 0$.*

- *Or we could just draw any arbitrary curve (not necessarily a straight line) s.t. $\log(1) = 0$ for instance, as long as the branch point is not encircled. However, such a nasty choice makes it difficult to write down the range of θ.*

Exactly the same considerations (and possible branch cuts) apply for $z^\alpha = r^\alpha e^{i\alpha\theta}$, $\alpha \notin \mathbb{Z}$. One way of seeing this is to note $z^\alpha = e^{\alpha \ln z}$.

12.1.3 Multiple branch points

When there is more than 1 branch points, we may need more than 1 branch cuts.

Example 12.3. *$f(z) = (z(z-1))^{1/3}$ has branch points at 0 and 1, so we need two branch cuts. One possibility is to have the two branches - one along the positive real axis from 1 to $+\infty$, and another on the negative real axis from 0 to $-\infty$. For any z, write $z = re^{i\theta}$, $z - 1 = r_1 e^{i\theta_1}$ with $\theta \in (-\pi, \pi]$, $\theta_1 \in [0, 2\pi)$ and define $f(z) = (rr_1)^{1/3} e^{i(\theta+\theta_1)/3}$. This is continuous as long as we don't cross either branch cut.*
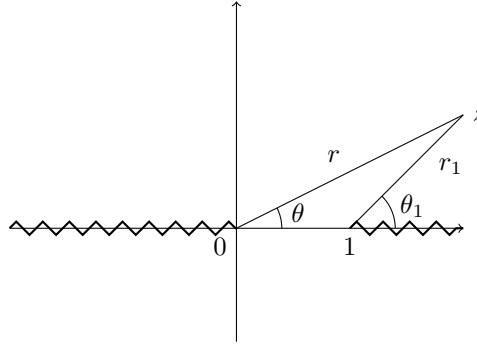
In fact, we might need fewer branch cuts (less than the number of branch points).

Example 12.4. $f(z) = \ln \frac{z-1}{z+1}$ *has branch points at ± 1. Let $z_1 - 1 = r_1 e^{i\theta_1}$ and $z_2 + 1 = r_2 e^{i\theta_2}$, then with our choice of branch cut (from 1 to ∞ and from -1 to $-\infty$), we choose $0 \leq \theta_1 < 2\pi$ and $-\pi < \theta_2 \leq \pi$. Then,*

$$f(z) = \ln \frac{r_1 e^{i\theta_1}}{r_2 e^{i\theta_2}} = \ln(r_1/r_2) + i(\theta_1 - \theta_2)$$

where we get a single-valued definition of $f(z)$. It is now impossible to wind around either of the two branch points. $f(z)$ is now analytic everywhere except along the branch cuts.

In that branch cut, we have included the point of ∞, which is not necessary since ∞ is not a branch point, and thus may allow an infinitely large circle for z. An alternative branch cut would be from -1 to $+1$, and choose both $\theta_1, \theta_2 \in [0, 2\pi)$. If z were to wind around just one branch point, then it would cross the branch cut.
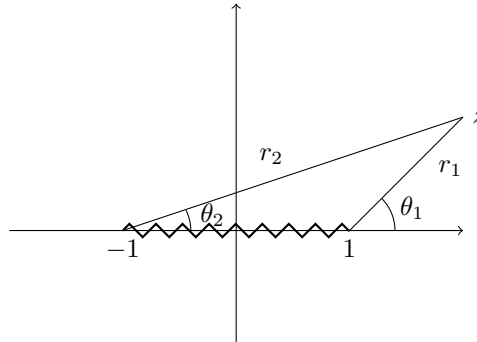
Let's say from above to below, then θ_1 is unchanged at π, while θ_2 jumps from 0 to 2π, causing a discontinuity in $f(z)$. Nevertheless, we may allow z to wind around both branch cuts together, i.e. for a sufficiently large curve that encircle both branch points exactly once. In that case, θ_1, θ_2 both suddenly jump from 2π back to 0 and so $\theta_1 - \theta_2$ does not jump.

Example 12.5. *Similarly, consider $g(z) = (z^2 - 1)^{1/2}$ with the same branch cut. The single-valued choice of $g(z)$ is $\sqrt{r_1 r_2} e^{i(\theta_1 + \theta_2)/2}$. When both θ_1, θ_2 jump by 2π, $\frac{1}{2}(\theta_1 + \theta_2)$ jumps by 2π also, i.e. $e^{2\pi i} = 1$, the branch cut prevents either θ_1 or θ_2 jump on its own.*

12.2   Contours and integrals

Definition 12.3 (Curve)**:** A curve $\gamma(t)$ is a continuous map $\gamma : [0, 1] \to \mathbb{C}$. A closed curve has $\gamma(0) = \gamma(1)$. A simple curve is one that doesn't intersect itself (except at $t = 0, 1$ if the curve is simple and closed).

Definition 12.4 (Contour)**:** A contour is a piecewise smooth curve.

Remarks 12.2.

1. *Given contours γ_1, γ_2 with matching endpoints, i.e. $\gamma_1(1) = \gamma_2(0)$, then $\gamma_1 + \gamma_2$ is the two contours joined end-to-end.*

2. *The contour $-\gamma$ is γ traversed in the opposite direction.*

Definition 12.5 (Contour integral)**:** The contour integral is defined to be

$$\int_\gamma f(z)dz := \int_0^1 f(\gamma(t))\gamma'(t)dt$$

Alternatively, and equivalently, for a simple contour, we dissect it at points $z_0, z_1, \ldots, z_N$ on the contour in that order, where $z_0 = \gamma(0)$ and $z_N = \gamma(1)$, and let $\delta_{z_n} = z_{n+1} - z_n$, $n = 0, \ldots, N - 1$. Then

$$\int_\gamma f(z)dz = \lim_{\Delta \to 0} \sum_{n=0}^{N-1} f(z_n)\delta_{z_n}$$

where $\Delta = \max_{n \in [0, \ldots, N-1]} |\delta_{z_n}|$ and as $\Delta \to 0$, $N \to \infty$.

Example 12.6. *The result of a contour integral between two points in $\mathbb{C}$ may depend on the choice of the contour. Consider $I_1 = \int_{\gamma_1} \frac{dz}{z}$ and $I_2 = \int_{\gamma_2} \frac{dz}{z}$, where in both cases we integrate from $z = -1$ to $+1$ around a unit semi-circle: γ_1 above and γ_2 below the real axis, and traversed clockwise and anti-clockwise respectively.*
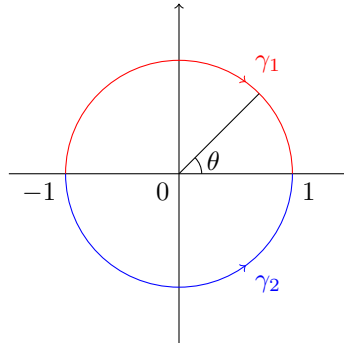
Substitute $z = e^{i\theta}$, $dz = e^{i\theta}id\theta$, then

$$I_1 = \int_\pi^0 \frac{ie^{i\theta}d\theta}{e^{i\theta}} = -i\pi, \quad I_2 = \int_{-\pi}^0 \frac{ie^{i\theta}d\theta}{e^{i\theta}}d\theta = +i\pi$$

Proposition 12.1 (Properties of contour integral)**:**

1. $-\int_\gamma f(z)dz = \int_{-\gamma} f(z)dz$;

2. $\int_{\gamma_1+\gamma_2} f(z)dz = \int_{\gamma_1} f(z)dz + \int_{\gamma_2} f(z)dz$;

3. If γ is a contour from a to b in $\mathbb{C}$, then $\int_\gamma f'(z)dz = f(b) - f(a)$ so long as f is differentiable at every point on γ. Note, we must not cross a branch cut of f;

4. Integration by substitution and by parts work exactly as for integrals on $\mathbb{R}$;

5. If γ has length L and $|f(z)| \leq M$ on γ_1, then

$$\left| \int_\gamma f(z)dz \right| \leq \int_\gamma |f(z)||dz| \leq M \int_\gamma |dz| = ML$$

This is also known as the ML-inequality, or estimation lemma.

Notation 12.1 (Integration on closed contours)**.** *The notation $\oint f(z)dz$ denotes an integral round a closed contour. Doesn't matter where we start. The usual direction for positive sense is anti-clockwise.*

Theorem 12.1 (Cauchy's theorem)**:** If $f(z)$ is analytic in a simply-connected domain D, then for every simple closed contour γ in D,

$$\oint_\gamma f(z)dz = 0$$

Proof. Let $f = u + iv$, then evaluate $\oint_\gamma f(z)dz$:

$$\oint_\gamma (u+iv)(dx+idy) = \oint_\gamma (udx-vdy)+i\oint_\gamma (vdx+udy) = \int_S \left(-\frac{\partial v}{\partial x}-\frac{\partial u}{\partial y}\right)dxdy+i\int_S \left(\frac{\partial u}{\partial x}-\frac{\partial v}{\partial y}\right)dxdy$$

where S is a region enclosed by γ, by applying Green's theorem in a plane

$$\oint_{\partial S} (Pdx + Qdy) = \int_S \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)dxdy$$

but both brackets $\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}$ and $-\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}$ vanish by CRE, since f is differentiable throughout S. The result follows.   □

Remarks 12.3. *We also require u and v to have continuous partial derivatives in S in order for Green's theorem to be true. There is a completely different proof available if we do not wish to make assumptions on u and v.*

Theorem 12.2 (Deformation theorem)**:** Suppose γ_1, γ_2 are two contours from a to b, and that f is analytic on and between them. Then

$$\int_{\gamma_1} f(z)dz = \int_{\gamma_2} f(z)dz$$

Proof. Suppose first that γ_1, γ_2 do not cross. Then $\gamma_1 - \gamma_2$ is a simple closed contour, so $\oint_{\gamma_1-\gamma_2} f(z)dz = 0$ by Cauchy's theorem and the result follows.

If γ_1 and γ_2 do cross, then dissect them at each crossing point, and apply the technique above to each section. Done.   □

Remarks 12.4.

1. *Essentially, if f has no singularities, $\int_a^b f(z)dz$ does not depend on the choice of contour.*

2. *Another way of thinking about path-independence and indeed Cauchy's theorem itself, is to consider $\int f(z)dz$ as a path integral in $\mathbb{R}^2$. Then*

$$f(z)dz = (u + iv)(dx + idy) = (u + iv)dx + (-v + iu)dy$$
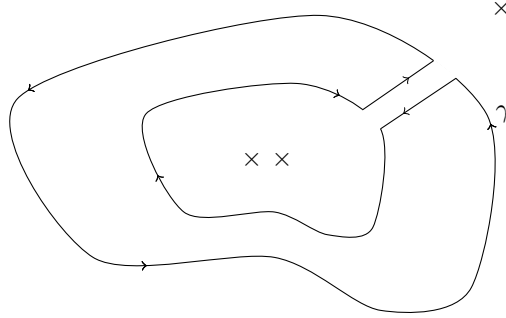
   is an exact differential. Hence, by CRE,

$$\frac{\partial}{\partial y}(u + iv) = \frac{\partial}{\partial x}(-v + iu)$$

3. *The deformation theorem also applies to closed contours. Let $\times$ denote a singularity in $\mathbb{C}$.*
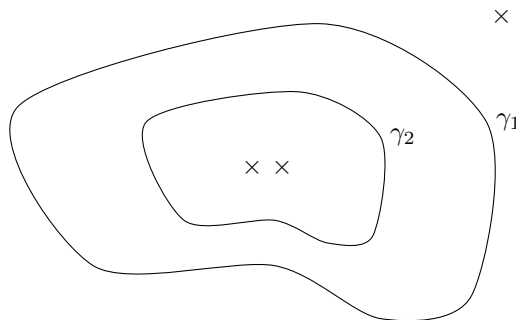
Suppose γ_1 is a closed contour that can be continuously deformed into another, γ_2, inside it, and there are no singularities in the region between them. Essentially, consider the contour γ shown:

Then by Cauchy's theorem,

$$\oint_\gamma f(z)dz = 0$$

Now let the distance between the two 'cross-cuts' tend to 0, then those contributions cancel out and, in this limit

$$\oint_{\gamma_1} f(z)dz - \oint_{\gamma_2} f(z)dz = 0$$

12.3   Residues

Definition 12.6 (Residue)**:** The residue of f at z_0 is the coefficient of a_{-1} of its Laurent series.

Notation 12.2. *We denote the residue as $\mathrm{res}_{z=z_0} f(z)$.*

Proposition 12.2:

1. At a simple pole, the residue is $\mathrm{res}_{z=z_0} f(z) = \lim_{z \to z_0}((z - z_0)f(z))$.

2. More generally, at a pole of order N, the residue is

$$\mathrm{res}_{z=z_0} f(z) = \lim_{z \to z_0}\left[\frac{1}{(N-1)!}\frac{d^{N-1}}{dz^{N-1}}((z - z_0)^N f(z))\right]$$

Proof. At a simple pole, $a_n = 0 \ \forall n < -1$, so

$$\lim_{z \to z_0}\left[(z - z_0)\left(\frac{a_{-1}}{z - z_0} + a_0 + a_1(z - z_0) + \dots\right)\right] = a_{-1}$$

For the more general case, we can prove it in a similar manner. $\square$

In practice, a variety of techniques can be used to evaluate residues, i.e. no single technique is optimal for all situations. Sometimes, L'Hopital rule is helpful.

Example 12.7 (Calculating residues)**.**

1. $\frac{e^z}{z^3}$ *at $z = 0$: Laurent series about $z = 0$ is*

$$z^{-3} + z^{-2} + \frac{1}{2}z^{-1} + \frac{1}{3!} + \dots$$

so residue is 0.5*. Alternatively, f has a pole of order 3 at $z = 0$, so*

$$\mathrm{res}_{z=0} f(z) = \lim_{z \to 0}\left\{\frac{1}{2!}\frac{d^2}{dz^2}(z^3 f(z))\right\} = \frac{1}{2}\lim_{z \to 0}\frac{d^2}{dz^2}e^z = \frac{1}{2}$$

2. $\frac{e^z}{z^2-1}$ *at* $z = 1$*: Laurent series about* $z = 1$ *is*

$$\frac{e}{2}\left(\frac{1}{z-1} + \frac{1}{2} + \dots\right)$$

so residue is $\frac{1}{2}e$*. Alternatively,*

$$\text{res}_{z=1}\, g(z) = \lim_{z\to 1} \frac{(z-1)e^z}{z^2-1} = \lim_{z\to 1} \frac{e^2}{z+1} = \frac{1}{2}e$$

3. $\frac{1}{z^8-w^8}$ *at* $z = w$*: we have 8 simple poles* $z = we^{n\pi i/4}$ *with* $n = 0, 1, \dots, 7$*. There are two ways to evaluate* $\text{res}_{z=w}\,\frac{1}{z^8-w^8} = \lim_{z\to w}\frac{z-w}{z^8-w^8}$*: either brute force*

$$\frac{1}{(w^2-(we^{i\pi/4})^2)(2w^2)(w^2-(we^{i3\pi/4})^2)(2w)} = \frac{1}{(1-i)2(1+i)2w^7} = \frac{1}{8w^7}$$

or just L'Hopital rule:
$$\lim_{z\to w}\frac{z-w}{z^8-w^8} = \lim_{z\to w}\frac{1}{8z^7} = \frac{1}{8w^7}$$

4. $\text{sech}(\pi z)$ *at* $z = (n+0.5)i$*: simple poles* $\forall n \in \mathbb{Z}$ *(since they are simple zeros for* $\cosh(\pi z)$*. Use L'Hopital rule to evaluate* $\text{res}_{z=(n+0.5)i}\,\text{sech}(\pi z)$

$$\lim_{z\to(n+0.5)i}\frac{z-(n+0.5)i}{\cosh(\pi z)} = \lim_{z\to(n+0.5)i}\frac{1}{\pi\sinh(\pi z)} = \frac{1}{\pi i\sinh(n+0.5)\pi} = \frac{-i(-1)^n}{\pi}$$

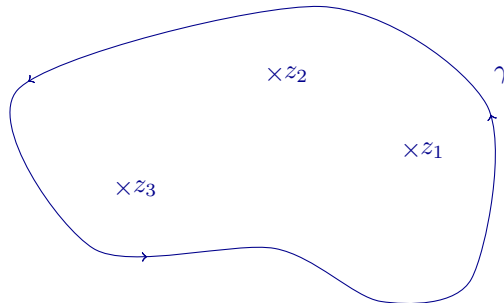5. $\frac{1}{\sinh^3 z}$ *at* $z = \pi i$*: zero of order 3. The Laurent series about* $z = \pi i$ *is*

$$\sinh^3 z = -(z-\pi i)^3 - \frac{1}{2}(z-\pi i)^5 - \dots$$

$$\implies \frac{1}{\sinh^3 z} = -(z-\pi i)^{-3}(1 + 0.5(z-\pi i)^2 + \dots)^{-1} = -(z-\pi i)^{-3} + 0.5(z-\pi i)^{-1} + \dots$$

Hence the residue is $\text{res}_{z=\pi i}\,\frac{1}{\sinh^3 z} = 0.5$*.*

Theorem 12.3 (Residue theorem)**:** Suppose f is analytic in a simply-connected domain except at a finite number of isolated singularities $\{z_1, \dots, z_n\}$. Suppose a simple closed contour γ encircles the origin anticlockwise, then

$$\oint_\gamma f(z)dz = 2\pi i \sum_{k=1}^{n} \text{res}_{z=z_k}\, f(z)$$

Proof. Consider $\hat{\gamma}$ shown, consisting of small clockwise circles $\gamma_1, \dots, \gamma_n$ around each singularity, cross-cuts which cancel in the limit as they approach each other in pairs, and the large outer curve (which is γ in this limit).
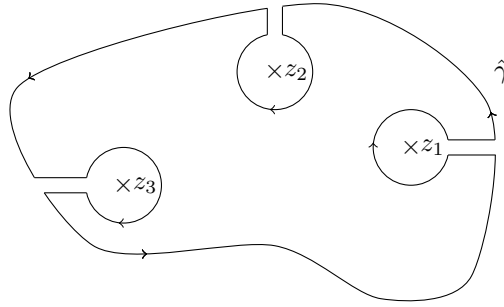
Since $\hat{\gamma}$ encircles no singularities, then by Cauchy's theorem $\oint_{\hat{\gamma}} f(z)dz = 0$. Hence, in the limit when the cross-cuts cancel, we have

$$\oint_{\gamma} f(z)dz + \sum_{k=1}^{n} \oint_{\gamma_k} f(z)dz = \oint_{\hat{\gamma}} f(z)dz = 0$$

But about each singularity z_k, there is a Laurent series valid locally in some annulus, so we have

$$\oint_{\gamma_k} f(z)dz = -2\pi i \operatorname{res}_{z=z_k} f(z)$$

where we used minus sign since γ_k is a clockwise contour encircling the singularity. The result follows.   □

Theorem 12.4 (Cauchy's integral formula)**:** Suppose f is analytic in a simply-connected domain D and $z \in D$, then for any simple closed contour γ in D encircling z anti-clockwise,

$$f(z) = \frac{1}{2\pi i} \oint_{\gamma} \frac{f(w)}{w - z} dw$$

Proof. The only singularity of the integrand is at $w = z$, and is a simple pole (since f is analytic there). Hence, using the residue theorem and the formula in Section 4.1, the integral equals

$$2\pi i \operatorname{res}_{w=z} \frac{f(w)}{w - z} = 2\pi i \lim_{w \to z} f(w) = 2\pi i f(z)$$

Result follows.   □

Remarks 12.5.

