Mathematical Methods

University of Cambridge Part IB Natural Sciences Tripos

Yue Wu

Yusuf Hamied Department of Chemistry
Lensfield Road,
Cambridge, CB2 1EW

yw628@cam.ac.uk

Acknowledgements

The structure and content of these notes are based on the following notes:

- *Mathematical Methods I & II* for Natural Sciences Tripos Part IB, Dr. Stephen Cowley;
- *Mathematical Methods III* for Natural Sciences Tripos Part IB, Prof. Matthew Wingate.

These notes also draw heavily on the following materials:

- *Vector Calculus* for Mathematical Tripos Part IA, Prof. David Tong;
- *Linear Algebra* for Mathematical Tripos Part IB, Prof. Simon Wadsley;
- *Analysis* for Mathematical Tripos Part IA, Prof. Timothy Gowers, notes taken by Dexter Chua;
- *Analysis and Topology* for Mathematical Tripos Part IB, Dr. Andras Zsak, notes taken by Zhiyuan Bai;
- *Methods* for Mathematical Tripos Part IB, Dr. Anthony Ashton;
- *Methods* for Mathematical Tripos Part IB, Prof. David Skinner;
- *Complex Methods* for Mathematical Tripos Part IB, Prof. Anthony Scholl;
- *Complex Analysis* for Mathematical Tripos Part IB, Prof. Ulrich Sperhake;
- *Variational Principles* for Mathematical Tripos Part IB, Prof. Paul Townsend;
- *Groups* for Mathematical Tripos Part IA, Prof. Oscar Randal-Williams;
- *Representation Theory* for Mathematical Tripos Part IID, Prof. Simon Wadsley.

These notes are not endorsed by the lecturers, and I have modified them (often significantly) after lectures. They are nowhere near accurate representations of what was actually lectured, and in particular, all errors are almost surely mine.

Preface

These notes were originally written as a personal source of revision, so they may not always be the most suitable resource if you are studying the material for the very first time. I have not always explained the motivation behind every step in detail, though I have added extra explanations while preparing these notes for publication. You will also find some topics that go beyond the syllabus, included because they are interesting and helpful for gaining a more systematic understanding of mathematics.

A final remark: these are the first formal lecture notes I have written. At the time of writing, I was still developing both my writing style and my LaTeX (especially TikZ) skills. Although I have revised the notes multiple times to improve them, they remain far from perfect. I apologise if you find any sections difficult to follow, and I appreciate your understanding.

Contents

0 Assumed Knowledge

0.1 Numbers and Sets

Definition 0.1. A *set* is a collection of objects without regard to order. We write $x \in X$ if x is an element of X.

Definition 0.2. Two sets A and B are *equal*, written as $A = B$, if for all x,

$$x \in A \iff x \in B\,.$$

Definition 0.3. A is a *subset* of B, written as $A \subseteq B$, if all elements in A are in B.

Definition 0.4. Given two sets A and B, we define the following:

- *Intersection.* $A \cap B \coloneqq \{x \mid x \in A \text{ and } x \in B\}$.
- *Union.* $A \cup B \coloneqq \{x \mid x \in A \text{ or } x \in B\}$.
- *Set difference.* $A \setminus B \coloneqq \{x \mid x \in A, x \notin B\}$.
- *Power set.* $\mathcal{P}(A) = \{X \mid X \subseteq A\}$.

Definition 0.5. Given two sets A, B, the *Cartesian product* of A and B is $A \times B = \{(a, b) : a \in A, b \in B\}$, where (a, b) is an *ordered pair* of two items in which order matters.

Definition 0.6. A *binary operation* $\cdot$ on a set X is a function $\cdot : X \times X \to X$.

Definition 0.7. A set F together with two binary operations, say addition and multiplication, is a *field* if elements in F satisfy the axioms:

(F1) *Closure.* $\forall a, b \in F$, $a + b$, $ab \in F$;

(F2) *Associativity.* $\forall a, b, c \in F$, $a + (b + c) = (a + b) + c$, $a(bc) = (ab)c$;

(F3) *Commutativity.* $\forall a, b \in F$, $a + b = b + a$, $ab = ba$;

(F4) *Distributivity.* $\forall a, b, c \in F$, $a(b + c) = ab + ac$;

(F5) *Additive identity.* $\exists e \in F$ such that $\forall a \in F$, $e + a = a$;

(F6) *Multiplicative identity.* $\exists u \in F$, $u \neq e$ such that $\forall a \in F$, $ua = a$;

(F7) *Additive inverse.* $\forall a \in F$, $\exists b \in F$ such that $a + b = e$;

(F8) *Multiplicative inverse.* $\forall a \in F$, $a \neq e$, $\exists b \in F$ such that $ab = u$.

Remark. When mentioning *fields* or *number fields* in these lecture notes, it is always fine to think of real numbers $\mathbb{R}$ or complex numbers $\mathbb{C}$. It is easy to check that they satisfy the field axioms, with additive identity 0 and multiplicative identity 1.

0.2 Functions

Definition 0.8. A *function*, or a *mapping*

$$f : X \to Y$$

is a rule to associate a single element $f(x)$ in set Y to every element x in set X. X is the *domain* and Y is the *codomain*. If the element $x_i \in X$ *maps to* $y_i \in Y$, we write $x_i \mapsto y_i$.

Definition 0.9. A function $f : X \to Y$ is *injective* if $f(x_1) = f(x_2)$ implies $x_1 = x_2$. f maps distinct elements to distinct elements.

Definition 0.10. A function $f : X \to Y$ is *surjective* if for every $y \in Y$, there is at least one $x \in X$ such that $f(x) = y$.

Definition 0.11. A function is *bijective* if it is both surjective and injective.

Definition 0.12. The *composition of two functions* is a function obtained by applying one after another. In particular, if $f : X \to Y$ and $g : Y \to Z$, then $g \circ f : X \to Z$ is defined by $g \circ f(x) = g(f(x))$.

Definition 0.13. An *inverse* of $f : X \to Y$ is a function $f^{-1} : Y \to X$ such that

$$f^{-1}(f(x)) = x \text{ and } f(f^{-1}(y)) = y \quad \forall\, x \in X\,, y \in Y\,.$$

Theorem 0.14. The inverse of f exists if and only if it is bijective, and the inverse of a function is necessarily unique.

Definition 0.15. The set A is *finite* if there exists a bijection $f : A \to \{1, 2, \ldots, n\}$ for some $n \in \mathbb{N}_0$. The *cardinality* or *size* of A, written as $|A|$, is n.

Definition 0.16. A set A is *countable* if A is finite or there is a bijection between A and $\mathbb{N}$. A set A is *uncountable* if A is not countable.

0.3 Calculus

Theorem 0.17 (The first fundamental theorem of calculus). The derivative of the integral of f is f if f if continuous.

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\int_a^x f(t)\,\mathrm{d}t\right) = f(x)\,.$$

Theorem 0.18 (The second fundamental theorem of calculus). The integral of the derivative of f is f if f is differentiable and f' is integrable.

$$\int_a^b \frac{\mathrm{d}f}{\mathrm{d}x}\,\mathrm{d}x = f(b) - f(a)\,.$$

Theorem 0.19 (Taylor's theorem). Let $f(x)$ be a function of a real variable x, which is differentiable at least n times in the interval $x_0 \le x \le x_0 + h$. Then the Taylor series of $f(x_0 + h)$ is given by

$$f(x_0 + h) = f(x_0) + hf'(x_0) + \frac{h^2}{2!}f''(x_0) + \cdots + \frac{h^{n-1}}{(n-1)!}f^{(n-1)}(x_0) + R_n\,,$$

where the remainder after n terms, R_n, is obtained by integration by parts to be

$$R_n = \int_{x_0}^{x_0+h} \frac{(x_0 + h - x)^{n-1}}{(n-1)!} f^{(n)}(x)\,\mathrm{d}x\ .$$

Corollary (Taylor's theorem of multiple variables). Let $f(x, y)$ be a function of two variables, then

$$f(x + \delta x, y + \delta y) = f(x, y) + \delta x\frac{\partial f}{\partial x} + \delta y\frac{\partial f}{\partial y} + \frac{1}{2!}\left((\delta x)^2\frac{\partial^2 f}{\partial x^2} + 2\delta x\delta y\frac{\partial^2 f}{\partial x\partial y} + (\delta y)^2\frac{\partial^2 f}{\partial y^2}\right) + \ldots$$

Theorem 0.20 (Chain rule). Suppose f is a function of n variables x_i, $f = f(x_1, x_2, \ldots, x_n)$, and x_i depends on m variables s_j, then

$$\frac{\partial f}{\partial s_j} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\frac{\partial x_i}{\partial s_j}\,.$$

0.4 Vectors

For vectors $\mathbf{a} = (a_1, a_2, a_3)$, $\mathbf{b} = (b_1, b_2, b_3)$ and $\mathbf{c} = (c_1, c_2, c_3) \in \mathbb{R}^3$.

Definition 0.21. The *scalar product* of $\mathbf{a}$ and $\mathbf{b}$ is

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

Definition 0.22. The *vector (cross) product* of $\mathbf{a}$ and $\mathbf{b}$ is

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} .$$

Theorem 0.23 (Scalar triple product).

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} .$$

Theorem 0.24 (Vector triple product).

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c} \,.$$

0.5 Vector Calculus

0.5.1 Line Integral

Consider a curve characterised by a map

$$f : \mathbb{R} \to \mathbb{R}^n \,.$$

Suppose the curve is parameterised by t, then the derivative is

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \lim_{\delta t \to 0} \frac{\delta \mathbf{x}}{\delta t} \,.$$

Theorem 0.25. The arc length is

$$s = \int_{t_0}^{t} \left| \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t'} \right| \mathrm{d}t' \,.$$

Definition 0.26. A *scalar field* $\phi(\mathbf{x})$ is a map

$$\phi : \mathbb{R}^n \to \mathbb{R} \,.$$

Definition 0.27. A *vector field* $\mathbf{F}(\mathbf{x})$ is a map

$$\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^n \,.$$

Theorem 0.28. The line integral of a scalar field $\phi(\mathbf{x})$ and a vector field $\mathbf{F}(\mathbf{x})$ along a curve γ parameterised by t in the interval $[t_a, t_b]$ is

$$\int_\gamma \phi \, \mathrm{d}s = \int_{t_a}^{t_b} \phi(\mathbf{x}(t)) \left| \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} \right| \mathrm{d}t \,,$$

$$\int_\gamma \mathbf{F}(\mathbf{x}) \cdot \mathrm{d}\mathbf{x} = \int_{t_a}^{t_b} \mathbf{F}(\mathbf{x}(t)) \cdot \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} \, \mathrm{d}t \,.$$

0.5.2 Multiple Integral

Consider a region $\Omega \subset \mathbb{R}^n$ and a scalar function $\phi : \mathbb{R}^n \to \mathbb{R}$.

Definition 0.29. The multiple integral of a function ϕ over a region is defined by the limit

$$\int_\Omega \phi(\mathbf{x})\,\mathrm{d}\mathbf{x} \coloneqq \lim_{\text{all } \delta V_i \to 0} \sum_i \phi(\mathbf{x}_i)\delta V_i\,,$$

where $\{\delta V_i\}$ partitions Ω.

Theorem 0.30. If a simple region $\Omega \subset \mathbb{R}^2$ has x in range $[a, b]$, and the upper and lower boundary of Ω are given by $y_2(x)$ and $y_1(x)$, then

$$\iint_\Omega \phi(x, y)\,\mathrm{d}A = \int_a^b \int_{y_1(x)}^{y_2(x)} \phi(x, y)\,\mathrm{d}y\,\mathrm{d}x\ .$$

Corollary. This can be generalised to n-dimensional integration:

$$\int_\Omega \phi(\mathbf{x})\,\mathrm{d}^n\mathbf{x} = \int_a^b \mathrm{d}x^1 \int_{x_1^2(x^1)}^{x_2^2(x^1)} \mathrm{d}x^2 \dots \int_{x_1^n(x^1,x^2,\dots,x^{n-1})}^{x_2^n(x^1,x^2,\dots,x^{n-1})} \mathrm{d}x^n\,\phi(x^1, x^2, \dots, x^n)\,.$$

0.5.3 Surface Integral

Consider a surface $\mathcal{S} \subset \mathbb{R}^3$ parameterised by u and v.

Theorem 0.31. The surface vector area element is given by

$$\mathrm{d}\mathbf{S} = \hat{\mathbf{n}}\,\mathrm{d}S = \frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v}\,\mathrm{d}u\,\mathrm{d}v\ ,$$

where $\hat{\mathbf{n}}$ is the local unit normal vector to the surface and $\mathrm{d}S$ is the scalar area element.

Definition 0.32. The *flux* of a vector field $\mathbf{F}(\mathbf{x})$ over a surface $\mathcal{S}$ is

$$\iint_\mathcal{S} \mathbf{F}(\mathbf{x}) \cdot \mathrm{d}\mathbf{S}\ .$$

Theorem 0.33. Let a surface $\mathcal{S}$ be parameterised by u and v, and let the unit normal vector to the surface be $\hat{\mathbf{n}}(\mathbf{x})$. The flux of $\mathbf{F}$ through $\mathcal{S}$ is

$$\iint_\mathcal{S} \mathbf{F}(\mathbf{x}) \cdot \mathrm{d}\mathbf{S} = \iint_\mathcal{S} \mathbf{F}(\mathbf{x}) \cdot \hat{\mathbf{n}}\,\mathrm{d}S = \iint_\mathcal{S} \mathbf{F}(\mathbf{x}(u, v)) \cdot \left(\frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v}\right) \mathrm{d}u\,\mathrm{d}v\ .$$

0.6 Fourier Series

Lemma 0.34. For $n, m \in \mathbb{N}_0$,

$$\int_0^L \sin\frac{2\pi n x}{L} \sin\frac{2\pi m x}{L}\,\mathrm{d}x = \begin{cases} \frac{L}{2} & \text{if } n = m \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\int_0^L \cos\frac{2\pi n x}{L} \cos\frac{2\pi m x}{L}\,\mathrm{d}x = \begin{cases} \frac{L}{2} & \text{if } n = m \neq 0 \\ L & \text{if } n = m = 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\int_0^L \sin\frac{2\pi n x}{L} \cos\frac{2\pi m x}{L}\,\mathrm{d}x = 0\,.$$

Theorem 0.35 (Fourier series). Let $f(x)$ be a function with period L, then the Fourier series expansion of $f(x)$ is given by

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

where

$$a_n = \frac{2}{L}\int_{x_0}^{x_0+L} f(x)\cos\frac{2\pi nx}{L}\,\mathrm{d}x\ ,$$

$$b_n = \frac{2}{L}\int_{x_0}^{x_0+L} f(x)\sin\frac{2\pi nx}{L}\,\mathrm{d}x\ .$$

1 Vector Calculus

In this chapter, we will consider the prototype $\mathbb{E}^3$ vector space without reference to general vector spaces and linear algebra in an abstract sense — this will be done in section 4. We will first quickly review some basic concepts that should be familiar at this stage.

1.1 Vectors and Basis

Definition 1.1. A *vector* is a quantity specified by a magnitude and a direction in space.

A 3D Euclidean space $\mathbb{E}^3$ is a close approximation to our physical space, with the following properties:

- points are the elements of the space;
- vectors are translatable, directed line segments;
- being *Euclidean* means lengths and angles obey the classical results of geometry.

Definition 1.2. Two vectors $\mathbf{u}$ and $\mathbf{v}$ are *linearly independent* if

$$\lambda\mathbf{u} + \mu\mathbf{v} = \mathbf{0} \implies \lambda = \mu = 0\,.$$

Definition 1.3. A *basis* is a set of linearly independent non-zero vectors, $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_3$, such that any vector $\mathbf{v}$ can be expressed uniquely as

$$\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + v_3\mathbf{e}_3\,.$$

The numbers (scalars) v_1, v_2 and v_3 are the *components* of the vector in this basis.

Remark. The choice of a basis is not unique. If we choose a different basis, then the components will be different.

Definition 1.4. A set of basis vectors is said to be *orthonormal* if

$$\mathbf{e}_i \cdot \mathbf{e}_i = 1,\ \mathbf{e}_i \cdot \mathbf{e}_j = 0 \text{ for } i \neq j\,.$$

We can choose the basis of an $\mathbb{E}^3$ space to be any three linearly independent vectors. But an orthonormal basis will make the calculations much simpler.

Definition 1.5. An orthonormal basis is *right-handed* if

$$\mathbf{e}_1 \times \mathbf{e}_2 = \mathbf{e}_3\,,$$

and therefore

$$[\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3] \coloneqq \mathbf{e}_1 \cdot (\mathbf{e}_2 \times \mathbf{e}_3) = 1\,.$$

Definition 1.6. Identify $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_3$ along x, y, z directions respectively. For a position vector $\mathbf{x}$ with

$$\mathbf{x} = x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3\,,$$

its *Cartesian coordinate* is given by (x, y, z).

1.2 Suffix Notation

Definition 1.7. An alternative notation for vectors is to write

$$\mathbf{v} = (v_1, v_2, v_3) =: \{v_i\} \quad \text{for} \quad i = 1, 2, 3\,.$$

Such notation is called the *suffix notation*.

We also denote the i^{th} component of $\mathbf{v}$ as $(\mathbf{v})_i$. $(\mathbf{v})_i \equiv v_i$.

Under suffix notation, we express position vectors as

$$\mathbf{x} = (x, y, z) = (x_1, x_2, x_3) = \{x_i\}\,.$$

If two vectors are equal, $\mathbf{a} = \mathbf{b}$, then all components of them should be equal, so we write

$$a_i = b_i \text{ for } i = 1, 2, 3\,.$$

This is a vector equation; when we omit the 'for $i = 1, 2, 3$', it is understood that the one free suffix i ranges through 1,2,3 (or 1,2 in 2D) so as to give three component equations. Similarly,

$$\mathbf{c} = \lambda\mathbf{a} + \mu\mathbf{b} \iff c_i = \lambda a_i + \mu b_i\,.$$

The symbol for the index is arbitrary. We can use whatever symbol we want, say

$$c_\$ = \lambda a_\$ + \mu b_\$\,,$$

and it is nothing different than using i.

More complicated expressions can also be expressed using suffix notations. For example, the scalar product is

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

However, the expression will soon get complicated when we have complex expressions, especially when we have a lot of summation signs to write. To express $(\mathbf{a}\cdot\mathbf{b})(\mathbf{c}\cdot\mathbf{d})$, we need to write

$$\sum_{i=1}^{3}\sum_{j=1}^{3} a_i b_i c_j d_j\,,$$

which is not any simpler than the usual expression. To make our life easier, we will introduce summation convention to simplify our expressions.

1.2.1 Summation Convention

Definition 1.8. The rules of *Einstein's summation convention* are

- if a suffix appears once, it is taken to be a *free suffix* and ranged through;
- if a suffix appears twice, it is taken to be a *dummy suffix* and summed over;
- if a suffix appears more than twice in one term, something has gone wrong unless there is an explicit sum.

Examples.

(i) Scalar product:
$$\mathbf{a} \cdot \mathbf{b} = a_i b_i$$

(ii) Transpose of a matrix:
$$(\mathsf{A}^{\mathrm{T}})_{ij} = A_{ji}$$

(iii) Trace of a matrix:
$$\mathrm{tr}(\mathsf{A}) = A_{ii}$$

(iv) Matrix times a vector:
$$\mathbf{y} = \mathsf{A}\mathbf{x} \quad \Longleftrightarrow \quad y_i = A_{ij} x_j$$

(v) Matrix times a matrix:
$$\mathsf{A} = \mathsf{B}\mathsf{C} \quad \Longleftrightarrow \quad A_{ij} = B_{ik} C_{kj}$$

1.2.2 Kronecker Delta and Levi–Civita Symbol

We introduce two extra symbols, motivated by dot product and cross product (which will be clear later), that will make our life even simpler.

Definition 1.9. The *Kronecker delta*, δ_{ij}, is defined as

$$\delta_{ij} \coloneqq \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

$$\begin{pmatrix} \delta_{11} & \delta_{12} & \delta_{13} \\ \delta_{21} & \delta_{22} & \delta_{23} \\ \delta_{31} & \delta_{32} & \delta_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \mathsf{I}$$

The Kronecker delta has the following properties.

(i) δ_{ij} is symmetric. $\delta_{ij} = \delta_{ji}$.

(ii) $a_i \delta_{ij} = a_j$

(iii) $\delta_{ij}\delta_{jk} = \delta_{ik}$

(iv) $a_i \delta_{ij} b_j = a_i b_i = \mathbf{a} \cdot \mathbf{b}$

(v) $\delta_{ii} = 3$

(vi) For an orthonormal basis, $\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$.

Definition 1.10. The *Levi–Civita symbol*, ε_{ijk}, is defined as

$$\varepsilon_{ijk} \coloneqq \begin{cases} 1 & \text{if } i\,j\,k \text{ is an even permutation of } 1\,2\,3 \\ -1 & \text{if } i\,j\,k \text{ is an odd permutation of } 1\,2\,3 \\ 0 & \text{otherwise}\,. \end{cases}$$

Lemma 1.11. The product of two Levi–Civita symbols is

$$\varepsilon_{ijk}\varepsilon_{lmn} = \begin{vmatrix} \delta_{il} & \delta_{im} & \delta_{in} \\ \delta_{jl} & \delta_{jm} & \delta_{jn} \\ \delta_{kl} & \delta_{km} & \delta_{kn} \end{vmatrix}.$$

Proof. We can observe that the value of both the LHS and the RHS:

(i) are 0 when any of (i, j, k) are equal (two rows equal in a determinant), or when any of (l, m, n) are equal (two columns equal in a determinant);

(ii) are 1 when $(i, j, k) = (l, m, n) = (1, 2, 3)$;

(iii) change sign when any of (i, j, k) are interchanged (row interchange in a determinant), or when any of (l, m, n) are interchanged (column interchange in a determinant). □

Contracting an increasing number of indices, we get the following handy identities.

Corollary.

$$\begin{aligned}\varepsilon_{ijk}\varepsilon_{imn} &= \delta_{jm}\delta_{kn} - \delta_{jn}\delta_{km}\\ \varepsilon_{ijk}\varepsilon_{ijn} &= 2\delta_{kn}\\ \varepsilon_{ijk}\varepsilon_{ijk} &= 6\,.\end{aligned}$$

A vector product can be expressed as

$$\mathbf{a}\times\mathbf{b} = \varepsilon_{ijk}\mathbf{e}_i a_j b_k\,,$$

or equivalently

$$(\mathbf{a}\times\mathbf{b})_i = \varepsilon_{ijk}a_j b_k\,.$$

The scalar triple product is

$$\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = a_i(\mathbf{b}\times\mathbf{c})_i = \varepsilon_{ijk}a_i b_j c_k$$

and the vector triple product is

$$\begin{aligned}(\mathbf{a}\times(\mathbf{b}\times\mathbf{c}))_i &= \varepsilon_{ijk}a_j(\mathbf{b}\times\mathbf{c})_k\\ &= \varepsilon_{ijk}a_j\varepsilon_{klm}b_l c_m\\ &= (\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl})a_j b_l c_m\\ &= a_j b_i c_j - a_j b_j c_i\\ &= ((\mathbf{a}\cdot\mathbf{c})\mathbf{b} - (\mathbf{a}\cdot\mathbf{b})\mathbf{c})_i\,.\end{aligned}$$

1.2.3 Cauchy–Schwarz Inequality

Theorem 1.12 (Cauchy–Schwarz inequality).

$$\|\mathbf{x}\|^2\|\mathbf{y}\|^2 - |\mathbf{x}\cdot\mathbf{y}|^2 \geq 0$$

Proof.

$$\begin{aligned}\|\mathbf{x}\|^2\|\mathbf{y}\|^2 - |\mathbf{x}\cdot\mathbf{y}|^2 &= x_i x_i y_j y_j - x_i y_i x_j y_j\\ &= \frac{1}{2}x_i x_i y_j y_j + \frac{1}{2}x_j x_j y_i y_i - x_i y_i x_j y_j\\ &= \frac{1}{2}(x_i y_j - x_j y_i)^2 \geq 0\,.\end{aligned}$$

□

1.3 Vector Calculus in Cartesian Coordinates

1.3.1 Gradient and Directional Derivative

For a scalar field $\psi(\mathbf{x})$, consider a small change to the position $\mathbf{x}$, say to $\mathbf{x}+\delta\mathbf{x}$. This small change in position will generally produce a small change in ψ. Estimating this change in ψ using the Taylor series to the first order, we get

$$\begin{aligned}\delta\psi \equiv \psi(\mathbf{x}+\delta\mathbf{x}) - \psi(\mathbf{x}) &= \frac{\partial\psi}{\partial x}\delta x + \frac{\partial\psi}{\partial y}\delta y + \frac{\partial\psi}{\partial z}\delta z + \dots \\ &= \left(\frac{\partial\psi}{\partial x}\mathbf{e}_x + \frac{\partial\psi}{\partial y}\mathbf{e}_y + \frac{\partial\psi}{\partial z}\mathbf{e}_z\right)\cdot(\delta x\mathbf{e}_x + \delta y\mathbf{e}_y + \delta z\mathbf{e}_z) + \dots\end{aligned}$$

The terms in the second bracket are just $\delta\mathbf{x}$, so the terms in the first bracket evaluate how the scalar field ψ changes to the first order in $\delta\mathbf{x}$.

Definition 1.13. The *gradient* of a scalar field ψ is

$$\boldsymbol{\nabla}\psi \coloneqq \frac{\partial\psi}{\partial x_i}\mathbf{e}_i\,.$$

When $\delta\mathbf{x}$ is infinitesimal, the higher order terms in the Taylor expansion becomes negligible, so we can write

$$\mathrm{d}\psi = \boldsymbol{\nabla}\psi\cdot\mathrm{d}\mathbf{x}\,.$$

Definition 1.14. The *differential operator*, $\boldsymbol{\nabla}$, is defined as

$$\boldsymbol{\nabla} \coloneqq \mathbf{e}_i\frac{\partial}{\partial x_i}\,.$$

Definition 1.15. The *directional derivative* of ψ in a direction $\hat{\mathbf{l}}$ is the rate of change of ψ in the direction of $\hat{\mathbf{l}}$. By doing a Taylor expansion of $\psi(\mathbf{x}+s\hat{\mathbf{l}})$ for small s, it's easy to see that this is

$$\left.\frac{\mathrm{d}}{\mathrm{d}s}\psi(\mathbf{x}+s\hat{\mathbf{l}})\right|_{s=0} = \hat{\mathbf{l}}\cdot\boldsymbol{\nabla}\psi\,.$$

Proposition 1.16. A unit normal to the surface $\phi = \text{const.}$ is given by

$$\hat{\mathbf{n}} = \frac{\boldsymbol{\nabla}\phi}{|\boldsymbol{\nabla}\phi|}\,.$$

Proof. When the directional derivative is zero, i.e. $\hat{\mathbf{l}}\cdot\boldsymbol{\nabla}\phi = 0$, it follows that if $\boldsymbol{\nabla}\phi \neq \mathbf{0}$, then ϕ does not change in the direction of $\mathbf{l}$; hence $\mathbf{l}$ is a tangent to the surface $\phi = \text{const.}$ $\boldsymbol{\nabla}\phi$ is orthogonal to all $\hat{\mathbf{l}}$ tangent to the surface, so it is the normal. □

1.3.2 Divergence and Curl

Here we introduce two differential operators that are ubiquitous in vector calculus.

Definition 1.17. For a vector field

$$\mathbf{F}(\mathbf{x}) = F_i(\mathbf{x})\mathbf{e}_i\,,$$

the *divergence* is defined as

$$\begin{aligned}\operatorname{div}\mathbf{F} \coloneqq \boldsymbol{\nabla}\cdot\mathbf{F} &= \left(\mathbf{e}_x\frac{\partial}{\partial x} + \mathbf{e}_y\frac{\partial}{\partial y} + \mathbf{e}_z\frac{\partial}{\partial z}\right)\cdot(F_x\mathbf{e}_x + F_y\mathbf{e}_y + F_z\mathbf{e}_z) \\ &= \frac{\partial F_i}{\partial x_i}\,,\end{aligned}$$

and the *curl* is defined as

$$\begin{aligned}\operatorname{curl}\mathbf{F} \coloneqq \boldsymbol{\nabla}\times\mathbf{F} &= \left(\mathbf{e}_x\frac{\partial}{\partial x}+\mathbf{e}_y\frac{\partial}{\partial y}+\mathbf{e}_z\frac{\partial}{\partial z}\right)\times(F_x\mathbf{e}_x+F_y\mathbf{e}_y+F_z\mathbf{e}_z)\\ &= \begin{vmatrix}\mathbf{e}_x & \mathbf{e}_y & \mathbf{e}_z\\ \partial_x & \partial_y & \partial_z\\ F_x & F_y & F_z\end{vmatrix}\\ &= \varepsilon_{ijk}\mathbf{e}_i\frac{\partial F_k}{\partial x_j}\,,\end{aligned}$$

where ∂_x is the shorthand notation of $\frac{\partial}{\partial x}$.

Definition 1.18. The scalar operator $\mathbf{F}\cdot\boldsymbol{\nabla}$ is defined as

$$\mathbf{F}\cdot\boldsymbol{\nabla} \coloneqq F_i\frac{\partial}{\partial x_i}\,.$$

It acts on both scalar and vector fields.

$$(\mathbf{F}\cdot\boldsymbol{\nabla})\psi = F_i\frac{\partial\psi}{\partial x_i} = \mathbf{F}\cdot(\boldsymbol{\nabla}\psi)$$

$$((\mathbf{F}\cdot\boldsymbol{\nabla})\mathbf{G})_i = F_j\frac{\partial G_i}{\partial x_j}$$

1.4 Second-order Vector Differential Operators

1.4.1 Curl Grad and Div Curl

Theorem 1.19. For any differentiable scalar field ψ and vector field $\mathbf{F}$

$$\boldsymbol{\nabla}\times\boldsymbol{\nabla}\psi \equiv \mathbf{0}$$

$$\boldsymbol{\nabla}\cdot\boldsymbol{\nabla}\times\mathbf{F} \equiv 0$$

Proof.

$$\begin{aligned}\boldsymbol{\nabla}\times\boldsymbol{\nabla}\psi &= \varepsilon_{ijk}\mathbf{e}_i\frac{\partial}{\partial x_j}\left(\frac{\partial}{\partial x_k}\psi\right)\\ &= \varepsilon_{ikj}\mathbf{e}_i\frac{\partial}{\partial x_k}\frac{\partial}{\partial x_j}\psi\\ &= -\varepsilon_{ijk}\mathbf{e}_i\frac{\partial}{\partial x_j}\frac{\partial}{\partial x_k}\psi\\ &= \mathbf{0}\end{aligned}$$

$$\begin{aligned}\boldsymbol{\nabla}\cdot\boldsymbol{\nabla}\times\mathbf{F} &= \frac{\partial}{\partial x_i}\varepsilon_{ijk}\frac{\partial}{\partial x_j}F_k\\ &= \frac{\partial}{\partial x_j}\varepsilon_{jik}\frac{\partial}{\partial x_i}F_k\\ &= -\frac{\partial}{\partial x_i}\varepsilon_{ijk}\frac{\partial}{\partial x_j}F_k\\ &= 0\end{aligned}$$

□

Definition 1.20. A vector field $\mathbf{F}(\mathbf{x})$ is *irrotational* if $\boldsymbol{\nabla} \times \mathbf{F} = \mathbf{0}$. A vector field is *conservative* if there exists a scalar field $\varphi(\mathbf{x})$ such that $\mathbf{F} = \boldsymbol{\nabla}\varphi$.

Theorem 1.21. The line integral around any closed curve vanishes if and only if $\mathbf{F}$ is conservative.

Proof.

($\Leftarrow$) Consider a conservative field $\mathbf{F} = \boldsymbol{\nabla}\varphi$. Consider a curve $\mathcal{C}$ that interpolates from $\mathbf{a}$ to $\mathbf{b}$, with parameterisation $\mathbf{x}(t)$. We have

$$\int_{\mathcal{C}} \mathbf{F} \cdot \mathrm{d}\mathbf{x} = \int_{\mathcal{C}} \boldsymbol{\nabla}\varphi \, \mathrm{d}\mathbf{x} = \int_{t_a}^{t_b} \frac{\partial \varphi}{\partial x_i} \frac{\mathrm{d}x_i}{\mathrm{d}t} \, \mathrm{d}t = \int_{t_a}^{t_b} \frac{\mathrm{d}}{\mathrm{d}t} \varphi(\mathbf{x}(t)) \, \mathrm{d}t \ .$$

We now have the integral of a total derivative, so

$$\int_{\mathcal{C}} \mathbf{F} \cdot \mathrm{d}\mathbf{x} = [\varphi(\mathbf{x}(t))]_{t_a}^{t_b} = \varphi(\mathbf{b}) - \varphi(\mathbf{a}) \, .$$

This means that the line integral is only dependent on the endpoints of the path. This also implies that the integral around any closed curve is 0.

($\Rightarrow$) We are able to construct a potential φ for any conservative field $\mathbf{F}$. Take an arbitrary value of φ at some point ($\varphi(\mathbf{0}) = 0$ for example), then for any other point $\mathbf{x} = \mathbf{y}$, its potential is given by

$$\varphi(\mathbf{y}) = \int_{\mathcal{C}(\mathbf{y})} \mathbf{F} \cdot \mathrm{d}\mathbf{x} \ ,$$

where $\mathcal{C}(\mathbf{y})$ is an arbitrary curve from $\mathbf{x} = \mathbf{0}$ to $\mathbf{x} = \mathbf{y}$. This is well-defined since a vanishing loop integral implies the line integral is only dependent on the endpoints of the curve.

We have

$$\begin{aligned} \frac{\partial \varphi(\mathbf{y})}{\partial x_i} &= \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[\int_{\mathcal{C}(\mathbf{y} + \epsilon \mathbf{e}_i)} \mathbf{F} \cdot \mathrm{d}\mathbf{x} - \int_{\mathcal{C}(\mathbf{y})} \mathbf{F} \cdot \mathrm{d}\mathbf{x} \right] \\ &= \lim_{\epsilon \to 0} \frac{1}{\epsilon} \int_{\mathcal{C}(\mathbf{y} + \epsilon \mathbf{e}_i) - \mathcal{C}(\mathbf{y})} \mathbf{F} \cdot \mathrm{d}\mathbf{x} \\ &= \mathbf{F}(\mathbf{y}) \cdot \mathbf{e}_i = F_i(\mathbf{y}) \, . \end{aligned}$$

This is our desired result $\boldsymbol{\nabla}\varphi = \mathbf{F}$. □

Theorem 1.22 (Poincaré lemma). For vector fields defined everywhere on $\mathbb{R}^3$, conservative is the same as irrotational.

$$\boldsymbol{\nabla} \times \mathbf{F} = \mathbf{0} \iff \mathbf{F} = \boldsymbol{\nabla}\varphi \text{ for some } \varphi \, .$$

Proof.

($\Rightarrow$) Corollary of Stokes' theorem (Theorem 1.28). Will be proved later.

($\Leftarrow$)

$$\boldsymbol{\nabla} \times \mathbf{F} = \boldsymbol{\nabla} \times \boldsymbol{\nabla}\varphi = \mathbf{0}$$

by Theorem 1.19. □

Definition 1.23. A vector field $\mathbf{F}(\mathbf{x})$ is *solenoidal* if $\boldsymbol{\nabla} \cdot \mathbf{F} = 0$.

Theorem 1.24. Any solenoidal field defined everywhere in $\mathbb{R}^3$ is the curl of some vector field.

$$\boldsymbol{\nabla} \cdot \mathbf{F} = 0 \iff \mathbf{F} = \boldsymbol{\nabla} \times \mathbf{A} \text{ for some } \mathbf{A} \, .$$

Proof.

($\Rightarrow$) We pick some arbitrary point $\mathbf{x}_0 = (x_0, y_0, z_0)$ and construct the following field

$$\mathbf{A}(\mathbf{x}) = \left(\int_{z_0}^{z} F_y(x, y, z')\,\mathrm{d}z' \,, \int_{x_0}^{x} F_z(x', y, z_0)\,\mathrm{d}x' - \int_{z_0}^{z} F_x(x, y, z')\,\mathrm{d}z' \,, 0 \right).$$

$$\boldsymbol{\nabla} \times \mathbf{A} = \left(-\frac{\partial A_y}{\partial z}, \frac{\partial A_x}{\partial z}, \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} \right).$$

If follows immediately from the fundamental theorem of calculus that the x and y components of $\boldsymbol{\nabla} \times \mathbf{A}$ is $\mathbf{F}$. Using the condition that $\boldsymbol{\nabla} \cdot \mathbf{F} = 0$, we have

$$\begin{aligned}(\boldsymbol{\nabla} \times \mathbf{A})_z &= F_z(x, y, z_0) - \int_{z_0}^{z} \frac{\partial F_x}{\partial x}\,\mathrm{d}z' - \int_{z_0}^{z} \frac{\partial F_y}{\partial y}\,\mathrm{d}z' \\ &= F_z(x, y, z_0) + \int_{z_0}^{z} \frac{\partial F_z}{\partial z}\,\mathrm{d}z' = F_z(x, y, z)\,.\end{aligned}$$

The existence of $\mathbf{A}$ is proved by construction.

($\Leftarrow$)

$$\boldsymbol{\nabla} \cdot \mathbf{F} = \boldsymbol{\nabla} \cdot \boldsymbol{\nabla} \times \mathbf{A} = 0$$

by Theorem 1.19. □

1.4.2 Laplacian

Consider

$$\begin{aligned}\boldsymbol{\nabla} \cdot \boldsymbol{\nabla}\psi &= \frac{\partial}{\partial x_i}\left(\frac{\partial}{\partial x_i}\psi \right) \\ &= \frac{\partial^2 \psi}{\partial x_i^2} \\ &= \left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2} \right)\psi\,.\end{aligned}$$

Definition 1.25. The *Laplacian operator*, ∇^2, is defined as

$$\nabla^2 \coloneqq \boldsymbol{\nabla} \cdot \boldsymbol{\nabla} = \frac{\partial^2}{\partial x_i^2}\,.$$

Remark. Although defined for scalar fields, the Laplacian can also act on a vector field componentwise. Its action on a vector field can also be defined via the identity

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

1.5 Integral Theorems

1.5.1 The Divergence Theorem (Gauss' Theorem)

Theorem 1.26 (The divergence theorem (Gauss' theorem)). Let $\mathcal{S}$ be a piecewise smooth surface enclosing a volume $\mathcal{V}$ in $\mathbb{R}^3$, with a normal $\hat{\mathbf{n}}$ that points outwards from $\mathcal{V}$. Let $\mathbf{F}$ be a smooth vector field. Then

$$\iiint_{\mathcal{V}} \boldsymbol{\nabla} \cdot \mathbf{F}\,\mathrm{d}V = \oiint_{\mathcal{S}} \mathbf{F} \cdot \mathrm{d}\mathbf{S}\,.$$

Proof. Consider $\mathbf{F} = F_x\mathbf{e}_x + F_y\mathbf{e}_y + F_z\mathbf{e}_z$. We have

$$\iiint_{\mathcal{V}} \boldsymbol{\nabla}\cdot\mathbf{F}\,\mathrm{d}V = \iiint_{\mathcal{V}} \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}\,\mathrm{d}V\ .$$

If the theorem holds for one component of $\mathbf{F}$, it must hold for the sum of all three dimensions. Let us consider the z component only.

Let us denote the projection of $\mathcal{V}$ on the xy plane as $\mathcal{D}$, and the upper and lower boundary of $\mathcal{V}$ as $z_+(x,y)$ and $z_-(x,y)$. We have

$$\begin{aligned}\iiint_{\mathcal{V}} \frac{\partial F_z}{\partial z}\,\mathrm{d}V &= \iint_{\mathcal{D}} \int_{z_-(x,y)}^{z_+(x,y)} \frac{\partial F_z}{\partial z}\,\mathrm{d}z\,\mathrm{d}A \\ &= \iint_{\mathcal{D}} (F_z(x,y,z_+(x,y)) - F_z(x,y,z_-(x,y)))\,\mathrm{d}A\ .\end{aligned}$$

Suppose that the normal of the upper surface makes an angle θ with $\mathbf{e}_z$. We have

$$\delta A = \cos\theta\delta S = \mathbf{e}_z\cdot\hat{\mathbf{n}}\delta S\,.$$

Similarly, for the lower bound, we have

$$\delta A = -\mathbf{e}_z\cdot\hat{\mathbf{n}}\delta S\,.$$

Therefore, we can write

$$\begin{aligned}\iiint_{\mathcal{V}} \frac{\partial F_z}{\partial z}\,\mathrm{d}V &= \iint_{\mathcal{D}} (F_z(x,y,z_+(x,y))\mathbf{e}_z\cdot\hat{\mathbf{n}} + F_z(x,y,z_-(x,y)))\mathbf{e}_z\cdot\hat{\mathbf{n}}\,\mathrm{d}S \\ &= \iint_{\mathcal{S}_+} F_z\mathbf{e}_z\cdot\hat{\mathbf{n}}\,\mathrm{d}S + \iint_{\mathcal{S}_-} F_z\mathbf{e}_z\cdot\hat{\mathbf{n}}\,\mathrm{d}S \\ &= \oiint_{\mathcal{S}} F_z\mathbf{e}_z\cdot\hat{\mathbf{n}}\,\mathrm{d}S\ .\end{aligned}$$

Summing over all three dimensions, we have

$$\begin{aligned}\iiint_{\mathcal{V}} \boldsymbol{\nabla}\cdot\mathbf{F}\,\mathrm{d}V &= \iiint_{\mathcal{V}} \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}\,\mathrm{d}V \\ &= \oiint_{\mathcal{S}} F_x\mathbf{e}_x\cdot\mathbf{n} + F_y\mathbf{e}_y\cdot\mathbf{n} + F_z\mathbf{e}_z\cdot\mathbf{n}\,\mathrm{d}S \\ &= \oiint_{\mathcal{S}} \mathbf{F}\cdot\mathrm{d}\mathbf{S}\ .\end{aligned}$$

□

Corollary. Generalisation for a scalar field. Let ψ be a smooth scalar field,

$$\iiint_{\mathcal{V}} \boldsymbol{\nabla}\psi\,\mathrm{d}V = \oiint_{\mathcal{S}} \psi\,\mathrm{d}\mathbf{S}\ .$$

Proof. Set $\mathbf{F} = \psi\mathbf{a}$, where $\mathbf{a}$ is an arbitrary constant vector. Then by the divergence theorem,

$$\mathbf{a}\cdot\iiint_{\mathcal{V}} \boldsymbol{\nabla}\psi\,\mathrm{d}V = \mathbf{a}\cdot\oiint_{\mathcal{S}} \psi\,\mathrm{d}\mathbf{S}\ .$$

Since $\mathbf{a}$ is arbitrary, the corollary follows. □

Alternative proof. Choose $\mathbf{a} = \mathbf{e}_i$ to obtain the component form

$$\iiint_{\mathcal{V}} \frac{\partial\psi}{\partial x_i}\,\mathrm{d}V = \oiint_{\mathcal{S}} \psi n_i\,\mathrm{d}S\ .$$

Then the corollary follows. □

Corollary. Generalisation for a vector field. Let $\mathbf{A}$ be a smooth vector field,

$$\iiint_{\mathcal{V}} \boldsymbol{\nabla} \times \mathbf{A} \, \mathrm{d}V = \oiint_{\mathcal{S}} \hat{\mathbf{n}} \times \mathbf{A} \, \mathrm{d}S \ .$$

Proof. Either set $\mathbf{F} = \mathbf{a} \times \mathbf{A}$ in the divergence theorem (Theorem 1.26), where $\mathbf{a}$ is an arbitrary constant vector, and then proceed as above, or let $\psi = \varepsilon_{ijk} A_j$ to recover this corollary in component form. □

1.5.2 Stokes' Theorem

Theorem 1.27 (Green's theorem). For smooth functions $P(x, y)$ and $Q(x, y)$ in a closed region $\mathcal{A} \subset \mathbb{R}^2$ bounded by piecewise smooth, non-intersecting closed curve $\mathcal{C} = \partial\mathcal{A}$,

$$\iint_{\mathcal{A}} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) \mathrm{d}x \, \mathrm{d}y = \oint_{\mathcal{C}} (P \, \mathrm{d}x + Q \, \mathrm{d}y) \, .$$

Remark. Green's theorem is equivalent to the 2D version of the Stokes' theorem, which we will see immediately after.

Proof. Let $\mathbf{F} = (Q, -P)$ be a vector field. We then have

$$\iint_{\mathcal{A}} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) \mathrm{d}A = \iint_{\mathcal{A}} \boldsymbol{\nabla} \cdot \mathbf{F} \, \mathrm{d}A \ .$$

Parameterise curve $\mathcal{C}$ by $\mathbf{x}(s) = (x(s), y(s))$, then the tangent vector is $\mathbf{t}(s) = (x'(s), y'(s))$ and the normal vector is $\hat{\mathbf{n}} = (y'(s), -x'(s))$. We then have

$$\mathbf{F} \cdot \hat{\mathbf{n}} = Q \frac{\mathrm{d}y}{\mathrm{d}s} + P \frac{\mathrm{d}x}{\mathrm{d}s} \, ,$$

and so the integral around $\mathcal{C}$ is

$$\oint_{\mathcal{C}} \mathbf{F} \cdot \hat{\mathbf{n}} \, \mathrm{d}s = \oint_{\mathcal{C}} P \, \mathrm{d}x + Q \, \mathrm{d}y \ .$$

The 2D divergence theorem states that

$$\iint_{\mathcal{A}} \boldsymbol{\nabla} \cdot \mathbf{F} \, \mathrm{d}A = \oint_{\mathcal{C}} \mathbf{F} \cdot \hat{\mathbf{n}} \, \mathrm{d}s \ ,$$

so the Green's theorem follows. □

Theorem 1.28 (Stokes' theorem). Let $\mathcal{C}$ be a piecewise smooth closed curve bounding a smooth open surface $\mathcal{S}$. Let $\mathbf{F}(\mathbf{x})$ be a smooth vector field. Then

$$\iint_{\mathcal{S}} \boldsymbol{\nabla} \times \mathbf{F} \cdot \mathrm{d}\mathbf{S} = \oint_{\mathcal{C}} \mathbf{F} \cdot \mathrm{d}\mathbf{x} \ ,$$

where the line integral is taken in the direction of $\mathcal{C}$ as specified by the right-hand rule.

Proof. Parameterise the surface $\mathcal{S}$ by $\mathbf{x}(u, v)$. Denote the associated area in the (u, v) plane as $\mathcal{A}$. Parameterise the boundary $\mathcal{C} = \partial\mathcal{S}$ as $\mathbf{x}(u(t), v(t))$ so that the corresponding boundary in (u, v) plane ∂A is $(u(t), v(t))$.

The key idea is to use Green's theorem in the (u, v) plane for the area $\mathcal{A}$ and then uplift this to prove Stokes' theorem for the surface $\mathcal{S}$.

The line integral around the boundary is

$$\oint_{\mathcal{C}} \mathbf{F} \cdot \mathrm{d}\mathbf{x} = \oint_{\mathcal{C}} \mathbf{F} \cdot \left(\frac{\partial \mathbf{x}}{\partial u} \, \mathrm{d}u + \frac{\partial \mathbf{x}}{\partial v} \, \mathrm{d}v \right) = \oint_{\partial\mathcal{A}} F_u \, \mathrm{d}u + F_v \, \mathrm{d}v \ ,$$

where $F_i = \mathbf{F} \cdot \frac{\partial \mathbf{x}}{\partial i}$. By Green's theorem,

$$\oint_{\partial \mathcal{A}} F_u \,\mathrm{d}u + F_v \,\mathrm{d}v = \iint_{\mathcal{A}} \left(\frac{\partial F_v}{\partial u} - \frac{\partial F_u}{\partial v} \right) \mathrm{d}A \ .$$

The partial derivatives are evaluated to be

$$\frac{\partial F_v}{\partial u} = \frac{\partial}{\partial u}\left(\mathbf{F} \cdot \frac{\partial \mathbf{x}}{\partial v} \right) = \frac{\partial}{\partial u}\left(F_i \frac{\partial x_i}{\partial v} \right) = \left(\frac{\partial F_i}{\partial x_j} \frac{\partial x_j}{\partial u} \right) \frac{\partial x_i}{\partial v} + F_i \frac{\partial^2 x_i}{\partial u \partial v} \,,$$

$$\frac{\partial F_u}{\partial v} = \frac{\partial}{\partial v}\left(\mathbf{F} \cdot \frac{\partial \mathbf{x}}{\partial u} \right) = \frac{\partial}{\partial v}\left(F_i \frac{\partial x_i}{\partial u} \right) = \left(\frac{\partial F_i}{\partial x_j} \frac{\partial x_j}{\partial v} \right) \frac{\partial x_i}{\partial u} + F_i \frac{\partial^2 x_i}{\partial v \partial u} \,.$$

The difference between the two partial derivatives becomes

$$\begin{aligned} \frac{\partial F_v}{\partial u} - \frac{\partial F_u}{\partial v} &= \frac{\partial x_j}{\partial u} \frac{\partial x_i}{\partial v} \left(\frac{\partial F_i}{\partial x_j} - \frac{\partial F_j}{\partial x_i} \right) \\ &= (\delta_{jk}\delta_{il} - \delta_{jl}\delta_{ik}) \frac{\partial x_k}{\partial u} \frac{\partial x_l}{\partial v} \frac{\partial F_i}{\partial x_j} \\ &= \varepsilon_{jip}\varepsilon_{pkl} \frac{\partial x_k}{\partial u} \frac{\partial x_l}{\partial v} \frac{\partial F_i}{\partial x_j} \\ &= (\boldsymbol{\nabla} \times \mathbf{F}) \cdot \left(\frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v} \right) . \end{aligned}$$

The Stokes' theorem follows.

$$\begin{aligned} \oint_{\mathcal{C}} \mathbf{F} \cdot \mathrm{d}\mathbf{x} &= \iint_{\mathcal{A}} (\boldsymbol{\nabla} \times \mathbf{F}) \cdot \left(\frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v} \right) \mathrm{d}u \,\mathrm{d}v \\ &= \iint_{\mathcal{S}} (\boldsymbol{\nabla} \times \mathbf{F}) \cdot \mathrm{d}\mathbf{S} \ . \end{aligned}$$

□

Corollary. An irrotational field defined everywhere on $\mathbb{R}^3$ is conservative. (Poincaré Lemma, Theorem 1.22, ⇒.)

Proof. It follows from Stokes' theorem that an irrotational vector field, obeying $\boldsymbol{\nabla} \times \mathbf{F} = \mathbf{0}$, necessarily has

$$\oint_{\mathcal{C}} \mathbf{F} \cdot \mathrm{d}x = 0$$

around any closed curve $\mathcal{C}$. By Theorem 1.21, it can be written as $\mathbf{F} = \boldsymbol{\nabla}\varphi$ for some potential. □

1.5.3 Alternative Interpretation of Divergence and Curl

The divergence theorem and Stokes' theorem can give us some more intuitive physical view of what the divergence and the curl are.

Theorem 1.29. Let a closed surface $\mathcal{S}$ enclose $\mathcal{V}$ with volume $|\mathcal{V}|$,

$$\boldsymbol{\nabla} \cdot \mathbf{u} = \lim_{|\mathcal{V}| \to 0} \frac{1}{|\mathcal{V}|} \oiint_{\mathcal{S}} \mathbf{u} \cdot \mathrm{d}\mathbf{S} \ .$$

Let an open smooth surface $\mathcal{S}$ of area $|\mathcal{S}|$ and normal direction $\hat{\mathbf{n}}$ be bounded by a curve $\mathcal{C}$,

$$\hat{\mathbf{n}} \cdot (\boldsymbol{\nabla} \times \mathbf{u}) = \lim_{|\mathcal{S}| \to 0} \frac{1}{|\mathcal{S}|} \oint_{\mathcal{C}} \mathbf{u} \cdot \mathrm{d}\mathbf{x} \ .$$

Proof. Follows directly from the divergence theorem and Stokes' theorem. □

Under such interpretations, the names 'divergence' and 'curl' should be obvious.

1.6 Coordinate Systems

From now on, we will move away from our good old Cartesian coordinate system and consider some more general coordinate systems, such as the spherical polar coordinates or the cylindrical coordinates.

In general, to describe a point in a n-dimensional space $\mathbb{R}^n$, we need n real numbers $\{q_i\}_{i=1}^n$ known as the *coordinates* of the point. So $\mathbf{x} = \mathbf{x}(q_1, \dots, q_n)$. As long as our coordinate system is defined in a good way, changing any one of these coordinates slightly, leaving the others fixed, will result in a small change in $\mathbf{x}$. We write

$$\mathrm{d}\mathbf{x} = \frac{\partial \mathbf{x}}{\partial q_i}\,\mathrm{d}q_i =: \mathbf{h_i}\,\mathrm{d}q_i \ .$$

Remark. $\frac{\partial \mathbf{x}}{\partial q_i}$ is the tangent vector to the lines formed when changing q_i while holding other coordinates constant.

Definition 1.30. The *metric coefficient* or *scale factor* of coordinate, h_i, is defined by

$$h_i \coloneqq |\mathbf{h}_i| = \left|\frac{\partial \mathbf{x}}{\partial q_i}\right| .$$

Definition 1.31. The *unit vector* of a coordinate, $\mathbf{e}_i$, is defined as

$$\mathbf{e}_i \coloneqq \frac{1}{h_i}\frac{\partial \mathbf{x}}{\partial q_i} .$$

1.6.1 The Jacobian

Definition 1.32. The *Jacobian matrix* J of the transformation from coordinates $(x_1, x_2, \dots, x_n)$ to $(q_1, q_2, \dots, q_n)$ is given by

$$\mathsf{J} \coloneqq \begin{pmatrix} \frac{\partial x_1}{\partial q_1} & \frac{\partial x_1}{\partial q_2} & \cdots & \frac{\partial x_1}{\partial q_n} \\ \frac{\partial x_2}{\partial q_1} & \frac{\partial x_2}{\partial q_2} & \cdots & \frac{\partial x_2}{\partial q_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial q_1} & \frac{\partial x_n}{\partial q_2} & \cdots & \frac{\partial x_n}{\partial q_n} \end{pmatrix} .$$

Definition 1.33. The *Jacobian* is the determinant of the Jacobian matrix:

$$J \coloneqq \frac{\partial(x_1, x_2, \dots, x_n)}{\partial(q_1, q_2, \dots, q_n)} = \det \mathsf{J} .$$

The columns of the Jacobian matrix are the vectors $\mathbf{h}_i$, so in three dimensional coordinates.

$$J = [\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3] = \mathbf{h}_1 \cdot (\mathbf{h}_2 \times \mathbf{h}_3) .$$

If we change the coordinates by $\{\mathrm{d}q_i\}$, this will be displacements $\{\mathrm{d}\mathbf{x}_i\} = \{\mathbf{h}_i\,\mathrm{d}q_i\}$ in the physical space along the three coordinate directions. They span a parallelepiped, which is the change in the physical volume when the coordinates change infinitesimally.

Proposition 1.34. For a coordinate system, the infinitesimal volume element in $\mathbb{R}^3$ is given by

$$\mathrm{d}V = |J|\,\mathrm{d}q_1\,\mathrm{d}q_2\,\mathrm{d}q_3 \ ,$$

so in volume integrals,

$$\iiint \Phi\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = \iiint \Phi |J|\,\mathrm{d}q_1\,\mathrm{d}q_2\,\mathrm{d}q_3 \ .$$

Under such interpretations of the Jacobian, the following two results should be obvious:

Proposition 1.35 (Chain rule for Jacobian).

$$\frac{\partial(\alpha_1, \dots \alpha_n)}{\partial(\gamma_1, \dots, \gamma_n)} = \frac{\partial(\alpha_1, \dots \alpha_n)}{\partial(\beta_1, \dots, \beta_n)} \frac{\partial(\beta_1, \dots \beta_n)}{\partial(\gamma_1, \dots, \gamma_n)}$$

Proposition 1.36 (Inverse transformation).

$$\frac{\partial(\alpha_1, \dots \alpha_n)}{\partial(\beta_1, \dots, \beta_n)} = \left[\frac{\partial(\beta_1, \dots \beta_n)}{\partial(\alpha_1, \dots, \alpha_n)}\right]^{-1}$$

The proofs are also straightforward. They follow from taking the determinant of the normal chain rule and the differentiation of inverse functions.

We have the following two ugly formulae that are rarely used.

Proposition 1.37. The surface area element for a surface in $\mathbb{R}^3$ is given by

$$\mathrm{d}\mathbf{S} = \operatorname{sign}(\hat{\mathbf{n}} \cdot \mathbf{e}_1) h_2 h_3 \,\mathrm{d}q_2 \,\mathrm{d}q_3 \,\mathbf{e}_1 + \operatorname{sign}(\hat{\mathbf{n}} \cdot \mathbf{e}_2) h_3 h_1 \,\mathrm{d}q_3 \,\mathrm{d}q_1 \,\mathbf{e}_2 + \operatorname{sign}(\hat{\mathbf{n}} \cdot \mathbf{e}_3) h_1 h_2 \,\mathrm{d}q_1 \,\mathrm{d}q_2 \,\mathbf{e}_3 \,,$$

$$\mathrm{d}S = \hat{\mathbf{n}} \cdot \mathrm{d}\mathbf{S} = h_2 h_3 \,\mathrm{d}q_2 \,\mathrm{d}q_3 \,|\hat{\mathbf{n}} \cdot \mathbf{e}_1| + h_3 h_1 \,\mathrm{d}q_3 \,\mathrm{d}q_1 \,|\hat{\mathbf{n}} \cdot \mathbf{e}_2| + h_1 h_2 \,\mathrm{d}q_1 \,\mathrm{d}q_2 \,|\hat{\mathbf{n}} \cdot \mathbf{e}_3| \,.$$

This one is much more common.

Corollary. If a surface in $\mathbb{R}^3$ is defined by holding a coordinate constant, so that the surface is parameterised by the other two coordinate u and v, then

$$\mathrm{d}\mathbf{S} = \frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v} \,\mathrm{d}u \,\mathrm{d}v \,.$$

1.7 Orthogonal Curvilinear Coordinates

It is very common that the three sets of basis vectors for the coordinates are always orthonormal. They are then called *orthogonal curvilinear coordinates*. Both spherical and cylindrical polar coordinates are examples of these.

1.7.1 Orthogonality

For orthogonal curvilinear coordinates, the $\mathbf{e}_i$ are required to be mutually orthogonal at all points in space:

$$\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij} \,,$$

and it is conventional to order q_i so that the coordinate system is right-handed:

$$\mathbf{e}_1 \times \mathbf{e}_2 = \mathbf{e}_3 \,.$$

Therefore in an orthogonal curvilinear coordinate system, the expression for the incremental distance is simplified to

$$\begin{aligned} |\mathrm{d}\mathbf{x}|^2 &= \mathrm{d}\mathbf{x} \cdot \mathrm{d}\mathbf{x} \\ &= \sum_{i,j} (h_i \,\mathrm{d}q_i)(h_j \,\mathrm{d}q_j)\delta_{ij} \\ &= \sum_i h_i^2 (\mathrm{d}q_i)^2 \,. \end{aligned}$$

1.7.2 Relationship between Coordinates

Suppose we have non-Cartesian coordinates $q_i\,(i = 1, 2, 3)$. There will be a functional dependence of q_i on Cartesian coordinates x_i:

$$q_i \equiv q_i(x, y, z)\,.$$

The position vector $\mathbf{x}$ is also a function of $\mathbf{q} = (q_1, q_2, q_3)$:

$$x_i \equiv x_i(q_1, q_2, q_3)\,.$$

Taking cylindrical polar coordinates and spherical polar coordinates as examples, we have:

	Cylindrical Polar Coordinates	Spherical Polar Coordinates
q_1	$\rho = \sqrt{x^2 + y^2}$	$r = \sqrt{x^2 + y^2 + z^2}$
q_2	$\phi = \tan^{-1}\left(\frac{y}{x}\right)$	$\theta = \tan^{-1}\left(\frac{\sqrt{x^2+y^2}}{z}\right)$
q_3	z	$\phi = \tan^{-1}\left(\frac{y}{x}\right)$

	Cylindrical Polar Coordinates	Spherical Polar Coordinates
x	$\rho\cos\phi$	$r\sin\theta\cos\phi$
y	$\rho\sin\phi$	$r\sin\theta\sin\phi$
z	z	$r\cos\theta$

1.7.3 Spherical Polar Coordinates

In spherical polar coordinates, $q_1 = r \in [0, \infty)$, $q_2 = \theta \in [0, \pi]$, $q_3 = \phi \in [0, 2\pi)$, and in terms of Cartesian coordinates,

$$\mathbf{x} = (r\sin\theta\cos\phi, r\sin\theta\sin\phi, r\cos\theta)\,.$$

Proposition 1.38. For spherical polar coordinates, we have

$$\begin{aligned}
\frac{\partial\mathbf{x}}{\partial q_1} &= \frac{\partial\mathbf{x}}{\partial r} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)\,;\\
\frac{\partial\mathbf{x}}{\partial q_2} &= \frac{\partial\mathbf{x}}{\partial\theta} = (r\cos\theta\cos\phi, r\cos\theta\sin\phi, -r\sin\theta)\,;\\
\frac{\partial\mathbf{x}}{\partial q_3} &= \frac{\partial\mathbf{x}}{\partial\phi} = (-r\sin\theta\sin\phi, r\sin\theta\cos\phi, 0)\,.
\end{aligned}$$

$$\begin{aligned}
h_1 = h_r &= \left|\frac{\partial\mathbf{x}}{\partial q_1}\right| = 1\,, & \mathbf{e}_1 = \mathbf{e}_r &= (\sin\theta\cos\phi,\ \sin\theta\sin\phi,\ \cos\theta)\,;\\
h_2 = h_\theta &= \left|\frac{\partial\mathbf{x}}{\partial q_2}\right| = r\,, & \mathbf{e}_2 = \mathbf{e}_\theta &= (\cos\theta\cos\phi,\ \cos\theta\sin\phi,\ -\sin\theta)\,;\\
h_3 = h_\phi &= \left|\frac{\partial\mathbf{x}}{\partial q_3}\right| = r\sin\theta\,, & \mathbf{e}_3 = \mathbf{e}_\phi &= (-\sin\phi,\ \cos\phi,\ 0)\,.
\end{aligned}$$

$$\mathrm{d}\mathbf{x} = \sum_j h_j\,\mathrm{d}q_j\,\mathbf{e}_j = \mathrm{d}r\,\mathbf{e}_r + r\,\mathrm{d}\theta\,\mathbf{e}_\theta + r\sin\theta\,\mathrm{d}\phi\,\mathbf{e}_\phi\,.$$

Remark. Note that the spherical polar coordinates are singular at $r = 0$, $\theta = 0$ and $\theta = \pi$.

1.7.4 Cylindrical Polar Coordinates

In cylindrical polar coordinates, $q_1 = \rho \in [0, \infty), q_2 = \phi \in [0, 2\pi), q_3 = z \in \mathbb{R}$, and in terms of Cartesian coordinates,

$$\mathbf{x} = (\rho \cos\phi, \rho \sin\phi, z)\,.$$

Proposition 1.39.

$$\begin{aligned}
\frac{\partial \mathbf{x}}{\partial q_1} &= \frac{\partial \mathbf{x}}{\partial \rho} = (\cos\phi, \sin\phi, 0)\,;\\
\frac{\partial \mathbf{x}}{\partial q_2} &= \frac{\partial \mathbf{x}}{\partial \phi} = (-\rho\sin\phi, \rho\cos\phi, 0)\,;\\
\frac{\partial \mathbf{x}}{\partial q_3} &= \frac{\partial \mathbf{x}}{\partial z} = (0, 0, 1)\,.
\end{aligned}$$

$$\begin{aligned}
h_1 = h_\rho &= \left|\frac{\partial \mathbf{x}}{\partial q_1}\right| = 1\,, & \mathbf{e}_1 = \mathbf{e}_\rho &= (\cos\phi, \sin\phi, 0)\,;\\
h_2 = h_\phi &= \left|\frac{\partial \mathbf{x}}{\partial q_2}\right| = \rho\,, & \mathbf{e}_2 = \mathbf{e}_\phi &= (-\sin\phi, \cos\phi, 0)\,;\\
h_3 = h_z &= \left|\frac{\partial \mathbf{x}}{\partial q_3}\right| = 1\,, & \mathbf{e}_3 = \mathbf{e}_z &= (0, 0, 1)\,.
\end{aligned}$$

$$\mathrm{d}\mathbf{x} = \sum_j h_j\,\mathrm{d}q_j\,\mathbf{e}_j = \mathrm{d}\rho\,\mathbf{e}_\rho + \rho\,\mathrm{d}\phi\,\mathbf{e}_\phi + \mathrm{d}z\,\mathbf{e}_z\,.$$

Remark. Note that cylindrical polar coordinates are singular at $\rho = 0$.

1.7.5 Volume and Surface Elements

Proposition 1.40. For orthogonal curvilinear coordinate systems, the volume element is given by

$$\mathrm{d}V = (h_1\,\mathrm{d}q_1\,\mathbf{e}_1)\cdot(h_2\,\mathrm{d}q_2\,\mathbf{e}_2)\times(h_3\,\mathrm{d}q_3\,\mathbf{e}_3) = h_1h_2h_3\,\mathrm{d}q_1\,\mathrm{d}q_2\,\mathrm{d}q_3\ .$$

- Spherical polar coordinates: $\mathrm{d}V = r^2\sin\theta\,\mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}\phi$
- Cylindrical polar coordinates: $\mathrm{d}V = \rho\,\mathrm{d}\rho\,\mathrm{d}\phi\,\mathrm{d}z$

Proposition 1.41. For an orthogonal curvilinear coordinate system, with a surface normal parallel to a basis vector (e.g. $\mathrm{d}\mathbf{S} \parallel \mathbf{e}_3$),

$$\begin{aligned}
\mathrm{d}\mathbf{S} &= (h_1\,\mathrm{d}q_1\,\mathbf{e}_1)\times(h_2\,\mathrm{d}q_2\,\mathbf{e}_2)\\
&= h_1h_2\,\mathrm{d}q_1\,\mathrm{d}q_2\,\mathbf{e}_3\,.
\end{aligned}$$

1.7.6 Gradient in Orthogonal Curvilinear Coordinates

Theorem 1.42. The gradient in an orthogonal curvilinear coordinate system is given by

$$\boldsymbol{\nabla}\psi = \sum_i \frac{\mathbf{e}_i}{h_i}\frac{\partial\psi}{\partial q_i} = \left(\frac{1}{h_1}\frac{\partial\psi}{\partial q_1}, \frac{1}{h_2}\frac{\partial\psi}{\partial q_2}, \frac{1}{h_3}\frac{\partial\psi}{\partial q_3}\right),$$

with differential operator

$$\boldsymbol{\nabla} = \sum_i \mathbf{e}_i \frac{1}{h_i} \frac{\partial}{\partial q_i} .$$

- Spherical polar coordinates:

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

- Cylindrical polar coordinates:

$$\boldsymbol{\nabla} = \mathbf{e}_\rho \frac{\partial}{\partial \rho} + \mathbf{e}_\phi \frac{1}{\rho} \frac{\partial}{\partial \phi} + \mathbf{e}_z \frac{\partial}{\partial z}$$

Proof. The gradient of a scalar field, $\boldsymbol{\nabla}\psi$, is defined to be the vector such that for all $\mathrm{d}\mathbf{x}$,

$$\mathrm{d}\psi = \boldsymbol{\nabla}\psi \cdot \mathrm{d}\mathbf{x} .$$

We write

$$\boldsymbol{\nabla}\psi = \sum_i \mathbf{e}_i \alpha_i ,$$

where the coefficients α_i for the gradient operator in this coordinate system are to be determined. Then

$$\mathrm{d}\psi = \boldsymbol{\nabla}\psi \cdot \mathrm{d}\mathbf{x} = \left(\sum_i \mathbf{e}_i \alpha_i \right) \cdot \left(\sum_j h_j \mathbf{e}_j \, \mathrm{d}q_j \right) = \sum_i \alpha_i h_i \, \mathrm{d}q_i .$$

We must also have

$$\mathrm{d}\psi = \sum_i \frac{\partial \psi}{\partial q_i} \, \mathrm{d}q_i ,$$

so the coefficients α_i are

$$\alpha_i = \frac{1}{h_i} \frac{\partial \psi}{\partial q_i} .$$

□

Corollary.

$$\boldsymbol{\nabla} q_i = \sum_j \mathbf{e}_j \frac{1}{h_j} \frac{\partial q_i}{\partial q_j} = \sum_j \frac{\mathbf{e}_j}{h_j} \delta_{ij} = \frac{\mathbf{e}_i}{h_i} ,$$

so

$$\mathbf{e}_i = h_i \boldsymbol{\nabla} q_i .$$

1.7.7 Divergence and Curl in Orthogonal Curvilinear Coordinates

Theorem 1.43. The divergence of a vector field $\mathbf{F}$ in an orthogonal curvilinear coordinate system is given by

$$\boldsymbol{\nabla} \cdot \mathbf{F} = \frac{1}{h_1 h_2 h_3} \left(\frac{\partial}{\partial q_1}(h_2 h_3 F_1) + \frac{\partial}{\partial q_2}(h_3 h_1 F_2) + \frac{\partial}{\partial q_3}(h_1 h_2 F_3) \right) .$$

- Spherical polar coordinates:

$$\boldsymbol{\nabla} \cdot \mathbf{F} = \frac{1}{r^2} \frac{\partial}{\partial r}\left(r^2 F_r\right) + \frac{1}{r \sin\theta} \frac{\partial}{\partial \theta}(\sin\theta F_\theta) + \frac{1}{r \sin\theta} \frac{\partial F_\phi}{\partial \phi}$$

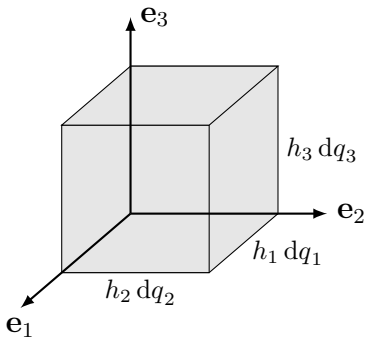
- Cylindrical polar coordinates:

$$\boldsymbol{\nabla} \cdot \mathbf{F} = \frac{1}{\rho} \frac{\partial}{\partial \rho}(\rho F_\rho) + \frac{1}{\rho} \frac{\partial F_\phi}{\partial \phi} + \frac{\partial F_z}{\partial z}$$

Proof.

$$\begin{aligned}\boldsymbol{\nabla}\cdot\mathbf{F} &= \boldsymbol{\nabla}\cdot\left(\sum_i F_i\mathbf{e}_i\right)\\ &= \boldsymbol{\nabla}\cdot\left((h_2h_3F_1)\left(\frac{\mathbf{e}_1}{h_2h_3}\right)\right) + \text{cyclic permutations}\\ &= \frac{\mathbf{e}_1}{h_2h_3}\cdot\boldsymbol{\nabla}(h_2h_3F_1) + h_2h_3F_1\boldsymbol{\nabla}\cdot\left(\frac{\mathbf{e}_1}{h_2h_3}\right) + \text{cyclic permutations}\\ &= \frac{\mathbf{e}_1}{h_2h_3}\cdot\sum_j\mathbf{e}_j\left(\frac{1}{h_j}\frac{\partial}{\partial q_j}(h_2h_3F_1)\right) + h_2h_3F_1\boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}q_2\times\boldsymbol{\nabla}q_3) + \text{cyclic permutations}\,,\end{aligned}$$

for which the latter terms vanish. □



Here is an easier way to interpret this result based on the alternative definition of divergence from Theorem 1.29. If we take an infinitesimal cuboid $\mathcal{V}$ of volume V at point (q_1, q_2, q_3) with sides parallel to the basis vectors $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_3$.

$$\boldsymbol{\nabla}\cdot\mathbf{F} = \lim_{V\to 0}\frac{1}{V}\oiint_{\mathcal{S}}\mathbf{F}\cdot\mathrm{d}S\,.$$

The volume of the cuboid is $V = h_1h_2h_3\delta q_1\delta q_2\delta q_3$. The area of the surfaces along directions $\mathbf{e}_i$ and $\mathbf{e}_j$ are given by $h_ih_j\delta q_i\delta q_j$. Therefore,

$$\begin{aligned}\oiint_{\mathcal{S}}\mathbf{F}\cdot\mathrm{d}\mathbf{S} &\approx [h_1h_2F_3(q_1,q_2,q_3+\delta q_3) - h_1h_2F_3(q_1,q_2,q_3)]\delta q_1\delta q_2 + \text{cyclic permutations}\\ &\approx \frac{\partial}{\partial q_3}(h_1h_2F_3)\delta q_1\delta q_2\delta q_3 + \text{cyclic permutations}\,.\end{aligned}$$

Dividing through the volume and taking the limit $V\to 0$ gives the formula of divergence as claimed.