1. *If z lies outside γ, RHS is 0 by Cauchy's theorem.*

2. *If we know f on γ, then we know it at all points within γ. Another way of looking at this is to write $f = u + iv$, with u, v harmonic, and are both specified on γ. Then we have Laplace's equation for u and v with Dirichlet b.c.s, which has a unique solution inside γ. Cauchy's formula thus gives us a way of calculating that solution explicitly.*

3. *Normally, Cauchy's integral formula is proved before the residue theorem, because it's needed to prove the existence of Laurent series.*

Corollary 12.1:

$$f^{(n)}(z) = \frac{n!}{2\pi i} \oint_{\gamma} \frac{f(w)}{(w - z)^{n+1}} dw$$

Proof. Differentiating Cauchy's formula n times give the result. Differentiation under the integral sign is valid because the integrand, both before and after, is a continuous function of w and z. Hence, at any point where f is analytic, all its derivatives exist, so it is differentiable infinitely many times.   □

Corollary 12.2 (Liouville's theorem)**:** Any bounded entire function is constant.

Proof. Suppose $|f(z)| \leq M \; \forall z$ and consider a circle of arbitrary radius R, centred at a given $z \in \mathbb{C}$, then

$$f'(z) = \frac{1}{2\pi i} \oint_{|w-z|=R} \frac{f(w)dw}{(w-z)^2} \implies |f'(z)| \leq \frac{1}{2\pi} 2\pi R \frac{M}{R^2} \implies \lim_{R \to \infty} |f'(z)| = 0$$

where we used ML inequality. Hence, $f'(z) = 0 \; \forall z \in \mathbb{C}$, so f is constant. $\qquad\square$

Corollary 12.3 (Maximum moduli principle)**:** If $f(z)$ is analytic within a bounded domain and on its boundary, then $|f(z)|$ attains its maximum on the boundary.

Proof. Let $0 < \rho < r$. Applying the Cauchy integral formula for $w = z + \rho e^{2\pi i \theta}$, we get

$$\begin{aligned}
|f(z)| &= \left| \frac{1}{2\pi i} \int_{|w-z|=\rho} \frac{f(w)}{w-z} \, dw \right| \\
&= \left| \int_0^1 f(z + \rho e^{2\pi i \theta}) \, d\theta \right| \\
&\leq f(z).
\end{aligned}$$

So we must have equality throughout. So $|f(w)|$ is constant on the circle $|z - w| = \rho$, and is equal to $f(z)$. Since this is true for all $\rho \in (0, r)$, it follows that $|f|$ is constant on this circular domain. Then the CRE entail that f must be constant. $\qquad\square$

12.4   Applications of the residue theorem

To illustrate the technique known as the 'calculus of residue', consider the following example:

Example 12.8. *Evaluate*

$$I = \int_0^\infty \frac{dx}{1 + x^2}$$

*Consider $\oint_\gamma \frac{dz}{1+z^2}$, where γ is the contour consisting of γ_0 followed by γ_R: from $-R$ to R along the real axis, then returning to $-R$ via a semicircle of radius R in the upper half plane. This is known as **closing in the upper half-plane** or closing above.*
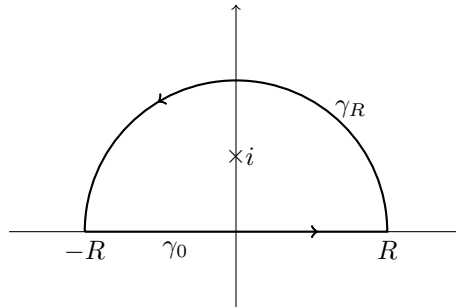
Now $\frac{1}{1+z^2} = \frac{1}{(z+i)(z-i)}$, so the only singularity enclosed by γ is a simple pole at $z = i$, whose residue is $\lim_{z \to i} \frac{z-i}{z^2+1} = \frac{1}{2i}$. Hence,

$$\oint_{\gamma_0} \frac{dz}{1+z^2} + \oint_{\gamma_R} \frac{dz}{1+z^2} = \oint_\gamma \frac{dz}{1+z^2} = 2\pi i \operatorname*{res}_{z=i} \frac{1}{1+z^2} = \pi$$

But

$$\oint_{\gamma_0} \frac{dz}{1+z^2} = \int_{-R}^R \frac{dx}{1+x^2} \to 2I, \; as \; R \to \infty$$

Also, we can justify $\int_{\gamma_R} \frac{dz}{1+z^2} \to 0$ as $R \to \infty$. Doing this formally:

$$|1 + z^2| \geq |1 - |z|^2| = |1 - R^2| = R^2 - 1$$

by triangle inequality and where R is large. So, $|\frac{1}{1+z^2}| \leq \frac{1}{R^2-1}$ and hence by the ML inequality,

$$\left| \int_{\gamma_R} \frac{dz}{1+z^2} \right| \leq \frac{\pi R}{R^2 - 1} \to 0, \; as \; R \to \infty$$

Or informally: for $z \in \gamma_R$, we use the ML inequality:

$$\left| \frac{1}{1+z^2} \right| = O(R^{-2}) \implies \left| \int_{\gamma_R} \frac{dz}{1+z^2} \right| \leq \pi R \, O(R^{-2}) = O(R^{-1}) \to 0, \ \text{as } R \to \infty$$

as we will see later, the informal argument is easier for more difficult integrals.

Example 12.9.

1. *Evaluate*

$$I = \int_0^\infty \frac{dx}{(x^2+a^2)^2}$$

where $a > 0$ is a real constant. Consider $\oint_\gamma \frac{dz}{(z^2+a^2)^2}$, with γ real axis closed in the upper half-plane (as before). The only singularity within γ is a pole of order 2 at $z = ia$, at which residue is

$$\operatorname{res}_{z=ia} \frac{1}{(z^2+a^2)^2} = \lim_{z \to ia} \frac{d}{dz} \frac{(z-ia)^2}{(z+ia)^2(z-ia)^2} = \lim_{z \to ia} \frac{-2}{(z+ia)^3} = \frac{-2}{-8ia^3} = -\frac{1}{4}ia^{-3}$$

The integral round γ_R will vanish as $R \to \infty$, since by the ML inequality,

$$\left| \int_{\gamma_R} \frac{dz}{(z^2+a^2)^2} \right| \leq \pi R \, O(R^{-4}) = O(R^{-3})$$

Hence, $2I = 2\pi i(-0.25 ia^{-3}) \implies I = \frac{\pi}{4a^3}$.

2. *For $I = \int_0^\infty \frac{dx}{1+x^4}$, use the same contour γ again. There are simple poles at $e^{\pm i\pi/4}$, $e^{\pm 3\pi i/4}$, but only the first two are enclosed, with residues $-\frac{1}{4}e^{i\pi/4}$ and $+\frac{1}{4}e^{-i\pi/4}$. The integral round the semicircle is bounded by $\pi R \, O(R^{-4}) \to 0$ as $R \to \infty$. Hence,*

$$2I = 2\pi i(-0.25 e^{\pi i/4} + 0.25 e^{-\pi i/4}) \implies I = \frac{\pi}{2\sqrt{2}}$$

Alternatively could use a quarter circle contour to enclose one pole only, say $e^{i\pi/4}$.

For trigonometric integrals of the form

$$\int_0^{2\pi} f(\sin\theta, \cos\theta) d\theta$$

substitute $z = e^{i\theta} \implies dz = iz d\theta$, and $\cos\theta = \frac{1}{2}(z+z^{-1})$ and $\sin\theta = \frac{1}{2i}(z-z^{-1})$ to obtain a closed contour integral.
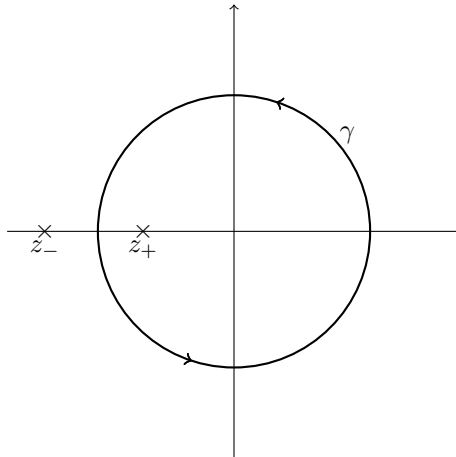
Example 12.10. *Evaluate*

$$I = \int_0^{2\pi} \frac{d\theta}{a+\cos\theta}, \quad a > 1$$

so that the integrand is always finite. Substitute $z = e^{i\theta}$,

$$I = \oint_\gamma \frac{(iz)^{-1} dz}{a + 0.5(z+z^{-1})} = -2i \oint_\gamma \frac{dz}{z^2 + 2az + 1}$$

where the integrand has poles $z_\pm = -a \pm \sqrt{a^2-1}$, both on the real axis. Since $a-1 < \sqrt{a^2-1} < a$, then $-1 < z_+ < 0$, i.e. z_+ is inside the unit circle.

The residue at $z = z_+$ is

$$\operatorname{res}_{z=z_+} \frac{1}{z^2 + 2az + 1} = \lim_{z \to z_+} \frac{z - z_+}{z^2 + 2az + 1} = \frac{1}{z_+ - z_-} = \frac{1}{2\sqrt{a^2 - 1}}$$

Hence, $I = -2i\frac{2\pi i}{2\sqrt{a^2-1}} = \frac{2\pi}{\sqrt{a^2-1}}$.

Example 12.11 (Integration around a branch cut). *Evaluate*

$$I = \int_0^\infty \frac{x^\alpha}{1 + \sqrt{2}x + x^2}dx, \quad -1 < \alpha < 1$$

so that the integral converges. Need branch cut for z^α. Take this along the positive real axis, $z^\alpha := r^\alpha e^{i\alpha\theta}$, where $0 \le \theta < 2\pi$. Consider

$$\oint_\gamma \frac{z^\alpha}{1 + \sqrt{2}z + z^2}dz$$

where γ is a keyhole contour, consisting of a large circle γ_R of radius R, a small circle γ_ε of radius ε (to avoid singularity of z^α at $z = 0$) and two lines just above and below the branch cut.
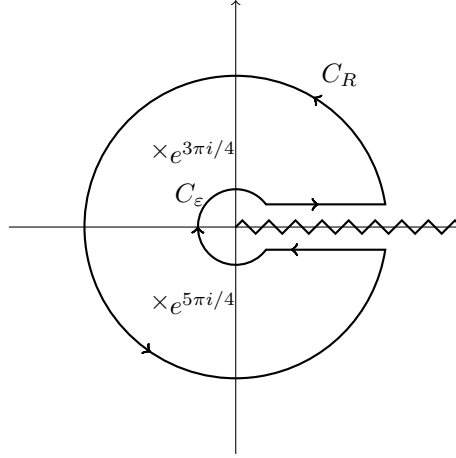
The contributions from γ_R and γ_ε respectively are

$$2\pi R \; O(R^{\alpha-2}) = O(R^{\alpha-1}) \to 0, \; \text{as } R \to \infty$$

$$2\pi\varepsilon \; O(\varepsilon^\alpha) = O(\varepsilon^{\alpha+1}) \to 0, \; \text{as } \varepsilon \to 0$$

where $\alpha < 1$ and $\alpha > -1$ respectively. Above and below the branch cut respectively:

$$\int_\varepsilon^R \frac{x^\alpha}{1 + \sqrt{2}x + x^2}dx \to I, \; \text{as } \varepsilon \to 0, R \to \infty$$

$$\int_R^\varepsilon \frac{x^\alpha e^{2\pi i\alpha}}{1 + \sqrt{2}x + x^2}dx \to -e^{2\pi i\alpha}I, \; \text{as } \varepsilon \to 0, R \to \infty$$

Thus,

$$\oint_\gamma \frac{z^\alpha}{1 + \sqrt{2}z + z^2}dz \to (1 - e^{2\pi i\alpha})I, \; \text{as } \varepsilon \to 0, R \to \infty$$

The integrand is equal to $\frac{z^\alpha}{(z - e^{3\pi i/4})(z - e^{i5\pi/4})}$. The poles inside γ have residues

$$\operatorname{res}_{z=e^{3\pi i/4}} \frac{z^\alpha}{(z - e^{3\pi i/4})(z - e^{i5\pi/4})} = \frac{e^{3\pi i\alpha/4}}{\sqrt{2}i}$$

$$\operatorname{res}_{z=e^{5\pi i/4}} \frac{z^\alpha}{(z - e^{3\pi i/4})(z - e^{i5\pi/4})} = \frac{e^{5\pi i\alpha/4}}{-\sqrt{2}i}$$

Note that it is not wise to write the location of the pole $e^{i5\pi/4}$ as $e^{-i3\pi/4}$, since this is inconsistent with the chosen branch for z^α. Hence, with $\varepsilon \to 0, R \to \infty$:

$$(1 - e^{2\pi i\alpha})I = 2\pi i\left(\frac{e^{3\alpha i\pi/4}}{\sqrt{2}i} + \frac{e^{5\alpha i\pi/4}}{-\sqrt{2}i}\right) \implies I = \sqrt{2}\pi\frac{\sin(\alpha\pi/4)}{\sin(\alpha\pi)}$$

Lemma 12.1: If $z = x + iy$, then
$$|\sinh y| \le f(z) \le \cosh y$$
where $f(z) = |\cos z|, |\sin z|,$ and
$$|\sinh x| \le g(z) \le \cosh x$$
where $g(z) = |\sinh z|, |\cosh z|$. Hence, as $y \to \pm\infty$, $f(z) \sim \frac{1}{2}e^{|y|}$, while as $x \to \pm\infty$, $g(z) \sim \frac{1}{2}e^{|x|}$.

Proof. For $\sin z$, note that $|e^{iz}| = |e^{ix}e^{-y}| = e^{-y}$. Apply the triangle inequality:
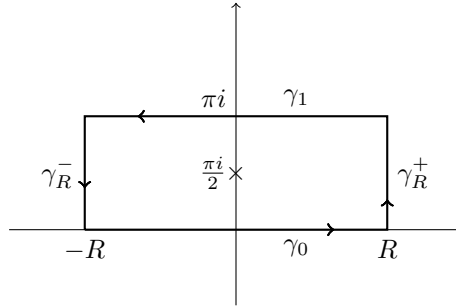$$||z_1| - |z_2|| \le |z_1 - z_2| \le |z_1| + |z_2|$$
with $z_1 = e^{iz}$ and $z_2 = e^{-iz}$. Similar proof for $\cos z$, $\sinh z$ and $\cosh z$. $\qquad\square$

Example 12.12 (Using rectangular contours). *Evaluate*
$$I = \int_{-\infty}^{\infty} \frac{e^{\alpha x}}{\cosh x}dx, \quad \alpha \in \mathbb{R}, \alpha \in (-1, 1)$$

Use a rectangular contour as shown:

We see that
$$\int_{\gamma_0} \frac{e^{\alpha z}}{\cosh z}dz \to I, \text{ as } R \to \infty$$
$$\int_{\gamma_1} \frac{e^{\alpha z}}{\cosh z}dz = \int_R^{-R} \frac{e^{\alpha(x+\pi i)}}{\cosh(x + \pi i)}dx = e^{\alpha\pi i}\int_R^{-R} \frac{e^{\alpha x}}{-\cosh x}dx \to e^{\alpha\pi i}I$$

On γ_R, $z = R + iy \implies |\cosh z| \sim \frac{1}{2}e^R$ from Lemma 12.1. Also $|e^{\alpha z}| = e^{\alpha R}$, hence
$$\left|\frac{e^{\alpha z}}{\cosh z}\right| = O(e^{(\alpha-1)R}) \to 0, \text{ as } R \to \infty$$

so $\int_{\gamma_R} \to 0$. Similar for $\int_{\gamma_{-R}}$. The only singularity enclosed is at $+\frac{\pi i}{2}$ with residue
$$\mathrm{res}_{z=\pi i/2} \frac{e^{\alpha z}}{\cosh z} = \frac{e^{\alpha\pi i/2}}{\sinh(\pi i/2)} = -ie^{\alpha\pi i/2}$$

Hence,
$$I(1 + e^{\alpha\pi i}) = 2\pi i(-ie^{\alpha\pi i/2}) \implies I = \pi\sec\frac{\alpha\pi}{2}$$

It is possible to use contour integration to sum infinite series. Traditionally, this is done using $\cot z$ and a large square contour that has been chosen carefully to avoid all singularities.

Example 12.13 (Summation of series using square contour). *Evaluate $\sum_{n=1}^{\infty} \frac{1}{n^2}$ by integrating suitable function around a square contour. Consider*
$$\oint_\gamma \frac{\cot\pi z}{z^2}dz$$

where γ is the square contour with vertices at $(N + 0.5)(\pm 1 \pm i)$. N is large, so γ avoids all singularities. The integrand have simple poles at $z = n \in \mathbb{Z}$, $n \ne 0$ with residues $\frac{1}{n^2\pi}$. There is also a triple pole at $z = 0$, i.e.
$$\frac{\cot\pi z}{z^2} = \frac{1}{z^2(\pi z + (\pi^3 z^3/3) + \dots)} = \frac{1}{\pi z^3}\left(1 - \frac{1}{3}\pi^2 z^2 + \dots\right)$$

with residue $-\frac{1}{3}\pi$.
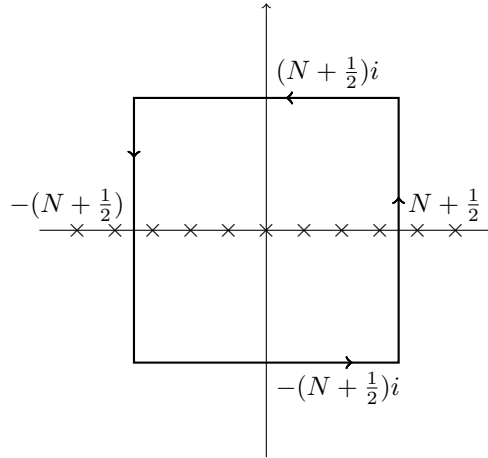
As $N \to \infty$, the integrals along the sides vanish (as we will justify below), so by the Residue theorem:

$$2\pi i \left(2 \sum_{n=1}^{N} \frac{1}{n^2 \pi} - \frac{\pi}{3} \right) \to 0 \implies \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$$

To demonstrate the behaviour of the integrals along the sides, show that on γ, $\cot \pi z$ is bounded by constant M that does not depend on N. By the ML inequality:

$$\left| \oint_{\gamma} \frac{\cot \pi z}{z^2} dz \right| \leq 4(2N+1) \frac{M}{(N+0.5)^2} \to 0, \ as \ N \to \infty$$

Consider the RHS, $z = N + 0.5 + iy$:

$$|\cot \pi z| = \frac{|\cos((N+0.5)\pi + i\pi y)|}{|\sin((N+0.5)\pi + i\pi y)|} = |-\tan i\pi y| = |\tanh \pi y| \leq 1$$

Similar for LHS. Consider top side, $z = (N + 0.5)i + x$:

$$|\cot \pi z| = \frac{|\cos \pi z|}{|\sin \pi z|} \leq \frac{\cosh(N+0.5)\pi}{|\sinh(N+0.5)\pi|} = \coth(N+0.5)\pi$$

where by Lemma 12.1, $|\sin z| \geq |\sinh Im[z]|$ and $|\cos z| \leq \cosh Im[z]$. Now $\coth \theta \to 1$ as $\theta \to \infty$, so in particular it is bounded (by say, 2) for sufficiently large θ; hence $\cot \pi z$ is bounded by 2 for sufficiently large N. Similar for bottom side. Here, we may choose $M = 2$ for sufficiently large N.

13   Transform methods

Fourier inversion by contour integration. Examples of simple linear differential equations, including diffusion equation.   **[2]**

13.1   Jordan's lemma

Lemma 13.1 (Jordan's Lemma)**:** Suppose f is analytic function, except for a finite number of singularities, and $f(z) \to 0$ as $|z| \to \infty$. Then for any real constant $\lambda > 0$,

$$\int_{\gamma_R} f(z)e^{i\lambda z}dz \to 0, \quad \text{as } R \to \infty$$

where γ_R is a semicircle of radius R in the upper half-plane. For $\lambda < 0$, same conclusion holds for semicircle γ_R in the lower half-plane.
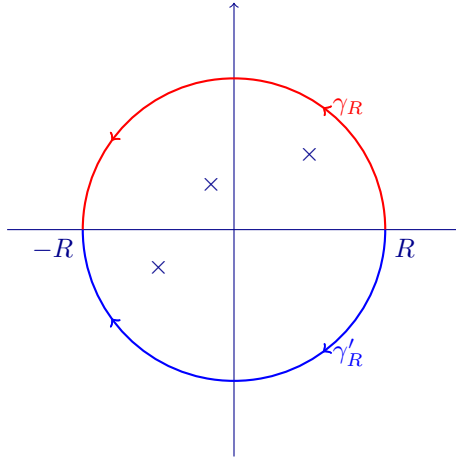
Such integrals are frequently in applications, particularly with Fourier transforms (see Chapter 5).

Remarks 13.1.

1. *Result easy to show in 'straightforward cases' when $f(z)$ decays more rapidly than $\frac{1}{z}$ as $|z| \to \infty$. Then $|f(z)| = o(1/R)$ on γ_R, and since $|e^{i\lambda z}| = e^{-\lambda Im[z]} \le 1$ on γ_R, we have*

$$\left| \int_{\gamma_R} f(z)e^{i\lambda z}dz \right| \le \pi R\, o(1/R) \to 0, \quad \text{as } R \to \infty$$

2. *But in 'sensitive cases', when f decays less rapidly than $o(1/z)$, the proof relies on the fact that for $\theta \in [0, \pi/2]$, $\sin\theta \ge \frac{2\theta}{\pi}$. Let $M_R = \sup_{z \in \gamma_R} |f(z)|$, then $M_R \to 0$ as $R \to \infty$. Now*

$$
\begin{aligned}
\left| \int_{\gamma_R} f(z)e^{i\lambda z}dz \right| &= \left| \int_0^\pi f(Re^{i\theta})e^{i\lambda Re^{i\theta}} iRe^{i\theta}d\theta \right| \\
&\le R\int_0^\pi |f(Re^{i\theta})||e^{i\lambda Re^{i\theta}}|d\theta \\
&\le 2RM_R \int_0^{\pi/2} e^{-\lambda R \sin\theta}d\theta \\
&\le 2RM_R \int_0^{\pi/2} e^{-\lambda R 2\theta/\pi}d\theta \\
&= \frac{\pi}{\lambda}(1 - e^{-\lambda R})M_R \to 0, \quad \text{as } R \to \infty
\end{aligned}
$$

Similar for γ_R' when $\lambda < 0$.

Example 13.1. *Evaluate*

$$I = \int_0^\infty \frac{\cos\alpha x}{1+x^2}dx = \frac{1}{2}\int_{-\infty}^\infty \frac{\cos\alpha x}{1+x^2}dx$$

by evenness, where $\alpha \in \mathbb{R}^+$, a constant. Try the following:

85

1. *Close contour above and try to show $\int_{\gamma_R} \frac{\cos \alpha z}{1+z^2} dz \to 0$ as $R \to \infty$. This approach fails because in fact $\cos \alpha z$ is exponentially large at ∞.*

2. *Write $I = \frac{1}{4} \int_{-\infty}^{\infty} \frac{e^{i\alpha x}}{1+x^2} dx + \frac{1}{4} \int_{-\infty}^{\infty} \frac{e^{-i\alpha x}}{1+x^2} dx$. We can close contour above for the first integral and use Jordan's Lemma on γ_R with $\lambda = \alpha > 0$. For the second one, close below using γ'_R, $\lambda = -\alpha < 0$.*

3. *A quicker method: note $I = \frac{1}{2} Re[\int_{-\infty}^{\infty} \frac{e^{i\alpha x}}{1+x^2} dx]$, so only need to use one contour - close above.*

4. *An even quicker method: note $\int_{-\infty}^{\infty} \frac{\sin \alpha x}{1+x^2} dx = 0$ by oddness, so in fact, $I = \frac{1}{2} \frac{e^{i\alpha x}}{1+x^2} dx$.*

So consider

$$\int_{\gamma_0 + \gamma_R} \frac{e^{i\alpha z}}{1 + z^2} dz$$
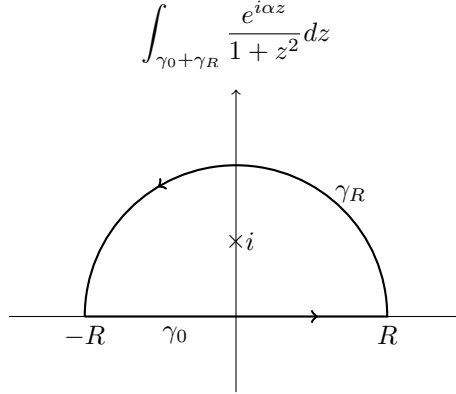
Along γ_0, we obtain $2I$ as $R \to \infty$, whereas on γ_R, we obtain 0 by Jordan's Lemma. This is a 'straightforward case' since $\frac{1}{1+z^2} = o(1/z)$ as $|z| \to \infty$. So

$$I = \frac{1}{2} 2\pi i \operatorname{res}_{z=i} \frac{e^{i\alpha z}}{1 + z^2} = \frac{e^{-\alpha}}{2i} \pi i = \frac{1}{2} \pi e^{-\alpha}$$

To find say $\int_{-\infty}^{\infty} \frac{\sin x}{x} dx$, we can also use Jordan's Lemma. This is a 'sensitive case' because the decay at ∞ is $O(1/z)$, not $o(1/z)$. But we have the complication of a singularity of $\frac{e^{iz}}{z}$ on the real axis.

Example 13.2 (Integration around singular point on real axis)**.** *Evaluate*

$$\int_{-\infty}^{\infty} \frac{\sin x}{x} dx$$

where the integrand is well-behaved at the origin. Use Jordan's Lemma: split $\sin z$ into exponentials, and then splitting the integral in two (closing $\int e^{iz}/z\,dz$ in the upper half-plane and $\int e^{-iz}/z\,dz$ in the lower half-plane). This will cause a problem because e^{iz}/z and e^{-iz}/z are singular at $z = 0$, so the contour will pass directly through a singularity in both integrals.

To avoid this, since $\frac{\sin x}{x}$ is bounded near the origin, we exclude a small interval $(-\varepsilon, \varepsilon)$ from the integral and later $\varepsilon \to 0$ without affecting value of the integral.

$$
\begin{aligned}
\int_{-\infty}^{\infty} \frac{\sin x}{x} dx &= \lim_{\varepsilon \to 0, R \to \infty} \left[\int_{-R}^{-\varepsilon} \frac{\sin x}{x} dx + \int_{\varepsilon}^{R} \frac{\sin x}{x} dx \right] \\
&= \frac{1}{2i} \lim_{\varepsilon \to 0, R \to \infty} \left[\int_{-R}^{-\varepsilon} \frac{e^{iz}}{z} dz - \int_{-R}^{-\varepsilon} \frac{e^{-iz}}{z} dz + \int_{\varepsilon}^{R} \frac{e^{iz}}{z} dz - \int_{\varepsilon}^{R} \frac{e^{-iz}}{z} dz \right] \\
&= \frac{1}{2i} \lim_{\varepsilon \to 0, R \to \infty} \left[\int_{-R}^{-\varepsilon} \frac{e^{iz}}{z} dz + \int_{\varepsilon}^{R} \frac{e^{iz}}{z} dz \right] \qquad\qquad (1) \\
&\quad - \frac{1}{2i} \lim_{\varepsilon \to 0, R \to \infty} \left[\int_{-R}^{-\varepsilon} \frac{e^{-iz}}{z} dz + \int_{\varepsilon}^{R} \frac{e^{-iz}}{z} dz \right] \qquad\qquad (2)
\end{aligned}
$$

Close (1) in the upper half-plane and close (2) in the lower half-plane. Alternatively, write

$$\int_{-\infty}^{\infty} \frac{\sin x}{x} dx = Im\left\{ \lim_{\varepsilon \to 0, R \to \infty} \left[\int_{-R}^{-\varepsilon} \frac{e^{iz}}{z} dz + \int_{\varepsilon}^{R} \frac{e^{iz}}{z} dz \right] \right\}$$

Let γ be a contour from $-R$ to $-\varepsilon$ along the real axis, then round the semicircle γ_ε of radius ε, then from ε to R along the real axis, returning via a semicircle γ_R of radius R.
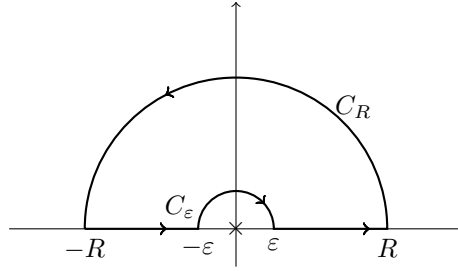
γ encloses no poles of e^{iz}/z, so

$$\int_{-R}^{-\varepsilon} \frac{e^{iz}}{z}dz + \int_{\varepsilon}^{R} \frac{e^{iz}}{z}dz = -\int_{\gamma_\varepsilon} \frac{e^{iz}}{z}dz - \int_{\gamma_R} \frac{e^{iz}}{z}dz$$

By Jordan's Lemma, the integral round γ_R vanish as $R \to \infty$. On γ_ε, $z = \varepsilon e^{i\theta}$, so $e^{iz} = 1 + O(\varepsilon)$

$$\int_{\gamma_\varepsilon} \frac{e^{iz}}{z}dz = \int_{\pi}^{0} \frac{1+O(\varepsilon)}{\varepsilon e^{i\theta}} i\varepsilon e^{i\theta} d\theta = -i\pi + O(\varepsilon) \qquad (*)$$

Take $\varepsilon \to 0$, $R \to \infty$ to get

$$\int_{-\infty}^{\infty} \frac{\sin x}{x}dx = Im[i\pi] = \pi$$

Remarks 13.2.

1. (*) is an example of the Indentation Lemma: which states if γ_ε is an anti-clockwise arc of radius ε and angle α centred around a simple pole z_0 of function $f(z)$, then

$$\int_{\gamma_\varepsilon} f(z)dz \to i\alpha \operatorname{res}_{z=z_0} f(z), \ \ as \ \varepsilon \to 0$$

   i.e. a fraction $\frac{\alpha}{2\pi}$ of the value of the integral around an entire cycle. This is easily proved using the Laurent series. In (*), residue of $\frac{e^{iz}}{z}$ is 1 with $\alpha = \pi$. The contour goes clockwise around $z = 0$, so obtain $-i\pi$.

2. A completely different approach for Example 13.2: the singularity of $\frac{\sin z}{z}$ at the origin is removable. With an analytic integrand, the original contour along the real axis can be moved to one which does not pass through the origin.
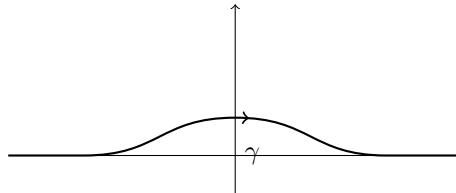
   In this case, it is now possible to write $\sin z = \frac{1}{2i}(e^{iz} - e^{-iz})$, split the integrand in two, and apply Jordan's Lemma to each part separately in the normal way, because our contour no longer passes through a singularity.

13.2 Inverting Fourier transforms using calculus of residues

When inverting $\mathcal{F}$, we generally use a semicircular contour (in the upper half-plane if $x > 0$, and lower half-plane if $x < 0$) and apply Jordan's Lemma.