Theorem 1.44. The curl of a vector field $\mathbf{F}$ in an orthogonal curvilinear coordinate system is given by

$$\boldsymbol{\nabla}\times\mathbf{F} = \frac{1}{h_1h_2h_3}\begin{vmatrix} h_1\mathbf{e}_1 & h_2\mathbf{e}_2 & h_3\mathbf{e}_3\\ \frac{\partial}{\partial q_1} & \frac{\partial}{\partial q_2} & \frac{\partial}{\partial q_3}\\ h_1F_1 & h_2F_2 & h_3F_3\end{vmatrix}\,.$$

- Spherical polar coordinates:

$$\boldsymbol{\nabla}\times\mathbf{F} = \frac{1}{r^2\sin\theta}\begin{vmatrix}\mathbf{e}_r & r\mathbf{e}_\theta & r\sin\theta\mathbf{e}_\phi\\ \frac{\partial}{\partial r} & \frac{\partial}{\partial\theta} & \frac{\partial}{\partial\phi}\\ F_r & rF_\theta & r\sin\theta F_\phi\end{vmatrix}$$

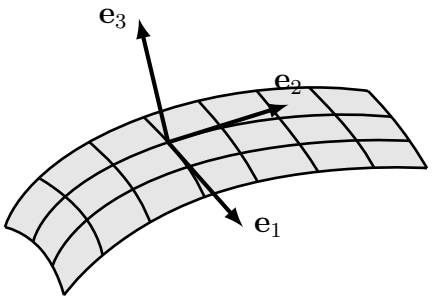
- Cylindrical polar coordinates:

$$\boldsymbol{\nabla}\times\mathbf{F}=\frac{1}{\rho}\begin{vmatrix}\mathbf{e}_\rho & \rho\mathbf{e}_\phi & \mathbf{e}_z\\ \frac{\partial}{\partial\rho} & \frac{\partial}{\partial\phi} & \frac{\partial}{\partial z}\\ F_\rho & \rho F_\phi & F_z\end{vmatrix}$$

Proof.

$$\begin{aligned}\boldsymbol{\nabla}\times\mathbf{F}&=\boldsymbol{\nabla}\times\left(\sum_i F_i\mathbf{e}_i\right)=\sum_i\boldsymbol{\nabla}\times\left((h_iF_i)\left(\frac{\mathbf{e}_i}{h_i}\right)\right)\\&=\sum_i\boldsymbol{\nabla}(h_iF_i)\times\frac{\mathbf{e}_i}{h_i}+\sum_i h_iF_i(\boldsymbol{\nabla}\times\boldsymbol{\nabla}q_i)\\&=\sum_i\sum_j\left(\frac{1}{h_ih_j}\frac{\partial(h_iF_i)}{\partial q_j}\right)\mathbf{e}_j\times\mathbf{e}_i\\&=\frac{\mathbf{e}_1}{h_2h_3}\left(\frac{\partial(h_3F_3)}{\partial q_2}-\frac{\partial(h_2F_2)}{\partial q_3}\right)+\frac{\mathbf{e}_2}{h_3h_1}\left(\frac{\partial(h_1F_1)}{\partial q_3}-\frac{\partial(h_3F_3)}{\partial q_1}\right)\\&\quad+\frac{\mathbf{e}_3}{h_1h_2}\left(\frac{\partial(h_2F_2)}{\partial q_1}-\frac{\partial(h_1F_1)}{\partial q_2}\right).\end{aligned}$$

□



Again, there is a more intuitive way to interpret this result based on Theorem 1.29. Take a surface $\mathcal{S}$ with normal $\hat{\mathbf{n}}=\mathbf{e}_i$ and area A at point (q_1,q_2,q_3), bounded by a rectangle with sides along $\mathbf{e}_j$ and $\mathbf{e}_k$. The component of $\boldsymbol{\nabla}\times\mathbf{F}$ along $\mathbf{e}_i$ is given by

$$\mathbf{e}_i\cdot(\boldsymbol{\nabla}\times\mathbf{F})=\lim_{A\to0}\frac{1}{A}\oint_{\mathcal{C}}\mathbf{F}\cdot\mathrm{d}\mathbf{x}\ .$$

The line integral can be approximated by

$$\begin{aligned}\oint_{\mathcal{C}}\mathbf{F}\cdot\mathrm{d}\mathbf{x}&\approx F_j(q_j,q_k)h_j\delta q_j+F_k(q_j+\delta q_j,q_k)h_k\delta q_k-F_j(q_j,q_k+\delta q_k)h_j\delta q_j-F_k(q_j,q_k)h_k\delta q_k\\&\approx\left[\frac{\partial}{\partial q_j}(h_kF_k)-\frac{\partial}{\partial q_k}(h_jF_j)\right]\delta q_j\delta q_k\,.\end{aligned}$$

Dividing by the area $A=h_jh_k\delta q_j\delta q_k$ and taking the limit $A\to0$ gives

$$\mathbf{e}_i\cdot(\boldsymbol{\nabla}\times\mathbf{F})=\frac{1}{h_jh_k}\left[\frac{\partial}{\partial q_j}(h_kF_k)-\frac{\partial}{\partial q_k}(h_jF_j)\right],$$

which is the components of $\boldsymbol{\nabla}\times\mathbf{F}$ as claimed.

1.7.8 Laplacian in Orthogonal Curvilinear Coordinates

Theorem 1.45. The Laplacian operator in an orthogonal curvilinear coordinate system is given by

$$\nabla^2=\frac{1}{h_1h_2h_3}\left(\frac{\partial}{\partial q_1}\left(\frac{h_2h_3}{h_1}\frac{\partial}{\partial q_1}\right)+\frac{\partial}{\partial q_2}\left(\frac{h_3h_1}{h_2}\frac{\partial}{\partial q_2}\right)+\frac{\partial}{\partial q_3}\left(\frac{h_1h_2}{h_3}\frac{\partial}{\partial q_3}\right)\right).$$

- Spherical polar coordinates:

$$\begin{aligned}\nabla^2\psi &= \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2\psi}{\partial\phi^2}\\ &= \frac{1}{r}\frac{\partial^2}{\partial r^2}(r\psi) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2\psi}{\partial\phi^2}\end{aligned}$$

- Cylindrical polar coordinates:

$$\nabla^2\psi = \frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial\psi}{\partial\rho}\right) + \frac{1}{\rho^2}\frac{\partial^2\psi}{\partial\phi^2} + \frac{\partial^2\psi}{\partial z^2}$$

Proof. We already have the divergence and gradient, so it is easy to work out.

$$\begin{aligned}\nabla^2\psi &= \boldsymbol{\nabla}\cdot\boldsymbol{\nabla}\psi\\ &= \frac{1}{h_1h_2h_3}\left(\frac{\partial}{\partial q_1}\left(\frac{h_2h_3}{h_1}\frac{\partial\psi}{\partial q_1}\right) + \frac{\partial}{\partial q_2}\left(\frac{h_3h_1}{h_2}\frac{\partial\psi}{\partial q_2}\right) + \frac{\partial}{\partial q_3}\left(\frac{h_1h_2}{h_3}\frac{\partial\psi}{\partial q_3}\right)\right).\end{aligned}$$

□

2 Green's Functions

2.1 The Dirac Delta Function

In the first part of this chapter, we will first try to define a mysterious object called the *Dirac delta function* which, technically, shouldn't even be called a 'function'.

2.1.1 Definition as Limit of Sequences

Consider the function $\delta_\epsilon(x)$ defined for $\epsilon > 0$ by

$$\delta_\epsilon(x) \coloneqq \begin{cases} 0 & x < -\epsilon \\ \frac{1}{2\epsilon} & -\epsilon \le x \le \epsilon \\ 0 & x > \epsilon\,, \end{cases}$$

then $\forall \epsilon > 0$,

$$\int_{-\infty}^{\infty} \delta_\epsilon(x)\,\mathrm{d}x = 1\,.$$

For any integrable function $f(x)$ and constant ξ,

$$\int_{-\infty}^{\infty} \delta_\epsilon(x-\xi) f(x)\,\mathrm{d}x = \frac{1}{2\epsilon}(F(\xi+\epsilon) - F(\xi-\epsilon))\,,$$

where F is the antiderivative of f. Then in the limit of $\epsilon \to 0^+$, we can recover

$$\begin{aligned} \lim_{\epsilon\to 0^+} \int_{-\infty}^{\infty} \delta_\epsilon(x-\xi) f(x)\,\mathrm{d}x &= \lim_{\epsilon\to 0^+} \frac{1}{2\epsilon}\Big(\quad F(\xi) + \epsilon f(\xi) + \frac{1}{2}\epsilon^2 f'(\xi) + \dots \\ &\qquad\qquad - F(\xi) + \epsilon f(\xi) - \frac{1}{2}\epsilon^2 f'(\xi) + \dots \quad \Big) \\ &= f(\xi)\,. \end{aligned}$$

This inspires us to make the following definition:

Definition 2.1. We can view the *Dirac delta function*, $\delta(x)$, as the limit as $\epsilon \to 0$ of $\delta_\epsilon(x)$:

$$\delta(x) \coloneqq \lim_{\epsilon\to 0^+} \delta_\epsilon(x)\,.$$

What are we doing here? When ϵ gets smaller, the function δ_ϵ has a higher and higher peak over a narrower and narrower range around $x = 0$. Although it is absolutely clear that neither putting $\epsilon = 0$ directly in the above sequence makes any sense, nor does the $\epsilon \to 0$ limit really converges to some well-defined function, what we are trying to do here is to create some object that has an infinitely high peak over an infinitely narrow range, and somehow carries a unit weight.

Of course, we can use another sequence of functions to make the same effect. For example, if we alternatively define $\delta_\epsilon(x)$ as

$$\delta_\epsilon(x) \coloneqq \frac{\epsilon}{\pi(x^2+\epsilon^2)}\,,$$

which gives

$$\int_{-\infty}^{\infty} \delta_\epsilon(x)\,\mathrm{d}x = 1\,.$$

Now δ_ϵ also has a peak of unit area centred at $x = 0$, which gets sharper as $\epsilon \to 0_+$. We can identify the Dirac delta function as the $\epsilon \to 0_+$ limit of this function as well, which also gives

$$\begin{aligned}\int_{-\infty}^{\infty} \delta(x-\xi) f(x)\,\mathrm{d}x &= \lim_{\epsilon\to 0^+} \int_{-\infty}^{\infty} \delta_\epsilon(x-\xi) f(x)\,\mathrm{d}x \\ &= f(\xi)\,.\end{aligned}$$

This sequence of functions even has the nice property that for any $\epsilon > 0$, δ_ϵ is smooth.

This definition of the delta function gives us an expression which turns out to be extremely useful later.

Proposition 2.2.

$$\delta(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx}\,\mathrm{d}k\ .$$

Proof. We note that

$$\begin{aligned}\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx-\epsilon|k|}\,\mathrm{d}k &= \frac{1}{2\pi}\left(\int_{-\infty}^{0} e^{ikx+\epsilon k}\,\mathrm{d}k + \int_0^{\infty} e^{ikx-\epsilon k}\,\mathrm{d}k\right) \\ &= \frac{\epsilon}{\pi(x^2+\epsilon^2)}\,.\end{aligned}$$

Take $\epsilon \to 0^+$. □

Of course we are doing something weird here. Putting $x = 0$, we get

$$\delta(0) \overset{?}{=} \frac{1}{2\pi}\int_{-\infty}^{\infty} 1\,\mathrm{d}x\ ,$$

which is a complete nonsense.

2.1.2 Properties

We may identify that the Dirac delta function has an infinitely sharp peak of zero width and has a unit area.

$$\delta(x) = \begin{cases}\infty & x = 0\\ 0 & x \neq 0\end{cases}\ ,$$

$$\int_a^b \delta(x)\,\mathrm{d}x = 1 \quad \forall a < 0, b > 0\,.$$

This is obviously not what we would normally call a function, but we can get a sense of what it is doing. It provides a surgical strike on the integrand to pick out its value at one particular point:

$$\int_{-\infty}^{\infty} \delta(x-\xi) f(x)\,\mathrm{d}x = f(\xi)\,.$$

This is the continuum analogue of the Kronecker delta.

Despite its absurdity, we can still derive some of its properties.

Proposition 2.3. $\delta(x)$ is symmetric.

Proof. By the substitution $k = -l$,

$$\delta(-x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ikx}\,\mathrm{d}k = -\frac{1}{2\pi}\int_{\infty}^{-\infty} e^{ilx}\,\mathrm{d}l = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ilx}\,\mathrm{d}l = \delta(x)\,.$$

□

Proposition 2.4. $\delta(x)$ is real.

Proof.

$$\delta^*(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ikx}\,\mathrm{d}k = \delta(-x) = \delta(x)\,.$$

□

2.1.3 Alternative Definition of the Dirac Delta Function

We have done a lot of whimsical things above that would certainly drive a mathematician crazy. The major issue of what we have done above is that we have encountered a lot of infinities — but there is one nice property of Dirac delta that stands out as it involves no infinity. This might help us to define the Dirac delta a bit more sensibly.

Definition 2.5. In an alternative (and better) view, $\delta(x)$ is defined as the *generalised function* (*distribution*) such that for all smooth functions $f(x)$,

$$\int_{-\infty}^{\infty} \delta(x-\xi)f(x)\,\mathrm{d}x = f(\xi)\,.$$

Remark. The upshot is that, in this way, $\delta(x)$ is defined within an integrand as a linear operator, and therefore should always be employed in an integrand. We should never consider taking it out of the integral.

2.1.4 Derivative of the Delta Function

We can be more brave and try to differentiate the Dirac delta. But to do this more legally, we will put it in an integral and see how it does to a well-behaved function. Using integration by parts, we see that

$$\int_{-\infty}^{\infty} \delta'(x-\xi)f(x)\,\mathrm{d}x = [\delta(x-\xi)f(x)]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \delta(x-\xi)f'(x)\,\mathrm{d}x = -f'(\xi)\,.$$

Definition 2.6. The derivative of $\delta(x)$ is defined as the generalised function such that for all differentiable functions $f(x)$,

$$\int_{-\infty}^{\infty} \delta'(x-\xi)f(x)\,\mathrm{d}x = -f'(\xi)\,.$$

Alternatively, the derivative of the delta function may again be defined as the limit of the derivatives of some sequence of functions, δ'_ϵ, similar to what we have done at the beginning of this section.

2.2 The Heaviside Step Function

Definition 2.7. The *Heaviside step function*, $H(x)$ is defined for $x \neq 0$ as

$$H(x) \coloneqq \begin{cases} 0 & x < 0 \\ 1 & x > 0\,. \end{cases}$$

There are various conventions for the value of the Heaviside step function at $x = 0$. It is not uncommon to choose $H(0) = \frac{1}{2}$. This is unimportant.

2.2.1 Properties

It is clear to see that this function is discontinuous at $x = 0$:

$$\lim_{x\to 0^-} H(x) = 0 \neq 1 = \lim_{x\to 0^+} H(x)\,.$$

Therefore, it seems that $H(x)$ is not differentiable at $x = 0$ — at least in a normal sense. However, we have just seen that $\delta(x)$ has a unit area within an infinitely narrow area near $x = 0$. This allows us to make the following identification.

Proposition 2.8.

$$H(x) = \int_{-\infty}^{x} \delta(\xi)\,\mathrm{d}\xi\ ,$$

and so we may identify

$$\frac{\mathrm{d}}{\mathrm{d}x}H(x) = \delta(x)\,.$$

Proof.

$$\begin{aligned}\int_{-\infty}^{\infty} H'(x-\xi)f(x)\,\mathrm{d}x &= [H(x-\xi)f(x)]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} H(x-\xi)f'(x)\,\mathrm{d}x\\ &= f(\infty) - \int_{\xi}^{\infty} f'(x)\,\mathrm{d}x\\ &= f(\xi)\,.\end{aligned}$$

□

2.3 Formal Theory of Distributions (Non-examinable)

After invented by physicist Dirac, the Dirac delta function quickly becomes ubiquitous in almost all fields in physics, especially in quantum mechanics and quantum field theory. Although we played around with infinities in very dangerous ways when defining this object, it just worked surprisingly well.

But mathematicians at that time were not satisfied with that. They were trying to find a rigorous mathematical theory to deal with objects like Dirac delta, and it was only achieved until mid-20$^{\text{th}}$ century when they invented the theory of distributions.

Of course we are not introducing distribution theory in this course — this is a very deep subject. Here, we will have a very brief look at some basic principles of it to get a sense of what a distribution really is.

2.3.1 Distributions

To define a distribution, we must first choose a class of *test functions*, which is the functions that our distribution will act on. For $\Omega \subseteq \mathbb{R}^n$, the simplest class of test functions are infinitely smooth functions $\phi \in C^\infty(\Omega)$ that have *compact support*, meaning that there exists a compact set $K \subset \Omega$ such that $\phi(\mathbf{x}) = 0$ whenever $x \notin K$. For our purposes, it is fine to just think of such functions to be the ones that take non-zero values only for a finite region in space. Let us denote the space of all test functions $\mathcal{D}(\Omega)$.

Definition 2.9. For a space of test functions $\mathcal{D}(\Omega)$, a *distribution* T is defined to be a linear map $T : \mathcal{D}(\Omega) \to \mathbb{R}$, given by

$$T : \phi \mapsto T[\phi]\,.$$

Remark. T is not a function on Ω itself but rather is a function on the infinite-dimensional space of test functions on Ω.

The space of distributions with test functions in $\mathcal{D}(\Omega)$ is denoted $\mathcal{D}'(\Omega)$. It is an infinite-dimensional vector space because we can add two distributions T_1 and T_2 together, defining the distribution $(T_1 + T_2)$ by

$$(T_1 + T_2)[\phi] \coloneqq T_1[\phi] + T_2[\phi]$$

for all $\phi \in \mathcal{D}(\Omega)$. We can multiply a distribution by a constant, defining the distribution (cT) by

$$(cT_1)[\phi] \coloneqq cT_1[\phi]$$

for all $\phi \in \mathcal{D}(\Omega)$ and $c \in \mathbb{R}$. We can also multiply distributions by smooth functions. If $\psi \in C^\infty(\Omega)$ and $T \in \mathcal{D}'(\Omega)$ then define the distribution (ψT) by

$$(\psi T)[\phi] \coloneqq T[\psi\phi]\,.$$

Remark. In general, there is no way to multiply two distributions together.

The simplest type of distribution is just an ordinary function $f : \Omega \to \mathbb{R}$ that is locally integrable, meaning that its integral over any compact set converges. To treat f as a distribution we must say how it acts on any test function $\phi \in \mathcal{D}(\Omega)$. For example, we can define

$$f[\phi] \coloneqq (f, \phi) = \int_\Omega f(x)\phi(x)\,\mathrm{d}V\ ,$$

which is the inner product of f with ϕ. This integral is guaranteed to be well-defined even when Ω is non-compact (say, the whole of $\mathbb{R}^n$) since ϕ has compact support and f is locally integrable. By definition, it is easy to check that $f[\phi]$ is a linear map from $\mathcal{D}(\Omega)$ to R. In the case where the generalised function is just an ordinary function, the map $T_f : D(\Omega) \to \mathbb{R}$ just corresponds to the usual inner product between functions.

The most important example of a generalised function that is not a function is the Dirac delta.

Definition 2.10. The *Dirac delta* δ is a distribution defined by

$$\delta[\phi] \coloneqq \phi(\mathbf{0})$$

for all $\phi \in \mathcal{D}(\Omega)$, where $\mathbf{0}$ is the origin in $\mathbb{R}^n$.

Proposition 2.11. $\delta : \mathcal{D}(\Omega) \to \mathbb{R}$ is a linear map.

Proof.

$$\delta[c_1\phi_1 + c_2\phi_2] = c_1\phi_1(\mathbf{0}) + c_2\phi_2(\mathbf{0})\,.$$

The addition on the left is in the vector space of test functions, while the addition on the right is addition in $\mathbb{R}$. □

By analogy with the case where the generalised function is itself a function, it is often convenient to abuse notation and write

$$T[\phi] = (T, \phi) = \int_\Omega T(\mathbf{x})\phi(\mathbf{x})\,\mathrm{d}V$$

even for general distributions that are not functions. However, for a general distribution the object $T(\mathbf{x})$ is not a function. There is no sense in which $T : \Omega \to \mathbb{R}$. For example, it is common to write

$$\delta[\phi] = \int_\Omega \delta(\mathbf{x})\phi(\mathbf{x})\,\mathrm{d}V$$

for some object $\delta(\mathbf{x})$. However, $\delta(\mathbf{x})$ cannot possibly be a genuine function. For the integral to be equal to $\phi(\mathbf{0})$, the value of $\delta(\mathbf{x})$ must vanish whenever $x \neq 0$. On the other hand, if $\delta(\mathbf{0})$ does indeed vanish everywhere except at one point, the integral cannot give the finite answer $\phi(\mathbf{0})$ if $\delta(\mathbf{x})$ takes any finite value at $\mathbf{x} = \mathbf{0}$. So it is not a genuine function in the sense of being a map from $\Omega \to \mathbb{R}$. One reason this abusive notation is convenient is that distributions can arise as the limit of a sequence of integrals of usual functions, just as what we have seen before.

2.3.2 Differentiation of Distributions

In the case that the distribution is just an ordinary function f, we have

$$\begin{aligned} T_{f'}[\phi] &= \int_\Omega f'(x)\phi(x)\,\mathrm{d}x \\ &= [f(x)\phi(x)]_\Omega - \int_\Omega f(x)\phi'(x)\,\mathrm{d}x \\ &= -T_f[\phi']\,, \end{aligned}$$

where the boundary term vanishes since ϕ has compact support inside Ω. Let us now define the derivative of a distribution.

Definition 2.12. The derivative of a generalised function T is defined as

$$T'[\phi] \coloneqq -T[\phi']$$

for all $\phi \in \mathcal{D}(\Omega)$.

Remark. The idea here is that if we think of our distribution as coming from the limit of a sequence of integrals involving only ordinary functions, this relation will hold for every member of the sequence, and so it will hold for the limiting value of the integrals.

Example. For the delta distribution, we have

$$\delta'[\phi] = -\delta[\phi'] = -\phi'(0)\,.$$

Proposition 2.13. The derivative of the Heaviside step function is the Dirac delta function.

Proof. $H(x)$ defines a generalised function on $\mathbb{R}$ by

$$H[\phi] = \int_{-\infty}^{\infty} H(x)\phi(x)\,\mathrm{d}x = \int_0^\infty \phi(x)\,\mathrm{d}x\ ,$$

which converges since ϕ has compact support. $H(x)$ is not differentiable, or even continuous as a function, but it is perfectly differentiable as a distribution. We have

$$\begin{aligned} H'[\phi] = -H[\phi'] &= -\int_{-\infty}^{\infty} H(x)\frac{\partial\phi}{\partial x}\,\mathrm{d}x \\ &= -\int_0^x \frac{\partial\phi}{\partial x}\,\mathrm{d}x \\ &= \phi(0) - \phi(\infty) = \phi(0)\,, \end{aligned}$$

since ϕ has compact support. Since $H'[\phi] = \phi(0) = \delta[\phi]$ holds for any test function ϕ, we can identify H' as the distribution δ. □

2.4 Second-order Linear Ordinary Differential Equations

One of the most important applications of the Dirac delta is in solving the differential equations.

A general second-order linear ODE for $y(x)$ can be written as

$$y'' + p(x)y' + q(x)y = f(x) \quad \text{or} \quad Ly(x) = f(x)\,,$$

where L is the differential operator

$$L \equiv \frac{\mathrm{d}^2}{\mathrm{d}x^2} + p(x)\frac{\mathrm{d}}{\mathrm{d}x} + q(x)\,.$$

If $f(x) = 0$, then the equation is *homogeneous*, otherwise it is *inhomogeneous*.

2.4.1 Homogeneous Second-order ODEs

Theorem 2.14 (The principle of superposition). If y_1 and y_2 are solutions to a homogeneous linear differential equation

$$Ly = 0\,,$$

where L is a linear differential operator, then they can be superposed to give a third. For any $\alpha, \beta \in \mathbb{R}$,

$$y = \alpha y_1 + \beta y_2$$

is also a solution.

Corollary. Suppose y_1 and y_2 are two linearly independent solutions, which means that

$$\alpha y_1(x) + \beta y_2(x) \equiv 0 \implies \alpha = \beta = 0\,.$$

Since the equation is second-order, the general solution will be of the form

$$y = \alpha y_1 + \beta y_2\,.$$

y_1 and y_2 are commonly referred to as *complementary functions*.

2.4.2 Inhomogeneous Second-order ODEs

Corollary. If $y_0(x)$ is any solution of the real inhomogeneous equation

$$Ly \equiv y'' + p(x)y' + q(x)y = f(x)\,,$$

then the general solution has the form

$$y(x) = y_0(x) + \alpha y_1(x) + \beta y_2(x)\,,$$

where y_1 and y_2 are any linearly independent solutions of $Ly = 0$. y_0 is referred to as a *particular solution*.

2.4.3 The Wronskian

If y_1 and y_2 are linearly dependent, then so are y_1' and y_2'. Hence y_1 and y_2 are linearly dependent only if the equation

$$\begin{pmatrix} y_1 & y_2 \\ y_1' & y_2' \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \mathbf{0}$$

has a non-zero solution for α and β. Conversely, non-zero functions y_1 and y_2 are linearly independent if and only if

$$\begin{pmatrix} y_1 & y_2 \\ y_1' & y_2' \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \mathbf{0} \implies \alpha = \beta = 0\,.$$

Definition 2.15. The *Wronskian*, $W(x)$, of the two solutions is defined to be

$$W[y_1, y_2] \coloneqq y_1 y_2' - y_2 y_1' \,.$$

Since $\mathsf{Ax} = \mathbf{0}$ has only the trivial solution iff $\det(\mathsf{A}) \neq 0$, we conclude that y_1 and y_2 are linearly independent iff

$$\begin{vmatrix} y_1 & y_2 \\ y_1' & y_2' \end{vmatrix} = y_1 y_2' - y_2 y_1' = W \neq 0 \,.$$

2.4.4 Initial-Value and Boundary-Value Problems

Two boundary conditions must be specified to fully determine the solution of a second-order ODE. The general form of a linear boundary condition at a point $x = a$ is

$$Ay(a) + By'(a) = E \,,$$

where A, B are not both zero. If $E = 0$ the boundary condition is said to be *homogeneous*.

If two boundary conditions are specified at the same point, then the problem is referred to as an *initial value problem*. If two conditions are specified at different points, then this is a *boundary value problem*.

2.5 Differential Equations containing Delta Functions

Consider the linear second-order ODE

$$\frac{\mathrm{d}^2 y}{\mathrm{d}x^2} + y = \delta(x) \,.$$

If x represent time, then this equation could represent the behaviour of a simple harmonic oscillator in response to an instantaneous force at $x = 0$ with unit impulse.

In regions $x < 0$ and $x > 0$ respectively, the right-hand side vanishes, so the general solution is given by a linear combination of $\cos x$ and $\sin x$

$$y = \begin{cases} \alpha_- \cos x + \beta_- \sin x & x < 0 \\ \alpha_+ \cos x + \beta_+ \sin x & x > 0 \,. \end{cases}$$

Since the general solution of a second-order ODE should contain only two arbitrary constants, it should be able to relate α_-, β_- with α_+, β_+.

Integrate the differential equation from $x = -\epsilon$ to $x = \epsilon$ to obtain

$$\int_{-\epsilon}^{\epsilon} \frac{\partial^2 y}{\partial x^2} \,\mathrm{d}x + \int_{-\epsilon}^{\epsilon} y(x) \,\mathrm{d}x = \int_{-\epsilon}^{\epsilon} \delta(x) \,\mathrm{d}x \,,$$

$$y'(\epsilon) - y'(-\epsilon) + \int_{-\epsilon}^{\epsilon} y(x) \,\mathrm{d}x = 1 \,.$$

Take the limit $\epsilon \to 0$, assume y is bounded,

$$\lim_{\epsilon \to 0} \int_{-\epsilon}^{\epsilon} y(x) \,\mathrm{d}x = 0 \,,$$

$$\implies \left[\frac{\mathrm{d}y}{\mathrm{d}x}\right] \coloneqq \lim_{\epsilon \to 0} \left[\frac{\mathrm{d}y}{\mathrm{d}x}\right]_{x=-\epsilon}^{x=\epsilon} = 1 \,.$$

Since there is only a finite jump in the derivative of y, y is continuous. The jump conditions are:

$$[y] = 0,\ \left[\frac{\mathrm{d}y}{\mathrm{d}x}\right] = 1 \text{ at } x = 0\,.$$

Applying these conditions, we obtain

$$\begin{cases} \alpha_+ - \alpha_- = 0 \\ \beta_+ - \beta_- = 1 \end{cases}.$$

Hence the general solution is

$$y = \begin{cases} \alpha\cos x + \beta\sin x & x < 0 \\ \alpha\cos x + (\beta+1)\sin x & x > 0\,. \end{cases}$$

2.6 Green's Functions

2.6.1 The Green's Function for Two-point Homogeneous Boundary-value Problems

Consider an ordinary differential equation

$$Ly(x) = f(x)\,,$$

where L is the general second-order linear differential operator in x:

$$L = \frac{\mathrm{d}^2}{\mathrm{d}x^2} + p(x)\frac{\mathrm{d}}{\mathrm{d}x} + q(x)$$

with p and q being continuous functions, under two homogeneous boundary conditions:

$$\begin{cases} Ay(a) + By'(a) = 0 \\ Cy(b) + Dy'(b) = 0\,. \end{cases}$$

Definition 2.16. The *Green's function*, $G(x;\xi)$, of a differential operator L for a given set of homogeneous boundary conditions is defined as the response of the system to forcing at a point ξ, such that

$$\mathcal{L}G(x;\xi) = \delta(x-\xi)\,,$$

subjected to homogeneous boundary conditions

$$\begin{cases} AG(a;\xi) + BG_x(a;\xi) = 0 \\ CG(b;\xi) + DG_x(b;\xi) = 0 \end{cases},$$

where

$$\mathcal{L} = \frac{\partial^2}{\partial x^2} + p(x)\frac{\partial}{\partial x} + q(x)$$
$$G_x(x;\xi) = \frac{\partial G}{\partial x}\,.$$

Theorem 2.17. The solution to the second-order linear differential equation $Ly = f$ is

$$y(x) = \int_a^b G(x;\xi)f(\xi)\,\mathrm{d}\xi\ ,$$

where $G(x;\xi)$ is the Green's function.

Remark. We may view Green's function as an inverse differential operator

$$Ly = f \implies y = L^{-1}f = \int_a^b \mathrm{d}\xi\, G(x;\xi)\ f(\xi)\,.$$

Proof. Our proposed solution satisfies both of the boundary conditions

$$Ay(a) + By'(a) = \int_a^b (AG(a;\xi) + BG_x(a;\xi))f(\xi)\,\mathrm{d}\xi = 0$$

$$Cy(b) + Dy'(b) = \int_a^b (CG(b;\xi) + DG_x(b;\xi))f(\xi)\,\mathrm{d}\xi = 0$$

and the inhomogeneous equation

$$\begin{aligned} Ly(x) &= \int_a^b \mathcal{L}G(x;\xi)f(\xi)\,\mathrm{d}\xi \\ &= \int_a^b \delta(x-\xi)f(\xi)\,\mathrm{d}\xi = f(x)\,. \end{aligned}$$

□

2.6.2 Properties of the Green's Functions

Lemma 2.18. G is continuous and there is a unit jump in $\frac{\partial G}{\partial x}$ at $x = \xi$.

Proof. Integrate both sides of

$$\mathcal{L}G(x;\xi) = \delta(x-\xi)$$

from $\xi - \epsilon$ to $\xi + \epsilon$ for $\epsilon > 0$ and consider the limit $\epsilon \to 0$:

$$\begin{aligned} 1 &= \lim_{\epsilon\to 0}\int_{\xi-\epsilon}^{\xi+\epsilon} \mathcal{L}G\,\mathrm{d}x \\ &= \lim_{\epsilon\to 0}\int_{\xi-\epsilon}^{\xi+\epsilon} \left(\frac{\partial^2 G}{\partial x^2} + p\frac{\partial G}{\partial x} + qG\right)\mathrm{d}x \\ &= \lim_{\epsilon\to 0}\int_{\xi-\epsilon}^{\xi+\epsilon} \frac{\partial}{\partial x}\left(\frac{\partial G}{\partial x} + pG\right)\mathrm{d}x + \lim_{\epsilon\to 0}\int_{\xi-\epsilon}^{\xi+\epsilon}\left(-\frac{\mathrm{d}p}{\mathrm{d}x}G + qG\right)\mathrm{d}x \\ &= \lim_{\epsilon\to 0}\left[\frac{\partial G}{\partial x} + pG\right]_{x=\xi-\epsilon}^{x=\xi+\epsilon} - \lim_{\epsilon\to 0}\int_{\xi-\epsilon}^{\xi+\epsilon}\left(\frac{\mathrm{d}p}{\mathrm{d}x} - q\right)G\,\mathrm{d}x\ . \end{aligned}$$

Suppose $G(x;\xi)$ is bounded near $x = \xi$, and since p and q are continuous, the latter term vanishes to give

$$\lim_{\epsilon\to 0}\left[\frac{\partial G}{\partial x} + pG\right]_{x=\xi-\epsilon}^{x=\xi+\epsilon} = 1\,.$$

This implies that the jump in the derivative of G is bounded, so G must be continuous. We conclude that

$$\lim_{\epsilon\to 0}\left[G(x;\xi)\right]_{x=\xi-\epsilon}^{x=\xi+\epsilon} = 0\,, \quad \lim_{\epsilon\to 0}\left[\frac{\partial G}{\partial x}\right]_{x=\xi-\epsilon}^{x=\xi+\epsilon} = 1\,.$$

□

2.6.3 Construction of the Green's Function

When $x \neq \xi$, G satisfies the homogeneous equation, so for both $x > \xi$ and $x < \xi$, we can express G in terms of solutions of the homogeneous equation. Suppose that $\{y_1, y_2\}$ are a basis of linearly independent solutions to the homogeneous equation $\mathcal{L}y = 0$ on $[a, b]$. We define this basis by requiring

$$\begin{cases} Ay_1(a) + By_1'(a) = 0 \\ Cy_2(b) + Dy_2'(b) = 0 \end{cases},$$

i.e. each of the solutions obeys one of the homogeneous boundary conditions. On $[a, \xi)$, the Green's function obeys $\mathcal{L}G = 0$ and

$$AG(a;\xi) + B\frac{\partial G}{\partial x}(a;\xi) = 0\,.$$

Since any homogeneous solution to $\mathcal{L}y = 0$ satisfying $Ay(a) + By'(a) = 0$ must be proportional to $y_1(x)$, with a proportionality constant independent of x. Thus we can set

$$G(x;\xi) = \alpha(\xi)y_1(x) \quad \text{for } x \in [a, \xi)\,.$$

Similarly on $(\xi, b]$ the Green's function must be proportional to $y_2(x)$ so we get

$$G(x;\xi) = \beta(\xi)y_2(x) \quad \text{for } x \in (\xi, b]\,.$$

Now we can determine how these constructions defined at $x \in [a, b] \setminus \{\xi\}$ can be joined together at $x = \xi$. From Lemma 2.18, we must have

$$\begin{cases} \beta(\xi)y_2(\xi) - \alpha(\xi)y_1(\xi) = 0 \\ \beta(\xi)y_2'(\xi) - \alpha(\xi)y_1'(\xi) = 1\,. \end{cases}$$

Rearranging gives

$$\begin{pmatrix} y_1 & y_2 \\ y_1' & y_2' \end{pmatrix}\begin{pmatrix} -\alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

So a solution exists if

$$W \equiv \begin{vmatrix} y_1 & y_2 \\ y_1' & y_2' \end{vmatrix} \neq 0\,,$$

which gives

$$\alpha(\xi) = \frac{y_2(\xi)}{W(\xi)} \quad \text{and} \quad \beta(\xi) = \frac{y_1(\xi)}{W(\xi)}\,.$$

Theorem 2.19. For a linear second-order differential operator L subjected to homogeneous boundary conditions, the Green's function is given by

$$G(x;\xi) = \begin{cases} \dfrac{y_1(x)y_2(\xi)}{W(\xi)} & \text{for } x \in [a, \xi) \\ \dfrac{y_1(\xi)y_2(x)}{W(\xi)} & \text{for } x \in [\xi, b]\,, \end{cases}$$

where y_1 and y_2 satisfy the boundary conditions at a and b respectively.

Example. Solve

$$y''(x) + y(x) = f(x), \quad y(0) = y(1) = 0\,.$$

The complementary functions satisfying left and right boundary conditions are

$$y_1 = \sin x,\ y_2 = \sin(x - 1)\,,$$

and the Wronskian is

$$W = y_1 y_2' - y_2 y_1' = \sin x \cos(x-1) - \sin(x-1)\cos x = \sin 1 \,.$$

Thus,

$$G(x;\xi) = \begin{cases} \dfrac{\sin x \sin(\xi - 1)}{\sin 1} & 0 \le x \le \xi \\ \dfrac{\sin \xi \sin(x-1)}{\sin 1} & \xi \le x \le 1 \,. \end{cases}$$

The solution is given by

$$\begin{aligned} y(x) &= \int_0^1 G(x;\xi) f(\xi)\,\mathrm{d}\xi \\ &= \frac{\sin(x-1)}{\sin 1}\int_0^x \sin\xi f(\xi)\,\mathrm{d}\xi + \frac{\sin x}{\sin 1}\int_x^1 \sin(\xi-1) f(\xi)\,\mathrm{d}\xi \,. \end{aligned}$$

2.6.4 The Green's Function for Homogeneous Initial-Value Problems

Suppose the boundary conditions are instead

$$y(a) = y'(a) = 0 \,.$$

For $x \in [a, \xi)$, choose the complementary functions such that $y_1(a) = 0$ and $y_2'(a) = 0$, and the Green's function is given by

$$G(x;\xi) = \alpha(\xi) y_1(x) + \beta(\xi) y_2(x) \,.$$

Apply boundary conditions to the Green's function, then we get

$$\begin{cases} \alpha y_1(a) + \beta y_2(a) = 0 \\ \alpha y_1'(a) + \beta y_2'(a) = 0 \end{cases} \implies \begin{cases} \alpha = 0 \\ \beta = 0 \,. \end{cases}$$

This implies that

$$G = 0 \quad \text{for } x \in [a, \xi) \,.$$

For $x \in [\xi, \infty)$, we again write the Green's function as

$$G(x;\xi) = \lambda(\xi) y_1(x) + \mu(\xi) y_2(x) \,.$$

Apply Lemma 2.18 at $x = \xi$, we get

$$\begin{cases} \lambda y_1(\xi) + \mu y_2(\xi) = 0 \\ \lambda y_1'(\xi) + \mu y_2'(\xi) = 1 \,, \end{cases}$$

which organises to

$$\begin{pmatrix} y_1(\xi) & y_2(\xi) \\ y_1'(\xi) & y_2'(\xi) \end{pmatrix} \begin{pmatrix} \lambda \\ \mu \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} ,$$

with solutions

$$\lambda = -\frac{y_2(\xi)}{W(\xi)} \quad \text{and} \quad \mu = \frac{y_1(\xi)}{W(\xi)} \,.$$

Theorem 2.20. For a linear second-order differential operator L subjected to homogeneous initial conditions, Green's function is given by

$$G(x;\xi) = \begin{cases} 0 & \text{for } a \le x < \xi \\ \dfrac{y_1(\xi) y_2(x) - y_1(x) y_2(\xi)}{W(\xi)} & \text{for } x \ge \xi \,. \end{cases}$$

where y_1 and y_2 satisfy the boundary conditions $y(a) = 0$ and $y'(a) = 0$ respectively.

2.6.5 Inhomogeneous Boundary Conditions

To solve an inhomogeneous equation under inhomogeneous boundary conditions

$$Ly = f\,,$$

first solve the homogeneous equation $Ly = 0$ for the inhomogeneous boundary conditions, which gives a solution y_{ibc}.

Then solve the inhomogeneous equation $Ly = f$ for the homogeneous boundary conditions (perhaps using Green's functions) to give a solution y_{hbc}.

By linearity, $y = y_{\text{ibc}} + y_{\text{hbc}}$ satisfies the inhomogeneous equation with inhomogeneous boundary conditions.

3 Fourier Transforms

3.1 Fourier Transforms

When we have a periodic function, we express it in terms of a Fourier series. In this chapter, we want to generalise this result to non-periodic functions.

Definition 3.1. For a suitably well-behaved function $f : \mathbb{R} \to \mathbb{C}$, its *Fourier transform* $\tilde{f} : \mathbb{R} \to \mathbb{C}$ is defined as

$$\mathcal{F}[f(x)] \equiv \tilde{f}(k) \coloneqq \int_{-\infty}^{\infty} e^{-ikx} f(x)\,\mathrm{d}x\ .$$

Remark. You may encounter a lot of different conventions when defining the Fourier transform. Sometimes there will be a normalising factor $\frac{1}{\sqrt{2\pi}}$, sometimes the $-ikx$ will be replaced by $+ikx$ — or sometimes even $\pm 2\pi ikx$. We will stick to the most common convention in the mathematical community defined above.

The Fourier transform transforms between complex-valued functions, so the transform of a real function does not necessarily remain real.

Proposition 3.2. If $f(x)$ is both real and even, then $\tilde{f}$ is real.

Proof.

$$\begin{aligned}\tilde{f}^*(k) &= \int_{-\infty}^{\infty} e^{ikx} f^*(x)\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} e^{ikx} f(-x)\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} e^{-iky} f(y)\,\mathrm{d}y \\ &= \tilde{f}(k)\end{aligned}$$

□

Proposition 3.3. If $f(x)$ is both real and odd, then $\tilde{f}$ is purely imaginary.

Proof. Similar to above. □

Remark. A necessary condition for $\tilde{f}(k)$ to exist (as a normal function, not as a distribution) for all real values of k is that $f(x) \to 0$ as $x \to \pm\infty$. Otherwise, the Fourier integral does not converge (e.g. for $k = 0$).

A set of sufficient conditions for $\tilde{f}(k)$ to exist is that $f(x)$ have bounded variation, a finite number of discontinuities and be absolutely integrable, i.e.

$$\int_{-\infty}^{\infty} |f(x)|\,\mathrm{d}x < \infty\,.$$

3.1.1 Examples of Fourier Transforms

(i) $e^{-b|x|}$, $b > 0$.

$$\begin{aligned}\mathcal{F}\left[e^{-b|x|}\right] &= \int_{-\infty}^{\infty} e^{-ikx-b|x|}\,\mathrm{d}x \\ &= \frac{2b}{k^2+b^2}\end{aligned}$$

(ii) $\cos(ax)e^{-b|x|}$.

$$\begin{aligned}\mathcal{F}\Big[\cos(ax)e^{-b|x|}\Big] &= \frac{1}{2}\int_{-\infty}^{\infty}(e^{iax}+e^{-iax})e^{-ikx-b|x|}\,\mathrm{d}x\\ &= b\bigg(\frac{1}{(a-k)^2+b^2}+\frac{1}{(a+k)^2+b^2}\bigg)\end{aligned}$$

(iii) $\sin(ax)e^{-b|x|}$.

$$\begin{aligned}\mathcal{F}\Big[\sin(ax)e^{-b|x|}\Big] &= \frac{1}{2i}\int_{-\infty}^{\infty}(e^{iax}-e^{-iax})e^{-ikx-b|x|}\,\mathrm{d}x\\ &= -ib\bigg(\frac{1}{(a-k)^2+b^2}-\frac{1}{(a+k)^2+b^2}\bigg)\end{aligned}$$

(iv) Gaussian.

$$\begin{aligned}\mathcal{F}\bigg[\frac{1}{\sqrt{2\pi}\epsilon}e^{-\frac{x^2}{2\epsilon^2}}\bigg] &= \frac{1}{\sqrt{2\pi}\epsilon}\int_{-\infty}^{\infty}\exp\bigg(-\frac{x^2}{2\epsilon^2}-ikx\bigg)\,\mathrm{d}x\\ &= \frac{1}{\sqrt{2\pi}\epsilon}\int_{-\infty}^{\infty}\exp\bigg(-\frac{1}{2}\Big(\frac{x}{\epsilon}+i\epsilon k\Big)^2-\frac{1}{2}\epsilon^2k^2\bigg)\,\mathrm{d}x\\ &= \frac{1}{\sqrt{2\pi}}\exp\bigg(-\frac{1}{2}\epsilon^2k^2\bigg)\int_{-\infty}^{\infty}\exp\bigg(-\frac{1}{2}y^2\bigg)\,\mathrm{d}y \qquad \text{substitution } x=\epsilon y-i\epsilon^2k\\ &= \exp\bigg(-\frac{1}{2}\epsilon^2k^2\bigg)\end{aligned}$$

Remark. The Fourier transform of a Gaussian of width ϵ is a Gaussian of width ϵ^{-1}.

(v) Dirac delta function.

$$\begin{aligned}\mathcal{F}[\delta(x-a)] &= \int_{-\infty}^{\infty}\delta(x-a)e^{-ikx}\,\mathrm{d}x\\ &= e^{-ika}\end{aligned}$$

Hence the Fourier transform of $\delta(x)$ is 1. Recall that the Dirac delta function can be considered as the limit of a Gaussian as $\epsilon\to 0^+$.

(vi) Constant function.

$$\mathcal{F}[a] = a\int_{-\infty}^{\infty}e^{-ikx}\,\mathrm{d}x\ .$$

This clearly violates our previous claim that in order to have a Fourier transform in the sense of a normal function, we must have $f(x)\to 0$ as $x\to\pm\infty$. Therefore, this Fourier transform only exists in the sense of a distribution. Recall the expression of the delta function in Proposition 2.2 — this is exactly what we have here. We can identify

$$\int_{-\infty}^{\infty}e^{-ikx} = 2\pi\delta(k)\,,$$

so

$$\mathcal{F}[a] = 2\pi a\delta(k)\,.$$

(vii) Heaviside step function. A direct Fourier transform is problematic:

$$\begin{aligned}\mathcal{F}[H(x-a)] &= \int_{-\infty}^{\infty}H(x-a)e^{-ikx}\,\mathrm{d}x\\ &= \int_{a}^{\infty}e^{-ikx}\\ &= \bigg[\frac{e^{-ikx}}{-ik}\bigg]_a^{\infty},\end{aligned}$$

but the limit $\lim_{x\to\infty} e^{-ikx}$ does not exist. This is another case that the function is not absolutely integrable, so we need a convergent regularisation of it.

We may first find the Fourier transform of $H(x-a)e^{-\epsilon(x-a)}$, then take the limit $\epsilon \to 0^+$. Doing so, we have

$$\begin{aligned}\mathcal{F}[H(x-a)e^{-\epsilon(x-a)}] &= \int_{-\infty}^{\infty} H(x-a)e^{-\epsilon(x-a)-ikx}\,\mathrm{d}x \\ &= \left[\frac{e^{-\epsilon(x-a)-ikx}}{-\epsilon-ik}\right]_a^{\infty} \\ &= \frac{e^{-ika}}{\epsilon+ik}\end{aligned}$$

For any $k \neq 0$, we are safe to ignore the ϵ in the denominator as we take the $\epsilon \to 0^+$ limit, and therefore we have

$$\mathcal{F}[H(x-a)] = \lim_{\epsilon\to 0^+} \mathcal{F}[H(x-a)e^{-\epsilon(x-a)}] = \frac{e^{-ika}}{ik}\,.$$

However, when $k = 0$, we are not allowed to do so. We have to work out the Fourier transform at this point separately. We have

$$\mathcal{F}[H(x-a)](0) = \int_a^{\infty} 1\,\mathrm{d}x = \frac{1}{2}\int_{-\infty}^{\infty} \mathrm{d}x\ .$$

This integral does not converge, but it is again exactly the expression of delta function we met in Proposition 2.2 with $x = 0$, so

$$\mathcal{F}[H(x-a)](0) = \pi\delta(0)\,.$$

Combining the above results, we have

$$\mathcal{F}[H(x-a)] = \frac{e^{-ika}}{ik} + \pi\delta(k)\,,$$

where we used the property that $\delta(k) = 0$ for $k \neq 0$.

If you think the proof above makes no sense, you absolutely are right. This result must be interpreted in the sense of a distribution, and is better proven using something called the Sokhotski–Plemelj identity in distribution theory. We are of course not doing this here.

Remark. $ik\mathcal{F}[H(x-a)] = \mathcal{F}[\delta(x-a)]$.

(viii) Top-hat function, $g(x)$, defined by

$$g(x) = \begin{cases} c & a < x < b \\ 0 & \text{otherwise}\,. \end{cases}$$

$$\tilde{g}(k) = \int_a^b ce^{-ikx}\,\mathrm{d}x = \frac{ic}{k}(e^{-ikb} - e^{-ika})$$

For instance, if $a = -1$, $b = 1$, $c = 1$,

$$\tilde{g}(k) = \frac{2\sin k}{k}\,.$$

3.2 Fourier Inversion Theorem

The nice thing about the Fourier transform is that you can transform it back easily.

Theorem 3.4 (Fourier inversion theorem). The inverse Fourier transform acting on $\tilde{f}(k)$ that recovers $f(x)$ is given by

$$\mathcal{I}[\tilde{f}] \coloneqq \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx}\tilde{f}(k)\,\mathrm{d}k = f(x)\,.$$

Proof.

$$\begin{aligned}\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx}\tilde{f}(k)\,\mathrm{d}k &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx}\left(\int_{-\infty}^{\infty} e^{-iks}f(s)\,\mathrm{d}s\right)\mathrm{d}k \\ &= \int_{-\infty}^{\infty}\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ik(x-s)}\,\mathrm{d}k\right)f(s)\,\mathrm{d}s \qquad \text{swap integration order} \\ &= \int_{-\infty}^{\infty} f(s)\delta(x-s)\,\mathrm{d}s \\ &= f(x)\,.\end{aligned}$$

□

Corollary. If $g(k) = \tilde{f}(k)$, then $\tilde{g}(k) = 2\pi f(-k)$.

We can take advantage of this when calculating the Fourier series.

Example. Find the Fourier transform of $(x^2+b^2)^{-1}$.

We have worked out

$$\mathcal{F}\left[e^{-b|x|}\right] = \frac{2b}{k^2+b^2}\,.$$

Applying the observation above, we get

$$\mathcal{F}\left[\frac{1}{x^2+b^2}\right](k) = \frac{\pi}{b}e^{-b|k|}\,.$$

3.3 Properties of Fourier Transforms

Proposition 3.5. The Fourier transform has the following properties.

(i) *Linearity.* For constants $\alpha, \beta \in \mathbb{C}$,

$$\mathcal{F}[\alpha f(x) + \beta g(x)] = \alpha\mathcal{F}[f(x)] + \beta\mathcal{F}[g(x)]\,.$$

(ii) *Rescaling.* For real constant $\alpha \in \mathbb{R}$,

$$\mathcal{F}[f(\alpha x)] = \frac{1}{|\alpha|}\tilde{f}\left(\frac{k}{\alpha}\right).$$

(iii) *Translation.* For real constant $\alpha \in \mathbb{R}$,

$$\mathcal{F}[f(x-\alpha)] = e^{-ik\alpha}\mathcal{F}[f(x)]$$

(iv) *Exponential.* For constant $\alpha \in \mathbb{C}$,

$$\mathcal{F}[e^{i\alpha x}f(x)](k) = \mathcal{F}[f(x)](k-\alpha)\,.$$

(v) *Duality.* If $g(x) = \tilde{f}(x)$, then
$$\tilde{g}(k) = 2\pi f(-k)\,.$$

(vi) *Complex conjugation and parity inversion.* For $k \in \mathbb{R}$,
$$\mathcal{F}[f^*](k) = \mathcal{F}[f](-k)^*\,.$$

(vii) *Symmetry.* If $f(-x) = \pm f(x)$ i.e. f is even or odd, then
$$\tilde{f}(-k) = \pm\tilde{f}(k)\,.$$

(viii) *Differentiation.*
$$\mathcal{F}\left[\frac{\mathrm{d}^n f}{\mathrm{d}x^n}\right] = (ik)^n \tilde{f}\,.$$

(ix) *Multiplication by x.*
$$\mathcal{F}[xf(x)] = i\frac{\mathrm{d}\tilde{f}}{\mathrm{d}k}\,.$$

Remark. Fourier transforms allow a simple representation of derivatives of $f(x)$ in Fourier space. This has important consequences for solving differential equations.

Proof.

(i) Trivial by the linearity of multiplication and integration.

(ii) Let $g(x) = f(\alpha x)$,
$$\begin{aligned}\tilde{g}(k) &= \int_{-\infty}^{\infty} e^{-ikx} f(\alpha x)\,\mathrm{d}x \\ &= \frac{\operatorname{sgn}\,\alpha}{\alpha}\int_{-\infty}^{\infty} e^{-i\frac{k}{\alpha}y} f(y)\,\mathrm{d}y \\ &= \frac{1}{|\alpha|}\tilde{f}\left(\frac{k}{\alpha}\right).\end{aligned}$$

(iii)
$$\begin{aligned}\mathcal{F}[f(x-\alpha)] &= \int_{-\infty}^{\infty} e^{-ikx} f(x-\alpha)\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} e^{-ik(y+\alpha)} f(y)\,\mathrm{d}y \\ &= e^{-ik\alpha}\mathcal{F}[f(x)]\,.\end{aligned}$$

(iv)
$$\begin{aligned}\mathcal{F}[e^{i\alpha x} f(x)](k) &= \int_{-\infty}^{\infty} e^{-i(k-\alpha)x} f(x)\,\mathrm{d}x \\ &= \mathcal{F}[f(x)](k-\alpha)\,.\end{aligned}$$

(v)
$$\begin{aligned}\tilde{g}(k) &= \int_{-\infty}^{\infty} e^{-ikx}\tilde{f}(x)\,\mathrm{d}x \\ &= 2\pi f(-k)\,.\end{aligned}$$

(vi)

$$\begin{aligned}\mathcal{F}[f^*](k) &= \int_{-\infty}^{\infty} e^{-ikx} f^*(x)\,\mathrm{d}x \\ &= \left(\int_{-\infty}^{\infty} e^{ikx} f(x)\,\mathrm{d}x\right)^* \\ &= \mathcal{F}[f](-k)^*\,.\end{aligned}$$

(vii)

$$\begin{aligned}\tilde{f}(-k) &= \int_{-\infty}^{\infty} f(x)e^{ikx} \\ &= \int_{-\infty}^{\infty} \pm f(-x)e^{ikx}\,\mathrm{d}x \\ &= \pm\int_{-\infty}^{\infty} f(y)e^{-iky}\,\mathrm{d}y \\ &= \pm\tilde{f}(k)\,.\end{aligned}$$

(viii) By differentiating the inverse Fourier theorem, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}x}f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx}\Big(ik\tilde{f}(k)\Big)\,\mathrm{d}k = \mathcal{I}\Big[ik\tilde{f}\Big]\,.$$

Fourier transform this equation, and we obtain

$$\mathcal{F}\left[\frac{\mathrm{d}f}{\mathrm{d}x}\right] = \mathcal{F}\Big[\mathcal{I}\Big[ik\tilde{f}\Big]\Big] = ik\tilde{f}\,,$$

and hence

$$\mathcal{F}\left[\frac{\mathrm{d}^n f}{\mathrm{d}x^n}\right] = (ik)^n\tilde{f}\,.$$

Alternatively, here is another proof.

$$\begin{aligned}\mathcal{F}\left[\frac{\mathrm{d}f}{\mathrm{d}x}\right] &= \int_{-\infty}^{\infty} f'(x)e^{-ikx}\,\mathrm{d}x \\ &= \big[f(x)e^{-ikx}\big]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} -ikf(x)e^{-ikx}\,\mathrm{d}x \\ &= ik\tilde{f}(k)\,,\end{aligned}$$

where the former part vanishes because $f(x) \to 0$ as $x \to \pm\infty$ for the Fourier transform to converge.

(ix) Differentiate the Fourier transform with respect to k, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}k}\tilde{f}(k) = \int_{-\infty}^{\infty} e^{-ikx}(-ixf(x))\,\mathrm{d}x\ .$$

After multiplying by i, we can deduce that

$$\mathcal{F}[xf(x)] = i\frac{\mathrm{d}\tilde{f}}{\mathrm{d}k}\,.$$

□

3.4 Fourier Series

We claimed that the Fourier transform is the generalisation of Fourier series to non-periodic functions. We will illustrate this connection here.

Lemma 3.6 (Fourier series). Suppose that $f : \mathbb{R} \to \mathbb{C}$ is a periodic function with period L, then f can be represented by a *Fourier series*

$$f(x) = \sum_{n=-\infty}^{\infty} a_n \exp\left(\frac{2\pi i n x}{L}\right),$$

where

$$a_n = \frac{1}{L}\int_{-\frac{1}{2}L}^{\frac{1}{2}L} f(x)\exp\left(-\frac{2\pi i n x}{L}\right)\mathrm{d}x\ .$$

As $L \to \infty$ ($f(x)$ becomes non-periodic), the increment between successive wavenumbers in its Fourier series, $\Delta k = \frac{2\pi}{L}$, becomes vanishingly small. Therefore, the spectrum for allowed wavenumbers k_n becomes a continuum.

Rewrite the formula of the Fourier series as

$$f(x) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty} \tilde{f}(k_n)\exp(ixk_n)\Delta k$$

$$\tilde{f}(k_n) = \int_{-\frac{1}{2}L}^{\frac{1}{2}L} f(x)\exp(-ixk_n)\,\mathrm{d}x\ ,$$

where

$$\tilde{f}(k_n) = La_n = \frac{2\pi a_n}{\Delta k}\,.$$

We then see that in the limit $\Delta k \to 0$ and $L \to \infty$

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{f}(k)\exp(ikx)\,\mathrm{d}k$$

$$\tilde{f}(k) = \int_{-\infty}^{\infty} f(x)\exp(-ikx)\,\mathrm{d}x\ .$$

3.5 Convolution

3.5.1 Definition of Convolution

Definition 3.7. The *convolution*, $f * g$, of a function $f(x)$ with another function $g(x)$ is defined by

$$(f * g)(x) \coloneqq \int_{-\infty}^{\infty} f(y)g(x-y)\,\mathrm{d}y\ .$$

Remark. The convolution expresses the amount of overlap of one function g as it is shifted over another function f.

Proposition 3.8. The convolution operator is commutative.

$$f * g = g * f$$

Proof. Let $z = x - y$.

$$\begin{aligned} f * g(x) &= \int_{-\infty}^{\infty} f(y)g(x-y)\,\mathrm{d}y \\ &= -\int_{+\infty}^{-\infty} f(x-z)g(z)\,\mathrm{d}z \\ &= \int_{-\infty}^{\infty} g(z)f(x-z)\,\mathrm{d}z\ . \end{aligned}$$

□

3.5.2 Interpretation

Proposition 3.9. If x and y are two random variables with probability densities $f(x)$ and $g(y)$. Let the distribution of their sum, $z = x + y$, be $h(z)$. The probability density function of z is given by

$$h = f * g\,.$$

Proof. For any given value of x, the probability that z lies in the range

$$z_0 < z < z_0 + \delta z$$

is the probability that y lies in the range

$$z_0 - x < y < z_0 - x + \delta_z\,,$$

which is $g(z_0 - x)\delta z$. Therefore, the probability that z lies in the same range for all x is

$$h(z_0)\delta z = \int_{-\infty}^{\infty} f(x)g(z_0 - x)\delta z\,\mathrm{d}x\ .$$

This implies that $h = f * g$. □

Remark. The effect of measuring, observing or processing scientific data can often be described as a convolution between the data with certain functions. For instance, the gravitational potential of an object

$$\Phi(\mathbf{x}) = -G\int \frac{\rho(\mathbf{y})}{|\mathbf{x}-\mathbf{y}|}\,\mathrm{d}\mathbf{y}$$

is the convolution of the mass density $\rho(\mathbf{x})$ with the potential of a point mass $-\frac{G}{|\mathbf{x}|}$.

3.5.3 The Convolution Theorem

Theorem 3.10 (The convolution theorem). If the functions f and g have Fourier transforms $\mathcal{F}[f]$ and $\mathcal{F}[g]$ respectively, then

$$\mathcal{F}[f * g] = \mathcal{F}[f]\mathcal{F}[g]\,.$$

Proof.

$$\begin{aligned} \mathcal{F}[f*g] &= \int_{-\infty}^{\infty} e^{-ikx}\left(\int_{-\infty}^{\infty} f(y)g(x-y)\,\mathrm{d}y\right)\mathrm{d}x \\ &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} e^{-ikx}g(x-y)\,\mathrm{d}x\right)f(y)\,\mathrm{d}y && \text{swap integration order} \\ &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} e^{-ik(z+y)}g(z)\,\mathrm{d}z\right)f(y)\,\mathrm{d}y && \text{substitution } z = x-y \\ &= \int_{-\infty}^{\infty} f(y)e^{-iky}\left(\int_{-\infty}^{\infty} e^{-ikz}g(z)\,\mathrm{d}z\right)\mathrm{d}y \\ &= \mathcal{F}[f]\mathcal{F}[g] \end{aligned}$$

□

Corollary. Conversely, the Fourier transform of the product fg is given by the convolution of Fourier transforms of f and g divided by 2π.

$$\mathcal{F}[fg] = \frac{1}{2\pi}\mathcal{F}[f] * \mathcal{F}[g]\,.$$

Proof.

$$\begin{aligned}
\mathcal{F}[fg](k) &= \int_{-\infty}^{\infty} e^{-ikx} f(x)g(x)\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty} e^{-ikx}\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ilx}\tilde{f}(l)\,\mathrm{d}l\right) g(x)\,\mathrm{d}x \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{f}(l)\left(\int_{-\infty}^{\infty} e^{-i(k-l)x} g(x)\,\mathrm{d}x\right)\mathrm{d}l \qquad \text{swap integration order} \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{f}(l)\tilde{g}(k-l)\,\mathrm{d}l \\
&= \frac{1}{2\pi}(\tilde{f} * \tilde{g})(k)
\end{aligned}$$

□

Remarks.

- Convolution is an operation best carried out as a multiplication in the Fourier domain.
- Convolution can be undone (deconvolution) by a division in the Fourier domain.

Example. Suppose a linear 'black box' has an output $G(\omega)\exp(i\omega t)$ for a periodic input $\exp(i\omega t)$. What is the output $r(t)$ corresponding to input $f(t)$?

Express the input as a Fourier transform:

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{i\omega t}\,\mathrm{d}\omega\ .$$

Then, since the 'black box' is linear, we can directly superpose the frequency space of the input to produce the output:

$$\begin{aligned}
r(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} G(\omega)F(\omega)e^{i\omega t}\,\mathrm{d}\omega \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} (\mathcal{F}[f * g])e^{i\omega t}\,\mathrm{d}\omega \\
&= (f * g)(t)\,.
\end{aligned}$$

3.6 Correlation

Definition 3.11. The *correlation* of two functions, $h = f \otimes g$, is defined by

$$h(x) = f(x) \otimes g(x) \coloneqq \int_{-\infty}^{\infty} f(y)^* g(x+y)\,\mathrm{d}y\ .$$

This is a way of quantifying the relationship between two oscillatory functions. If two signals oscillating about an average value of zero are in phase, their correlation will be positive. If they are in opposite phases, the correlation will be negative. If they are completely unrelated, their correlation will be zero.

Lemma 3.12. The Fourier transform of a correlation is given by

$$\mathcal{F}[f(x) \otimes g(x)] = [\tilde{f}(k)]^* \tilde{g}(k)\,.$$

Proof.

$$\begin{aligned}
\mathcal{F}[f(x) \otimes g(x)] &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(y)^* g(x+y)\,\mathrm{d}y \right) e^{-ikx}\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y)^* g(z) e^{iky} e^{-ikz}\,\mathrm{d}z\,\mathrm{d}y \qquad \text{substitution } z = x+y \\
&= \left[\int_{-\infty}^{\infty} f(y) e^{-iky}\,\mathrm{d}y \right]^* \int_{-\infty}^{\infty} g(z) e^{-ikz}\,\mathrm{d}z \\
&= [\tilde{f}(k)]^* \tilde{g}(k)\,.
\end{aligned}$$

□

Definition 3.13. The quantity

$$\Phi(k) = \left| \tilde{f}(k) \right|^2$$

is the *power spectrum* (*power spectral density*) of the function $f(x)$.

Theorem 3.14 (Wiener–Khinchin theorem). The Fourier transform of the autocorrelation of a function is its power spectrum.

$$\mathcal{F}[f \otimes f](k) = \left| \tilde{f}(k) \right|^2$$

Proof. The special case of Lemma 3.12 when $g(x) = f(x)$. □

Remark. The spectrum of a perfectly periodic signal consists of a series of delta functions at the principal frequency and its harmonics (if present). Its autocorrelation does not decay as $t \to \infty$.

White noise is an ideal random signal with an autocorrelation function proportional to $\delta(t)$: the signal is perfectly decorrelated and therefore has a flat spectrum ($\Phi = \text{const.}$).

3.7 Parseval's Theorem

Theorem 3.15 (Parseval's theorem). Fourier transform is a *unitary transform* that preserves the inner product between two functions up to a multiplicative constant.

$$\int_{-\infty}^{\infty} [f(x)]^* g(x)\,\mathrm{d}x = \frac{1}{2\pi} \int_{-\infty}^{\infty} [\tilde{f}(k)]^* \tilde{g}(k)\,\mathrm{d}k\ .$$

Proof. Apply inverse Fourier transform to Lemma 3.12 to obtain

$$\int_{-\infty}^{\infty} [f(y)]^* g(x+y)\,\mathrm{d}y = \frac{1}{2\pi} \int_{-\infty}^{\infty} [\tilde{f}(k)]^* \tilde{g}(k) e^{ikx}\,\mathrm{d}k\ .$$

Set $x = 0$ and relabel $y \to x$ to obtain Parseval's theorem

$$\int_{-\infty}^{\infty} [f(x)]^* g(x)\,\mathrm{d}x = \frac{1}{2\pi} \int_{-\infty}^{\infty} [\tilde{f}(k)]^* \tilde{g}(k)\,\mathrm{d}k\ .$$

□

Corollary. The special case is used most frequently when $g = f$:

$$\int_{-\infty}^{\infty} |f(x)|^2\,\mathrm{d}x = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \tilde{f}(k) \right|^2 \mathrm{d}k\ .$$

Alternative proof.

$$\begin{aligned}\int_{-\infty}^{\infty} |f(x)|^2 \,\mathrm{d}x &= \int_{-\infty}^{\infty} f(x) f^*(x) \,\mathrm{d}x \\ &= \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} e^{ikx} \tilde{f}(k) \,\mathrm{d}k \right) \left(\int_{-\infty}^{\infty} e^{-ilx} \tilde{f}^*(l) \,\mathrm{d}l \right) \mathrm{d}x \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i(k-l)x} \,\mathrm{d}x \right) \tilde{f}^*(l) \tilde{f}(k) \,\mathrm{d}l \,\mathrm{d}k \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{f}^*(l) \delta(k-l) \,\mathrm{d}l \, \tilde{f}(k) \,\mathrm{d}k \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{f}(k) \tilde{f}^*(k) \,\mathrm{d}k = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \tilde{f}(k) \right|^2 \mathrm{d}k\end{aligned}$$

□

Example. Heisenberg's Principle of Uncertainty.

Suppose that

$$\psi(x) = \frac{1}{(2\pi\Delta_x^2)^{\frac{1}{4}}} \exp\left(-\frac{x^2}{4\Delta_x^2} \right)$$

is a real wave function. Then

$$\left|\psi^2(x)\right| = \frac{1}{\sqrt{2\pi\Delta_x^2}} \exp\left(-\frac{x^2}{2\Delta_x^2} \right)$$

is the probability density of finding the particle at position x, and Δ_x is the root mean square deviation in position.

There is a unit probability of finding the particle since $\left|\psi^2\right|$ is a Gaussian of width Δ_x and

$$\int_{-\infty}^{\infty} \left|\psi^2(x)\right| \mathrm{d}x = \frac{1}{\sqrt{2\pi\Delta_x^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2\Delta_x^2} \right) \mathrm{d}x = 1 \,.$$

The Fourier transform of the wave function gives

$$\begin{aligned}\tilde{\psi}(k) &= \left(8\pi\Delta_x^2\right)^{\frac{1}{4}} \exp\left(-\Delta_x^2 k^2\right) \\ &= \left(\frac{2\pi}{\Delta_k^2} \right)^{\frac{1}{4}} \exp\left(-\frac{k^2}{4\Delta_k^2} \right),\end{aligned}$$

where $\Delta_k = \frac{1}{2\Delta_x}$. $\tilde{\psi}^2$ is another Gaussian with root mean square deviation in wavenumber of Δ_k. In agreement with Parseval's theorem, it has an area of 2π.

Therefore, in the case of a Gaussian wave packet, $\Delta_k \Delta_x = \frac{1}{2}$. More generally, for any wavefunction $\psi(x)$,

$$\Delta_k \Delta_x \geq \frac{1}{2} \,.$$

In quantum mechanics, the momentum of a particle is given by $p = \hbar k$. Therefore, if we interpret $\Delta x = \Delta_x$ and $\Delta p = \hbar \Delta_k$ to be the uncertainty in the particle's position and momentum, we can obtain Heisenberg's Uncertainty Principle

$$\Delta p \Delta x \geq \frac{1}{2} \hbar \,.$$

3.8 Solution of Ordinary Differential Equations using Fourier Transforms

Suppose $\psi(x)$ satisfies

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2} - a^2\psi = -f(x)\,,$$

where a is a constant and f is a known function. Suppose also that ψ satisfies the boundary conditions $|\psi| \to 0$ as $|x| \to \pm\infty$ (required for the Fourier transform to converge).

If we multiply the LHS by $\exp(-ikx)$ and integrate over x, then we obtain

$$\begin{aligned}\int_{-\infty}^{\infty} e^{-ikx}\left(\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2} - a^2\psi\right)\mathrm{d}x &= \mathcal{F}\left[\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2}\right] - a^2\mathcal{F}[\psi] \\ &= -k^2\mathcal{F}[\psi] - a^2\mathcal{F}[\psi]\,.\end{aligned}$$

The same action on the RHS yields $-\mathcal{F}[f]$. Hence, by taking the Fourier transform of the whole equation we have

$$-k^2\mathcal{F}[\psi] - a^2\mathcal{F}[\psi] = -\mathcal{F}[f]\,,$$

and rearrangement gives

$$\mathcal{F}[\psi] = \frac{\mathcal{F}[f]}{k^2 + a^2}\,.$$

Taking the inverse Fourier transform, we obtain the solution

$$\psi = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx}\frac{\mathcal{F}[f]}{k^2 + a^2}\,\mathrm{d}k\,.$$

We will explore this technique in much greater detail in section 13.

4 Linear Algebra

4.1 Vector Spaces

Definition 4.1. A *vector space* over a field $\mathbb{F}$ is a non-empty set V together with

- a binary operation, *vector addition* $V \times V \to V$, $(\mathbf{u}, \mathbf{v}) \mapsto \mathbf{u} + \mathbf{v}$,
- a binary function, *scalar multiplication* $\mathbb{F} \times V \to V$, $(\lambda, \mathbf{v}) \mapsto \lambda\mathbf{v}$,

that satisfy the eight axioms listed below. The elements of V are called *vectors*, and the elements of $\mathbb{F}$ are called *scalars*.

The eight axioms satisfied for every $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $\lambda, \mu \in \mathbb{F}$ are

(V1) the vector addition is *associative*

$$(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})\,;$$

(V2) the vector addition is *commutative*

$$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}\,;$$

(V3) there exists a *null vector*, or *zero vector*, $\mathbf{0} \in V$ such that,

$$\mathbf{v} + \mathbf{0} = \mathbf{v}\,;$$

(V4) for every $\mathbf{v} \in V$ there exists a *negative vector*, or *inverse vector*, $-\mathbf{v} \in V$ such that

$$\mathbf{v} + (-\mathbf{v}) = 0\,;$$

(V5) the scalar multiplication is *compatible with field multiplication*

$$\lambda(\mu\mathbf{v}) = (\lambda\mu)\mathbf{v}\,;$$

(V6) for the multiplicative identity $1 \in \mathbb{F}$,

$$1\mathbf{v} = \mathbf{v}\,;$$

(V7) the scalar multiplication is *distributive* with respect to vector addition

$$\lambda(\mathbf{u} + \mathbf{v}) = \lambda\mathbf{u} + \lambda\mathbf{v}\,;$$

(V8) the scalar multiplication is *distributive* with respect to field addition

$$(\lambda + \mu)\mathbf{u} = \lambda\mathbf{u} + \mu\mathbf{u}\,.$$

Remarks.

- The field $\mathbb{F}$ is commonly $\mathbb{R}$ or $\mathbb{C}$ — this will always be the case in our course.
- The zero vector $\mathbf{0}$ is unique.
- The additive inverse of a vector $\mathbf{v}$ is unique.
- The existence of a negative vector allows us to define the subtraction of vectors

$$\mathbf{u} - \mathbf{v} \equiv \mathbf{u} + (-\mathbf{v})\,.$$

- Vector multiplication is not defined in general.