Example 13.3 (Using contour integration to invert Fourier transforms). *Consider* $f(x) = e^{-\alpha x} H(x)$, *where* $a > 0$ *is a real constant. Then,*

$$\tilde{f}(k) = \int_0^\infty e^{-ax-ikx} dx = -\frac{1}{a+ik}[e^{-ax-ikx}]_0^\infty = \frac{1}{a+ik}$$

Verify this result by evaluating $\frac{1}{2\pi}\int_{-\infty}^\infty \tilde{f}(k)e^{ikx}dk$. *In complex k-plane, let γ_0 be contour from $-R$ to R on real axis, γ_R be semicircle of radius R in the upper half-plane, γ_R' be semicircle of radius R in the lower half-plane. Let γ be γ_0 followed by γ_R, and let γ' be γ_0 followed by γ_R'.*
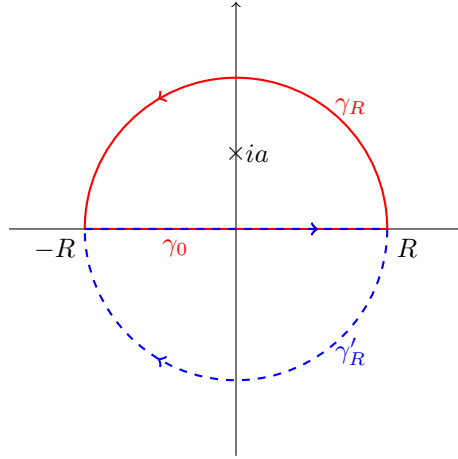
Now $\tilde{f}(k)$ has only one pole, $k = ia$. This pole is simple.

$$\oint_\gamma \tilde{f}(k)e^{ikx}dk = 2\pi i \operatorname{res}_{k=ia} \frac{e^{ikx}}{i(k-ia)} = 2\pi e^{-ax}$$

$$\oint_{\gamma'} \tilde{f}(k)e^{ikx}dk = -0$$

Now if $x > 0$, we can apply the Jordan's Lemma with $\lambda = x$, to γ_R to show that $\int_{\gamma_R} \tilde{f}(k)e^{ikx}dk \to 0$ as $R \to \infty$ since $\tilde{f}(k) = O(1/k)$ as $|k| \to \infty$. Hence for $x > 0$,

$$\frac{1}{2\pi}\int_{-\infty}^\infty \tilde{f}(k)e^{ikx}dk = \frac{1}{2\pi} \lim_{R\to\infty} \oint_\gamma \tilde{f}(k)e^{ikx}dk = e^{-ax}$$

For $x < 0$, close the lower half-plane instead and use γ':

$$\frac{1}{2\pi}\int_{-\infty}^\infty \tilde{f}(k)e^{ikx}dk = \frac{1}{2\pi} \lim_{R\to\infty} \oint_{\gamma'} \tilde{f}(k)e^{ikx}dk = 0$$

Example 13.4 (Damped harmonic oscillator). *The amplitude of a driven damped harmonic oscillator satisfy*

$$\ddot{x}(t) + 2\gamma\dot{x}(t) + \omega_0^2 x(t) = f(t)$$

where $f(t)$ is the forcing function, $\omega_0 > 0$ is real and $\gamma > 0$ is real and represents the effects of damping. Assume $x(t) \to 0$ as $|t| \to \infty$ so that we can define the Fourier transform of $x(t)$. After Fourier transformin g the PDE we arrive at

$$\tilde{x}(\omega) = -\frac{\tilde{f}(\omega)}{(\omega - \omega_+)(\omega - \omega_-)} := \tilde{f}(\omega)\tilde{g}(\omega)$$

where $\omega_\pm = i\gamma \pm \sqrt{\omega_0^2 - \gamma^2}$. Find $x(t)$ by taking inverse Fourier transform and invoke convolution theorem:

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

where $g(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{-e^{i\omega\tau} d\omega}{(\omega-\omega_+)(\omega-\omega_-)}$. *This is equivalent to a solution using the Green function* $G(t, s) = g(t - s)$ *of* $\mathcal{L} = \frac{d^2}{dt^2} + 2\gamma\frac{d}{dt} + \omega_0^2$. *Use contour integration to evaluate. If* $\tau < 0$, *close the lower half-plane (invoke Jordan's Lemma) and if* $\tau > 0$, *close the upper half-plane (invoke Jordan's Lemma). There are two poles enclosed in* $\tau > 0$, *so* $g(\tau > 0) = \frac{1}{2\pi} \oint_C \tilde{g}(\omega)e^{i\omega\tau} d\omega$ *where* C *is the semi-circle in the upper-half plane. But there are no poles enclosed in* $\tau < 0$, *so* $g(\tau < 0) = 0 \implies x(t < 0) = 0$. *This is consistent with causal behaviour. Evaluating the residues, we then use residue theorem to show for the different regimes of damping:*

$$g(\tau > 0) = \begin{cases} \frac{e^{-\gamma\tau}}{\sqrt{\omega_0^2-\gamma^2}} \sin(\tau\sqrt{\omega_0^2 - \gamma^2}) & \gamma < \omega_0 \\ \frac{e^{-\gamma\tau}}{\sqrt{-\omega_0^2+\gamma^2}} \sinh(\tau\sqrt{-\omega_0^2 + \gamma^2}) & \gamma > \omega_0 \\ \tau e^{-\gamma\tau} & \gamma = \omega_0 \end{cases}$$

Example 13.5. *If* $f(x) = e^{-x^2/2}$, *then*

$$\tilde{f}(k) = \int_{-\infty}^{\infty} e^{-x^2/2} e^{-ikx} dx = \int_{-\infty}^{\infty} e^{-(x+ik)^2/2} e^{-k^2/2} dx = e^{-k^2/2} \int_{\gamma_0} e^{-z^2/2} dz$$

where $z = x + ik$, *and* γ_0 *runs along* $\mathrm{Im}[z] = k$, *in the limit* $R \to \infty$.
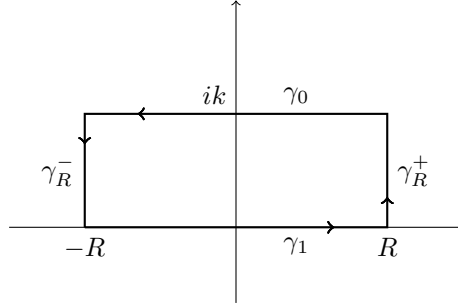
Now, $\int_{\gamma_R} \to 0$ *and* $\int_{\gamma_{-R}} \to 0$ *(both bounded by* $ke^{-(R^2-k^2)/2}$, *which is* $O(e^{-R^2/2})$ *for fixed* k*), and there are no singularities, so* $\int_{\gamma_0} = -\int_{\gamma_1} = \int_{-\infty}^{\infty}$ *in the limit. Hence, using a standard result from real analysis:*

$$\tilde{f}(k) = e^{-k^2/2} \int_{-\infty}^{\infty} e^{-z^2/2} dz = \sqrt{2\pi} e^{-k^2/2}$$

Lemma 13.2 (Gaussian integration lemma)**:** Extending $I = \int_{-\infty}^{\infty} e^{-(u+a)^2} du = \sqrt{\pi}$ to $a \in \mathbb{C}$, we have the exact same result.

Proof. Define $z := u + a \in \mathbb{C}$, and consider $I = \oint_C e^{-z^2} dz$. Evaluate this using the same contour in Example 13.5.

$$0 = \int_{-R}^{R} e^{-x^2} dx + \int_0^{\mathrm{Im}[a]} e^{-(R+iy)^2} i\,dy - \int_{\gamma_0} e^{-z^2} dz + \int_{\mathrm{Im}[a]}^0 e^{-(-R+iy)^2} i\,dy = \sqrt{\pi} - I + 2e^{-R^2} \int_0^{\mathrm{Im}[a]} e^{y^2} \sin(2Ry) dy$$

Final term tends to zero as $R \to \infty$. $\qquad\square$

Example 13.6 (Diffusion equation)**.** *Consider diffusion equation* $\frac{\partial T}{\partial t} = \lambda\frac{\partial^2 T}{\partial x^2}$ *with Fourier transform* $\frac{\partial \tilde{T}}{\partial t} = -\lambda k^2 \tilde{T} \implies \tilde{T}(k, t) = \tilde{T}(k, 0)e^{-\lambda k^2 t}$ *where* $\tilde{T}(k, 0) = \int_{-\infty}^{\infty} e^{-ikx} T(x, 0) dx$. *The Green's function (also known as the heat kernel in this context) is* $\tilde{G}(k, t) = e^{-\lambda k^2 t}$, *then we have*

$$T(x, t) = \int_{-\infty}^{\infty} T(y, 0) G(x - y, t) dy, \quad G(x, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ikx - \lambda k^2 t} dk$$

Let $u = \sqrt{\lambda t} k$ *and use the Gaussian integration lemma to get*

$$G(x, t) = \frac{e^{-x^2/4\lambda t}}{2\pi\sqrt{\lambda t}} \int_{-\infty}^{\infty} e^{-(u-(ix/2\sqrt{\lambda t}))^2} du = \frac{e^{-x^2/4\lambda t}}{\sqrt{4\pi\lambda t}} \implies T(x, t) = \frac{1}{\sqrt{4\pi\lambda t}} \int_{-\infty}^{\infty} T(y, 0) e^{-(x-y)^2/4\lambda t} dy$$

14   Small oscillations

Small oscillations and equilibrium; normal modes, normal coordinates, examples, e.g. vibrations of linear molecules such as CO_2. Symmetries of normal modes **[2]**

Definition 14.1 (Normal modes and normal frequency)**:** A normal mode of an oscillating system is a pattern of motion in which all parts of the system move sinusoidally with the same frequency and with a fixed phase relation.

Definition 14.2 (Generalized coordinates)**:** In analytical mechanics, the term generalized coordinates refers to the parameters that uniquely describe the configuration of the system relative to some reference configuration.

Proposition 14.1: A system has n degrees of freedom and undergoes small oscillations about an equilibrium point. The equations of motion in terms of the generalized coordinates is a system of differential equation

$$\frac{d}{dt}T_{ij}\dot{q}_j = -V_{ij}q_j$$

where T_{ij} and V_{ij} are entries of $n \times n$ matrices for the kinetic energy and potential energy respectively.

Proof. Let each particle be described by a generalized coordinate q_i, and assuming there are no velocity-dependent potentials and the kinetic energies are quadratic in the speeds, then the Lagrangian is the kinetic energy minus the potential energy $\mathcal{L} = T - V$, and has the form

$$\mathcal{L}(\mathbf{q}, \dot{\mathbf{q}}, t) = \sum_{i=1}^{n} \frac{1}{2}T_{ij}(t)\dot{q}_i\dot{q}_j - V(\mathbf{q}, t)$$

If we further assume $\mathcal{L}$ to be time-independent and expand the potential around equilibrium positions (minima). At equilibrium, $\frac{\partial \mathcal{L}}{\partial \mathbf{q}} = \frac{\partial \mathcal{T}}{\partial \mathbf{q}} - \frac{\partial \mathcal{V}}{\partial \mathbf{q}} = 0 - 0 = 0$. When we have a deviation $\Delta\mathbf{q} = (q_1, q_2, q_3)$ from $\mathbf{q_0} = \mathbf{0}$, then the Taylor expansion of $\mathcal{V}$ (valid since oscillation amplitudes are small) is

$$\mathcal{V}(\Delta\mathbf{q}) = \mathcal{V}(\mathbf{0}) + 0 + \frac{1}{2}(\Delta\mathbf{q}\cdot\boldsymbol{\nabla})^2\mathcal{V} + O(\Delta\mathbf{q})^3 = \mathcal{V}(\mathbf{0}) + \frac{1}{2}\begin{pmatrix} q_1 & q_2 & q_3 \end{pmatrix}\begin{pmatrix} \frac{\partial^2\mathcal{V}}{\partial q_1^2} & \frac{\partial^2\mathcal{V}}{\partial q_1 q_2} & \frac{\partial^2\mathcal{V}}{\partial q_1 q_3} \\ \frac{\partial^2\mathcal{V}}{\partial q_1 q_2} & \frac{\partial^2\mathcal{V}}{\partial q_2^2} & \frac{\partial^2\mathcal{V}}{\partial q_2 q_3} \\ \frac{\partial^2\mathcal{V}}{\partial q_1 q_3} & \frac{\partial^2\mathcal{V}}{\partial q_2 q_3} & \frac{\partial^2\mathcal{V}}{\partial q_3^2} \end{pmatrix}\begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} + O(\Delta\mathbf{q}^2)$$

so the quadratic approximation to the Lagrangian is

$$\mathcal{L} \approx \sum_{i=1}^{n} \frac{1}{2}T_{ij}\dot{q}_i\dot{q}_j - \sum_{i=1}^{n} \frac{1}{2}\frac{\partial^2 V}{\partial q_i \partial q_j}q_iq_j$$

unique up to an additive constant (potential at the equilibrium position), which we can set to zero without loss of generality. Also, we assumed for any principal directions for which V_{ij} has a zero eigenvalue, all subsequent derivatives are also zero. When the Lagrangian $\mathcal{L}$ is extremized, it satisfies the Euler-Lagrange equations

$$\frac{d}{dt}\frac{\partial\mathcal{L}}{\partial\dot{q}_i} = \frac{\partial\mathcal{L}}{\partial q_i}$$

to give the equations of motion

$$\frac{d}{dt}T_{ij}\dot{q}_j = -V_{ij}q_j$$

If T is time-independent, then LHS becomes $T_{ij}\ddot{q}_j$. $\qquad\square$

Remarks 14.1. *Let* (X_i, Y_i) *be the vector* $\mathbf{X_i}$. *The quadratic approximation to the potential energy for each string (of original length ℓ_0) is*

$$\frac{1}{2}k\left(|\mathbf{X_i} - \mathbf{X_j}| - \ell_0\right)^2 \approx \frac{k}{2}\left(\frac{\mathbf{a}\cdot\mathbf{v}}{\ell_0}\right)^2$$

where we write $\mathbf{X_i} - \mathbf{X_j}$ *as* $\mathbf{v} + \mathbf{a}$ *where* $|\mathbf{a}| = \ell_0 >> |\mathbf{v}|$ *and thus,*

$$|\mathbf{a} + \mathbf{v}| = (\mathbf{a}\cdot\mathbf{a} + 2\mathbf{a}\cdot\mathbf{v} + \mathbf{v}\cdot\mathbf{v})^{1/2} = \ell_0\left(1 + \frac{\mathbf{a}\cdot\mathbf{v}}{\ell_0^2} + \dots\right)$$

Corollary 14.1: The normal frequencies ω can be obtained by solving the characteristic equation

$$\det[\omega^2 T - V] = 0$$

which is a polynomial of degree N in ω^2.

Proof. A normal mode is a solution to this generalized eigenvalue of the form $\mathbf{q} = \mathbf{q_0}e^{i\omega t}$ or $\mathbf{q} = \mathbf{a} + t\mathbf{b}$, where latter is known as the zero mode. To find the normal mode frequencies, we solve the matrix equation

$$\omega^2 T\mathbf{q} = -V\mathbf{q}$$

and for each $\omega = 0$, we need a linearly independent zero mode. Looking for the non-trivial solutions $\mathbf{Q}$ is equivalent to finding the eigenfrequencies ω by solving $\det(-\omega^2\mathcal{T} + \mathcal{V}) = 0$. $\qquad\square$

Remarks 14.2. *Periodicity is attained only if all the frequency ratios are rational. Zero modes occur when that normal mode has symmetry (translational or rotational, etc).*

Definition 14.3 (Normal coordinates)**:** Normal coordinates are linear combinations of the original generalized coordinates $q(t)$ which oscillate with a single pure frequency.

Remarks 14.3. *V must be a semi-positive matrix, i.e. $V_{ij}q_iq_j \geq 0 \ \forall q_i$ since the forces derived from the potential are either restoring forces or zero (no force). T however must be positive definite, i.e. $T_{ij}\dot{q}_i\dot{q}_j > 0 \ \forall \dot{\mathbf{q}} \neq \mathbf{0}$.*

Definition 14.4 (Generalized eigenvectors and eigenvalues)**:** Suppose the non-zero solutions of $\omega^2 T\mathbf{q} = -V\mathbf{q}$ are $\mathbf{Q}$, then $\mathbf{Q}$ are said to be the generalized eigenvectors (generalized as long as $T \neq I$ identity) while

$$\omega^2 = \frac{V_{ij}Q_iQ_j}{T_{ij}Q_iQ_j}$$

are the generalized eigenvalues.

Proposition 14.2: The orthogonality relation of two normal modes $\mathbf{Q^{(i)}}$ and $\mathbf{Q^{(j)}}$ is

$$(\mathbf{Q^{(i)}})^T\mathcal{T}\mathbf{Q^{(j)}} = \delta_{ij}$$

where the orthogonal relation is with respect to the matrix with entries $(\mathcal{T})_{ij} = T_{ij}$.

Proof. Suppose there are two generalized eigenvectors of distinct normal frequencies, namely $\mathbf{Q^{(i)}}$ and $\mathbf{Q^{(j)}}$ with $\omega_i \neq \omega_j$. Then, they separately satisfy

$$(-\omega_i^2\mathcal{T} + \mathcal{V})\mathbf{Q^{(i)}} = \mathbf{0}, \quad (-\omega_j^2\mathcal{T} + \mathcal{V})\mathbf{Q^{(j)}} = \mathbf{0}$$

This gives $(\omega_i^2 - \omega_j^2)(\mathbf{Q^{(i)}})^T\mathcal{T}\mathbf{Q^{(j)}} = 0$. Since $\omega_i^2 - \omega_j^2 \neq 0$, we must have $(\mathbf{Q^{(i)}})^T\mathcal{T}\mathbf{Q^{(j)}} = 0$, i.e. the normal modes are orthogonal with respect to $\mathcal{T}$. $\qquad\square$

Remarks 14.4. *The general solution for this small oscillation problem is a linear combination of the generalized eigenvectors multiplied by a phase related to the generalized eigenfrequencies, i.e.*

$$\mathbf{q}(t) = Re\left[\sum_j c_j\mathbf{e_i}e^{i\omega_j t}\right]$$

Suppose the initial conditions are $\mathbf{q}(0)$ and $\dot{\mathbf{q}}(0)$ then we can find the constants by exploiting orthogonality and solve

$$c_i\langle\mathbf{e_i}|\mathbf{e_i}\rangle_T = \langle\mathbf{e_i}|\mathbf{q}(0)\rangle_T$$

where the dot product is taken with respect to the matrix T.

15   Group theory

Idea of an algebra of symmetry operations; symmetry operations on a square. Definition of a group: group table. Subgroups; homomorphic and isomorphic groups.   **[5]**

15.1   Definitions and examples

Definition 15.1 (Binary Operation)**:** A binary operation on a set X is a function $X \times X \to X$.