Example. The basic example of a vector space is $\mathbb{F}^n$. An element of $\mathbb{F}^n$ is an ordered list of n scalars, $(x_1, \ldots, x_n)$, where $x_i \in \mathbb{F}$, called an n-tuple. Vector addition and scalar multiplication are defined component-wise:

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n)$$
$$\alpha(x_1, \ldots, x_n) = (\alpha x_1, \ldots, \alpha x_n)$$

4.1.1 Span and Linear Independence

Definition 4.2. Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ be a subset of vectors in V. A *linear combination* of S is any vector of the form

$$a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_m\mathbf{u}_m \equiv a_i\mathbf{u}_i\,,$$

where $a_1, a_2, \ldots, a_m \in \mathbb{F}$.

Definition 4.3. The *span* of S is the set of all vectors that are linear combinations of S, written as

$$\text{span}(S) \equiv \langle S \rangle\,.$$

Definition 4.4. A set of m non-zero vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ is *linearly independent* if

$$a_i\mathbf{u}_i = 0 \implies a_i = 0\,.$$

Otherwise, the vectors are *linearly dependent*. There exists scalars a_i, at least one of which is non-zero, such that

$$a_i\mathbf{u}_i = 0\,.$$

4.1.2 Basis and Dimension

Definition 4.5. $S = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is a *basis* of the vector space V if it is linearly independent and spans V.

Lemma 4.6. The set of vectors $S = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ form a basis of V if and only if for all vectors $\mathbf{v} \in V$, there exists a unique set of scalars $v_i \in \mathbb{F}$ such that

$$\mathbf{v} = v_i\mathbf{e}_i\,.$$

The v_i are said to be the *components* of $\mathbf{v}$ with respect to the basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$.

Proof. ($\Rightarrow$): Since $\{\mathbf{e}_i\}$ span V, there exists $\{v_i\}$ such that

$$\mathbf{v} = v_i\mathbf{e}_i$$

for all $\mathbf{v} \in V$. Suppose also

$$\mathbf{v} = w_i\mathbf{e}_i\,,$$

then the difference

$$\sum (w_i - v_i)\mathbf{e}_i = \mathbf{0}\,.$$

Since $\{\mathbf{e}_i\}$ are linearly independent, $w_i = v_i\,\forall\, i$. The expression is therefore unique.

($\Leftarrow$): By assumption, $\{\mathbf{e}_i\}$ span V. Suppose that

$$v_i\mathbf{e}_i = \mathbf{0}\,,$$

and since $\mathbf{0} = \sum_i 0 \cdot \mathbf{e}_i$, by the uniqueness of $\mathbf{0}$, $v_i = 0$ for all i. $\{\mathbf{e}_i\}$ is a basis. □

Lemma 4.7 (Steinitz Exchange Lemma). Let V be a finite-dimensional vector space. Take $\mathbf{u}_1, \ldots, \mathbf{u}_m$ to be linearly independent, $\mathbf{v}_1, \ldots, \mathbf{v}_n$ to span V, then

(i) $m \leq n$

(ii) reordering the $\mathbf{v}_i$ if needed, $\{\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{v}_{m+1}, \ldots, \mathbf{v}_n\}$ spans V.

Theorem 4.8. If V is a finite-dimensional vector space over $\mathbb{F}$, then any two bases of V have the same cardinality, which is called the *dimension* of V, denoted as

$$\dim_{\mathbb{F}} V\,.$$

Proof. Suppose that $\{\mathbf{u}_1,\ldots,\mathbf{u}_n\}$ and $\{\mathbf{v}_1,\ldots,\mathbf{v}_m\}$ are both bases of V. Since $\{\mathbf{u}_i\}$ are linearly independent and $\{\mathbf{v}_i\}$ span V, $n \leq m$ by Lemma 4.7. Similarly $m \leq n$. Therefore, the bases of V have the same cardinality. □

Remarks. For a set of vectors, $\{\mathbf{u}_1,\ldots,\mathbf{u}_m\}$, in an n-dimensional vector space:

- If $m < n$ then there exists a vector that cannot be expressed as a linear combination of $\mathbf{u}_i$.
- If $m > n$ then there exists some vector that, when expressed as a linear combination of $\mathbf{u}_i$, has a non-unique scalar coefficient, whether or not the $\mathbf{u}_i$ span V.
- Vector spaces can have infinite dimensions, e.g. the function defined on $0 \leq x < 2\pi$ with a Fourier series
$$f(x) = \sum_{-\infty}^{\infty} f_n e^{inx},.$$
Here $f(x)$ is the vector and f_n are its components with respect to its basis of functions e^{inx}.

Examples.

(i) *3D Euclidean space* $\mathbb{E}^3$. In this case the scalars are real and V is three dimensional.

(ii) *The complex numbers.* Here we have two possibilities.

Suppose we are considering a complex linear vector space (linear vector space over $\mathbb{C}$). Then every complex number z can be written uniquely as

$$z = \alpha \cdot 1 \text{ where } \alpha \in \mathbb{C}\,.$$

and moreover,

$$\alpha \cdot 1 = 0 \implies \alpha = 0 \text{ for } \alpha \in \mathbb{C}\,.$$

We conclude that the single vector $\{1\}$ constitutes a basis for $\mathbb{C}$.

We might alternatively consider the complex numbers as a linear vector space over $\mathbb{R}$, so the scalars are real. Then the pair of vectors $\{1, i\}$ constitute a basis since every complex number z can be written uniquely as

$$z = \alpha \cdot 1 + \beta \cdot i \text{ where } \alpha, \beta \in \mathbb{R}\,,$$

and

$$\alpha \cdot 1 + \beta \cdot i = 0 \implies \alpha = \beta = 0 \text{ if } \alpha, \beta \in \mathbb{R}\,.$$

Thus we have

$$\dim_{\mathbb{C}} \mathbb{C} = 1 \text{ and } \dim_{\mathbb{R}} \mathbb{C} = 2\,.$$

Remarks.

- $\mathbb{R}^3$ is not the same as the physical space because the physical space has a rule for the distance between two points (metric).
- $\mathbb{R}^2$ is not the same as $\mathbb{C}$ because $\mathbb{C}$ has a rule for multiplication.

4.2 Vector Subspace and Direct Sum (Non-examinable)

Definition 4.9. Suppose V is a vector space over $\mathbb{F}$. A subset $U \subseteq V$ is a *subspace* over $\mathbb{F}$ if

(i) for all $\mathbf{u}_1, \mathbf{u}_2 \in U$, $\mathbf{u}_1 + \mathbf{u}_2 \in U$.

(ii) for all $\lambda \in \mathbb{F}$ and $\mathbf{u} \in U$, $\lambda u \in U$.

(iii) $\mathbf{0} \in U$.

Remark. We say U is a *proper subspace* of V to exclude the cases $U = \{\mathbf{0}\}$ and $U = V$.

Definition 4.10. Suppose that V is a vector space over $\mathbb{F}$ and U, W are subspaces of V. The *sum* of U and W is defined to be the set

$$U + W \coloneqq \{\mathbf{u} + \mathbf{w} \mid \mathbf{u} \in U, \mathbf{w} \in W\}.$$

Definition 4.11. We say that V is the *(internal) direct sum* of U and W, written as $V = U \oplus W$, if

$$V = U + W \text{ and } U \cap W = 0.$$

Remark. Equivalently, $V = U \oplus W$ if every element $\mathbf{v} \in V$ can be written uniquely as $\mathbf{u} + \mathbf{w}$ with $\mathbf{u} \in U$ and $\mathbf{w} \in W$.

Definition 4.12. Given any two vector spaces U and W over $\mathbb{F}$, the *(external) direct sum*, $U \oplus W$ of U and W is defined to be the set of pairs

$$\{(u, w) \mid \mathbf{u} \in U, w \in \mathbf{W}\}$$

with addition given by

$$(\mathbf{u}_1, \mathbf{w}_1) + (\mathbf{u}_2, \mathbf{w}_2) = (\mathbf{u}_1 + \mathbf{u}_2, \mathbf{w}_1 + \mathbf{w}_2)$$

and scalar multiplication given by

$$\lambda(\mathbf{u}, \mathbf{w}) = (\lambda\mathbf{u}, \lambda\mathbf{w})\,.$$

More generally, we can make the following definitions.

Definition 4.13. If $U_1, U_2, \ldots, U_n$ are subspaces of V, then V is the *(internal) direct sum* of $U_1, \ldots, U_n$, denoted as

$$V = U_1 \oplus \cdots \oplus U_n = \bigoplus_{i=1}^{n} U_i\,,$$

if every element $\mathbf{v}$ in V can be written uniquely as

$$\mathbf{v} = \sum_{i=1}^{n} \mathbf{u}_i$$

with $\mathbf{u}_1 \in U_i$.

Definition 4.14. If $U_1, \ldots, U_n$ are vector spaces over $\mathbf{F}$, their *(external) direct sum* is the vector space

$$\bigoplus_{i=1}^{n} U_i \coloneqq \{(\mathbf{u}_1, \ldots, \mathbf{u}_n) \mid \mathbf{u}_i \in U_i\}$$

with coordinate-wise operations.

4.3 Matrices

4.3.1 Linear Maps

Definition 4.15. Suppose that U and V are vector spaces over a field $\mathbb{F}$. A function $\mathcal{A} : U \to V$ is a *linear map* if

(i) $\mathcal{A}(\mathbf{u}_1 + \mathbf{u}_2) = \mathcal{A}(\mathbf{u}_1) + \mathcal{A}(\mathbf{u}_2)$ for all $\mathbf{u}_1, \mathbf{u}_2 \in U$.

(ii) $\mathcal{A}(\lambda\mathbf{u}) = \lambda\mathcal{A}(\mathbf{u})$ for all $\mathbf{u} \in U$ and $\lambda \in \mathbb{F}$.

We say U is the *domain* of $\mathcal{A}$ and V is the *codomain* of $\mathcal{A}$. We denote the vector space of linear maps from U to V as $\mathcal{L}(U, V)$.

Definition 4.16. For a linear map $\mathcal{A} : U \to V$,

- The *image* of $\mathcal{A}$,
$$\operatorname{Im}\mathcal{A} \coloneqq \{\mathcal{A}(\mathbf{u}) \mid \mathbf{u} \in U\}\,.$$
- The *kernel* of $\mathcal{A}$,
$$\ker\mathcal{A} \coloneqq \{\mathbf{u} \in U \mid \mathcal{A}(\mathbf{u}) = \mathbf{0}\}\,.$$

Remark. A linear map, or linear operator, has an existence without reference to any basis.

Proposition 4.17. For a linear map $\mathcal{A} : U \to V$, $\ker\mathcal{A}$ is a subspace of U and $\operatorname{Im}\mathcal{A}$ is a subspace of V.

Proof. First of all, $\mathbf{0}$ is in both $\ker\mathcal{A}$ and $\operatorname{Im}\mathcal{A}$. For all $\lambda, \mu \in \mathbb{F}$ and $\mathbf{u}_1, \mathbf{u}_2 \in \ker\mathcal{A}$,

$$\mathcal{A}(\lambda\mathbf{u}_1 + \mu\mathbf{u}_2) = \lambda\mathcal{A}(\mathbf{u}_1) + \mu\mathcal{A}(\mathbf{u}_2) = \mathbf{0} + \mathbf{0} = \mathbf{0}\,,$$

so $\ker\mathcal{A}$ is a subspace of U. Similarly, for $\mathbf{u}_1, \mathbf{u}_2 \in U$

$$\lambda\mathcal{A}(\mathbf{u}_1) + \mu\mathcal{A}(\mathbf{u}_2) = \mathcal{A}(\lambda\mathbf{u}_1 + \mu\mathbf{u}_2) \in \operatorname{Im}\mathcal{A}\,.$$

□

Definition 4.18. Suppose that $\mathcal{A} : U \to V$ is a linear map between finite dimensional vector spaces.

- The number $n(\mathcal{A}) \coloneqq \dim\ker\mathcal{A}$ is called the *nullity* of $\mathcal{A}$.
- The number $r(\mathcal{A}) \coloneqq \dim\operatorname{Im}\mathcal{A}$ is called the *rank* of $\mathcal{A}$.

4.3.2 Matrix Representations of Linear Operators

Proposition 4.19. Suppose that U and V are vector spaces over $\mathbb{F}$ and let S be a basis for U: $S \coloneqq \{\mathbf{e}_1, \dots \mathbf{e}_n\}$. Every function $f : S \to V$ extends uniquely to a linear map $\mathcal{A} : U \to V$.

Proof. First we prove uniqueness: suppose that $f : S \to V$, and $\mathcal{A}$ and $\mathcal{B}$ are two linear maps $U \to V$ extending f. Let $\mathbf{u} \in U$ so that $\mathbf{u} = \sum_i u_i\mathbf{e}_i$ for some $u_i \in \mathbb{F}$. Then

$$\begin{aligned}\mathcal{A}(\mathbf{u}) &= \mathcal{A}\left(\sum_{i=1}^{n} u_i\mathbf{e}_i\right)\\ &= \sum_{i=1}^{n} u_i\mathcal{A}(\mathbf{e}_i)\,.\end{aligned}$$

Similarly,

$$\mathcal{B}(\mathbf{u}) = \sum_{i=1}^{n} u_i \mathcal{B}(\mathbf{e}_i)\,.$$

Since $\mathcal{A}(\mathbf{e}_i) = f(\mathbf{e}_i) = \mathcal{B}(\mathbf{e}_i)$ for each i, we see that $\mathcal{A}(\mathbf{u}) = \mathcal{B}(\mathbf{u})$ for all $\mathbf{u} \in U$ and so $\mathcal{A} \equiv \mathcal{B}$.

That argument also shows us how to construct a linear map $\mathcal{A}$ that extends f. Every $\mathbf{u} \in U$ can be written uniquely as $\mathbf{u} = \sum_i u_i \mathbf{e}_i$ with $u_i \in \mathbb{F}$. Thus we can define $\mathcal{A}(\mathbf{u}) = \sum_i \mathbf{u}_i f(\mathbf{e}_i)$ without ambiguity. It remains to show that $\mathcal{A}$ is linear. We compute for $\mathbf{u} = \sum_i u_i \mathbf{e}_i$ and $\mathbf{v} = \sum_i v_i \mathbf{e}_i$,

$$\begin{aligned}
\mathcal{A}(\lambda\mathbf{u} + \mu\mathbf{v}) &= \mathcal{A}\left(\sum_{i=1}^{n} (\lambda \mathbf{u}_i + \mu \mathbf{v}_i)\mathbf{e}_i\right) \\
&= \sum_{i=1}^{n} (\lambda u_i + \mu v_i) f(\mathbf{e}_i) \\
&= \lambda \sum_{i=1}^{n} u_i f(\mathbf{e}_i) + \mu \sum_{i=1}^{n} v_i f(\mathbf{e}_i) \\
&= \lambda \mathcal{A}(\mathbf{u}) + \mu \mathcal{A}(\mathbf{v})\,.
\end{aligned}$$

□

Remark. To define a linear map, it suffices to specify its values on a basis.

Corollary. If U and V are finite dimensional vector spaces over $\mathbb{F}$ with ordered bases $(\mathbf{e}_1, \dots, \mathbf{e}_m)$ and $(\mathbf{f}_1, \dots, \mathbf{f}_n)$ respectively then there is a bijection

$$\mathrm{Mat}_{n\times m}(\mathbb{F}) \leftrightarrow \mathcal{L}(U, V)$$

that sends a $n \times m$ matrix A to the unique linear map $\mathcal{A}$ such that $\mathcal{A}(\mathbf{e}_i) = \sum_j A_{ji} \mathbf{f}_j$.

Remark. The i-th column of the matrix A tells where the i-th basis vector of U goes as a linear combination of the basis vectors of V.

Corollary. The sum of two linear operators is defined by

$$(\mathcal{A} + \mathcal{B})\mathbf{x} = \mathcal{A}\mathbf{x} + \mathcal{B}\mathbf{x} = \mathbf{e}_i (A_{ij} + B_{ij}) \mathbf{x}_j\,.$$

The product, or composition, of two linear operators has the action

$$\mathcal{A}\mathcal{B}\mathbf{x} = \mathcal{A}(\mathcal{B}\mathbf{x}) = \mathcal{A}(\mathbf{e}_k B_{kj} x_j) = (\mathcal{A}\mathbf{e}_k) B_{kj} x_j = \mathbf{e}_i A_{ik} B_{kj} x_j\,.$$

The components therefore satisfy the rules of matrix addition and multiplication:

$$(\mathsf{A} + \mathsf{B})_{ij} = A_{ij} + B_{ij} \quad (\mathsf{AB})_{ij} = A_{ik} B_{kj}\,.$$

Note that $\mathcal{A}\mathcal{B} \neq \mathcal{B}\mathcal{A}$ in general, and matrix multiplication is not commutative.

Remark. A matrix is the components of the linear operator with respect to a given basis.

We will focus on linear maps that transform within the same vector space.

Definition 4.20. Let V be a finite dimensional vector space over $\mathbb{F}$. An *endomorphism* of V is a linear map $\mathcal{A} : V \to V$.

Remark. The endomorphisms of V form a vector space, which is usually denoted by $\mathrm{End}(V)$.

4.3.3 Transformation Matrices

Let $\{\mathbf{e}_i\}_{i=1}^n$ and $\{\mathbf{e}'_i\}_{i=1}^n$ be two sets of basis vectors for an n-dimensional vector space V over $\mathbb{F}$. Since $\{\mathbf{e}_i\}_{i=1}^n$ is a basis of V, each basis vector $\mathbf{e}'_i$ can be written in the $\{\mathbf{e}_i\}$ basis as

$$\mathbf{e}'_j = \mathbf{e}_i A_{ij}$$

for some numbers A_{ij}. A_{ij} is the i^{th} component of the vector $\mathbf{e}'_j$ in the basis $\{\mathbf{e}_i\}_{i=1}^n$.

Proposition 4.21. The numbers A_{ij} can be represented by a $n \times n$ transformation matrix A

$$\mathsf{A} = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{pmatrix},$$

where the j^{th} column of A consists of the components of $\mathbf{e}'_j$ in the $\{\mathbf{e}_i\}_{i=1}^n$ basis.

Similarly, the individual basis vectors of the basis $\{\mathbf{e}_i\}_{i=1}^n$ can be written as

$$\mathbf{e}_i = \mathbf{e}'_k B_{ki}\,,$$

for some numbers B_{ki}. Here B_{ki} is the k^{th} component of the vector $\mathbf{e}_i$ in the basis $\{\mathbf{e}'_i\}_{i=1}^n$. Again the B_{ki} can be viewed as the entries of a matrix B

$$\mathsf{B} = \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1n} \\ B_{21} & B_{22} & \cdots & B_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ B_{n1} & B_{n2} & \cdots & B_{nn} \end{pmatrix}.$$

From the above two transformations of basis, we have that

$$\mathbf{e}'_j = (\mathbf{e}'_k B_{ki}) A_{ij} = \mathbf{e}'_k (B_{ki} A_{ij})\,.$$

Since

$$\mathbf{e}'_j = \mathbf{e}'_k \delta_{kj}\,,$$

it follows that

$$B_{ki} A_{ij} = \delta_{kj}\,.$$

Hence in matrix notation, $\mathsf{BA} = \mathsf{I}$, where I is the identity matrix. Conversely, we can also prove that $\mathsf{AB} = \mathsf{I}$.

Proposition 4.22. For the transformation matrices between two sets of bases, A and B,

$$\mathsf{B} = \mathsf{A}^{-1}\,,$$

and

$$\det \mathsf{A} \neq 0 \text{ and } \det \mathsf{B} \neq 0\,.$$

Remark. A transformation matrix uniquely determines an endomorphism.

4.3.4 Transformation Law for Vector Components

Proposition 4.23. Let v and v' be the column matrices of the components of a vector in the two bases $\{\mathbf{e}_i\}$ and $\{\mathbf{e}'_i\}$ respectively, and let A be the transformation matrix of the endomorphism between the two bases. We have

$$\mathsf{v} = \mathsf{A}\mathsf{v}'\,,$$
$$\mathsf{v}' = \mathsf{A}^{-1}\mathsf{v}\,.$$

Proof. For a vector $\mathbf{v}$, in the $\{\mathbf{e}_i\}_{i=1}^n$ basis we have

$$\mathbf{v} = v_i \mathbf{e}_i \,.$$

Similarly, in the $\{\mathbf{e}'_i\}_{i=1}^n$ basis we can write

$$\begin{aligned} \mathbf{v} &= v'_j \mathbf{e}'_j \\ &= v'_j \mathbf{e}_i A_{ij} \\ &= \mathbf{e}_i (A_{ij} v'_j) \,, \end{aligned}$$

so it follows that

$$v_i = A_{ij} v'_j \,,$$

which relates the components of $\mathbf{v}$ in the basis $\{\mathbf{e}_i\}_{i=1}^n$ to those in the basis $\{\mathbf{e}'_i\}_{i=1}^n$ as claimed. The second equation follows as A is invertible. □

Remark. In matrix notation, the transformation between bases is expressed as

$$\mathsf{e}' = \mathsf{eA} \,.$$

We can see that the components of $\mathbf{v}$ transform inversely to the way that the basis vectors transform, so that the vector $\mathbf{v}$ is unchanged:

$$\begin{aligned} \mathbf{v} &= v'_j \mathbf{e}_j \\ &= ((\mathsf{A}^{-1})_{jk} v_k)(\mathbf{e}_i A_{ij}) \\ &= \mathbf{e}_i (v_k (A_{ij} (\mathsf{A}^{-1})_{jk})) \\ &= \mathbf{e}_i (v_k \delta_{ik}) \\ &= v_i \mathbf{e}_i \,. \end{aligned}$$

4.3.5 Transformation of Matrices Representing Linear Maps

Proposition 4.24. Let M and M' be the matrices representing an endomorphism $\mathcal{M} \in \mathrm{End}(\mathbb{F}^n)$ in two bases $\{\mathbf{e}_i\}$ and $\{\mathbf{e}'_i\}$ where the transformation matrix between the two bases is A. Then, we have

$$\mathsf{M}' = \mathsf{A}^{-1} \mathsf{M} \mathsf{A} \,.$$

Proof. Let $\mathbf{v} \mapsto \mathcal{M}(\mathbf{v}) = \mathbf{u}$. In terms of matrices, this can be expressed as

$$\mathsf{u} = \mathsf{Mv} \,,$$

where u, v are the component column matrices of $\mathbf{u}$ and $\mathbf{v}$ with respect to the basis $\{\mathbf{e}_i\}$. Let u' and v' be the component column matrices of $\mathbf{u}$ and $\mathbf{v}$ in an alternative basis $\{\mathbf{e}'_i\}$. Then it follows that

$$\begin{aligned} \mathsf{Au}' &= \mathsf{MAv}' \\ \implies \mathsf{u}' &= (\mathsf{A}^{-1}\mathsf{MA})\mathsf{v}' \,. \end{aligned}$$

Therefore, in the new basis $\{\mathbf{e}'_i\}$, the matrix representing the linear map $\mathcal{M}$ becomes

$$\mathsf{M}' = \mathsf{A}^{-1}\mathsf{MA}$$

as claimed. □

4.4 Some Definitions of Special Matrices

We define the following special real matrices:

Definition 4.25. A square matrix is *symmetric* if it is equal to its transpose:

$$\mathsf{A}^{\mathrm{T}} = \mathsf{A} \quad \text{or} \quad A_{ij} = A_{ji}\,.$$

Definition 4.26. A square matrix is *anti-symmetric* if it is equal to the negative of its transpose:

$$\mathsf{A}^{\mathrm{T}} = -\mathsf{A} \quad \text{or} \quad A_{ij} = -A_{ji}\,.$$

Definition 4.27. A square matrix is *orthogonal* if its transpose is equal to its inverse:

$$\mathsf{A}^{\mathrm{T}} = \mathsf{A}^{-1} \quad \text{or} \quad \mathsf{A}\mathsf{A}^{\mathrm{T}} = \mathsf{A}^{\mathrm{T}}\mathsf{A} = \mathsf{I}\,.$$

These ideas can be generalised to a complex vector space.

Definition 4.28. The *Hermitian conjugate* of a matrix is the complex conjugate of its transpose:

$$\mathsf{A}^{\dagger} \coloneqq (\mathsf{A}^{\mathrm{T}})^* = (\mathsf{A}^*)^{\mathrm{T}} \quad \text{or} \quad (\mathsf{A}^{\dagger})_{ij} \coloneqq A^*_{ji}\,.$$

Proposition 4.29.

(i) The Hermitian conjugate of a Hermitian conjugate:

$$\mathsf{A}^{\dagger\dagger} = \mathsf{A}\,.$$

(ii) The Hermitian conjugate of a product:

$$(\mathsf{A}\mathsf{B})^{\dagger} = \mathsf{B}^{\dagger}\mathsf{A}^{\dagger}\,.$$

Proof.

(i)

$$\mathsf{A}^{\dagger\dagger} = (\mathsf{A}^{*\mathrm{T}})^{\mathrm{T}*} = \mathsf{A}\,.$$

(ii)

$$\begin{aligned}(\mathsf{A}\mathsf{B})^{\dagger} &= ((\mathsf{A}\mathsf{B})^{\mathrm{T}})^* \\ &= (\mathsf{B}^{\mathrm{T}}\mathsf{A}^{\mathrm{T}})^* \\ &= (\mathsf{B}^{\mathrm{T}})^*(\mathsf{A}^{\mathrm{T}})^* = \mathsf{B}^{\dagger}\mathsf{A}^{\dagger}\,.\end{aligned}$$

□

Definition 4.30. A $n \times n$ matrix is *positive definite* if for all column matrices v of length n,

$$\mathsf{v}^{\dagger}\mathsf{A}\mathsf{v} \geq 0\,, \text{ with equality iff } \mathsf{v} = \mathsf{0}\,.$$

Remark. If equality to zero were possible for non-zero v, then A is said to be *positive semi-definite* rather than positive definite.

Definition 4.31. A square matrix is *Hermitian* if it is equal to its Hermitian conjugate:

$$\mathsf{A}^{\dagger} = \mathsf{A} \quad \text{or} \quad A^*_{ji} = A_{ij}\,.$$

Definition 4.32. A square matrix is *anti-Hermitian* if it is equal to the negative of its Hermitian conjugate:

$$\mathsf{A}^{\dagger} = -\mathsf{A} \quad \text{or} \quad A^*_{ji} = -A_{ij}\,.$$

Definition 4.33. A square matrix is *unitary* if its Hermitian conjugate is equal to its inverse:

$$\mathsf{A}^{\dagger} = \mathsf{A}^{-1} \quad \text{or} \quad \mathsf{A}\mathsf{A}^{\dagger} = \mathsf{A}^{\dagger}\mathsf{A} = \mathsf{I}\,.$$

Definition 4.34. A square matrix is *normal* if it commutes with its Hermitian conjugate:

$$\mathsf{A}\mathsf{A}^{\dagger} = \mathsf{A}^{\dagger}\mathsf{A}\,.$$

4.5 Scalar Product

4.5.1 Definition of a Scalar Product

The prototype vector space $\mathbb{E}^3$ has the additional property that any two vectors $\mathbf{u}$ and $\mathbf{v}$ can be combined to form a scalar $\mathbf{u} \cdot \mathbf{v}$. This is generalised to an n-dimensional vector space over $\mathbb{C}$. From now on, let $\mathbb{F}$ be $\mathbb{C}$ or $\mathbb{R}$.

Definition 4.35. An *inner product space* is a vector space over $\mathbb{C}$ equipped with an *inner product*. An *inner product* is a map $\cdot : V \times V \to \mathbb{C}$ with the properties.

(i) *Sesquilinear.* For $a, b \in \mathbb{C}$, $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$,

$$(a\mathbf{x} + b\mathbf{y}) \cdot \mathbf{z} = a^*(\mathbf{x} \cdot \mathbf{z}) + b^*(\mathbf{y} \cdot \mathbf{z}) \,,$$

$$\mathbf{x} \cdot (a\mathbf{y} + b\mathbf{z}) = a(\mathbf{x} \cdot \mathbf{y}) + b(\mathbf{x} \cdot \mathbf{z}) \,.$$

(ii) *Hermitian.*

$$\mathbf{x} \cdot \mathbf{y} = (\mathbf{y} \cdot \mathbf{x})^* \,.$$

(iii) *Positive definite.* For all $\mathbf{x} \in V \setminus \{\mathbf{0}\}$, $\mathbf{x} \cdot \mathbf{x} > 0$, and $\mathbf{0} \cdot \mathbf{0} = 0$.

Remark. A scalar product has existence without reference to any basis.

Definition 4.36. The *(Euclidean) norm* of a vector $\mathbf{v} \in V$ is defined as

$$\|\mathbf{v}\| = (\mathbf{v} \cdot \mathbf{v})^{\frac{1}{2}} \,.$$

Remark. A norm in general is a function $V \to \mathbb{R}_{\geq 0}$ satisfying several properties, and the Euclidean norm is a specific example of a norm. A norm can induce a notion of distance (*metric*) in the vector space, which makes a normed vector space also a *metric space*.

Definition 4.37. Two vectors are *orthogonal* if $\mathbf{u} \cdot \mathbf{v} = 0$.

Two orthogonal vectors are *orthonormal* if $\|\mathbf{u}\| = \|\mathbf{v}\| = 1$.

Lemma 4.38. Orthogonal vectors are linearly independent.

Proof. For two orthogonal vectors $\mathbf{u}$ and $\mathbf{v}$, suppose that there exists α and β such that

$$\alpha\mathbf{u} + \beta\mathbf{v} = 0 \,.$$

By pre-multiplying $\mathbf{u}$, we have

$$\alpha(\mathbf{u} \cdot \mathbf{u}) + \beta(\mathbf{u} \cdot \mathbf{v}) = \alpha\|\mathbf{u}\|^2 + 0 = 0 \,.$$

Since u is non-zero, $\alpha = 0$, and similarly $\beta = 0$, so they are linearly independent. □

Alternative notations.

$$\langle \mathbf{u} | \mathbf{v} \rangle \equiv \mathbf{u} \cdot \mathbf{v} \,, \ \|\mathbf{v}\| \equiv |\mathbf{v}| = (\mathbf{v} \cdot \mathbf{v})^{\frac{1}{2}} \,.$$

4.5.2 Some Inequalities

Theorem 4.39 (Cauchy–Schwarz inequality).

$$|\langle \mathbf{u} | \mathbf{v} \rangle| \leq \|\mathbf{u}\| \|\mathbf{v}\| \,,$$

with equality only when $\mathbf{u}$ is a scalar multiple of $\mathbf{v}$.

Proof. Write $\langle \mathbf{u}|\mathbf{v}\rangle = |\langle \mathbf{u}|\mathbf{v}\rangle| e^{i\alpha}$, and for $\lambda \in \mathbb{C}$, consider

$$\begin{aligned}\|\mathbf{u}+\lambda\mathbf{v}\|^2 &= \langle \mathbf{u}+\lambda\mathbf{v}|\mathbf{u}+\lambda\mathbf{v}\rangle \\ &= \langle \mathbf{u}|\mathbf{u}\rangle + \lambda\langle \mathbf{u}|\mathbf{v}\rangle + \lambda^*\langle \mathbf{v}|\mathbf{u}\rangle + |\lambda|^2\langle \mathbf{v}|\mathbf{v}\rangle \\ &= \langle \mathbf{u}|\mathbf{u}\rangle + (\lambda e^{i\alpha} + \lambda^* e^{-i\alpha})|\langle \mathbf{u}|\mathbf{v}\rangle| + |\lambda|^2\langle \mathbf{v}|\mathbf{v}\rangle \ .\end{aligned}$$

First, suppose that $\mathbf{v} = \mathbf{0}$. The RHS of the equation simplifies to an expression linear in λ. If $\langle \mathbf{u}|\mathbf{v}\rangle \neq 0$ we then have a contradiction since for a certain choice of λ this expression can be negative. Hence we conclude that

$$\langle \mathbf{u}|\mathbf{v}\rangle = 0 \text{ if } \mathbf{v} = \mathbf{0}\,,$$

which satisfies the Cauchy–Schwarz inequality as an equality.

Next suppose $\mathbf{v} \neq \mathbf{0}$ and choose $\lambda = re^{-i\alpha}$ so that

$$0 \leq \|\mathbf{u}+\lambda\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + 2r|\langle \mathbf{u}|\mathbf{v}\rangle| + r^2\|\mathbf{v}\|^2\,.$$

The RHS is a quadratic in r that has a minimum when $r\|\mathbf{v}\|^2 = -|\langle \mathbf{u}|\mathbf{v}\rangle|$. Cauchy–Schwarz inequality follows by substituting this value of r, with equality if $\mathbf{u} = -\lambda\mathbf{v}$. □

Theorem 4.40 (The triangle inequality).

$$\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|\,.$$

Proof.

$$\begin{aligned}\|\mathbf{u}+\mathbf{v}\|^2 &= \langle \mathbf{u}|\mathbf{u}\rangle + \langle \mathbf{u}|\mathbf{v}\rangle + \langle \mathbf{u}|\mathbf{v}\rangle^* + \langle \mathbf{v}|\mathbf{v}\rangle \\ &= \|\mathbf{u}\|^2 + 2\,\mathrm{Re}\{\langle \mathbf{u}|\mathbf{v}\rangle\} + \|\mathbf{v}\|^2 \\ &\leq \|\mathbf{u}\|^2 + 2|\langle \mathbf{u}|\mathbf{v}\rangle| + \|\mathbf{v}\|^2 \\ &\leq \|\mathbf{u}\|^2 + 2\|\mathbf{u}\|\|\mathbf{v}\| + \|\mathbf{v}\|^2 \\ &= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2\,.\end{aligned}$$

□

4.5.3 The Scalar Product in Terms of Components

Suppose that we have a scalar product defined on a vector space with a given basis $\{\mathbf{e}_i\}_{i=1}^n$. The scalar product is determined for all pairs of vectors by its value for all pairs of basis vectors. Define the complex number G_{ij} by

$$G_{ij} = \mathbf{e}_i \cdot \mathbf{e}_j \ (i,j = 1,\ldots,n)\,.$$

Then, for any two vectors

$$\mathbf{v} = v_i\mathbf{e}_i \text{ and } \mathbf{w} = w_j\mathbf{e}_j\,,$$

we have that

$$\begin{aligned}\mathbf{v}\cdot\mathbf{w} &= (v_i\mathbf{e}_i)\cdot(w_j\mathbf{e}_j) \\ &= v_i^* w_j \mathbf{e}_i\cdot\mathbf{e}_j \\ &= v_i^* G_{ij} w_j\,.\end{aligned}$$

Definition 4.41. In matrix notation, the scalar product is written as

$$\mathbf{v}\cdot\mathbf{w} = \mathsf{v}^\dagger \mathsf{G}\mathsf{w}\,,$$

where G is the matrix with entries G_{ij}, called the *metric*.

4.5.4 Properties of the Metric

Proposition 4.42. A metric is Hermitian.

Proof. The elements of the Hermitian conjugate of the metric G are the complex numbers

$$\begin{aligned}(\mathsf{G}^\dagger)_{ij} &= (G_{ji})^* \\ &= (\mathbf{e}_j \cdot \mathbf{e}_i)^* \\ &= \mathbf{e}_i \cdot \mathbf{e}_j \\ &= G_{ij}\,.\end{aligned}$$

Hence G is Hermitian. $\mathsf{G}^\dagger = \mathsf{G}$. □

Remark. The property that G is Hermitian is consistent with the requirement that $|\mathbf{v}|^2 = \mathbf{v}\cdot\mathbf{v}$ is real.

$$\begin{aligned}(\mathbf{v}\cdot\mathbf{v})^* &= ((\mathbf{v}\cdot\mathbf{v})^*)^{\mathrm{T}} \\ &= (\mathbf{v}\cdot\mathbf{v})^\dagger \\ &= (\mathsf{v}^\dagger\mathsf{G}\mathsf{v})^\dagger \\ &= \mathsf{v}^\dagger\mathsf{G}^\dagger\mathsf{v} \\ &= \mathsf{v}^\dagger\mathsf{G}\mathsf{v} \\ &= \mathbf{v}\cdot\mathbf{v}\,.\end{aligned}$$

Proposition 4.43. A metric is positive definite.

Proof. For any $\mathbf{v}$, we have

$$|\mathbf{v}|^2 \geq 0 \text{ with equality iff } \mathbf{v} = 0\,.$$

Hence, for any $\mathbf{v}$,

$$\mathsf{v}^\dagger\mathsf{G}\mathsf{v} \geq \mathsf{0} \text{ with equality iff } \mathsf{v} = \mathsf{0}\,.$$

Therefore, G is positive definite by definition. □

4.5.5 The Gram–Schmidt Process

For a $\mathbb{F}^n$ space, a set of mutually orthogonal vectors can be generated using the following process.

Proposition 4.44 (The Gram–Schmidt process). Let $\mathbb{F}^n$ be a vector space equipped with an inner product. Given a set of linearly independent vectors $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$, a set of mutually orthogonal vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ as

$$\begin{aligned}\mathbf{v}_1 &= \mathbf{w}_1\,, \\ \mathbf{v}_2 &= \mathbf{w}_2 - \mathcal{P}_{\mathbf{v}_1}(\mathbf{w}_2)\,, \\ \mathbf{v}_3 &= \mathbf{w}_3 - \mathcal{P}_{\mathbf{v}_1}(\mathbf{w}_3) - \mathcal{P}_{\mathbf{v}_2}(\mathbf{w}_3)\,, \\ &\vdots \\ \mathbf{v}_n &= \mathbf{w}_n - \sum_{i=1}^{n-1} \mathcal{P}_{\mathbf{v}_i}(\mathbf{w}_n)\,,\end{aligned}$$

where $\mathcal{P}$ is the operator that projects $\mathbf{w}$ orthogonally on $\mathbf{v}$:

$$\mathcal{P}_{\mathbf{v}}(\mathbf{w}) = \frac{\mathbf{v}\cdot\mathbf{w}}{\mathbf{v}\cdot\mathbf{v}}\mathbf{v}\,.$$

Proof. First, we have that

$$\begin{aligned}\mathbf{v}_1 \cdot \mathbf{v}_2 &= \mathbf{v}_1 \cdot \left(\mathbf{w}_2 - \frac{\mathbf{v}_1 \cdot \mathbf{w}_2}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1\right) \\ &= 0\,.\end{aligned}$$

Assume that $\{\mathbf{v}_1, \dots \mathbf{v}_{k-1}\}$ are mutually orthogonal, then for $1 \leq i \leq k-1$,

$$\begin{aligned}\mathbf{v}_i \cdot \mathbf{v}_k &= \mathbf{v}_i \cdot \left(\mathbf{w}_k - \sum_{j=1}^{k-1} \mathcal{P}_{\mathbf{v}_j}(\mathbf{w}_k)\right) \\ &= \mathbf{v}_i \cdot \mathbf{w}_k - \sum_{j=1}^{k-1} \frac{\mathbf{v}_j \cdot \mathbf{w}_k}{\mathbf{v}_j \cdot \mathbf{v}_j}\mathbf{v}_i \cdot \mathbf{v}_j \\ &= \mathbf{v}_i \cdot \mathbf{w}_k - \frac{\mathbf{v}_i \cdot \mathbf{w}_k}{\mathbf{v}_i \cdot \mathbf{v}_i}\mathbf{v}_i \cdot \mathbf{v}_i \\ &= 0\,.\end{aligned}$$

The proposition is therefore proved by induction. □

4.6 Eigenvalues, Eigenvectors and Diagonalisation

Theorem 4.45 (The fundamental theorem of algebra). Let $p(z)$ be a polynomial of degree $m \geq 1$

$$p(z) = \sum_{j=0}^{m} c_j z^j\,,$$

with $c_j \in \mathbb{C}$ and $c_m \neq 0$. Then $p(z)$ can be factorised as

$$p(z) = c_m \prod_{j=1}^{m} (z - \omega_j)\,,$$

where $\omega_j \in \mathbb{C}$. $p(z)$ is guaranteed to have m roots in $\mathbb{C}$. The number of times an ω is repeated in the factorisation is called its *multiplicity*.

Definition 4.46. Let $\mathcal{A} \in \mathrm{End}(\mathbb{F}^n)$ be an endomorphism, where $\mathbb{F}$ is $\mathbb{R}$ or $\mathbb{C}$. Then a non-zero vector $\mathbf{x} \in \mathbb{F}^n$ that satisfies the *eigenvalue equation*

$$\mathcal{A}\mathbf{x} = \lambda\mathbf{x}\,,$$

where $\lambda \in \mathbb{F}$, is said to be an *eigenvector* of the endomorphism $\mathcal{A}$ with *eigenvalue* λ.

Definition 4.47. For a $n \times n$ square matrix M, its *characteristic polynomial* is

$$p(\lambda) \coloneqq \det(\mathsf{M} - \lambda\mathsf{I})\,,$$

which is a polynomial of order n. The *characteristic equation* of M is

$$p(\lambda) = 0\,.$$

Proposition 4.48. The eigenvalues of a square matrix M are given by the roots of its characteristic polynomial. A $n \times n$ matrix has n eigenvalues.

Proof. Rewrite the eigenvalue equation as

$$(\mathsf{M} - \lambda\mathsf{I})\mathsf{x} = 0\,.$$

Since x is non-zero, a non-trivial linear combination of the columns of the matrix $(\mathsf{M} - \lambda\mathsf{I})$ is zero, so the columns of the matrix are linearly dependent. Therefore,

$$\det(\mathsf{M} - \lambda\mathsf{I}) = 0\,.$$

Since an n^{th} order polynomial has exactly n complex roots (might be repeated) by the fundamental theorem of algebra, there are always n eigenvalues. □

If the eigenvalues are not all distinct, then the repeated eigenvalues are said to be *degenerate*. If an eigenvalue occurs m times, there may be any number between 1 and m of linearly independent eigenvectors corresponding to it.

Definition 4.49. The multiplicity of an eigenvalue as a root of the characteristic polynomial is called the *algebraic multiplicity* of λ, which is denoted by a_λ. If the characteristic polynomial has degree n, then

$$\sum_\lambda a_\lambda = n\,.$$

Definition 4.50. The maximum number, m_λ, of linearly independent eigenvectors corresponding to λ is called the *geometric multiplicity* of λ.

Any linear combination of these eigenvectors with the same eigenvalue is also an eigenvector with this eigenvalue, and the space spanned by these vectors is called an *eigenspace*.

Proposition 4.51. The set of all eigenvectors corresponding to an eigenvalue, together with $\mathbf{0}$, is a vector subspace of $\mathbb{F}^n$ called the *eigenspace* of λ, denoted as E_λ.

Proof. The set of all eigenvectors corresponding to an eigenvalue λ_i with $\mathbf{0}$ are the kernel of the linear map $(\mathsf{M} - \lambda_i\mathsf{I})$, so is a vector subspace of $\mathbb{F}^n$ by Proposition 4.17. □

Corollary. $m_\lambda = \dim_{\mathbb{F}} E_\lambda$.

Definition 4.52. The difference $\Delta_\lambda = a_\lambda - m_\lambda$ is called the *defect* of λ.

Remark. We shall see below that if the eigenvectors of a map form a basis of $\mathbb{F}^n$ (i.e. if there is no eigenvalue with strictly positive defect), then it is possible to analyse the behaviour of that map and associated matrices in terms of these eigenvectors.

Lemma 4.53. If the n eigenvalues of a square $n \times n$ matrix are all distinct, then there are n linearly independent eigenvectors, each of which is determined uniquely up to an arbitrary multiplicative constant.

4.6.1 Similarity Transformation

Definition 4.54. Two $n \times n$ matrices A and B are *similar*, or *conjugate*, if there exists a invertible matrix P such that

$$\mathsf{B} = \mathsf{P}^{-1}\mathsf{A}\mathsf{P}\,.$$

A map from A to $\mathsf{P}^{-1}\mathsf{A}\mathsf{P}$ is a *similarity transformation*.

Remark. The matrices representing the same linear map $\mathcal{A}$ with respect to different bases are similar.

Proposition 4.55. Similar matrices have the same determinant and trace.

Proof.

$$\begin{aligned}\det\big(\mathsf{P}^{-1}\mathsf{A}\mathsf{P}\big) &= \det\mathsf{P}^{-1}\det\mathsf{A}\det\mathsf{P}\\ &= \det\mathsf{A}\det\big(\mathsf{P}^{-1}\mathsf{P}\big)\\ &= \det A\,.\end{aligned}$$

$$\begin{aligned}\mathrm{Tr}\big(\mathsf{P}^{-1}\mathsf{AP}\big) &= P^{-1}_{ij}A_{jk}P_{ki}\\ &= A_{jk}P_{ki}P^{-1}_{ij}\\ &= A_{jk}\delta_{kj}\\ &= \mathrm{Tr}(\mathsf{A})\,.\end{aligned}$$

□

Remark. The determinants and traces of matrices representing a map A with respect to different bases are the same. We can therefore talk of the determinant and trace of a map directly.

Proposition 4.56. Two similar matrices have the same characteristic polynomial, and hence the same eigenvalues.

Proof. Suppose that $\mathsf{B} = \mathsf{P}^{-1}\mathsf{AP}$, then

$$\begin{aligned}p_B(\lambda) &= \det(\mathsf{B} - \lambda\mathsf{I})\\ &= \det\big(\mathsf{P}^{-1}\mathsf{AP} - \lambda\mathsf{P}^{-1}\mathsf{IP}\big)\\ &= \det\big(\mathsf{P}^{-1}(\mathsf{A} - \lambda\mathsf{I})\mathsf{P}\big)\\ &= \det\big(\mathsf{P}^{-1}\big)\det(\mathsf{A} - \lambda\mathsf{I})\det\mathsf{P}\\ &= p_A(\lambda)\,.\end{aligned}$$

□

4.6.2 Diagonalisation

Theorem 4.57. If a $n \times n$ matrix M has n linearly independent eigenvectors $\mathbf{x}^i$ with corresponding eigenvalues λ_i, then M can be diagonalised by a similarity transformation

$$\mathsf{X}^{-1}\mathsf{MX} = \mathsf{\Lambda}\,,$$

where X is a square matrix whose columns are the eigenvectors $\mathbf{x}^i$ and $\mathsf{\Lambda}$ is the diagonal matrix with diagonal entries $\Lambda_{ii} = \lambda_i$.

Proof. We have

$$\mathsf{M}\mathsf{x}^i = \lambda_i\mathsf{x}^i\,, \quad \text{(no summation convention)}$$

or in index notation for the j^{th} component,

$$\sum_{k=1}^{n} M_{jk}x^i_k = \lambda_i x^i_j\,.$$

Let X be the $n \times n$ matrix whose columns are the eigenvectors of M, then

$$(\mathsf{X})_{ij} = X_{ij} = x^j_i\,,$$

$$\mathsf{X} = \begin{pmatrix} x^1_1 & x^2_1 & \cdots & x^n_1\\ x^1_2 & x^2_2 & \cdots & x^n_2\\ \vdots & \vdots & \ddots & \vdots\\ x^1_n & x^2_n & \cdots & x^n_n\end{pmatrix}.$$

The eigenvalue equation can be rewritten as

$$\sum_{k=1}^{n} M_{jk}X_{ki} = \lambda_i X_{ji} = \sum_{k=1}^{n} X_{jk}\delta_{ki}\lambda_i\,,$$

or, in matrix notation, as

$$\mathsf{M}\mathsf{X} = \mathsf{X}\mathsf{\Lambda}\,,$$

where $\mathsf{\Lambda}$ is the diagonal matrix

$$\mathsf{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

The theorem follows if X is invertible, which is equivalent to the requirement that the columns of X are linearly independent, i.e. M has n linearly independent eigenvectors. □

Remark. If any eigenvalue of the matrix has a non-zero defect, then it is not diagonalisable.

4.6.3 Eigenvalues and Eigenvectors of a Hermitian Matrix

Let H be a Hermitian matrix, and consider two eigenvectors x and y corresponding eigenvalues λ and μ:

$$\mathsf{H}\mathsf{x} = \lambda\mathsf{x}\,,$$

$$\mathsf{H}\mathsf{y} = \mu\mathsf{y}\,.$$

Since H is Hermitian,

$$\mathsf{y}^\dagger\mathsf{H} = \mu^*\mathsf{y}^\dagger\,.$$

We can therefore construct two expressions for $\mathsf{y}^\dagger\mathsf{H}\mathsf{x}$:

$$\mathsf{y}^\dagger\mathsf{H}\mathsf{x} = \lambda\mathsf{y}^\dagger\mathsf{x} = \mu^*\mathsf{y}^\dagger\mathsf{x}\,,$$

and hence

$$(\lambda - \mu^*)\mathsf{y}^\dagger\mathsf{x} = 0\,.$$

Theorem 4.58. The eigenvalues of a Hermitian matrix are real.

Proof. Suppose that x and y are the same eigenvector. Then $\lambda = \mu$, so

$$(\lambda - \lambda^*)\mathsf{x}^\dagger\mathsf{x} = 0\,.$$

Since $\mathsf{x} \neq \mathsf{0}$, $\mathsf{x}^\dagger\mathsf{x} = x_i^* x_i = |\mathbf{x}|^2 \neq 0$, and so $\lambda = \lambda^*$, i.e. the eigenvalues are real. □

Theorem 4.59. The eigenvectors of a Hermitian matrix are orthogonal.

Proof. Since the eigenvalues of a Hermitian matrix are real,

$$(\lambda - \mu)\mathsf{y}^\dagger\mathsf{x} = 0\,.$$

If y and x are different eigenvectors, we can deduce that $\mathsf{y}^\dagger\mathsf{x} = 0$ provided $\mu \neq \lambda$. Therefore, the eigenvectors with different eigenvalues are orthogonal.

When there is a repeated eigenvalue, the proof of orthogonality is more difficult and we will not do it here. The basic idea of the proof is to use the Gram–Schmidt procedure to extend the set of eigenvectors of non-degenerate eigenvalues to a complete set of orthonormal basis of $\mathbb{C}^n$ and use it to attempt to diagonalise H. We can iteratively diagonalise the undiagonalised part until the matrix is fully diagonalised. □

If we want to diagonalise a Hermitian matrix with a repeated eigenvalue, we can directly use the Gram–Schmidt procedure to generate an arbitrary orthogonal basis in the eigenspace.

Corollary. A $n \times n$ Hermitian matrix has n orthonormal eigenvectors.

Proof. For any $\mu \in \mathbb{C}$, if $\mathsf{H}\mathsf{x} = \lambda\mathsf{x}$, then $\mathsf{H}(\mu\mathsf{x}) = \lambda(\mu\mathsf{x})$. This allows us to normalise the eigenvectors so that

$$\mathsf{x}^\dagger\mathsf{x} = 1\,.$$

Therefore, for Hermitian matrices, it is always possible to find n orthonormal eigenvectors that are linearly independent. □

Corollary. The eigenvectors of a Hermitian matrix are linearly independent.

Proof. By Lemma 4.38, orthogonal vectors are linearly independent. Therefore, a $n \times n$ Hermitian matrix has n orthonormal eigenvectors that are linearly independent. □

Corollary. It can be proved similarly that the eigenvalues of anti-Hermitian and unitary matrices are imaginary and of unit modulus, respectively.

Corollary. The eigenvectors of normal matrices corresponding to distinct eigenvalues are orthogonal. Moreover, if a repeated eigenvalue λ occurs m times, then there are m corresponding linearly independent eigenvectors.

Therefore, even if the eigenvalues of a Hermitian matrix are degenerate, it is possible to find n mutually orthogonal eigenvectors, which form a basis for the vector space.

4.6.4 Diagonalisation of Hermitian Matrices

Theorem 4.60. Every Hermitian matrix, H, is diagonalisable by a transformation

$$\mathsf{X}^\dagger\mathsf{H}\mathsf{X} = \mathsf{\Lambda}\,,$$

where X is a unitary matrix.

Proof. As shown above, a $n \times n$ Hermitian matrix H has n linearly independent eigenvectors, so it must be diagonalisable to the matrix $\mathsf{\Lambda}$ by the transformation $\mathsf{X}^{-1}\mathsf{H}\mathsf{X}$, where the columns of X are the eigenvectors of H:

$$\mathsf{X} = \begin{pmatrix} x_1^1 & x_1^2 & \cdots & x_1^n \\ x_2^1 & x_2^2 & \cdots & x_2^n \\ \vdots & \vdots & \ddots & \vdots \\ x_n^1 & x_n^2 & \cdots & x_n^n \end{pmatrix}.$$

If the x^i are the orthonormal eigenvectors of H then X is a unitary matrix since:

$$(\mathsf{X}^\dagger\mathsf{X})_{ij} = (\mathsf{X}^\dagger)_{ik}(\mathsf{X})_{kj} = (x_k^i)^*(x_k^j) = \mathsf{x}^{i\dagger}\mathsf{x}^j = \delta_{ij}\,,$$

or, in matrix notation

$$\mathsf{X}^\dagger\mathsf{X} = \mathsf{I}\,.$$

Hence X is a unitary matrix. □

Corollary. If we restrict ourselves to real matrices, we conclude that for every real symmetric matrix, S, is diagonalisable by a transformation $\mathsf{R}^{\mathrm{T}}\mathsf{S}\mathsf{R}$, where R is an orthogonal matrix.

Corollary. As noted above, normal matrices always have n linearly independent eigenvectors, and hence can always be diagonalised. So, in addition to Hermitian matrices, anti-Hermitian matrices and unitary matrices can always be diagonalised.

4.7 Application of Diagonalisation

4.7.1 Transformation Law for Metrics

For an arbitrary vector $\mathbf{v}$, its components in the two bases transform according to $\mathsf{v} = \mathsf{A}\mathsf{v}'$, where v and v' are column matrices containing the components. Taking the Hermitian conjugate of this expression, we have that

$$\mathsf{v}^\dagger = \mathsf{v}'^\dagger \mathsf{A}^\dagger \,.$$

Hence for arbitrary vectors $\mathbf{v}$ and $\mathbf{w}$

$$\begin{aligned} \mathbf{v}\cdot\mathbf{w} &= \mathsf{v}^\dagger \mathsf{G}\mathsf{w} \\ &= \mathsf{v}'^\dagger \mathsf{A}^\dagger \mathsf{G}\mathsf{A}\mathsf{w}' \,. \end{aligned}$$

We also have that in terms of the new basis,

$$\mathbf{v}\cdot\mathbf{w} = \mathsf{v}'^\dagger \mathsf{G}'\mathsf{w}' \,,$$

where G' is the metric in the new $\{\mathbf{e}'_i\}_{i=1}^n$ basis. Since $\mathbf{v}$ and $\mathbf{w}$ are arbitrary we conclude that the metric in the new basis is given by

$$\mathsf{G}' = \mathsf{A}^\dagger \mathsf{G}\mathsf{A} \,.$$

Alternative derivation.

$$\begin{aligned} (\mathsf{G}')_{ij} \equiv G'_{ij} &= \mathbf{e}'_i \cdot \mathbf{e}'_j \\ &= (\mathbf{e}_k A_{ki}) \cdot (\mathbf{e}_l A_{lj}) \\ &= A^*_{ki} (\mathbf{e}_k \cdot \mathbf{e}_l) A_{lj} \\ &= A^\dagger_{ik} G_{kl} A_{lj} \\ &= (\mathsf{A}^\dagger \mathsf{G}\mathsf{A})_{ij} \,. \end{aligned}$$

Remark. G' is also Hermitian since

$$(\mathsf{G}')^\dagger = (\mathsf{A}^\dagger \mathsf{G}\mathsf{A})^\dagger = \mathsf{A}^\dagger \mathsf{G}^\dagger \mathsf{A} = \mathsf{A}^\dagger \mathsf{G}\mathsf{A} = \mathsf{G}' \,.$$

4.7.2 Diagonalisation of the Metric

If we identify A with X, the matrix with columns consisting of the orthonormal eigenvectors of G, then

$$\mathsf{G}' = \mathsf{X}^\dagger \mathsf{G}\mathsf{X} = \Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} ,$$

where the λ_i are the real eigenvalues of the Hermitian matrix G.

Proposition 4.61. The eigenvalues of a metric are strictly positive.

Proof. Writing $\Lambda_{ij} = \lambda_i \delta_{ij}$, from the positive definiteness of a metric, we have that

$$0 \le \mathsf{v}'^\dagger \mathsf{G}'\mathsf{v}' = \sum_{i,j=1}^n v'^*_i \lambda_i \delta_{ij} v'_j = \sum_{i=1}^n \lambda_i |v'_i|^2 \,,$$

with equality only if $\mathbf{v}' = 0$. This can only be true for all vectors $\mathbf{v}'$ if

$$\lambda_i > 0 \text{ for } i = 1, \ldots, n \,,$$

i.e. the diagonal entries λ_i are strictly positive. □

4.7.3 Orthonormal Bases

For a diagonalised metric, the new basis vectors $\{\mathbf{e}'_i\}_{i=1}^n$ are the eigenvectors of G since

$$\mathbf{e}'_j = \mathbf{e}_i X_{ij} = \mathbf{e}_i x_i^j \quad (j = 1, \ldots, n)\,.$$

Hence the new basis vectors are orthogonal. Further,

$$\mathbf{e}'_i \cdot \mathbf{e}'_j = G'_{ij} = \Lambda_{ij} = \lambda_{ij}\,.$$

Hence, because the λ_i are strictly positive, we can normalise the basis

$$\mathbf{e}''_i = \frac{1}{\sqrt{\lambda_i}}\mathbf{e}'_i$$

so that

$$\mathbf{e}''_i \cdot \mathbf{e}''_j = \delta_{ij}\,.$$

The $\{\mathbf{e}''_i\}_{i=1}^n$ are thus an orthonormal basis. We therefore conclude that any vector space with a scalar product has an orthonormal basis.

Theorem 4.62. Let column matrices v and w contain the components of two vectors $\mathbf{v}$ and $\mathbf{w}$, in an orthonormal basis $\{\mathbf{e}_i\}_{i=1}^n$. Then

$$\mathbf{v} \cdot \mathbf{w} = \mathsf{v}^\dagger \mathsf{I} \mathsf{w} = \mathsf{v}^\dagger \mathsf{w}\,.$$

Corollary. If the two vectors $\mathbf{v}$ and $\mathbf{w}$ are orthogonal, i.e. $\mathbf{v} \cdot \mathbf{w} = 0$, then the components in an orthonormal basis are such that

$$\mathsf{v}^\dagger \mathsf{w} = 0\,.$$

4.7.4 Transformation between Orthonormal Bases

Given an orthonormal basis $\{\mathbf{e}_i\}_{i=1}^n$, a question is what change of basis maintains orthonormality. Suppose that $\{\mathbf{e}'_i\}_{i=1}^n$ is a new orthonormal basis, and suppose that in terms of the original orthonormal basis (with summation convention)

$$\mathbf{e}'_i = \mathbf{e}_k U_{ki}\,,$$

where U is the transformation matrix. Then the metric of the new basis is given by

$$\mathsf{G}' = \mathsf{U}^\dagger \mathsf{I} \mathsf{U} = \mathsf{U}^\dagger \mathsf{U}\,.$$

For the new basis to be orthonormal we require that the new metric be the identity matrix:

$$\mathsf{U}^\dagger \mathsf{U} = \mathsf{I}\,.$$

Since $\det \mathsf{U} \neq 0$, the inverse U^{-1} exists and hence U must be unitary:

$$\mathsf{U}^\dagger = \mathsf{U}^{-1}\,.$$

Corollary. An analogous result applies to vector spaces over $\mathbb{R}$. Then because the transformation matrix $\mathsf{U} = \mathsf{R}$ is real,

$$\mathsf{U}^\dagger = \mathsf{R}^{\mathrm{T}}\,,$$

and so R must be orthogonal:

$$\mathsf{R}^{\mathrm{T}} = \mathsf{R}^{-1}\,.$$

4.7.5 Uses of Diagonalisation

Because diagonal matrices can be multiplied easily component-wise along the diagonal, certain operations on diagonalisable matrices are more easily carried out using the representations:

$$\mathsf{X}^{-1}\mathsf{M}\mathsf{X} = \mathsf{\Lambda} \text{ and } \mathsf{M} = \mathsf{X}\mathsf{\Lambda}\mathsf{X}^{-1}\,.$$

Proposition 4.63. For a diagonalisable matrix M,

(i) $\mathsf{M}^n = \mathsf{X}\mathsf{\Lambda}^n\mathsf{X}^{-1}$.

(ii) $\det \mathsf{M} = \prod_i \lambda_i$.

(iii) $\operatorname{tr} \mathsf{M} = \sum_i \lambda_i$.

(iv) $\operatorname{tr}(\mathsf{M}^n) = \sum_i \lambda_i^n$.

Proof.

(i)

$$\begin{aligned}\mathsf{M}^n &= (\mathsf{X}\mathsf{\Lambda}\mathsf{X}^{-1})\dots(\mathsf{X}\mathsf{\Lambda}\mathsf{X}^{-1})\\ &= \mathsf{X}\mathsf{\Lambda}^n\mathsf{X}^{-1}\,.\end{aligned}$$

(ii)

$$\begin{aligned}\det \mathsf{M} &= \det\bigl(\mathsf{X}\mathsf{\Lambda}\mathsf{X}^{-1}\bigr)\\ &= \det \mathsf{X} \det \mathsf{\Lambda} \det \mathsf{X}^{-1}\\ &= \det \mathsf{\Lambda} = \prod_i \lambda_i\,.\end{aligned}$$

(iii)

$$\begin{aligned}\operatorname{tr} \mathsf{M} &= \operatorname{tr}\bigl(\mathsf{X}\mathsf{\Lambda}\mathsf{X}^{-1}\bigr)\\ &= \operatorname{tr}\bigl(\mathsf{\Lambda}\mathsf{X}^{-1}\mathsf{X}\bigr)\\ &= \operatorname{tr} \mathsf{\Lambda} = \sum_i \lambda_i\,.\end{aligned}$$

(iv)

$$\begin{aligned}\operatorname{tr}(\mathsf{M}^n) &= \operatorname{tr}\bigl(\mathsf{X}\mathsf{\Lambda}^n\mathsf{X}^{-1}\bigr)\\ &= \operatorname{tr}\bigl(\mathsf{\Lambda}^n\mathsf{X}^{-1}\mathsf{X}\bigr)\\ &= \operatorname{tr}(\mathsf{\Lambda}^n) = \sum_i \lambda_i^n\,.\end{aligned}$$

□

Remark. The results (ii) and (iii) are in fact true for all matrices, whether or not they are diagonalisable.

4.7.6 Cayley–Hamilton Theorem (Non-examinable)

Theorem 4.64 (Cayley–Hamilton theorem). Every endomorphism satisfies its own characteristic equation. If $\mathcal{A}$ has a characteristic polynomial $p(\lambda)$, then

$$p(\mathcal{A}) = 0\,.$$

A proof is beyond the scope of this course.

We can take advantage of the Cayley–Hamilton theorem to calculate the inverse of a matrix. Suppose a $n \times n$ square matrix A has a characteristic polynomial

$$p(\lambda) = \sum_{i=0}^{n} c_i \lambda^i \,.$$

Then by Cayley–Hamilton theorem,

$$\sum_{i=0}^{n} c_i \mathsf{A}^i = \mathsf{0} \,.$$

By acting a A^{-1} on both sides, we have

$$\sum_{i=0}^{n-1} c_{i+1} \mathsf{A}^i + c_0 \mathsf{A}^{-1} = \mathsf{0}$$

$$\implies \mathsf{A}^{-1} = -\frac{1}{c_0} \sum_{i=0}^{n-1} c_{i+1} \mathsf{A}^i \,.$$

These powers of A are generally much easier to calculate, especially when A is diagonalisable.

4.8 Jordan Normal Form (Non-examinable)

We will quote a few interesting results in this section, but will prove none of them.

4.8.1 Minimal Polynomial

The Cayley–Hamilton theorem states that

$$p(\mathsf{A}) = \prod_{i=1}^{n} (\mathsf{A} - \lambda_i \mathsf{I}) = 0 \,,$$

but this polynomial is not necessarily the 'minimal' one to make A vanish.

Definition 4.65. The *minimal polynomial* of an endomorphism $\mathcal{A} \in \mathrm{End}(V)$ is the non-zero monic (leading coefficient is 1) polynomial $m_A(t)$ of the least degree with coefficients in $\mathbb{F}$ such that $m_A(\mathcal{A}) = 0$.

Lemma 4.66. Let $f(x)$ be a non-zero polynomial with coefficient in $\mathbb{F}$. $f(\mathcal{A}) = 0$ if and only if m_A is a factor of f.

Remark. In particular, the minimal polynomial is a factor of the characteristic polynomial.

Lemma 4.67. For an endomorphism $\mathcal{A} \in \mathbb{F}^n$ and $\lambda \in \mathbb{F}$, the following are equivalent.

(i) λ is an eigenvalue of $\mathcal{A}$;

(ii) λ is a root of the characteristic polynomial p_A;

(iii) λ is a root of the minimal polynomial m_A.

Let us extend our definition on the multiplicity of an eigenvalue.

Definition 4.68. Let $\mathcal{A} \in \mathrm{End}(\mathbb{F}^n)$ be an endomorphism with an eigenvalue λ.

(i) The *algebraic multiplicity* of λ is

$$a_\lambda \coloneqq \text{the multiplicity of } \lambda \text{ as a root of } p_A\,.$$

(ii) The *geometric multiplicity* of λ is

$$g_\lambda \coloneqq \dim E_A(\lambda)\,.$$

(iii) Another useful number is

$$c_\lambda \coloneqq \text{the multiplicity of } \lambda \text{ as a root of } m_A\,.$$

Lemma 4.69. Let $\mathcal{A} \in \mathrm{End}(\mathbb{F}^n)$ be an endomorphism and λ be an eigenvalue of A.

(i) $1 \le g_\lambda \le a_\lambda$;

(ii) $1 \le c_\lambda \le a_\lambda$.

Lemma 4.70. For a square matrix over $\mathbb{C}$, the following are equivalent

(i) A is diagonalisable;

(ii) $a_\lambda = g_\lambda$ for all eigenvalues of A, i.e. $\Delta_\lambda = 0$ for all λ;

(iii) $c_\lambda = 1$ for all eigenvalues of A.

Now, we are ready to introduce the Jordan normal form of a matrix.

Definition 4.71. A matrix $\mathsf{A} \in \mathrm{M}_{n\times n}(\mathbb{C})$ is in *Jordan normal form* (*Jordan canonical form*) if it is a block diagonal matrix

$$\mathsf{A} = \begin{pmatrix} \mathsf{J}_{n_1}(\lambda_1) & 0 & \cdots & 0 \\ 0 & \mathsf{J}_{n_2}(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \mathsf{J}_{n_k}(\lambda_k) \end{pmatrix},$$

where $k \ge 1$, $n_1, \dots, n_k \in \mathbb{N}$ such that $n_1, \dots, n_k \in \mathbb{N}$ such that $\sum_{i=1}^k n_i = n$ and $\lambda_1, \dots, \lambda_k \in \mathbb{C}$ (not necessarily distinct) and $\mathsf{J}_m(\lambda) \in \mathrm{Mat}_{m\times m}(\mathbb{C})$ has the form

$$\mathsf{J}_m(\lambda) \coloneqq \begin{pmatrix} \lambda & 1 & \cdots & 0 \\ 0 & \lambda & \ddots & 0 \\ \vdots & \vdots & \ddots & 1 \\ 0 & 0 & 0 & \lambda \end{pmatrix}.$$

We call $\mathsf{J}_m(\lambda)$ the *Jordan blocks*.

Finally, we have the theorem.

Theorem 4.72 (Jordan normal form). Every matrix $\mathsf{A} \in \mathrm{Mat}_{n\times n}(\mathbb{C})$ is similar to a matrix in Jordan normal form. This matrix in Jordan normal form is uniquely determined by A up to reordering the Jordan blocks.

The Jordan normal form of a matrix satisfies the following properties:

(i) a_λ is the total number of λ on the diagonal.

(ii) g_λ is the number of Jordan normal blocks with diagonal entries λ

(iii) c_λ is the size of the largest block with eigenvalue λ.

Remark. If the matrix is diagonalisable, then its Jordan normal form is trivially a diagonal matrix of eigenvalues. If not, then the matrix in Jordan normal form can immediately show us the multiplicities of its eigenvalues.

Example. Consider a matrix with Jordan normal form

$$\begin{pmatrix} \boxed{1} & \boxed{1} & 0 & 0 & 0 & 0 \\ 0 & \boxed{1} & 0 & 0 & 0 & 0 \\ 0 & 0 & \boxed{1} & 0 & 0 & 0 \\ 0 & 0 & 0 & \boxed{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \boxed{3} & 0 \\ 0 & 0 & 0 & 0 & 0 & \boxed{3} \end{pmatrix}$$

Then we can see that the eigenvalue $\lambda_1 = 1$ has algebraic multiplicity of 3, geometric multiplicity of 2 and $c_1 = 2$. Its eigenspace has defect $\Delta_1 = 1$. The next eigenvalue $\lambda_2 = 2$ is non-degenerate. The eigenvalue $\lambda_3 = 3$ has algebraic multiplicity of 2, geometric multiplicity of 2 and $c_3 = 1$. The eigenspace is 2 dimensional with no defect. The characteristic polynomial is

$$p(t) = (t-1)^3(t-2)(t-3)^2\,,$$

and the minimal polynomial is

$$m(t) = (t-1)^2(t-2)(t-3)\,.$$

4.9 Duality (Non-examinable)

This concept will be especially important if your are doing quantum mechanics.

4.9.1 Dual Spaces

To specify a subspace of $\mathbb{F}^n$, we can write down a set of linear equations that every vector in the subspace satisfies. For example, let

$$U = \left\langle \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \right\rangle \subset \mathbb{F}^3\,,$$

we can see that

$$U = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \,\middle|\, 2x_1 - x_2 = 0\,,\ x_1 - x_3 = 0 \right\}\,.$$

These equations are determined by the kernels of linear maps $\theta : \mathbb{F}^n \to \mathbb{F}$. Moreover if $\theta_1, \theta_2 : \mathbb{F}^n \to \mathbb{F}$ are linear maps that vanish on U and $\lambda, \mu \in \mathbb{F}$, then $\lambda\theta_1 + \mu\theta_2$ vanishes on U. One may study the subspace of linear maps $\mathbb{F}^n \to \mathbb{F}$ that vanish on U.

Definition 4.73. Let V be a vector space over $\mathbb{F}$. The *dual space* of V is the vector space

$$V^* \coloneqq \mathcal{L}(V, \mathbb{F}) = \{\theta : V \to \mathbb{F} \text{ linear}\}$$

with pointwise addition and scalar multiplication.

The elements of V^* are sometimes called *linear forms* or *linear functionals* on V.

Lemma 4.74. Suppose that V is a finite dimensional vector space over $\mathbb{F}$ with basis $(\mathbf{e}_1, \dots, \mathbf{e}_n)$, then its dual space V^* has a basis $(\epsilon_1, \dots, \epsilon_n)$ such that

$$\epsilon_i(\mathbf{e}_j) = \delta_{ij}\,.$$

Proof. We know that to define a linear map it suffices to define it on a basis so there are unique elements $\epsilon_1, \dots, \epsilon_n$ such that $\epsilon_i(\mathbf{e}_j) = \delta_{ij}$. We must show that they span V^* and are linearly independent.

To show that $\{\epsilon_i\}$ do span V^*, we need to show that any linear map $\theta \in V^*$ can be written as

$$\theta = \sum_{i=1}^{n} \lambda_i \epsilon_i \,.$$

Let $\lambda_i = \theta(\mathbf{e}_i) \in \mathbb{F}$, it suffices to show that the two representations agree on the basis $(\mathbf{e}_1, \dots, \mathbf{e}_n)$ of V. We have

$$\sum_{i=1}^{n} \lambda_i \epsilon_i(\mathbf{e})_i = \lambda_j = \theta(\mathbf{e}_j) \,,$$

so the claim is true.

Next, suppose that $\sum \mu_i \epsilon_i = 0 \in V^*$ for some $\mu_1, \dots, \mu_n \in F$. Then $0 = \sum \mu_i \epsilon_i(\mathbf{e}_j) = \mu_j$ for each $j = 1, \dots, n$. Thus $\epsilon_1, \dots, \epsilon_n$ are linearly independent as claimed. □

Corollary. If V is a finite dimensional vector space over $\mathbb{F}$, then $\dim_{\mathbb{F}} V = \dim_{\mathbb{F}} V^*$.

Definition 4.75. We call the basis $(\epsilon_1, \dots, \epsilon_n)$ of V^* such that

$$\epsilon_i(\mathbf{e}_j) = \delta_{ij}$$

the *dual basis* with respect to the basis $(\mathbf{e}_i, \dots, \mathbf{e}_n)$ of V.

Remark. If we think of the elements of V as column vectors with respect to some basis

$$\sum x_i \mathbf{e}_i = \begin{pmatrix} x_i \\ \vdots \\ x_n \end{pmatrix} \,,$$

then we can view elements of V^* as row vectors with respect to the dual basis

$$\sum a_i \epsilon_i = \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix}$$

such that

$$\Big(\sum a_i \epsilon_i\Big)\Big(\sum x_j \mathbf{e}_j\Big) = \sum a_i x_i = \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix} \begin{pmatrix} x_i \\ \vdots \\ x_n \end{pmatrix} \,.$$

Proposition 4.76. Suppose that V is a finite dimensional vector space over $\mathbb{F}$ with bases $(\mathbf{e}_1, \dots, \mathbf{e}_n)$ and $(\mathbf{f}_1, \dots, \mathbf{f}_n)$, and that P is the change of basis matrix from $\{\mathbf{e}_i\}$ to $\{\mathbf{f}_i\}$ such that $\mathbf{f}_i = P_{ki}\mathbf{e}_k$. Let $(\epsilon_1, \dots, \epsilon_n)$ and $(\eta_1, \dots, \eta_n)$ be the corresponding dual bases such that

$$\epsilon_i(\mathbf{e}_j) = \delta_{ij} = \eta_i(\mathbf{f}_j) \,.$$

The change of basis matrix from $\{\epsilon_i\}$ to $\{\eta_i\}$ is given by $(\mathsf{P}^{-1})^{\mathrm{T}}$ such that

$$\epsilon_i = \sum \eta_l P_{il} \,.$$

Proof. Let $\mathsf{Q} = \mathsf{P}^{-1}$. Then $\mathbf{e}_j = Q_{kj}\mathbf{f}_k$, so we can compute

$$\left(\sum_l \eta_l P_{il}\right)(\mathbf{e}_j) = \sum_{k,l} (P_{il}\eta_l)(Q_{jk}\mathbf{f}_k) = \sum_{k,l} P_{il}\delta_{kl}Q_{kj} = \delta_{ij} \,,$$

so $\epsilon_i = \sum_l \eta_l P_{il}$ as claimed. □

4.9.2 Dual Maps

Definition 4.77. Let V and W be vector spaces over $\mathbf{F}$ and suppose that $\alpha : V \to W$ is a linear map. The *dual map* to α is defined to be the map $\alpha^* : W^* \to V^*$, $\theta \mapsto \theta\alpha$.

Proposition 4.78. Let α^* be the dual map $\alpha^* : W^* \to V^*$,

$$\alpha^* \in \mathcal{L}(W^*, V^*)\,.$$

Proof. $\theta\alpha$ is the composite of two linear maps and so is linear. Moreover, if $\lambda, \mu \in \mathbb{F}$, $\theta_1, \theta_2 \in W^*$ and $\mathbf{v} \in V$, then

$$\begin{aligned}\alpha^*(\lambda\theta_1 + \mu\theta_2)(\mathbf{v}) &= (\lambda\theta_1 + \mu\theta_2)\alpha(\mathbf{v}) \\ &= \lambda\theta_1\alpha(\mathbf{v}) + \mu\theta_2\alpha(\mathbf{v}) \\ &= (\lambda\alpha^*(\theta_1) + \mu\alpha^*(\theta_2))(\mathbf{v})\,.\end{aligned}$$

Therefore, $\alpha^*(\lambda\theta_1 + \mu\theta_2) = \lambda\alpha^*(\theta_1) + \mu\alpha^*(\theta_2)$, and α^* is linear, so $\alpha^* \in \mathcal{L}(W^*, V^*)$. □

Lemma 4.79. Suppose that V and W are finite dimensional vector spaces with bases $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ and $(\mathbf{f}_1, \ldots, \mathbf{f}_m)$ respectively. Let $(\epsilon_1, \ldots, \epsilon_n)$ and $(\eta_1, \ldots, \eta_m)$ be the corresponding dual bases. If $\alpha : V \to W$ is represented by A with respect to $\{\mathbf{e}_i\}$ and $\{\mathbf{f}_j\}$, then α^* is represented by A^{T} with respect to $\{\eta_i\}$ and $\{\epsilon_i\}$.

Proof. We are given $\alpha(\mathbf{e}_i) = \sum A_{ki}\mathbf{f}_k$ and must compute $\alpha^*(\eta_i)$ in terms of $\epsilon_1, \ldots, \epsilon_n$.

$$\begin{aligned}\alpha^*(\eta_i)(\mathbf{e}_j) &= \eta_i(\alpha(\mathbf{e}_j)) \\ &= \eta_i\left(\sum_k A_{jk}\mathbf{f}_k\right) \\ &= \sum_k A_{kj}\delta_{ik} = A_{ij}\,.\end{aligned}$$

Thus $\alpha^*(\eta_i)(\mathbf{e}_j) = \sum_k A_{ik}\epsilon_k(\mathbf{e}_j) = \sum_k A^{\mathrm{T}}_{ki}\epsilon_k(\mathbf{e}_k)$ and so $\alpha^*(\eta_i) = \sum_k A^{\mathrm{T}}_{ki}\epsilon_k$ as required. □

4.10 Forms

Definition 4.80. Let U and V be vector spaces of $\mathbb{F}$. A map $\mathcal{F} : U \times V \to \mathbb{F}$ is a *bilinear form* if it is linear in both of its arguments. For $a, b \in \mathbb{C}$, $\mathbf{u}_1, \mathbf{u}_2 \in U$ and $\mathbf{v}_1, \mathbf{v}_2$ in V,

$$\mathcal{F}(a\mathbf{u}_1 + b\mathbf{u}_2, \mathbf{v}_1) = a\mathcal{F}(\mathbf{u}_1, \mathbf{v}_1) + b\mathcal{F}(\mathbf{u}_2, \mathbf{v}_1)\,,$$

$$\mathcal{F}(\mathbf{u}_1, a\mathbf{v}_1 + b\mathbf{v}_2) = a\mathcal{F}(\mathbf{u}_1, \mathbf{v}_1) + b\mathcal{F}(\mathbf{u}_1, \mathbf{v}_2)\,.$$

Definition 4.81. A map $\mathcal{F} : U \times V \to \mathbb{C}$ is a *sesquilinear form* if it is linear in its second argument and anti-linear in its first argument. For $a, b \in \mathbb{C}$, $\mathbf{u}_1, \mathbf{u}_2 \in U$ and $\mathbf{v}_1, \mathbf{v}_2$ in V,

$$\mathcal{F}(a\mathbf{u}_1 + b\mathbf{u}_2, \mathbf{v}_1) = a^*\mathcal{F}(\mathbf{u}_1, \mathbf{v}_1) + b^*\mathcal{F}(\mathbf{u}_2, \mathbf{v}_1)\,,$$

$$\mathcal{F}(\mathbf{u}_1, a\mathbf{v}_1 + b\mathbf{v}_2) = a\mathcal{F}(\mathbf{u}_1, \mathbf{v}_1) + b\mathcal{F}(\mathbf{u}_1, \mathbf{v}_2)\,.$$

Lemma 4.82. Let $\mathcal{F} : U \times V \to \mathbb{C}$ be a sesquilinear form and let $\{\mathbf{u}_i\}$ and $\{\mathbf{v}_j\}$ be the bases of U and V. The form can be represented by a matrix A defined by

$$A_{ij} = \mathcal{F}(\mathbf{u}_i, \mathbf{v}_j)$$

such that for $\mathbf{u} \in U$ and $\mathbf{v} \in V$,

$$\mathcal{F}(\mathbf{u}, \mathbf{v}) = \mathsf{u}^\dagger\mathsf{A}\mathsf{v}\,.$$

Proof. Let

$$\mathbf{u} = \sum_i \lambda_i \mathbf{u}_i \,,\ \mathbf{v} = \sum_j \mu_j \mathbf{v}_j \,.$$

$$\begin{aligned}\mathcal{F}(\mathbf{u},\mathbf{v}) &= \mathcal{F}\left(\sum_i \lambda_i \mathbf{u}_i, \sum_j \mu_j \mathbf{v}_j\right)\\ &= \sum_i \sum_j \lambda_i^* A_{ij} \mu_j\\ &= \mathsf{u}^\dagger \mathsf{A} \mathsf{v}\,.\end{aligned}$$

□

Definition 4.83. A sesquilinear form $\mathcal{F} : V \times V \to \mathbb{C}$ is *Hermitian* if

$$\mathcal{F}(\mathbf{x},\mathbf{y}) = \mathcal{F}(\mathbf{y},\mathbf{x})^*$$

for $\mathbf{x}, \mathbf{y} \in V$.

Remark. An inner product is a positive definite Hermitian form.

Lemma 4.84. A form is Hermitian if and only if its representing matrix is Hermitian.

Proof. If $\mathcal{F}$ is Hermitian then

$$H_{ij} = \mathcal{F}(\mathbf{v}_i,\mathbf{v}_j) = \mathcal{F}(\mathbf{v}_j,\mathbf{v}_i)^* = H_{ji}^* \,.$$

Conversely, if $\mathsf{H} = \mathsf{H}^\dagger$,

$$\mathcal{F}(\mathbf{x},\mathbf{y}) = \mathsf{x}^\dagger \mathsf{H} \mathsf{y} = \mathsf{y}^{\mathrm{T}} \mathsf{H}^{\mathrm{T}} \mathsf{x}^* = (\mathsf{y}^\dagger \mathsf{H} \mathsf{x})^* = \left(\mathcal{F}(\mathbf{y},\mathbf{x})\right)^* .$$

□

Lemma 4.85. A Hermitian form of a vector with itself $\mathcal{F} : V \to \mathbb{C}$ is real.

Proof.

$$\begin{aligned}(\mathsf{x}^\dagger \mathsf{H} \mathsf{x})^* &= (\mathsf{x}^\dagger \mathsf{H} \mathsf{x})^\dagger\\ &= \mathsf{x}^\dagger \mathsf{H}^\dagger \mathsf{x}\\ &= \mathsf{x}^\dagger \mathsf{H} \mathsf{x}\,.\end{aligned}$$

□

An important special case is obtained by restriction to real vector spaces.

Definition 4.86. If $\phi : V \times V \to \mathbb{F}$ is a bilinear form, then we call the map $\mathcal{F} : V \to \mathbb{F}$; $\mathbf{v} \mapsto \phi(\mathbf{v},\mathbf{v})$ a *quadratic form*.

Lemma 4.87. A quadratic form on real vector spaces can be represented by a real, symmetric matrix.

$$\mathcal{F}(\mathbf{x}) = \mathsf{x}^{\mathrm{T}} \mathsf{S} \mathsf{x} = \sum_{i,j=1}^{n} x_i S_{ij} x_j \,.$$

Proof. Restricting to real vector space, a bilinear map of a vector with itself is clearly sesquilinear and Hermitian. Its representing matrix is Hermitian, but should also be real. Therefore, a quadratic form is represented by a real symmetric matrix. □

4.10.1 Eigenvectors and Principal Axes

The coefficient matrix, H, of a Hermitian form can be written as

$$\mathsf{H} = \mathsf{U}\mathsf{\Lambda}\mathsf{U}^\dagger\,,$$

where U is unitary and $\mathsf{\Lambda}$ is a diagonal matrix of eigenvalues. Let

$$\mathsf{x}' = \mathsf{U}^\dagger\mathsf{x}\,,$$

then the Hermitian form of a vector with itself can be written as

$$\begin{aligned}\mathcal{F}(\mathbf{x}) &= \mathsf{x}^\dagger\mathsf{U}\mathsf{\Lambda}\mathsf{U}^\dagger\mathsf{x}\\ &= \mathsf{x}'^\dagger\mathsf{\Lambda}\mathsf{x}'\\ &= \sum_{i=1}^{n}\lambda_i|x_i'|^2\,.\end{aligned}$$

Transforming to a basis of orthonormal eigenvectors transforms the Hermitian form to a standard form with no off-diagonal terms. The orthonormal basis vectors that coincide with the eigenvectors of the coefficient matrix, which lead to the simplified version of the form, are known as the *principal axes*.

Proposition 4.88. For a Hermitian form

$$\mathcal{F}(\mathbf{x}) = \mathsf{x}^\dagger\mathsf{H}\mathsf{x}\,,$$

its principal axes are given by the eigenvectors of the Hermitian matrix H.

Example. Let $\mathcal{F}(\mathbf{x})$ be the quadratic form

$$\mathcal{F}(\mathbf{x}) = 2x^2 - 4xy + 5y^2 = \mathsf{x}^{\mathrm{T}}\mathsf{S}\mathsf{x}\,,$$

where

$$\mathsf{x} = \begin{pmatrix}x\\y\end{pmatrix} \quad\text{and}\quad \mathsf{S} = \begin{pmatrix}2 & -2\\ -2 & 5\end{pmatrix}.$$

Find the surface described by $\mathcal{F}(x) =$ constant.

The eigenvalues of the real symmetric matrix S are $\lambda_1 = 1$ and $\lambda_2 = 6$, with corresponding unit eigenvectors

$$\mathsf{u}_1 = \frac{1}{\sqrt{5}}\begin{pmatrix}2\\1\end{pmatrix} \quad\text{and}\quad \mathsf{u}_2 = \frac{1}{\sqrt{5}}\begin{pmatrix}1\\-2\end{pmatrix}.$$

The orthogonal matrix

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

transforms the original orthonormal basis to a basis of principal axes. Hence $\mathsf{S} = \mathsf{Q}\mathsf{\Lambda}\mathsf{Q}^{\mathrm{T}}$, where $\mathsf{\Lambda}$ is a diagonal matrix of eigenvalues. It follows that $\mathcal{F}$ can be rewritten in the normalised form

$$\mathcal{F} = \mathsf{x}^{\mathrm{T}}\mathsf{Q}\mathsf{\Lambda}\mathsf{Q}^{\mathrm{T}}\mathsf{x} = \mathsf{x}'^{\mathrm{T}}\mathsf{\Lambda}\mathsf{x}' = x'^2 + 6y'^2\,,$$

where

$$\mathsf{x}' = \mathsf{Q}^{\mathrm{T}}\mathsf{x}, \quad \text{i.e.} \quad \begin{pmatrix}x'\\y'\end{pmatrix} = \frac{1}{\sqrt{5}}\begin{pmatrix}2 & 1\\ 1 & -2\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}.$$

Such a surface is therefore an ellipse.

Remark. In diagonalising S by transforming to its eigenvector basis, we are rotating the coordinates to reduce the quadratic form to its simplest form.

4.10.2 Quadrics and Conics

Definition 4.89. A *quadric*, or a *quadric surface*, is the n-dimensional hypersurface defined by the zeros of a real quadratic polynomial. For coordinates $(x_1, \ldots, x_n)$ the general quadric is defined by

$$x_i A_{ij} x_j + b_i x_i + c \equiv \mathsf{x}^{\mathrm{T}} \mathsf{A} \mathsf{x} + \mathsf{b}^{\mathrm{T}} \mathsf{x} + c = 0 \,,$$

under summation convention, where A is a $n \times n$ matrix, b is a $n \times 1$ column vector and c is a constant.

Let

$$\mathsf{S} = \frac{1}{2}(\mathsf{A} + \mathsf{A}^{\mathrm{T}}) \,,$$

then from the above two equations of the quadric,

$$\mathsf{x}^{\mathrm{T}} \mathsf{S} \mathsf{x} + \mathsf{b}^{\mathrm{T}} \mathsf{x} + c = 0 \,.$$

Note that S is real symmetric. By taking the principal axes as basis vectors it follows that

$$\mathsf{x}'^{\mathrm{T}} \mathsf{\Lambda} \mathsf{x}' + \mathsf{b}'^{\mathrm{T}} \mathsf{x}' + c = 0 \,,$$

where $\mathsf{\Lambda} = \mathsf{Q}^{\mathrm{T}} \mathsf{S} \mathsf{Q}$, $\mathsf{b}' = \mathsf{Q}^{\mathrm{T}} \mathsf{b}$ and $\mathsf{x}' = \mathsf{Q}^{\mathrm{T}} \mathsf{x}$. If $\mathsf{\Lambda}$ does not have a zero eigenvalue, then it is invertible and the equation can be simplified further by a translation of the origin

$$\mathsf{x}' \to \mathsf{x}' - \frac{1}{2} \mathsf{\Lambda}^{-1} \mathsf{b}' \,,$$

to obtain

$$\mathsf{x}'^{\mathrm{T}} \mathsf{\Lambda} \mathsf{x}' = k \,,$$

where k is a constant.

Conic sections. First suppose that $n = 2$ and that $\mathsf{\Lambda}$ does not have a zero eigenvalue, then with

$$\mathsf{x}' = \begin{pmatrix} x' \\ y' \end{pmatrix} \quad \text{and} \quad \mathsf{\Lambda} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \,,$$

the simplified quadric formula becomes

$$\lambda_1 x'^2 + \lambda_2 y'^2 = k \,,$$

which is the normalised equation for a conic section.

(i) $\lambda_1 \lambda_2 > 0$. If $\lambda_1 \lambda_2 > 0$, then k must have the same sign as λ_1, and this is an equation of an ellipse with principal axes coinciding with the x' and y' axes.

Scale. The scale of the ellipse is determined by k.

Shape. The shape of the ellipse is determined by the ratio of the eigenvalues.

Orientation. The orientation of the ellipse in the original basis is determined by the eigenvectors of S.

In the degenerate case, $\lambda_1 = \lambda_2$, the ellipse becomes a circle with no preferred principal axes. Any two orthogonal (or just linearly independent) vectors may be chosen as the principal axes.

(ii) $\lambda_1 \lambda_2 < 0$. If $\lambda_1 \lambda_2 < 0$ then this is an equation for a hyperbola with principal axes coinciding with the x' and y' axes.

(iii) $\lambda_1\lambda_2 = 0$. If $\lambda_1 = \lambda_2 = 0$ then there is no quadratic term.

We assume that only one eigenvalue is zero; wlog $\lambda_2 = 0$. Then we alternatively translate the origin according to

$$x' \to x' - \frac{b_1'}{2\lambda_1}\,,\; y' - \frac{c}{b_2'} + \frac{b_1'^2}{4\lambda_1 b_2'}\,,$$

assuming $b_2' \neq 0$, to obtain

$$\lambda_1 x'^2 + b_2' y' = 0\,.$$

This is the equation of a parabola with principal axes coinciding with the x' and y' axes.

If $b_2' = 0$, the equation for the conic section can be reduces to $\lambda_1 x'^2 = k$, with possible solutions of zero ($\lambda_1 k < 0$), one ($k = 0$) or two ($\lambda_1 k > 0$) lines.

Three Dimensions. If $n = 3$ and Λ does not have a zero eigenvalue, then with

$$\mathsf{x}' = \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} \quad \text{and} \quad \Lambda = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}.$$

The simplified quadric equation becomes

$$\lambda_1 x'^2 + \lambda_2 y'^2 + \lambda_3 z'^2 = k\,.$$

When $\lambda_i k > 0$, the distance to the surface along the i^{th} principal axes $= \sqrt{\frac{k}{\lambda_i}}$.

This equation describes a number of characteristic surfaces.

Coefficients	Quadric Surface
$\lambda_1 > 0, \lambda_2 > 0, \lambda_3 > 0, k > 0$	Ellipsoid. This include the case of metric matrices, since S is then positive definite and the λ_i are all positive.
$\lambda_1 = \lambda_2$	Surface of revolution about the z' axis.
$\lambda_1 = \lambda_2 > 0, \lambda_3 > 0, k > 0$	Spheroid: A prolate spheroid if $\lambda_1 = \lambda_2 > \lambda_3$ and an oblate spheroid if $\lambda_1 = \lambda_2 < \lambda_3$.
$\lambda_1 = \lambda_2 = \lambda_3 > 0, k > 0$	Sphere.
$\lambda_1 = \lambda_2 > 0, \lambda_3 = 0, k > 0$	Cylinder.
$\lambda_1 > 0, \lambda_2 > 0, \lambda_3 = 0, k > 0$	Elliptic cylinder.
$\lambda_1 > 0, \lambda_2 > 0, \lambda_3 < 0, k > 0$	Hyperboloid of one sheet.
$\lambda_1 > 0, \lambda_2 > 0, \lambda_3 < 0, k = 0$	Elliptical conical surface.
$\lambda_1 > 0, \lambda_2 < 0, \lambda_3 < 0, k > 0$	Hyperboloid of two sheet.
$\lambda_1 > 0, \lambda_2 = \lambda_3 = 0, \lambda_1 k \geq 0$	Planes $x = \pm\sqrt{\frac{k}{\lambda_1}}$.

4.10.3 The Stationary Properties of the Eigenvalues

Suppose that we have an orthonormal basis, and let x be a point on $\mathsf{x}^{\mathrm{T}}\mathsf{S}\mathsf{x} = k$ where k is a constant. Then the distance squared from the origin to the quadric surface is $\mathsf{x}^{\mathrm{T}}\mathsf{x}$. This distance naturally depends on the value of k. This dependence on k can be removed by considering the square of the relative distance to the surface:

$$(\text{relative distance to surface})^2 = \frac{\mathsf{x}^{\mathrm{T}}\mathsf{x}}{\mathsf{x}^{\mathrm{T}}\mathsf{S}\mathsf{x}}\,.$$

Let us consider the directions for which this relative distance, or equivalently its inverse (referred to as the *Rayleigh quotient*)

$$\lambda(\mathsf{x}) = \frac{\mathsf{x}^{\mathrm{T}}\mathsf{S}\mathsf{x}}{\mathsf{x}^{\mathrm{T}}\mathsf{x}},$$

is stationary. We can find the first variation in $\lambda(x)$ by letting

$$\mathsf{x} \to \mathsf{x} + \delta\mathsf{x} \text{ and } \mathsf{x}^{\mathrm{T}} \to \mathsf{x}^{\mathrm{T}} + \delta\mathsf{x}^{\mathrm{T}},$$

using Taylor expansion, and ignoring terms quadratic or higher in $|\delta\mathsf{x}|$. First note that

$$\begin{aligned}(\mathsf{x}^{\mathrm{T}} + \delta\mathsf{x}^{\mathrm{T}})(\mathsf{x} + \delta\mathsf{x}) &= \mathsf{x}^{\mathrm{T}}\mathsf{x} + \mathsf{x}^{\mathrm{T}}\delta\mathsf{x} + \delta\mathsf{x}^{\mathrm{T}}\mathsf{x} + \dots \\ &= \mathsf{x}^{\mathrm{T}}\mathsf{x} + 2\delta\mathsf{x}^{\mathrm{T}}\mathsf{x} + \dots\end{aligned}$$

Hence

$$\begin{aligned}\frac{1}{(\mathsf{x}^{\mathrm{T}} + \delta\mathsf{x}^{\mathrm{T}})(\mathsf{x} + \delta\mathsf{x})} &= \frac{1}{\mathsf{x}^{\mathrm{T}}\mathsf{x} + 2\delta\mathsf{x}^{\mathrm{T}}\mathsf{x} + \dots} \\ &= \frac{1}{\mathsf{x}^{\mathrm{T}}\mathsf{x}}\left(1 + \frac{2\delta\mathsf{x}^{\mathrm{T}}\mathsf{x}}{\mathsf{x}^{\mathrm{T}}\mathsf{x}} + \dots\right)^{-1} \\ &= \frac{1}{\mathsf{x}^{\mathrm{T}}\mathsf{x}}\left(1 - \frac{2\delta\mathsf{x}^{\mathrm{T}}\mathsf{x}}{\mathsf{x}^{\mathrm{T}}\mathsf{x}} + \dots\right)\end{aligned}$$

Similarly,

$$\begin{aligned}(\mathsf{x}^{\mathrm{T}} + \delta\mathsf{x}^{\mathrm{T}})\mathsf{S}(\mathsf{x} + \delta\mathsf{x}) &= \mathsf{x}^{\mathrm{T}}\mathsf{S}\mathsf{x} + \mathsf{x}^{\mathrm{T}}\mathsf{S}\delta\mathsf{x} + \delta\mathsf{x}^{\mathrm{T}}\mathsf{S}\mathsf{x} + \dots \\ &= \mathsf{x}^{\mathrm{T}}\mathsf{S}\mathsf{x} + 2\delta\mathsf{x}^{\mathrm{T}}\mathsf{S}\mathsf{x} + \dots\end{aligned}$$

Putting them together we have

$$\begin{aligned}\delta\lambda(\mathsf{x}) \equiv \lambda(\mathsf{x} + \delta\mathsf{x}) - \lambda(\mathsf{x}) &= \frac{(\mathsf{x}^{\mathrm{T}} + \delta\mathsf{x}^{\mathrm{T}})\mathsf{S}(\mathsf{x} + \delta\mathsf{x})}{(\mathsf{x}^{\mathrm{T}} + \delta\mathsf{x}^{\mathrm{T}})(\mathsf{x} + \delta\mathsf{x})} - \frac{\mathsf{x}^{\mathrm{T}}\mathsf{S}\mathsf{x}}{\mathsf{x}^{\mathrm{T}}\mathsf{x}} \\ &= \frac{\mathsf{x}^{\mathrm{T}}\mathsf{S}\mathsf{x} + 2\delta\mathsf{x}^{\mathrm{T}}\mathsf{S}\mathsf{x} + \dots}{\mathsf{x}^{\mathrm{T}}\mathsf{x}}\left(1 - \frac{2\delta\mathsf{x}^{\mathrm{T}}\mathsf{x}}{\mathsf{x}^{\mathrm{T}}\mathsf{x}} + \dots\right) - \frac{\mathsf{x}^{\mathrm{T}}\mathsf{S}\mathsf{x}}{\mathsf{x}^{\mathrm{T}}\mathsf{x}} \\ &= \frac{2\delta\mathsf{x}^{\mathrm{T}}\mathsf{S}\mathsf{x}}{\mathsf{x}^{\mathrm{T}}\mathsf{x}} - \frac{\mathsf{x}^{\mathrm{T}}\mathsf{S}\mathsf{x}}{\mathsf{x}^{\mathrm{T}}\mathsf{x}}\frac{2\delta\mathsf{x}^{\mathrm{T}}\mathsf{x}}{\mathsf{x}^{\mathrm{T}}\mathsf{x}} \\ &= \frac{2}{\mathsf{x}^{\mathrm{T}}\mathsf{x}}(\delta\mathsf{x}^{\mathrm{T}}\mathsf{S}\mathsf{x} - \lambda(\mathsf{x})\delta\mathsf{x}^{\mathrm{T}}\mathsf{x}) \\ &= \frac{2}{\mathsf{x}^{\mathrm{T}}\mathsf{x}}\delta\mathsf{x}^{\mathrm{T}}(\mathsf{S}\mathsf{x} - \lambda(\mathsf{x})\mathsf{x})\end{aligned}$$

Theorem 4.90 (The Rayleigh–Ritz variational principle). The eigenvectors of S are the directions that make the relative distance to the quartic surface stationary, and the eigenvalues are the values of

$$\lambda(x) = \frac{\mathsf{x}^{\mathrm{T}}\mathsf{x}}{\mathsf{x}^{\mathrm{T}}\mathsf{x}}$$

at the stationary points.

Proof. The first variation of $\lambda(\mathsf{x})$ is zero for all possible $\delta\mathsf{x}$ when

$$\mathsf{S}\mathsf{x} = \lambda(\mathsf{x})\mathsf{x},$$

i.e. when x is an eigenvector of S and λ is the associated eigenvalue. □

Corollary. Similarly, it can be shown that the eigenvalues of a Hermitian matrix H are the values of the function

$$\lambda(x) = \frac{\mathsf{x}^{\mathrm{T}}\mathsf{H}\mathsf{x}}{\mathsf{x}^{\mathrm{T}}\mathsf{x}}$$

at its stationary points.

5 Elementary Analysis

5.1 Sequences and Limits

5.1.1 Sequences

Definition 5.1. A *sequence* is a function $s : \mathbb{N} \to \mathbb{R}$ (or $\mathbb{C}$) usually written as s_n instead of $s(n)$ as an ordered list of numbers.

5.1.2 Behaviour of sequences

Possible behaviours of a sequence s_n as n increases are:

(i) s_n tends towards a particular value;

(ii) s_n does not tend to any value but remains limited in magnitude;

(iii) s_n is unlimited in magnitude.

Definition 5.2. A sequence s_n is said to tend to the *limit* s if given any positive ϵ, there exists $N \equiv N(\epsilon)$ such that $|s_n - s| < \epsilon$ for all $n > N$. We write this as

$$\lim_{n\to\infty} s_n = s \quad \text{or as} \quad s_n \to s \text{ as } n \to \infty\,.$$

In other words, the members of the sequence are eventually contained within an arbitrarily small disk centred on s.

Theorem 5.3 (Cauchy's principle of convergence). A sequence $s_n \in \mathbb{R}$ or $\mathbb{C}$ is convergent if and only if for any positive number ϵ, there exists $N > 0$ such that for all $n \geq N$ and $m \geq 1$, $\left|s_{(n+m)} - s_n\right| < \epsilon$. Such a sequence is said to be *Cauchy*.

Remark. For sequences in other domains, convergence $\implies$ Cauchy but a Cauchy sequence is not necessarily convergent.

Definition 5.4. The sequence s_n is *bounded* as $n \to \infty$ if there exists a positive number K and a positive integer N such that $|s_n| \leq K$ for all $n \geq N$.

Definition 5.5. A sequence a_n is *increasing* if $a_n \leq a_{n+1}$ for all n. It is *strictly increasing* if $a_n < a_{n+1}$ for all n. *(Strictly) decreasing* sequences are defined analogously.

A sequence is *(strictly) monotone* if it is (strictly) increasing or (strictly) decreasing.

Definition 5.6. A sequence is said to *tend to infinity* if given any $A \in \mathbb{R}$ (however large), there exists $N \in \mathbb{N}$ such that $s_n > A$ for all $n > N$. We then write $s_n \to \infty$ as $n \to \infty$.

Similarly, we say that $s_n \to -\infty$ as $n \to \infty$ if given any A (however large), there exists $N \equiv N(A)$ such that $s_n < -A$ for all $n > N$.

Proposition 5.7. If $s_{n+1} > s_n$, and $s_n < K \in \mathbb{R}\,\forall\, n$, then $s = \lim_{n\to\infty} s_n$ exists.

Proof. A monotone sequence tends either to a limit or to $\pm\infty$. Hence a bounded monotone sequence tends to a limit. □

Definition 5.8. If a sequence does not tend to a limit or $\pm\infty$, then it is said to *oscillate*. If s_n oscillates and is bounded, it oscillates finitely, otherwise, it oscillates infinitely.

5.2 Convergence of Infinite Series

5.2.1 Convergent and Divergent Series

Definition 5.9. Given an infinite sequence of numbers $u_1, u_2, \ldots$, the *partial sum* s_n is defined by

$$s_n = \sum_{r=1}^{n} u_r \,.$$

Definition 5.10. If as $n \to \infty$, s_n tends to a finite limit s, then we say that the infinite series

$$s = \sum_{r=1}^{\infty} u_r \,,$$

converges, and that s is its sum.

An infinite series which is not *convergent* is called *divergent*.

Remarks.

- Whether a series converges or diverges depends on the behaviour of the terms u_n as n tends to infinity.
- According to Cauchy's principle of convergence, a necessary and sufficient condition for $\sum u_r$ to converge is that, for any positive number ϵ,

  $$|s_{n+m} - s_n| = |u_{n+1} + u_{n+2} + \cdots + u_{n+m}| < \epsilon$$

  for all positive integers m, for sufficiently large n.

Lemma 5.11. The geometric series

$$\sum_{r=0}^{\infty} z^r = 1 + z + z^2 + z^3 + \ldots \,,$$

converges to $(1-z)^{-1}$ provided that $|z| < 1$.

Proof. Consider the partial sum

$$s_n = 1 + z + \cdots + z^{n-1} = \frac{1 - z^n}{1 - z} \,.$$

If $|z| < 1$, then we have that $z^n \to 0$ as $n \to \infty$, and hence

$$s = \lim_{n\to\infty} s_n = \frac{1}{1-z} \quad \text{for } |z| < 1 \,.$$

However if $|z| \geq 1$ the series diverges. □

Lemma 5.12. The harmonic series,

$$\sum_{n=1}^{\infty} \frac{1}{n} \,,$$

diverges.

Proof. Consider the partial sum

$$s_n = \sum_{r=1}^{n} \frac{1}{r} = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \,.$$

Then

$$s_n > \int_1^{n+1} \frac{1}{x}\,\mathrm{d}x = \ln(n+1) \,.$$

Therefore s_n increases without bound and does not tend to a limit as $n \to \infty$. □

Lemma 5.13. A necessary condition for s to converge is that $u_r \to 0$ as $r \to \infty$.

Proof. Use the fact that $u_r = s_r - s_{r-1}$ we have that

$$\lim_{r\to\infty} u_r = \lim_{r\to\infty} (s_r - s_{r-1}) = \lim_{r\to\infty} s_r - \lim_{r\to\infty} s_{r-1} = s - s = 0\,.$$

□

Remark. However, $u_r \to 0$ as $r \to \infty$ is not a sufficient condition for convergence. The harmonic series is an example.

5.2.2 Absolute and Conditional Convergence

Definition 5.14. A series $\sum u_r$ is said to *converge absolutely* if

$$\sum_{r=1}^{\infty} |u_r|$$

converges.

Lemma 5.15. If $\sum |u_r|$ converges, then $\sum u_r$ also converges.

Proof. If $\sum |u_r|$ converges, then for any positive number ϵ, $|u_{n+1}| + |u_{n+2}| + \cdots + |u_{n+m}| < \epsilon$ For all positive integers m, for sufficiently large n. But then

$$|u_{n+1} + u_{n+2} + \cdots + u_{n+m}| \leq |u_{n+1}| + |u_{n+2}| + \cdots + |u_{n+m}| < \epsilon\,,$$

and so $\sum u_r$ also converges. □

Definition 5.16. If $\sum |u_r|$ diverges but $\sum u_r$ converges, then the series is said to *converge conditionally.*

Lemma 5.17. The *Alternating harmonic series*,

$$\sum_{r=1}^{\infty} (-1)^{r-1} \frac{1}{r} = 1 - \frac{1}{2} + \frac{1}{3} + \cdots\,,$$

converges conditionally.

Proof. From the Taylor expansion

$$\log(1+x) = -\sum_{r=1}^{\infty} \frac{(-x)^r}{r}\,,$$

we spot that $s = \log 2$. Hence $\sum_{r=1}^{\infty} u_r$ converges. However, $\sum_{r=1}^{\infty} |u_r|$ diverges, so the series is conditionally convergent. □

5.3 Tests of Convergence

5.3.1 The Comparison Test

This test applies to series of non-negative real numbers.

Theorem 5.18 (The comparison test). Suppose given that $v_r > 0$ and that $S = \sum_{r=1}^{\infty} v_r$ is convergent. If

$$0 \leq u_r \leq K v_r$$

for some K independent of r. Then the series $\sum_{r=1}^{\infty} u_r$ is also convergent.

Proof. Since $u_r > 0$, $s_n = \sum_{r=1}^{n} u_r$ is an increasing sequence. Further

$$s_n = \sum_{r=1}^{n} u_r < K \sum_{r=1}^{n} v_r \,,$$

and thus

$$\lim_{n\to\infty} s_n < K \sum_{r=1}^{\infty} v_r = KS \,,$$

i.e. s_n is an increasing bounded sequence. Therefore, $\sum_{r=1}^{\infty} u_r$ is convergent. □

Remark. Similarly, if $\sum_{r=1}^{\infty} v_r$ diverges, $v_r > 0$ and $u_r > K v_r$ for some K independent of r, then $\sum_{r=1}^{\infty} u_r$ diverges.

5.3.2 D'Alembert's Ratio Test

Theorem 5.19 (D'Alembert's ratio test). Suppose that the u_r are real and positive, $u_r > 0$. Define the ratio of successive terms to be

$$\varrho_r = \frac{u_{r+1}}{u_r} \,,$$

and suppose that ϱ_r tends to a limit ϱ as $r \to \infty$, i.e.

$$\lim_{r\to\infty} \frac{u_{r+1}}{u_r} = \varrho \,.$$

Then $\sum u_r$ converges if $\varrho < 1$ and diverges if $\varrho > 1$.

Proof.

- $\varrho < 1$. For the case $\varrho < 1$, choose σ with $\varrho < \sigma < 1$. Then there exists $N \equiv N(\sigma)$ such that

  $$\frac{u_{r+1}}{u_r} < \sigma \text{ for all } r > N \,.$$

  It follows that

  $$\begin{aligned} \sum_{r=1}^{\infty} u_r &= \sum_{r=1}^{N} u_r + u_{N+1}\left(1 + \frac{u_{N+2}}{u_{N+1}} + \frac{u_{N+3}}{u_{N+2}} + \dots\right) \\ &< \sum_{r=1}^{N} u_r + u_{N+1}\left(1 + \sigma + \sigma^2 + \dots\right) \\ &< \sum_{r=1}^{N} u_r + \frac{u_{N+1}}{1-\sigma} \,. \end{aligned}$$

  We conclude that $\sum_{r=1}^{\infty} u_r$ is bounded. Then, since $s_n = \sum_{r=1}^{n} u_r$ is an increasing sequence, it follows that $\sum u_r$ converges.

- $\varrho > 1$. For the case $\varrho > 1$, choose τ with $\varrho > \tau > 1$. Then there exists $M \equiv M(\tau)$ such that

  $$\frac{u_{r+1}}{u_r} > \tau > 1 \text{ for all } r > M \,,$$

  and hence

  $$\frac{u_r}{u_M} > \tau^{r-M} > 1 \text{ for all } r > M \,.$$

  Thus, since $u_r \not\to 0$ as $r \to \infty$, we conclude that $\sum u_r$ diverges. □

Corollary. A series $\sum u_r$ of complex terms converges if the limit of the absolute ratio of successive terms is less than one:

$$\lim_{r\to\infty}\left|\frac{u_{r+1}}{u_r}\right| = \rho < 1\,.$$

Proof. D'Alembert's ratio test shows that $\sum u_r$ converges absolutely, so $\sum u_r$ must converge. □

Remarks.

- If $\varrho = 1$, then nothing can be concluded. The series may converge or not, and a different test is required.
- The ratio test cannot be used for series in which some of the terms are zero. However, it may be adapted by relabelling the series to remove the vanishing terms.

5.3.3 Cauchy's Test

Theorem 5.20 (Cauchy's test). Suppose that the $u_r > 0$ and that

$$\lim_{r\to\infty} u_r^{\frac{1}{r}} = \varrho\,.$$

Then $\sum u_r$ converges if $\varrho < 1$, while $\sum u_r$ diverges if $\varrho > 1$.

Proof. First, suppose that $\varrho < 1$. Choose σ with $\varrho < \sigma < 1$. Then there exists $N \equiv N(\sigma)$ such that

$$u_r^{\frac{1}{r}} < \sigma, \text{ i.e. } u_r < \sigma^r \text{ for all } r > N\,.$$

It follows that

$$\sum_{r=1}^{\infty} u_r < \sum_{r=1}^{N} u_r + \sum_{r=N+1}^{\infty} \sigma^r\,.$$

We conclude that $\sum_{r=1}^{\infty} u_r$ is bounded (since $\sigma < 1$). Moreover $s_n = \sum_{r=1}^{n} u_r$ is an increasing sequence, and hence $\sum u_r$ converges.

Next, suppose that $\varrho > 1$. Choose τ with $1 < \tau < \varrho$. Then there exists $M \equiv M(\tau)$ such that

$$u_r^{\frac{1}{r}} > \tau > 1, \text{ i.e. } u_r > \tau^r > 1, \text{ for all } r > M\,.$$

Thus, since $u_r \not\to 0$ as $r \to \infty$, $\sum u_r$ must diverge. □

5.4 Functions of a Continuous Variable

5.4.1 Limits and Continuity

Let $f : U \to \mathbb{R}$ or $\mathbb{C}$, where $U \subseteq \mathbb{R}$ or $\mathbb{C}$.

Definition 5.21. The function $f(z)$ tends to the *limit* L as $z \to z_0$ if for any $\epsilon > 0$, $\exists\,\delta > 0$ such that $|f(z) - L| < \epsilon$ for all $0 < |z - z_0| < \delta$. We write this as

$$\lim_{z\to z_0} f(z) = L \quad \text{or} \quad f(z) \to L \text{ as } z \to z_0\,.$$

Definition 5.22. The function $f(z)$ is *continuous* at the point $z = z_0$ if $f(z) \to f(z_0)$ as $z \to z_0$.

Remark. The notion of limit and continuity can be easily generalised to $f : U \to \mathbb{F}^n$, where $\mathbb{F}$ is $\mathbb{R}$ or $\mathbb{C}$ and $U \subseteq \mathbb{F}^m$.

Definition 5.23. The function $f(z)$ is *bounded* as $z \to z_0$ if there exist positive numbers K and δ such that $|f(z)| < K$ for all z with $0 < |z - z_0| < \delta$.

Definition 5.24. The function $f(z)$ tends to the *limit* L as $z \to \infty$ if for any $\epsilon > 0$, $\exists\, R > 0$ such that $|f(z) - L| < \epsilon$ for all $|z| > R$. We write this as

$$\lim_{z\to\infty} f(z) = L \quad \text{or} \quad f(z) \to L \text{ as } z \to \infty\,.$$

Definition 5.25. The function $f(z)$ is *bounded* as $z \to \infty$ (or *eventually bounded*) if there exist positive numbers K and R such that $|f(z)| < K$ for all z with $|z| > R$.

Warning: approaches to a point. There are different ways in which z can approach z_0 or ∞, especially in the complex plane. Sometimes the limit or bound applies only if the point is approached in a particular way.

For example, consider $\tanh(z)$ as $|z| \to \infty$ for real z:

$$\lim_{z\to+\infty} \tanh z = 1\,, \qquad \lim_{z\to-\infty} \tanh z = -1\,.$$

However, if z approaches infinity along the imaginary axis ($z \to \pm i\infty$), the limit of $\tanh z$ is not defined.

Remark. In the context of real variables, $x \to \infty$ usually means specifically $x \to +\infty$. One-side limits of a function $f(z)$ at $z = z_0$ are denoted by

$$\lim_{z\to z_0^+} f(z) \text{ and } \lim_{z\to z_0^-} f(z)\,.$$

5.4.2 The O Notation

The symbols O, o and $\sim$ are often used to compare the rate of growth or decay of different functions.

Definition 5.26. Suppose that $f(z)$ and $g(z)$ are functions of z, then

(i) if $\dfrac{f(z)}{g(z)}$ is bounded	as $z \to z_0$, we say that $f(z) = O(g(z))$	as $z \to z_0$;
(ii) if $\dfrac{f(z)}{g(z)} \to 0$	as $z \to z_0$, we say that $f(z) = o(g(z))$	as $z \to z_0$;
(iii) if $\dfrac{f(z)}{g(z)} \to 1$	as $z \to z_0$, we say that $f(z) \sim g(z)$	as $z \to z_0$.

Remarks.

- Only $f(z) \sim g(z)$ is a symmetric relation.
- If $f(z) \sim g(z)$ we say that $f(z)$ is asymptotically equal to $g(z)$.

5.5 Differentiability

Definition 5.27. Let $U \subseteq \mathbb{R}$. A function $f : U \to \mathbb{R}$ is *differentiable* at $x = x_0$ with a *derivative* $f'(x_0)$ if the limit

$$\lim_{x\to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

exists and is equal to $f'(x_0)$.

The derivative of f as a function of x is denoted

$$f'(x) \quad \text{or} \quad \frac{\mathrm{d}f}{\mathrm{d}x}\,.$$

Alternatively, we can write

$$\frac{f(x+h)-f(x)}{h} = f'(x) + \epsilon(h)\,,$$

where $\epsilon(h) \to 0$ as $h \to 0$.

Proposition 5.28.

$$f(x+h) = f(x) + hf'(x) + o(h)\,.$$

Proof. Rearranging the above expression. □

Remark. This can be seen as an approximation of $f(x+h)$ for small h.

Definition 5.29. The *multiple derivatives* are defined recursively. f is $n+1$-times differentiable if it is n-times differentiable and its n-th derivative, denoted

$$f^{(n)}(x) \quad \text{or} \quad \frac{\mathrm{d}^n f}{\mathrm{d}x^n}\,,$$

is differentiable. The $(n+1)$-th derivative of f is the derivative of its n-th derivative.

Definition 5.30. A function f is of C^n class if $f', \ldots, f^{(n)}$ exist and are continuous.

We can extend this to higher dimensions (multiple variables).

Definition 5.31. Let $U \subset \mathbb{R}^m$ and $f : U \to \mathbb{R}^n$. We say f is *differentiable* at $\mathbf{x}_0 \in U$ if there is a linear map $T : \mathbb{R}^m \to \mathbb{R}^n$ and a function $\epsilon : \{\mathbf{h} \in \mathbb{R}^m \mid \mathbf{x}_0 + \mathbf{h} \in U\} \to \mathbb{R}^n$ such that

$$f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + T(\mathbf{h}) + \epsilon(\mathbf{h})\|\mathbf{h}\|\,,$$

where $\epsilon \to 0$ as $\mathbf{h} \to \mathbf{0}$. T is the derivative of f.

Remark. These definitions continue to work well when we extend the codomain of our functions to $\mathbb{C}$. However, we still require our domain to be a subset of $\mathbb{R}$ or $\mathbb{R}^n$.

Differentiating functions with a complex variable is a bit more subtle. We will do this in later chapters.

Proposition 5.32 (Sum, product, quotient and chain rule). Let f, g be differentiable.

$$(f+g)'(x) = f'(x) + g'(x)$$

$$(f \times g)'(x) = f'(x)g(x) + f(x)g'(x)$$

$$\left(\frac{f}{g}\right)' = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}$$

$$(f \circ g)'(x) = f'(g(x))g'(x)$$

5.6 Taylor's Theorem for Functions of a Real Variable

Theorem 5.33 (Taylor's theorem). Let $f(x)$ be a (real or complex) function of a real variable x, and

$$f(x) \in C^n[x_0, x_0 + h]\,.$$

Then the Taylor series of $f(x_0 + h)$ is given by

$$f(x_0 + h) = f(x_0) + hf'(x_0) + \frac{h^2}{2!}f''(x_0) + \cdots + \frac{h^{n-1}}{(n-1)!}f^{(n-1)}(x_0) + R_n\,,$$

where the remainder after n terms, R_n, is obtained by integration by parts to be

$$R_n = \int_{x_0}^{x_0+h} \frac{(x_0 + h - x)^{n-1}}{(n-1)!} f^{(n)}(x)\,\mathrm{d}x\ .$$

Remark. The remainder term can be expressed in alternative ways. Lagrange's expression for the remainder is

$$R_n = \frac{h^n}{n!}f^{(n)}(\xi)\,,$$

where ξ is an unknown number in the interval $x_0 < \xi < x_0 + h$. It follows that

$$R_n = O(h^n)\,.$$

Definition 5.34. A function $f(x)$ is *smooth* in some region if it is infinitely differentiable there. We denote this as

$$f(x) \in C^\infty\,.$$

Definition 5.35. A function $f(x)$ smooth in $x_0 \le x \le x_0 + h$ is *analytic* if it is locally given by a convergent infinite Taylor series:

$$f(x_0 + h) = \sum_{n=0}^{\infty} \frac{h^n}{n!} f^{(n)}(x_0)\,.$$

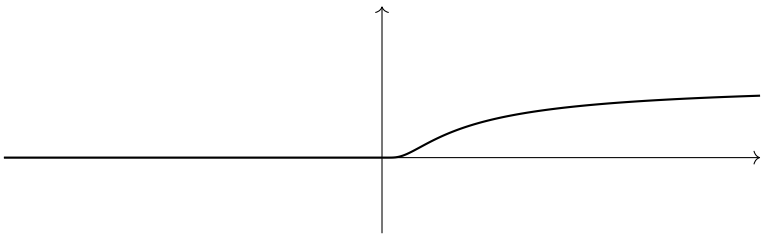
This power series in h converges for sufficiently small h.

Remark. A smooth function is not necessarily analytic.

Example. Consider

$$f(x) = \begin{cases} 0 & x \le 0 \\ e^{-\frac{1}{x}} & x > 0\,. \end{cases}$$

This function is smooth everywhere and its derivatives of all orders are 0 at $x = 0$. Therefore, the Taylor series of f about $x = 0$ is $f = 0$, which fails to converge to f for arbitrarily small $x > 0$.



We can even construct functions that are smooth but not analytic anywhere. An example is the Fourier series

$$g(x) = \sum_{k\in\mathbb{N}} e^{-\sqrt{2^k}} \cos\bigl(2^k x\bigr)\,.$$

5.7 Riemann Integration

Before introducing integration, let us first introduce some definitions.

Definition 5.36. A *dissection*, or *partition* D of the interval $[a, b]$ is a finite set of points $\{t_0, \ldots, t_N\}$ such that

$$a = t_0 < t_1 < \cdots < t_N = b\,.$$

Definition 5.37. The *supremum* of a set S is the smallest number a such that $s \leq a$ for all $s \in S$, written as

$$\sup S = a\,.$$

The *infimum* of a set S is the largest number b such that $s \geq b$ for all $s \in S$, written as

$$\inf S = b\,.$$

Definition 5.38. The *modulus*, *gauge* or *norm* of a dissection D, written $|D|$, is defined to be the supremum of the sub-intervals $t_j - t_{j-1}$ of D.

$$|D| \coloneqq \sup_{1 \leq j \leq N} |t_j - t_{j-1}|\,.$$

We might have seen the following definition of integral.

Definition 5.39 (Ill-defined). The *integral* of a function $f(x)$ in the interval $[a, b]$ is defined as

$$\int_a^b f(t)\,\mathrm{d}t \coloneqq \lim_{N\to\infty} \sum_{j=1}^{N} f(a + jh)h\,, \text{ where } h = \frac{b-a}{N}\,.$$

Remark. This definition of integral involves uniform dissections of the interval $[a, b]$.

This definition of integration is fine for well-behaved functions. However, for some pathological functions, this definition fails to work. Consider the following function.

Definition 5.40. The *Dirichlet function* is defined as

$$D(x) \coloneqq \begin{cases} 0 & \text{if } x \text{ is irrational} \\ 1 & \text{if } x \text{ is rational}\,. \end{cases}$$

Consider the integral

$$\int_0^b D(x)\,\mathrm{d}x\ .$$

If

- $b = \pi$, then the integral evaluates to 0;
- b is a rational approximation of π, then the integral evaluates to b.

We can choose our upper bound to be arbitrarily close to π, but the integral does not approach 0, but rather approaches π. This may suggest that our definition of the integral (Definition 5.39) is ill-defined.

We may suggest a better definition of integration.

Definition 5.41. A *Riemann sum*, $\sigma(D, \zeta)$ for a bounded function $f(t)$ is the sum

$$\sigma(D, \zeta) \coloneqq \sum_{j=1}^{N} f(\zeta_j)(t_j - t_{j-1}) \text{ where } \zeta_j \in [t_{j-1}, t_j]\,.$$

Remark. Different from the ill-defined integral (Definition 5.39), the dissection D and the choice ζ_j are arbitrary for a Riemann sum.

Definition 5.42. A bounded function $f(t)$ is *Riemann integrable* if given any $\epsilon > 0$, there exists $I \in \mathbb{R}$ and $\delta > 0$ such that for all dissections D with $|D| < \delta$ and all choices of ζ,

$$|\sigma(D, \zeta) - I| < \epsilon \,.$$

Definition 5.43. For a Riemann integrable function f, the *Riemann definite integral* of f over the interval $[a, b]$ is the limiting value of the Riemann sum

$$\int_a^b f(t)\,\mathrm{d}t \coloneqq \lim_{|D|\to 0} \sigma(D, \zeta) = I \,.$$

5.7.1 Properties of the Riemann Integral

Proposition 5.44. If f and g are Riemann integrable, $a < c < b$ and $k \in \mathbb{R}$, we have the following properties of Riemann integrals.

$$\int_a^b f(t)\,\mathrm{d}t = -\int_b^a f(t)\,\mathrm{d}t \ ,$$

$$\int_a^b f(t)\,\mathrm{d}t = \int_a^c f(t)\,\mathrm{d}t + \int_c^b f(t)\,\mathrm{d}t \ ,$$

$$\int_a^b kf(t)\,\mathrm{d}t = k\int_a^b f(t)\,\mathrm{d}t \ ,$$

$$\int_a^b f(t) + g(t)\,\mathrm{d}t = \int_a^b f(t)\,\mathrm{d}t + \int_a^b g(t)\,\mathrm{d}t \ ,$$

$$\left|\int_a^b f(t)\,\mathrm{d}t\right| \leq \int_a^b |f(t)|\,\mathrm{d}t \ .$$

Theorem 5.45 (Cauchy–Schwarz inequality). For real, integrable functions f and g,

$$\left(\int_a^b fg\,\mathrm{d}t\right)^2 \leq \left(\int_a^b f^2\,\mathrm{d}t\right)\left(\int_a^b g^2\,\mathrm{d}t\right).$$

Proof. For arbitrary $\lambda \in \mathbb{R}$, we have

$$0 \leq \int_a^b (\lambda f + g)^2\,\mathrm{d}t = \lambda^2 \int_a^b f^2\,\mathrm{d}t + 2\lambda \int_a^b fg\,\mathrm{d}t + \int_a^b g^2\,\mathrm{d}t \ .$$

If $\int_a^b f^2\,\mathrm{d}t = 0$, then

$$2\lambda \int_a^b fg\,\mathrm{d}t + \int_a^b g^2\,\mathrm{d}t \geq 0 \,.$$

This can only be true for all λ if $\int_a^b fg\,\mathrm{d}t = 0$. The equality follows.

If $\int_a^b f^2\,\mathrm{d}t \neq 0$, then choose

$$\lambda = -\frac{\displaystyle\int_a^b fg\,\mathrm{d}t}{\displaystyle\int_a^b f^2\,\mathrm{d}t} \,,$$

and the inequality again follows. □

5.7.2 The Fundamental Theorems of Calculus

Lemma 5.46. If f is bounded and Riemann integrable on an interval $[a, b]$, then

$$F(x) = \int_a^x f(t)\,\mathrm{d}t$$

is continuous on $[a, b]$.

Proof.

$$\begin{aligned} |F(x+h) - F(x)| &= \left| \int_x^{x+h} f(t)\,\mathrm{d}t \right| \\ &\leq \int_x^{x+h} |f(t)|\,\mathrm{d}t \\ &\leq \left(\sup_{t\in[x,x+h]} |f(t)| \right) |h|\,, \end{aligned}$$

and hence

$$\lim_{h\to 0} |F(x+h) - F(x)| = 0\,.$$

Therefore $F(x)$ is continuous by definition. □

Theorem 5.47 (The first fundamental theorem of calculus). If $f(x)$ is continuous, then

$$\frac{\mathrm{d}F}{\mathrm{d}x} \equiv \frac{\mathrm{d}}{\mathrm{d}x}\left(\int_a^x f(t)\,\mathrm{d}t \right) = f(x)\,.$$

Proof.

$$\frac{F(x+h) - F(x)}{h} = \frac{1}{h}\int_x^{x+h} f(t)\,\mathrm{d}t\ .$$

Let $\epsilon > 0$. Since f is continuous at x, then there exists δ such that $|y - x| < \delta$ implies $|f(y) - f(x)| < \epsilon$. If $|h| < \delta$, then

$$\begin{aligned} \left| \frac{1}{h}\int_x^{x+h} f(t)\,\mathrm{d}t - f(x) \right| &= \left| \frac{1}{h}\int_x^{x+h} (f(t) - f(x))\,\mathrm{d}t \right| \\ &\leq \frac{1}{|h|}\left| \int_x^{x+h} |f(t) - f(x)|\,\mathrm{d}t \right| \\ &\leq \frac{\epsilon|h|}{|h|} = \epsilon\,. \end{aligned}$$

□

Theorem 5.48 (The second fundamental theorem of calculus). If g is differentiable then

$$\int_a^x \frac{\mathrm{d}g}{\mathrm{d}t}\,\mathrm{d}t = g(x) - g(a)\,.$$

Proof. Let

$$f(x) = \frac{\mathrm{d}g}{\mathrm{d}x}(x)\,,\ F(x) = \int_a^x f(t)\,\mathrm{d}t\ .$$

Using Theorem 5.47, we have

$$\frac{\mathrm{d}}{\mathrm{d}x}(F - g) = 0\,.$$

This implies that $F - g$ is a constant function, so

$$F(x) - g(x) = F(a) - g(a) = -g(a)\,,$$

and therefore

$$F(x) = \int_a^x \frac{\mathrm{d}g}{\mathrm{d}t}\,\mathrm{d}t = g(x) - g(a)\,.$$

□

5.7.3 Indefinite and Improper Integrals

Definition 5.49. Suppose f is integrable and $f(x) = F'(x)$ for some function F. We define the *indefinite integral* of f to be

$$\int^x f(t)\,\mathrm{d}t \coloneqq F(x) + c\,,$$

where c is an arbitrary constant.

Definition 5.50. Suppose that we have a function $f : [a, b] \to \mathbb{R}$ such that, for every $\epsilon > 0$, f is integrable on $[a + \epsilon, b]$ and

$$\lim_{\epsilon \to 0} \int_{a+\epsilon}^b f(x)\,\mathrm{d}x$$

exists. Then we define the improper integral

$$\int_a^b f(x)\,\mathrm{d}x \coloneqq \lim_{\epsilon \to 0} \int_{a+\epsilon}^b f(x)\,\mathrm{d}x$$

even if the Riemann integral does not exist.

Definition 5.51. The integral to infinity is defined as

$$\int_a^\infty f(x)\,\mathrm{d}x \coloneqq \lim_{b \to \infty} \int_a^b f(x)\,\mathrm{d}x\ ,$$

$$\int_{-\infty}^\infty f(x)\,\mathrm{d}x \coloneqq \lim_{\substack{a \to -\infty \\ b \to \infty}} \int_a^b f(x)\,\mathrm{d}x\ .$$

Remark. When integrating from $-\infty$ to ∞, the two limits need to both converge. Integral like

$$\int_{-\infty}^\infty \frac{x}{1 + x^2}$$

does not converge as both

$$\int_a^\infty \frac{x}{1 + x^2} \text{ and } \int_{-\infty}^b \frac{x}{1 + x^2}\,\mathrm{d}x$$

does not converge for any finite $a, b \in \mathbb{R}$, although this function is odd and it is tempting to say that the integral is 0.

5.8 Convergence of Functions (Non-examinable)

Definition 5.52. A sequence of functions $\{f_n\}$ with the same domain X and codomain Y is said to *converge pointwise* to a given function $f : X \to Y$, written as

$$\lim_{n \to \infty} f_n = f \text{ pointwise}\,,$$

if and only if

$$\lim_{n \to \infty} f_n(x) = f(x)$$

for every x in the domain of f.

This is an easy definition that is simple to check. However, there is a problem. Ideally, we want to deduce properties of f from properties of f_n. For example, it would be great if the continuity of all f_n implies continuity of f, and similarly for integrability and values of derivatives and integrals. However, it turns out we cannot. The notion of pointwise convergence is too weak.

Example. Consider a sequence of functions $f_n : [-1, 1] \to \mathbb{R}$ defined by $f_n = x^{1/(2n+1)}$. These are all continuous, but their pointwise limit function is

$$f_n(x) \to f(x) = \begin{cases} 1 & x \in (0, 1] \\ 0 & x = 0 \\ -1 & x \in [-1, 0) \end{cases},$$

which is not continuous. The continuity of functions is not preserved.

Example. Consider a sequence of functions $f_n : [0, 1] \to \mathbb{R}$ be the piecewise linear function formed by joining the points $(0, 0)$, $(\frac{1}{n}, n)$, $(\frac{2}{n}, 0)$, $(1, 0)$. The pointwise limit of this function is $f_n(x) \to f(x) = 0$. However, we have

$$\int_0^1 f_n(x)\,\mathrm{d}x = 1 \text{ for all } n, \text{ but } \int_0^1 f(x)\,\mathrm{d}x = 0\,.$$

The limit of the integral is not the integral of the limit.

Example. Let $f_n : [0, 1] \to \mathbb{R}$ be

$$f_n(x) = \begin{cases} 1 & \text{if } n!x \in \mathbb{Z} \\ 0 & \text{otherwise} \end{cases},$$

which all have finitely many discontinuities, so are Riemann integrable. However, its limit is

$$f_n(x) \to f(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \notin \mathbb{Q} \end{cases},$$

which is not integrable. So the integrability of a function is not preserved by pointwise limits.

Definition 5.53. A sequence of functions $\{f_n\}$ with the same domain X and codomain Y is said to *converge uniformly* to a given function $f : X \to Y$, written as

$$\lim_{n\to\infty} f_n = f \text{ uniformly}\,,$$

if and only if for any $\epsilon > 0$, there exists a $N \in \mathbb{N}$ such that for all $n \geq N$ and for all $x \in X$,

$$|f_n(x) - f(x)| < \epsilon\,.$$

Here is a useful test on whether the limit of a sequence of functions is uniform or not.

Theorem 5.54 (Weierstrass M-test). Suppose that f_n is a sequence of real or complex-valued functions defined on a set X, and that there is a sequence of non-negative numbers M_n satisfying the conditions

$$|f_n(x)| \leq M_n$$

for all $n \geq 1$ and all $x \in X$, and

$$\sum_{n=1}^{\infty} M_n$$

converges. Then the series

$$\sum_{n=1}^{\infty} f_n(x)$$

converges absolutely and uniformly on X.

We now move on to show that uniform convergence tends to preserve properties of functions.

Theorem 5.55. Let $f_n, f : X \to \mathbb{R}$, where $X \subseteq \mathbb{R}$. Suppose $f_n \to f$ uniformly and f_n are continuous at x for all n, then f is also continuous at x. In particular, if f_n are continuous everywhere, then f is continuous everywhere.

Proof. Let $\epsilon > 0$. Choose N such that for all $n > N$, we have

$$\sup_{y \in X} |f_n(y) - f(y)| < \epsilon \,.$$

Since f_N is continuous at x, there is some δ such that

$$|x - y| < \delta \implies |f_N(x) - f_N(y)| < \epsilon \,.$$

Then for each y such that $|x - y| < \delta$, we have

$$|f(x) - f(y)| \leq |f(x) - f_N(x)| + |f_N(x) - f_N(y)| + |f_N(y) - f(y)| < 3\epsilon \,.$$

□

Remark. The uniform limit of continuous functions is continuous.

We will state but not prove that the uniform limit of a sequence of integrable functions is also integrable.

Lemma 5.56. If a sequence of functions $f_n : [a, b] \to \mathbb{R}$ are integrable in $[a, b]$, then their uniform limit f is also integrable in $[a, b]$.

Theorem 5.57. Let $f_n, f : [a, b] \to \mathbb{R}$ be Riemann integrable, with $f_n \to f$ uniformly. Then

$$\int_a^b f_n(x)\,\mathrm{d}x \to \int_a^b f(x)\,\mathrm{d}x \,.$$

Proof.

$$\begin{aligned}
\left| \int_a^b f_n(x)\,\mathrm{d}x - \int_a^b f(x)\,\mathrm{d}x \right| &= \left| \int_a^b f_n(x) - f(x)\,\mathrm{d}x \right| \\
&\leq \int_a^b |f_n(x) - f(x)|\,\mathrm{d}x \\
&\leq \sup_{x \in [a,b]} |f_n(x) - f(x)|(b - a) \\
&\to 0 \text{ as } n \to \infty \,.
\end{aligned}$$

□

Remark. We are allowed to exchange the order of a limit and an integral if the function converges uniformly. Similarly, we are allowed to exchange the order of a sum and an integral if the sum converges uniformly.

Corollary. If $f_n : [a, b] \to \mathbb{R}$ is a sequence of integrable functions whose partial sum converges uniformly to some function f, then f is integrable and

$$\int_a^b f(x)\,\mathrm{d}x = \sum_{n=1}^{\infty} \int_a^b f_n(x)\,\mathrm{d}x \,.$$

However, the relationship between uniform convergence and differentiability is more subtle. The uniform limit of differentiable functions need not be differentiable. Even if it were, the limit of the derivative is not necessarily the same as the derivative of the limit.

Example. Let $f_n, f : [-1,1] \to \mathbb{R}$ be defined by

$$f_n(x) = |x|^{1+\frac{1}{n}}\,,\ f(x) = |x|\,.$$

Then $f_n \to f$ uniformly. Each f is differentiable. At $x = 0$,

$$f_n'(0) = \lim_{x\to 0} \frac{f_n(x) - f_n(0)}{x} = \lim_{x\to 0} \operatorname{sgn}(x)|x|^{\frac{1}{n}} = 0\,.$$

However, the limit f is not differentiable at $x = 0$.

We need a condition even stronger than uniform convergence.

Theorem 5.58. Let $f_n : [a,b] \to \mathbb{R}$ be a sequence of functions differentiable on $[a,b]$. If

(i) for some $c \in [a,b]$, $f_n(c)$ converges;

(ii) the sequence of derivatives f_n' converges uniformly on $[a,b]$,

then f_n converges uniformly on $[a,b]$. If $f = \lim_{n\to\infty} f_n$, then f is differentiable with derivative $f'(x) = \lim_{n\to\infty} f_n'(x)$.

6 Complex Analysis

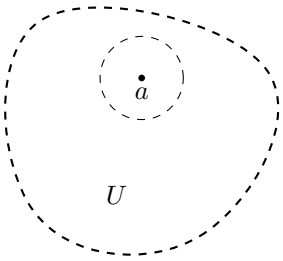
6.1 Complex Differentiation

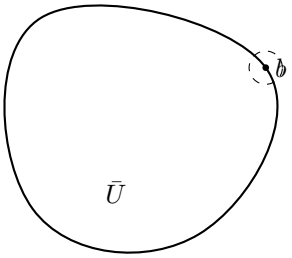
Let us first make a few definitions.

Definition 6.1. An *open ball* of radius $r > 0$ centred at $a \in \mathbb{C}$ is

$$D(a, r) \coloneqq \{z \in \mathbb{C} \mid |z - a| < r\}.$$

Definition 6.2. A subset $U \subset \mathbb{C}$ is *open* if for every $a \in U$, there exists $\epsilon > 0$ such that $D(a, \epsilon) \subset U$.



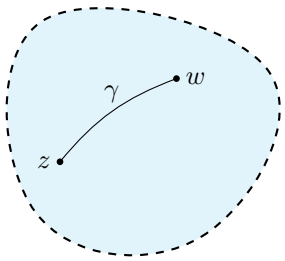


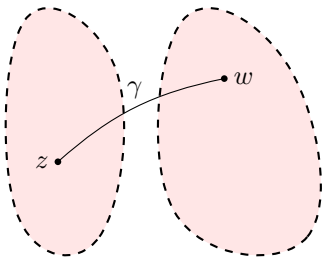
For example, the subset U on the left, excluding its boundary, is open. For any $a \in U$, we can find small enough $r > 0$ such that $D(a, r)$ is completely contained within U. While the subset $\bar{U}$ on the right, which includes its boundary, is not open because if we take b to be on the boundary, $b \in \partial\bar{U}$, then $D(b, \epsilon) \not\subset \bar{U}$ for any $\epsilon > 0$, no matter how small ϵ is. By the same logic, any $V \subset \mathbb{C}$ that contains its boundary (even a section of it) is not open.

Definition 6.3. A *curve* is a continuous map from a closed interval to the complex plane $\gamma : [a, b] \to \mathbb{C}$. A curve is *continuously differentiable* (C^1) if γ' exists and is continuous on $[a, b]$ (at endpoints a, b this means one-sided derivative).

Definition 6.4. An open set $U \subset \mathbb{C}$ is *path-connected* if for every $z, w \in U$ there exists a curve $\gamma : [0, 1] \to U$ with endpoints z, w.

Definition 6.5. A *domain* is a non-empty path-connected open subset of $\mathbb{C}$.





As an example, the region shaded in blue on the left is path-connected, and therefore is a domain because for any two points in the region, we can connect them with a curve. While for the two-pieces subset on the right, if we take one point from each piece, then they cannot be connected by a curve completely within the subset. It is therefore not path connected and not a domain.

Remark. We make so many definitions above just to make sure that whenever we say 'domain' in this chapter, it is a 'single piece' of region without a boundary so that any point has a neighbourhood.

Definition 6.6. The *complex derivative* of the function $f : U \to \mathbb{C}$ at a point $z = z_0 \in U$ is

$$f'(z_0) \coloneqq \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

if such a limit exists, and the function $f(z)$ is said to be *complex differentiable* at $z = z_0$.

Remark. Complex differentiation satisfies the same formal rules (derivatives of sum, product, quotient, chain rule, and inverse functions) as the differentiation of functions of a real variable because they are defined by the same limit.

6.1.1 The Cauchy–Riemann Equations

Being complex differentiable is actually a very strict condition, as you will see later. It means that if we move by a small amount h in all possible directions in the complex plane, we must have

$$f(z_0 + h) = f(z_0) + hf'(z_0) + o(|h|)\,.$$

This is stricter than viewing a complex function as a function defined on $\mathbb{R}^2$ and ask it to be differentiable along x and y direction respectively. It must be differentiable, with the same derivative f' along all possible directions.

This extra requirement can be captured by the Cauchy–Riemann equations.

Theorem 6.7 (The Cauchy–Riemann equations). $f : U \to \mathbb{C}$ is differentiable at $w = c + id \in U$ if and only if, writing $f(x + iy) = u(x, y) + iv(x, y)$, u and v are real differentiable at (c, d) and

$$\begin{cases} \dfrac{\partial u}{\partial x} = \dfrac{\partial v}{\partial y} \\ \dfrac{\partial v}{\partial x} = -\dfrac{\partial u}{\partial y}\,. \end{cases}$$

Proof. f is differentiable at $w = c + id$ with $f'(w) = p + iq$

$$\begin{aligned} &\iff \lim_{z \to w} \frac{f(z) - f(w) - (z - w)(p + iq)}{z - w} = 0 \\ &\iff \lim_{z \to w} \frac{f(z) - f(w) - (z - w)(p + iq)}{|z - w|} = 0\,. \end{aligned}$$

Writing $f = u + iv$ and evaluating the real and imaginary parts, this holds

$$\iff \begin{cases} \displaystyle\lim_{(x,y)\to(c,d)} \frac{u(x, y) - u(c, d) - [p(x - c) - q(y - d)]}{\sqrt{(x - c)^2 + (y - d)^2}} = 0 \\ \displaystyle\lim_{(x,y)\to(c,d)} \frac{v(x, y) - v(c, d) - [q(x - c) + p(y - d)]}{\sqrt{(x - c)^2 + (y - d)^2}} = 0\,. \end{cases}$$

Therefore, f is differentiable at w with derivative $f'(w) = p + iq$ if and only if u, v are differentiable at (c, d) with $\nabla u(c, d) = (p, -q)$ and $\nabla v(c, d) = (q, p)$. □

6.1.2 Holomorphic Functions

Definition 6.8. A function $f : U \to \mathbb{C}$ is *holomorphic* at $z_0 \in U$ if there exists $\epsilon > 0$ such that f is differentiable for all $z \in D(z_0, \epsilon)$. f is *holomorphic on* U if it is differentiable at all $z_0 \in U$. The function is *entire* if it is holomorphic in the whole complex plane.

Remark. To be holomorphic at some point, the function must be differentiable on a small neighbourhood of that point.

The existence of a complex derivative in a neighbourhood is a very strong condition: it implies that a holomorphic function is infinitely differentiable.

Proposition 6.9. Let $f = u + iv : U \to \mathbb{C}$. Suppose that the functions u, v have continuous partial derivatives everywhere on U and that they satisfy the Cauchy–Riemann equations, then f is holomorphic on U.

Example. Entire functions:

(i) $f(z) = c$, where $c \in \mathbb{C}$.

(ii) $f(z) = z$.

(iii) $f(z) = \exp(z)$.

(iv) $f(z) = z^n$, where $n \in \mathbb{N}_0$.

Property: Sums, products and compositions of holomorphic functions are also holomorphic.

Example. Non-holomorphic functions:

(i) $f(z) = \operatorname{Re} z$.

(ii) $f(z) = z^*$.

(iii) $f(z) = |z|$.

(iv) $f(z) = |z|^2$.

Remark. In the last case, the Cauchy–Riemann equations are satisfied only at $x = y = 0$ and we can say that $f'(0) = 0$. However, $f(z)$ is not holomorphic even at $z = 0$ because it is not differentiable throughout any neighbourhood $|z| < \epsilon$ of 0.

Definition 6.10. Many complex functions are holomorphic everywhere in the complex plane except at some isolated points, which are called the *singular points* or *singularities* of the function.

Example. $f(z) = z^c$, where $c \in \mathbb{C}$, is holomorphic except at $z = 0$. (Strictly speaking for non-integer c we need a branch choice, so z^c is holomorphic on any simply connected domain avoiding the branch cut and 0.)

6.1.3 Consequences of the Cauchy–Riemann Equations

If we know the real part of a holomorphic function in some region, we can find its imaginary part (or vice versa) up to an additive constant by integrating the Cauchy–Riemann equations.

Proposition 6.11. The real and imaginary parts of a holomorphic function satisfy Laplace's equation, i.e. u and v are *harmonic functions*.

Proof.

$$\begin{aligned}\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} &= \frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial y}\left(\frac{\partial u}{\partial y}\right) \\ &= \frac{\partial}{\partial x}\left(\frac{\partial v}{\partial y}\right) + \frac{\partial}{\partial y}\left(-\frac{\partial u}{\partial x}\right) \\ &= 0\,.\end{aligned}$$

Similarly, $\nabla^2 v = 0$. □

Proposition 6.12. The curves of constant u and those of constant v are orthogonal. They are said to be *conjugate harmonic functions*.