Definition 15.2 (Group)**:** A group consists of a set G, a binary operation $\cdot$ on G, and an element $e \in G$ s.t. the elements satisfy the three group axioms:

1. Associativity: $(a \cdot b) \cdot c = a \cdot (b \cdot c) \ \forall a, b, c \in G$;

2. Identity: $a \cdot e = a \ \forall a \in G$;

3. Inverse: For each $a \in G$, $\exists b \in G$ s.t. $a \cdot b = e$.

Definition 15.3 (Group Table)**:** A group table describes the structure of a finite group by arranging all the possible products of all the group's elements in a square table. It illustrates closure (an important requirement for a group) where all possible products of the binary operation must still be in the group. Also due to closure, each row of the group table is just a permutation of the group elements.

Theorem 15.1: Let $(G, \cdot, e)$ be a group.

1. For $a, b \in G$, if $a \cdot b = e$, then $b \cdot a = e$;

2. If $a \in G$, then $e \cdot a = a$;

3. If $a, b, b' \in G$ are s.t. $a \cdot b = e$ and $a \cdot b' = e$, then $b = b'$, i.e. inverse is unique;

4. If $e' \in G$ is s.t. $a \cdot e' = a$ for some $a \in G$, then $e' = e$, i.e. identity is unique.

Proof. We repeatedly invoke the three group axioms:

1. Assume $a \cdot b = e$, then observe that $b = b \cdot e = b \cdot (a \cdot b) = (b \cdot a) \cdot b$. But $\exists c \in G$ s.t. $b \cdot c = e$, then
$$e = ((b \cdot a) \cdot b) \cdot c = (b \cdot a) \cdot (b \cdot c) = (b \cdot a) \cdot e = b \cdot a$$

2. $\exists b \in G$ s.t. $a \cdot b = e$ and $b \cdot a = e$ both hold, then
$$e \cdot a = (a \cdot b) \cdot a = a \cdot (b \cdot a) = a \cdot e = a$$

3. Since we have $a \cdot b = e \implies b \cdot a = e$, and assuming $a \cdot b' = e$,
$$b' = e \cdot b' = (b \cdot a) \cdot b' = b \cdot (a \cdot b') = b \cdot e = b$$

4. $\exists b \in G$ s.t. $a \cdot b = e$ and $b \cdot a = e$ both hold, then
$$e = b \cdot a = b \cdot (a \cdot e') = (b \cdot a) \cdot e' = e \cdot e' = e$$

$\square$

Corollary 15.1:
$$(a^{-1})^{-1} = a$$

It is convenient to extend this notation. For any $a \in G$, let $a^0 = e$ and for $n \in \{1, 2, 3, ...\}$, inductively define $a^n = a \cdot a^{n-1}$. Similarly for $n \in \mathbb{Z}^-$. It follows that $a^n \cdot a^m = a^{n+m}$ and $(a^n)^m = a^{nm}$ for $n, m \in \mathbb{Z}$.

Corollary 15.2: Given $(G, \cdot)$ and two elements $e, e' \in G$ s.t. $(G, \cdot, e)$ and $(G, \cdot, e')$ are both groups, then $e' = e$.

Proof. By Theorem 15.1, $e' = e \cdot e'$, then by the second group axiom, $e \cdot e' = e$. $\square$

Definition 15.4 (Abelian)**:** A group $(G, \cdot, e)$ is abelian if $\forall a, b \in G$, we have $a \cdot b = b \cdot a$.

Definition 15.5 (Finite)**:** A group $(G, \cdot, e)$ is finite if the set G has finitely-many elements.

Definition 15.6 (Order)**:** The order of a group is the number of elements of G. This is written as $|G|$.

Example 15.1. *We consider a few examples and counterexamples of groups.*

- *If $\{e\}$ is the set with a single element, then $(\{e\}, \cdot, e)$ is the trivial group;*

- *$(\mathbb{Z}, +, 0)$, $(\mathbb{Q}, +, 0)$, $(\mathbb{R}, +, 0)$, $(\mathbb{C}, +, 0)$ are all abelian groups;*

- *$(\mathbb{N}, +, 0)$ is not a group, as there is no natural number N s.t. $1 + N = 0$;*

- *$(\mathbb{Z}, -, 0)$ is not a group since subtraction is not associative;*

- *$(\mathbb{Q}, \times, 1)$ is not a group, as there is no rational number q s.t. $0 \times q = 1$. But $(\mathbb{Q}\backslash\{0\}, \times, 1)$ is a group, which is also abelian;*

- *If $n \in \mathbb{N}$, then the set $\mathbb{Z}_n = \{0, 1, 2, .., n-1\}$ can be given the structure of a group by letting $a +_n b$ be the remainder when $a + b$ is divided by n. Thus, $(\mathbb{Z}_n, +_n, 0)$ is an abelian group, which is finite of order n;*

- *Let Isom($\mathbb{R}$) be the set of functions $f : \mathbb{R} \to \mathbb{R}$ which are distance-preserving s.t. $|f(x) - f(y)| = |x - y| \ \forall x, y \in \mathbb{R}$. The identity e is the identity function $\mathrm{Id}(x) = x$, and the binary operation is the composition of function $\circ$. Then, this is a group. We can check that this is not abelian, e.g. $t \circ r \neq r \circ t$, where $r(x)$ and $t(x)$ are reflection and translation operations respectively;*

- *Let $\mathrm{GL}_2(\mathbb{R})$ denote the set of all 2×2 matrices with real entries which are invertible. The binary operation is the matrix multiplication, and e is the identity matrix. This is a group.*

Definition 15.7 (Subgroup)**:** We say that a group $\{H, \cdot_H, e_H\}$ is a subgroup of $(G, \cdot_G, e_G)$ if

1. H is a subset of G, i.e. $H \subset G$;

2. $e_H = e_G$;

3. $a \cdot_G b = a \cdot_H b \ \forall a, b \in H$.

We write $(H, \cdot_H, e_H) \leq (G, \cdot_G, e_G)$.

Proposition 15.1: Let $(G, \cdot_G, e_G)$ be a group and $H \subset G$ be a non-empty subset. If $a \cdot_G b^{-1} \in H$ $\forall a, b \in H$, then there is a unique binary operation $\cdot_H$ and $e_H \in H$, making $(H, \cdot_H, e_H)$ a subgroup of $(G, \cdot_G, e_G)$.

Proof. As $H \subset G$ is non-empty we may choose a $x \in H$ s.t. $e_G = x \cdot_G x^{-1} \in H$. Then it follows that $a^{-1} = e_G \cdot_G a^{-1} \in H \ \forall a \in H$. Now for any $a, b \in H$, we have $a \cdot_G b = a \cdot_G (b^{-1})^{-1} \in H$. The binary operation can be defined as

$$\cdot_H : H \times H \quad \to \quad H$$
$$(a, b) \quad \mapsto \quad a \cdot_G b$$

and define the identity $e_H := e_G \in H$. Now, check the group axioms. $\cdot_G$ is associative $\implies \cdot_H$ as well. Also, $a \cdot_H e_H = a \cdot_G e_G = a$, so the identity axiom holds. For any $a \in H$, we have shown $a^{-1} \in H$ and $a \cdot_H a^{-1} = a \cdot_G a^{-1} = e_G = e_H$, so the inverse axiom holds. $\square$

Example 15.2. *Below are some examples of subgroups.*

- *$(\mathbb{Z}, +, 0) \leq (\mathbb{Q}, +, 0) \leq (\mathbb{R}, +, 0) \leq (\mathbb{C}, +, 0)$;*

- *any group is a subgroup of itself;*

- *the trivial subgroup is the subgroup for any group;*

- *if n divides m, then $C_n \leq C_m$;*

- *the subset $\mathrm{SL}_2(\mathbb{R}) \subset \mathrm{GL}_2(\mathbb{R})$ of those 2×2 matrices with determinant equal to 1 is a subgroup.*

Proposition 15.2: The subgroups of $(\mathbb{Z}, +, 0)$ are given precisely by the subsets $n\mathbb{Z} \subset \mathbb{Z}$ for some $n \in \mathbb{N}$, where $n\mathbb{Z} := \{kn \in \mathbb{Z} | k \in \mathbb{Z}\}$ denotes the set of multiples of n.

Proof. We first check closure. For $a, b \in n\mathbb{Z}$, $\exists a', b' \in \mathbb{Z}$, then

$$a + (-b) = na' - nb' = n(a' - b') \in n\mathbb{Z}$$

We can show using Proposition 15.1 that $n\mathbb{Z} \subset \mathbb{Z}$ is indeed a subgroup for each $n \in \mathbb{N}$.

Finally, we are left to show that the only subgroups of $\mathbb{Z}$ are precisely these subsets $n\mathbb{Z}$. Let $S \subset \mathbb{Z}$ be a subgroup. If $S = \{0\}$ (trivial subgroup), we are done. Let $n \in S$ be the smallest positive element of S. As S is a subgroup, it is closed under negation and addition, so $n\mathbb{Z} \subset S$.

Suppose $n\mathbb{Z} \neq S$, $\exists x \in S \backslash n\mathbb{Z}$ s.t. $nk < x < n(k+1)$ for some k. This gives $0 < x - nk < n$, showing that $x - nk$ is a positive element, that is smaller than n. This contradicts our construction. Hence, $n\mathbb{Z} = S$. $\square$

From now on, we will just write $(G, \cdot, e_G)$ as G.

Definition 15.8 (Direct Product)**:** If $(G, \cdot_G, e_G)$ and $(H, \cdot_H, e_H)$ are groups then we define a binary operation $\cdot_{G \times H}$ on the set $G \times H$ via

$$(g_1, h_1) \cdot_{G \times H} (g_2, h_2) := (g_1 \cdot_G g_2, h_1 \cdot_H h_2)$$

and $(G \times H, \cdot_{G \times H}, (e_G, e_H))$ is a group, called the direct product of G and H.

Dihedral Group

Definition 15.9 (Dihedral Group)**:** The dihedral group D_{2n} (sometimes written as D_n if specified) consists of the set of symmetries of the regular n-gon which has vertices $\{e^{i2\pi j/n}\}_{i=1}^{n-1}$ in the complex plane $\mathbb{C}$. Essentially, the dihedral group consists of the isometries of the complex plane $\mathbb{C}$ which send the n-gon to itself.

Theorem 15.2: The set of symmetries of the regular n-gon is indeed a group and has $2n$ elements.

Proof. Check the group axioms. The elements of D_{2n} are isometries which send the n-gon to itself. Clearly, if $f, g : \mathbb{C} \to \mathbb{C}$, then $f \circ g$ is also an isometry, hence D_{2n} is associative for $\circ$. The identity element is just the identity function which $x \mapsto x$, s.t. $\forall f \in D_{2n}$, $(f \circ \mathrm{Id})(x) = f(x) \; \forall x \in \mathbb{C}$. To show the inverse of all elements always exist, we need to figure the elements in D_{2n}.

One can construct the elements of D_{2n} using the rotation map $r : z \mapsto ze^{2\pi i/n}$ and reflection map (about the real axis) $s : z \mapsto \overline{z}$. If $z, w \in \mathbb{C}$, we can check r and s are indeed isometries.

$$|r(z) - r(w)| = |ze^{2\pi i/n} - we^{i2\pi/n}| = |z - w||e^{2\pi i/n}| = |z - w|$$

$$|s(z) - s(w)| = |\overline{z} - \overline{w}| = \sqrt{(\overline{z} - \overline{w})\overline{(\overline{z} - \overline{w})}} = |z - w|$$

In addition, $r^n = \mathrm{Id}$, $s^2 = \mathrm{Id}$ s.t. $r^{-1} = r^{n-1}$ and $s^{-1} = s$. We want to show that the only isometries of D_{2n} are combinations of r and s, i.e. Let $f \in D_{2n}$ be some isometry, then f must have the form $f = r^k$ or $r^k s$, $\forall k \in [1, n-1]$.

To start, consider the vertex 1, then $f(1)$ must also be a vertex, i.e. $e^{2\pi ik/n}$ for some k. But $r^k(1) = e^{2\pi ik/n}$, and so $(r^{-k} \circ f)(1) = 1$, i.e. $g := r^{-k} \circ f$ is an isometry which fixes 1. The neighbouring vertex $e^{2\pi i/n}$ shares an edge with 1, so since g is an isometry, $g(e^{2\pi i/n})$ must also be a vertex which shares an edge with $g(1) = 1$. There are two possible neighbours $e^{2\pi i/n}$ and $e^{2\pi i(n-1)/n}$.

- $e^{2\pi i/n}$: Then we have shown g fixes both $e^{2\pi i/n}$ and 1 (effectively interchanging vertices). Also, observe that g fixes the centre of mass, which is at position:

$$\frac{1}{n} \sum_{k=0}^{n-1} e^{2\pi ik/n} = \frac{1}{n} \frac{1 - e^{2\pi i/n}}{1 - e^{2\pi i/n}} = 0$$

  Hence, g is the identity and $f = r^k$.

- $e^{2\pi i(n-1)/n}$: Observe that $(s \circ g)(e^{2\pi i/n}) = s(e^{2\pi i \frac{n-1}{n}}) = e^{2\pi i(1-n)/n} = e^{2\pi i/n}$. Also, $(s \circ g)(1) = s(1) = 1$. Thus, $s \circ g$ fixes $e^{2\pi i/n}$ and 1. Then, $s \circ g = \mathrm{Id}$ and we have $s \circ r^{-k} \circ f = \mathrm{Id} \implies f = r^k \circ s$.

$\square$

Corollary 15.3:
$$s \circ r = r^{-1} \circ s$$

Proof. Observe
$$(s \circ r)(1) = s(e^{2\pi i/n}) = e^{2\pi i(n-1)/n} = r^{n-1}(1) = r^{-1}(1)$$

where the inverse of r is r^{n-1}. Then,

$$(r^{-(-1)} \circ s \circ r)(e^{2\pi i/n}) = e^{2\pi i(1-n)/n}\overline{e^{2\pi i/n}e^{2\pi i/n}} = e^{-2\pi i(1+n)/n} = e^{2\pi i(n-1)/n} = s(e^{2\pi i/n})$$

Hence $s \circ r^{-(-1)} \circ s \circ r = \mathrm{Id}$, and the result follows. $\qquad\square$

In another words, D_{2n} is not abelian for any $n \geq 3$.

Symmetric Group

Definition 15.10 (Permutation)**:** For a set X, a permutation of X is a function $f : X \to X$ which is invertible, i.e. $\exists g : X \to X$ satisfying $(f \circ g)(x) = x$ and $(g \circ f)(x) = x \ \forall x \in X$.

Theorem 15.3: For any set X, the set of permutations of X is called $\mathrm{Sym}(X)$. $(\mathrm{Sym}(X), \circ, \mathrm{Id})$ is a group, called the symmetric group of X.

Proof. Check group axioms again! Function composition is associative. Id acts as an identity for composition of functions. By assumption, the elements of $\mathrm{Sym}(X)$ have inverses. All we need to check is closure: Suppose $f, f' \in \mathrm{Sym}(X)$, then by assumption $\exists g, g'$ s.t. $f \circ g = \mathrm{Id}$ and $f' \circ g' = \mathrm{Id}$, then

$$(f \circ f') \circ (g' \circ g) = f \circ (f' \circ g') \circ g = f \circ g = \mathrm{Id}$$
$$(g' \circ g) \circ (f \circ f') = g' \circ (g \circ f) \circ f' = g' \circ f' = \mathrm{Id}$$

Hence, $f \circ f' \in \mathrm{Sym}(X)$. $\qquad\square$

Proposition 15.3: $|\mathrm{Sym}(X)| = |X|!$.

Proof. Combinatorics problem. Arrange n objects in n prescribed positions. $\qquad\square$

Quaternion Group

Definition 15.11 (Quaternion group)**:** $(\mathcal{Q}_8, \cdot, \mathbf{1})$ is the quaternion group with 8 elements $\{\pm\mathbf{1}, \pm\boldsymbol{i}, \pm\boldsymbol{j}, \pm\boldsymbol{k}\}$ where

$$\mathbf{1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{i} = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, \quad \boldsymbol{j} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad \boldsymbol{k} = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}$$

and are related by:

$$(-\mathbf{1})^2 = \mathbf{1}, \quad \boldsymbol{i}^2 = \boldsymbol{j}^2 = \boldsymbol{k}^2 = -\mathbf{1}, \quad \boldsymbol{ij} = \boldsymbol{k}, \quad \boldsymbol{jk} = \boldsymbol{i}, \quad \boldsymbol{ki} = \boldsymbol{j}, \quad \boldsymbol{ji} = -\boldsymbol{k}, \quad \boldsymbol{kj} = -\boldsymbol{i}, \quad \boldsymbol{ik} = -\boldsymbol{j}$$

15.2   Homomorphisms and cyclic groups

Definition 15.12 (Homomorphism)**:** If H and G are groups then a function $\phi : H \to G$ is called a group homomorphism, if $\forall a, b \in H$, we have $\phi(a \cdot_H b) = \phi(a) \cdot_G \phi(b)$.

In another words, homomorphisms preserves structure.

Definition 15.13 (Isomorphism)**:** If the homomorphism ϕ is invertible, then it is called a group isomorphism. If H is isomorphic to G, then $H \simeq G$.