Proof. Using the Cauchy–Riemann equations, we have

$$\begin{aligned}\boldsymbol{\nabla} u \cdot \boldsymbol{\nabla} v &= \frac{\partial u}{\partial x}\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\frac{\partial v}{\partial y} \\ &= \frac{\partial v}{\partial y}\frac{\partial v}{\partial x} - \frac{\partial v}{\partial x}\frac{\partial v}{\partial y} \\ &= 0\,.\end{aligned}$$

□

Remark. Since u and v satisfy Laplace's equation, which is an elliptic PDE, it is guaranteed that u and v are smooth (and in fact analytic). This is a result of the elliptic regularity theorem.

This is a (quite convoluted) way to prove that a holomorphic complex function is always smooth.

Theorem 6.13 (Elliptic regularity theorem). For a differential equation

$$\mathcal{L}\psi(\mathbf{x}) = f(\mathbf{x})\,,$$

where $\mathcal{L}$ is an elliptic differential operator, if f is infinitely differentiable, then ψ is also infinitely differentiable.

6.2 Power Series of a Complex Variable

6.2.1 Convergence of Power Series

A complex power series about $z = z_0$ has the form

$$f(z) = \sum_{r=0}^{\infty} a_r (z - z_0)^r \text{ where } a_r \in \mathbb{C}\,.$$

Many of the tests of convergence for real series can be generalised for complex series. Note that if the sum of the absolute values of a complex series converges, then so does the series. Hence if $\sum |a_r(z - z_0)^r|$ converges, so does $\sum a_r(z - z_0)^r$.

Lemma 6.14. If the power series converges for $z = z_1$, then the series converges absolutely for all z such that $|z - z_0| < |z_1 - z_0|$.

Proof. Since $\sum a_r(z_1 - z_0)^r$ converges, then from the necessary condition for convergence,

$$\lim_{r\to\infty} a_r(z_1 - z_0)^r = 0\,.$$

Hence for a given ϵ there exists $N = N(\epsilon)$ such that if $r > N$ then

$$|a_r(z_1 - z_0)| < \epsilon\,.$$

Thus for $r > N$,

$$\begin{aligned}|a_r(z - z_0)^r| &= |a_r(z_1 - z_0)^r| \left|\frac{z - z_0}{z_1 - z_0}\right|^r \\ &< \epsilon\varrho^r\,, \quad \text{where } \varrho = \left|\frac{z - z_0}{z_1 - z_0}\right|.\end{aligned}$$

Hence, by comparison with a geometric series, $\sum a_r(z - z_0)^r$ converges for $\varrho < 1$, i.e. $|z - z_0| < |z_1 - z_0|$. □

Corollary. If the power series diverges for $z = z_1$ then it diverges for all $|z - z_0| > |z_1 - z_0|$.

Proof. This can be proven by contradiction. If the series converges for the sum z_2 such that $|z_2 - z_0| > |z_1 - z_0|$, then the series would converge for $z = z_1$; this is a contradiction. □

Definition 6.15. Let c_n be a sequence of complex numbers. There must exist a unique real number $R \in [0, \infty]$ such that the power series $\sum c_n(z - z_0)^n$

- converges absolutely if $0 < |z - z_0| < R$,
- diverges if $|z - z_0| > R$.

This R is called the *radius of convergence*.

Proof. Trivial by Lemma 6.14 and its corollary. □

Remark. On the circle of convergence, the series may either converge or diverge.

Corollary. Let the complex power series $\sum c_n(z - z_0)^n$ have a radius of convergence $R > 0$. Let $0 < r < R$, then the power series converges uniformly on $D(z_0, r)$.

Proof. We know that $\sum |c_n| r^n$ is convergent. If $|z - z_0| \leq r$, then

$$|c_n(z - z_0)^n| \leq |c_n| r^n \,.$$

So the result follows from the Weierstrass M-test by taking $M_n = |c_n| r^n$. □

Remark. This can be seen as the uniform convergence of a geometric progression.

6.2.2 Determination of the Radius of Convergence

Without loss of generality take $z_0 = 0$, so that the power series becomes

$$f(z) = \sum_{r=0}^{\infty} u_r = \sum_{r=0}^{\infty} a_r z^r \,.$$

Proposition 6.16. If the limit exists, then

$$\lim_{r\to\infty} \left| \frac{a_{r+1}}{a_r} \right| = \frac{1}{R} \,.$$

Proof. For the series to converge, we must have

$$\lim_{r\to\infty} \left| \frac{u_{r+1}}{u_r} \right| = \lim_{r\to\infty} \left| \frac{a_{r+1}}{a_r} \right| |z| < 1 \,.$$

Therefore by the D'Alembert ratio test, the power series

$$\text{converges if} \quad \frac{1}{|z|} > \lim_{r\to\infty} \left| \frac{u_{r+1}}{u_r} \right| \quad \text{and diverges if} \quad \frac{1}{|z|} < \lim_{r\to\infty} \left| \frac{u_{r+1}}{u_r} \right| .$$

□

Proposition 6.17. If the limit exists, then

$$\lim_{r\to\infty} |a_r|^{\frac{1}{r}} = \frac{1}{R} \,.$$

Proof. For the series to converge, we must have

$$\lim_{r\to\infty} |u_r|^{\frac{1}{r}} = \lim_{r\to\infty} |a_r|^{\frac{1}{r}} |z| < 1$$

by Cauchy's test. □

6.2.3 Holomorphicity of Analytic Functions (Non-examinable)

Theorem 6.18. Let $f(z) = \sum c_n(z - z_0)^n$ be a complex power series with radius of convergence $R > 0$, then

(i) f is holomorphic on $D(z_0, R)$;

(ii) its derivative is given by the series

$$\sum_{n=0}^{\infty} nc_n(z - z_0)^{n-1},$$

which also has a radius of convergence R;

(iii) f has derivatives of all orders on $D(z_0, R)$ and $f^{(n)}(z_0) = n!c_n$.

Proof. We will assume wlog that $z_0 = 0$. Consider the function

$$\sum_{n=1}^{\infty} nc_n z^{n-1}.$$

Since $|nc_n| \geq |c_n|$, this series has radius of convergence $R' \leq R$. If $0 < R_1 < R$, then for $|z| < R_1$, we have

$$\left|nc_n z^{n-1}\right| < nc_n R_1^{n-1} \frac{|z|^{n-1}}{R_1^{n-1}},$$

so since $n\frac{|z|^{n-1}}{R_1^{n-1}} \to 0$ as $n \to \infty$, for suitably large n, $|c_n|R_1^{n-1}$ provides an upper bound for $\left|nc_n z^{n-1}\right|$. By the Weierstrass M-test, the series converges absolutely and uniformly on $0 < |z - z_0| < R_1$, so the radius of convergence of this series is R.

Consider

$$\begin{aligned}\frac{f(z) - f(w)}{z - w} &= \sum_{n=0}^{\infty} c_n \frac{z^n - w^n}{z - w} \\ &= \lim_{N\to\infty} \sum_{n=0}^{N} c_n \left[\sum_{j=0}^{n-1} z_j w^{n-1-j}\right]. \qquad (*)\end{aligned}$$

For $|z|, |w| < r < R$, we have

$$\left|c_n \left[\sum_{j=0}^{n-1} z_j w^{n-1-j}\right]\right| < |c_n| n r^{n-1},$$

so $(*)$ converges uniformly on $|z|, |w| < r$, so the series has a continuous limit. We call it $g(z, w)$. When $z = w$,

$$g(z, z) = \sum_{n=0}^{N} c_n n z^{n-1}.$$

Therefore, f is differentiable with this derivative. This proves (i), (ii). (iii) is induction. □

6.3 Contour Integration

Definition 6.19. A path (curve) $\gamma : [a, b] \to \mathbb{C}$ is *closed* if $\gamma(a) = \gamma(b)$, and is *simple* if γ is injective, possibly except at endpoints.

This means that a simple path never crosses itself.

Definition 6.20. A *contour* is a simple piecewise differentiable path.

Definition 6.21. Let $f : U \to \mathbb{C}$ be a continuous complex function and $\gamma : [a, b] \to U$ be a contour. The *contour integral* of f along γ is

$$\int_\gamma f(z)\,\mathrm{d}z = \int_a^b f(\gamma(t))\gamma'(t)\,\mathrm{d}t\ .$$

Example. Consider the integral

$$\int_\gamma \frac{1}{z}\,\mathrm{d}z$$

from $z = -1$ to $z = 1$ along paths around half the unit circle (i). clockwise; (ii). anticlockwise. Making the substitution $z = e^{i\theta}$, $\mathrm{d}z = ie^{i\theta}$, we have

$$I_1 = \int_{\gamma_1} \frac{1}{z}\,\mathrm{d}z = \int_\pi^0 \frac{ie^{i\theta}}{e^{i\theta}}\,\mathrm{d}\theta = \int_\pi^0 i\,\mathrm{d}\theta = -i\pi\,,$$

$$I_2 = \int_{\gamma_2} \frac{1}{z}\,\mathrm{d}z = \int_\pi^{2\pi} i\,\mathrm{d}\theta = i\pi\,.$$

Remark. The result of a contour integration may depend on the contour.

Proposition 6.22. Basic properties of contour integration:

(i) *Linearity.*
$$\int_\gamma c_1 f_1(z) + c_2 f_2(z)\,\mathrm{d}z = c_1 \int_\gamma f_1(z)\,\mathrm{d}z + c_2 \int_\gamma f_2(z)\,\mathrm{d}z\ .$$

(ii) *Additivity.* If γ_1 is a contour from $z = a$ to $z = b$, γ_2 is a contour from $z = b$ to $z = c$, and γ is γ_1 followed by γ_2, then
$$\int_\gamma f(z)\,\mathrm{d}z = \int_{\gamma_1} f(z)\,\mathrm{d}z + \int_{\gamma_2} f(z)\,\mathrm{d}z\ .$$

(iii) *Inverse path.* If γ_+ is a contour from α to β, and γ_- is the contour in reverse, then
$$\int_{\gamma_+} f(z)\,\mathrm{d}z = -\int_{\gamma_-} f(z)\,\mathrm{d}z\ .$$

(iv) *Reparameterisation.* If $\gamma : [a, b] \to U$ is a contour, $\phi : [a', b'] \to [a, b] \in C^1$ with $\phi(a') = a$ and $\phi(b') = b$ so that $\delta = \gamma \circ \phi : [a', b'] \to U$ is a different parameterisation of the same curve, then
$$\int_\gamma f(z)\,\mathrm{d}z = \int_\delta f(z)\,\mathrm{d}z\ .$$

Lemma 6.23 (ML estimation lemma). If a contour γ has length L, then

$$\left|\int_\gamma f(z)\,\mathrm{d}z\right| \le L \sup_{z\in\gamma} |f(z)|\,.$$

Theorem 6.24 ('Fundamental theorem of calculus'). If $F : U \to \mathbb{C}$ is holomorphic with a continuous derivative and $\gamma : [a, b] \to U$ is a curve, then

$$\int_\gamma F'(z)\,\mathrm{d}z = F(\gamma(b)) - F(\gamma(a))\,.$$

Proof.

$$\int_\gamma F'(z)\,\mathrm{d}z = \int_a^b F'(\gamma(t))\gamma'(t)\,\mathrm{d}t = \int_a^b \frac{\mathrm{d}}{\mathrm{d}t}(F(\gamma(t))) = [F(\gamma(t))]_a^b\,.$$

□

6.4 Cauchy–Goursat Theorem

Definition 6.25. A subset $U \subseteq \mathbb{C}$ is *simply connected* if it is path-connected and any loop $\gamma : S^1 \to U$ path can be continuously contracted to a point: for any γ, there exists an extended continuous map $\hat{\gamma} : D^2 \to U$ such that $\hat{\gamma}|_{S_1} = \gamma$. Here, S^1 and D^2 denote the unit circle and closed unit disk respectively.

Remark. A simply connected domain has no 'hole' in it.

Theorem 6.26 (Cauchy–Goursat theorem). If a function f is holomorphic and has continuous partial derivatives in a simply connected domain U, then for any closed contour γ in U,

$$\oint_\gamma f(z)\,\mathrm{d}z = 0\,.$$

Proof. Green's theorem states that, for a vector field $\mathbf{p} = (p, q)$,

$$\oint_\gamma (p\,\mathrm{d}x + q\,\mathrm{d}y) = \int_\Omega \left(\frac{\partial q}{\partial x} - \frac{\partial p}{\partial y}\right) \mathrm{d}x\,\mathrm{d}y\ ,$$

where Ω is the simply connected region bounded by a simple closed curve γ. Hence, by expanding into real and imaginary parts and using the Cauchy–Riemann equations,

$$\begin{aligned}\oint_\gamma f(z)\,\mathrm{d}z &= \oint_\gamma (u + iv)(\mathrm{d}x + i\,\mathrm{d}y)\\ &= \oint_\gamma (u\,\mathrm{d}x - v\,\mathrm{d}y) + i\oint_\gamma (v\,\mathrm{d}x + u\,\mathrm{d}y)\\ &= \int_\Omega \left(-\frac{\partial u}{\partial y} - \frac{\partial v}{\partial x}\right)\mathrm{d}x\,\mathrm{d}y + i\int_\Omega \left(-\frac{\partial v}{\partial y} + \frac{\partial u}{\partial x}\right)\mathrm{d}x\,\mathrm{d}y = 0\,.\end{aligned}$$

□

Remark. This is a weaker version of this theorem proven by Cauchy. It requires f to be holomorphic and have continuous partial derivatives. A stronger version of this theorem, proven by Goursat, removes the need for partial-derivative continuity.

6.4.1 Deforming Contours

Proposition 6.27. Let γ_1 and γ_2 be two different contours both from α to β. If $f(z)$ is a function that is holomorphic on both contours and inside the region bounded by the contours, then

$$\int_{\gamma_1} f(z)\,\mathrm{d}z = \int_{\gamma_2} f(z)\,\mathrm{d}z\ .$$

Proof. Consider the closed curve $\gamma = \gamma_1 - \gamma_2$. It follows from Cauchy's theorem that

$$\oint_\gamma f(z)\,\mathrm{d}z = \int_{\gamma_1} f(z)\,\mathrm{d}z - \int_{\gamma_2} f(z)\,\mathrm{d}z = 0\,.$$

The proposition hence follows. □

Corollary. We can deform a contour without changing the value of the integral as long as we do not move the contour across a singularity.

Corollary. We can deform a closed contour if we are not passing it through any singularity.

Remark. This concept of continuously deforming a contour is formally known as a *homotopy* in algebraic topology.

This is possible even if the contour have (the same) singular points enclosed before and after deformation. Consider two closed contours γ_1 and γ_2 shown below. Let $f(z)$ be holomorphic within the region bounded between γ_1 and γ_2, but has singularities enclosed by γ_2. Consider cutting off a small piece of each contour and joining them together by two bridges γ_ϵ and γ'_ϵ to form a closed contour Γ.

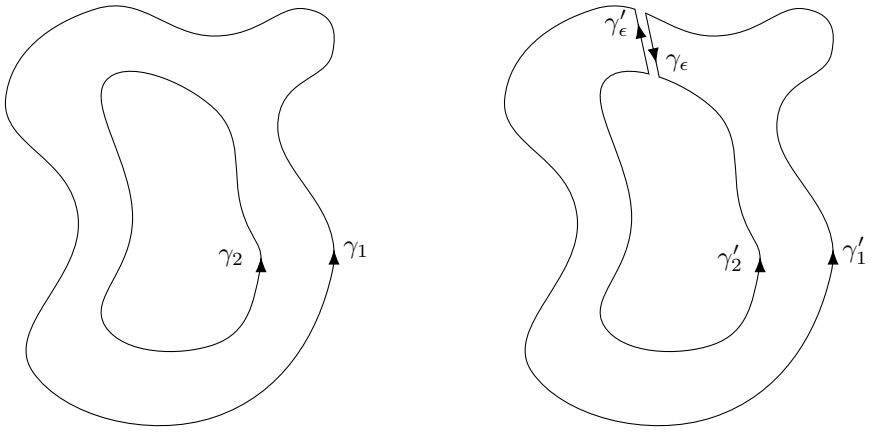
We have $\Gamma = \gamma'_1 - \gamma'_2 + \gamma_\epsilon + \gamma'_\epsilon$. Since $f(z)$ is holomorphic in the region bounded by Γ by assumption, we have

$$\int_\Gamma f(z)\,\mathrm{d}z = 0\,.$$

If we take the limit such that the width of the cut $\to 0$, we have $\gamma'_1 \to \gamma$, $\gamma'_2 \to \gamma_2$ and $\gamma'_\epsilon \to \gamma_\epsilon$. Therefore,

$$\int_\Gamma f(z)\,\mathrm{d}z \to \int_{\gamma_1} f(z)\,\mathrm{d}z - \int_{\gamma_2} f(z)\,\mathrm{d}z = 0\,,$$

so the integral along two closed contours enclosing the same singularities is the same.



Remark. This is connected with Theorem 6.24. If we can find a function $F(z)$ that is holomorphic in a simply connected domain U and $F'(z) = f(z)$, then

$$\int_\gamma f(z)\,\mathrm{d}z = F(\gamma(b)) - F(\gamma(a))$$

for all $\gamma : [a, b] \to U$. The integral is invariant.

6.5 Cauchy's Integral Formula

Theorem 6.28 (Cauchy's integral formula). If $f(z)$ is holomorphic on a domain $U \subseteq \mathbb{C}$, $z_0 \in U$, and $\gamma \subset U$ is an arbitrary simple closed contour that encircles z_0 counter-clockwise. We have Cauchy's integral formula

$$f(z_0) = \frac{1}{2\pi i} \oint_\gamma \frac{f(z)}{z - z_0}\,\mathrm{d}z\ .$$

Proof. $\frac{f(z)}{z-z_0}$ is holomorphic everywhere except at $z = z_0$. Therefore, we can deform γ to an arbitrarily small contour, say a circle of radius ϵ around z_0, γ_ϵ contained completely within γ. By substituting $z = z_0 + \epsilon e^{i\theta}$, we have

$$\begin{aligned}\oint_\gamma \frac{f(z)}{z - z_0}\,\mathrm{d}z &= \oint_{\gamma_\epsilon} \frac{f(z)}{z - z_0}\,\mathrm{d}z \\ &= \int_0^{2\pi} \frac{f(z_0 + \epsilon e^{i\theta})}{\epsilon e^{i\theta}} i\epsilon e^{i\theta}\,\mathrm{d}\theta\ .\end{aligned}$$

Take the limit $\epsilon \to 0$, we have

$$\oint_\gamma \frac{f(z)}{z - z_0}\,\mathrm{d}z = i \lim_{\epsilon \to 0} \int_0^{2\pi} f(z_0 + \epsilon e^{i\theta})\,\mathrm{d}\theta$$
$$= i \int_0^{2\pi} f(z_0)\,\mathrm{d}\theta = 2\pi i f(z_0)\,.$$

□

Remarks.

- If we know $f(z)$ on γ, then from the Cauchy's formula, we know $f(z)$ throughout the interior of γ.
- Since the real and imaginary parts of an analytic function, u and v, satisfy Laplace's equation, this statement is equivalent to the uniqueness of the solutions to Laplace's equation with Dirichlet boundary conditions (Theorem 10.9).

  In particular, if we specify u and v on γ, then there is a unique solution for u and v inside γ. This is equivalent to the integral solution of Poisson's equation that we will see later (Theorem 14.26).

Corollary (The mean-value property). If $f : D(z_0, r) \to \mathbb{C}$ is holomorphic, then

$$f(z_0) = \int_0^1 f(z_0 + re^{2\pi i t})\,\mathrm{d}t\ .$$

Proof. Change the variable $\theta = 2\pi t$ in the above proof and done. □

Remark. $f(w)$ equals the average value of f on any circle with centre w.

Theorem 6.29 (Liouville's theorem). Every bounded entire function must be constant.

Proof. Let $f : \mathbb{C} \to \mathbb{C}$ be an entire function such that $|f| < M$. For any $w \in \mathbb{C}$, let $R > |w|$,

$$|f(w) - f(0)| = \frac{1}{2\pi}\left|\int_{|z|=R} f(z)\left(\frac{1}{z - w} - \frac{1}{z}\right)\mathrm{d}z\right|$$
$$= \frac{1}{2\pi}\left|\int_{|z|=R} \frac{wf(z)}{z(z - w)}\,\mathrm{d}z\right|$$
$$\leq \frac{1}{2\pi} \times 2\pi R \times \frac{M|w|}{R(R - |w|)} \to 0 \text{ as } R \to \infty\,,$$

so $f(w) = f(0)$ by Cauchy's formula and ML estimation lemma. □

6.6 Taylor Expansion

Theorem 6.30 (Complex Taylor expansion). Let $f : D(z_0, r) \to \mathbb{C}$ be holomorphic, then f has a convergent power series representation on $D(z_0, r)$:

$$f(z) = \sum_{n=0}^{\infty} c_n (z - z_0)^n\,,$$

where

$$c_n = \frac{f^{(n)}(z_0)}{n!} = \frac{1}{2\pi i}\int_{|z - z_0| = \rho} \frac{f(z)}{(z - z_0)^{n+1}}\,\mathrm{d}z$$

for arbitrary $0 < \rho < r$.

Proof (Non-examinable). Let $|w - z_0| < \rho < r$. Using the geometric series, we have

$$\frac{1}{z-w} = \frac{1}{(z-z_0)\left(1-\frac{w-z_0}{z-z_0}\right)} = \sum_{n=0}^{\infty} \frac{(w-z_0)^n}{(z-z_0)^{n+1}}$$

which converges uniformly for $|z - z_0| = \rho$. By the Cauchy integral formula,

$$\begin{aligned} f(w) &= \frac{1}{2\pi i}\int_{|z-z_0|=\rho} \frac{f(z)}{z-w}\,\mathrm{d}z \\ &= \frac{1}{2\pi i}\int_{|z-z_0|=\rho} f(z)\sum_{n=0}^{\infty}\frac{(w-z_0)^n}{(z-z_0)^{n+1}}\,\mathrm{d}z \\ &= \sum_{n=0}^{\infty}\left(\frac{1}{2\pi i}\int_{|z-z_0|=\rho} f(z)\frac{1}{(z-z_0)^{n+1}}\,\mathrm{d}z\right)(w-z_0)^n\,, \end{aligned}$$

where the interchange of integration and summation is justified by the uniform convergence of the geometric progression. So f has a convergent power series representation on $D(z_0, \rho)$ for any $\rho < r$.

Differentiate the Cauchy's integral formula n times to get

$$f(z_0)^{(n)} = \frac{n!}{2\pi i}\int_{|z-z_0|=\rho}\frac{f(z)}{(z-z_0)^{n+1}}\,\mathrm{d}z\,,$$

so the coefficients of the power series are $f^{(n)}(z_0)/n!$. □

Corollary. If $f : U \to \mathbb{C}$ is holomorphic then its derivatives of all orders exist and are holomorphic.

Definition 6.31. A function $f : U \to \mathbb{C}$ is said to be *analytic* if every $z_0 \in U$, f can be represented by a convergent power series on some $D(z_0, r) \subset U$.

Theorem 6.32. For a complex function $f : U \to \mathbb{C}$, holomorphic $\iff$ analytic $\implies$ infinitely differentiable.

Note that 'smooth' in complex analysis means infinitely differentiable in $\mathbb{R}^2$ sense, not infinitely complex differentiable, so smoothness does not imply holomorphicity (e.g. $f(z) = z^*$).

Proof. Analytic $\implies$ holomorphic by Theorem 6.18. Holomorphic $\implies$ infinitely differentiable and analytic by the previous theorem. □

Remark. This is not the case in real analysis. A once differentiable real function is far from being infinitely differentiable. Even if a function is infinitely differentiable in some interval, it can be nowhere analytic. The Taylor series may have a radius of convergence of zero, or the function defined by its Taylor series fails to converge to f.

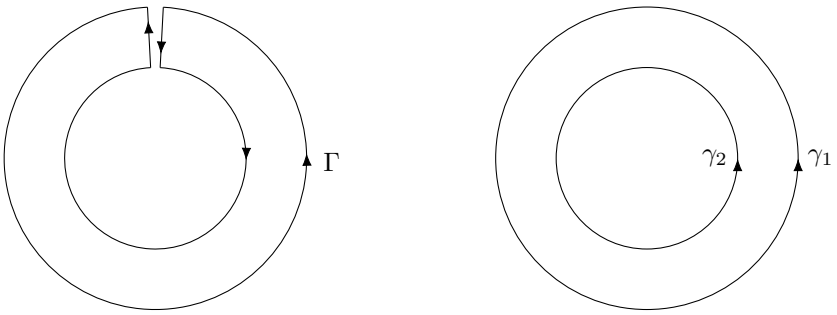
From now on, we shall use the terms 'holomorphic' and 'analytic' interchangeably.

6.7 Analytic Continuation (Non-examinable)

The fact that holomorphic functions are analytic has an interesting and important consequence: a holomorphic function on a domain U is determined by its restriction to a subdomain in U.

Definition 6.33. Let $U' \subset U$ be domains and $f : U' \to \mathbb{C}$ be analytic. A function $g(z) : U \to \mathbb{C}$ is called the *analytic continuation* of f if it is analytic and $f(z) = g(z)$ for all $z \in U'$

Theorem 6.34. The analytic continuation of a function is unique (if exists).



Proof. Let $g_1, g_2 : U \to \mathbb{C}$ be analytic continuations of $f : U' \to \mathbb{C}$ to U. Then $h = g_1 - g_2 : U \to \mathbb{C}$ is analytic and $h(z) = 0$ on U'. It suffices to show that h is identically zero on U. Define

$$U_0 = \{w \in U \mid h \text{ is identically 0 on some open disk } D(w, r)\}$$
$$U_1 = \{w \in U \mid h^{(n)}(w) \neq 0 \text{ for some } n > 0\}.$$

Then since h has a convergent power series expansion about each point $w \in U$, we see that $U = U_0 \cup U_1$ and $U_0 \cap U_1 = \varnothing$. Moreover, both U_0 and U_1 are open subsets of $\mathbb{C}$. So as U is connected, one of U_i is empty, and as $U_0 \supset U' \neq \varnothing$ we must have $U_1 = \varnothing$, so $U = U_0$ and $h = 0$ on all of U. □

6.8 Singularities and the Laurent Expansion

Theorem 6.35. Let f be holomorphic on an annulus $A = \{z \in \mathbb{C} \mid r < |z - z_0| < R\}$, where $0 \leq r < R \leq \infty$. Then

(i) There is a unique convergent *Laurent series expansion* on A

$$f(z) = \sum_{n=-\infty}^{\infty} c_n (z - z_0)^n .$$

(ii) For any $\rho \in (r, R)$, the coefficient c_n of the Laurent series is given by

$$c_n = \frac{1}{2\pi i} \int_{|z-z_0|=\rho} \frac{f(z)}{(z - z_0)^{n+1}} \, \mathrm{d}z \ .$$

(iii) If $r < \rho' \leq \rho < R$, then the Laurent series converges uniformly on $\{z \in \mathbb{C} \mid \rho' \leq |z - z_0| \leq \rho\}$, and hence on any compact subdomain of A.

Proof (Non-examinable). For a given $w \in A$, choose $r < \rho_2 < |w - z_0| < \rho_1 < R$ and consider the circular paths γ_1, γ_2, where γ_i is the circle $|z - z_0| = \rho_i$. Construct a path Γ by joining γ_1 and γ_2 together with two straight paths of width ϵ. By the Cauchy's integral formula and taking $\epsilon \to 0$, we have

$$\begin{aligned} f(w) &= \frac{1}{2\pi i} \int_{\Gamma} \frac{f(z)}{z - w} \, \mathrm{d}z \\ &= \frac{1}{2\pi i} \int_{\gamma_1} \frac{f(z)}{z - w} \, \mathrm{d}z - \frac{1}{2\pi i} \int_{\gamma_2} \frac{f(z)}{z - w} \, \mathrm{d}z \\ &=: f_1(w) + f_2(w) . \end{aligned}$$

For the first integral term $f_1(w)$, expand as in the proof of the Taylor series to get $f_1(w) = \sum_{n=0}^{\infty} c_n (w - z_0)^n$, where

$$c_n = \frac{1}{2\pi i} \int_{|z-z_0|=\rho_1} \frac{f(z)}{(z - z_0)^{n+1}} \, \mathrm{d}z \ , \text{for all } n \geq 0 .$$

For the second integral term $f_2(w)$, use the geometric series

$$\frac{-1}{z-w} = \frac{\frac{1}{w-z_0}}{1-\frac{z-z_0}{w-z_0}} = \sum_{m=1}^{\infty} \frac{(z-z_0)^{m-1}}{(w-z_0)^m}$$

which converges uniformly for $|z-z_0| = \rho_2$. This gives $f_2(w) = \sum_{m=1}^{\infty} d_m (w-z_0)^{-m}$, where

$$d_m = \frac{1}{2\pi i} \int_{|z-z_0|=\rho_2} \frac{f(z)}{(z-z_0)^{-m+1}} \,\mathrm{d}z \text{ for all } m \geq 1\,.$$

Combining these two results gives (i).

To show (ii) and (iii), suppose that we have any convergent series $\sum_{n=-\infty}^{\infty} c_n (z-z_0)^n$ on A, and let $r < \rho' \leq \rho < R$. Then the power series $\sum_{n=0}^{\infty} c_n (z-z_0)^n$ must have radius of convergence $\geq R$, so converges uniformly on $\{|z-z_0| \leq \rho\}$. Likewise, let $u = (z-z_0)^{-1}$, then the series $\sum_{n=1}^{\infty} c_{-n} u^n$ must have a radius of convergence $\geq \frac{1}{r}$, so converges uniformly on $\{|u| \leq \frac{1}{\rho'}\}$. Therefore, the series $\sum_{n=-\infty}^{\infty} c_n (z-z_0)^n$ converges uniformly on $\{\rho' \leq |z-z_0| \leq \rho\}$ and therefore in particular can be integrated term-by-term along any curve in this set; so

$$\int_{|z-z_0|=\rho} \frac{f(z)}{(z-z_0)^{m+1}} \,\mathrm{d}z = \sum_{n=-\infty}^{\infty} c_n \int_{|z-z_0|=\rho} (z-z_0)^{n-m-1} = 2\pi i c_m\,.$$

□

Definition 6.36. The *zeros* of a holomorphic function $f(z)$ are the points $z = z_0$ where $f(z_0) = 0$. A zero is of *order* k if the first non-zero term in the Taylor expansion of $f(z)$ about z_0 is $c_k(z-z_0)^k$.

Definition 6.37. A *singularity* of a function f is a point $z = z_0$ where f is not holomorphic. If f has a singularity at z_0, but f is holomorphic in a neighbourhood of z_0 except at z_0 itself, then z_0 is an *isolated singularity* of f. If there exists no such neighbourhood, z_0 is a non-isolated singularity.

Examples.

(i) $f = \operatorname{cosech} z = \frac{1}{\sinh z}$ has isolated singularities at $z = in\pi$, $n \in \mathbb{Z}$ because $\sinh x = 0$ at these points.

(ii) $f = \operatorname{cosech} \frac{1}{z}$ has isolated singularities at $z = \frac{1}{in\pi}$, $n \in \mathbb{Z} \setminus \{0\}$. f also has a non-isolated singularity at $z = 0$; for any disk $D(\epsilon, 0)$ we can find a large enough n such that the disk contains another singularity at $z = \frac{1}{in\pi}$.

We shall usually be concerned with isolated singularities, for which $f(z)$ is holomorphic on the punctured disk $D(z_0, R)^{\times} := D(z_0, R) \setminus \{z_0\}$.

Definition 6.38. There are three possible behaviours of f with isolated singularities:

(i) If the first non-zero term in the Laurent series of f has $n \geq 0$, then the Laurent series converges throughout the unpunctured disk $D(z_0, R)$. We say f has a *removable singularity* at $z = z_0$.

(ii) If there exists some finite $k > 0$ such that $c_{-k} \neq 0$ but $c_n = 0$ for all $n < -k$, then f has a *pole of order* k at $z = z_0$.

(iii) Otherwise, if the Laurent series centred at $z = z_0$ involves an infinite number of terms with $n < 0$, we say f has an *essential singularity* at $z = z_0$.

Examples.

(i) Removable singularity typically arises when f is given by some formula which is not well-defined at $z = z_0$; for example, take $z_0 = 0$ and $f(z) = (e^z - 1)/z$.

(ii) An example of essential singularity is $f(z) = e^{\frac{1}{z}}$ at $z = 0$.

$$e^{\frac{1}{z}} = \sum_{n=0}^{\infty} \frac{1}{n!}\left(\frac{1}{z}\right)^n = \sum_{n=-\infty}^{0} \frac{1}{(-n)!} z^n \,.$$

Remark. The behaviour of a function near an essential singularity is remarkably complicated.

Theorem 6.39 (Picard's theorem). In any neighbourhood of an essential singularity, the function takes all possible complex values (possibly with one exception) at infinitely many points.

Example. In the case of $e^{\frac{1}{z}}$, the exceptional value 0 is never obtained.

Proposition 6.40. f has a removable singularity at $z = z_0$ if and only if

$$\lim_{z \to z_0} (z - z_0) f(z) = 0 \,.$$

Proof.

($\Rightarrow$) Write

$$(z - z_0) f(z) = \sum_{n=0}^{\infty} c_n (z - z_0)^{n+1}$$

so it vanishes as $z \to z_0$.

($\Leftarrow$) Consider

$$g(z) = \begin{cases} (z - z_0)^2 f(z) & \text{if } z \neq z_0 \\ 0 & \text{if } z = z_0 \,. \end{cases}$$

We see that g is holomorphic and $g'(z_0) = 0$ so

$$g(z) = \sum_{n=2}^{\infty} c_n (z - z_0)^n$$

$$f(z) = \sum_{n=0}^{\infty} c_{n+2} (z - z_0)^n$$

and hence f has a removable singularity at $z = z_0$.

□

Proposition 6.41. f has a pole at $z = z_0$ if and only if $|f(z)| \to \infty$ as $z \to z_0$. Moreover, the following statements are equivalent:

(i) f has a pole of order k at $z = z_0$.

(ii) $f = (z - a)^{-k} g(z)$, where $g : D(z_0, R) \to \mathbb{C}$ is holomorphic and $g(z_0) \neq 0$.

(iii) $f(z) = \frac{1}{h(z)}$, where h is holomorphic at $z = z_0$ with a zero of order k.

Proof. First, prove (i) $\Leftrightarrow$ (ii). Given f with a pole, multiplying the Laurent series by $(z - z_0)^k$ gives a power series with a non-zero constant term, defining g, and the converse is clear. The Taylor series for g multiplied with $(z - z_0)^{-k}$ gives the Laurent series for f.

Next, (ii) $\Leftrightarrow$ (iii). This is because g is holomorphic and non-zero $z = z_0$ if and only if $1/g$ is holomorphic and non-zero at $z = z_0$.

Suppose f has a pole at $z = z_0$. Then by (ii) $|f| \to \infty$ as $z \to z_0$. Conversely if $|f| \to \infty$ as $z \to z_0$, then for some $r > 0$, f is non-zero for $0 < |z - z_0| < r$. Therefore $1/f$ is holomorphic for $0 < |z - z_0| < r$ and $1/f \to 0$ as $z \to z_0$. By the previous proposition, $1/f$ has a removable singularity at $z = z_0$. Thus there is a holomorphic h on $D(z_0, r)$ with $1/h = f$ for $0 < |z - z_0| < r$. As $1/f \to 0$ as $z \to z_0$, h has a zero at $z = z_0$. □

6.8.1 Meromorphic Functions

Definition 6.42. Let U be a domain and $S \subset U$ is a set of isolated points in U, then a function $f : U \setminus S \to \mathbb{C}$ with at worst poles (poles and removable singularities) at the points in S is said to be *meromorphic*.

Remark. A function f meromorphic on U can be written as $f = \frac{g}{h}$, where g and h are holomorphic on U.

6.8.2 Behaviour at Infinity

We can examine the behaviour of a function $f(z)$ as $z \to \infty$ by defining a new variable $\xi = \frac{1}{z}$ and a new function $g(\xi) = f(z)$. The $z = \infty$ maps to a single point $\xi = 0$, the point at infinity.

Examples.

(i) $f(z) = e^z = e^{\frac{1}{\xi}} = g(\xi)$ has an essential singularity at $z = \infty$.

(ii) $f(z) = z^2 = \frac{1}{\xi^2} = g(\xi)$ has a double pole at $z = \infty$.

(iii) $f(z) = e^{\frac{1}{z}} = e^{\xi} = g(\xi)$ is analytic at $z = \infty$.

Remark. All entire functions $f(z)$ have essential singularities at $z = \infty$ unless they are polynomials, and all polynomials have poles at $z = \infty$ unless they are constant.

7 Series Solutions of Ordinary Differential Equations

7.1 Linear Independence and the Wronskian

7.1.1 Linearly Independent Solutions

Consider homogeneous second-order linear ordinary differential equations of the form

$$y'' + p(x)y' + q(x)y = 0\,. \tag{†}$$

Recall that two solutions y_1 and y_2 are linearly independent if and only if

$$\alpha y_1(x) + \beta y_2(x) = 0 \implies \alpha = \beta = 0\,.$$

If $y_1(x)$ and $y_2(x)$ are linearly independent solutions, then the general solution of the ODE is given by

$$y(x) = \alpha y_1(x) + \beta y_2(x)\,,$$

where α and β are arbitrary constants.

7.1.2 The Wronskian

Recall that the Wronskian of two solutions $y_1(x)$ and $y_2(x)$ of a second-order ODE of the form (†) is the determinant of the Wronskian matrix

$$W[y_1, y_2] = \begin{vmatrix} y_1 & y_2 \\ y_1' & y_2' \end{vmatrix} = y_1 y_2' - y_2 y_1'\,.$$

If $W \neq 0$, the solutions y_1 and y_2 must be linearly independent.

Lemma 7.1. The Wronskian, W, of a homogeneous second-order linear ODE

$$y'' + p(x)y' + q(x)y = 0 \tag{†}$$

satisfies the first-order equation

$$W' + p(x)W = 0\,,$$

with solution

$$W(x) = \kappa \exp\left(-\int^x p(\xi)\,\mathrm{d}\xi\right),$$

where κ is a constant.

Proof.

$$\begin{aligned} W' &= y_1 y_2'' - y_1'' y_2 \\ &= -y_1(py_2' + qy_2) + (py_1' + qy_1)y_2 \\ &= -p(y_1 y_2' - y_1' y_2) \\ &= -pW\,. \end{aligned}$$

□

Remarks.

- Up to the multiplicative κ, the Wronskian W is the same for any two linearly independent solutions y_1 and y_2, and hence it is an intrinsic property of the ODE.
- If $W \neq 0$ for one value of x, then $W \neq 0$ for all x. Hence if y_1 and y_2 are linearly independent for one value of x, they are linearly independent for all values of x; it follows that linear independence needs to be checked at only one value of x.

7.1.3 A Second Solution via Wronskian

Suppose that we already have one solution, y_1, to the homogeneous equation. Then we can calculate a second linearly independent solution, y_2, using the Wronskian.

Lemma 7.2. For a given solution y_1 to the equation (†), a linearly independent solution is given by

$$\begin{aligned} y_2(x) &= y_1(x) \int^x \frac{W(\eta)}{y_1^2(\eta)}\,\mathrm{d}\eta \\ &= y_1(x) \int^x \frac{\kappa}{y_1^2(\eta)} \exp\left(-\int^\eta p(\xi)\,\mathrm{d}\xi\right)\mathrm{d}\eta\,. \end{aligned}$$

Proof. The definition of the Wronskian provides a first-order linear ODE for the unknown y_2:

$$y_1 y_2' - y_1' y_2 = W(x)\,.$$

To solve, divide by y_1^2 to obtain

$$\left(\frac{y_2}{y_1}\right)' = \frac{y_2'}{y_1} - \frac{y_2 y_1'}{y_1^2} = \frac{W}{y_1^2}\,,$$

and hence

$$y_2(x) = y_1(x) \int^x \frac{W(\eta)}{y_1^2(\eta)}\,\mathrm{d}\eta\,.$$

□

Remarks.

- The indefinite integral involves an arbitrary additive constant since any amount of y_1 can be added to y_2.
- W involves an arbitrary multiplicative constant, since y_2 can be multiplied by any constant.
- The same result can be obtained by writing $y_2(x) = y_1(x)u(x)$ and obtaining a first order linear ODE for u'. This method applies to higher-order linear ODEs and is reminiscent of the factorisation of polynomial equations.

7.2 Taylor Series Solutions

7.2.1 Ordinary and Singular Points

Now generalise the ODE to complex functions $y(z)$ of a complex variable z. The homogeneous linear second-order ODE in the standard form then becomes

$$y''(z) + p(z)y'(z) + q(z)y(z) = 0\,. \qquad (\dagger\dagger)$$

Definition 7.3. For ODE of the form (††), if $p(z)$ and $q(z)$ are both analytic at $z = z_0$, then $z = z_0$ is called an *ordinary point* of the ODE. A point at which p and/or q is singular is called a *singular point* of the ODE.

Definition 7.4. A singular point $z = z_0$ is *regular* if $(z - z_0)p(z)$ and $(z - z_0)^2 q(z)$ are both analytic at $z = z_0$.

Example. Legendre's equation.

Consider *Legendre's equation*

$$(1 - z^2)y'' - 2zy' + \ell(\ell + 1)y = 0\,, \qquad (*)$$

where ℓ is a constant. To identify the singular points and their nature, we obtain the standard form with

$$p(z) = -\frac{2z}{1-z^2}\,, \quad q(z) = \frac{\ell(\ell+1)}{1-z^2}\,.$$

Both $p(z)$ and $q(z)$ are analytic for all z except $z = \pm 1$, which are the singular points. However, they are both regular since

$$(z-1)p(z) = \frac{2z}{1+z}\,, \text{ and } (z-1)^2 q(z) = \ell(\ell+1)\left(\frac{1-z}{1+z}\right)$$

are both analytic at $z = 1$, and similarly for $z = -1$.

7.2.2 The Solution at Ordinary Points in terms of a Power Series

Claim 7.5. If $z = z_0$ is an ordinary point of the complex ODE (††), then we claim that the solution $y(z)$ is analytic at $z = z_0$, and consequently, the equation has two linearly independent solutions of the form

$$y = \sum_{n=0}^{\infty} a_n (z - z_0)^n \text{ when } |z - z_0| < R\,,$$

where R is the radius of convergence.

The coefficients can be determined by substituting the series into the equation and comparing powers of $(z - z_0)$. The radius of convergence turns out to be the distance to the nearest singular point of the equation in the complex plane.

For simplicity, we will assume henceforth wlog that $z_0 = 0$ (corresponding to a shift in the origin, e.g. define $z' = z - z_0$). Hence we seek solutions of the form

$$y = \sum_{n=0}^{\infty} a_n z^n\,,$$

for which

$$y' = \sum_{n=1}^{\infty} n a_n z^{n-1} = \sum_{m=0}^{\infty} (m+1) a_{m+1} z^m\,,$$

$$y'' = \sum_{n=2}^{\infty} n(n-1) a_n z^{n-2} = \sum_{r=0}^{\infty} (r+2)(r+1) a_{r+2} z^r\,.$$

At an ordinary point $p(z)$ and $q(z)$ are analytic so we can write

$$p(z) = \sum_{n=0}^{\infty} p_n z^n \text{ and } q(z) = \sum_{n=0}^{\infty} q_n z^n\,.$$

On substituting the above series into the ODE, we need a rule for multiplying double sums of the form

$$\sum_{n=0}^{\infty} A_n z^n \sum_{m=0}^{\infty} B_m z^m$$

to only include powers like z^r. Let $r = n + m$, and we then have

$$\sum_{n=0}^{\infty} A_n z^n \sum_{m=0}^{\infty} B_m z^{=} \sum_{r=0}^{\infty} \left(\sum_{m=0}^{r} A_{r-m} B_m \right) z^r\,.$$

Hence we have

$$p(z)y'(z) = \sum_{r=0}^{\infty}\left(\sum_{m=0}^{r} p_{r-m}(m+1)a_{m+1}\right)z^r\,,$$

$$q(z)y(z) = \sum_{r=0}^{\infty}\left(\sum_{m=0}^{r} q_{r-m}a_m\right)z^r\,.$$

Now substitute series into the ODE, and group powers of z^r, we have

$$\sum_{r=0}^{\infty}\left((r+2)(r+1)a_{r+2} + \sum_{m=0}^{r}\left((m+1)a_{m+1}p_{r-m} + a_m q_{r-m}\right)\right)z^r = 0\,.$$

Since the equation is true for all $|z| < R$, each coefficient of z^r $(r = 0, 1, \ldots)$ must be zero. Thus we deduce the recurrence relation

$$a_{r+2} = -\frac{1}{(r+2)(r+1)}\sum_{m=0}^{r}\left((m+1)a_{m+1}p_{r-m} + a_m q_{r-m}\right) \text{ for } r \geq 0\,.$$

This is a recurrence relation that determines a_{r+2} (for $r \geq 0$) in terms of preceding coefficients $a_0, a_1, \ldots, a_{r+1}$. This means that if a_0 and a_1 are known then so are all the a_r. The first two coefficients a_0 and a_1 play the role of the two integration constants in the general solution.

7.2.3 Example

Consider

$$y'' - \frac{2}{(1-z)^2}y = 0\,.$$

$z = 0$ is an ordinary point so try

$$y = \sum_{n=0}^{\infty} a_n z^n\,.$$

We note that

$$p = 0\,,\ q = -\frac{2}{(1-z)^2} = -2\sum_{m=0}^{\infty}(m+1)z^m\,,$$

and hence we have $p_m = 0$ and $q_m = -2(m+1)$. Substitution into the general result we obtain the recurrence relation

$$a_{r+2} = \frac{2}{(r+1)(r+2)}\sum_{n=0}^{r} a_n(r-n+1) \text{ for } r \geq 0\,.$$

However, without the reference to the standard result, we may obtain a simpler recurrence relation with a small amount of forethought. We can simplify the ODE to

$$(1-z)^2 y'' - 2y = 0\,.$$

Then the substitution of the series of derivatives gives

$$\sum_{n=2}^{\infty} n(n-1)a_n z^{n-2} - 2\sum_{n=1}^{\infty} n(n-1)a_n z^{n-1} + \sum_{n=0}^{\infty}(n^2-n-2)a_n z^n = 0\,.$$

After the substitutions $r = n-2$, $r = n-1$ and $r = n$ in the three terms respectively, we obtain

$$\sum_{r=0}^{\infty}(r+1)[(r+2)a_{r+2} - 2ra_{r+1} + (r-2)a_r]z^r = 0\,,$$

which leads to the recurrence relation

$$a_{r+2} = \frac{1}{r+2}(2ra_{r+1} - (r-2)a_r) \text{ for } r \geq 0\,.$$

For $r = 0$ the recurrence relation yields $a_2 = a_0$, while for $r = 1$ and $r = 2$ we obtain

$$a_3 = \frac{1}{3}(2a_2 + a_1) \text{ and } a_4 = a_3\,.$$

First we note that if $2a_2 + a_1 = 0$, then $a_3 = a_4 = 0$, and hence $a_r = 0$ for all $r \geq 3$. We thus have our first solution (with $a_0 = \alpha \neq 0$)

$$y_1 = \alpha(1-z)^2\,.$$

Next, we note that $a_r = a_0$ for all r is also a solution to the recurrence relation. In this case, we have

$$y_2 = \beta\sum_{n=0}^{\infty} z^n = \frac{\beta}{1-z}\,.$$

Linear independence. We can check that y_1 and y_2 are linearly independent by calculating the Wronskian:

$$W = \alpha(1-z)^2\frac{\beta}{(1-z)^2} + 2\alpha(1-z)\frac{\beta}{1-z} = 3\alpha\beta \neq 0\,.$$

Hence the general solution is given by

$$y(z) = \alpha(1-z)^2 + \frac{\beta}{1-z}\,,$$

for constant α and β.

Radius of convergence. The radius of convergence of y_2 is $R = 1$, which is consistent with the general solution being singular at $z = 1$, and the equation having a singular point at $z = 1$ since $q(z) = -2(1-z)^{-2}$.

7.2.4 Example: Legendre's Equation

Consider Legendre's equation

$$(1-z^2)y'' - 2zy' + \ell(\ell+1)y = 0\,, \qquad (*)$$

where $\ell \in \mathbb{R}$. The points $z = \pm 1$ are singular points but $z = 0$ is an ordinary point, so for smallish z we seek a power series solution

$$y = \sum_{n=0}^{\infty} a_n z^n\,.$$

On substituting this into the Legendre's equation, we have

$$\sum_{n=2}^{\infty} n(n-1)a_n z^{n-2} - \sum_{n=0}^{\infty} n(n-1)a_n z^n - 2\sum_{n=0}^{\infty} na_n z^n + \sum_{n=0}^{\infty} \ell(\ell+1)a_n z^n = 0\,.$$

From substituting $r = n-2$ in the first sum and $r = n$ in the next three sums, and from grouping powers of z^r, we obtain

$$\sum_{r=0}^{\infty} [(r+2)(r+1)a_{r+2} - (r(r+1) - \ell(\ell+1))a_r]z^r = 0\,.$$

The recurrence relationship is therefore

$$a_{r+2} = \frac{r(r+1) - \ell(\ell+1)}{(r+1)(r+2)} a_r \text{ for } r \in \mathbb{N}_0 \,.$$

a_0 and a_1 are arbitrary constants. For instance:

- if $a_0 = 1$ and $a_1 = 0$, then
$$y_1 = 1 - \frac{\ell(\ell+1)}{2} z^2 + O(z^4)$$
is an even solution;
- if $a_0 = 0$ and $a_1 = 1$, then
$$y_2 = z + \frac{2 - \ell(\ell+1)}{6} z^3 + O(z^5)$$
is an odd solution.

Linear independence. By checking the Wronskian, we can confirm that y_1 and y_2 are linearly independent.

Radius of convergence. The two solutions are effectively power series in z^2 rather than z. Hence to find the radius of convergence, we may re-express our series (e.g. $z^2 \to y$ and $a_{2n} \to b_n$), or use a slightly modified D'Alembert's ratio test. We observe that

$$\lim_{n\to\infty} \left| \frac{a_{n+2} z^{n+2}}{a_n z^n} \right| = \lim_{n\to\infty} \left| \frac{n(n+1) - \ell(\ell+1)}{(n+1)(n+2)} \right| |z|^2 = |z|^2 \,.$$

It then follows from an extension of D'Alembert's ratio test that the series converges for $|z| < 1$.

Remark. The radius of convergence is the distance to the nearest singularity of the ODE.

Legendre Polynomials

In a generic situation, the power series of the solution has an infinite number of terms. However, for $\ell \in \mathbb{N}_0$, it follows that

$$a_{\ell+2} = \frac{\ell(\ell+1) - \ell(\ell+1)}{(\ell+1)(\ell+2)} a_\ell = 0 \,,$$

and so the series terminates. For instance,

$$\begin{aligned} \ell = 0 &: y = a_0 \,, \\ \ell = 1 &: y = a_1 z \,, \\ \ell = 2 &: y = a_0(1 - 3z^2) \,. \end{aligned}$$

These functions are proportional to the *Legendre polynomials*, $P_\ell(z)$, which are conventionally normalized so that $P_\ell(1) = 1$. Thus, the first few Legendre polynomials are

$$\begin{aligned} P_0(z) &= 1 \,, \\ P_1(z) &= z \,, \\ P_2(z) &= \frac{1}{2}(3z^2 - 1) \,, \\ &\dots \end{aligned}$$

7.3 Regular Singular Points

Let $z = z_0$ be a regular singular point of the ODE

$$y''(z) + p(z)y'(z) + q(z)y(z) = 0$$

where wlog we can take $z_0 = 0$. If we write

$$p(z) = \frac{1}{z}s(z) \quad \text{and} \quad q(z) = \frac{1}{z^2}t(z)\,,$$

then the homogeneous equation becomes

$$z^2y'' + zs(z)y' + t(z)y = 0\,,$$

where, from the definition of a regular singular point, $s(z)$ and $t(z)$ are both analytic at $z = 0$. It follows that $s_0 \equiv s(0)$ and $t_0 \equiv t(0)$ are finite.

7.3.1 The Indicial Equation

Theorem 7.6 (Fuchs' theorem). A second-order differential equation of the form

$$y'' + p(z)y' + q(z)y = g(z)$$

has at least one solution expressible by a *Frobenius series* of the form

$$y = \sum_{n=0}^{\infty} a_n(z - z_0)^{n+\sigma}, \quad a_0 \neq 0 \text{ and } \sigma \in \mathbb{C}$$

when $p(z)$, $q(z)$ and $g(z)$ are analytic at $z = z_0$ or $z = z_0$ is a regular singular point.

If $z = 0$ is a regular singular point, Fuchs' theorem guarantees that there is at least one solution of the form

$$y = z^\sigma \sum_{n=0}^{\infty} a_n z^n, \quad a_0 \neq 0 \text{ and } \sigma \in \mathbb{C}\,.$$

Remarks.

- This is a Taylor series only if σ is a non-negative integer.
- There may be one or two solutions of this form.
- The condition $a_0 \neq 0$ is required to define σ uniquely.

Substitute the solution into the homogeneous equation (††), after the division of z^σ, we have

$$\sum_{n=0}^{\infty}((\sigma + n)(\sigma + n - 1) + (\sigma + n)s(z) + t(z))a_n z^n = 0\,.$$

We now evaluate this sum at $z = 0$. Since $z^n = 0$ except for $n = 0$, we have

$$(\sigma(\sigma - 1) + \sigma s_0 + t_0)a_0 = 0\,.$$

Since by definition $a_0 \neq 0$, we obtain the indicial equation for σ:

$$\sigma^2 + \sigma(s_0 - 1) + t_0 = 0\,.$$

The roots σ_1 and σ_2 are called the indices of the regular singular point.

7.3.2 Series Solutions

For each choice of σ from σ_1 and σ_2 we can find a recurrence relation for a_n by comparing powers of z.

- $\sigma_1 - \sigma_2 \notin \mathbb{Z}$. If $\sigma_1 - \sigma_2 \notin \mathbb{Z}$ we can find both linearly independent solutions in this way.
- $\sigma_1 - \sigma_2 \in \mathbb{Z}$. If $\sigma_1 = \sigma_2$ we note that we can find only one solution by the ansatz. The ansatz also fails (in general) to give both solutions when σ_1 and σ_2 differ by an integer (although there are some exceptions).

Frobenius' method is used to find the series of solutions about a regular singular point. This is demonstrated by the example below.

7.3.3 Bessel's Equation of Order ν

Definition 7.7. *Bessel's equation of order ν* is

$$y'' + \frac{1}{z}y' + \left(1 - \frac{\nu^2}{z^2}\right)y = 0\,, \tag{$**$}$$

where $\nu \geq 0$ wlog.

The origin $z = 0$ is a singular point with

$$s(z) = 1 \quad \text{and} \quad t(z) = z^2 - \nu^2\,.$$

A Frobenius series solution solves the Bessel's equation if

$$\sum_{n=0}^{\infty}((\sigma+n)(\sigma+n-1) + (\sigma+n) - \nu^2)a_n z^n + \sum_{n=0}^{\infty} a_n z^{n+2} = 0\,.$$

By a transformation $n \to n-2$ in the second sum, this simplifies to

$$\sum_{n=0}^{\infty}((\sigma+n)^2 - \nu^2)a_n z^n + \sum_{n=2}^{\infty} a_{n-2} z^n = 0\,.$$

Comparing the powers of z gives

$$\begin{aligned} &n = 0 : \sigma^2 - \nu^2 = 0 \\ &n = 1 : ((\sigma+1)^2 - \nu^2)a_1 = 0 \\ &n \geq 2 : ((\sigma+n)^2 - \nu^2)a_n + a_{n-2} = 0\,. \end{aligned}$$

The $n = 0$ case is the indicial equation, and it implies that

$$\sigma = \pm\nu\,.$$

Substituting this into the $n = 1$ and $n \geq 2$ equations yields

$$\begin{aligned} (1 + 2\nu)a_1 &= 0 \\ n(n \pm 2\nu)a_n &= -a_{n-2} \quad \text{for} \quad n \geq 2\,. \end{aligned}$$

This gives us a recurrence relation to solve for a_n from a_{n-2}.

Radius of convergence. The radius of convergence of the solution is infinity since

$$\lim_{n\to\infty}\left|\frac{a_n}{a_{n-2}}\right| = \lim_{n\to\infty}\left|\frac{1}{n(n\pm 2\nu)}\right| = 0\,.$$

This is consistent with p and q having no singularities other than at $z = 0$.

Remark. We note that there is no difficulty in solving a_n from a_{n-2} using the recurrence relation if $\sigma = +\nu$. However, if $\sigma = -\nu$ the recursion will fail with a_n predicted to be infinite if at any point $n = 2\nu$. There are hence potential problems if $\sigma_1 - \sigma_2 = 2\nu \in \mathbb{Z}$, i.e. if the indices differ by an integer.

- $2\nu \notin \mathbb{Z}$. First, suppose that $2\nu \notin \mathbb{Z}$ so that σ_1 and σ_2 do not differ by an integer. In this case, a_n is solved to be
$$a_n = \begin{cases} 0 & n = 1, 3, 5\ldots \\ -\dfrac{a_{n-2}}{n(n\pm 2\nu)} & n = 2, 4, 6, \ldots \end{cases}$$
and so we get two linearly independent solutions
$$y_1 = a_0 z^{+\nu}\left(1 - \frac{z^2}{4(1+\nu)} + \frac{z^4}{32(1+\nu)(2+\nu)} + \ldots\right),$$
$$y_2 = a_0 z^{-\nu}\left(1 - \frac{z^2}{4(1-\nu)} + \frac{z^4}{32(1-\nu)(2-\nu)} + \ldots\right).$$
- $2\nu = 2m + 1$, $m \in \mathbb{N}$. In this case, even though σ_1 and σ_2 differ by an odd integer there is no problem. The solutions are still as above. This is because for Bessel's equation, the power series proceed in even powers of z, and hence the problem recursion when $n = 2\nu = 2m + 1$ is never encountered.
- $2\nu = 0$. If $\nu = 0$ then $\sigma_1 = \sigma_2$ and we can only find one power series solution of the proposed form
$$y = a_0\left(1 - \frac{1}{4}z^2 + \ldots\right).$$
- $2\nu = 2m$, $m \in \mathbb{N}$. If ν is a positive integer, m, then we can find one solution by choosing $\nu = \sigma$. However, if we take $\sigma = -\nu$ then a_{2m} is predicted to be infinite. The second series solution fails.

Remark. The existence of a second power series solution for $2\nu = 2m + 1$, $m \in \mathbb{N}$ is a lucky accident. In general, there exists only one solution of the proposed form whenever the indices σ_1 and σ_2 differ by an integer.

Bessel's equation of zeroth order

In order to obtain an idea of how to proceed when $\sigma_1 - \sigma_2 \in \mathbb{Z}$, first consider the example of Bessel's equation of zeroth order. Let y_1 denote the power series solution obtained

$$y_1 = a_0\left(1 - \frac{1}{4}z^2 + \ldots\right).$$

Then from what we derived before, a second linearly independent solution is given by

$$y_2 = \kappa y_1(z)\int^z \frac{1}{\eta y_1^2(\eta)}\,\mathrm{d}\eta\,.$$

For small positive z we can deduce that

$$\begin{aligned} y_2(z) &= \kappa a_0(1+O(z^2))\int^z \frac{1}{\eta a_0^2}(1+O(\eta^2))\,\mathrm{d}\eta \\ &= \frac{\kappa}{a_0}\log z + \dots \end{aligned}$$

We conclude that the second solution contains a logarithm.

Claim 7.8. Let σ_1, σ_2 be two solutions to the indicial equation for a regular singular point at $z=0$. Order them so that

$$\mathrm{Re}\{\sigma_1\} \geq \mathrm{Re}\{\sigma_2\}\,.$$

Then we can always find one solution of the form

$$y_1(z) = z^{\sigma_1}\sum_{n=0}^{\infty} a_n z^n\,.$$

If $\sigma_1 - \sigma_2 \in \mathbb{Z}$ we claim that the second solution takes the form

$$y_2(z) = z^{\sigma_2}\sum_{n=0}^{\infty} b_n z^n + k y_1(z)\log z\,,$$

for some number of k. The coefficient b_n can be found by substitution into the ODE. In some very special cases, k may vanish, but in general $k \neq 0$.

Bessel's Equation of Integer Order

Suppose that y_1 is the series solution with $\sigma = +m$ to

$$z^2y'' + zy' + (z^2-m^2)y = 0\,,$$

Hence,

$$y_1 = z^m\sum_{l=0}^{\infty} a_{2l}z^{2l}\,,$$

since $a_{2l+1}=0$ for integer l. Let

$$y = ky_1\log z + w\,,$$

then

$$y' = ky_1'\log z + \frac{ky_1}{z} + w'\,,$$

$$y'' = ky_1''\log z + \frac{2ky_1'}{z} - \frac{ky_1}{z^2} + w''\,.$$

On substituting into Bessel's equation, and using the fact that y_1 is a solution of the equation, we find that

$$z^2w'' + zw' + (z^2-m^2)w = -2kzy_1'\,.$$

Based on our claim, we now seek a series solution of the form

$$w = kz^{-m}\sum_{n=0}^{\infty} b_n z^n\,.$$

Substitution gives

$$k\sum_{n=1}^{\infty} n(n-2m)b_n z^{n-m} + k\sum_{n=0}^{\infty} b_n z^{n-m+2} = -2k\sum_{l=0}^{\infty}(2l+m)a_{2l}z^{2l+m}\,.$$

After multiplying by z^m and making transformations $n \to n-2$ and $2l \to n-2m$ in the second and third sums respectively, it follows that

$$\sum_{n=1}^{\infty} n(n-2m)b_n z^n + \sum_{n=2}^{\infty} b_{n-2} z^n = -2 \sum_{n=2m,\, n \text{ even}}^{\infty} (n-m) a_{n-2m} z^n \,.$$

We now demand the combined coefficient of z^n is zero. Consider the even and odd powers of z^n in turn.

- *Odd n.* From equating powers of z^1 it follows that $b_1 = 0$. Next, from writing $n = 2l+1$ ($l = 1, 2, \ldots$) and equating the powers of z^{2l+1}, we obtain the recurrence relation:

  $$(2l+1)(2l+1-2m)b_{2l+1} = -b_{2l+1} \,.$$

  Since $b_1 = 0$, we conclude that $b_{2l+1} = 0$ ($l = 1, 2, \ldots$).

- *Even n.* Let $n = 2l$ ($l = 1, 2, \ldots$), then from equating powers of z^{2l} we obtain

  $$\begin{aligned} b_{2l-2} &= -4l(l-m)b_{2l} && \text{for } 1 \le l \le m-1 \\ b_{2m-2} &= -2m a_0 && \text{for } l = m \\ b_{2l} &= -\frac{1}{4l(l-m)} b_{2l-2} - \frac{2l-m}{2l(l-m)} a_{2l-2m} && \text{for } l \ge m+1 \,. \end{aligned}$$

  To determine the even coefficients, b_{2l},

  - first, after noting that $2m-2 \ge 0$, solve for b_{2m-2} in terms of a_0 from the second equation;
  - next, if $m \ge 2$, solve for the b_{2l} ($l = m-2, m-3, \ldots, 0$) recurrently using the first recurrence relation;
  - then, having noted that a non-zero value simply generates a solution proportional to y_1, choose a value for b_{2m}, e.g., wlog, $b_{2m} = 0$;
  - finally, having fixed b_{2m}, solve for the b_{2l} ($l = m+1, m+2, \ldots$) using the third recurrence relation.

Remark. These examples illustrate a feature that is commonly encountered in scientific applications: one solution is regular (analytic) and the other is singular. Often only the regular solution is an acceptable solution for the scientific problem.

Theorem 7.9. The Bessel equation of order ν always has a solution of the form

$$J_\nu(z) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!\Gamma(n+\nu+1)} \Big(\frac{z}{2}\Big)^{2n+\nu} \,,$$

known as the *Bessel function of the first kind* of order ν. The k-th zero of the function is denoted as $j_{\nu k}$. The second solution is singular at $z = 0$, known as the *Bessel function of the second kind*.

Here, $\Gamma(z)$ is the generalisation of the factorial to the complex numbers (except non-positive integers). It is defined by

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} \, \mathrm{d}t$$

with the property $\Gamma(n) = (n-1)!$ for positive integers n.

7.3.4 Irregular Singular Points

If either $(z-z_0)p(z)$ or $(z-z_0)^2 q(z)$ is not analytic at the point $z = z_0$, it is an irregular singular point of the equation. The solution can have worse kinds of singular behaviours there.

Example. The equation $z^4 y'' + 2z^3 y' - y = 0$ has an irregular singular point at $z = 0$. Its solutions are $\exp\big(\pm z^{-1}\big)$, both of which have an essential singularity at $z = 0$.

7.4 The Method of Variation of Parameters (Non-examinable)

To solve an inhomogeneous ODE

$$y''(x) + p(x)y'(x) + q(x)y(x) = f(x)\,,$$

the question that remains is how to find a particular solution. First, suppose that we have solved the homogeneous equation and found two linearly-independent solutions y_1 and y_2. Then in order to find a particular solution consider

$$y_0(x) = u(x)y_1(x) + v(x)y_2(x)\,.$$

If u and v are constants, y_0 would solve the homogeneous equation. However, we allow these parameters to vary, i.e. to be functions of x, in such a way that y_0 solves the inhomogeneous equation.

Remark. We have gone from one unknown function (y_0) and one equation to two unknown functions (u, v) and one equation. We will need to choose another equation.

We now differentiate our expression of y_0 to find that

$$y_0' = (uy_1' + vy_2') + (u'y_1 + v'y_2)$$

$$y_0'' = (uy_1'' + vy_2'' + u'y_1' + v'y_2') + (u''y_1 + v''y_2 + u'y_1' + v'y_2')\,.$$

If we directly substitute this into the equation, we will not make much progress. However, we can demand u and v to satisfy the extra equation

$$u'y_1 + v'y_2 = 0\,.$$

Then the derivatives of y_0 become

$$y_0' = uy_1' + vy_2'$$
$$y_0'' = uy_1'' + vy_2'' + u'y_1' + v'y_2'$$

Therefore

$$\begin{aligned} y_0'' + py_0' + qy_0 &= u(y_1'' + py_1' + qy_1) + v(y_2'' + py_2' + qy_2) + u'y_1' + v'y_2' \\ &= u'y_1' + v'y_2'\,, \end{aligned}$$

since y_1 and y_2 solve the homogeneous equation. Hence y_0 solves the inhomogeneous equation if

$$u'y_1' + v'y_2' = f\,,$$

We now have two simultaneous equations for u' and v'

$$\begin{cases} u'y_1 + v'y_2 = 0 \\ u'y_1' + v'y_2' = f\,, \end{cases}$$

with solution

$$\begin{cases} u' = -\dfrac{fy_2}{W} \\ v' = \dfrac{fy_1}{W}\,, \end{cases}$$

where W is the Wronskian

$$W = y_1y_2' - y_2y_1'\,.$$

Integrating we obtain

$$\begin{cases} u = -\displaystyle\int_a^x \frac{y_2(\xi)f(\xi)}{W(\xi)}\,\mathrm{d}\xi \\ v = \displaystyle\int_a^x \frac{y_1(\xi)f(\xi)}{W(\xi)}\,\mathrm{d}\xi\ , \end{cases}$$

where the lower bound of integration is arbitrary. Substituting this result back into the expression of y_0 we obtained

$$y_0(x) = \int_a^x \frac{f(\xi)}{W(\xi)}(y_1(\xi)y_2(x) - y_1(x)y_2(\xi))\,\mathrm{d}\xi\ .$$

Remark. We observe that, since the integrand is zero when $\xi = x$,

$$y_0'(x) = \int_a^x \frac{f(\xi)}{W(\xi)}(y_1(\xi)y_2'(x) - y_1'(x)y_2(\xi))\,\mathrm{d}\xi\ .$$

Hence the particular solution $y = y_0$ we obtained satisfies the initial value boundary conditions

$$y(a) = y'(a) = 0\,.$$

More general initial value boundary conditions would be inhomogeneous:

$$y(a) = k_1\,,\ y'(a) = k_2\,,$$

where k_1 and k_2 are constants. Such inhomogeneous boundary conditions are obtained by adding suitable multiples of the linearly independent solutions of the homogeneous equation, i.e. y_1 and y_2.

8 Sturm–Liouville Theory

8.1 Abstract Eigenvalue Problems

8.1.1 Eigenfunctions

Suppose we want to solve an inhomogeneous ordinary differential equation of the form

$$\tilde{\mathcal{L}}y(x) = f(x)\,,$$

where $\tilde{\mathcal{L}}$ is a general second-order linear differential operator in the form

$$\tilde{\mathcal{L}} = p(x)\frac{\mathrm{d}^2}{\mathrm{d}x^2} + r(x)\frac{\mathrm{d}}{\mathrm{d}x} + s(x)\,,$$

with p, r and s being real functions, and boundary conditions on the solutions are specified at $x = \alpha$ and $x = \beta$.

Except for simple $f(x)$, it will generally not be possible to find a particular integral in a closed form. However, we can exploit the linearity of $\tilde{\mathcal{L}}$ further to find the solution in terms of a superposition of a set of solutions. A convenient choice for the set of basis functions is the set of *eigenfunctions* of $\tilde{\mathcal{L}}$ that satisfy the boundary conditions.

Definition 8.1. The *eigenfunctions*, $\{y_i\}$, of an operator $\tilde{\mathcal{L}}$ are the functions that satisfy the *eigenvalue equation*

$$\tilde{\mathcal{L}}y_i(x) = \lambda_i y_i(x)\,,$$

where the constants $\{\lambda_i\}$ are the *eigenvalues* of $\tilde{\mathcal{L}}$.

Remark. Note the close analogy between matrices and differential operators: functions form a vector space and differential operators are linear maps.

8.1.2 Inner Products of Functions

Let V be a vector space of functions $[\alpha, \beta] \to \mathbb{C}$. Let us equip V with an inner product.

Definition 8.2. For two complex functions $u(x)$ and $v(x)$ defined for $\alpha \le x \le \beta$, an *inner product* is defined as

$$\langle u|v\rangle_w \coloneqq \int_\alpha^\beta u^*(x)w(x)v(x)\,\mathrm{d}x\ ,$$

where the real, positive function $w : (a, b) \to \mathbb{R}_{>0}$ is called the *weight function*.

Remark. When the weight function $w = 1$, abbreviate

$$\langle u|v\rangle_w =: \langle u|v\rangle\ .$$

Proposition 8.3. For complex functions $u(x)$, $v(x)$ and $t(x)$ and complex numbers a and b:

- $\langle u|v\rangle_w = \langle v|u\rangle_w^*$.
- $\langle u|av + bt\rangle_w = a\langle u|v\rangle_w + b\langle u|t\rangle_w$.
- $\langle au + bv|t\rangle_w = a^*\langle u|t\rangle_w + b^*\langle v|t\rangle_w$.

Remark. An inner product is a sesquilinear and Hermitian form.

Definition 8.4. Two functions u and v are *orthogonal* if $\langle u|v\rangle_w = 0$.

Definition 8.5. The *norm* of $u(x)$, $\|u\|_w$, is defined by

$$\|u\|_w \coloneqq \sqrt{\langle u|u\rangle_w} = \sqrt{\int_\alpha^\beta w(x)|u(x)|^2\,\mathrm{d}x}\,.$$

Note that $\|u\|_w$ is always real and non-negative.

Remark. We will restrict ourselves to reasonably well-behaved functions such that

$$\|u\| = 0 \implies u(x) = 0\,.$$

However, for some less well-behaved functions, $\|u\| = 0$ may not imply $u(x) = 0$. An example would be the *Dirichlet function*, $D(x)$, which is unity when x is rational and zero otherwise.

Definition 8.6. A *normalised function* $y(x)$ is a function with a unit norm, i.e.

$$\|y\|_w = 1\,.$$

8.1.3 Adjointness

Definition 8.7. For a general differential operator $\tilde{\mathcal{L}}$, and a given inner product $\langle u|v\rangle_w$, the *adjoint operator* of $\tilde{\mathcal{L}}$, $\tilde{\mathcal{L}}^\dagger$, is defined to be the operator such that

$$\left\langle u\middle|\tilde{\mathcal{L}}v\right\rangle_w = \left\langle \tilde{\mathcal{L}}^\dagger u\middle|v\right\rangle_w\,.$$

Definition 8.8. A differential operator $\tilde{\mathcal{L}}$ is *self-adjoint* (*Hermitian*) on $(V, \langle\,\cdot\,|\,\cdot\,\rangle_w)$ if

$$\left\langle u\middle|\tilde{\mathcal{L}}v\right\rangle_w = \left\langle \tilde{\mathcal{L}}u\middle|v\right\rangle_w \quad \forall\, u, v \in V\,.$$

Remark. Self-adjoint operators are analogous to Hermitian matrices. Suppose that an inner product for column vectors u and v is defined by

$$\langle \mathsf{u}|\mathsf{v}\rangle = \mathsf{u}^\dagger\mathsf{v}\,.$$

Then for a Hermitian matrix H,

$$\begin{aligned}\langle \mathsf{u}|\mathsf{H}\mathsf{v}\rangle &= \mathsf{u}^\dagger\mathsf{H}\mathsf{v} = \mathsf{u}^\dagger\mathsf{H}^\dagger\mathsf{v}\\ &= (\mathsf{H}\mathsf{u})^\dagger\mathsf{v} = \langle \mathsf{H}\mathsf{u}|\mathsf{v}\rangle\,.\end{aligned}$$

Theorem 8.9. If $\tilde{\mathcal{L}}$ is self-adjoint on $(V, \langle\,\cdot\,|\,\cdot\,\rangle_w)$, then

(i) eigenvalues are real.