Example 15.3. *Some examples of homomorphisms:*

- *If n divides m, the function $\phi : C_m \to C_n, \quad z \mapsto z^{m/n}$ is a homomorphism.*

- *The determinant gives a homomorphism:*

$$\det : (\mathrm{GL}_2(\mathbb{R}), \cdot, I_{2\times 2}) \to (\mathbb{R}\backslash\{0\}, \times, 1)$$

Lemma 15.1: If $\phi : H \to G$ is a group homomorphism, then

1. $\phi(e_H) = e_G$

2. $\forall a \in H, \phi(a^{-1}) = \phi(a)^{-1}$.

Proof. This follows from the definition of homomorphism.

1. Apply $\phi(e_H)^{-1}$ to both sides of the statement below:

$$\phi(e_H) \cdot_G \phi(e_H) = \phi(e_H \cdot_H e_H) = \phi(e_H) \implies \phi(e_H) = e_G$$

2. The defining property of $\phi(a)^{-1}$ is also satisfied by $\phi(a^{-1})$:

$$\phi(a) \cdot_G \phi(a^{-1}) = \phi(a \cdot_H a^{-1}) = \phi(e_H) = e_G$$
$$\phi(a^{-1}) \cdot_G \phi(a) = \phi(a^{-1} \cdot_H a) = \phi(e_H) = e_G$$

$\square$

Corollary 15.4: The composition of two homormophisms is a homomorphism.

Proof. Take $\phi : H \to G$ and $\psi : G \to K$ as homomorphisms, then

$$\psi(\phi(a \cdot_H b)) = \psi(\phi(a) \cdot_G \phi(b)) = \psi(\phi(a)) \cdot_K \psi(\phi(b)) = (\psi \circ \phi)(a) \cdot_K (\psi \circ \phi)(b)$$

where $\psi \circ \phi$ represents the composition. $\square$

Definition 15.14 (Image)**:** For the group homomorphism $\phi : H \to G$, the image of ϕ is

$$\mathrm{Im}(\phi) := \{g \in G | g = \phi(h) \text{ for some } h \in H\}$$

Definition 15.15 (Kernel)**:** For the group homomorphism $\phi : H \to G$, the kernel of ϕ is

$$\mathrm{Ker}(\phi) := \{h \in H | \phi(h) = e_G \text{ for some } h \in H\}$$

Proposition 15.4: If $\phi : H \to G$ is a group homomorphism, then $\mathrm{Im}(\phi) \leq G$ and $\mathrm{Ker}(\phi) \leq H$.

Proof. For both, we check that the set is non-empty before checking that it is a subgroup.

For $\mathrm{Im}(\phi)$, we have $\phi(e_H) = e_G \implies e_G \in \mathrm{Im}(\phi)$, so this set is non-empty. If $a = \phi(x), b = \phi(y) \in \mathrm{Im}(\phi)$, then

$$a \cdot_G b^{-1} = \phi(x) \cdot_G \phi(y)^{-1} = \phi(x) \cdot_G \phi(y^{-1}) = \phi(x \cdot_H y^{-1}) \in \mathrm{Im}(\phi)$$

For $\mathrm{Ker}(\phi)$, we have $\phi^{-1}(e_G) = e_H \in \mathrm{Ker}(\phi)$, so this set is non-empty. If $a, b \in \mathrm{Ker}(\phi)$, then

$$\phi(a \cdot_H b^{-1}) = \phi(a) \cdot_G \phi(b^{-1}) = \phi(a) \cdot_G \phi(b)^{-1} = e_G \cdot_G e_G^{-1} = e_G$$

so $a \cdot_H b^{-1} \in \mathrm{Ker}(\phi)$. $\square$

Lemma 15.2: If $\phi : H \to G$ is a homomorphism, then it is an isomorphism iff $\text{Im}(\phi) = G$ and $\text{Ker}(\phi) = \{e_H\}$. Also, $\phi^{-1} : G \to H$ is also an isomorphism.

Proof. Suppose first that ϕ is invertible. It follows that:

- $g = \phi(\phi^{-1}(g)) \; \forall g \in G$. Hence, $\text{Im}(\phi) = G$;

- if $\phi(h) = e_G$ and since identities are unique, $h = e_H \implies \text{Ker}(\phi) = \{e_H\}$.

On the other hand, suppose the contrary $\text{Im}(\phi) = G$ and $\text{Ker}(\phi) = \{e_H\}$. Then, every element of G is of the form $\phi(h_g)$ for some $h_g \in H$. We can define a function $\psi : G \to H$ by $\psi(g) := h_g$. By construction, this satisfies $(\phi \circ \psi)(g) = g$. Now

$$\phi(h^{-1} \cdot_H (\psi \circ \phi)(h)) = \phi(h)^{-1} \cdot_G (\phi \circ \psi \circ \phi)(h) = \phi(h)^{-1} \cdot_G \phi(h) = e_G$$

Hence, $h^{-1} \cdot_H (\psi \circ \phi)(h) \in \text{Ker}(\phi)$. This must be e_H, and so $(\psi \circ \phi)(h) = h$. Hence, $\psi = \phi^{-1}$, and thus ϕ is invertible.

For the last part, let $a, b \in G$ s.t. $a = \phi(\phi^{-1}(a))$ and $b = \phi(\phi^{-1}(b))$, then

$$\phi^{-1}(a \cdot_G b) = \phi^{-1}(\phi(\phi^{-1}(a))) \cdot_G \phi(\phi^{-1}(b)) = \phi^{-1}(\phi(\phi^{-1}(a) \cdot_H \phi^{-1}(b))) = \phi^{-1}(a) \cdot_H \phi^{-1}(b)$$

$\square$

Definition 15.16 (Cyclic): A group G is cyclic if $\exists \; g \in G$ s.t. every element of G has the form g^k for some $k \in \mathbb{Z}$. We call such a g a generator of G. We write this as $G = \langle g \rangle$.

Example 15.4. *Some examples of cyclic groups:*

- C_n *and* $\mathbb{Z}_n$ *are cyclic groups* $\forall n \in \mathbb{Z}^+$;

- $(\mathbb{Z}, +, 0)$ *is cyclic.*

Theorem 15.4: A cyclic group is isomorphic to some C_n or to $(\mathbb{Z}, +, 0)$, i.e. any two cyclic groups of the same order are isomorphic.

Proof. Let G be a cyclic group and $g \in G$ be a generator, and consider the set

$$S := \{k \in \mathbb{N}_{>0} | g^k = e\}$$

Consider the case where S is empty, then consider the map (which we will show that it is an isomorphism)

$$\phi : \mathbb{Z} \to G, \quad k \mapsto g^k$$

Then, $\phi(k + l) = g^{k+l} = g^k g^l = \phi(k)\phi(l)$ and thus ϕ is a homomorphism. ϕ must be surjective since g is a generator of G. If it is not injective, then $\text{Ker}(\phi) \leq \mathbb{Z}$, so is $\text{Ker}(\phi) \leq k\mathbb{Z}$ for some $k \in \mathbb{N}_{>0}$. But then, $e = \phi(k) = g^k$, and thus $k \in S$ (a contradiction). So, ϕ must be injective.

Consider the separate case where S is non-empty and n is the smallest number it contains

$$\phi : C_n = \{\xi^0, \xi^1, ..., \xi^{n-1}\} \to G, \quad \xi^k \mapsto g^k$$

Let $l < n$, then if say $k + l \geq n$, $\exists r$ s.t. $k + l = n + r$ with $0 \leq r < n$. Since $\xi^n = 1 \implies \xi^{k+1} = \xi^r$. If instead, $k + l < n$, then $\exists r = k + l$ s.t. $\xi^{k+l} = \xi^r$.

$\xi^{k+1} = \xi^r$ is true for both cases. If $g^k = e$, then $g^{k+l} = g^r$. Then, $\phi(\xi^{k+l}) = \phi(\xi^r) = g^r = g^{k+l} = \phi(\xi^k)\phi(\xi^l)$, hence ϕ is a homomorphism. Like before, ϕ must be surjective. If ϕ was not injective, then $\exists 0 < k < n$ with $\phi(\xi^k) = e \implies g^k = e$. Thus, $\exists k \leq n$ (since $g^n = e$) that is in S. (a contradiction). Again, ϕ must be injective.

For both cases, ϕ is an isomorphism. $\square$

Definition 15.17 (Order): If G is a group and $g \in G$, the order of g is the smallest $k \in \mathbb{N}$ s.t. $g^k = e$. We write this as $\text{ord}(g) = k$. If no such k exists, then $\text{ord}(g) = \infty$.

Corollary 15.5: $\langle g \rangle \simeq C_n$ with $\text{ord}(g) = n$.

Proof. Consider the generator $\langle g \rangle$, for $g^k, g^l \in \langle g \rangle \implies g^k(g^l)^{-1} = g^{l-l} \in \langle g \rangle \implies \langle g \rangle \leq G$. Then, $\langle g \rangle$ is cyclic by definition. Thus, we have $\langle g \rangle \simeq C_n$ and by definition, n is the order. $\square$

15.3   Conjugation classes and Cosets

Definition 15.18 (Cosets)**:** The coset gT is defined to be

$$gT = \{gt | t \in T\}$$

where $T \leq G$ and $g \in G$. gT is said to be a left coset of T in G, while Tg is said to be a right coset of T in G. We call g a representative of gT (and of Tg).

Lemma 15.3:

1. For $H \leq G$, $gH = Hg$ if G is abelian.

2. Since $H = e_G H = H e_G$, then H is both a left coset and a right coset (with e_G its representative).

3. $g \in gH$ and $g \in Hg$.

Theorem 15.5: Let G be a group and $H \leq G$ and $g_i \in G$. We say $g_1 \sim_H g_2$ if $g_1 = g_2 h$ for some $h \in H$. Then,

1. $\sim_H$ is an equivalence relation.

2. The equivalence class containing g is just gH.

3. G is a disjoint union of left cosets. There is $\{g_i\} \subseteq G$ such that G is a disjoint union of all $g_i H$ (decomposition of G into left cosets of H in G). Can define a similar thing for right cosets.

Corollary 15.6: For $H \leq G$ and $g \in G$, the following are equivalent:

1. $g \in H$

2. $gH = H$ and $Hg = H$

3. $gH \subseteq H$ and $Hg \subseteq H$

4. $Hg \supseteq H$ and $gH \supseteq H$.

Theorem 15.6:

1. Two cosets are either disjoint or identical.

2. Two cosets Hg_1 and Hg_2 are identical iff $g_1 g_2^{-1} \in H$.

3. Every element of G is in some coset.

Proof.

1. Suppose the cosets Hg_1 and Hg_2 have one element in common, then for some $h_1, h_2 \in H$, can write this as either $h_1 g_1$ or $h_2 g_2$. Then

$$Hg_1 = Hh_1^{-1} h_2 g_2 = Hg_2$$

since $h_1^{-1} h_2 \in H$. So if there is one element in common, the two cosets are identical.

2. If $g_1 g_2^{-1} \in H$, then let $g_1 g_2^{-1} = h$ for some $h \in H$. Hence, $Hg_1 = Hhg_2 = Hg_2$. Conversely, suppose $Hg_1 = Hg_2$, then $Hg_1 g_2^{-1} = H \implies hg_1 g_2^{-1} \in H \ \forall h \in H$.

3. Since H contains I, then for any $g \in G$, the coset Hg contains g.

$\square$

Remarks 15.1. *As a consequence of Theorem 15.6, the subgroup H and its left cosets partition G, i.e. divide G into disjoint subsets. Similar for right cosets. We can thus pick out any element of a set as its representative.*

Definition 15.19 (Conjugation)**:** $g_1, g_2 \in G$ are said to be conjugate each other if $\exists$ any group element $g \in G$ such that

$$g_2 = g g_1 g^{-1}$$

Definition 15.20 (Conjugacy classes)**:** The conjugacy class of $h \in G$ is written as $\operatorname{ccl}(h)$ and is

$$\operatorname{ccl}(h) := \{k \in G \text{ s.t. } k = g \cdot h \cdot g^{-1} \text{ for some } g \in G\}$$

We can always rearrange the group table to visually partition the group table into the distinct conjugacy classes in the group.

Remarks 15.2. *Again, conjugacy is an equivalence relation and we can write $g_1 \sim g_2$ (different equivalence relation from that of cosets). By a similar argument, the conjugacy relation partitions any group G into disjoint classes of elements, called conjugacy classes. We can also show each element must be in a unique conjugacy class.*

This partitioning consequence is a result of group actions (a group element $g \in G$ acting on a set X, the resultant products are called orbits). Orbits always partition X. A coset is an orbit formed from the regular action of H on G where $H \leq G$. A conjugacy class of $h \in G$ in G is the orbit formed from the conjugation action of G on G.

Corollary 15.7: Each element of an abelian group is in a class by itself. Trivially, the identity of any group is in a conjugacy class of its own.

Definition 15.21 (Conjugate of H by g)**:** If $H \leq G$ and $g \in G$, then the conjugate of H by g is

$$gHg^{-1} := \{k \in G \text{ s.t } k = g \cdot h \cdot g^{-1} \text{ for some } h \in H\}$$

Lemma 15.4: If $H \leq G$ and $g \in G$, then the conjugate $gHg^{-1} \leq G$.

Proof. We apply Proposition 15.1 to check that gHg^{-1} is non-empty. As $geg^{-1} = e$, we have $e \in gHg^{-1}$, so $gHg^{-1} \neq \phi$.

If $a, b \in gHg^{-1}$, then $a = ghg^{-1}$ and $b = gkg^{-1}$ for some $h, k \in H$. Then, to check subgroup:

$$a \cdot b^{-1} = (ghg^{-1}) \cdot (gkg^{-1})^{-1} = ghg^{-1}gk^{-1}g^{-1} = ghk^{-1}g^{-1} \in gHg^{-1}$$

since $hk^{-1} \in H$. $\qquad\square$

15.4   Normal subgroup and quotient group

Definition 15.22 (Normal subgroup)**:** A subgroup $H \leq G$ is called normal if for every $h \in H$ and $g \in G$, we have $ghg^{-1} \in H$. We write it as $H \lhd G$. Normal subgroup of G thus consists of complete conjugacy classes of G.

Example 15.5. *Some examples of normal subgroups:*

- *In any group G, the subgroups $\{e\}$ and G are normal.*

- *The subgroups $\langle r \rangle \leq D_{2n}$ is normal, but the subgroup $\langle s \rangle \leq D_{2n}$ is normal for $n \geq 3$.*

- *If G is abelian, then every subgroup is normal.*

Lemma 15.5: A subgroup $H \leq G$ is normal iff $Hg = gH \ \forall g \in G$.

Proof. Suppose H is normal and let $h \in H$. Then $ghg^{-1} \in H$, so $gh \in Hg$. This holds $\forall h$ so $gH \subset Hg$. Similarly, $Hg \subset gH$. Hence, $Hg = gH \ \forall g$.

Conversely, suppose $Hg = gH \ \forall g \in G$. If $h \in H$, then $gh \in gH = Hg$, so $gh = h'g$ for some $h' \in H$, so $ghg^{-1} = h' \in H$. This is true $\forall h \in H$, so H is normal. $\qquad\square$

Corollary 15.8: If $H \leq G$ has index 2, i.e. $|G/H| = 2$, then $H \lhd G$.

Proof. Since H has index 2, it has 2 left cosets in G. We write $G = H \cup gH$ for some $g \in G$. It also has 2 right cosets in G. We write $G = H \cup Hg'$ for some $g' \in G$. Now $g \notin H$, so $g \in Hg'$, so $Hg = Hg'$. But then Hg and gH are both equal to the complement of H in G, so they are equal to each other. $\qquad\square$

Proposition 15.5: If $\phi : G \to K$ is a group homomorphism, then $\operatorname{Ker} \phi \lhd G$.

Proof. Let $k \in \operatorname{Ker} \phi$ s.t. $\phi(k) = e$. If $g \in G$,

$$\phi(gkg^{-1}) = \phi(g)\phi(k)\phi(g^{-1}) = \phi(g)\phi(g^{-1}) = \phi(e) = e$$

hence, $gkg^{-1} \in \operatorname{Ker} \phi \implies \operatorname{Ker} \phi \lhd G$. $\qquad\square$

Theorem 15.7: If $H \lhd G$, then

$$(g_1 H) \cdot (g_2 H) := g_1 g_2 H$$

is a well-defined binary operation on the set G/H of left cosets, and $(G/H, \cdot, eH)$ is a group.

Proof. We first show it is well-defined. Let $g_1 H = g_1' H$, $g_2 H = g_2' H$, then $g_1' = g_1 h_1$ and $g_2' = g_2 h_2$, thus

$$g_1' g_2' H = g_1 h_1 g_2 h_2 H = g_1 h_1 g_2 H = g_1 g_2 (g_2^{-1} h_1 g_2) H$$

which is equal to $g_1 g_2 H$ iff $g_2^{-1} h_1 g_2 \in H$, i.e. $H \lhd G$. Next, check the group axioms. Associativity axiom:

$$(g_1 H \cdot g_2 H) \cdot g_3 H = (g_1 g_2 H) \cdot g_3 H = ((g_1 g_2) g_3) H = (g_1 (g_2 g_3)) H = g_1 H \cdot (g_2 H \cdot g_3 H)$$

Identity axiom: $gH \cdot eH = (g \cdot e)H = gH$, inverse axiom: $gH \cdot g^{-1}H = (g \cdot g^{-1})H = eH$ $\qquad\square$

Definition 15.23 (Quotient group)**:** If $H \lhd G$, then the quotient group is the set G/H of left cosets with the group structure described in Theorem 15.7.

15.5   Lagrange's Theorem

Theorem 15.8 (Lagrange's theorem)**:** If G is a finite group and $H \leq G$ is a subgroup, then $|G| = |H||G/H|$, i.e. $|H|$ divides $|G|$.