(ii) eigenfunctions with different eigenvalues are orthogonal.

Proof. (i) Let $\tilde{\mathcal{L}}y = \lambda y$,

$$\begin{aligned}(\lambda^* - \lambda)\|y\|_w^2 &= (\lambda^* - \lambda)\,\langle y|y\rangle_w\\ &= \langle \lambda y|y\rangle_w - \langle y|\lambda y\rangle_w\\ &= \left\langle \tilde{\mathcal{L}}y\middle|y\right\rangle_w - \left\langle y\middle|\tilde{\mathcal{L}}y\right\rangle_w = 0\,.\end{aligned}$$

(ii) Let $\tilde{\mathcal{L}}y_1 = \lambda_1 y_1$, $\tilde{\mathcal{L}}y_2 = \lambda_2 y_2$, $\lambda_1 \neq \lambda_2$.

$$\begin{aligned}(\lambda_1 - \lambda_2)\,\langle y_1|y_2\rangle_w &= \langle \lambda_1 y_1|y_2\rangle_w - \langle y_1|\lambda_2 y_2\rangle_w\\ &= \left\langle \tilde{\mathcal{L}}y_1\middle|y_2\right\rangle - \left\langle y_1\middle|\tilde{\mathcal{L}}y_2\right\rangle_w = 0\\ \implies \langle y_1|y_2\rangle_w &= 0\,.\end{aligned}$$

□

Remark. We claim without proof that mutually orthogonal eigenfunctions can always be constructed, even for repeated eigenvalues. Further, if we normalize all eigenfunctions to have unit norm then we have an orthonormal set of eigenfunctions.

8.2 Sturm–Liouville Operators

Definition 8.10. A *Sturm–Liouville operator* is a second-order differential operator, defined on the range $\alpha \le x \le \beta$, of the form,

$$\mathcal{L} = -\frac{\mathrm{d}}{\mathrm{d}x}\left(\rho(x)\frac{\mathrm{d}}{\mathrm{d}x}\right) + \sigma(x)\,,$$

where σ and ρ are real, smooth functions, and $\rho(x) > 0$ for $\alpha < x < \beta$.

Definition 8.11. For a Sturm–Liouville problem on (a, b), we say an endpoint $c \in \{\alpha, \beta\}$ is *singular* if $\rho(c) = 0$, and *non-singular* if $\rho(c) \neq 0$.

Remark. We only need to specify boundary conditions at non-singular endpoints.

At a non-singular endpoint $c \in \{\alpha, \beta\}$, we will impose boundary condition of the form

$$a_c y(c) + b_c y'(c) = 0\,,$$

where $a_c, b_c \in \mathbb{R}$, and a_c and b_c are not both zero.

We call these real, homogeneous boundary conditions — latter because if y_1, y_2 satisfy the boundary conditions, then so does $c_1 y_1 + c_2 y_2$.

We will work on vector spaces of the form

$$V = \left\{ y \in C^2[\alpha, \beta] \;\middle|\; \begin{array}{c} y \text{ satisfies real, homogeneous boundary conditions} \\ \text{at each non-singular endpoint.} \end{array} \right\},$$

where $C^2[\alpha, \beta]$ means the vector space of all twice differentiable functions in $[\alpha, \beta]$.

Lemma 8.12. A Sturm–Liouville operator of the form

$$\mathcal{L} = -\frac{\mathrm{d}}{\mathrm{d}x}\left(\rho(x)\frac{\mathrm{d}}{\mathrm{d}x}\right) + \sigma(x)$$

is self-adjoint if the boundary conditions are such that

$$[\rho W(v, u^*)]_\alpha^\beta = 0\,,$$

where W is the Wronskian.

Proof. Sturm–Liouville operators satisfy

$$\begin{aligned}
\langle u|\mathcal{L}v\rangle &= \int_\alpha^\beta u^* \mathcal{L}v \,\mathrm{d}x = -\int_\alpha^\beta u^* (\rho v')' \,\mathrm{d}x + \int_\alpha^\beta u^* \sigma v \,\mathrm{d}x \\
&= -[u^* \rho v']_\alpha^\beta + \int_\alpha^\beta \rho v' u^{*\prime} \,\mathrm{d}x + \int_\alpha^\beta v \sigma u^* \,\mathrm{d}x \\
&= -[u^* \rho v']_\alpha^\beta + \left[\rho v u^{*\prime}\right]_\alpha^\beta - \int_\alpha^\beta v(\rho u^{*\prime})' \,\mathrm{d}x + \int_\alpha^\beta v \sigma u^* \,\mathrm{d}x \\
&= \int_\alpha^\beta (\mathcal{L}u)^* v \,\mathrm{d}x + \left[\rho(v u^{*\prime} - u^* v')\right]_\alpha^\beta \\
&= \langle \mathcal{L}u|v\rangle + [\rho W(v, u^*)]_\alpha^\beta\,,
\end{aligned}$$

where $[\rho W(v,u^*)]_\alpha^\beta$ is the boundary term. Therefore we have

$$\langle u|\mathcal{L}v\rangle = \langle \mathcal{L}u|v\rangle$$

if the boundary terms vanish. □

Theorem 8.13. If $y_1, y_2 \in V$, then $\mathcal{L}$ is self-adjoint on $\langle y_1|y_2\rangle$.

Proof. If the endpoint c is singular, then $\rho(c) = 0$, so $\rho W(v,u^*)$ is trivially zero.

If the endpoint c is non-singular, then we have the boundary conditions

$$\begin{cases} a_c y_1(c) + b_c y_1'(c) = 0 \\ a_c y_2(c) + b_c y_2'(c) = 0 \end{cases} \implies a_c y_2^*(c) + b_c {y_2'}^*(c) = 0$$

$$\implies \begin{pmatrix} y_1(c) & y_1'(c) \\ y_2^*(c) & {y_2'}^*(c) \end{pmatrix} \begin{pmatrix} a_c \\ b_c \end{pmatrix} = \mathbf{0}\,.$$

Since a_c, b_c are not both zero,

$$\det \begin{pmatrix} y_1(c) & y_1'(c) \\ y_2^*(c) & {y_2'}^*(c) \end{pmatrix} = W(y_1, y_2^*) = 0\,.$$

Therefore, $[\rho W(v,u^*)]_\alpha^\beta = 0$ is satisfied. □

8.2.1 Reduction to Sturm–Liouville Form

Lemma 8.14. For a general second-order linear differential operator, $\tilde{\mathcal{L}}$, that is not in the Sturm–Liouville form, we can always find a weight function $w(x)$ such that $\mathcal{L} = w\tilde{\mathcal{L}}$ is in the Sturm–Liouville form.

Proof. Consider a general second-order linear differential operator defined for $\alpha \le x \le \beta$

$$\tilde{\mathcal{L}} = p\frac{\mathrm{d}^2}{\mathrm{d}x^2} + r\frac{\mathrm{d}}{\mathrm{d}x} + s = -\frac{\mathrm{d}}{\mathrm{d}x}\left(a(x)\frac{\mathrm{d}}{\mathrm{d}x}\right) - b(x)\frac{\mathrm{d}}{\mathrm{d}x} - c(x)\,,$$

where a, b, c are all real, smooth functions and $a(x) > 0$ for $\alpha < x < \beta$. Then

$$w\tilde{\mathcal{L}} = -\frac{\mathrm{d}}{\mathrm{d}x}\left(aw\frac{\mathrm{d}}{\mathrm{d}x}\right) + (aw' - bw)\frac{\mathrm{d}}{\mathrm{d}x} - wc\,.$$

Choose $w(x)$ such that $aw' = bw$, i.e.

$$w(x) = C\exp\left(\int^x \frac{b(u)}{a(u)}\,\mathrm{d}u\right),$$

where the arbitrary constant is chosen to be positive and hence $w(x)$ is positive. This gives an operator of Sturm–Liouville form

$$\mathcal{L} = w\tilde{\mathcal{L}} = -\frac{\mathrm{d}}{\mathrm{d}x}\left(aw\frac{\mathrm{d}}{\mathrm{d}x}\right) - wc\,.$$

□

Theorem 8.15. A general differential operator,

$$\tilde{\mathcal{L}} = -\frac{\mathrm{d}}{\mathrm{d}x}\left(a(x)\frac{\mathrm{d}}{\mathrm{d}x}\right) - b(x)\frac{\mathrm{d}}{\mathrm{d}x} - c(x)\,,$$

is self-adjoint with respect to an inner product of u and v with weight function w, given by

$$w(x) = C \exp\left(\int^x \frac{b(u)}{a(u)} \, \mathrm{d}u \right),$$

if the boundary conditions are real, homogeneous at each non-singular endpoint such that

$$[waW(y_1, y_2^*)]_\alpha^\beta = 0 \,.$$

Corollary. If $y(x)$ satisfies the eigenvalue equation for $\tilde{\mathcal{L}}$,

$$\tilde{\mathcal{L}} y = \lambda y \,,$$

then it also satisfies a generalised eigenvalue equation for $\mathcal{L}$ reduced to Sturm–Liouville form with weight function $w(x)$,

$$\mathcal{L} y = \lambda w(x) y \,.$$

8.2.2 Hermite's Equation

The *Hermite's equation*,

$$y'' - 2xy' + 2ny = 0 \,,$$

arises when solving the quantum harmonic oscillator.

This can be rewritten as an eigenvalue equation, $\tilde{\mathcal{L}} y = \lambda y$, with

$$\tilde{\mathcal{L}} = -\frac{\mathrm{d}^2}{\mathrm{d}x^2} + 2x \frac{\mathrm{d}}{\mathrm{d}x} = -e^{x^2} \frac{\mathrm{d}}{\mathrm{d}x} \left(e^{-x^2} \frac{\mathrm{d}}{\mathrm{d}x} \right),$$

$$\lambda = 2n \,.$$

This $\tilde{\mathcal{L}}$ is not in the Sturm–Liouville form. Multiplying by $w(x) = e^{-x^2}$ gives

$$\mathcal{L} = w\tilde{\mathcal{L}} = -\frac{\mathrm{d}}{\mathrm{d}x} \left(e^{-x^2} \frac{\mathrm{d}}{\mathrm{d}x} \right),$$

which is in the Sturm–Liouville form. The Hermite's equation can therefore be rewritten as

$$\mathcal{L} y = \lambda w y \,.$$

If we require that the norm $\|y\|_w$ is finite, then non-zero solutions exist only when n is a non-negative integer: these are n^{th} order polynomials called *Hermite polynomials*.

8.2.3 Legendre's Equation

The *Legendre's equation*,

$$(1 - x^2) y'' - 2xy' + \ell(\ell + 1) y = 0 \,,$$

arises when solving Laplace's equation with axial symmetry or Schrödinger equation in 3D with a central potential. We may view this equation as an eigenvalue equation, $\mathcal{L} y = \lambda y$, with

$$\mathcal{L} = -\frac{\mathrm{d}}{\mathrm{d}x} \left[(1 - x^2) \frac{\mathrm{d}}{\mathrm{d}x} \right], \ \lambda = \ell(\ell + 1) \,,$$

where $\mathcal{L}$ is of the Sturm–Liouville form. This operator is self-adjoint when acting on functions $y(x)$ that are finite at $x = \pm 1$.

In previous sections, we have found that the only non-zero solutions for which y is finite at both $x = \pm 1$ are polynomials. This only happens if ℓ is an integer.

These polynomials are known as *Legendre polynomials*, $P_\ell(x)$, with conventional normalisation such that $P_\ell(1) = 1$. Note that Legendre polynomials are orthogonal, but with conventional normalisation, they are not orthonormal:

$$\int_{-1}^{1} P_\ell(x) P_k(x)\,\mathrm{d}x = \frac{2}{2\ell+1}\delta_{\ell k}\,.$$

8.3 Eigenfunction Expansion

Theorem 8.16. If $\tilde{\mathcal{L}}$ is self-adjoint on $(V, \langle\,\cdot\,|\,\cdot\,\rangle_w)$, then the countable set of its eigenfunctions, $\{y_n\}_{n=1}^{\infty}$, form a *complete orthogonal set* on V. This means that for any $f \in V$, we can write

$$f = \sum_{n=1}^{\infty} \tilde{f}_n y_n\,,$$

where

$$\begin{aligned}\tilde{f}_n &= \frac{\langle f|y_n\rangle_w}{\langle y_n|y_n\rangle_w}\\ &= \langle f|y_n\rangle_w \text{ if } y_n \text{ are normalised.}\end{aligned}$$

We call the $\tilde{f}_n$ the *generalised Fourier coefficients* of f.

Remark. We claim that the eigenfunctions of a self-adjoint operator form a basis on V.

Proof. We cannot prove that such an expansion always exists (David Hilbert famously claimed it was true). However, if we assume such an expansion does exist, by the orthogonality of the eigenfunctions of a self-adjoint operator, the coefficients must be given by $\langle y_n|f\rangle_w$ because

$$\begin{aligned}\langle y_n|f\rangle_w &= \left\langle y_n \middle| \sum_{m=1}^{\infty} a_m y_m(x) \right\rangle\\ &= \sum_{m=1}^{\infty} a_m \langle y_n|y_m\rangle_w\\ &= \sum_{m=1}^{\infty} a_m \delta_{nm} = a_n\end{aligned}$$

□

Proposition 8.17 (The completeness relation). Let $\{y_n\}$ be a complete orthonormal set of eigenfunctions of an operator,

$$\sum_{n=1}^{\infty} y_n(x) y_n^*(\xi) = \frac{1}{w(\xi)}\delta(x-\xi)\,.$$

Proof. If the eigenfunctions are complete, we must have

$$\begin{aligned}f(x) = \sum_{n=1}^{\infty} a_n y_n(x) &= \sum_{n=1}^{\infty} y_n(x)\,\langle y_n|f\rangle_w\\ &= \sum_{n=1}^{\infty} y_n(x) \int_\alpha^\beta w(\xi) y_n^*(\xi) f(\xi)\,\mathrm{d}\xi\\ &= \int_\alpha^\beta f(\xi)\left[w(\xi)\sum_{n=1}^{\infty} y_n^*(\xi) y_n(x)\right]\mathrm{d}\xi\,.\end{aligned}$$

Since

$$f(x) = \int_\alpha^\beta f(\xi)\delta(x-\xi)\,\mathrm{d}\xi\ ,$$

if our proposition holds true for all f, we have

$$\sum_{n=1}^{\infty} y_n(x)y_n^*(\xi) = \frac{1}{w(\xi)}\delta(x-\xi)\,.$$

□

Remark. The completeness relation defines a complete set of functions $\{y_n\}$ such that any function $f \in V$ can be expressed on the basis of the eigenfunctions y_n. □

Example. Fourier Series.

Again, consider the Sturm–Liouville operator

$$\mathcal{L} = -\frac{\mathrm{d}^2}{\mathrm{d}x^2}\,.$$

In this case assume that the operator acts on functions that are 2π-periodic, i.e. $y(x) = y(x+2\pi)$. Write the general solution to the eigenvalue equation $\mathcal{L}y = \lambda y$ as

$$y = Ae^{i\sqrt{\lambda}x} + Be^{-i\sqrt{\lambda}x}\,,$$

where A and B are constants. This solution is 2π-periodic if $\lambda = n^2$ for integer n. Label the eigenfunctions by y_n for $n = \ldots, -1, 0, 1, \ldots$, with corresponding eigenvalues $\lambda_n = n^2$. Although there are repeated eigenvalues ($n = \pm k$), there still exists an orthonormal set of eigenfunctions as claimed, i.e.

$$y_n = \frac{1}{\sqrt{2\pi}}e^{inx}\,,\ n \in \mathbb{Z}\,.$$

Hence, a 2π-periodic function f has an eigenfunction expansion

$$f(x) = \frac{1}{\sqrt{2\pi}}\sum_{n=-\infty}^{\infty} a_n e^{inx}$$

for some complex coefficients a_k, which is the Fourier series expansion of f. To the extent that any such function has a Fourier series expansion, the set $\{y_k \mid k \in \mathbb{Z}\}$ is complete, with completeness relation

$$\sum_{k\in\mathbb{Z}} e^{ik(x-\xi)} = 2\pi\delta(x-\xi)\,.$$

Remark. The Fourier series is a particular example of an expansion in terms of the eigenfunctions of a self-adjoint operator.

8.4 Solution of Differential Equations

8.4.1 Green's Functions for Sturm–Liouville Operators

Consider the differential equation of the form

$$\mathcal{L}y(x) = f(x)$$

with homogeneous boundary conditions for some forcing function $f(x)$. We require $\mathcal{L}$ to be in Sturm–Liouville form and have a complete set of normalised eigenfunctions $\{y_n\}_{n=1}^\infty$ with eigenvalues $\{\lambda_n\}_{n=1}^\infty$ such that

$$\mathcal{L}y_n = \lambda_n w y_n$$
$$\langle y_n | y_m \rangle_w = \delta_{mn}$$

Lemma 8.18. A formal solution is given by

$$y(x) = \int_\alpha^\beta G(x;\xi) f(\xi)\,\mathrm{d}\xi\ ,$$

where the Green's function $G(x;\xi)$ satisfies the boundary conditions when considered both as a function of x and ξ, and it is the response of the system to a point-like source:

$$\mathcal{L}G(x;\xi) = \delta(x-\xi)\,.$$

Proof.

$$\begin{aligned}\mathcal{L}y &= \mathcal{L}\int_\alpha^\beta G(x;\xi) f(\xi)\,\mathrm{d}\xi \\ &= \int_\alpha^\beta \mathcal{L}G(x;\xi) f(\xi)\,\mathrm{d}\xi \\ &= \int_\alpha^\beta \delta(x-\xi) f(\xi)\,\mathrm{d}\xi = f(x)\,.\end{aligned}$$

□

Recall that if $\mathcal{L}$ is a general second-order linear differential operator, then Green's function subjected to homogeneous boundary conditions at a and b can be written as

$$G(x;\xi) = \begin{cases} \dfrac{y_1(x)y_2(\xi)}{W(\xi)} & \text{for } x \in [a,\xi) \\ \dfrac{y_1(\xi)y_2(x)}{W(\xi)} & \text{for } x \in [\xi,b]\,, \end{cases}$$

where y_1 and y_2 satisfy the boundary conditions at a and b respectively (Theorem 2.19).

However, if $\mathcal{L}$ is a Sturm–Liouville operator, the Green's function would have a much better form.

Lemma 8.19. For a Sturm–Liouville operator $\mathcal{L}$ with $\{y_n\}_{n=1}^\infty$ and eigenvalues $\{\lambda_n\}_{n=1}^\infty$, the eigenfunction expansion of the Green's function is given by

$$G(x;\xi) = \sum_{n=1}^\infty \frac{1}{\lambda_n} y_n(x) y_n^*(\xi)\,.$$

Proof. The Green's function constructed in this way satisfies the boundary conditions, and satisfies

$$\begin{aligned}\mathcal{L}G(x;\xi) &= \sum_{n=1}^\infty \frac{\mathcal{L}y_n}{\lambda_n} y_n^*(\xi) = \sum_{n=1}^\infty \frac{\lambda_n w(x) y_n}{\lambda_n} y_n^*(\xi) \\ &= w(x)\sum_{n=1}^\infty y_n(x) y_n^*(\xi) \\ &= w(x)\frac{1}{w(x)}\delta(x-\xi) && \text{by the completeness relation} \\ &= \delta(x-\xi)\,.\end{aligned}$$

□

Remark. This can be seen as an eigenfunction expansion of Green's function, although paradoxically it is not twice differentiable in $[a,b]$.

Remark. By the above equation of the Green's function, we can observe that

$$G(x;\xi) = G^*(\xi;x)\,.$$

Theorem 8.20. The solution to the equation

$$\mathcal{L}y(x) = f(x)\,,$$

where $\mathcal{L}$ is a differential operator in the Sturm–Liouville form, is

$$y = \sum_{n=1}^{\infty} y_n(x)\frac{\langle y_n|f\rangle}{\lambda_n}\,,$$

where y_n and λ_n are the corresponding eigenvectors and eigenvalues of $\mathcal{L}$.

Proof.

$$\begin{aligned} y(x) &= \int_\alpha^\beta \sum_{n=1}^{\infty} \frac{1}{\lambda_n} y_n(x) y_n^*(\xi) f(\xi)\,\mathrm{d}\xi \\ &= \sum_{n=1}^{\infty} y_n(x)\frac{1}{\lambda_n}\int_\alpha^\beta y_n^*(\xi) f(\xi)\,\mathrm{d}\xi \\ &= \sum_{n=1}^{\infty} y_n(x)\frac{\langle y_n|f\rangle}{\lambda_n}\,. \end{aligned}$$

□

Remark. If $\mathcal{L}$ has a zero eigenvalue then $G(x;\xi)$ will not exist and there is no finite solution for y for a general f. In other words, there is no solution to the forced problem if there is a solution to the homogeneous equation, $\mathcal{L}y = 0$, satisfying the boundary conditions. The vanishing of one or more of the eigenvalues is related to the phenomenon called *resonance*. If a solution to the problem exists in the absence of the forcing f, then any non-zero forcing elicits an infinite response.

If instead λ_1 is very small compared to others, then

$$y(x) = \sum_{n=1}^{\lambda} y_n(x)\frac{\langle y_n|f\rangle}{\lambda_n} \approx \frac{y_1(x)}{\lambda_1}\langle y_1|f\rangle + \dots$$

and the omitted terms are all suppressed as long as $\langle y_1|f\rangle$ is not too small. Any forcing function with non-zero y_1-component will cause a large resonant response $\propto y_1(x)$.

Example. Solve the equation

$$-\frac{\mathrm{d}^2y}{\mathrm{d}x^2} - \frac{\mathrm{d}y}{\mathrm{d}x} - \frac{1}{4}y = -e^{-\frac{x}{2}}$$

subjected to the boundary conditions

$$\begin{cases} y(0) = 0 \\ \dfrac{\mathrm{d}y}{\mathrm{d}x}(1) + \dfrac{1}{2}y(1) = 0\,. \end{cases}$$

For the differential operator

$$\tilde{\mathcal{L}} = -\frac{\mathrm{d}^2}{\mathrm{d}x^2} - \frac{\mathrm{d}}{\mathrm{d}x} - \frac{1}{4}\,,$$

an appropriate weight function is

$$w(x) = \exp\left(\int \mathrm{d}x\right) = e^x\,,$$

so that a Sturm–Liouville form operator would be

$$\mathcal{L} = w(x)\tilde{\mathcal{L}} = -\frac{\mathrm{d}}{\mathrm{d}x}\left(e^x\frac{\mathrm{d}}{\mathrm{d}x}\right) - \frac{1}{4}e^x\,,$$

and the equation becomes

$$\mathcal{L}y = -e^{\frac{x}{2}} .$$

First, solve the eigenvalue equation

$$\tilde{\mathcal{L}}y_n = \lambda_n y_n , \text{ or equivalently } \mathcal{L}y_n = \lambda_n w y_n .$$

A trial solution $y = e^{kx}$ yields

$$k = -\frac{1}{2} \pm i\sqrt{\lambda} .$$

Therefore,

$$y_n(x) = e^{-\frac{x}{2}} (A \sin \sqrt{\lambda_n} x + B \cos \sqrt{\lambda_n} x) .$$

Applying the boundary conditions gives

$$\begin{cases} B = 0 \\ \sqrt{\lambda_n} e^{-\frac{1}{2}} \cos \sqrt{\lambda_n} = 0 , \end{cases}$$

so

$$\lambda_n = \left(n + \frac{1}{2}\right)^2 \pi^2 ,$$

$$y_n = A_n e^{-\frac{x}{2}} \sin\left[\left(n + \frac{1}{2}\right)\pi x\right] \quad \text{for } n \in \mathbb{N}_0 .$$

With these eigenfunctions, the boundary terms vanish, so $\mathcal{L}$ is self-adjoint. Therefore, for normalised y_n,

$$\langle y_n | y_m \rangle_w = \int_0^1 y_n(x) y_m(x) e^x \, \mathrm{d}x = \delta_{nm} .$$

To normalise y_n,

$$\langle y_n | y_n \rangle_w = 1$$

gives $A_n = \sqrt{2}$, so the normalised eigenfunctions are

$$y_n(x) = \sqrt{2} e^{-\frac{x}{2}} \sin\left[\left(n + \frac{1}{2}\right)\pi x\right] .$$

Since $\mathcal{L}y_n = \lambda_n e^x y_n$ and so writing $y = \sum_n a_n y_n$, we require

$$\mathcal{L}y = \sum_n a_n \lambda_n e^x y_n = -e^{\frac{x}{2}} .$$

Multiplying by $y_m^*(x)$ and integrating gives

$$\begin{aligned} \int_0^1 \sum_n a_n \lambda_n e^x y_n(x) y_m^*(x) \, \mathrm{d}x &= \sum_n a_n \lambda_n \int_0^1 e^x y_m^*(x) y_n(x) \, \mathrm{d}x \\ &= \sum_n a_n \lambda_n \langle y_m | y_n \rangle_w \\ &= \sum_n a_n \lambda_n \delta_{mn} \\ &= a_m \lambda_m , \end{aligned}$$

$$\int_0^1 -e^{\frac{x}{2}} y_m^*(x)\,\mathrm{d}x = -\sqrt{2}\int_0^1 \sin\left[\left(m+\frac{1}{2}\right)\pi x\right]\mathrm{d}x$$
$$= -\frac{\sqrt{2}}{\left(m+\frac{1}{2}\right)\pi}\,,$$

which gives

$$a_m = -\frac{\sqrt{2}}{\lambda_m\left(m+\frac{1}{2}\right)\pi} = -\frac{\sqrt{2}}{\left(m+\frac{1}{2}\right)^3\pi}\,.$$

Finally, this gives us the solution

$$y(x) = -\frac{2}{\pi^3}e^{-\frac{x}{2}}\sum_{n=0}^{\infty}\frac{1}{\left(n+\frac{1}{2}\right)^3}\sin\left[\left(n+\frac{1}{2}\right)\pi x\right].$$

8.5 Bessel's Equation

Consider the eigenvalue problem

$$-\frac{\mathrm{d}}{\mathrm{d}r}\left(r\frac{\mathrm{d}y}{\mathrm{d}r}\right)+\frac{m^2}{r}y = \lambda r y\,, \tag{†}$$

where the boundary conditions are such that we work on the vector space

$$y \in V = \{y \in C^2[0,1] : y(1) = 0\}\,.$$

This differential operator is naturally in Sturm–Liouville form with weight function $w = r$. Let $z = r\sqrt{\lambda}$, and set $y(r) = R(z) = R(r\sqrt{\lambda})$, we can expand equation (†) to

$$\begin{cases} z^2R'' + zR' + (z^2 - m^2)R = 0 & \text{for } z \in (0,\sqrt{\lambda}) \\ R(\sqrt{\lambda}) = 0\,. \end{cases}$$

This is the form of Bessel's equation of order m we have seen before in Definition 7.7.

Lemma 8.21. The Bessel's equation of order m has two series solutions. The first solution is

$$J_m(z) = \left(\frac{z}{2}\right)^m\sum_{k=0}^{\infty}\frac{(-1)^k}{k!\Gamma(k+m+1)}\left(\frac{z}{2}\right)^{2k},$$

known as *Bessel function of the first kind* of order m. The other solution, known as *Bessel function of the second kind* of order m, is singular as $z \to 0$.

Here, we will focus on the Bessel function of the first kind. We can show that

$$J_m(z) = \sqrt{\frac{2}{\pi z}}\cos\Big(z - \frac{m\pi}{2} - \frac{\pi}{4}\Big) + O\left(\frac{1}{z^{\frac{3}{2}}}\right)$$

as $z \to \infty$. This shows that J_m has infinitely many zeroes on $z > 0$. We call them $\{j_{mk}\}$ for $k = 1, 2, \ldots$ for each m. To fix the boundary conditions, we must have

$$R(\sqrt{\lambda}) = J_m(\sqrt{\lambda}) = 0\,,$$

so the eigenvalues must be

$$\lambda_k = j_{mk}^2\,, \quad k = 1, 2, \ldots$$

The eigenfunctions are therefore

$$y_k(r) = J_m(j_{mk}r)\,.$$

We also have the orthogonality relations.

$$\begin{aligned}\langle y_k|y_l\rangle_w &= \int_0^1 J_m(j_{mk}r)J_m(j_{ml}r)r\,\mathrm{d}r\\ &= \frac{1}{2}\delta_{kl}[J'_m(j_{mk})]^2\\ &= \frac{1}{2}\delta_{kl}[J_{m+1}(j_{mk})]^2\,.\end{aligned}$$

Remark. The Bessel functions $J_m(j_{mk}r)$ are orthogonal for each fixed m, but between different values k, l of the index labeling the roots.

8.6 Approximation via Eigenfunction Expansions

If we only keep the first N terms in an expansion and write

$$y(x) \approx \sum_{n=1}^{N} a_n y_n(x)\,,$$

it is not obvious how to determine a_n for this truncated series to best approximate $y(x)$. We define the error of such an approximation to be

$$\begin{aligned}E_N(a_1, a_2, \dots, a_N) &\coloneqq \left\| y(x) - \sum_{n=1}^{N} a_n y_n(x)\right\|_w^2\\ &= \left\langle y - \sum_{n=1}^{N} a_n y_n \middle| y - \sum_{m=1}^{N} a_m y_m \right\rangle\\ &= \|y\|_w^2 - \sum_{n=1}^{N} a_n^* \langle y_n|y\rangle_w - \sum_{m=1}^{N} a_m \langle y|y_m\rangle_w + \sum_{n,m=1}^{N} a_N^* a_m \langle y_n|y_m\rangle_w\\ &= \|y\|_w^2 - \sum_{n=1}^{N} [a_n^* \langle y_n|y\rangle_w + a_n \langle y|y_n\rangle_w] + \sum_{n=1}^{N} |a_n|^2\,.\end{aligned}$$

Let $a_n = \mathrm{Re}_n + i\,\mathrm{Im}_n$, then $a_n^* = \mathrm{Re}_n - i\,\mathrm{Im}_n$. Since $\langle y|y_n\rangle_w = \langle y_n|y\rangle_w^*$, we can write

$$\begin{aligned}E_N &= \|y\|_w^2 + \sum_{n=1}^{N} |a_n|^2 - \sum_{n=1}^{N} [\mathrm{Re}_n\,(\langle y_n|y\rangle_w + \langle y|y_n\rangle_w) + i\,\mathrm{Im}_n\,(\langle y|y_n\rangle_w - \langle y_n|y\rangle_w)]\\ &= \|y\|_w^2 + \sum_{n=1}^{N} |a_n|^2 - 2\sum_{n=1}^{N} [\mathrm{Re}_n\,\mathrm{Re}\,\langle y_n|y\rangle_w + \mathrm{Im}_n\,\mathrm{Im}\,\langle y_n|y\rangle_w]\,.\end{aligned}$$

Consider E_N as a function of Re_n and Im_n, taking derivatives gives

$$\begin{aligned}\frac{\partial E_N}{\partial\,\mathrm{Re}_n} &= 2\,\mathrm{Re}_n - 2\,\mathrm{Re}\,\langle y_n|y\rangle_w\,,\\ \frac{\partial E_N}{\partial\,\mathrm{Im}_n} &= 2\,\mathrm{Im}_n - 2\,\mathrm{Im}\,\langle y_n|y\rangle_w\,.\end{aligned}$$

Those derivatives are zero when the error is minimized, which gives us $a_n = \langle y_n|f\rangle_w$, just as the value in the untruncated expansion.

The minimum error is

$$\min E_N = \|y\|_w^2 - \sum_{n=1}^{N} \left[|a_n|^2 + |a_n|^2\right] + \sum_{n=1}^{N} |a_n|^2 = \|y\|_w^2 - \sum_{n=1}^{N} |a_n|^2\,.$$

Corollary. Because $E_N \geq 0$ from its definition, we deduce the *Bessel's Inequality*

$$\|y\|_w^2 \geq \sum_{n=1}^{N} |a_n|^2 .$$

Remark. This becomes an equality in the limit $N \to \infty$.

$$\|y\|_w^2 = \sum_{n=1}^{N} |a_n|^2 ,$$

where $y = \sum_{n=1}^{\infty} a_n y_n$. This is a generalisation of Parseval's theorem.

9 Calculus of Variations

9.1 Functionals

Definition 9.1. A real *function* of many variables $\{x_k \mid k = 1, 2, \ldots, n\}$ maps the ordered pair $(x_1, \ldots, x_n)$ to a real number $f : \mathbb{R}^n \to \mathbb{R}$,

$$f : (x_1, \ldots, x_n) \mapsto f(x_1, \ldots, x_n) \in \mathbb{R}\,.$$

We may take the number of variables to uncountably many: $y = \{y(x) \mid x \in \mathbb{R}\}$.

Definition 9.2. A real *functional* maps multiple functions of multiple variables to a real number.

$$G : (y_1(\mathbf{x}_1), \ldots, y_n(\mathbf{x}_n)) \mapsto G[(y_1(\mathbf{x}_1), \ldots, y_n(\mathbf{x}_n))] \in \mathbb{R}\,.$$

We shall usually be concerned with functionals of the form

$$G[y] = \int_\alpha^\beta f(y, y'; x)\,\mathrm{d}x\ . \tag{†}$$

9.2 Functional Derivatives

9.2.1 Functional Derivatives

Consider the effect of changing a function $y(x)$ to a nearby function $y(x) + \delta y(x)$.

Definition 9.3. The *variation* of a functional G is defined as

$$\delta G \coloneqq G[y + \delta y] - G[y]\,.$$

The variation of G of the above integral form (†) is

$$\begin{aligned}
\delta G &= \int_\alpha^\beta f(y + \delta y, y' + \delta y'; x)\,\mathrm{d}x - \int_\alpha^\beta f(y, y'; x)\,\mathrm{d}x \\
&= \int_\alpha^\beta f(y, y'; x) + \delta y \frac{\partial f}{\partial y} + (\delta y)' \frac{\partial f}{\partial y'} + \ldots \mathrm{d}x - \int_\alpha^\beta f(y, y'; x)\,\mathrm{d}x \\
&= \int_\alpha^\beta \delta y \frac{\partial f}{\partial y}\,\mathrm{d}x + \left[\delta y \frac{\partial f}{\partial y'}\right]_\alpha^\beta - \int_\alpha^\beta \delta y \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial y'}\right)\mathrm{d}x + \ldots
\end{aligned}$$

where the terms of $O(\delta y)^2$ are omitted.

Definition 9.4. The *functional derivative* of G with respect to a function y is defined as

$$\frac{\delta G}{\delta y(x)} \coloneqq \frac{\partial f}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial y'}\right).$$

Therefore, if y is fixed on the boundaries, then

$$\delta G = \int_\alpha^\beta \delta y \left[\frac{\partial f}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial y'}\right)\right] = \int_\alpha^\beta \delta y \frac{\delta G}{\delta y(x)}\,\mathrm{d}x\ .$$

Remark. Compare this with the variation of a function $f(\{y_i\})$:

$$\delta f = \sum_i \delta y_i \frac{\partial f}{\partial y_i}\,.$$

Theorem 9.5 (Euler–Lagrange equation). Let $y(x)$ be a real, smooth function with fixed values at $x = \alpha$ and $x = \beta$. The functional

$$G[y] = \int_\alpha^\beta f(y, y'; x)\,\mathrm{d}x \tag{\dagger}$$

is stationary if and only if

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial y'}\right) = \frac{\partial f}{\partial y}\,.$$

Proof. The functional is stationary when $\delta G = 0$ for any change δy, so $\frac{\delta G}{\delta y(x)} = 0$. □

Remark. $\frac{\partial f}{\partial y'}$ may look strange since it seems impossible to change y' while not changing y. Here $\frac{\partial}{\partial y}$ and $\frac{\partial}{\partial y'}$ are just formal derivatives and we can pretend that y and y' are not connected.

Corollary. This can be generalised to functionals $F[\mathbf{y}]$ for $\mathbf{y}(x) \in \mathbb{R}^n$:

$$\frac{\partial f}{\partial y_i} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial y_i'}\right) \text{ for each } i\,.$$

9.2.2 Geodesics of the Euclidean Plane

Definition 9.6. A *geodesic* is a curve representing in some sense the shortest path between two points in a surface (or more generally in a *Riemannian manifold*).

What is the geodesic between two points A and B on the Euclidean plane?

There are two ways to do this.

(i) We restrict to curves for which y can be made a function of x. The length of the curve is given by

$$L = \int_A^B \mathrm{d}l = \int_A^B \sqrt{\mathrm{d}x^2 + \mathrm{d}y^2} = \int_{x_A}^{x_B} \sqrt{1 + \left(\frac{\partial y}{\partial x}\right)^2}\,\mathrm{d}x\,.$$

This is a functional of $y(x)$:

$$L(y) = \int_{x_A}^{x_B} f(y')\,\mathrm{d}x\,,\ f(y') = \sqrt{1 + y'^2}\,.$$

The Euler–Lagrange equation is

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial y'}\right) = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{y'}{\sqrt{1 + y'^2}}\right) = \frac{\partial f}{\partial y} = 0\,.$$

$$\frac{y'}{\sqrt{1 + y'^2}} = \text{const.}\,,$$

so we must have

$$y' = \text{const.} \implies y = ax + b\,.$$

The geodesic on an Euclidean plane is a straight line.

(ii) This can be done more generally without the restriction by choosing a parameterisation $\mathbf{x} = (x(t), y(t))$ for $t \in [0, 1]$ such that $\mathbf{x}(0) = A$, $\mathbf{x}(1) = B$, so

$$L[x, y] = \int \mathrm{d}l = \int_0^1 \sqrt{\dot{x}^2 + \dot{y}^2}\,\mathrm{d}t\,.$$

We have

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0$$
$$\implies \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial f}{\partial \dot{x}}\right) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial f}{\partial \dot{y}}\right) = 0\,,$$

so we obtain the solutions

$$\frac{\dot{x}}{\sqrt{\dot{x}^2+\dot{y}^2}} = c\,,\ \frac{\dot{y}}{\sqrt{\dot{x}^2+\dot{y}^2}} = s\,,$$

where c, s are constants. They satisfy $c^2 + s^2 = 1$, so we can write $c = \cos\theta$, $s = \sin\theta$, and the conditions are equivalent to

$$(\dot{x}\sin\theta)^2 = (\dot{y}\cos\theta)^2\,.$$

Hence,

$$\dot{x}\sin\theta = \pm\dot{y}\cos\theta\,.$$

We can choose a θ such that we have a positive sign. So

$$y\cos\theta = x\sin\theta + A$$

for a constant A. This is a straight line with gradient $\tan\theta$.

9.2.3 First Integral Forms

Proposition 9.7. If $\frac{\partial f}{\partial y} = 0$, the Euler–Lagrange equation reduces to

$$\frac{\partial f}{\partial y'} = \text{const.}$$

known as a *first integral*.

Proof.

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial y'}\right) = \frac{\partial f}{\partial y} = 0\,.$$

□

Proposition 9.8. If $\frac{\partial f}{\partial x} = 0$, the Euler–Lagrange equation reduces to a *first integral*.

$$y'\frac{\partial f}{\partial y'} - f = \text{const.}\,.$$

Proof. From the chain rule, we have

$$\begin{aligned}\frac{\mathrm{d}f}{\mathrm{d}x} &= \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial x} + \frac{\partial f}{\partial y'}\frac{\partial y'}{\partial x}\\ &= \frac{\partial f}{\partial x} + y'\frac{\partial f}{\partial y} + y''\frac{\partial f}{\partial y'}\,.\end{aligned}$$

Euler–Lagrange equation gives

$$\begin{aligned}\frac{\mathrm{d}f}{\mathrm{d}x} &= \frac{\partial f}{\partial x} + y'\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial y'}\right) + y''\frac{\partial f}{\partial y'}\\ &= \frac{\partial f}{\partial x} + \frac{\mathrm{d}}{\mathrm{d}x}\left(y'\frac{\partial f}{\partial y'}\right),\end{aligned}$$

and hence

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(f - y'\frac{\partial f}{\partial y'}\right) = \frac{\partial f}{\partial x} = 0\,.$$

□

Remark. The first integral corresponds to a conserved quantity of the system.

9.2.4 The Brachistochrone

Definition 9.9. The *Brachistochrone* is the smooth curve joining two points A and B (not underneath one another) along which a particle will slide from A to B under gravity in the shortest possible time.

The particle starts with speed $v = 0$, and so from the conservation of energy we have

$$\frac{1}{2}mv^2 = mgy \implies v = \sqrt{2gy}\,.$$

We also have

$$v = \sqrt{\dot{x}^2 + \dot{y}^2} = \dot{x}\sqrt{1 + \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2} = \frac{\mathrm{d}x}{\mathrm{d}t}\sqrt{1 + y'^2}\,,$$

so

$$\mathrm{d}t = \frac{1}{v}\sqrt{1 + y'^2}\,\mathrm{d}x\ .$$

The total time is

$$\begin{aligned} T[y] = \int_A^B \mathrm{d}t &= \int_A^B \sqrt{\frac{1 + y'^2}{2gy}}\,\mathrm{d}x \\ &= \frac{1}{\sqrt{2g}}\int_A^B f(y, y')\,\mathrm{d}x\ , \text{ where } f(y, y') = \sqrt{\frac{1 + y'^2}{y}}\,. \end{aligned}$$

Here $\frac{\partial f}{\partial x} = 0$, so the first integral gives

$$\begin{aligned} f - y'\frac{\partial f}{\partial y'} &= \sqrt{\frac{1 + y'^2}{y}} - y'\frac{y'}{\sqrt{y(1 + y'^2)}} \\ &= \frac{1}{\sqrt{y(1 + y'^2)}} = \text{const.} \end{aligned}$$

Let

$$y(1 + y'^2) = 2c\,,$$

where c is a constant, then

$$y' = \sqrt{\frac{2c - y}{y}}\,.$$

By parameterising $y = c(1 - \cos\theta) = 2c\sin^2\left(\frac{\theta}{2}\right)$, we have

$$\begin{aligned} y' &= \sqrt{\frac{2c - 2c\sin^2\left(\frac{\theta}{2}\right)}{2c\sin^2\left(\frac{\theta}{2}\right)}} \\ &= \frac{\cos\frac{\theta}{2}}{\sin\frac{\theta}{2}}\,, \end{aligned}$$

and

$$y' = \frac{\mathrm{d}y}{\mathrm{d}\theta}\frac{\mathrm{d}\theta}{\mathrm{d}x} = 2c\sin\frac{\theta}{2}\cos\frac{\theta}{2}\frac{\mathrm{d}\theta}{\mathrm{d}x}\,,$$

so

$$\frac{\mathrm{d}x}{\mathrm{d}\theta} = 2c\sin^2\left(\frac{\theta}{2}\right) = c(1 - \cos\theta)\,.$$

Therefore, with $y(0) = 0$, the solution is given parametrically by

$$\begin{cases} x = c(\theta - \sin\theta) \\ y = c(1 - \cos\theta)\,. \end{cases}$$

This is an inverted *cycloid*, the curve traced by a point on the rim of a circular wheel as the wheel rolls along a straight line without slippage.

9.2.5 Geodesics on the Surface of a Sphere

In a spherical coordinate,

$$\mathrm{d}\mathbf{x} = \mathrm{d}r\,\mathbf{e}_r + r\,\mathrm{d}\theta\,\mathbf{e}_\theta + r\sin\theta\,\mathrm{d}\phi\,\mathbf{e}_\phi\,.$$

On the surface of a sphere, $\mathrm{d}r = 0$, so the length of a path from A to B on the surface of a sphere with radius r is

$$\begin{aligned} L &= \int_A^B |\mathrm{d}\mathbf{x}| \\ &= \int_A^B \sqrt{r^2\,\mathrm{d}\theta^2 + r^2\sin^2\theta\,\mathrm{d}\phi^2} \\ &= r\int_A^B \sqrt{1+\sin^2\theta\left(\frac{\mathrm{d}\phi}{\mathrm{d}\theta}\right)^2}\,\mathrm{d}\theta \\ &= r\int_{\theta_A}^{\theta_B} \sqrt{1+\sin^2\theta\,\phi'^2}\,\mathrm{d}\theta\,. \end{aligned}$$

The Euler–Lagrange equation gives, for $L[\phi]$,

$$f(\phi,\phi',\theta) = \sqrt{1+\sin^2\theta\,\phi'^2}\,,$$

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{\partial f}{\partial \phi}\right) = \frac{\partial f}{\partial \phi} = 0\,,$$

and so

$$\frac{\partial f}{\partial \phi'} = \frac{\sin^2\theta\,\phi'}{1+\sin^2\theta\,\phi'^2} = \text{const.} = c\,.$$

Rearrangement gives

$$\sin^4\theta\,\phi'^2 = c^2 + c^2\sin^2\theta\,\phi'^2\,,$$

$$\phi' = \frac{c}{\sin\theta\sqrt{\sin^2\theta - c^2}} = \frac{c}{\sin^2\theta\sqrt{1-c^2\csc^2\theta}}\,.$$

Therefore,

$$\begin{aligned} \phi &= \int \frac{c}{\sin^2\theta\sqrt{1-c^2\csc^2\theta}}\,\mathrm{d}\theta \\ &= \int -\frac{c}{\sin^2\theta\csc^2\theta\sqrt{1-c^2\csc^2\theta}}\,\mathrm{d}u && \text{substitute } u = \cot\theta \\ &= \int -\frac{c}{1-c^2(1+u^2)}\,\mathrm{d}u \\ &= \int -\frac{1}{\sqrt{\frac{1-c^2}{c^2}-u^2}}\,\mathrm{d}u \\ &= \int -\frac{1}{\sqrt{a^2-u^2}}\,\mathrm{d}u && \text{set } a^2 = \frac{1-c^2}{c^2} \\ &= a\arccos\left(\frac{u}{a}\right) + \phi_0\,. \end{aligned}$$

The path is therefore given by

$$\cot\theta = a\cos(\phi-\phi_0)\,,$$

or equivalently

$$\begin{cases} x = r\sin\theta\cos\phi \\ y = r\sin\theta\sin\phi \\ z = r\cos\theta\,. \end{cases}$$

Remark. This path is along a *great circle* passing through A and B. It is also the intersection of the sphere and a plane crossing the origin.

9.3 Variational Principles

9.3.1 Fermat's Principle

Theorem 9.10 (Fermat's principle). The path taken by a light ray from point A to point B in a material of variable *refractive index*, $\mu(\mathbf{x})$, makes the *optical path length*, P, stationary, where

$$P = \int_A^B \mu(\mathbf{x})\,\mathrm{d}l\ ,$$

and $\mathrm{d}l$ is the length element

$$\mathrm{d}l = \sqrt{\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2}\,.$$

Remark. Fermat's principle applies only in geometric optics approximations when the wavelength of light is small compared to the physical dimensions of the system. The approximation fails in occasions like diffraction.

Using the x-coordinate to parameterise position along the path and assuming there is no doubling back, the optical path length is a functional of $y(x)$ and $z(x)$.

$$P[y,z] = \int_A^B \mu(x,y,z)\sqrt{1+y'^2+z'^2}\,\mathrm{d}x = \int_A^B f(y,y',z,z';x)\,\mathrm{d}x\ .$$

Looking for stationary points of $P[y,z]$ with respect to variations of $y(x)$ and $z(x)$ gives

$$\delta P = \int_{x_A}^{x_B} \delta y \frac{\delta P}{\delta y}\,\mathrm{d}x + \int_{x_A}^{x_B} \delta z \frac{\delta P}{\delta z}\,\mathrm{d}x\ ,$$

where

$$\frac{\delta P}{\delta y} = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial y'}\right) - \frac{\partial f}{\partial y}\,,$$
$$\frac{\delta P}{\delta z} = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial z'}\right) - \frac{\partial f}{\partial z}\,.$$

Example. Path of light in a uniform medium.

If $\mu \equiv 1$, then

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial y'}\right) = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial f}{\partial z'}\right) = 0$$

$$\implies \quad \frac{y'}{\sqrt{1+y'^2+z'^2}} = c_1\,,\ \frac{z'}{\sqrt{1+y'^2+z'^2}} = c_2$$
$$\implies \quad y' = \text{const.}\,,\ z' = \text{const.}\,.$$

This implies that the path of light in a uniform medium is the intersection of two planes, i.e. a straight line.

Example. Snell's law.

We suppose that the path is in the xy plane and

$$\mu \equiv \mu(y) = \begin{cases} \mu_1\,, & y < 0 \\ \mu_2\,, & y \geq 0\,, \end{cases}$$

where μ_1 and μ_2 are constants. The integrand to make the integral stationary is

$$f(y,y') = \mu(y)\sqrt{1+y'^2}\,.$$

There is no explicit x dependence, so the first integral gives

$$\begin{aligned} f - y' \frac{\partial f}{\partial y'} &= \mu\sqrt{1+y'^2} - y' \frac{\mu y'}{1+y'^2} \\ &= \frac{\mu(y)}{\sqrt{1+y'^2}} = c\,, \end{aligned}$$

where c is a constant. For regions where μ is constant, this says that the path is straight. Let the angle the light paths make with the y axis be θ, then we have

$$y' = \cot\theta\,.$$

Therefore,

$$\begin{aligned} c = \frac{\mu}{\sqrt{1+y'^2}} &= \frac{\mu}{\sqrt{1+\cot^2\theta}} \\ &= \mu\sin\theta\,. \end{aligned}$$

Since the constant c is the same for the entire path, we can deduce Snell's law

$$\mu_1 \sin\theta_1 = \mu_2 \sin\theta_2\,.$$

Example. There is an analogous principle for sound waves where the acoustic path is stationary. This can be used to explain why distant sounds are better heard at night.

The speed of sound, v, depends on the absolute air temperature, T,

$$v \propto \sqrt{T}\,.$$

After sunset, the ground cools faster than the air, setting up a temperature gradient. Assume such variation is linear

$$T = T_0 + \alpha z\,,$$

where z is the height above the ground. This leads to a variational problem for

$$P[z] = \int_A^B \frac{\mathrm{d}l}{v} \propto \int_A^B \frac{\sqrt{1+z'^2}}{\sqrt{\alpha z + T_0}}\,\mathrm{d}x = \int_A^B f(z,z')\,\mathrm{d}x\ ,$$

where

$$f(z,z') = \sqrt{\frac{1+z'^2}{\alpha z + T_0}}\,.$$

This is now equivalent to the Brachistochrone problem.

9.3.2 Hamilton's Principle

Lagrangian and Hamiltonian mechanics reformulate Newtonian mechanics in terms of the *principle of least action*, based on energy rather than force. The time evolution of a system is viewed as the motion of a point in a multi-dimensional *configuration space* described by some *generalised coordinates* $\{q_i\}$.

Definition 9.11. The *Lagrangian* of a system, L, is defined as

$$L \coloneqq T - V\,,$$

where T is the kinetic energy and V is the potential energy.

Definition 9.12. The *action* of a path, starting at time t_i and ending at t_f, is given by

$$S[q_i] \coloneqq \int_{t_i}^{t_f} L(\{q_i\}, \{\dot{q}_i\}; t)\,\mathrm{d}t\ .$$

Theorem 9.13 (Hamilton's principle (Principle of the least action)). The motion in configuration space extremises the action functional S

Theorem 9.14 (Lagrange's equations). For $L(\{q_i\}, \{\dot{q}_i\}; t)$ with N generalised coordinates and fixed starting and ending points,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} = 0\,,\ i = 1, \ldots, N\,.$$

Proof. First proof that Lagrange's equations are equivalent to Newtonian mechanics when the generalised coordinates $\{q_i\}$ are set to be the Cartesian coordinates $\{x_i\}$.

From the definition of the Lagrangian, we have

$$\frac{\partial L}{\partial x_i} = -\frac{\partial V}{\partial x_i}\,,$$

since T has no direct dependence on x_i. while

$$\frac{\partial L}{\partial \dot{x}_i} = p_i\,,$$

so we have

$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = -\frac{\partial V}{\partial x_i}\,,$$

which implies that Lagrangian mechanics is indeed equivalent to Newtonian mechanics.

Next, we need to show that the form of the equation holds when shifting from Cartesian coordinates to any generalised coordinates.

Let

$$q_i = q_i(\ldots, x_i, \ldots, x_n, t)\,,$$

then by chain rule, we can write

$$\dot{q}_i = \frac{\mathrm{d}q_i}{\mathrm{d}t} = \frac{\partial q_i}{\partial x_j}\dot{x}_j + \frac{\partial q_i}{\partial t}\,.$$

To be a proper coordinate system, we should be able to invert the relationship so that $x_j = x_j(\ldots, q_i, \ldots, q_n, t)$, which we can do as long as we have a non-zero Jacobian. Then we have

$$\dot{x}_j = \frac{\partial x_j}{\partial q_i}\dot{q}_i + \frac{\partial x_j}{\partial t}\,.$$

Therefore,

$$\frac{\partial L}{\partial q_i} = \frac{\partial L}{\partial x_j}\frac{\partial x_j}{\partial q_i} + \frac{\partial L}{\partial \dot{x}_j}\left(\frac{\partial^2 x_j}{\partial q_i \partial q_k}\dot{q}_k + \frac{\partial^2 x_j}{\partial t \partial q_i}\right),$$

while

$$\frac{\partial L}{\partial \dot{q}_i} = \frac{\partial L}{\partial \dot{x}_j}\frac{\partial \dot{x}_j}{\partial \dot{q}_i}\,.$$

Now use the fact that

$$\frac{\partial \dot{x}_j}{\partial \dot{q}_i} = \frac{\partial x_j}{\partial q_i}\,,$$

we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_i}\right) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{x}_j}\right)\frac{\partial x_j}{\partial q_i} + \frac{\partial L}{\partial \dot{x}_j}\left(\frac{\partial^2 x_j}{\partial q_i \partial q_k}\dot{q}_k + \frac{\partial^2 x_j}{\partial q_i \partial t}\right),$$

so we have that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} = \left[\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{x}_j}\right) - \frac{\partial L}{\partial x_j}\right]\frac{\partial x_j}{\partial q_i}\,.$$

Therefore, once Lagrange's equation is solved in the $\{q_i\}$ system, it is also solved in the $\{x_i\}$ system. □

Corollary. Given that the Lagrangian has no explicit dependence on time, i.e. $\frac{\partial L}{\partial t} = 0$,

$$\sum_{i=1}^{N} \dot{q}_i \frac{\partial L}{\partial \dot{q}_i} - L = \text{const.}$$

Proof. The chain rule and Lagrange's equations give

$$\begin{aligned}\frac{\mathrm{d}L}{\mathrm{d}t} &= \frac{\partial L}{\partial t} + \sum_{i=1}^{N}\left(\dot{q}_i\frac{\partial L}{\partial q_i} + \ddot{q}_i\frac{\partial L}{\partial \dot{q}_i}\right)\\ &= \frac{\partial L}{\partial t} + \sum_{i=1}^{N}\left(\dot{q}_i\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_i}\right) + \ddot{q}_i\frac{\partial L}{\partial \dot{q}_i}\right),\end{aligned}$$

and hence,

$$\begin{aligned}\frac{\mathrm{d}L}{\mathrm{d}t} &= \frac{\partial L}{\partial t} + \frac{\mathrm{d}}{\mathrm{d}t}\sum_{i=1}^{N}\dot{q}_i\frac{\partial L}{\partial \dot{q}_i}\,,\\ \frac{\partial L}{\partial t} &= \frac{\mathrm{d}}{\mathrm{d}t}\left(L - \sum_{i=1}^{N}\dot{q}_i\frac{\partial L}{\partial \dot{q}_i}\right).\end{aligned}$$

□

Theorem 9.15 (Energy conservation). The conserved quantity

$$\sum_{i=1}^{N} \dot{q}_i \frac{\partial L}{\partial \dot{q}_i} - L$$

is equivalent to energy if the generalised coordinates are *natural coordinates* that have no explicit time dependence: $\mathbf{r} = \mathbf{r}(q_1, \ldots, q_n)$.

Proof. Under such assumption, T is a homogeneous quadratic in the generalised velocities $\{\dot{q}_i\}$, i.e.

$$T = \frac{1}{2}mv^2 = \frac{m}{2}\sum \frac{\partial \mathbf{x}}{\partial q_i}\dot{q}_i\frac{\partial \mathbf{x}}{\partial q_j}\dot{q}_j = \sum_{i,j} a_{ij}(q_1, \ldots, q_N)\dot{q}_i\dot{q}_j\,,$$

Then the conserved quantity:

$$\begin{aligned}\sum_{i=1}^{N}\dot{q}_i\frac{\partial L}{\partial \dot{q}_i} - L &= \sum_{u=1}^{N}\dot{q}_u\frac{\partial}{\partial \dot{q}_u}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}a_{ij}\dot{q}_i\dot{q}_j - V\right) - \sum_{i=1}^{N}\sum_{j=1}^{N}a_{ij}\dot{q}_i\dot{q}_j + V\\ &= \sum_{i=1}^{N}\sum_{j=1}^{N}a_{ij}\dot{q}_i\dot{q}_j + V = T + V = E\,,\end{aligned}$$

i.e. The total energy $E = T + V$ is conserved. □

Remark. This is an example of *Noether's theorem*. The uniformity of time corresponds to energy conservation.

Theorem 9.16 (Noether's theorem). Every differentiable symmetry of the action of a physical system with conservative forces has a corresponding conservation law.

- Time invariance $\implies$ conservation of energy;
- Translational invariance $\implies$ conservation of momentum;
- Rotational invariance $\implies$ conservation of angular momentum.

Example. Consider a particle of mass m subjected to a conservative force field

$$\mathbf{F}(\mathbf{x}) = -\boldsymbol{\nabla} V(\mathbf{x})\,.$$

We have

$$\begin{aligned} L &= \frac{1}{2} m |\dot{\mathbf{x}}|^2 - V(\mathbf{x}) \\ &= \frac{1}{2} m \sum_{i=1}^{3} \dot{x_i}^2 - V(x_1, x_2, x_3)\,. \end{aligned}$$

$$\begin{aligned} \frac{\partial L}{\partial x_i} &= -\frac{\partial V}{\partial x_i} \\ \frac{\partial L}{\partial \dot{x_i}} &= m\dot{x_i}\,. \end{aligned}$$

The Euler–Lagrange equations give

$$m\ddot{x_i} = -\frac{\partial V}{\partial x_i} \implies m\ddot{\mathbf{x}} = -\boldsymbol{\nabla} V = \mathbf{F}\,.$$

This is Newton's second law. We also found that, as expected,

$$\begin{aligned} \sum_{i=1}^{3} \dot{x_i} \frac{\partial L}{\partial \dot{x_i}} - L &= m \sum_{i=1}^{3} \dot{x_i}^2 - \frac{1}{2} m \sum_{i=1}^{3} \dot{x_i}^2 + V \\ &= \frac{1}{2} m |\dot{\mathbf{x}}|^2 + V = E = \text{const.}\,. \end{aligned}$$

Example. Suppose that the central force field depends only on $r = |\mathbf{x}|$. For a planar motion, work in polar coordinate, we have

$$\begin{aligned} L &= \frac{1}{2} m (v_r^2 + v_\varphi^2) - V(r) \\ &= \frac{1}{2} m \dot{r}^2 + \frac{1}{2} m r^2 \dot{\varphi}^2 - V(r)\,. \end{aligned}$$

The Euler–Lagrange equations are

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{r}}\right) = \frac{\partial L}{\partial r} &\implies \frac{\mathrm{d}}{\mathrm{d}t}(m\dot{r}) = m\ddot{r} = mr\dot{\varphi}^2 - \frac{\partial V}{\partial r} \\ \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{\varphi}}\right) = \frac{\partial L}{\partial \varphi} &\implies \frac{\mathrm{d}}{\mathrm{d}t}\big(mr^2\dot{\varphi}\big) = 0\,. \end{aligned}$$

The second equation gives $mr^2\dot{\boldsymbol{\varphi}} = \mathbf{J} = \text{const.}$, where $\mathbf{J}$ is the angular momentum. Define specific angular momentum, h, such that

$$h = \frac{J}{m} = r^2\dot{\varphi}\,.$$

The first equation reduces to

$$\begin{aligned} m\ddot{r} &= -\frac{\partial V}{\partial r} + \frac{mh^2}{r^3} \\ &= -\frac{\partial}{\partial r}\left(V + \frac{mh^2}{2r^2}\right) = -\frac{\partial V_{\text{eff}}}{\partial r}\,. \end{aligned}$$

For $h \neq 0$, the effective potential has a centrifugal barrier. For example, if the gravitational potential

$$V = -\frac{GMm}{r}\,,$$

then we have

$$V_{\text{eff}}(r) = m\left(-\frac{GM}{r} + \frac{h^2}{2r^2}\right).$$

When $h \neq 0$, the centrifugal barrier prevents an approach to $r = 0$. There will be three types of stable orbits depending on the V_{eff} of the particle.

- $V_{\text{eff}} < 0$. Elliptical orbits;
- $V_{\text{eff}} = 0$. Parabolic orbits;
- $V_{\text{eff}} > 0$. Hyperbolic orbits.

Example. Consider two particles of masses m_1 and m_2, interacting via a potential $V(\mathbf{x}_1 - \mathbf{x}_2)$. A point in configuration space can be specified by two position vectors $\mathbf{x}_1$, $\mathbf{x}_2$, but we can also use the centre of mass $\mathbf{R}$ and relative position $\mathbf{r}$.

$$\mathbf{R} = \frac{m_1\mathbf{x}_1 + m_2\mathbf{x}_2}{m_1 + m_2}\,,$$

$$\mathbf{r} = \mathbf{x}_1 - \mathbf{x}_2\,.$$

Then we have

$$\mathbf{x}_1 = \mathbf{R} + \frac{m_2}{m_1 + m_2}\mathbf{r}\,,$$

$$\mathbf{x}_2 = \mathbf{R} + \frac{m_1}{m_1 + m_2}\mathbf{r}\,.$$

The kinetic energy is

$$\begin{aligned} T &= \frac{1}{2}m_1|\dot{\mathbf{x}}_1|^2 + \frac{1}{2}m_2|\dot{\mathbf{x}}_2|^2 \\ &= \frac{1}{2}m_1\left[\dot{R}^2 + 2\frac{m_2}{m_1+m_2}\dot{\mathbf{r}}\cdot\dot{\mathbf{R}} + \left(\frac{m_2}{m_1+m_2}\right)^2\dot{r}^2\right] \\ &\quad + \frac{1}{2}m_2\left[\dot{R}^2 - 2\frac{m_1}{m_1+m_2}\dot{\mathbf{r}}\cdot\dot{\mathbf{R}} + \left(\frac{m_1}{m_1+m_2}\right)^2\dot{r}^2\right] \\ &= \frac{1}{2}(m_1+m_2)\dot{R}^2 + \frac{1}{2}\frac{m_1m_2}{m_1+m_2}\dot{r}^2\,. \end{aligned}$$

Let

$$M = m_1 + m_2\,,\ \mu = \frac{m_1m_2}{m_1+m_2}\,,$$

then

$$T = \frac{1}{2}M\dot{R}^2 + \frac{1}{2}\mu\dot{r}^2\,.$$

The Lagrangian equation for $\mathbf{R}$ gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial\dot{\mathbf{R}}}\right) = \frac{\partial L}{\partial\mathbf{R}} = 0\,,$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(M\dot{\mathbf{R}}\right) = 0\,,$$

$$\dot{\mathbf{R}} = \text{const.}\,,$$

i.e. the centre of mass moves with constant velocity. The Lagrangian equation for $\mathbf{r}$ gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{\mathbf{r}}}\right) = \frac{\partial L}{\partial \mathbf{r}}\,,$$
$$\frac{\mathrm{d}}{\mathrm{d}t}(\mu\dot{\mathbf{r}}) = -\boldsymbol{\nabla} V\,,$$
$$\mu\ddot{\mathbf{r}} = -\boldsymbol{\nabla} V(\mathbf{r})\,.$$

Because T is a homogeneous quadratic in the generalised velocities, V does not depend on velocities and contains no explicit t-dependence, $E = T + V$ is constant.

9.4 Constrained Variation and Lagrange Multipliers

Lemma 9.17. For any differentiable $f : \mathbb{R}^n \to \mathbb{R}$,

$$\mathrm{d}f = \boldsymbol{\nabla} f \cdot \mathrm{d}x\ .$$

Proof. Taylor's theorem states that

$$\begin{aligned}\delta f &= f(\mathbf{x} + \delta\mathbf{x}) - f(\mathbf{x}) \\ &= \frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial y}\delta y + \frac{\partial f}{\partial z}\delta z + \dots \\ &= \boldsymbol{\nabla} f \cdot \delta\mathbf{x}\end{aligned}$$

in the limit $|\delta x| \to 0$. □

Consider extremising $f(x, y)$ along a path specified by $p(x, y) = 0$. We will require

$$\mathrm{d}f = \mathrm{d}l \cdot \boldsymbol{\nabla} f = 0$$

for $\mathrm{d}l$ along the path. Such $\mathrm{d}l$ along the path would naturally require

$$\mathrm{d}p = 0 \implies \boldsymbol{\nabla} p \cdot \mathrm{d}l = 0\,.$$

This is a constraint on $\mathrm{d}l$. At the extremum point on the path, $\boldsymbol{\nabla} f$ will be orthogonal to all $\mathrm{d}l$ that are orthogonal to $\boldsymbol{\nabla} p$. Therefore $\boldsymbol{\nabla} f$ and $\boldsymbol{\nabla} p$ are parallel or anti-parallel. For some λ,

$$\begin{cases}\boldsymbol{\nabla} f - \lambda \boldsymbol{\nabla} p = 0 \\ p(x, y) = 0\,.\end{cases}$$

These equations arise from extremisation without constraint of a function of three variables

$$\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda p(x, y)\,.$$

Variation with respect to the *Lagrange multiplier*, λ, gives the constraint $p = 0$. Variation with respect to x, y gives the other equations.

Remark. By introducing the Lagrange multiplier, we turned a constrained variation problem into an unconstrained variation problem.

We can extend this method to higher numbers of variables and constraints.

Theorem 9.18 (Lagrange multiplier). To find the stationary point of a function $f(\{\xi\})$ of n variables $(\xi_1, \dots, \xi_n)$ under m constraints $p_i(\{\xi\}) = 0$ $(i = 1, \dots, m)$, we need to extremise, with respect to $n + m$ variables,

$$\mathcal{L}(\{\xi\}; \{\lambda\}) = f(\{\xi\}) - \sum_{i=1}^{m} \lambda_i p_i(\{\xi\})\,.$$

The generalisation to functionals ($n \to \infty$) is straightforward.

Corollary. To maximize $G[y]$ subject to the constraint $P[y] = 0$, we may generalise without constraint

$$\mathcal{L}[y] = G[y] - \lambda P[y]$$

with respect to the function y and the variable λ.

9.4.1 Catenary

Definition 9.19. A *catenary* is the curve that an idealized hanging chain assumes under its own weight when supported only at its ends in a uniform gravitational field.

Let the two fixed points be at $x = \pm L$ with $y > 0$. The chain has a fixed length $l_0 > 2L$ and a constant mass per unit length ρ. The potential energy of an element $\mathrm{d}l$ is $\mathrm{d}V = \rho g y \,\mathrm{d}l$, where $y(x)$ is the height of the chain above the ground. Therefore, the potential energy is

$$\begin{aligned} V &\propto \int_{\text{chain}} y \,\mathrm{d}l \\ &= \int_{-L}^{L} y\sqrt{1+y'^2}\,\mathrm{d}x = G[y]\,. \end{aligned}$$

We must minimize V subject to the constraint

$$l_0 = \int_{\text{chain}} \mathrm{d}l = \int_{-L}^{L} \sqrt{1+y'^2}\,\mathrm{d}x\ ,$$

$$\implies P[y] = \int_{-L}^{L} \sqrt{1+y'^2}\,\mathrm{d}x - l_0 = 0\,.$$

This is equivalent to extremising without constraint

$$\begin{aligned} \mathcal{L}[y] &= G[y] - \lambda P[y] \\ &= \int_{-L}^{L} (y-\lambda)\sqrt{1+y'^2}\,\mathrm{d}x + \lambda l_0 \\ &= \int_{-L}^{L} f(y, y'; \lambda)\,\mathrm{d}x + \lambda l_0\,, \end{aligned}$$

where λ is the Lagrange multiplier. Because there is no explicit dependence on x in the integrand, the first integral gives

$$\begin{aligned} \text{const.} &= y'\frac{\partial f}{\partial y'} - f \\ &= y'\frac{(y-\lambda)y'}{\sqrt{1+y'^2}} - (y-\lambda)\sqrt{1+y'^2} \\ &= (y-\lambda)\left(\frac{y'^2}{\sqrt{1+y'^2}} - \sqrt{1+y'^2}\right) \\ &= \frac{\lambda - y}{\sqrt{1+y'^2}} = c\,. \end{aligned}$$

$$\implies y' = \frac{1}{c}\sqrt{(y-\lambda)^2 - c^2}\,.$$

This gives us the solution

$$y(x) = \lambda + c\cosh\frac{x+a}{c}\,.$$

For simplicity, suppose both ends are at height $h = \lambda + c\cosh\frac{L}{c}$ above the ground such that $a = 0$. We have

$$y(x) = c\cosh\frac{x}{c}\,,$$
$$l_0 = 2c\sinh\frac{L}{c}\,.$$

9.4.2 Isoperimetric Problem

Find the simple closed plane curve $\mathcal{C}$ of fixed length L in a plane that maximizes the enclosed area A. This implies that the curve does not intersect itself and the area it encloses is simply connected.

It is obvious that the inside region must be convex to maximise the area enclosed, otherwise, a curve with a larger area can easily be constructed. Assume that the curve has x values in the range $[x_1, x_2]$. Then, the curve can be divided into an 'upper' part and a 'lower' part, given by $y_1(x)$ and $y_2(x)$ for $x \in [x_1, x_2]$. The area of the curve is

$$A = \int_{x_1}^{x_2} y_2(x) - y_1(x)\,\mathrm{d}x = \oint_{\mathcal{C}} y(x) = A[y]\,.$$

We want to maximize A subject to the constraint

$$\begin{aligned} L &= \oint_{\mathcal{C}} \mathrm{d}l \\ &= \oint_{\mathcal{C}} \sqrt{1+y'^2}\,\mathrm{d}x\ , \end{aligned}$$
$$P[y] = \oint_{\mathcal{C}} \sqrt{1+y'^2}\,\mathrm{d}x - L = 0\,.$$

Therefore we have to extremise without constraint

$$\begin{aligned} \mathcal{L}[y] &= \oint_{\mathcal{C}} y - \lambda\sqrt{1+y'^2}\,\mathrm{d}x + \lambda L \\ &= \oint_{\mathcal{C}} f(y, y'; \lambda)\,\mathrm{d}x + \lambda L \end{aligned}$$

with respect to the function y and the real variable λ.

$f(y, y'; \lambda)$ has no explicit dependence on x, so the first integral gives

$$\begin{aligned} \text{const.} &= y'\frac{\partial f}{\partial y'} - f \\ &= y'\frac{-\lambda y'}{\sqrt{1+y'^2}} - y + \lambda\sqrt{1+y'^2} \\ &= \frac{\lambda}{\sqrt{1+y'^2}} - y = -y_0\,, \end{aligned}$$
$$\frac{\lambda}{\sqrt{1+y'^2}} = y - y_0\,.$$

This is equivalent to

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\sqrt{\lambda^2 - (y-y_0)^2}}{y-y_0}$$
$$\implies x - x_0 = \int \frac{y-y_0}{\sqrt{\lambda^2-(y-y_0)^2}}$$

for some constant y_0. This ODE has solution

$$y = y_0 \pm \sqrt{\lambda^2 - (x - x_0)^2}$$

for some constant x_0, so

$$(x - x_0)^2 + (y - y_0)^2 = \lambda^2 \,.$$

This is a circle of radius λ. Varying Φ with respect to λ gives the original constant $2\pi\lambda = L$.