Proof. Since the left cosets $\{gH\}$ partition the set G, then observe that each gH is in bijection with H, i.e. can show

$$\phi_1 : H \mapsto gH, \quad \phi_2 : gH \mapsto H$$

to be inverses of each other. $\phi_1(h) = gh$, $\phi_2(gh) = g^{-1}gh = h$. Hence, each of these coset has $|H|$ elements. Thus, G is partitioned into $|G/H|$ number of sets, each having $|H|$ elements. $\qquad\square$

Definition 15.24 (Index of subgroup)**:** If G is a group and $H \leq G$ is a subgroup, then the index of H in G is $|G/H|$, as long as the set G/H is finite (and ∞ otherwise).

Corollary 15.9: If G is a finite group and $g \in G$, then $\mathrm{ord}(g)$ divides $|G|$.

Proof. We have $\mathrm{ord}(g) = |\langle g \rangle|$. Since $\langle g \rangle \leq G$, then by Lagrange's theorem, $|\langle g \rangle| = \mathrm{ord}(g)$ divides $|G|$. $\qquad\square$

Corollary 15.10: If G is a finite group and $g \in G$, then $g^{|G|} = e$.

Proof. By Corollary 15.8, we can write $|G| = \mathrm{ord}(g)N$ for some $N \in \mathbb{Z}$. Then $g^{|G|} = (g^{\mathrm{ord}(g)})^N$ but $g^{\mathrm{ord}(g)} = e$ by definition, and $e^N = e$. $\qquad\square$

Corollary 15.11: If G is a finite group with order $|G|$ equal to a prime number, then G is cyclic, and is generated by any non-identity element.

Proof. As $|G|$ is a prime number, say $p > 1$, so G contains some non-identity element, say $g \neq e$. Its order is then not 1, but divides p by Corollary 15.8. Thus, $\mathrm{ord}(g) = |G|$ by definition of prime numbers. The cyclic group $\langle g \rangle$ has $|G|$ elements and since $\langle g \rangle \leq G$, $\langle g \rangle = G$. $\qquad\square$

Theorem 15.9 (Cauchy's theorem)**:** Let G be a finite group and p be a prime number which divides $|G|$. Then G has an element of order p.

No proof. Just note the result.

15.6   Permutations

15.6.1   Cycles and Transpositions

Definition 15.25 (Cycles)**:** Given a list $a_1, a_2, ..., a_k \in \{1, 2, \ldots, n\}$ of distinct elements, the k-cycle $(a_1 a_2 \ldots a_k) \in S_n$ is the permutation given by

$$(a_1 a_2 \ldots a_k)(i) = \begin{cases} a_{j+1} & i = a_j \text{ for } j < k \\ a_1 & i = a_k \\ i & i \neq a_j \text{ for any } j \end{cases}$$

The k-cycle moves every element in the subset $\{a_1, a_2, \ldots, a_k\} \subset \{1, 2, \ldots, n\}$, and fixes every element outside of this subset. This is indeed an invertible function from the set $\{1, 2, \ldots, n\}$ to itself, as the k-cycle $(a_k a_{k-1} \ldots a_2 a_1)$ is an inverse for it.

Definition 15.26 (Transposition)**:** A 2-cycle is often called a transposition as (ab) is the permutation which transposes a and b, and leaves all other elements fixed.

Definition 15.27 (Disjoint cycles)**:** Cycles $(a_1 a_2 \ldots a_k)$ and $(b_1 b_2 \ldots b_l)$ are called disjoint if $a_i \neq b_j$ for any i and j.

Example 15.6. *Consider* $(1234)(324)$. *Then, this permutation* $1 \mapsto 2 \mapsto 1$ *and fixes 3 and 4, hence is the transposition* (12). *If instead, consider* $(324)(1234)$, *then* $1 \mapsto 4 \mapsto 1$ *and fixes 2 and 3, hence is the transposition* (14). *Clearly, the group* S_n *is not abelian for* $n \geq 3$.

Lemma 15.6:

1. Cycles can be cycled, i.e. $(a_1 a_2 \ldots a_k) = (a_k a_1 a_2 \ldots a_{k-1})$;

2. Disjoint cycles commute, i.e. $\sigma, \tau \in S_n$ are disjoint if $\sigma\tau = \tau\sigma$.

Proof. 1. We see that $(a_k a_1 a_2 \ldots a_{k-1})$

$$(a_k a_1 a_2 \ldots a_{k-1})(i) = \begin{cases} a_{j+1} & i = a_j \text{ for } 1 \leq j < k \\ a_1 & i = a_k \\ a_k & i = a_{k-1} \\ i & i \neq a_j \text{ for any } j \end{cases}$$

But the case for $i = a_k$ and $i = a_{k-1}$ is the same, so this is precisely the same as $(a_1 a_2 \ldots a_k)(i)$.

2. Let $\sigma = (a_1 a_2 \ldots a_k)$ and $\tau = (b_1 b_2 \ldots b_l)$ be disjoint cycles. For $i \in \{a_1, a_2, \ldots a_k\}$, we have τ fixing i:

$$\sigma\tau(i) = \sigma(i)$$

but τ fixes $\sigma(i)$ as well

$$\tau\sigma(i) = \sigma(i)$$

so $\sigma\tau(i) = \tau\sigma(i)$. If $i \notin \{a_1, a_2, \ldots a_k\}$, then $\tau(i) \notin \{a_1, a_2, \ldots, a_k\}$ too. Both are fixed by σ, so

$$\sigma\tau(i) = \tau(i), \quad \tau\sigma(i) = \tau(i)$$

hence, $\sigma\tau = \tau\sigma$.   □

Lemma 15.7: There is a minimal $k \in \mathbb{N}$ for $\sigma^k(1) = 1$, i.e. $\{1, \sigma(1), \sigma^2(1), \ldots, \sigma^{k-1}(1)\}$ are all distinct.

Proof. As the set $\{1, 2, \ldots, n\}$ is finite, we must eventually find $\sigma^a(1) = \sigma^b(1)$ for some $a \geq b$, but then $\sigma^{a-b}(1) = 1$.

Let k be this minimal value s.t. $\sigma^k(1) = 1$. If $0 \leq i < j < k$ are s.t. $\sigma^i(1) = \sigma^j(1)$, then $\sigma^{j-i}(1) = 1$, which contradicts k being minimal.   □

Theorem 15.10 (Disjoint cycle decomposition)**:** Every permutation in S_n is a composition of disjoint cycles. The expression of an element $\sigma \in S_n$ as a composition of disjoint cycles

$$\sigma = (a_1^1 a_2^1 \ldots a_{k_1}^1)(a_1^2 a_2^2 \ldots a_{k_2}^2) \ldots (a_1^r a_2^r \ldots a_{k_r}^r)$$

in which every element of $\{1, 2, \ldots, n\}$ appears, is unique up to

- cycling the terms in a cycle, and

- reordering the cycles.

Proof. We proceed by induction on n. If $n = 1$, the group S_1 is trivial. Let $\sigma \in S_n$ and consider the sequence $1, \sigma(1), \sigma^2(1), \ldots$ From Lemma 5.2, $\exists k \in \mathbb{N}$ s.t. $\sigma^k(1) = 1$. Then the permutation

$$\tau = \sigma \circ (1\sigma(1)\sigma^2(1)\ldots\sigma^{k-1}(1))^{-1}$$

fixes the set $\{1, \sigma(1), \sigma^2(1), \ldots, \sigma^{k-1}(1)\}$, so is a permutation of the set

$$X := \{1, 2, \ldots, n\}\backslash\{1, \sigma(1), \sigma^2(1), \ldots, \sigma^{k-1}(1)\}$$

This has strictly fewer than n elements, so by induction τ may be written as a composition of disjoint cycles in the set X. But then

$$\sigma = \tau \circ (1\sigma(1)\sigma^2(1)\ldots\sigma^{k-1}(1))$$

is a composition of disjoint cycles, as the sets X and $\{1, \sigma(1), \sigma^2(1), \ldots, \sigma^{k-1}(1)\}$ are disjoint. Thus, we may decompose σ as a composition of disjoint cycles.

Next, we will show uniqueness. Suppose there are two ways of disjoint cycle decomposition in which every element of $\{1, 2, \ldots, n\}$ appears:

$$(a_1^1 a_2^1 \ldots a_{k_1}^1)(a_1^2 a_2^2 \ldots a_{k_2}^2)\ldots(a_1^r a_2^r \ldots a_{k_r}^r) = (b_1^1 b_2^1 \ldots b_{l_1}^1)(b_1^2 b_2^2 \ldots b_{l_2}^2)\ldots(b_1^s b_2^s \ldots b_{l_s}^s)$$

We must have $a_1^1 = b_i^j$ for some i and j. Applying σ then shows $a_2^1 = b_{i+1}^j$, and so on, cycling the indices round if necessary. As we require precisely k_1 iterations of σ to bring a_1^1 back to itself, the same is true of b_i^j, and so $k_1 = l_j$. But then the sequence $(a_1^1, a_2^1, \ldots, a_{k_1}^1)$ and the sequence $(b_1^j, b_2^j, \ldots, b_{l_j}^j)$ are related by cycling ($j - 1$ times). We may then cancel the cycle $(a_1^1 a_2^1 \ldots a_{k_1}^1) = (b_1^j b_2^j \ldots b_{l_j}^j)$ from both sides, obtaining a simple problem. We repeat this for a finite number of times, and eventually the two collections of cycles are the same up to reordering and cycling. $\square$

Lemma 15.8: If $\sigma = (a_1^1 a_2^1 \ldots a_{k_1}^1)(a_1^2 a_2^2 \ldots a_{k_2}^2)\ldots(a_1^r a_2^r \ldots a_{k_r}^r)$ is a disjoint cycle decomposition, the order of σ is the lowest common multiple of the k_i's.

Proof. We have

$$\sigma^j = (a_1^1 a_2^1 \ldots a_{k_1}^1)^j(a_1^2 a_2^2 \ldots a_{k_2}^2)^j \ldots (a_1^r a_2^r \ldots a_{k_r}^r)^j$$

as disjoint cycles commute with each other. Since the order of a cycle $(b_1 b_2 \ldots b_r)$ is r, then when $l := \mathrm{lcm}\{k_i\}$, we have

$$(a_1^i a_2^i \ldots a_{k_i}^i)^l = ((a_1^i a_2^i \ldots a_{k_i}^i)^{k_i})^{l/k_i} = e^{l/k_i} = e$$

and so the order of σ divides l. On the other hand, if σ has order m, then rearranging, we have

$$(a_1^1 a_2^1 \ldots a_{k_1}^1)^m = ((a_1^2 a_2^2 \ldots a_{k_2}^2)^m \ldots (a_1^r a_2^r \ldots a_{k_r}^r)^m)^{-1}$$

but the LHS fixes all a_i^j with $j \neq 1$ and the RHS fixes all a_i^j with $j = 1$, so both sides must be the identity permutation. Thus, k_1 divides m. The same goes for all k_i, so the lcm of the k_i divides m. it follows that $m = l$. $\square$

15.6.2   Sign of permutation

Proposition 15.6: Every permutation in S_n is a composition of transpositions.

Proof. We prove by induction on n. S_1 is trivial and every permutation $\{1\}$ is the composition of no transpositions. For $\sigma \in S_n$, we have a transposition $(n\sigma(n))$. Now the permutation $(n\sigma(n)) \circ \sigma$ fixes n, so we may consider it as a permutation of $\{1, 2, \ldots, n-1\}$, i.e. an element of S_{n-1}. By induction $\tau = (n\sigma(n)) \circ \sigma$ is a composition of transpositions, so $\sigma = (\sigma(n)n) \circ \tau$ is too. $\square$

Definition 15.28 (Sign): We define the sign of a permutation σ as

$$\mathrm{sgn}(\sigma) := (-1)^n$$

where n is the number of transpositions needed to write σ as a composition of transpositions.

If $\mathrm{sgn}(\sigma) = +1$, it is an even permutation, otherwise, it is an odd permutation. Although $\sigma \in S_n$ cannot be written uniquely as a composition of transpositions, sgn is actually well-defined.

Theorem 15.11: The function $\mathrm{sgn} : S_n \to \{1, -1\}$ is well-defined.

Proof. By Theorem 5.1, elements of S_n can be written as a composition of disjoint cycles in which every element of $\{1, 2, \ldots, n\}$ appears, unique up to reordering the cycles and cycling the terms in a cycle. Let $l(\sigma)$ be the number of cycles when σ is written in this way. For an arbitrary transposition (cd), where wlog $c < d$, then consider the composition $\sigma(cd)$:

- Case 1: c and d lie in the same cycle of σ, i.e. $(ca_2a_3 \ldots a_{i-1}da_{i+1} \ldots a_{i+j})$, then

$$(ca_2a_3 \ldots a_{i-1}da_{i+1} \ldots a_{i+j})(cd) = (ca_{i=1} \ldots a_{i+j})(da_2a_3 \ldots a_{i-1})$$

  Then, $l(\sigma(cd)) = l(\sigma) + 1$;

- Case 2: c and d lie in different cycles of σ, i.e. $(ca_{i+1} \ldots a_{i+j})(da_2a_3 \ldots a_{i-1})$, then

$$(ca_{i+1} \ldots a_{i+j})(da_2a_3 \ldots a_{i-1})(cd) = (ca_2a_3 \ldots a_{i-1}da_{i+1} \ldots a_{i+j})$$

  Then, $l(\sigma(cd)) = l(\sigma) - 1$.

In either case, we see that $l(\sigma(cd)) \equiv l(\sigma) + 1 \mod 2$. If σ can be written as the composition of k transformations, then

$$l(\sigma) \equiv l(e) + k \mod 2$$

where $e = (1)(2) \ldots (n)$ in S_n and hence $l(e) = n$. In that case,

$$\mathrm{sgn}(\sigma) = (-1)^k = (-1)^{n+l(\sigma)}$$

Since the RHS only depends on the permutation σ, and not on how we choose to write it as a composition of transpositions, then sgn is well-defined. $\square$

Corollary 15.12: The function $\mathrm{sgn} : S_n \to \{1, -1\} = C_2$ is a group homomorphism.

Proof. Take $\sigma, \tau \in S_n$ s.t. they consist of a composition of a and b transpositions respectively, then $\sigma\tau$ is a composition of $a + b$ transpositions, then

$$\mathrm{sgn}(\sigma\tau) = (-1)^{a+b} = (-1)^a(-1)^b = \mathrm{sgn}(\sigma)\,\mathrm{sgn}(\tau)$$

which satisfies the defining property of a group homomorphism. $\square$

Definition 15.29: The alternating group $A_n \leq S_n$ consists of only even permutations.

Corollary 15.13: $A_n = \mathrm{Ker}(\mathrm{sgn})$ and $A_n \lhd S_n$.

Proof. The identity of C_2 is $+1$, so $A_n = \mathrm{Ker}(\mathrm{sgn})$. By Proposition 4.1, since sgn is a group homomorphism (Corollary 5.1), then $A_n \lhd S_n$. $\square$

Example 15.7. A_n has index 2 in S_n as long as $n \geq 2$.

Lemma 15.9:

$$\mathrm{sgn}((a_1a_2 \ldots a_r)) = (-1)^{r-1}$$

Proof. We have $(a_1a_2 \ldots a_r) = (a_1a_r)(a_1a_{r-1}) \ldots (a_1a_3)(a_1a_2)$ to be a composition of $r-1$ transpositions. $\square$

15.6.3   Conjugation in permutation group

Proposition 15.7: If $\sigma \in S_n$, then $\sigma(a_1 a_2 \dots a_k)\sigma^{-1} = (\sigma(a_1)\sigma(a_2)\dots\sigma(a_k))$.

Proof. We need to check separately for the cases where $\sigma^{-1}(i)$ is in and outside of the cycle. If $\sigma^{-1}(i) \notin \{a_1, a_2, \dots a_k\}$, then it is a fixed point to the cycle $(a_1 \dots a_k)$, i.e.

$$(a_1 a_2 \dots a_k)(\sigma^{-1}(i)) = \sigma^{-1}(i)$$

Applying σ will give back i. Since $i \notin \{\sigma(a_1), \sigma(a_2), \dots, \sigma(a_k)\}$, we obviously have i to be a fixed point of $(\sigma(a_1)\sigma(a_2)\dots\sigma(a_k))$.

On the other hand, if $\sigma^{-1}(i) = a_j$, then

$$(\sigma(a_1 a_2 \dots a_k)\sigma^{-1})(i) = (\sigma(a_1 a_2 \dots a_k))(\sigma^{-1}(i)) = (\sigma(a_1 a_2 \dots a_k))(a_j) = \sigma(a_{j+1})$$

but this is also the result of applying $(\sigma(a_1)\sigma(a_2)\dots\sigma(a_k))$ to $i = \sigma(a_j)$. $\qquad\square$

Corollary 15.14: Elements $\tau, \tau' \in S_n$ are conjugate iff when they are written as a composition of disjoint cycles (in which every element of $\{1, 2, \dots, n\}$ appears) they have the same number of cycles of each length.