9.4.3 Sturm–Liouville Theory

Consider a Sturm–Liouville operator $\mathcal{L}$ and the following real functionals of the real function y

$$\begin{aligned} F[y] = \langle y|\mathcal{L}y\rangle &= \int_\alpha^\beta \left\{-y\frac{\mathrm{d}}{\mathrm{d}x}[\rho(x)y'] + \sigma(x)y^2\right\}\mathrm{d}x \\ &= \int_\alpha^\beta \{\rho(x)(y')^2 + \sigma(x)y^2\}\,\mathrm{d}x \ , \end{aligned}$$

$$G[y] = \langle y|y\rangle_w = \int_\alpha^\beta w(x)y^2\,\mathrm{d}x \ ,$$

where $\rho(x)$, $w(x) > 0$ for $\alpha < x < \beta$. Assume that the boundary conditions lead to vanishing boundary terms, we have the functional derivatives

$$\begin{aligned} \frac{\delta F}{\delta y} &= 2\mathcal{L}y \\ \frac{\delta G}{\delta y} &= 2wy \,. \end{aligned}$$

Consider the formulation of the Sturm–Liouville problem as an extremisation of F subjected to the constraint $G = 1$. This is equivalent to extremising without constraint

$$\Phi[y] = F[y] - \lambda(G[y] - 1) = (F[y] - \lambda G[y]) + \lambda$$

with respect to the function y and the real variable λ. Extremising Φ, we obtain

$$\begin{aligned} \frac{\delta\Phi}{\delta y} &= \frac{\delta F}{\delta y} - \lambda\frac{\delta G}{\delta y} \\ &= 2\mathcal{L}y - 2\lambda wy = 0 \\ &\mathit{implies} \mathcal{L}y = \lambda wy \,, \end{aligned}$$

i.e. the eigenvalue equation, where now the Lagrange multiplier λ is the eigenvalue.

We can view this problem in a different way. Notice that $\mathcal{L}y = \lambda wy$ is linear in y. Hence if y is a solution, then so is ay. But if $G[y] = 1$, then $G[ay] = a^2$. Hence the condition $G[y] = 1$ is simply a normalisation condition. We can get around this problem by asking for the minimum of the functional

$$\Lambda[y] = \frac{F[y]}{G[y]} \,.$$

$$\begin{aligned} \delta\Lambda[y] = \frac{F[y+\delta y]}{G[y+\delta y]} - \frac{F[y]}{G[y]} &= \frac{F[y] + \delta F}{G[y] + \delta G} - \frac{F[y]}{G[y]} \\ &= \frac{1}{G}\left[(F + \delta F)\left(1 - \frac{\delta G}{G} + \dots\right)\right] - \frac{F}{G} && \text{expansion} \\ &= \frac{F - \left(\frac{F}{G}\right)\delta G + \delta F}{G} - \frac{F}{G} && \text{only keep first order terms} \\ &= \frac{1}{G}\left(\delta F - \frac{F}{G}\delta G\right) . \end{aligned}$$

When Λ is minimised, we have

$$\delta\Lambda = 0 \quad \iff \quad \frac{\delta F}{\delta y} = \Lambda \frac{\delta G}{\delta y} \quad \iff \quad \mathcal{L}y = \Lambda w y\,.$$

So at stationary values of $\Lambda[y]$, Λ is the associated Sturm–Liouville eigenvalue.

9.5 Rayleigh–Ritz Method

The eigenvalues of a Sturm–Liouville problem are the extremal values of $\Lambda = \frac{F}{G}$, where

$$F[y] = \int_\alpha^\beta \left[\rho(x)y'^2 + \sigma(x)y^2\right] \mathrm{d}x = \langle y|\mathcal{L}y\rangle$$
$$G[y] = \int_\alpha^\beta w(x)y^2 \,\mathrm{d}x = \langle y|y\rangle_w\,.$$

Suppose $\rho > 0$ and $\sigma \geq 0$ so that $F \geq 0$, so that $\Lambda \geq 0$. Let one of the extremal values of Λ, λ_0 be an *absolute minimum*.

$$\lambda_0 = \Lambda[y_0]\,,$$

where y_0 is the eigenfunction corresponding to λ_0. For simplicity, we assume that there is no degeneracy.

Proposition 9.20. We have the inequality

$$\lambda_0 \leq \Lambda[y] \quad \text{for all } y(x)\,,$$

with equality if and only if $y = y_0$.

Proof. Let

$$y = \sum_{n=0}^\infty a_n y_n\,,$$

$$\Lambda[y] = \frac{\langle y|\mathcal{L}y\rangle}{\langle y|y\rangle_w} = \frac{\left\langle \sum_{n=0}^\infty a_n y_n \middle| \mathcal{L}\sum_{m=0}^\infty a_m y_m \right\rangle}{\left\langle \sum_{n=0}^\infty a_n y_n \middle| \sum_{m=0}^\infty a_m y_m \right\rangle}$$
$$= \frac{\sum_{n=0}^\infty \sum_{m=0}^\infty a_n^* a_m \langle y_n|\mathcal{L}y_m\rangle}{\sum_{n=0}^\infty \sum_{m=0}^\infty a_n^* a_m \langle y_n|y_m\rangle_w}\,.$$

Since

$$\langle y_n|y_m\rangle_w = \delta_{nm}\,,$$
$$\mathcal{L}y_m = \lambda_m w y_m\,,$$

we have

$$\langle y_n|\mathcal{L}y_m\rangle = \langle y_n|\lambda_m w y_n\rangle$$
$$= \lambda_m \langle y_n|y_m\rangle_w = \lambda_m \delta_{nm}\,.$$

Therefore,

$$\Lambda[y] = \frac{\sum_{n=0}^\infty |a_n|^2 \lambda_n}{\sum_{n=0}^\infty |a_n|^2} \geq \frac{\sum_{n=0}^\infty |a_n|^2 \lambda_0}{\sum_{n=0}^\infty |a_n|^2} = \lambda_0\,.$$

□

We can first make a guess, y_{trial}, for y_0 and evaluate $\Lambda[y_{\text{trial}}]$

$$\lambda_0 \leq \Lambda[y_{\text{trial}}]\,.$$

The better y_{trial} is, the closer $\Lambda[y_{\text{trial}}]$ will be to λ_0. Because $\Lambda[y]$ is stationary at $y = y_0$, a moderately good guess would give us a reasonable approximation to λ_0.

Theorem 9.21 (Rayleigh–Ritz method). Choose y_{trial} to depend on one or more parameters $(a_1, a_2, \dots)$. Since $\lambda_0 \leq \Lambda[y(\{a_i\})] \equiv \Lambda(\{a_i\})$ for all choices of parameters, we can get the lowest upper bound by minimizing $\Lambda(\{a_i\})$ with respect to $\{a_i\}$,

$$\lambda_0 \leq \min_{\{a_i\}} \Lambda(\{a_i\})\,.$$

Example. Quantum harmonic oscillator.

Consider solutions to

$$\mathcal{L}\psi = \lambda\psi = 2E\psi\,,$$
$$\mathcal{L} \equiv -\frac{\mathrm{d}^2}{\mathrm{d}x^2} + x^2\,,$$

subjected to boundary conditions that $\psi \to 0$ as $|x| \to \infty$, which makes the Sturm–Liouville operator $\mathcal{L}$ self-adjoint. The solution of this equation maximizes $\Lambda[\psi] = \frac{F[\psi]}{G[\psi]}$, where

$$F[\psi] = \int_{-\infty}^{\infty} \left[(\psi')^2 + x^2\psi^2\right] \mathrm{d}x$$
$$G[\psi] = \int_{-\infty}^{\infty} \psi^2 \,\mathrm{d}x\ .$$

For a suitable choice of unit, this is the Schrödinger equation for a particle of energy E in a harmonic oscillator potential.

Since $\psi(x) \to 0$ as $|x| \to \infty$, we take the trial solution as

$$\psi_{\text{trial}}(x) = \exp\Big(-\frac{\alpha}{2}x^2\Big)\,,$$

so

$$\psi'_{\text{trial}} = -\alpha x \exp\Big(-\frac{\alpha}{2}x^2\Big)\,,$$

with parameter α. Using the result that

$$\int_{-\infty}^{\infty} x^{2n} e^{-\alpha x^2} = \frac{(2n)!}{2^{2n}n!}\sqrt{\frac{\pi}{\alpha^{2n+1}}}\,,$$

we find

$$F[\psi_{\text{trial}}] = (\alpha^2 + 1)\int_{-\infty}^{\infty} x^2 e^{-\alpha x^2}\,,$$
$$G[\psi_{\text{trial}}] = \int_{-\infty}^{\infty} e^{-\alpha x^2}\,\mathrm{d}x\ ,$$
$$\implies \Lambda[\psi_{\text{trial}}] = \frac{\alpha^2 + 1}{2\alpha}\,.$$

This is a minimum when $\alpha = 1$, where

$$\min_{\alpha} \Lambda[\psi_{\text{trial}}(x;\alpha)] = 1$$

for

$$\psi_{\text{trial}}(x) = e^{-\frac{x^2}{2}}\,.$$

We deduce that $\lambda_0 \leq 1$.

Remark. $\lambda_0 = 1$ is actually the exact answer, which happens when the exact eigenfunction corresponding to the lowest eigenvalue happens to be of the form of the trial solution.

For some less-inspired guesses, the answer would not be exact. For example, let the trial solution be

$$\psi_{\text{trial}}(x) = (1 + ax^2)e^{-x^2},$$

which gives an upper bound of $\lambda_0 \leq \Lambda(a_{\min}) \approx 1.03$. The accuracy of this upper bound can be improved by adding more parameters to our trial solution. An improved trial solution might be

$$\psi_{\text{trial}}(x) = (1 + ax^2 + bx^4)e^{-x^2}.$$

Example. Circularly symmetric vibration of a circular drum.

Consider a drum with a unit radius fixed at $r = 1$. The amplitude $y(r)$ of small-amplitude vibrations satisfies *Bessel's equation*

$$y'' + \frac{1}{r}y' + \lambda y = 0,$$

subject to $y(1) = 0$ and finite $y(0)$. λ goes like the square of the angular frequency, and the dominant effect is from the lowest frequency, so we want to estimate λ_0.

To put the differential operator into a Sturm–Liouville form, we need to multiply $w(r) = r$.

$$F[y] = \int_0^1 r(y')^2 \,\mathrm{d}r$$
$$G[y] = \int_0^1 ry^2 \,\mathrm{d}r.$$

Try

$$y_{\text{trial}} = a + br^2 + cr^4,$$

where $a + b + c = 0$ to satisfy $r(1) = 0$. We include only even powers of r because the equation has the same form when $r \to -r$. Using $a = -b - c$, we get

$$F[y_{\text{trial}}] = f(b, c) = b^2 + \frac{8}{3}bc + 2c^2$$
$$G[y_{\text{trial}}] = g(b, c) = \frac{1}{6}b^2 + \frac{5}{12}bc + \frac{4}{15}c^2.$$

This gives us our approximation to λ_0: $\lambda_0 \approx 5.784$. This is close to the true value of $\lambda_0 = 5.7832\ldots$.

9.5.1 Extension to Higher Eigenvalues

Suppose we already have a good approximation to λ_0 and y_0, and we want an approximation to the next-lowest eigenvalue λ_1. The orthogonality suggests that we should consider a trial function $y^{(1)}_{\text{trial}}$ orthogonal to y_0:

$$y^{(1)}_{\text{trial}} = \sum_{n=1}^{\infty} a_n y_n(x),$$
$$\left\langle y_0 \middle| y^{(1)}_{\text{trial}} \right\rangle_w = 0$$

for some coefficients a_n. The $a_0 y_0$ term is missing because of the required orthogonality. We do not know what the functions y_n are, just that they are a complete set of eigenfunctions. Now,

$$\Lambda\left[y^{(1)}_{\text{trial}}\right] = \frac{\sum_{n=1}^{\infty} |a_n|^2 \lambda_n}{\sum_{n=1}^{\infty} |a_n|^2}.$$

Since $|a_n|^2 \geq 0$ and $\lambda_1 \leq \lambda_n$ for $n = 1, 2, \ldots$, we have

$$\sum_{n=1}^{\infty} |a_n|^2 \lambda_n \geq \lambda_1 \sum_{n=1}^{\infty} |a_n|^2 ,$$

and hence $\Lambda\left[y^{(1)}_{\text{trial}}\right] \geq \lambda_1$.

Therefore, we will have an upper bound on λ_1 from any trial function orthogonal to y_0, and a reasonably good guess will give a good estimate for λ_1.

An obvious problem is: how do we find a trial function orthogonal to y_0 if we only have an approximation to y_0? In general we cannot, but it can be done in some exceptional cases. For example, a theorem in quantum mechanics states that the ground-state wavefunction of a particle in a symmetric potential, $V(x) = V(-x)$, is a symmetric function. Since any anti-symmetric function is orthogonal to any symmetric function, we can find a bound on λ_1 by choosing any anti-symmetric trial function.

10 Partial Differential Equations and Separation of Variables

10.1 Nomenclature

Definition 10.1. *Partial differential equations* are equations relating one or more unknown functions (dependent variables) of two or more independent variables with one or more of the functions' partial derivatives with respect to those variables.

$$F\left(\psi, \frac{\partial\psi}{\partial x}, \frac{\partial\psi}{\partial y}, \frac{\partial^2\psi}{\partial x^2}, \frac{\partial^2\psi}{\partial x\partial y}, \frac{\partial^2\psi}{\partial y^2}, \ldots, x, y\right) = 0$$

The *order* of the PDE is the order of the highest derivative in the equation.

If the system of differential equations is of the first degree in the dependent variables and all its derivatives, then the system is said to be *linear*.

10.1.1 Linear Second-order Partial Differential Equations

The most general linear second-order PDE in two variables is

$$\mathcal{L}\psi(x,y) = g(x,y)\,,$$

where $\mathcal{L}$ is a differential operator such that

$$\mathcal{L}\psi \equiv a(x,y)\frac{\partial^2\psi}{\partial x^2} + b(x,y)\frac{\partial^2\psi}{\partial x\partial y} + c(x,y)\frac{\partial^2\psi}{\partial y^2} + d(x,y)\frac{\partial\psi}{\partial x} + e(x,y)\frac{\partial\psi}{\partial y} + f(x,y)\psi\,.$$

Remarks.

- If $g = 0$, then the equation is said to be homogeneous.
- We will focus on examples where the coefficients are independent of x and y i.e. constant coefficients.
- These ideas can be generalised to more than two independent variables or systems of PDEs with more than one dependent variable.

Definition 10.2. The linear second-order PDEs are often classified as:

- Elliptic: $b^2 - 4ac < 0$;
- Parabolic: $b^2 - 4ac = 0$;
- Hyperbolic: $b^2 - 4ac > 0$.

Remark. In hyperbolic PDEs (e.g. wave equations), the smoothness of the solution depends on the smoothness of the initial and boundary conditions. In the worst case, the solution even may be differentiable nowhere. In a system modeled with a hyperbolic PDE, information travels at a finite speed referred to as the wave speed.

In contrast, the solutions of elliptic PDEs (e.g. Laplace's equation) are always smooth, even if the initial and boundary conditions are rough. This is known as *elliptic regularity*. In addition, boundary data at any point affect the solution at all points in the domain.

Parabolic PDEs are usually time-dependent and represent diffusion-like processes. Solutions are smooth in space but may possess singularities. However, information travels at infinite speed in a parabolic system.

10.1.2 Boundary Conditions

Definition 10.3. The boundary conditions may be classified as the following:

- Dirichlet condition:
$$\psi = g(\mathbf{x})$$
where $g(\mathbf{x})$ is a known function.
- Neumann condition:
$$\frac{\partial \psi}{\partial n} \equiv \hat{\mathbf{n}} \cdot \boldsymbol{\nabla} \psi = h(\mathbf{x}) \,,$$
where $h(\mathbf{x})$ is a known function.
- Robin (mixed) condition:
$$\alpha(\mathbf{x}) \frac{\partial \psi}{\partial n} + \beta(\mathbf{x}) \psi = d(\mathbf{x})$$
where $\alpha(\mathbf{x}), \beta(\mathbf{x})$ and $d(\mathbf{x})$ are known functions.

10.1.3 Superposition

$\mathcal{L}$ is a linear operator since
$$\mathcal{L}(\alpha\psi + \beta\phi) = \alpha\mathcal{L}\psi + \beta\mathcal{L}\phi$$
where ψ and ϕ are any functions of x and y, and α and β are any constants.

Theorem 10.4 (The principle of superposition). If $\mathcal{L}$ is a linear operator and both ψ and ϕ satisfy the homogeneous equation
$$\mathcal{L}\psi = \mathcal{L}\phi = 0 \,,$$
then $\alpha\psi + \beta\phi$ also satisfies the homogeneous equation.

Consider the following boundary value problem
$$\begin{cases} \mathcal{L}\psi(\mathbf{x}) = F(\mathbf{x}) & \mathbf{x} \in \Omega \\ \psi = f(\mathbf{x}) & \mathbf{x} \in \partial\Omega \,, \end{cases} \tag{A}$$
and the following initial, boundary value problem
$$\begin{cases} \mathcal{L}\phi(\mathbf{x}) = G(\mathbf{x}) & \mathbf{x} \in \Omega, t > 0 \\ \phi(\mathbf{x}, t) = g(\mathbf{x}, t) & \mathbf{x} \in \partial\Omega, t > 0 \\ \phi(\mathbf{x}) = h(\mathbf{x}) & \mathbf{x} \in \Omega, t = 0 \,. \end{cases} \tag{B}$$

Solve (A) by considering
$$\begin{cases} \mathcal{L}\psi_1(\mathbf{x}) = F(\mathbf{x}) & \mathbf{x} \in \Omega \\ \psi_1 = 0 & \mathbf{x} \in \partial\Omega \end{cases} \quad \begin{cases} \mathcal{L}\psi_2(\mathbf{x}) = 0 & \mathbf{x} \in \Omega \\ \psi_2 = f(\mathbf{x}) & \mathbf{x} \in \partial\Omega \,. \end{cases}$$
Then $\psi = \psi_1 + \psi_2$ solves (A).

Proof.
$$\mathcal{L}\psi = \mathcal{L}\psi_1 + \mathcal{L}\psi_2 = F \,, \ \mathbf{x} \in \Omega$$
$$\psi|_{\partial\Omega} = \psi_1|_{\partial\Omega} + \psi_2|_{\partial\Omega} = f \,.$$
□

Solve (B) by considering

$$\begin{cases} \mathcal{L}\phi_1(\mathbf{x}) = G(\mathbf{x}) & \\ \phi_1(\mathbf{x}, t) = 0 & \mathbf{x} \in \partial\Omega \\ \phi_1(\mathbf{x}) = 0 & t = 0 \end{cases} \quad \begin{cases} \mathcal{L}\phi_2(\mathbf{x}) = 0 & \\ \phi_2(\mathbf{x}, t) = g(\mathbf{x}, t) & \mathbf{x} \in \partial\Omega \\ \phi_2(\mathbf{x}) = 0 & t = 0 \end{cases} \quad \begin{cases} \mathcal{L}\phi_3(\mathbf{x}) = 0 & \\ \phi_3(\mathbf{x}, t) = 0 & \mathbf{x} \in \partial\Omega \\ \phi_3(\mathbf{x}) = h(\mathbf{x}) & t = 0\,. \end{cases}$$

Then $\phi = \phi_1 + \phi_2 + \phi_3$ solves (B).

Example. Consider a boundary value problem on a square

$$\begin{cases} \mathcal{L}\psi = 0 & \mathbf{x} \in (0,1) \times (0,1) \\ \psi = f_i(\mathbf{x}) & \text{on side } i,\ i = 1,2,3,4\,. \end{cases}$$

Then for $i = 1, 2, 3, 4$, solve

$$\begin{cases} \mathcal{L}\psi_i = 0 & \mathbf{x} \in (0,1) \times (0,1) \\ \psi_i = f_i(\mathbf{x}) & \text{on side } i \\ \psi_i = 0 & \text{on side } \neq i\,. \end{cases}$$

Then $\psi = \psi_1 + \psi_2 + \psi_3 + \psi_4$.

10.2 Common Partial Differential Equations

Definition 10.5. There are some common types of partial differential equations.

- *Laplace's equation.*
$$\nabla^2 \Psi(\mathbf{x}) = 0\,.$$
- *Poisson's equation.*
$$\nabla^2 \Psi(\mathbf{x}) = f(\mathbf{x})\,.$$
- *Diffusion equation.*
$$\frac{\partial \Psi(\mathbf{x})}{\partial t} = D\nabla^2 \Psi\,.$$
- *Wave equation.*
$$\frac{\partial^2 \Psi(\mathbf{x})}{\partial t^2} = c^2 \nabla^2 \Psi(\mathbf{x})\,.$$

10.3 Physical Examples and Applications

10.3.1 Waves on a Violin String

Consider small displacements on a stretched elastic string of line density μ. Assume that all displacements $y(x, t)$ are vertical. Resolve forces to obtain

$$T(x + \mathrm{d}x)\cos\theta(x + \mathrm{d}x) = T(x)\cos\theta(x) = T$$

$$\begin{aligned} (\mu\,\mathrm{d}x)\frac{\partial^2 y}{\partial t^2} &= T(x + \mathrm{d}x)\sin\theta(x + \mathrm{d}x) - T(x)\sin\theta(x) \\ &= T(x + \mathrm{d}x)\cos\theta(x + \mathrm{d}x)(\tan\theta(x + \mathrm{d}x) - \tan\theta(x))\,. \end{aligned}$$

And we observe that

$$\tan\theta = \frac{\partial y}{\partial x}\,.$$

From Taylor's theorem,

$$\begin{aligned}\mu\,\mathrm{d}x\,\frac{\partial^2 y}{\partial t^2} &= T(\tan\theta(x+\mathrm{d}x)-\tan\theta(x))\\ &= T\left(\frac{\partial}{\partial x}y(x+\mathrm{d}x\,,t)-\frac{\partial}{\partial x}y(x,t)\right)\\ &= T\frac{\partial^2 y}{\partial x^2}\,\mathrm{d}x+\dots\end{aligned}$$

and hence, in the infinitesimal limit,

$$\frac{\partial^2 y}{\partial t^2}=\frac{T}{\mu}\frac{\partial^2 y}{\partial x^2}\,.$$

This is the one-dimensional wave equation with wave speed $c=\sqrt{\frac{T}{\mu}}$.

10.3.2 Electromagnetic Waves

The theory of electromagnetism is based on *Maxwell's equations*

$$\begin{aligned}\boldsymbol{\nabla}\cdot\mathbf{E}&=\frac{\rho}{\epsilon_0}\\ \boldsymbol{\nabla}\cdot\mathbf{B}&=0\\ \boldsymbol{\nabla}\times\mathbf{E}&=-\frac{\partial\mathbf{B}}{\partial t}\\ \boldsymbol{\nabla}\times\mathbf{B}&=\mu_0\mathbf{J}+\mu_0\epsilon_0\frac{\partial\mathbf{E}}{\partial t}\,,\end{aligned}$$

which relate the electric field $\mathbf{E}$, magnetic field $\mathbf{B}$, the charge density ρ, the current density $\mathbf{J}$ and two fundamental constants μ_0 and ϵ_0.

In a vacuum where there is no free charge or current,

$$\begin{aligned}\mu_0\epsilon_0\frac{\partial^2\mathbf{E}}{\partial t^2}&=\boldsymbol{\nabla}\times\frac{\partial\mathbf{B}}{\partial t}\\ &=-\boldsymbol{\nabla}\times\boldsymbol{\nabla}\times\mathbf{E}\\ &=\nabla^2\mathbf{E}-\boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\mathbf{E})\\ &=\nabla^2\mathbf{E}\,.\end{aligned}$$

Therefore, we recovered the three-dimensional wave equation

$$\frac{\partial^2\mathbf{E}}{\partial t^2}=\frac{1}{\mu_0\epsilon_0}\nabla^2\mathbf{E}\,,$$

in which the wave speed is $c=\frac{1}{\sqrt{\mu_0\epsilon_0}}\approx 3\times10^8\,\mathrm{m\,s^{-1}}$.

Remarks.

- $\mathbf{B}$ obeys the same equation.
- The pressure perturbation of gas (sound waves) satisfies the scalar equivalent of this equation, with $c\approx 300\,\mathrm{m\,s^{-1}}$.

10.3.3 Electrostatics

In the absence of a magnetic field, Maxwell's equations for a static electric field $\mathbf{E}(\mathbf{x})$ give

$$\begin{cases}\boldsymbol{\nabla}\times\mathbf{E}=\mathbf{0}\\ \boldsymbol{\nabla}\cdot\mathbf{E}=\frac{\rho(\mathbf{x})}{\epsilon_0}\,.\end{cases}$$

Because $\mathbf{E}$ is irrotational, we can write $\mathbf{E} = -\boldsymbol{\nabla}\Phi$, where Φ is the *electrical potential*. We obtain the Poisson's equation

$$\nabla^2\Phi = -\frac{\rho}{\epsilon_0}\,.$$

In regions where there is no electric charge, this reduces to Laplace's equation

$$\nabla^2\Phi = 0\,.$$

Remark. In absence of currents, a static magnetic field $\mathbf{B}(\mathbf{x})$ satisfies

$$\begin{cases}\boldsymbol{\nabla}\times\mathbf{B} = \mathbf{0}\\ \boldsymbol{\nabla}\cdot\mathbf{B} = 0\,,\end{cases}$$

and so there is also a *magnetostatic potential* ψ satisfying the Laplace's equation

$$\nabla^2\psi = 0\,.$$

10.3.4 Gravitational Fields

The Newtonian gravitational field $\mathbf{g}$ of mass density ρ follows

$$\boldsymbol{\nabla}\cdot\mathbf{g} = -4\pi G\rho$$

and

$$\boldsymbol{\nabla}\times\mathbf{g} = \mathbf{0}\,.$$

Therefore, there exists a gravitational potential φ such that

$$\mathbf{g} = -\boldsymbol{\nabla}\varphi\,,$$

and it satisfies Poisson's equation

$$\nabla^2\varphi = 4\pi G\rho\,.$$

10.3.5 Diffusion of a Passive Tracer

Suppose we want to describe the diffusion of an inert chemical. Denote the mass concentration of the chemical by $C(\mathbf{x},t)$, and the material *flux vector* of the chemical by $\mathbf{q}(\mathbf{x},t)$. Then the amount of chemical crossing a small surface $\mathrm{d}\mathbf{S}$ in time δt is

$$\text{local flux} = (\mathbf{q}\cdot\mathrm{d}\mathbf{S})\delta t\,.$$

Hence the flux of chemical out of a closed surface $\mathcal{S}$ enclosing a volume $\mathcal{V}$ in time δt is

$$\text{surface flux} = \left(\oiint_{\mathcal{S}}\mathbf{q}\cdot\mathrm{d}\mathbf{S}\right)\delta t\,.$$

Let $Q(\mathbf{x},t)$ denote the chemical mass source per unit time per unit volume of the media. Then since the change of chemical within the volume is to be equal to the flux of the chemical out of the surface in time δt,

$$\left(\oiint_{\mathcal{S}}\mathbf{q}\cdot\mathrm{d}\mathbf{S}\right)\delta t = -\left(\frac{\mathrm{d}}{\mathrm{d}t}\iiint_{\mathcal{V}}C\,\mathrm{d}V\right)\delta t + \left(\iiint_{\mathcal{V}}Q\,\mathrm{d}V\right)\delta t\,.$$

Then by the divergence theorem, and exchanging the order of differentiation and integration,

$$\iiint_{\mathcal{V}}\left(\boldsymbol{\nabla}\cdot\mathbf{q} + \frac{\partial C}{\partial t} - Q\right)\mathrm{d}V = 0\,.$$

and since this is true for any volume,

$$\frac{\partial C}{\partial t} = -\boldsymbol{\nabla}\cdot\mathbf{q} + Q\,.$$

The simplest empirical law that relates concentration flux to concentration gradient is Fick's first law,

$$\mathbf{q} = -D\boldsymbol{\nabla}C\,,$$

where D is the diffusion constant.

The PDE governing the concentration is therefore

$$\frac{\partial C}{\partial t} = D\nabla^2 C + Q\,.$$

- *Diffusion Equation.* If there is no chemical source ($Q = 0$), then the governing equation becomes the diffusion equation
$$\frac{\partial C}{\partial t} = D\nabla^2 C\,.$$
- *Poisson's Equation.* If the system has reached a steady state ($\frac{\partial C}{\partial t} = 0$), then with $f(\mathbf{x}) = \frac{Q(\mathbf{x})}{D}$, the governing equation is the Poisson's equation
$$\nabla^2 C = -f\,.$$
- *Laplace's Equation.* If the system has reached a steady state and there is no chemical source then the concentration is governed by Laplace's equation
$$\nabla^2 C = 0\,.$$

10.3.6 Heat Flow

Let $\mathbf{q}(\mathbf{x}, t)$ denote the heat flux vector. Then the heat energy flowing out of a closed surface $\mathcal{S}$ enclosing a volume $\mathcal{V}$ in time δt is again

$$\text{surface flux} = \left(\iint_{\mathcal{S}} \mathbf{q}\cdot\mathrm{d}\mathbf{S}\right)\delta t\,.$$

Also, let $E(\mathbf{x}, t)$ denote the internal energy per unit mass, let $Q(\mathbf{x}, t)$ denote the heat source per unit time per unit volume, and let $\rho(\mathbf{x}, t)$ denote the mass density.

The heat flowing in and out of $\mathcal{S}$ must balance the change in internal energy and the heat source over a time δt, so

$$\left(\iint_{\mathcal{S}} \mathbf{q}\cdot\mathrm{d}\mathbf{S}\right)\delta t = -\left(\frac{\mathrm{d}}{\mathrm{d}t}\iiint_{\mathcal{V}} \rho E\,\mathrm{d}V\right)\delta t + \left(\iiint_{\mathcal{V}} Q\,\mathrm{d}V\right)\delta t\,.$$

From the first law of thermodynamics, for slow changes at constant volume,

$$E(\mathbf{x}, t) = c_v\theta(\mathbf{x}, t)\,,$$

where θ is the temperature, and c_v is the specific heat capacity (assumed constant). Hence,

$$\iiint_{\mathcal{V}} \left(\boldsymbol{\nabla}\cdot\mathbf{q} + \rho c_v\frac{\partial\theta}{\partial t} - Q\right)\mathrm{d}V = 0\,,$$

followed by

$$\rho c_v \frac{\partial \theta}{\partial t} = -\boldsymbol{\nabla} \cdot \mathbf{q} + Q \,.$$

Similarly, the simplest empirical law relating heat flow to temperature gradient is Fourier's law

$$\mathbf{q} = -k \boldsymbol{\nabla} \theta \,,$$

where k is the heat conductivity. If k is constant, then the PDE governing the temperature is

$$\frac{\partial \theta}{\partial t} = \nu \nabla^2 \theta + \frac{Q}{\rho c_v}$$

where $\nu = \frac{k}{\rho c_v}$ is the diffusivity.

Remark. The heat diffusion is remarkably similar to chemical diffusion.

10.3.7 Schrödinger Equation

The *Schrödinger equation* is

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{x}) \right] \psi = E \psi \,.$$

10.3.8 Ideal Fluid Flow

The flow of a fluid can be described by a vector field of the fluid's velocity $\mathbf{u}(t, \mathbf{x})$. If the flow is *irrotational* (no *vortex*, where the *vorticity* is given by $\boldsymbol{\omega} = \boldsymbol{\nabla} \times \mathbf{u}$) and non-viscous, then

$$\mathbf{u} = \boldsymbol{\nabla} \Phi \,,$$

where Φ is the *velocity potential*. The *continuity equation*, given by the conservation of mass,

$$\frac{\partial \rho}{\partial t} + \boldsymbol{\nabla} \cdot (\rho \mathbf{u}) = 0$$

reduces to

$$\boldsymbol{\nabla} \cdot \mathbf{u} = 0$$

if the fluid is incompressible and therefore has a uniform constant density ρ. Therefore the velocity potential satisfies Laplace's equation

$$\nabla^2 \Phi = 0 \,.$$

10.3.9 Other Equations

PDEs are also useful in non-scientific areas, such as the Black–Scholes equation for call option pricing

$$\frac{\partial w}{\partial t} = rw - rx \frac{\partial w}{\partial x} - \frac{1}{2} v^2 x^2 \frac{\partial^2 w}{\partial x^2} \,,$$

where $w(x, t)$ is the price of the call option of the stock, x is the variable market price of the stock, r is the fixed interest rate and v^2 is the variance rate of the stock price.

Despite all the equations above being linear, many other interesting equations are nonlinear, such as Euler's equation for an inviscid fluid

$$\rho \left(\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} \right) = -\boldsymbol{\nabla} p \,,$$

and the non-linear Schrödinger equation

$$i \frac{\partial A}{\partial t} + \frac{\partial^2 A}{\partial x^2} = A |A|^2 \,.$$

10.4 Wave Equation

10.4.1 One Dimensional Wave Equation

We have the initial, boundary value problem

$$\left\{\begin{array}{ll} \dfrac{\partial^2 y}{\partial t^2} = c^2 \dfrac{\partial^2 y}{\partial x^2} & (x,t) \in (0,L)\times(0,\infty) \\ y(0,t) = y(L,t) = 0 & t \in (0,\infty) \\ \left.\begin{array}{l} y(x,0) = f(x) \\ \dfrac{\partial y}{\partial t}(x,0) = g(x) \end{array}\right\} & x \in (0,L)\,. \end{array}\right. \tag{$\dagger$}$$

Try the solution of the form

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

Substituting the solution into the equation (†), we obtain

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

where a $\dot{}$ and a $'$ denote differentiation by t and x respectively. Rearrangement gives

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

where λ can only be a constant. We have therefore split the PDE into two ODEs:

$$\begin{cases} \ddot{T} - c^2\lambda T = 0 \\ X'' - \lambda X = 0\,. \end{cases}$$

There are three cases to consider based on the sign of λ.

- $\lambda = 0$. In this case,
$$\ddot{T}(t) = X''(x) = 0 \implies \begin{cases} T = A_0 + B_0 t \\ X = C_0 + D_0 x\,, \end{cases}$$
where A_0, B_0, C_0, D_0 are constants.
$$y = (A_0 + B_0 t)(C_0 + D_0 x)$$

- $\lambda = \sigma^2 > 0$. In this case,
$$\begin{cases} \ddot{T} - \sigma^2 c^2 T = 0 \\ X'' - \sigma^2 X = 0 \end{cases} \implies \begin{cases} T = A_\sigma e^{\sigma c t} + B_\sigma e^{-\sigma c t} \\ X = C_\sigma e^{\sigma x} + D_\sigma e^{-\sigma x}\,, \end{cases}$$
where $A_\sigma, B_\sigma, C_\sigma, D_\sigma$ are constants.
$$y = (A_\sigma e^{\sigma c t} + B_\sigma e^{-\sigma c t})(C_\sigma e^{\sigma x} + D_\sigma e^{-\sigma x})$$
or alternatively
$$y = (\tilde{A}_\sigma \cosh \sigma c t + \tilde{B}_\sigma \sinh \sigma c t)(\tilde{C}_\sigma \cosh \sigma x + \tilde{D}_\sigma \sinh \sigma x)$$

- $\lambda = -k^2 < 0$. In this case,
$$\begin{cases} \ddot{T} + k^2 c^2 T = 0 \\ X'' + k^2 X = 0 \end{cases} \implies \begin{cases} T = A_k \cos kct + B_k \sin kct \\ X = C_k \cos kx + D_k \sin kx\,, \end{cases}$$
where A_k, B_k, C_k, D_k are constants.
$$y = (A_k \cos kct + B_k \sin kct)(C_k \cos kx + D_k \sin kx)$$

Now substitute the boundary conditions in.

- $\lambda = 0$. If the homogeneous boundary conditions are to be satisfied for all time, then we must have $C_0 = D_0 = 0$. This gives a trivial solution of $y = 0$.
- $\lambda > 0$. Again, for the homogeneous boundary conditions to be satisfied, $C_\sigma = D_\sigma = 0$.
- $\lambda < 0$. Applying the boundary conditions yields

$$C_k = 0 \text{ and } D_k \sin kL = 0$$

If $D_k = 0$ then this solution is trivial (as for $\lambda \geq 0$), so the only non-trivial solution has

$$\sin kL = 0 \implies k = \frac{n\pi}{L}\,,$$

where n is a non-zero integer. These special values of k are eigenvalues and the corresponding eigenfunctions, or normal modes, are

$$X_{\frac{n\pi}{L}} = D_{\frac{n\pi}{L}} \sin \frac{n\pi x}{L}\,.$$

Hence, solutions to the wave equation that satisfy the homogeneous boundary conditions are

$$y_n(x,t) = \left(\mathcal{A}_n \cos \frac{n\pi ct}{L} + \mathcal{B}_n \sin \frac{n\pi ct}{L}\right) \sin \frac{n\pi x}{L}\,.$$

Since the wave equation is linear, we may superimpose solutions to get the general solution

$$y(x,t) = \sum_{n=1}^{\infty} \left(\mathcal{A}_n \cos \frac{n\pi ct}{L} + \mathcal{B}_n \sin \frac{n\pi ct}{L}\right) \sin \frac{n\pi x}{L}\,.$$

Note that the solution has the form of a Fourier series. To satisfy the initial conditions, we require that

$$y(x,0) = f(x) = \sum_{n=1}^{\infty} \mathcal{A}_n \sin \frac{n\pi x}{L}\,,$$

$$\frac{\partial y}{\partial t}(x,0) = g(x) = \sum_{n=1}^{\infty} \mathcal{B}_n \frac{n\pi c}{L} \sin \frac{n\pi x}{L}\,.$$

$\mathcal{A}_n$ and $\mathcal{B}_n$ can be found using the orthogonality of sin

$$\int_0^L \sin \frac{n\pi x}{L} \sin \frac{m\pi x}{L}\,\mathrm{d}x = \frac{L}{2}\delta_{nm}\,.$$

Hence for an integer $m > 0$,

$$\begin{aligned}
\frac{2}{L}\int_0^L f(x) \sin \frac{m\pi x}{L}\,\mathrm{d}x &= \frac{2}{L}\int_0^L \left(\sum_{n=1}^{\infty} \mathcal{A}_n \sin \frac{n\pi x}{L}\right) \sin \frac{m\pi x}{L}\,\mathrm{d}x \\
&= \sum_{n=1}^{\infty} \frac{2\mathcal{A}_n}{L} \int_0^L \sin \frac{n\pi x}{L} \sin \frac{m\pi x}{L}\,\mathrm{d}x \\
&= \sum_{n=1}^{\infty} \mathcal{A}_n \delta_{mn} \\
&= \mathcal{A}_m\,.
\end{aligned}$$

Similarly,

$$\mathcal{B}_m = \frac{2}{m\pi c}\int_0^L g(x) \sin \frac{m\pi x}{L}\,\mathrm{d}x\,.$$

10.4.2 Waves on a Drum

Consider the wave propagating on the surface of a drum, on a region $\Omega = \{(r,\theta) \mid r \in [0,1], \theta \in [0, 2\pi)\}$, where (r,θ) are the plane polar coordinates. This is an initial-boundary value problem

$$\left\{\begin{array}{ll} \dfrac{\partial^2\psi}{\partial t^2} = c^2\nabla^2\psi & (\mathbf{x},t) \in \Omega \times (0,\infty) \\ \psi(\mathbf{x},t) = 0 & (\mathbf{x},t) \in \partial\Omega \times (0,\infty) \\ \left.\begin{array}{l} \psi(\mathbf{x},0) = f(\mathbf{x}) \\ \dfrac{\partial\psi}{\partial t}(\mathbf{x},0) = g(\mathbf{x}) \end{array}\right\} & \mathbf{x} \in \Omega\,. \end{array}\right. \tag{††}$$

For simplicity, we will assume $f = f(r)$ and $g = g(r)$, so we expect $\psi = \psi(r,t)$. Try the solution of the form

$$\psi = R(r)T(t)\,,$$

and the wave equation becomes

$$\frac{\ddot{T}}{c^2 T} - \frac{(rR')'}{rR} = 0\,.$$

There must be a constant λ such that

$$\begin{cases} -(rR')' = \lambda rR \\ \ddot{T} + \lambda c^2 T = 0\,. \end{cases}$$

The R equation is the Bessel equation of order zero. We know that the solution nonsingular at $r = 0$ are

$$R_k(r) = J_0(j_{0k}r)\,,\ \lambda_k = j_{0k}^2\,,\ \text{for } k = 1, 2, \ldots$$

The solutions to the T equation are then

$$T_k(t) = A_k \cos(j_{0k}ct) + B_k \sin(j_{0k}ct)\,.$$

This gives the following general solution that satisfies the homogeneous boundary conditions

$$\psi(r,t) = \sum_{k=1}^{\infty} J_0(j_{0k}r)[A_k \cos(j_{0k}ct) + B_k \sin(j_{0k}ct)]\,.$$

To match the initial conditions, we require

$$f(r) = \sum_{k=1}^{\infty} A_k J_0(j_{0k}r)\,,$$

$$g(r) = \sum_{k=1}^{\infty} B_k c j_{0k} J_0(j_{0k}r)\,.$$

We can solve these equations by the orthogonality conditions for Bessel functions

$$\int_0^1 J_0(j_{0k}r)J_0(j_{0l}r)r\,\mathrm{d}r = \frac{1}{2}J_0'(j_{0k})^2\delta_{kl}\,.$$

This allows us to find A_k and B_k as follows

$$A_k = \frac{2}{J_0'(j_{0k})^2}\int_0^1 J_0(j_{0k}r)f(r)r\,\mathrm{d}r\ ,$$

$$B_k = \frac{2}{cj_{0k}J_0'(j_{0k})^2}\int_0^1 J_0(j_{0k}r)g(r)r\,\mathrm{d}r\ .$$

10.4.3 Energy Conservation and Uniqueness of Solution

For the wave equation on the string, define the *energy*

$$E(t) = \frac{1}{2}\mu \int_0^L \left(\frac{\partial y}{\partial t}\right)^2 \mathrm{d}x + \frac{1}{2}T \int_0^L \left(\frac{\partial y}{\partial x}\right)^2 \mathrm{d}x \ ,$$

where the former part is the kinetic energy and the latter part is the potential energy. For simplicity, take $\mu = 1$, then $T = c^2$, and the energy simplifies to

$$E(t) = \frac{1}{2}\int_0^L \left[\left(\frac{\partial y}{\partial t}\right)^2 + c^2\left(\frac{\partial y}{\partial x}\right)^2\right] \mathrm{d}x \ .$$

This can be generalised to higher dimensions

$$E(t) = \frac{1}{2}\int_\Omega \left[\left(\frac{\partial \psi}{\partial t}\right)^2 + c^2|\boldsymbol{\nabla}\psi|^2\right] \mathrm{d}V \ ,$$

where ψ satisfies the wave equation for higher dimensions

$$\begin{cases} \dfrac{\partial^2 \psi}{\partial t^2} = c^2\nabla^2\psi & (\mathbf{x}, t) \in \Omega \times (0, \infty) \\ \psi(\mathbf{x}, t) = 0 & (\mathbf{x}, t) \in \partial\Omega \times (0, \infty) \\ \left.\begin{array}{l} \psi(\mathbf{x}, 0) = f(\mathbf{x}) \\ \dfrac{\partial \psi}{\partial t}(\mathbf{x}, 0) = g(\mathbf{x}) \end{array}\right\} & \mathbf{x} \in \Omega \, . \end{cases} \tag{††}$$

Theorem 10.6. The energy is conserved for a solution of the wave equation.

Proof. Consider how $E(t)$ changes as ψ evolves with t.

$$\begin{aligned} \dot{E}(t) &= \int_\Omega \left[\frac{\partial \psi}{\partial t}\frac{\partial^2 \psi}{\partial t^2} + c^2 \boldsymbol{\nabla}\psi \cdot \boldsymbol{\nabla}\frac{\partial \psi}{\partial t}\right] \mathrm{d}V \\ &= c^2 \int_\Omega \left[\frac{\partial \psi}{\partial t}\nabla^2\psi + \boldsymbol{\nabla}\psi \cdot \boldsymbol{\nabla}\frac{\partial \psi}{\partial t}\right] \mathrm{d}V \\ &= c^2 \int_\Omega \boldsymbol{\nabla} \cdot \left(\frac{\partial \psi}{\partial t}\boldsymbol{\nabla}\psi\right) \mathrm{d}V \\ &= c^2 \int_{\partial\Omega} \frac{\partial \psi}{\partial t}\boldsymbol{\nabla}\psi \cdot \mathbf{n} \, \mathrm{d}\mathbf{S} \ . \end{aligned}$$

Since $\psi(\mathbf{x}, t) = 0$ for $\mathbf{x} \in \partial\Omega$,

$$\frac{\partial \psi}{\partial t} = 0 \text{ for } \mathbf{x} \in \partial\Omega \, ,$$

so the integral vanishes. □

Theorem 10.7. The solution to the wave equation (††) is unique.

Proof. Suppose the solution is not unique. Let ψ_1 and ψ_2 solve (††).

Set $\psi = \psi_1 - \psi_2$. By linearity, ψ satisfies (††), but with initial conditions $f = g = 0$. Energy associated with ψ is

$$\begin{aligned} E(t) &= \frac{1}{2}\int_\Omega \left(\frac{\partial \psi}{\partial t}\right)^2 + c^2|\boldsymbol{\nabla}\psi|^2 \, \mathrm{d}V \\ &= E(0) = 0 \, , \end{aligned}$$

so

$$\frac{\partial \psi}{\partial t} = |\boldsymbol{\nabla}\psi| = 0$$

throughout Ω and $t \geq 0$, i.e. ψ is constant. Since $\psi = 0$ on $\partial\Omega$, $\psi = 0$ everywhere. $\psi_1 = \psi_2$, i.e. the solution is unique. □

10.5 Diffusion Equation

10.5.1 1D Diffusion of Chemical

Consider the problem of a solvent occupying the region between $x = 0$ and $x = L$. Suppose that at $t = 0$, there is no chemical in the solvent. Suppose also for $t > 0$, the concentration of the chemical is maintained at C_0 at $x = 0$, and is 0 at $x = L$. This is an initial, boundary value problem

$$\begin{cases} \dfrac{\partial C}{\partial t} = D\dfrac{\partial^2 C}{\partial x^2} & (x,t) \in (0,L) \times (0,\infty) \\ C(0,t) = C_0 & t \in (0,\infty) \\ C(L,t) = 0 & t \in (0,\infty) \\ C(x,0) = 0 & x \in (0,L)\,. \end{cases} \tag{$\dagger$}$$

Seek solutions to (†) of the separable form

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

On substitution we obtain

$$X\dot{T} = DTX''\,,$$

and after arrangement we have

$$\frac{1}{D}\frac{\dot{T}(t)}{T(t)} = \frac{X''(x)}{X(x)} = \lambda\,,$$

where λ is again a constant. We therefore have

$$\begin{cases} \dot{T} - D\lambda T = 0 \\ X'' - \lambda X = 0\,, \end{cases}$$

with again three cases to consider.

- $\lambda = 0$. In this case
$$\begin{cases} \dot{T}(t) = 0 \\ X''(x) = 0 \end{cases} \implies \begin{cases} T = a_0 \\ X = \beta_0 + \gamma_0 x \end{cases}$$
where α_0, β_0 and γ_0 are constants. Combining these results we obtain (with $\alpha_0 = 1$ wlog)
$$C = \beta_0 + \gamma_0 x\,.$$

- $\lambda = \sigma^2 > 0$.
$$\begin{cases} \dot{T} - D\sigma^2 T = 0 \\ X'' - \sigma^2 X = 0 \end{cases} \implies \begin{cases} T = \alpha_\sigma \exp\big(D\sigma^2 t\big) \\ X = \beta_\sigma \cosh\sigma x + \gamma_\sigma \sinh\sigma x\,, \end{cases}$$
where $\alpha_\sigma, \beta_\sigma$ and γ_σ are all constants. We therefore obtain
$$C = \exp\big(D\sigma^2 t\big)(\beta_\sigma \cosh\sigma x + \gamma_\sigma \sinh\sigma x)\,.$$

- $\lambda = -k^2 < 0$.
$$\begin{cases} \dot{T} + Dk^2 T = 0 \\ X'' + k^2 X = 0 \end{cases} \implies \begin{cases} T = \alpha_k \exp\big(-Dk^2 t\big) \\ X = \beta_k \cos kx + \gamma_k \sin kx\,, \end{cases}$$
where α_k, β_k and γ_k are all constants. We therefore obtain
$$C = \exp\big(-Dk^2 t\big)(\beta_k \cos kx + \gamma_k \sin kx)\,.$$

Note that the separable solutions with $\lambda \neq 0$ depend on time, while the solution with $\lambda = 0$ does not. The cases $\lambda \leq 0$ are suitable for the initial and boundary conditions, so we need to consider all.

We may first try to fit the $\lambda = 0$ solution to the boundary condition. We call this part of the total solution $C_\infty(x)$, then

$$C_\infty(x) = C_0\Big(1 - \frac{x}{L}\Big)\,,$$

which is just a linear variation in C from C_0 at $x = 0$ to 0 at $x = L$.

Write

$$C(x,t) = C_\infty(x) + \tilde{C}(x,t)\,,$$

where $\tilde{C}$ is a sum of the separable time-dependent solutions with $\lambda \neq 0$. Then from (†), the conditions for $\tilde{C}$ is

$$\begin{cases} \dfrac{\partial \tilde{C}}{\partial t} = D\dfrac{\partial^2 \tilde{C}}{\partial x^2} & (x,t) \in (0,L)\times(0,\infty) \\ \tilde{C}(0,t) = \tilde{C}(L,t) = 0 & t \in (0,\infty) \\ \tilde{C}(x,0) = -C_0\Big(1 - \dfrac{x}{L}\Big) & x \in (0,L)\,. \end{cases}$$

The homogeneous boundary conditions are satisfied by $\lambda = -k^2 < 0$ if $\beta_k = 0$ and $\gamma_k \sin kL = 0$, with

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

The corresponding eigenfunctions are

$$X_n = \Gamma_n \sin\frac{n\pi x}{L}\,,$$

where $\Gamma_n = \gamma_{\frac{n\pi}{L}}$.

Since the equation is linear, we can add them to get the general solution

$$\tilde{C}(x,t) = \sum_{n=1}^{\infty} \Gamma_n \exp\Big(-\frac{n^2\pi^2 Dt}{L^2}\Big)\sin\frac{n\pi x}{L}\,.$$

The Γ_n are fixed by the initial condition

$$-C_0\Big(1 - \frac{x}{L}\Big) = \sum_{n=1}^{\infty}\Gamma_n \sin\frac{n\pi x}{L}\,.$$

Hence

$$\Gamma_m = -\frac{2C_0}{L}\int_0^L \Big(1 - \frac{x}{L}\Big)\sin\frac{m\pi x}{L}\,\mathrm{d}x = -\frac{2C_0}{m\pi}\,.$$

The solution is thus given by

$$\begin{aligned} C &= C_0\Big(1 - \frac{x}{L}\Big) - \sum_{n=1}^{\infty}\frac{2C_0}{n\pi}\exp\Big(-\frac{n^2\pi^2 Dt}{L^2}\Big)\sin\frac{n\pi x}{L} \\ &= \sum_{n=1}^{\infty}\frac{2C_0}{n\pi}\left(1 - \exp\Big(-\frac{n^2\pi^2 Dt}{L^2}\Big)\right)\sin\frac{n\pi x}{L}\,. \end{aligned}$$

Remarks.

- As $t \to \infty$,
$$C \to C_0\Big(1 - \frac{x}{L}\Big) = C_\infty(x)\,.$$
- This solution is odd and has a period of $2L$. We are in effect solving the $2L$-periodic diffusion problem where C is initially zero. Then, at $t = 0^+$, C is raised to 1 at $2nL^+$ and lowered to -1 at $2nL^-$, and kept zero everywhere else.

10.5.2 Heating along a Square Sheet

Consider the heat distributed over a square $\Omega = \{(x, y) \mid x \in (0, L), y \in (0, L)\}$ with heat loss on its boundary. This is an initial-boundary value problem

$$\begin{cases} \dfrac{\partial \varphi}{\partial t} = \kappa \nabla^2 \varphi & (\mathbf{x}, t) \in \Omega \times (0, \infty) \\ \varphi(\mathbf{x}, 0) = f & \mathbf{x} \in \Omega \\ \varphi(\mathbf{x}, t) = 0 & (\mathbf{x}, t) \in \partial\Omega \times (0, \infty)\,. \end{cases} \tag{††}$$

We write $\varphi = T(t)X(x)Y(y)$ to find

$$\frac{\dot{T}}{\kappa T} = \frac{X''}{X} + \frac{Y''}{Y} = -\mu\,.$$

This gives us a set of differential equations

$$\begin{cases} \dot{T} + \kappa\mu T = 0 \\ X'' + \lambda X = 0 \\ Y'' + (\mu - \lambda)Y = 0\,. \end{cases}$$

T equation solves to

$$T(t) = Ae^{-\mu\kappa t}\,.$$

The boundary conditions are $X(0) = X(L) = Y(0) = Y(L) = 0$, so

$$X_n(x) = \sin\Big(\frac{n\pi x}{L}\Big)\,,\ \lambda_n = \Big(\frac{n\pi}{L}\Big)^2\,,\ n = 1, 2, \ldots$$

$$Y_m(y) = \sin\Big(\frac{m\pi y}{L}\Big)\,,\ \mu_{mn} - \lambda_n = \Big(\frac{m\pi}{L}\Big)^2\,,\ m = 1, 2, \ldots$$

We therefore have the general solution that satisfies the homogeneous boundary conditions

$$\varphi(x, y, t) = \sum_{n,m=1}^{\infty} A_{mn} e^{-\kappa\mu_{mn}t} \sin\Big(\frac{n\pi x}{L}\Big) \sin\Big(\frac{m\pi y}{L}\Big)\,,$$

where

$$\mu_{mn} = \Big(\frac{n\pi}{L}\Big)^2 + \Big(\frac{m\pi}{L}\Big)^2\,.$$

Using the orthogonality, the initial condition gives

$$A_{mn} = \frac{4}{L^2} \int_0^L \int_0^L \sin\Big(\frac{n\pi x}{L}\Big) \sin\Big(\frac{m\pi y}{L}\Big) f(x, y)\,\mathrm{d}x\,\mathrm{d}y\ .$$

Because μ_{mn} increase quadratically with m and n, as t increases, the dominant contribution comes from the first $m = n = 1$ mode.

$$\varphi(x, y, t) \sim A_{11} e^{-\frac{2\kappa\pi^2 t}{L^2}} \sin\Big(\frac{\pi x}{L}\Big) \sin\Big(\frac{\pi y}{L}\Big) \quad \text{as } t \to \infty\,.$$

10.5.3 Energy Loss and Uniqueness of Solution

For a heat equation problem as in (††), the total energy is defined as

$$Q(t) = \frac{1}{2} \int_\Omega \varphi^2\,\mathrm{d}V\ .$$

Note that

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = \int_\Omega \varphi \frac{\mathrm{d}\varphi}{\mathrm{d}t}\,\mathrm{d}V = \kappa \int_\Omega \varphi \nabla^2 \varphi \,\mathrm{d}V\ .$$

Since we have

$$\begin{aligned}\boldsymbol{\nabla}\cdot(\Psi\boldsymbol{\nabla}\Psi) &= \boldsymbol{\nabla}\Psi\cdot\boldsymbol{\nabla}\Psi + \Psi\boldsymbol{\nabla}\cdot\boldsymbol{\nabla}\Psi\\ &= |\boldsymbol{\nabla}\Psi|^2 + \Psi\nabla^2\Psi\,,\end{aligned}$$

the integral becomes

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = \kappa\int_{\partial\Omega}\varphi\boldsymbol{\nabla}\varphi\cdot\mathrm{d}S - \kappa\int_\Omega |\boldsymbol{\nabla}\varphi|^2\,\mathrm{d}V\ .$$

The first term vanishes since $\varphi = 0$ on the boundary, so

$$Q'(t) = -\kappa\int_\Omega |\boldsymbol{\nabla}\varphi|^2\,\mathrm{d}V \le 0\,.$$

Theorem 10.8. The solutions to heat equations with initial and boundary conditions are unique.

Proof. Assume there are two solutions, φ_1 and φ_2 to a heat equation with inhomogeneous boundary conditions and initial conditions. Then $\psi = \varphi_1 - \varphi_2$ would satisfy the heat equation with homogeneous boundary conditions and zero initial data. Since $Q(0) = 0$, we can get

$$Q(t) \le 0\,.$$

But clearly $Q(t) \ge 0$ by definition, so $Q(t) = 0$ for all t.

$$\int_\Omega \psi^2\,\mathrm{d}V = 0\,.$$

This is only possible when $\psi = 0$, and so $\varphi_1 = \varphi_2$. The solution is unique. □

10.6 Laplace's Equation and Poisson's Equation

10.6.1 Uniqueness of Solutions of the Poisson's Equation

Theorem 10.9. The solution to Poisson's equation in a domain Ω with a Dirichlet boundary condition on its boundary $\partial\Omega$ is unique.

$$\begin{cases}\nabla^2\Phi(\mathbf{x}) = \rho(\mathbf{x}) & \mathbf{x}\in\Omega\\ \Phi(\mathbf{x}) = f(\mathbf{x}) & \mathbf{x}\in\partial\Omega\,.\end{cases}$$

Proof. Suppose that there are two solutions, $\Phi_1(\mathbf{x})$ and $\Phi_2(\mathbf{x})$. Let $\Psi = \Phi_1 - \Phi_2$, then Ψ satisfies Laplace's equation with zero boundary conditions.

Now consider

$$\begin{aligned}\boldsymbol{\nabla}\cdot(\Psi\boldsymbol{\nabla}\Psi) &= |\boldsymbol{\nabla}\Psi|^2 + \Psi\nabla^2\Psi\\ &= |\boldsymbol{\nabla}\Psi|^2\,.\end{aligned}$$

Therefore,

$$\begin{aligned}\int_\Omega |\boldsymbol{\nabla}\Psi|^2\,\mathrm{d}V &= \int_\Omega \boldsymbol{\nabla}\cdot(\Psi\boldsymbol{\nabla}\Psi)\,\mathrm{d}V\\ &= \oint_{\partial\Omega}\Psi\boldsymbol{\nabla}\Psi\cdot\mathrm{d}\mathbf{S}\\ &= 0\end{aligned}$$

since $\Psi = 0$ on $\partial\Omega$. This integral can only be 0 if $\boldsymbol{\nabla}\Psi = \mathbf{0}$ everywhere in Ω, so Ψ is a constant. Since Ψ is 0 on $\partial\Omega$, Ψ can only be 0 throughout Ω. The solution must be unique. □

Theorem 10.10. The solution to a Poisson's equation in a volume Ω with a Neumann boundary condition on its boundary $\partial\Omega$ is unique up to a constant.

$$\begin{cases} \nabla^2\Phi(\mathbf{x}) = \rho(\mathbf{x}) & \mathbf{x} \in \Omega \\ \dfrac{\partial\Phi}{\partial n} \coloneqq \mathbf{n}(\mathbf{x}) \cdot \boldsymbol{\nabla}\Phi = f(\mathbf{x}) & \mathbf{x} \in \partial\Omega\,. \end{cases}$$

Proof. Note that

$$\left.\frac{\partial\Psi}{\partial n}\right|_{\partial\Omega} = (\boldsymbol{\nabla}\Phi_1 \cdot \mathbf{n})|_{\partial\Omega} - (\boldsymbol{\nabla}\Phi_2 \cdot \mathbf{n})|_{\partial\Omega} = f(\mathbf{x}) - f(\mathbf{x}) = 0\,,$$

then the rest of the proof is similar to the above. □

10.6.2 Poisson's Equation on a Semi-infinite Rod

Consider uniformly heating a semi-infinite rod with $x \geq 0$ and $0 \leq y \leq 1$. This is an initial, boundary value problem:

$$\begin{cases} \nabla^2\theta = -1 & (x,y) \in (0,\infty) \times (0,1) \\ \theta(x,0) = \theta(x,1) = 0 & x \in (0,\infty) \\ \theta(0,y) = 0 & y \in (0,1) \\ \lim\limits_{x\to\infty} \dfrac{\partial\theta}{\partial x} = 0 & y \in (0,1)\,. \end{cases} \tag{†}$$

Let us first find out a particular solution independent of x. Poisson's equation then reduces to

$$\frac{\mathrm{d}^2\theta_s}{\mathrm{d}y^2} = -1\,,$$

which has solution

$$\theta_s = a_0 + b_0 y - \frac{1}{2}y^2\,,$$

where a_0 and b_0 are constants. Using the boundary conditions on y, we obtained the particular solution

$$\theta_s = \frac{1}{2}y(1-y) \geq 0\,.$$

Define $\varphi = \theta - \theta_s$. Now by linearity, we only need to solve the following boundary value problem

$$\begin{cases} \nabla^2\varphi = 0 & (x,y) \in (0,\infty) \times (0,1) \\ \varphi(x,0) = \varphi(x,1) = 0 & x \in (0,\infty) \\ \varphi(0,y) = -\dfrac{1}{2}y(1-y) & y \in (0,1) \\ \lim\limits_{x\to\infty} \dfrac{\partial\varphi}{\partial x} = 0 & y \in (0,1)\,. \end{cases} \tag{††}$$

By writing $\varphi(x,y) = X(x)Y(y)$ and substituting into (††), it follows that

$$\frac{X''(x)}{X(x)} = -\frac{Y''(y)}{Y(y)} = \lambda\,,$$

so that

$$\begin{cases} X'' - \lambda X = 0 \\ Y'' + \lambda Y = 0\,. \end{cases}$$

This leaves us with three possibilities

(i) $\lambda = 0$.

$$\varphi = (A_0 + B_0 x)(C_0 + D_0 y)\,.$$

(ii) $\lambda = \sigma^2 > 0$.

$$\varphi = (A_\sigma \cosh \sigma x + B_\sigma \sinh \sigma x)(C_\sigma \cos \sigma y + D_\sigma \sin \sigma y)\,.$$

(iii) $\lambda = -k^2 < 0$.

$$\varphi = (A_k \cos kx + B_k \sin kx)(C_k \cosh ky + D_k \sinh ky)\,.$$

The boundary conditions $\varphi(x, 0) = 0$ and $\varphi(x, 1) = 0$ implies that only solutions proportional to $\sin n\pi y$ are appropriate. Hence we try $\lambda = n^2\pi^2$ where n is an integer. The eigenfunctions are thus

$$\varphi_n = (\mathcal{A}_n e^{n\pi x} + \mathcal{B}_n e^{-n\pi x}) \sin(n\pi y)\,,$$

where $\mathcal{A}_n$ and $\mathcal{B}_n$ are constants. However, if boundary conditions as $x \to \infty$ are to be satisfied then $\mathcal{A}_n = 0$. Hence the solution has the form

$$\varphi = \sum_{n=1}^{\infty} \mathcal{B}_n e^{-n\pi x} \sin(n\pi y)\,.$$

$\mathcal{B}_n$ are fixed by the first boundary condition

$$-\frac{1}{2} y(1 - y) = \sum_{n=1}^{\infty} \mathcal{B}_n \sin(n\pi y)\,.$$

Using the orthogonality relations, it follows that

$$\mathcal{B}_m = 2\frac{(-1)^m - 1}{m^3\pi^3}\,,$$

and hence

$$\theta = \frac{1}{2} y(1 - y) - \sum_{l=0}^{\infty} \frac{4}{\pi^3(2l + 1)^3} \sin((2l + 1)\pi y) e^{-(2l+1)\pi x}\,.$$

10.6.3 Laplace's Equation in Plane Polar Coordinates

In the plane polar coordinates, the Laplace's equation is

$$\nabla^2 \Psi = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \Psi}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 \Psi}{\partial \phi^2} = 0\,.$$

Remark. The same equation arises in cylindrical polar coordinates (r, ϕ, z) when $\frac{\partial \Psi}{\partial z} = 0$.

If we consider separable solutions of the form $\Psi(r, \phi) = R(r)\Phi(\phi)$, then the Laplace's equation reduces to

$$\frac{\Phi}{r}\frac{\partial}{\partial r}\left(r\frac{\partial R}{\partial r}\right) + \frac{R}{r^2}\frac{\partial^2 \Phi}{\partial \phi^2} = 0\,.$$

Rearrangement gives

$$\frac{r}{R}\frac{\partial}{\partial r}\left(r\frac{\partial R}{\partial r}\right) = -\frac{1}{\Phi}\frac{\partial^2 \Phi}{\partial \phi^2}\,.$$

The LHS is a function of r only and the RHS is a function of ϕ only, so both must equal to a constant λ.

The equation for $\Phi(\phi)$ gives

$$\frac{\partial^2 \Phi}{\partial \phi^2} + \lambda \Phi = 0\,,$$

and so

$$\Phi(\phi) = \begin{cases} A + B\phi & \lambda = 0 \\ A\cos\sqrt{\lambda}\phi + B\sin\sqrt{\lambda}\phi & \lambda \neq 0\,. \end{cases}$$

In some cases Ψ corresponds to some physical quantity, and so Ψ must be periodic:

$$\Psi(r, \phi) = \Psi(r, \phi + 2\pi)\,.$$

However, in some situations, Ψ is not a physical quantity (e.g. potential) but $\boldsymbol{\nabla}\Psi$ is. This more general case requires

$$\Phi'(\phi) = \Phi'(\phi + 2\pi)\,,$$

which gives

$$\lambda = n^2\,,\ n \in \mathbb{N}\,.$$

Hence,

$$\Phi_n(\phi) = \begin{cases} A + B\phi & n = 0 \\ A\cos n\phi + B\sin n\phi & n \neq 0,\ n \in \mathbb{Z}\,. \end{cases}$$

Returning to $R(r)$, the equation

$$\frac{r}{R}\frac{\mathrm{d}}{\mathrm{d}r}\left(r\frac{\mathrm{d}R}{\mathrm{d}r}\right) = n^2$$
$$\implies r^2 R'' + rR' - n^2 R = 0$$

gives the solutions

$$R_n(r) = \begin{cases} C + D\ln r & n = 0 \\ Cr^n + Dr^{-n} & n \neq 0\,. \end{cases}$$

Combining R and Φ gives

$$\Psi_n(r, \phi) = R_n(r)\Phi_n(\phi) = \begin{cases} (C_0 + D_0 \ln r)(A_0 + B_0\phi) & n = 0 \\ (C_n r^n + D_n r^{-n})(A_n \cos n\phi + B_n \sin n\phi) & n \in \mathbb{N}\,. \end{cases}$$

The $\phi \ln r$ combination has to be excluded because it does not satisfy the periodic requirement of $\boldsymbol{\nabla}\Psi$.

Theorem 10.11. The general solutions to Laplace's equation in planar polar coordinates are

$$\Psi = A_0 + B_0\phi + C_0 \ln r + \sum_{n=1}^{\infty}(A_n r^n + C_n r^{-n})\cos n\phi + \sum_{n=1}^{\infty}(B_n r^n + D_n r^{-n})\sin n\phi\,,$$

or equivalently

$$\Psi = A_0 + B_0\phi + C_0 \ln r + \sum_{n=-\infty,\, n\neq 0}^{\infty} r^n (A_n \cos n\phi + B_n \sin n\phi)\,.$$

Example. Steady state temperature distribution in a cylinder.

An infinitely long cylinder of radius a centred at the origin is heated on its boundary with the boundary conditions

$$T(a, \phi) = \begin{cases} T_0 & 0 \leq \phi < \pi \\ -T_0 & \pi \leq \phi < 2\pi\,. \end{cases}$$

The steady-state temperature $T(r, \phi)$ for $r < a$ satisfies

$$\nabla^2 T = 0 \,.$$

The temperature T must be finite and single-valued, so a general solution would be given by

$$T = A_0 + \sum_{n=1}^{\infty} r^n (A_n \cos n\phi + B_n \sin n\phi) \,.$$

At $r = a$,

$$T(a, \phi) = A_0 + \sum_{n=1}^{\infty} a^n (A_n \cos n\phi + B_n \sin n\phi) \,,$$

which is a Fourier series. Applying the boundary conditions gives the final solution of

$$T(r, \phi) = \frac{4T_0}{\pi} \sum_{n=1}^{\infty} \frac{r^{2n-1}}{(2n-1)a^{2n-1}} \sin[(2n-1)\phi] \,.$$

Example. 2D fluid flow past a circular barrier.

Find the 2D velocity field $\mathbf{u}$ of an ideal (irrotational and non-viscous), incompressible fluid in steady flow past a circular barrier of radius r_0. The fluid has constant velocity $\mathbf{U} = U\hat{\mathbf{x}}$ at infinity.

The fluid flow can be described by a velocity potential Φ such that $\mathbf{u} = \boldsymbol{\nabla}\Phi$, satisfying the Laplace's equation

$$\nabla^2 \Phi = 0 \,.$$

At infinity we require $\boldsymbol{\nabla}\Phi = U\hat{\mathbf{x}}$, so

$$\Phi \to Ux = Ur\cos\phi \,.$$

It also implies that, in our general solution, $B_0 = C_0 = B_n = A_{n\neq 1} = 0$ and $A_1 = U$, with arbitrary A_0. Therefore, we may write our general solution as

$$\Phi(r, \phi) = Ur\cos\phi + \sum_{n=1}^{\infty} r^{-n} (C_n \cos n\phi + D_n \sin n\phi) \,.$$

At the surface of the cylinder, the flow must be tangent to it, so the radial component of the fluid must be 0.

$$\left.\frac{\partial \Phi}{\partial r}\right|_{r=r_0} = U\cos\phi - \sum_{n=1}^{\infty} n r_0^{-(n+1)} (C_n \cos n\phi + D_n \sin n\phi) = 0 \,.$$

This is true for all Φ, so we must have

$$D_n = C_{n>1} = 0 \,,\ C_1 = Ur_0^2 \,.$$

Therefore, we have the solution

$$\Phi(r, \phi) = \left(r + \frac{r_0^2}{r}\right) U\cos\phi = \mathbf{U}\cdot\mathbf{r}\left(1 + \frac{r_0^2}{r^2}\right) ,$$

$$\mathbf{u}(r, \phi) = \left(1 + \frac{r_0^2}{r^2}\right)\mathbf{U} - \frac{2r_0^2}{r^4}(\mathbf{U}\cdot\mathbf{r})\mathbf{r} \,.$$

10.6.4 Laplace's Equation in Spherical Polar Coordinates with Axial Symmetry

In spherical polar coordinates (r, θ, ϕ), when $\Psi(r, \theta, \phi)$ is axisymmetric ($\frac{\partial \Psi}{\partial \phi} = 0$), the Laplace's equation is

$$\nabla^2 \Psi = \frac{1}{r^2} \frac{\partial}{\partial r}\left(r^2 \frac{\partial \Psi}{\partial r}\right) + \frac{1}{r^2 \sin\theta} \frac{\partial}{\partial \theta}\left(\sin\theta \frac{\partial \Psi}{\partial \theta}\right) = 0\,.$$

If we look for separable solutions of the form $\Psi(r, \theta) = R(r)\Theta(\theta)$, then we must have

$$\frac{1}{R}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2 \frac{\mathrm{d}R}{\mathrm{d}r}\right) = -\frac{1}{\Theta \sin\theta}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\sin\theta \frac{\mathrm{d}\Theta}{\mathrm{d}\theta}\right) = \lambda\,,$$

where λ is a constant.

The equation of $\Theta(\theta)$ gives

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\sin\theta \frac{\mathrm{d}\Theta}{\mathrm{d}\theta}\right) = -\lambda \sin\theta \Theta\,.$$

By the substitution of $u = \cos\theta$, we have

$$-\sin\theta \frac{\mathrm{d}}{\mathrm{d}u}\left(-\sin^2\theta \frac{\mathrm{d}\Theta}{\mathrm{d}u}\right) = -\lambda \Theta \sin\theta\,,$$

which simplifies to

$$\frac{\mathrm{d}}{\mathrm{d}u}\left[(1-u^2)\frac{\mathrm{d}\Theta}{\mathrm{d}u}\right] + \lambda\Theta = 0\,.$$

This is the *Legendre's equation*.

Lemma 10.12. The Legendre's equation,

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[(1-x^2)\frac{\mathrm{d}y}{\mathrm{d}x}\right] + \lambda y = 0\,,$$

has solutions non-singular at $x = \pm 1$ only if $\lambda = \ell(\ell+1)$, where $\ell \in \mathbb{N}_0$. The resulting polynomial solutions are referred to as Legendre's polynomials, $P_\ell(x)$, with conventional normalisation of $P_\ell(1) = 1$.

Therefore, we obtain the Solutions

$$\Theta_\ell(\theta) = P_\ell(\cos\theta)\,,$$

where $P_\ell(x)$ are the Legendre polynomials.

The equation of $R(r)$ gives

$$r^2 R'' + 2rR' - \ell(\ell+1)R = 0\,,$$

which has the solutions

$$R_\ell(r) = Ar^\ell + Br^{-(\ell+1)}\,.$$

Theorem 10.13. The general solutions to Laplace's equation in spherical polar coordinates with axial symmetry are

$$\Psi(r, \theta) = \sum_{\ell=0}^{\infty} (A_\ell r^\ell + B_\ell r^{-(\ell+1)}) P_\ell(\cos\theta)\,.$$

Remark. In the non-axisymmetric case, a similar analysis would give an extra equation involving ϕ, and the Legendre polynomials would be replaced by the *associated Legendre polynomials*, which are the solutions of the *associated Legendre's equation*.