Proof. Let the disjoint cycle decomposition of $\tau \in S_n$ be

$$\tau = (a_1^1 a_2^1 \dots a_{k_1}^1)(a_1^2 a_2^2 \dots a_{k_2}^2) \dots (a_1^r a_2^r \dots a_{k_r}^r)$$

Then, we see that its conjugate

$$\sigma\tau\sigma^{-1} = (\sigma(a_1^1)\sigma(a_2^1)\dots\sigma(a_{k_1}^1))(\sigma(a_1^2)\sigma(a_2^2)\dots\sigma(a_{k_2}^2))\dots(\sigma(a_1^r)\sigma(a_2^r)\dots\sigma(a_{k_r}^r))$$

which is also a disjoint cycle decomposition. Thus, τ and $\sigma\tau\sigma^{-1}$ have the same number of cycles of each length.

Conversely, if $\tau, \tau' \in S_n$ have the same number of cycles of each length, then after reordering their cycles, we may find disjoint cycle decompositions

$$\tau = (a_1^1 a_2^1 \dots a_{k_1}^1)(a_1^2 a_2^2 \dots a_{k_2}^2) \dots (a_1^r a_2^r \dots a_{k_r}^r)$$

$$\tau' = (b_1^1 b_2^1 \dots b_{k_1}^1)(b_1^2 b_2^2 \dots b_{k_2}^2) \dots (b_1^r b_2^r \dots b_{k_r}^r)$$

in which every element of $\{1, 2, \dots, n\}$ appears. We may then define a mapping

$$\begin{aligned} \sigma : \{1, 2, \dots, n\} &\rightarrow \{1, 2, \dots, n\} \\ a_j^i &\mapsto b_j^i \end{aligned}$$

This is indeed a permutation by construction. Clearly, $\tau' = \sigma\tau\sigma^{-1}$, as required. $\qquad\square$

Definition 15.30 (Cycle type)**:** For a permutation $\sigma \in S_n$, we may record its conjugacy class as $1^{a_1} 2^{a_2} 3^{a_3} \dots n^{a_n}$, where a_i denotes the number of i-cycles when σ is written as a composition of disjoint cycles in which every element of $\{1, 2, \dots, n\}$ appears.

Corollary 15.15: If $\tau \in S_n$ has a cycle type $1^{a_1} 2^{a_2} 3^{a_3} \dots n^{a_n}$, then the size of its conjugacy class is

$$|\operatorname{ccl}(\tau)| = \frac{n!}{1^{a_1}(a_1!)2^{a_2}(a_2!)3^{a_3}(a_3!)\dots n^{a_n}(a_n!)}$$

Example 15.8 (Conjugacy classes in S_4)**.** *The conjugacy classes in S_4 have the cycle type $1^{a_1} 2^{a_2} 3^{a_3} 4^{a_4}$ with $\sum_{i=1}^{4} ia_i = 4$, i.e. $\{1^4, 1^2 2^1, 2^2, 1^1 3^1, 4^1\}$.*

16   Representation theory

Representation of groups; reducible and irreducible representations; basic theorems of representation theory. Classes, characters. Examples of character tables of point groups. Applications in molecular physics.

[3]

16.1   Definitions and examples

Notation 16.1 (General linear group)**.** *We write V for a finite-dimensional vector space over the field $\mathbb{F}$. We write $\mathrm{GL}(V)$ for the group of invertible linear maps $\theta : V \mapsto V$. This is a group with the operation given by composition of maps, with the identity as the identity map (and inverse by inverse). As groups, $\mathrm{GL}(V) \simeq \mathrm{GL}_n(\mathbb{F})$ ($n \times n$ invertible square matrices over the field $\mathbb{F}$).*

Definition 16.1 (Representation)**:** Let V be a finite-dimensional vector space over $\mathbb{F}$. A linear representation of G on V is a group homomorphism

$$D = D_V : \ G \to \mathrm{GL}(V)$$

We sometimes write D_g for $D_V(g)$, so for each $g \in G$, $D_g \in \mathrm{GL}(V)$, and $D_g D_h = D_{gh}$ and $D_{g^{-1}} = (D_g)^{-1}$ for all $g, h \in G$.

Definition 16.2 (Dimension of representation)**:** The dimension of a representation $D : \ G \to \mathrm{GL}(V)$ is $\dim_{\mathbb{F}}(V)$.

Definition 16.3 (Faithful representation)**:** A faithful representation is a representation D such that $\mathrm{Ker}\, D = 1$, and is thus an isomorphism.

Definition 16.4 (Regular representation)**:** A regular representation of a group G is a faithful representation of dimension $|G|$.

Definition 16.5 (Linear action)**:** A group G acts linearly on a vector space V if it acts on V such that
$$g(\mathbf{v}_1 + \mathbf{v}_2) = g\mathbf{v}_1 + g\mathbf{v}_2, \quad g(\lambda\mathbf{v}_1) = \lambda(g\mathbf{v}_1)$$

for all $g \in G$, $\mathbf{v}_1, \mathbf{v}_2 \in V$ and $\lambda \in \mathbb{F}$. We call this a linear action.

Definition 16.6 (Matrix representation)**:** R is a *matrix representation* of G of degree n if R is a homomorphism $G \to \mathrm{GL}_n(\mathbb{F})$.

Remarks 16.1 (Equivalent definitions of representations)**.** *Now if g acts linearly on V, the map $G \to \mathrm{GL}(V)$ defined by $g \mapsto D_g$, with $D_g : \ \mathbf{v} \mapsto g\mathbf{v}$, is a representation in the previous sense. Conversely, given a representation $D : \ G \to \mathrm{GL}(V)$, we have a linear action of G on V via $g\mathbf{v} = D(g)\mathbf{v}$. In other words, a representation is just a linear action.*

We can view this as a representation that acts on $\mathbb{F}^n$. Since all finite-dimensional vector spaces are isomorphic to $\mathbb{F}^n$ for some n, every representation is equivalent to some matrix representation.

In particular, given a linear representation $D : \ G \to \mathrm{GL}(V)$ with $\dim V = n$, we can get a matrix representation by fixing a basis $\mathcal{B}$, and then define the matrix representation $G \to \mathrm{GL}_n(\mathbb{F})$ by $g \mapsto [\rho(g)]_{\mathcal{B}}$. Conversely, given a matrix representation R, we get a linear representation D in the obvious way, i.e. $D : \ G \to \mathrm{GL}(\mathbb{F}^n)$ by $g \mapsto D_g$ via $D_g(\mathbf{v}) = R_g\mathbf{v}$.

Example 16.1 (Trivial representation)**.** *Given any group G, take $V = \mathbb{F}$ (the one-dimensional space), and $D : G \to \mathrm{GL}(V)$ by $g \mapsto (\mathrm{Id} : \mathbb{F} \to \mathbb{F}) \ \forall g$. This is the trivial representation of G, and has degree 1.*

Example 16.2. *Let $G = C_4 = \langle x : x^4 = 1 \rangle$. Let $n = 2$, and work over $\mathbb{F} = \mathbb{C}$. Then we can define a representation by picking a matrix A, and then define $R : \ x \mapsto A$. Then the action of other elements follows directly by $x^j \mapsto A^j$. Of course, we cannot choose A arbitrarily. We need to have $A^4 = I_2$, and this is easily seen to be the only restriction. So we have the following possibilities:*

1. *A is diagonal: the diagonal entries can be chosen freely from $\{\pm 1, \pm i\}$. Since there are two diagonal entries, we have 16 choices.*

2. *A is not diagonal: then it will be equivalent to a diagonal matrix since $A^4 = I_2$. So we don't really get anything new.*

Definition 16.7 (Equivalent/isomorphic representations)**:** Two representations D, D' are equivalent if in terms of matrix representations, if there exists some non-singular matrix $X \in \mathrm{GL}_n(\mathbb{F})$ such that the representations $R: G \to \mathrm{GL}_n(\mathbb{F})$ and $R': G \to \mathrm{GL}_n(\mathbb{F})$ satisfy

$$R'(g) = XR(g)X^{-1} \ \forall g$$

Definition 16.8 (Irreducible representation)**:** A representation $D: G \to \mathrm{GL}(V)$ is irreducible if there are no proper non-zero subspaces $W \leq V$ that is $D(G)$-invariant, i.e. $D_g(W) \leq W \ \forall g \in G$.

Definition 16.9 (Decomposable representation)**:** A representation $D: G \to \mathrm{GL}(V)$ is decomposable if there are proper $D(G)$-invariant subspaces $U, W \leq V$ with

$$V = U \oplus W.$$

We say D is a direct sum $D_u \oplus D_w$. If no such decomposition exists, we say that D is indecomposable.

Definition 16.10 (Completely reducible representation)**:** A representation $D: G \to \mathrm{GL}(V)$ is completely reducible if it is the direct sum of irreducible representations.

Definition 16.11 (G-invariant inner product)**:** The Hermitian inner product $\langle \, , \, \rangle$ is G-invariant if

$$[D(h)x, D(h)y] = [x, y]$$

for any fixed $h \in G$, where $[x, y]$ is defined to be $\sum_{g \in G} \langle D(g)x, D(g)y \rangle$.

Definition 16.12 (Character)**:** The character $\chi: G \to \mathbb{C}$ of a representation $D: G \to \mathrm{GL}(V)$, written $\chi_D = \chi_v = \chi$, is defined as

$$\chi(g) = \mathrm{Tr}\, D(g)$$

Alternatively, the character is $\mathrm{Tr}\, R(g)$, where $R(g)$ is any matrix representing $D(g)$ with respect to any basis. We usually write it as a set of traces for a representation D.

Definition 16.13 (Degree of character)**:** The degree of χ_v is $\dim V$.

Definition 16.14 (Character table)**:** The character table of G is the $k \times k$ matrix $X = (\chi_i(g_j))$, where $1 = \chi_1, \chi_2, \cdots, \chi_k$ are the irreducible characters of G.

16.2   Quick theorems with no proofs

Proposition 16.1: Let W be G-invariant subspace of V, and V have a G-invariant inner product. Then $W^\perp$ is also G-invariant.

Theorem 16.1 (Weyl's unitary trick)**:** Let D be a complex representation of a finite group G on the complex vector space V. Then there is a G-invariant Hermitian inner product on V.

Corollary 16.1: Every finite subgroup of $\mathrm{GL}_n(\mathbb{C})$ is conjugate to a subgroup of $\mathrm{U}(n)$, where the unitary group is defined by

$$
\begin{aligned}
\mathrm{U}(V) &= \{ f \in \mathrm{GL}(V) : \langle f(\mathbf{u}), f(\mathbf{v}) \rangle = \langle \mathbf{u}, \mathbf{v} \rangle \text{ for all } \mathbf{u}, \mathbf{v} \in V \} \\
&= \{ A \in \mathrm{GL}_n(\mathbb{C}) : AA^\dagger = I \} \\
&= \mathrm{U}(n).
\end{aligned}
\tag{3}
$$

Theorem 16.2 (Schur's lemma)**:** Given any finite group G, consider two of its inequivalent irreducible representations with underlying vector spaces V and W over a field $\mathbb{F}$..

1. Any G-homomorphism (preserves all vector space operations and group operations) $\theta : V \to W$ is either zero or an isomorphism.

2. Any mapping $V \to V$ is a scalar multiple of the identity map.

Corollary 16.2: The irreducible complex representations of a finite abelian group G are all one-dimensional.

Theorem 16.3:

1. $\chi_V(1) = \dim V$.

2. χ_V is a class function, namely it is conjugation invariant, i.e. class function

$$
\chi_V(hgh^{-1}) = \chi_V(g), \ \forall g, h \in G
$$

   Thus χ_V is constant on conjugacy classes.

3. $\chi_V(g^{-1}) = \overline{\chi_V(g)}$.

4. For two representations V, W, we have

$$
\chi_{V \oplus W} = \chi_V + \chi_W
$$

Theorem 16.4 (Completeness of characters)**:** The complex irreducible characters of G form an orthonormal basis of the conjugacy class of G, namely

1. If $D : G \to \mathrm{GL}(V)$ and $D' : G \to \mathrm{GL}(V')$ are two complex irreducible representations affording characters χ, χ' respectively, then

$$
\langle \chi, \chi' \rangle =
\begin{cases}
1 & \text{if } \rho \text{ and } \rho' \text{ are isomorphic representations} \\
0 & \text{otherwise}
\end{cases}
$$

   This is the (row) orthogonality of characters. Can also define column orthogonality of characters.

2. Each class function of G can be expressed as a linear combination of irreducible characters of G.

Theorem 16.5: Let $D_1, \ldots, D_k$ be the irreducible complex representations of G, and let their dimensions be $n_1, \ldots, n_k$. Then the order of the group is the sum of the dimensions of the irreducible representations squared.

$$
|G| = \sum n_i^2
$$

Corollary 16.3: The number of irreducible characters of G (up to equivalence) is k, the number of conjugacy classes.

Theorem 16.6: Let G be a given finite group with $|G|$ elements. Let $D^{(1)}, \ldots, D^{(N)}$ be the inequivalent irreducible representations of G, with dimensions $n_1, \ldots, n_N$ respectively. Then for any two fo these, the matrix elements satisfy

$$\sum_g D^{(\alpha)}_{ij}(g) D^{(\beta)}_{kl}(g^{-1}) = \frac{|G|}{n_\alpha} \delta_{\alpha\beta} \delta_{il} \delta_{jk}$$

16.3   Examples

Example 16.3.

1. *Consider $G = C_4 = \langle x \rangle$. The four 1-dimensional irreducible representations are given by*

	1	x	x^2	x^3
D_1	1	1	1	1
D_2	1	i	-1	$-i$
D_2	1	-1	1	-1
D_2	1	$-i$	-1	i

2. *Consider the Klein four group $G = V_R = \langle x_1 \rangle \times \langle x_2 \rangle$. The irreducible representations are*

	1	x_1	x_2	$x_1 x_2$
D_1	1	1	1	1
D_2	1	1	-1	-1
D_2	1	-1	1	-1
D_2	1	-1	-1	1

Example 16.4. *Consider symmetries of an equilateral triangle (which is isomorphic to S_3). This is $\langle r, s : r^3 = s^2 = 1, rs = sr^{-1} \rangle$. We have previously seen that each irreducible representation has dimension at most 2. We spot at least three irreducible representations:*

$$\begin{array}{lll} 1 & triad & r \mapsto 1 \quad s \mapsto 1 \\ S & sign & r \mapsto 1 \quad s \mapsto -1 \\ W & \text{2-dimensional} \end{array}$$

The last representation is the action of the group on $\mathbb{R}^2$ in the natural way, i.e. the rotations and reflections of the plane that corresponds to the symmetries of the triangle. It is helpful to view this as a complex representation in order to make the matrix look nice. The 2-dimensional representation (D, W) is defined by $W = \mathbb{C}^2$, where r and s act on W as

$$D(r) = \begin{pmatrix} \omega & 0 \\ 0 & \omega^2 \end{pmatrix}, \quad \rho(s) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

and $\omega = e^{2\pi i/3}$ is a third root of unity. We will now show that these are indeed all the irreducible representations, by decomposing any representation into sum of these.

So let's decompose an arbitrary representation. Let (D', V) be any complex representation of G. Since $D'(r)$ has order 3, it is diagonalizable has eigenvalues $1, \omega, \omega^2$. We diagonalize $D'(r)$ and then V splits as a vector space into the eigenspaces

$$V = V(1) \oplus V(\omega) \oplus V(\omega^2)$$

Since $srs^{-1} = r^{-1}$, we know $\rho'(s)$ preserves $V(1)$ and interchanges $V(\omega)$ and $V(\omega^2)$.

Now we decompose $V(1)$ into $D'(s)$ eigenspaces, with eigenvalues ± 1. Since r has to act trivially on these eigenspaces, $V(1)$ splits into sums of copies of the irreducible representations 1 and S.

For the remaining mess, choose a basis $\mathbf{v}_1, \cdots, \mathbf{v}_n$ of $V(\omega)$, and let $\mathbf{v}'_j = \rho'(s)\mathbf{v}_j$. Then $D'(s)$ acts on the two-dimensional space $\langle \mathbf{v}_j, \mathbf{v}'_j \rangle$ as $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, while $D'(r)$ acts as $\begin{pmatrix} \omega & 0 \\ 0 & \omega^2 \end{pmatrix}$. This means $V(\omega) \oplus V(\omega^2)$ decomposes into many copies of W.

Example 16.5. *Consider the conjugacy classes in Example 16.4,*

$$\{1\}, \quad \{s, sr, sr^2\}, \quad \{r, r^{-1}\}$$

We have found the three representations: 1, the trivial representation; S, the sign; and W, the two-dimensional representation. In W, the reflections sr^j acts by matrix with eigenvalues ± 1. So the sum of eigenvalues is 0, hence $\chi_w(sr^2) = 0$. It also also not hard to see

$$r^m \mapsto \begin{pmatrix} \cos\frac{2m\pi}{3} & -\sin\frac{2m\pi}{3} \\ \sin\frac{2m\pi}{3} & \cos\frac{2m\pi}{3} \end{pmatrix}$$

So $\chi_w(r^m) = -1$. After finding all the irreducible representations, the character table is

	1	$\mathcal{C}_2$	$\mathcal{C}_3$
1	1	1	1
S	1	−1	1
χ_w	2	0	−1

We see that the sum of the squares of the first column is $1^2 + 1^2 + 2^2 = 6 = |D_6|$, as expected. The rows in the character table are indeed mutually orthogonal.

16.4   Applications to normal modes

The study of representation theory is useful in simplifying the analysis of normal modes problems. For instance, to figure out the degeneracy of a normal mode. The general procedure is

1. Construct the symmetry group elements of the problem.

2. Take the action of this group to work out the matrix representation of the symmetry.

3. Use theorem 16.5 and Corollary 16.3 to figure out the number of irreducible representations and their dimensions.

4. Use orthogonality theorem to figure out the multiplicity.