Laplace's equation in spherical polar coordinates is

$$\nabla^2\Psi = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\Psi}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\Psi}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2\Psi}{\partial\phi^2} = 0\,.$$

Consider solutions of the form $\Psi = R(r)\Theta(\theta)\Phi(\phi)$, substitution gives

$$\frac{\sin^2\theta}{R}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\frac{\mathrm{d}R}{\mathrm{d}r}\right) + \frac{\sin\theta}{\Theta}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\sin\theta\frac{\mathrm{d}\Theta}{\mathrm{d}\theta}\right) = -\frac{1}{\Phi}\frac{\mathrm{d}^2\Phi}{\mathrm{d}\phi^2} = \lambda_1$$

$$\implies \frac{1}{R}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\frac{\mathrm{d}R}{\mathrm{d}r}\right) = -\frac{1}{\Theta\sin\theta}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\sin\theta\frac{\mathrm{d}\Theta}{\mathrm{d}\theta}\right) + \frac{\lambda_1}{\sin^2\theta} = \lambda_2\,.$$

Solving this system of ODEs gives us the final solution.

Example. A hollow conducting sphere of radius a, centred at the origin, has its top hemisphere held at an electric potential $\Phi = V_0$. The bottom hemisphere, separated from the top by an insulating layer, is earthed ($\Phi = 0$). Find the electric potential Φ inside and outside the sphere.

The problem is to find the axisymmetric solution to

$$\nabla^2\Psi = 0$$

with boundary conditions

$$\Psi(a,\theta) = \begin{cases} V_0 & 0 < \theta < \frac{\pi}{2} \\ 0 & \frac{\pi}{2} < \theta < \pi\,. \end{cases}$$

Inside the sphere, Φ must be finite at $r = 0$, so $B_\ell = 0\ \forall\ell$ in the general solution.

$$\Phi(r,\theta) = \sum_{\ell=0}^{\infty} A_\ell r^\ell P_\ell(\cos\theta)\,.$$

Using the orthogonality of the Legendre polynomials and by writing $u = \cos\theta$, the boundary conditions give

$$\begin{aligned}\int_{-1}^{1}\Phi(a,\theta)P_m(u)\,\mathrm{d}u &= \int_{-1}^{1}\sum_{\ell=0}^{\infty}A_\ell a^\ell P_\ell(u)P_m(u)\,\mathrm{d}u \\ &= \sum_{\ell=0}^{\infty}A_\ell a^\ell\frac{2}{2m+1}\delta_{\ell m} \\ &= A_m a^m\frac{2}{2m+1}\,.\end{aligned}$$

$$\begin{aligned}\implies A_m a^m &= \frac{2m+1}{2}\int_{-1}^{1}\Phi(a,\theta)P_m(u)\,\mathrm{d}u \\ &= \frac{2m+1}{2}V_0\int_0^1 P_m(u)\,\mathrm{d}u\ .\end{aligned}$$

Therefore, inside the sphere,

$$\Phi(r,\theta) = \sum_{\ell=0}^{\infty}\frac{2\ell+1}{2a^\ell}V_0\int_0^1 P_\ell(u)\,\mathrm{d}u\,r^\ell P_\ell(\cos\theta)\,,$$

where the integrals can be evaluated.

Outside the sphere, we require that Φ is bounded as $r\to\infty$, so $A_\ell = 0\ \forall\ell$. Using a similar method, we may obtain that

$$\Phi(r,\theta) = \sum_{\ell=0}^{\infty}\frac{2\ell+1}{2a^{-(\ell+1)}}V_0\int_0^1 P_\ell(u)\,\mathrm{d}u\,r^{-(\ell+1)}P_\ell(\cos\theta)\,.$$

10.7 General Method

- In the case of an inhomogeneous equation, use the principle of superposition to seek a particular solution to reduce the equation to one that is homogeneous.
- Seek separable solutions to the homogeneous equation.
- In the case of inhomogeneous boundary conditions consider seeking a separable solution to reduce the boundary conditions to ones that are homogeneous.
- Use the boundary conditions to rule out certain separable solutions and to identify eigenvalues.
- Using the principle of superposition, seek a solution that is a sum of eigenfunctions.
- Determine unknown constants using the boundary conditions.

11 Cartesian Tensors

11.1 Vectors

A *vector* is a particular example of a *tensor*: a first-order tensor. Before discussing the general tensor, it will be useful to review vectors.

A vector has a physical meaning, direction and magnitude, independent of the coordinate system, but we can also think of a vector as a set of components, (v_1, v_2, v_3) with respect to some coordinate system. The components will generally be different in different coordinate systems, but the vector will be the same.

For any coordinate system with a basis of unit vectors $\{\mathbf{e}_i\}$, we can write a vector $\mathbf{v}$ as

$$\mathbf{v} = v_i \mathbf{e}_i \,.$$

In an orthonormal coordinate system, $\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$, so

$$v_i = \mathbf{e}_i \cdot \mathbf{v} \,.$$

In particular, we shall consider only Cartesian coordinate systems: orthonormal coordinate systems where $\{\mathbf{e}_i\}$ are independent of position.

11.1.1 Transformation of Basis

In a Cartesian coordinate system,

$$\mathbf{v} = v_i \mathbf{e}_i \,,$$

and in a different set of Cartesian coordinate system with basis vectors $\{\mathbf{e}'_i\}$,

$$\mathbf{v} = v'_i \mathbf{e}'_i \,, \text{ with } v'_i = \mathbf{e}'_i \cdot \mathbf{v} \,.$$

We also have

$$v'_i = \mathbf{e}'_i \cdot \mathbf{v} = \mathbf{e}'_i \cdot \mathbf{e}_j v_j = R_{ij} v_j \,,$$

where the matrix R with entries R_{ij} is defined by

$$R_{ij} = \mathbf{e}'_i \cdot \mathbf{e}_j \,.$$

Therefore,

$$v'_i = R_{ij} v_j \text{ or } \mathbf{v}' = \mathsf{R}\mathbf{v} \,,$$

where $\mathbf{v}'$ and $\mathbf{v}$ are column vectors with components v'_i and v_i. Note also that

$$\mathbf{e}'_i = (\mathbf{e}'_i \cdot \mathbf{e}_j)\mathbf{e}_j = R_{ij}\mathbf{e}_j \,.$$

Reversing the argument gives that

$$v_i = \mathbf{e}_i \cdot \mathbf{v} = \mathbf{e}_i \cdot v'_j \mathbf{e}'_j = (\mathbf{e}'_j \cdot \mathbf{e}_i) v'_j = R_{ji} v'_j = \mathsf{R}^{\mathrm{T}}_{ij} v'_j \,.$$

We therefore have

$$\mathbf{v} = \mathsf{R}^{\mathrm{T}}\mathsf{R}\mathbf{v} \,,$$

so

$$\mathsf{R}^{\mathrm{T}}\mathsf{R} = \mathsf{I} \,,$$

i.e. R is orthogonal.

Definition 11.1. A Cartesian *vector* $\mathbf{v}$ is a set of coefficients v_i, defined with respect to a set of orthonormal basis vectors $\{\mathbf{e}_i\}$, such that the coefficients v_i' with respect to another orthonormal basis $\{\mathbf{e}_i'\}$ are given by

$$v_i' = R_{ij} v_j \, .$$

Example. Consider the differential operator

$$\nabla = \mathbf{e}_i \frac{\partial}{\partial x_i} \, .$$

Since $x_i = R_{ji} x_j'$ and R_{ji} is constant, we have

$$\frac{\partial x_j}{\partial x_i'} = \frac{\partial R_{kj} x_k'}{\partial x_i'} = R_{kj} \frac{\partial x_k'}{\partial x_i'} = R_{kj} \delta_{ki} = R_{ij} \, ,$$

so using the chain rule,

$$\nabla_i' = \frac{\partial}{\partial x_i'} = \frac{\partial x_i}{\partial x_i'} \frac{\partial}{\partial x_j} = R_{ij} \frac{\partial}{\partial x_j} = R_{ij} \nabla_j \, .$$

Therefore, ∇ is a vector.

Remark. For more general straight line coordinate systems, $\mathsf{R}^{\mathrm{T}} \neq \mathsf{R}^{-1}$, and then ∇ is not a vector. One has then to distinguish between vectors and *co-vectors*, but there is no such distinction for Cartesian coordinates.

11.1.2 Axial-vectors

An orthogonal matrix has determinant ± 1. Those with $\det \mathsf{R} = 1$ are rotation matrices (*proper rotations*) and those with $\det \mathsf{R} = -1$ are the composition of a rotation with a reflection in some plane (*improper rotations*).

If we transform a basis $\{\mathbf{e}_i\}$ to $\{\mathbf{e}_i'\}$, and then to $\{\mathbf{e}_i''\}$, the components of a vector will then transform as

$$v_i' = R_{ij}^{(1)} v_j \, , \; v_i'' = R_{ij}^{(2)} v_j' \, ,$$

$$\implies v_i'' = R_{ij}^{(2)} R_{jk}^{(1)} v_k = R_{ik} v_k \, ,$$

where $\mathsf{R} = \mathsf{R}^{(2)} \mathsf{R}^{(1)}$ and $\det \mathsf{R} = \det \mathsf{R}^{(2)} \det \mathsf{R}^{(1)}$.

Remark. If both $\mathsf{R}^{(1)}$ and $\mathsf{R}^{(2)}$ are proper rotations, then so is the composite transformation $\mathsf{R} = \mathsf{R}^{(1)} \mathsf{R}^{(2)}$. We could then consistently restrict our attention to proper rotations. They form a *subgroup* $SO(3)$ of the orthogonal group $O(3)$.

This is not true for improper rotations.

Definition 11.2. A Cartesian *axial-vector*, or *pseudo-vector*, $\mathbf{a}$, is a set of coefficients a_i defined with respect to a set of orthonormal basis vectors $\{\mathbf{e}_i\}$ such that the coefficients a_i' with respect to another orthonormal basis $\{\mathbf{e}_i'\}$ are given by

$$a_i' = \det \mathsf{R} \, R_{ij} a_j \, .$$

When $\det \mathsf{R} = 1$, i.e. we don't change the handedness of the coordinate system, this is the same as for a vector. However, it differs in sign when $\det \mathsf{R} = -1$.

Example. An example of an axial-vector is the angular momentum

$$\mathbf{J} = \mathbf{r} \times \mathbf{p} \, ,$$

of a particle with momentum $\mathbf{p}$ at position $\mathbf{r}$. Compare the behaviour of $\mathbf{p}$ and $\mathbf{J}$ under a reflection in the xy-plane:

$$p_i' = \begin{pmatrix} p_x' \\ p_y' \\ p_z' \end{pmatrix} = \begin{pmatrix} p_x \\ p_y \\ -p_z \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} p_x \\ p_y \\ p_z \end{pmatrix} = R_{ij} p_j \,.$$

$$J_i' = \begin{pmatrix} J_x' \\ J_y' \\ J_z' \end{pmatrix} = \begin{pmatrix} -J_x \\ -J_y \\ J_z \end{pmatrix} = - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} J_x \\ J_y \\ J_z \end{pmatrix} = \det \mathsf{R}\, R_{ij} J_j \,.$$

Remark. We will see later that the cross product of two vectors is an axial-vector. Because ∇ is a vector, the curl of a vector is an axial-vector.

11.2 Tensors

Tensors are generalisations of vectors. They have some physical meaning independent of the coordinate basis and we can measure their components in some coordinate system.

Example. The electric current density $\mathbf{J}$ arising when an electric field $\mathbf{E}$ is applied to a material with isotropic conductivity *sigma* is given by

$$\mathbf{J} = \sigma \mathbf{E} \,.$$

However, the conductivity of a material may be anisotropic. Then we may have

$$J_i = \sigma_{ij} E_j \,,$$

where σ_{ij} are the components of the conductivity tensor. In a different basis,

$$J_i' = \sigma_{ij}' E_j' \,.$$

From the transformation law for vectors,

$$J_i' = R_{il} J_l = R_{il} \sigma_{lm} E_m = R_{il} \sigma_{lm} R^{\mathrm{T}}_{mj} E_j' \,.$$

Therefore, for arbitrary R and E_j',

$$\sigma_{ij}' = R_{il} \sigma_{lm} R^{\mathrm{T}}_{mj} = R_{il} R_{jm} \sigma_{lm} \,.$$

This can also be written in matrix notation as

$$\boldsymbol{\sigma}' = \mathsf{R} \boldsymbol{\sigma} \mathsf{R}^{\mathrm{T}} \,.$$

Having two indices, $\boldsymbol{\sigma}$ is a tensor of rank two.

Definition 11.3. A Cartesian *tensor* T of *rank* (*order*) n is a set of coefficients $T_{i_1 \ldots i_n}$ labelled by n indices defined with respect to a set of orthonormal basis vectors $\{\mathbf{e}_i\}$, and such that the coefficients with respect to another orthonormal basis $\{\mathbf{e}_i' : \mathbf{e}_i' = R_{ij}\mathbf{e}_i\}$ are given by the transformation law

$$T'_{i_1 \ldots i_n} = R_{i_1 j_1} \ldots R_{i_n j_n} T_{j_1 \ldots j_n} \,.$$

Remark. A tensor of rank zero is a scalar. A tensor of rank one is a vector.

Definition 11.4. A Cartesian *pseudo-tensor* E of *rank* n is a set of coefficients $E_{i_1 \cdots_n}$ labelled by n indices defined with respect to a set of orthonormal basis vectors $\{\mathbf{e}_i\}$, and such that the coefficients with respect to another orthonormal basis $\{\mathbf{e}_i' : \mathbf{e}_i' = R_{ij}\mathbf{e}_i\}$ are given by the transformation law

$$E'_{i_1 \ldots i_n} = \det \mathsf{R}\, R_{i_1 j_1} \ldots R_{i_n j_n} E_{j_1 \ldots j_n} \,.$$

When $\det \mathsf{R} = 1$, i.e. we don't change the handedness of the coordinate system, this is the same for a tensor. However, it differs in sign when $\det \mathsf{R} = -1$.

Remark. A pseudo-tensor of rank zero is a pseudo-scalar, a scalar that changes signs under reflections. e.g. $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$. A pseudo-tensor of rank one is a pseudo-vector.

11.2.1 Kronecker Delta, δ_{ij}, and Levi–Civita Symbol ε_{ijk}.

The Kronecker delta, δ_{ij}, is a rank two tensor defined without reference to a frame: its component should be the same in all frames.

$$\delta'_{ij} = \delta_{ij}$$

We can check that it does indeed transform in the way of a tensor.

$$\delta'_{ij} = R_{ip}R_{jq}\delta_{pq} = R_{ip}R_{jp} = R_{ip}R^{\mathrm{T}}_{pj} = \delta_{ij}\,,$$

as $\mathsf{R}\mathsf{R}^{\mathrm{T}} = \mathsf{I}$.

Likewise, the Levi–Civita symbol, ε_{ijk}, should be the same in all coordinate systems. It only has one independent non-zero component which we may choose to be $\varepsilon_{123} = 1$. If it transforms as a tensor, we would have

$$\varepsilon'_{123} = R_{1p}R_{2q}R_{3r}\varepsilon_{pqr}\,,$$

which is the definition of the determinant of a 3×3 matrix R. However, this would imply that $\varepsilon'_{123} = -1$ under a reflection, which is not true. The Levi–Civita symbol instead transforms as a pseudo-tensor:

$$\varepsilon'_{123} = \det\mathsf{R}\, R_{1p}R_{2q}R_{3r}\varepsilon_{pqr} = (\det\mathsf{R})^2 = 1 = \varepsilon_{123}\,,$$

and $\varepsilon'_{123} = 1$ in all frames as required. Therefore, ε_{ijk} is a pseudo-tensor of rank 3.

Remark. δ_{ij} and ε_{ijk} are examples of *isotropic tensors*.

11.2.2 Inertia Tensors

Consider a rigid body of variable density $\rho(\mathbf{x})$ within a volume $\mathcal{V}$ rotating with angular velocity ω. Then the angular momentum of an infinitesimal mass element $\mathrm{d}m = \rho(\mathbf{x})\,\mathrm{d}V$ is

$$\mathrm{d}m\,\mathbf{x}\times\mathbf{v} = \mathrm{d}m\,\mathbf{x}\times(\omega\times\mathbf{x}) = \mathrm{d}\mathbf{m}\,(|\mathbf{x}|^{\mathbf{2}}\omega - (\omega\cdot\mathbf{x})\mathbf{x})\,.$$

The total angular momentum $\mathbf{J}$ is given by

$$J_i = \int_{\mathcal{V}} \rho(\mathbf{x})(x_k x_k\omega_i - \omega_j x_j x_i)\,\mathrm{d}V = \int_{\mathcal{V}} (x_k x_k\delta_{ij} - x_j x_i)\omega_j\,\mathrm{d}V = I_{ij}\omega_j\,,$$

where I is the *inertia tensor* of the rigid body, given by

$$I_{ij} = \int_{\mathcal{V}} \rho(\mathbf{x})(x_k x_k\delta_{ij} - x_i x_j)\,\mathrm{d}V\,.$$

It can be checked that I is a rank 2 tensor.

11.2.3 Electric and Magnetic Susceptibility Tensor

Consider an electric insulator (*dielectric*) in an external electric field $\mathbf{E}$. No current flows because the charges are not free to move. However, the field does induce an electric polarisation density (dipole moment density) $\mathbf{P}$, given by

$$P_i = \epsilon_0\chi_{ij}E_j\,,$$

where χ is the *electric susceptibility* tensor. A related quantity is the *molecular polarisability*, α, that gives the dipole moment of a molecule induced by a local electric field.

$$p_i = \epsilon_0\alpha_{ij}E^{\mathrm{local}}_j\,.$$

The magnetic susceptibility is defined in a similar way:

$$M_i = \xi^M_{ij} H_j \,,$$

where $\mathbf{M}$ is the magnetisation (magnetic dipole moment per unit volume) and $\mathbf{H}$ is the magnetic field, both of which are pseudo-vectors.

Note that the electric susceptibility, molecular polarizability and magnetic susceptibility are all tensors.

11.2.4 Stress and Elastic Strain Tensors

In an elastic body, a local deformation due to applied forces (stresses) can be described by an elastic *strain tensor*

$$e_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right),$$

where $\mathbf{u}(\mathbf{x})$ is the displacement vector of a small volume element whose unstrained position is $\mathbf{x}$.

The elements T_{ij} of the stress tensor, T, are defined as the x_j component of forces acting on a plane perpendicular to the x_i axis. A generalisation of Hooke's law gives, for certain materials,

$$T_{ij} = C_{ijkl} e_{kl} \,,$$

where C is the rank 4 stiffness tensor.

11.2.5 Piezo-electric Strain Tensor

The application of stress to certain materials produces an electric polarisation that results in an electric field. The polarisation density $\mathbf{P}$ is related to the applied stress T by

$$P_i = D_{ijk} T_{ji} \,,$$

where D is the rank 3 piezo-electric strain tensor.

11.3 Properties of Tensors

11.3.1 Tensors as Maps

We first pick a set of coordinates, and the transformation law then requires that the tensor transforms nicely so that, ultimately, nothing depends on these coordinates. If that is the case, surely there should be a definition of a tensor that does not rely on coordinates at all.

Theorem 11.5. A tensor T of rank p is a multi-linear map that maps p vectors to a number in $\mathbb{R}$. $\mathsf{T} : V^p \to \mathbb{R}$.

$$\mathsf{T}(\mathbf{a}, \mathbf{b}, \ldots, \mathbf{c}) = T_{i_1 i_2 \ldots i_p} a_{i_1} b_{i_2} \ldots c_{i_p} \in \mathbb{R} \,.$$

Multi-linearity means that T is linear in each of the entries individually.

Proof. A tensor defined so does transform as a tensor.

$$\begin{aligned}
\mathsf{T}(\mathbf{a}, \mathbf{b}, \ldots, \mathbf{c}) &= T'_{i_1 i_2 \ldots i_p} a'_{i_1} b'_{i_2} \ldots c'_{i_p} \\
&= (R_{i_1 j_1} R_{i_2 j_2} \ldots R_{i_p j_p} T'_{i_1 i_2 \ldots i_p})(R_{i_1 k_1} a'_{k_1})(R_{i_2 k_2} b'_{k_1}) \ldots (R_{i_p k_p} c'_{k_p}) \\
&= T_{j_1 j_2 \ldots j_p} a_{j_1} b_{j_2} \ldots c_{j_p} \,,
\end{aligned}$$

since R is orthogonal. □

Rather than thinking of a tensor as a map from many vectors to $\mathbb{R}$, it is often more convenient to think of it as a map from some lower-rank tensor to another.

$$a_i = T_{ij_1\dots j_{p-1}} b_{j_1} \dots c_{j_{p-1}} .$$

This is the way that tensors typically arise in physics or applied mathematics, where the most common example is a rank 2 tensor, defined as a map from one vector to another

$$\mathbf{u} = \mathsf{T}\mathbf{v} \implies u_i = T_{ij} v_j .$$

Second-order tensors are usually given the name matrix but for the equation $\mathbf{u} = \mathsf{T}\mathbf{v}$ to make sense, T must transform as a tensor. This is inherited from the transformation rules of the vectors: $u_i' = R_{ij} u_j$ and $v_i' = R_{ij} v_j$, giving

$$u_i' = T_{ij}' v' j \quad \text{with} \quad T_{ij}' = R_{ik} R_{jl} T_{kl} .$$

Written as a matrix equation, this is

$$\mathsf{T}' = \mathsf{R}\mathsf{T}\mathsf{R}^{\mathrm{T}} .$$

11.3.2 Tensor Operations

Definition 11.6. If A and B are tensors or rank n, then we define sum of two tensors, $\mathsf{C} = \mathsf{A} + \mathsf{B}$, to be

$$C_{i_1\dots i_n} \coloneqq A_{i_1\dots i_n} + B_{i_1\dots i_n} .$$

Proposition 11.7. The sum of the scalar multiples of two tensors of rank n is a tensor of rank n.

Proof.

$$\begin{aligned} C'_{i_1\dots i_n} &= \alpha' A'_{i_1\dots i_n} + \beta' B'_{i_1\dots i_n} \\ &= \alpha R_{i_1 j_1} \dots R_{i_n j_n} A_{j_1,\dots,j_n} + \beta R_{i_1 j_1} \dots R_{i_n j_n} B_{j_1\dots j_n} \\ &= R_{i_1 j_1} \dots R_{i_n j_n} A_{j_1\dots j_n} (\alpha A_{i_1\dots i_n} + \beta B_{i_1\dots i_n}) \\ &= R_{i_1 j_1} \dots R_{i_n j_n} C_{i_1\dots i_n} \end{aligned}$$

do transforms as a tensor. □

Definition 11.8. The *tensor product* (*outer product*) of two tensors, A and B, of rank n and m respectively, is defined as

$$C_{i_1\dots i_n i_{n+1}\dots i_{n+m}} \coloneqq A_{i_1\dots i_n} B_{i_{n+1}\dots i_{n+m}} .$$

This is denoted as

$$\mathsf{C} = \mathsf{A} \otimes \mathsf{B} .$$

Proposition 11.9. If A is a tensor of rank n and B is a tensor of rank m, then their tensor product $\mathsf{C} = \mathsf{A} \otimes \mathsf{B}$ is a tensor of rank $n + m$. If instead A is a tensor of rank n and B is a pseudo-tensor of rank m, then their tensor product $\mathsf{C} = \mathsf{A} \otimes \mathsf{B}$ is a pseudo-tensor of rank $n + m$.

Remark. We can write a tensor as

$$\mathsf{T} = T_{i_1 i_2 \dots i_n} \otimes \mathbf{e}_{i_1} \otimes \mathbf{e}_{i_2} \cdots \otimes \mathbf{e}_{i_n} .$$

Proposition 11.10 (Inner product). If T is a tensor of rank n, then if we contract two indices (set equal and sum over), we get a tensor of rank $n - 2$. If $i_k = i_l$, then

$$T_{i_1\dots i_{k-1} i_k i_{k+1}\dots i_{l-1} i_k i_{l+1}\dots i_n} = S_{i_1\dots i_{k-1} i_{k+1}\dots i_{l-1} i_{l+1}\dots i_n} .$$

Proof.

$$\begin{aligned} S'_{\alpha_1\ldots\alpha_{k-1}\alpha_{k+1}\ldots\alpha_{l-1}\alpha_{l+1}\ldots\alpha_{n-2}} &= T'_{\alpha_1\ldots\alpha_{k-1}\alpha_k\alpha_{k+1}\ldots\alpha_{l-1}\alpha_k\alpha_{l+1}\ldots\alpha_n} \\ &= R_{\alpha_1 i_1}\ldots R_{\alpha_k i_k}\ldots R_{\alpha_k i_l}\ldots R_{\alpha_n i_n} T_{i_1\ldots i_k\ldots i_l\ldots i_n} \\ &= (R_{\alpha_k i_k} R_{\alpha_k i_l}) R_{\alpha_1 i_1}\ldots R_{\alpha_n i_n} T_{i_1\ldots i_k\ldots i_l\ldots i_n} \\ &= \delta_{i_k i_l} R_{\alpha_1 i_1}\ldots R_{\alpha_n i_n} T_{i_1\ldots i_k\ldots i_l\ldots i_n} \\ &= R_{\alpha_1 i_1}\ldots R_{\alpha_n i_n} S_{i_1\ldots i_n}\,. \end{aligned}$$

□

Examples.

(i) If T is a tensor of rank two, then from the definition of the inner product, $T_{11} = \mathrm{tr}(\mathsf{T})$ is a scalar.

(ii) Suppose $\mathbf{u}$ and $\mathbf{v}$ are vectors, then from the definition of the outer product, $T_{ij} = u_i v_j$ is a tensor of rank two. It follows that $T_{ii} = u_i v_i = \mathbf{u}\cdot\mathbf{v}$ is a scalar.

(iii) If A is a rank-two tensor and $\mathbf{u}$ is a vector, then $v_i = A_{ij}u_j$ is a vector, since $2+1-2=1$.

(iv) If A and B are tensors of rank two, then $C_{ij} = A_{ik}B_{kj}$ is a tensor of rank two, since $2+2-2=2$.

(v) The cross product of two vectors,

$$\mathbf{u}\times\mathbf{v} = \varepsilon_{ijk}\mathbf{e}_i u_j v_k$$

is an axial-vector, since ε_{ijk} is a rank-three pseudo-tensor and $3+2+1+1-6=1$.

11.3.3 Symmetric and Anti-symmetric Tensors

Definition 11.11. A tensor T is *symmetric* in a pair of indices α and β if

$$T_{\ldots\alpha\ldots\beta\ldots} = T_{\ldots\beta\ldots\alpha\ldots}\,,$$

and is *anti-symmetric* in α and β if

$$T_{\ldots\alpha\ldots\beta\ldots} = -T_{\ldots\beta\ldots\alpha\ldots}\,.$$

A tensor that is (anti-)symmetric in all pairs of indices is said to be *totally (anti-)symmetric*.

Proposition 11.12. Symmetry and anti-symmetry are invariant under a change of coordinate.

Proof. If $T_{ijk\ldots}$ is symmetric in i and j, then

$$\begin{aligned} T'_{ijk\ldots} &= R_{ip}R_{jq}R_{kr}\ldots T_{pqr\ldots} \\ &= R_{ip}R_{jq}R_{kr}\ldots T_{qpr\ldots} \\ &= R_{jq}R_{ip}R_{kr}\ldots T_{qpr\ldots} \\ &= T'_{jik\ldots}\,. \end{aligned}$$

□

Proposition 11.13. If $S_{ijk..}$ is symmetric in, say, i and j, and $A_{pqr\ldots}$ is anti-symmetric in, say, p and q, then

$$S_{ijk\ldots}A_{ijr\ldots} = 0\,.$$

Proof.

$$\begin{aligned} S_{ijk\ldots}A_{ijr\ldots} &= -S_{jik\ldots}A_{jir\ldots} \\ &= -S_{ijk\ldots}A_{ijr\ldots}\,. \end{aligned}$$

$$\implies S_{ijk\ldots}A_{ijr\ldots} = 0\,.$$

□

Remarks.

- The Kronecker delta δ_{ij} is symmetric, and the Levi–Civita symbol ε_{ijk} is anti-symmetric, in any pair of indices.
- The inertia tensor and strain tensors are symmetric from their definitions.
- In most situations, but not all, the stress tensor is also symmetric. The conductivity tensor and susceptibility tensors are usually symmetric.

11.4 Rank Two Tensors

Since a rank two tensor only has one pair of indices, if it is symmetric or anti-symmetric in these indices we can refer to the tensor as symmetric or anti-symmetric. The matrices corresponding to symmetric and anti-symmetric rank two tensors are symmetric and anti-symmetric respectively.

$$\mathsf{S}^{\mathrm{T}} = \mathsf{S} \qquad \text{if } \mathsf{S} \text{ is symmetric;}$$
$$\mathsf{A}^{\mathrm{T}} = -\mathsf{A} \qquad \text{if } \mathsf{A} \text{ is anti-symmetric.}$$

Proposition 11.14 (Symmetric / anti-symmetric decomposition). Any rank two tensor T_{ij} can be uniquely decomposed into the sum of a symmetric and an anti-symmetric tensor:

$$T_{ij} = S_{ij} + A_{ij}\,, \text{where } S_{ij} = \frac{1}{2}(T_{ij} + T_{ji}) \text{ and } A_{ij} = \frac{1}{2}(T_{ij} - T_{ji})\,.$$

Proposition 11.15 (Duality). An anti-symmetric 2-tensor A is equivalent to an axial-vector

$$\omega_k = \frac{1}{2}\varepsilon_{klm}A_{lm}\,.$$

$\boldsymbol{\omega}$ is its *dual vector*, such that

$$\mathsf{A}\mathbf{v} = \boldsymbol{\omega} \times \mathbf{v}\,.$$

Proof.

$$\begin{aligned}\varepsilon_{ijk}\omega_k &= \frac{1}{2}\varepsilon_{ijk}\varepsilon_{klm}A_{lm} = \frac{1}{2}(\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl})A_{lm} \\ &= \frac{1}{2}(A_{ij} - A_{ji}) = A_{ij}\,,\end{aligned}$$

So we must have $\mathsf{A}\mathbf{v} = \boldsymbol{\omega} \times \mathbf{v}$. Since ε_{ijk} is a pseudo-tensor, $\boldsymbol{\omega}$ must be an axial-vector. □

Remark. An anti-symmetric 2-tensor has three independent components, and can be written as

$$A_{ij} = \varepsilon_{ijk}\omega_k = \begin{pmatrix} 0 & \omega_3 & -\omega_2 \\ -\omega_3 & 0 & \omega_1 \\ \omega_2 & -\omega_1 & 0 \end{pmatrix}$$

in terms of its dual vector.

Proposition 11.16 (Symmetric tensor decomposition). Any symmetric rank two tensor, S, has a unique decomposition in terms of a symmetric traceless tensor, $\tilde{\mathsf{S}}$, and a scalar multiple of the identity I.

$$\mathsf{S} = \tilde{\mathsf{S}} + \frac{1}{3}\operatorname{tr}(\mathsf{S})\mathsf{I} = \tilde{\mathsf{S}} + \frac{1}{3}Q\mathsf{I}\,,$$

where Q is the trace of S.

Proof.

$$\operatorname{tr}(\tilde{\mathsf{S}}) = \operatorname{tr}(\mathsf{S}) - \frac{1}{3}\operatorname{tr}(\mathsf{S})\operatorname{tr}(\mathsf{I}) = 0\,.$$

$\tilde{\mathsf{S}}$ is traceless and clearly symmetric. Using the transformation law of tensors, it can be shown further that $\operatorname{tr}(\tilde{\mathsf{S}})$ is traceless in any Cartesian coordinate system. □

Corollary. Any 3×3 matrix can be decomposed as

$$T_{ij} = \tilde{S}_{ij} + \varepsilon_{ijk}\boldsymbol{\omega}_k + \frac{1}{3}\delta_{ij}Q\,.$$

Example. An elastic body is subjected to a simple shear so that the displacement $\mathbf{u}(\mathbf{x})$ at position $\mathbf{x} = (x, y, z)$ is given by $\mathbf{u} = (\gamma y, 0, 0)$ for some constant γ.

We can decompose $\frac{\partial u_i}{\partial x_j}$ into symmetric and anti-symmetric parts,

$$\frac{\partial u_i}{\partial x_j} = \begin{pmatrix} 0 & \gamma & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{2}\gamma & 0 \\ \frac{1}{2}\gamma & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & \frac{1}{2}\gamma & 0 \\ -\frac{1}{2}\gamma & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where the symmetric part is the strain tensor e_{ij}, and the anti-symmetric part can be written as $\varepsilon_{ijk}\omega_k$, where $\boldsymbol{\omega} = (0, 0, \frac{1}{2}\gamma)$.

This also corresponds to writing

$$\mathbf{u} = \left(\frac{1}{2}\gamma y, \frac{1}{2}\gamma x, 0\right) + \left(\frac{1}{2}\gamma y, -\frac{1}{2}\gamma x, 0\right).$$

The first term corresponds to a stretch at 45° to the x and y axis, while the second term is a rotation.

11.4.1 Diagonalisation of Symmetric Rank Two Tensor

Suppose that S_{ij} is a symmetric rank two tensor with components relative to a coordinate system with basis vectors $\{\mathbf{e}_i\}$. The matrix of components, S, is symmetric, i.e. Hermitian. We know that S has real eigenvalues λ_i and orthonormal vectors $\mathbf{e}'_i$, which can be rearranged in a right-handed set.

Now transform from the coordinate system with the basis $\mathbf{e}_i$ to $\mathbf{e}'_i$. The transformation matrix R is the matrix with the components of $\mathbf{e}'_i$ as rows, i.e.

$$R_{ij} = \mathbf{e}'_i \cdot \mathbf{e}_j\,.$$

Hence,

$$\mathsf{S}\mathsf{R}^{\mathrm{T}} = \mathsf{S}(\mathbf{e}'_1|\mathbf{e}'_2|\mathbf{e}'_3) = (\lambda_1\mathbf{e}'_1|\lambda_2\mathbf{e}'_2|\lambda_3\mathbf{e}'_3)\,.$$

Lemma 11.17. We can diagonalise a symmetric matrix S with some appropriate rotation of the coordinate axes.

Proof.

$$\begin{aligned} \mathsf{S}' &= \mathsf{R}\mathsf{S}\mathsf{R}^{\mathrm{T}} \\ &= \begin{pmatrix} — & \mathbf{e}_1'^{\,\mathrm{T}} & — \\ — & \mathbf{e}_2'^{\,\mathrm{T}} & — \\ — & \mathbf{e}_3'^{\,\mathrm{T}} & — \end{pmatrix} \begin{pmatrix} | & | & | \\ \lambda_1\mathbf{e}'_1 & \lambda_2\mathbf{e}'_2 & \lambda_3\mathbf{e}'_3 \\ | & | & | \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}. \end{aligned}$$

□

Definition 11.18. The eigenvalues λ_i of a symmetric real tensor are known as its *principal values*.

Remark. The principal values do not depend on the frame: they are the properties of tensors, not the coordinate system.

Definition 11.19. The Cartesian coordinate axes for which S_{ij} is diagonal, i.e. the eigenvectors of S_{ij}, are known as the *principal axes*.

11.5 Isotropic Tensors

Definition 11.20. A tensor or a pseudo-tensor is *invariant* under some rotation R if its components are the same under the rotation, i.e.

$$T'_{i_1\dots i_n} = R_{i_1 j_1}\dots R_{i_n j_n} T_{i_1\dots i_n} = T_{i_1\dots i_n}\,.$$

Definition 11.21. If a tensor or a pseudo-tensor is invariant under any rotation, then it is called *isotropic*.

Remarks.

- Isotropic tensors do not have any 'preferred' direction. For example, the conductivity of an isotropic medium is an isotropic tensor.
- *Isotropic* means the same in all directions, and *homogeneous* means the same at all points in space.

Theorem 11.22. Consider the non-zero isotropic tensors in $\mathbb{R}^3$.

(i) All tensors of rank zero (scalars) are isotropic.

(ii) There are no rank one isotropic tensors.

(iii) The only second order isotropic tensors are scalar multiples of δ_{ij}.

(iv) The only third order isotropic tensors are the scalar multiples of ε_{ijk}.

Proof. The idea is simply to look at how tensors transform under a bunch of specific rotations by π or $\pi/2$ about certain axes.

(i) Trivial.

(ii) Consider a tensor of rank 1, so that

$$T'_i = R_{ij}T_j\,,\ \text{where } \mathsf{R} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Requiring $T'_i = T_i$ gives $T_1 = T_2 = 0$. A similar argument, using a different R, also gives $T_3 = 0$.

(iii) For a tensor of rank 2, consider the transformation

$$T'_{ij} = \tilde{R}_{ik}\tilde{R}_{jl}T_{kl}\,,\ \text{where } \tilde{\mathsf{R}} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The rotation gives $T'_{13} = T_{23}$ and $T'_{23} = -T_{13}$ so if $T'_{ij} = T_{ij}$, we must have $T_{13} = T_{23} = 0$. Meanwhile $T'_{11} = T_{22}$. Similar arguments tell us that all off-diagonal elements must vanish and all diagonal elements must be equal: $T_{11} = T_{22} = T_{33} = \lambda$ for some constant λ. Hence $T_{ij} = \lambda\delta_{ij}$.

(iv) For a rank 3 tensor we have

$$T'_{ijk} = R_{il}R_{jp}R_{kq}T_{lpq}\,,\ \text{where } \mathsf{R} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

We find that $T'_{133} = -T_{133}$ and $T'_{111} = -T_{111}$. Similar arguments show that an isotropic tensor must have $T_{ijk} = 0$ unless i, j, k are all distinct.

Meanwhile, if we pick

$$\tilde{\mathsf{R}} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} ,$$

then we get $T'_{123} = -T_{213}$. We end up with the result that T_{ijk} is isotropic if and only if $T_{ijk} = \mu\varepsilon_{ijk}$ for some constant μ. □

11.5.1 Application to Integrals

Example. Integration over a sphere

Consider the integral over the sphere

$$\mathbf{X} = \int_{r\le a} \mathbf{x}\rho(r)\,\mathrm{d}V \text{ or } X_i = \int_{r\le a} x_i\rho(r)\,\mathrm{d}V .$$

Relabelling the integration variables we can write

$$X_i = \int_{r'\le a} x'_i\rho(r')\,\mathrm{d}V' .$$

Now make the substitution $x'_i = R_{ij}x_j$ for a rotation matrix R. The integration volume and the function ρ is spherically symmetric, so $\rho(r') = \rho(r)$ and $\mathrm{d}V' = \mathrm{d}V$, and since R is an orthogonal matrix,

$$X_i = R_{ij}\int_{r\le a} x_j\rho(r)\,\mathrm{d}V = R_{ij}X_j = X'_i .$$

From the definition of a vector, Definition 11.1, this equation says that $X_i = X'_i$, i.e. $\mathbf{X}$ is an isotropic vector. However, the only isotropic vector is the zero vector, so we deduced that

$$\mathbf{X} = 0 .$$

Example. A rank 2 tensor integral over a sphere

Consider the integral

$$K_{ij} = \int_{r\le a} x_ix_j\rho(r)\,\mathrm{d}V .$$

A similar argument as above shows that, using the definition of a rank two tensor (Definition 11.3),

$$K_{ij} = R_{ik}R_{jl}K_{kl} = K'_{ij} ,$$

which means that K is an isotropic tensor, and so

$$K_{ij} = \lambda\delta_{ij}$$

for some scalar λ. Take the trace to deduce that

$$\begin{aligned} \lambda = \frac{1}{3}\operatorname{tr}(\mathsf{K}) &= \frac{1}{3}\int_{r\le a} x_ix_i\rho(r)\,\mathrm{d}V \\ &= \frac{1}{3}\int_{r\le a} r^2\rho(r)\,\mathrm{d}V , \end{aligned}$$

and hence,

$$K_{ij} = \left(\int_{r\le a}\frac{1}{3}r^2\rho(r)\,\mathrm{d}V\right)\delta_{ij} .$$

Example. A rank two tensor integral over all space

Consider the integral,

$$K_{ij} = \int_{\mathcal{V}} x_i x_j e^{-r^2} \, \mathrm{d}V \ ,$$

where $\mathcal{V}$ is all of space. Using the result derived above,

$$\begin{aligned} K_{ij} &= \left(\int_{\mathcal{V}} \frac{1}{3} r^2 e^{-r^2} \, \mathrm{d}V \right) \delta_{ij} \\ &= \frac{1}{2} \sqrt{\pi^3} \delta_{ij} \, . \end{aligned}$$

Example. An inertia tensor

Consider the inertia tensor of a sphere of radius a and mass M. The density per unit volume is $\rho = 3M/4\pi a^3$. Hence we have

$$\begin{aligned} I_{ij} &= \rho \int_{\mathcal{V}} (x_k x_k \delta_{ij} - x_i x_j) \, \mathrm{d}V \\ &= \frac{3M}{4\pi a^2} \left(\int_{\mathcal{V}} r^2 - \frac{1}{3} r^2 \, \mathrm{d}V \right) \delta_{ij} \\ &= \frac{M}{2\pi a^3} \left(\int_{r \leq a} r^2 \, \mathrm{d}V \right) \delta_{ij} \\ &= \frac{2M}{a^3} \int_0^a r^4 \, \mathrm{d}r \, \delta_{ij} \\ &= \frac{2}{5} M a^2 \delta_{ij} \, . \end{aligned}$$

11.6 Tensor Fields

Definition 11.23. A *tensor field* is the assignment of a tensor T to every point $\mathbf{x}$

$$\mathsf{T} : \mathbb{R}^n \to \mathbb{R}^m \, .$$

Remark. A *scalar field* is a rank 0 tensor field, and a *vector field* is a rank 1 tensor field.

11.6.1 Tensor Differential Operators

Given a tensor field, we can always construct higher rank tensors by taking derivatives.

Proposition 11.24. For a scalar field $\phi(\mathbf{x})$, its gradient $\boldsymbol{\nabla}\phi$ is a vector field.

Proof. For two orthonormal bases $\{\mathbf{e}_i\}$ and $\{\mathbf{e}'_i\}$, any vector can be decomposed as

$$\mathbf{v} = v_i \mathbf{e}_i = v'_i \mathbf{e}'_i \, .$$

If we expand $\mathbf{x}$ in this way,

$$\mathbf{x} = x_i \mathbf{e}_i = x'_i \mathbf{e}'_i \implies x_i = (\mathbf{e}_i \cdot \mathbf{e}'_j) x'_j \implies \frac{\partial x_i}{\partial x'_j} = \mathbf{e}_i \cdot \mathbf{e}'_j \, .$$

Here $\mathbf{e}_i \cdot \mathbf{e}'_j$ is the rotation matrix that takes us from one basis to the other. Meanwhile, we can always expand one set of basis vectors in terms of the other:

$$\mathbf{e}_i = (\mathbf{e}_i \cdot \mathbf{e}'_j) \mathbf{e}'_j = \frac{\partial x_i}{\partial x'_j} \mathbf{e}'_j \, .$$

This tells us that we could equally as well write the gradient as

$$\boldsymbol{\nabla}\phi = \frac{\partial\phi}{\partial x_i}\mathbf{e}_i = \frac{\partial\phi}{\partial x_i}\frac{\partial x_i}{\partial x'_j}\mathbf{e}'_j = \frac{\partial\phi}{\partial x'_j}\mathbf{e}'_j\,.$$

If you work in a different primed basis, then you have the same definition of gradient. The components transform correctly under a rotation, so $\boldsymbol{\nabla}\phi$ is indeed a vector. □

We can extend the result above to any suitably smooth tensor field. We can differentiate it by any number of times to get a new tensor field of higher order.

Proposition 11.25. If a suitably smooth tensor field $\mathsf{T}(\mathbf{x})$ of rank p is differentiated q times, we will get a new tensor field of rank $p+q$:

$$X_{i_1\ldots i_q j_1\ldots j_p} = \frac{\partial}{\partial x_{i_1}}\cdots\frac{\partial}{\partial x_{i_q}}T_{j_1\ldots j_p}(\mathbf{x})\,.$$

Proof. In a new basis, we have $x'_i = R_{ij}x_j$, where $R_{ij} = \mathbf{e}'_i\cdot\mathbf{e}_j$, and so

$$\frac{\partial x'_i}{\partial x_j} = R_{ij} \implies \frac{\partial}{\partial x'_i} = \frac{\partial x_j}{\partial x'_i}\frac{\partial}{\partial x_j} = R_{ij}\frac{\partial}{\partial x_j}\,.$$

$\mathsf{X}(\mathbf{x})$ is indeed a tensor field. □

Examples.

(i) The divergence of a vector field $\mathbf{F}$ is a scalar field $\frac{\partial}{\partial x_i}F_i$, since the contraction of a rank 2 tensor is a zeroth tensor.

(ii) The curl of a vector field $\mathbf{F}$ is an axial-vector field, $\varepsilon_{ijk}\frac{\partial}{\partial x_j}F_k$, since the double contraction of a rank 5 pseudo-tensor field is a pseudo-vector field.

(iii) The Laplacian of a scalar field, $\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_i}\Phi$, is a scalar field.

(iv) The derivative of a rank 2 tensor, σ, is a rank 3 tensor field, $\frac{\partial}{\partial x_i}\sigma_{jk}$.

We can implement any of the tensorial manipulations that we met previously for tensor fields. Consider the rank 2 tensor field

$$T_{ij}(\mathbf{x}) = \frac{\partial F_i}{\partial x_j}$$

defined for a vector field $\mathbf{F}(\mathbf{x})$. We have seen that any rank 2 tensor can be decomposed into various pieces. There is an anti-symmetric piece

$$A_{ij}(\mathbf{x}) = \varepsilon_{ijk}B_k(\mathbf{x})\,,\ \text{where } B_k = \frac{1}{2}\varepsilon_{ijk}\frac{\partial F_i}{\partial x_j} = -\frac{1}{2}(\boldsymbol{\nabla}\times\mathbf{F})_k\,,$$

a trace piece

$$Q = \frac{\partial F_i}{\partial x_i} = \boldsymbol{\nabla}\cdot\mathbf{F}\,,$$

and a symmetric traceless piece

$$\tilde{S}_{ij}(\mathbf{x}) = \frac{1}{2}\left(\frac{\partial F_i}{\partial x_j} + \frac{\partial F_j}{\partial x_i}\right) - \frac{1}{3}\boldsymbol{\nabla}\cdot\mathbf{F}\,.$$

11.7 A Unification of the Integral Theorems (Non-examinable)

It is obvious that the three integral theorems in the vector calculus are closely related. We will show how they can be presented in a unified framework in this section.

11.7.1 Integrating in Higher Dimensions

This unified framework will give us integral theorems in any dimension $\mathbb{R}^n$. Note that the divergence theorem already holds in any $\mathbb{R}^n$. However, Stokes' theorem is restricted to surfaces in $\mathbb{R}^3$ since the cross product is only defined in $\mathbb{R}^3$. Therefore, we must first extend the cross product into higher dimensions.

Definition 11.26. The *Levi–Civita symbol* in n dimensions is defined as

$$\varepsilon_{a_1 a_2 \dots a_n} \coloneqq \begin{cases} 1 & \text{if } (a_1, a_2, \dots, a_n) \text{ is an even permutation of } (1, 2, \dots, n) \\ -1 & \text{if } (a_1, a_2, \dots, a_n) \text{ is an odd permutation of } (1, 2, \dots, n) \\ 0 & \text{otherwise}\,. \end{cases}$$

Using this, we can define the cross product in any dimension $\mathbb{R}^n$.

Definition 11.27. A *cross product* is a map from two vectors in $\mathbb{R}^n$ to an anti-symmetric rank $(n-2)$ tensor.

$$(\mathbf{a} \times \mathbf{b})_{i_1 \dots i_{n-2}} = \varepsilon_{i_1 \dots i_n} a_{i_{n-1}} b_{i_n}\,.$$

Remark. The Levi–Civita symbol can be thought of as a map from anti-symmetric rank p tensors to anti-symmetric rank $(n-p)$ tensor by contracting indices,

$$\varepsilon : T_{i_1 \dots i_p} \mapsto \frac{1}{(n-p)!} \varepsilon_{i_1 \dots i_n} T_{i_{n-p+1} \dots i_n}\,.$$

This map is known as a *Hodge dual*.

Next, we need to think about what this has to do with integration. We have found two natural ways to integrate vector fields in $\mathbb{R}^3$.

- *Line integral.*
$$\int_{\mathcal{C}} \mathbf{F} \cdot \mathrm{d}\mathbf{x}\,. \tag{$\dagger$}$$
This captures the component of the vector field tangent to the line. We can perform this procedure in any dimension $\mathbf{R}^n$.
- *Surface integral.*
$$\int_{\mathcal{S}} \mathbf{F} \cdot \mathrm{d}\mathbf{S}\,, \tag{$\dagger\dagger$}$$
where $\mathrm{d}\mathbf{S}$ points in the direction normal to the surface. The integration captures the component of the vector field normal to the surface and only makes sense in $\mathbb{R}^3$. This is because it is only in $\mathbb{R}^3$ that a two-dimensional surface has a unique normal.

Let us expand $\mathrm{d}\mathbf{S}$ to clearly see what is going on. For a parameterised surface $\mathbf{x}(u, v)$, the vector area element is

$$\mathrm{d}\mathbf{S} = \frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v}\,\mathrm{d}u\,\mathrm{d}v\,,$$

or, in component forms,

$$\mathrm{d}S_i = \varepsilon_{ijk} \frac{\partial x_j}{\partial u} \frac{\partial x_k}{\partial v}\,\mathrm{d}u\,\mathrm{d}v\,.$$

Rather than thinking of equation ($\dagger$) as the integral of a vector field projected normal to the surface, instead think of it as the integral of an anti-symmetric rank 2 tensor $F_{ij} = \varepsilon_{ijk} F_k$ integrated tangent to the surface. We then have

$$\int_{\mathcal{S}} \mathbf{F} \cdot \mathrm{d}\mathbf{S} = \int_{\mathcal{S}} F_{ij}\,\mathrm{d}S_{ij}\,,$$

where

$$\mathrm{d}S_{ij} = \frac{1}{2}\left(\frac{\partial x_i}{\partial u}\frac{\partial x_j}{\partial v} - \frac{\partial x_i}{\partial v}\frac{\partial x_j}{\partial u}\right)\mathrm{d}u\,\mathrm{d}v\ .$$

This is the same equation as before, just with the epsilon symbol viewed as part of the integrand F_{ij} rather than as part of the measure $\mathrm{d}S_i$. Note that we have retained the anti-symmetry of the area element $\mathrm{d}S_{ij}$ that was inherent in our original cross product definition of $\mathrm{d}\mathbf{S}$. Strictly speaking this is not necessary because we are contracting with anti-symmetric indices in F_{ij}, but it turns out that it is best to think of both objects F_{ij} and $\mathrm{d}S_{ij}$ as individually anti-symmetric.

This new perspective suggests a way to generalise to higher dimensions. In the line integral (†) we are integrating a vector field over a line. In the surface integral (††), we are really integrating an anti-symmetric 2-tensor over a surface. The key idea is that one can integrate a totally anti-symmetric p-tensor over a p-dimensional subspace.

- *Generalisation of line integral* (†). Let $\Omega \subset \mathbb{R}^n$ be a p-dimensional subspace. We can then integrate an anti-symmetric p-tensor over Ω

$$\int_\Omega T_{i_1\ldots i_p}\,\mathrm{d}S_{i_1\ldots i_p}\ .$$

- *Generalisation of the surface integral* (††). First map the anti-symmetric p-tensor to an anti-symmetric $(n-p)$-tensor using the Hodge dual, then this can be integrated over an $(n-p)$-dimensional subspace $\tilde{\Omega} \subset \mathbb{R}^n$:

$$\int_{\tilde{\Omega}} T_{i_1\ldots i_p}\varepsilon_{i_1\ldots i_p j_1\ldots j_p}\,\mathrm{d}\tilde{S}_{j_1\ldots j_{n-p}}\ .$$

11.7.2 Differentiating Anti-symmetric Tensors

We have already noted in Proposition 11.25 that we can differentiate a p-tensor once to get a tensor of rank $p+1$, but in general differentiating loses the anti-symmetry property. There is a way to restore it so that when we differentiate a totally anti-symmetric p tensor, we end up with a totally anti-symmetric $(p+1)$-tensor.

First consider a scalar field. This is trivial since its gradient is an vector field by Proposition 11.24, and this is automatically 'anti-symmetric' because there is nothing to anti-symmetrise.

If we are given a vector field $\mathbf{F}$, we can differentiate and then anti-symmetrise using the operation

$$(\mathcal{D}\mathbf{F})_{ij} \coloneqq \frac{1}{2}\left(\frac{\partial F_i}{\partial x_j} - \frac{\partial F_j}{\partial x_i}\right).$$

Remark. This is formally known as the *differential form* and is usually written as $\mathrm{d}F$, but the notation $\mathrm{d}F$ is loaded with all sorts of other connotations which are best ignored at this stage. Hence we will temporarily use the notation $\mathcal{D}\mathbf{F}$ here.

In $\mathbb{R}^3$, this anti-symmetric differentiation is equivalent to the curl using the Hodge map

$$(\boldsymbol{\nabla}\times\mathbf{F})_i = \varepsilon_{ijk}(\mathcal{D}\mathbf{F})_{jk}\,.$$

Now we can extend this definition to any anti-symmetric p-tensor. We can always differentiate and anti-symmetrise to get a $(p+1)$-tensor defined by

$$(\mathcal{D}\mathsf{T})_{i_1\ldots i_{p+1}} = \frac{1}{p+1}\left(\frac{\partial T_{i_1\ldots i_p}}{\partial x_{i_{p+1}}} + p \text{ further terms}\right),$$

where the further terms involve replacing the derivative $\frac{\partial}{\partial x_{i_{p+1}}}$ with one of the other coordinates $\frac{\partial}{\partial x_j}$ so that the whole derivative is fully anti-symmetric.

Note that with this definition of $\mathcal{D}$, a second derivative of a p-tensor is a $(p+2)$-tensor, but this tensor always vanishes

$$(\mathcal{D}\mathcal{D}\mathsf{T})_{i_1\dots i_{p+2}} = 0$$

for any tensor T. This is because we will have two derivatives contracted with an epsilon and is the higher dimensional generalisation of the statements that $\boldsymbol{\nabla} \times \boldsymbol{\nabla}\phi = 0$ and $\boldsymbol{\nabla} \cdot \boldsymbol{\nabla} \times \mathbf{F} = 0$.

Remark. Here our anti-symmetric derivative obeys $\mathcal{D}^2(-) = 0$. We can link this with the fact that the boundary of a boundary is always zero. If a higher dimensional space (a manifold) M has boundary ∂M then $\partial^2 M = \partial(\partial M) = 0$. Conceptually, these two ideas are very different but one can't help but be struck by the similarity of the equations $\mathcal{D}^2(\text{anything}) = 0$ and $\partial^2(\text{anything}) = 0$, even though they are acting on very different objects. It turns out that this similarity is pointing at a deep connection between the topology of spaces and the kinds of tensors that one can put on these spaces. In mathematical terms, this is the link between homology and cohomology

Finally, we can now state the general integration theorem.

Theorem 11.28 (The general integration theorem). Given an anti-symmetric p-tensor T, then

$$\int_\Omega (\mathcal{D}\mathsf{T})_{i_1\dots i_{p+1}}\, \mathrm{d}\mathsf{S}_{i_1\dots i_{p+1}} = \int_{\partial\Omega} T_{i_1\dots i_p}\, \mathrm{d}\mathsf{S}_{i_1\dots i_p}\ ,$$

where $\dim(\Omega) = p+1$ and hence $\dim(\partial\Omega) = p$.

Remark. This is a unification of all integration theorems. It contains the fundamental theorem of calculus (when $p = 0$), the divergence theorem (when $p = n-1$) and Stokes' theorem (when $p = 1$ and $\mathbb{R}^n = \mathbb{R}^3$).

12 Further Contour Integrations

12.1 Residues

Lemma 12.1. Any function holomorphic and single-valued throughout an annulus $\alpha < |z - z_0| < \beta$ centred on $z = z_0$ has a unique *Laurent series* about $z = z_0$ that uniformly converges for all values of z within any compact subset of the annulus

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

Remark. If $f(z)$ has a single isolated singularity at $z = z_0$, then $\alpha > 0$ can be made arbitrarily small.

Definition 12.2. The coefficient a_{-1} in the Laurent series is called the *residue* of the function at the pole, denoted as

$$\operatorname*{res}_{z=z_0} f(z) \,.$$

Proposition 12.3. For a function f with a simple pole at $z = z_0$,

$$\operatorname*{res}_{z=z_0} f(z) = \lim_{z \to z_0} (z - z_0) f(z) \,.$$

Proof.

$$\lim_{z \to z_0} (z - z_0) f(z) = \lim_{z \to z_0} [a_{-1} + a_0 (z - z_0) + a_1 (z - z_0)^2] = a_{-1} \,.$$

□

Proposition 12.4. For a function f with a pole of order N at $z = z_0$,

$$\operatorname*{res}_{z=z_0} f(z) = \lim_{z \to z_0} \left[\frac{1}{(N-1)!} \frac{\mathrm{d}^{N-1}}{\mathrm{d}z^{N-1}} \big((z - z_0)^N f(z) \big) \right] .$$

Proof.

$$\begin{aligned}
&\lim_{z \to z_0} \left[\frac{1}{(N-1)!} \frac{\mathrm{d}^{N-1}}{\mathrm{d}z^{N-1}} \big((z - z_0)^N f(z) \big) \right] \\
&= \lim_{z \to z_0} \left[\frac{1}{(N-1)!} \frac{\mathrm{d}^{N-1}}{\mathrm{d}z^{N-1}} \big(a_{-N} + \cdots + a_{-1} (z - z_0)^{N-1} + a_0 (z - z_0)^N + \dots \big) \right] \\
&= \lim_{z \to z_0} [a_{-1} + N a_0 (z - z_0) + \dots] \\
&= a_{-1} \,.
\end{aligned}$$

□

12.2 Calculus of Residues

Lemma 12.5. Let $f(z)$ be analytic within a simply connected domain U except for an isolated singularity at $z = z_0$. Let γ be a simple closed contour in U counterclockwise around z_0. Then

$$\oint_\gamma f(z) \, \mathrm{d}z = 2\pi i \operatorname*{res}_{z=z_0} f(z) \,.$$

Proof. Since the Laurent series of f converges uniformly in $z \in U \setminus \{z_0\}$,

$$\begin{aligned}
\oint_\gamma f(z) \, \mathrm{d}z &= \oint_\gamma \sum_{n=-\infty}^{\infty} a_n (z - z_0)^n \\
&= \sum_{n=-\infty}^{\infty} \oint_\gamma a_n (z - z_0)^n \,.
\end{aligned}$$

Consider each term of the Laurent series separately.

For $n \geq 0$, each term is holomorphic throughout U, so by Cauchy's theorem,

$$\oint_\gamma a_n (z - z_0)^n \, \mathrm{d}z = 0 \, .$$

For $n < 0$, shrink the contour to a circle of radius ϵ about z_0 and substitute $z = z_0 + \epsilon e^{i\theta}$ to obtain

$$\begin{aligned}
\oint_\gamma a_n (z - z_0)^n \, \mathrm{d}z &= \oint_{|z-z_0|=\epsilon} a_n (z - z_0)^n \, \mathrm{d}z \\
&= \int_0^{2\pi} a_n \epsilon^n e^{in\theta} i\epsilon e^{i\theta} \, \mathrm{d}\theta \\
&= i a_n \epsilon^{n+1} \int_0^{2\pi} e^{i(n+1)\theta} \, \mathrm{d}\theta \\
&= \begin{cases} i a_n \epsilon^{n+1} \left[\dfrac{e^{i(n+1)\theta}}{i(n+1)} \right]_0^{2\pi} & \text{if } n \leq -2 \\ i a_n \epsilon^{n+1} 2\pi & \text{if } n = -1 \end{cases} \\
&= \begin{cases} 0 & \text{if } n \leq -2 \\ 2\pi i a_{-1} & \text{if } n = -1 \end{cases} \quad \text{as } \epsilon \to 0 \, .
\end{aligned}$$

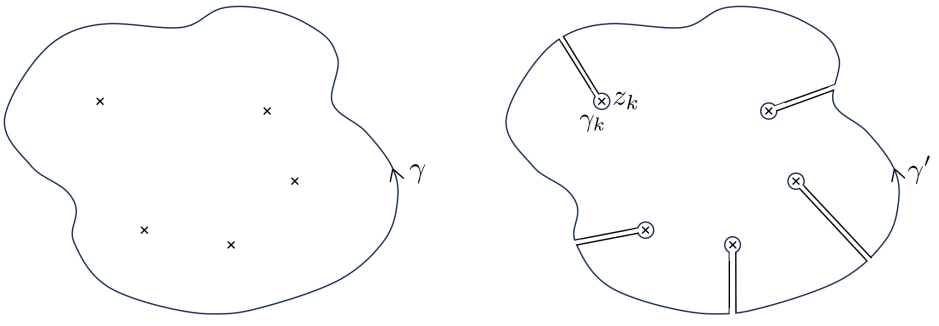
Therefore, summing over the results, we have

$$\oint_\gamma f(z) \, \mathrm{d}z = \sum_{n=-\infty}^{\infty} \oint_\gamma a_n (z - z_0)^n \, \mathrm{d}z = 2\pi i a_{-1} = 2\pi i \operatorname*{res}_{z=z_0} f(z) \, .$$

□

Theorem 12.6 (Residue theorem). Let f be meromorphic on a simply-connected domain U. Let γ be a simple closed counterclockwise contour in U encircling a finite number of isolated singularities at $z = z_1, z_2, \ldots, z_n$ and there is no singularity on γ, then

$$\oint_\gamma f(z) \, \mathrm{d}z = 2\pi i \sum_{k=1}^{n} \operatorname*{res}_{z=z_k} f(z) \, .$$



Proof. Consider a new contour γ' by joining γ with contours around each singularity γ_k using 'bridges' of width ϵ as shown. γ' does not enclose any pole so

$$\oint_{\gamma'} f(z) \, \mathrm{d}z = 0 \, .$$

At the limit $\epsilon \to 0$, the 'bridges' in the two opposite directions cancel out, so $\gamma' = \gamma - \sum_k \gamma_k$, and by Lemma 12.5, we have

$$\begin{aligned}\oint_{\gamma'} f(z)\,\mathrm{d}z &= \oint_{\gamma} f(z)\,\mathrm{d}z - \sum_k \oint_{\gamma_k} f(z)\,\mathrm{d}z \\ &= \oint_{\gamma} f(z)\,\mathrm{d}z - 2\pi i \sum_k \operatorname*{res}_{z=z_k} f(z) = 0\,.\end{aligned}$$

Therefore we have

$$\oint_{\gamma} f(z)\,\mathrm{d}z = 2\pi i \sum_k \operatorname*{res}_{z=z_k} f(z)\,.$$

□

12.3 The Point at Infinity

Some functions tend to a definite limit as $z \to \infty$ irrespective of the direction from which the infinity is approached. e.g. $f(z) = 1/z$ goes to 0 as $|z| \to \infty$. Therefore, it sometimes makes sense to think of ∞ as a single point, as illustrated by the *stereographic projection* of the complex plane onto *Riemann sphere*.

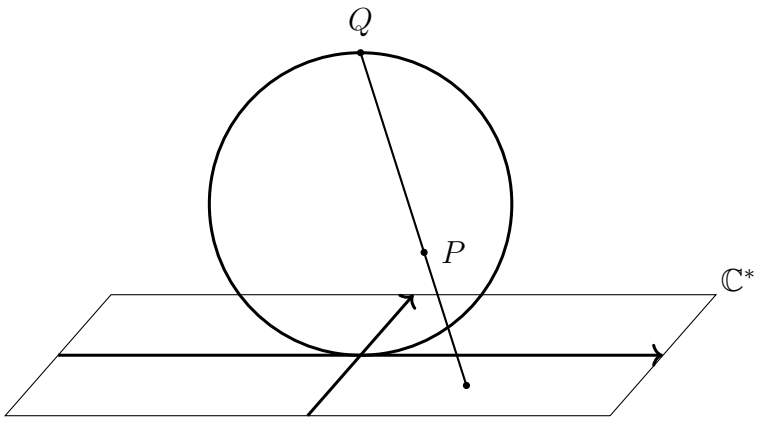
Definition 12.7. *Stereographic projection* may be applied to a unit n-sphere S^n in an $(n+1)$-dimensional Euclidean space $\mathbb{E}^{n+1}$. If Q is a point of S^n and E is a hyperplane in $\mathbb{E}^{n+1}$, then the stereographic projection of a point $P \in S^n - \{Q\}$ is the point P', intersection of the line QP with E. The hypersphere of projection is the *Riemann sphere*.

Remark. The point Q in stereographic projection is projected to all points on a circle of infinite radius. We can think of this as a single *point at infinity*.

Definition 12.8. The *extended complex plane* $\mathbb{C}_\infty \coloneqq \mathbb{C} \cup \{\infty\}$

Remark. Stereographic projection is a smooth, bijective function from the entire sphere except for the centre of projection to the entire plane. It maps circles on the sphere to circles or lines on the plane, and is conformal (angle-preserving).

The Riemann sphere of the extended complex plane is shown below.



We can study the behaviour of $f(z)$ near the point at infinity by defining a new complex variable,

$$\zeta = \frac{1}{z}\,.$$

The point at infinity in the z-plane is the origin in the ζ-plane, and vice versa. Setting

$$g(\zeta) = f\left(\frac{1}{\zeta}\right),$$

we can find a Laurent expansion for g about $\zeta = 0$. If g has a singularity at $\zeta = 0$ then f has this singularity at infinity.

Remark. Care must be taken when combining the idea of a point at infinity with the residue theorem because the residue is not strictly a property of f but of $f\,\mathrm{d}z$. For example, if γ is the anticlockwise unit circle in the z-plane, then

$$\frac{1}{2\pi i}\oint_\gamma \frac{\mathrm{d}z}{z} = 1\,.$$

But γ is also the clockwise unit circle in the ζ-plane. γ may also be viewed as a simple closed contour that encloses the point at infinity in the z-plane. The integral is therefore equal to minus the sum of the residues outside γ. As $1/z$ has no singularities in the complex plane away from $z = 0$, the residue at $z = \infty$ must be -1, even though the function $1/z$ is not singular there.

12.4 Applications of the Calculus of Residues

12.4.1 Integrals Involving Trigonometric Functions

Consider the integral

$$I = \int_0^{2\pi} \frac{\mathrm{d}\theta}{2(a - \cos\theta)}\,,$$

where $a > 1$ is a real constant. Consider the substitution

$$z = e^{i\theta}\,.$$

This gives us $\mathrm{d}z = iz\,\mathrm{d}\theta$ and $\cos\theta = \frac{1}{2}(z + z^{-1})$, while the integral between $\theta \in [0, 2\pi]$ corresponds to the integral over z around a unit circle γ in the complex plane. Then

$$\begin{aligned} I &= \oint_\gamma \frac{\mathrm{d}z}{2iz\big(a - \frac{1}{2}(z + z^{-1})\big)} \\ &= i\oint_\gamma \frac{\mathrm{d}z}{z^2 - 2az + 1} \\ &= i\oint_\gamma \frac{\mathrm{d}z}{(z - z_+)(z - z_-)}\,, \end{aligned}$$

where the integrand has simple poles at $z_\pm = a \pm \sqrt{a^2 - 1}$.

Since $a > 1$, it follows that $z_- \in (0, 1)$ and $z_+ > 1$. Hence the pole z_- is inside γ and z_+ is outside it. The residue is

$$\frac{i}{z_- - z_+} = -\frac{i}{2\sqrt{a^2 - 1}}\,,$$

so from the residue theorem (Theorem 12.6),

$$I = \frac{\pi}{\sqrt{a^2 - 1}}\,.$$

12.4.2 Closing a Contour at Infinity

Suppose that we wish to calculate the integral

$$I = \int_0^\infty \frac{\mathrm{d}x}{x^2 + 1}\,.$$

Consider

$$\oint_\gamma \frac{\mathrm{d}z}{z^2+1} = \oint_{\gamma_0+\gamma_R} \frac{\mathrm{d}z}{(z+i)(z-i)}\,,$$

where $\gamma = \gamma_0 + \gamma_R$ consists of two parts: first a contour, γ_0, from $-R$ to R along the real axis, and a second contour, γ_R, counterclockwise along a semicircle of radius R in the upper half plane.

The integrand has two simple poles, but only the one at $z = i$ is enclosed by γ. Hence, from the residue theorem (Theorem 12.6),

$$\oint_\gamma \frac{\mathrm{d}z}{z^2+1} = \oint_{\gamma_0+\gamma_R} \frac{\mathrm{d}z}{(z+i)(z-i)} = 2\pi i \frac{1}{2i} = \pi\,.$$

We also have that, using the symmetry of the integrand,

$$\int_{\gamma_0} \frac{\mathrm{d}z}{z^2+1} \equiv \int_{-R}^{R} \frac{\mathrm{d}z}{z^2+1} = 2\int_0^R \frac{\mathrm{d}z}{z^2+1} \to 2I \text{ as } R \to \infty\,.$$

Finally, we consider the value of the integral along γ_R. On this semicircle, the integrand is $O(R^{-2})$, while the contour has length πR. Hence,

$$\left|\int_{\gamma_R} \frac{\mathrm{d}z}{z^2+1}\right| \le \int_{\gamma_R} \frac{|\mathrm{d}z|}{\min|z^2+1|} \le \frac{\pi R}{R^2-1} \to 0 \text{ as } R \to \infty\,.$$

Combining the above result and taking the limit $R \to \infty$, we conclude that

$$I = \frac{\pi}{2}\,.$$

Remark. This method can be easily generalised to contour integrals containing multiple poles.

12.5 Multi-valued Functions and Branch Cuts

Not all complex functions have a single value for each complex point $z = re^{i\theta}$. For instance, the complex function $\log z = \ln r + i\theta$ has infinitely many values, or branches, since θ can take infinitely many values.

If a contour γ does not enclose the origin, then we can always choose some range of θ so that $\ln z$ is continuous and single-valued. However, if the contour γ encloses the origin, then $\ln z$ on the contour can only be multivalued or discontinuous.

Definition 12.9. A point that cannot be encircled by a curve where the function is continuous and single-valued is called a *branch point*. The function has a *branch point singularity* at that point.

In order to make a function with a branch point continuous and single-valued on a curve, it is necessary that the curve does not encircle the branch point. To do this, we have to introduce a *branch cut* that no curve is permitted to cross. Having a branch cut, a *branch* of a function is defined such that in the neighbourhood of the branch point, values of θ in a 2π range are chosen.

Remark. If a curve did cross the cut, the function would be discontinuous and not analytic.

Example. The canonical branch cut for $\log z$ is along the half real axis from $-\infty$ to the origin, so that $\theta \in (-\pi, \pi)$. With this choice of branch cut, the value of $\log z$ is called the *principal value* of the logarithm, often denoted as $\mathrm{Log}\, z$.

Remarks.

- $\log z$ has an infinite number of branches.

- The complex logarithm is holomorphic everywhere on each branch except on the branch cut.
- The function is single-valued and continuous on any curve that does not cross the cut.
- Branch cuts need not be straight lines. Any continuous non-intersecting curve from the branch point to infinity can be a branch cut.

Corollary. Multivalued functions have no Laurent expansions about the branch points, since any annulus $|z - z_0| \in (\alpha, \beta)$ would be crossed by the branch cut so the function would not be analytic in the annulus.

Remark. Riemann introduced a different idea where the different branches of a function are regarded as separate copies of the complex plane $\mathbb{C}$ stacked onto each other and each connected to its neighbours at the respective branch cuts. This is known as the *Riemann surface*.

Example. Consider

$$f(z) = \sqrt{z^2 - 1} = \sqrt{z - 1}\sqrt{z + 1}\,,$$

a function that has two branch points at $z = \pm 1$. Setting

$$z - 1 = r_1 e^{i\theta_1} \text{ and } z + 1 = r_2 e^{i\theta_2}\,,$$

we see that

$$f(z) = \sqrt{r_1 r_2} e^{\frac{1}{2} i(\theta_1 + \theta_2)}\,.$$

If z_1 is enclosed by a small curve γ_1, then

$$\theta_1 \to \theta_1 + 2\pi\,,\ \theta_2 \to \theta_2 \text{ and } \frac{1}{2}(\theta_1 + \theta_2) \to \frac{1}{2}(\theta_1 + \theta_2) + \pi\,.$$

Hence $f(z)$ changes the sign. The same applies to a small curve γ_2 encircling $z = -1$. However, going around a curve γ_3 encircling both branch points has the effect of

$$\theta_1 \to \theta_1 + 2\pi\,,\ \theta_2 \to \theta_2 + 2\pi \text{ and } \frac{1}{2}(\theta_1 + \theta_2) \to \frac{1}{2}(\theta_1 + \theta_2) + 2\pi\,.$$

Hence $f(z)$ does not change the sign.

Therefore, we can introduce a branch cut that goes from $z = -1$ to $z = 1$, with the simplest case being a cut on the real axis. Alternatively, we can introduce two separate branch cuts, one from each branch point to infinity.

Remark. The two branch cuts to infinity can be seen as a single branch cut that happens to pass through the point at infinity. This is because the cuts can be smoothly deformed. In general, when there is more than one branch point, we may need more than one branch cut.

12.6 Contour Integration around a Branch Cut

Evaluate the integral

$$I = \int_0^\infty \frac{x^\alpha}{1 + \sqrt{2}x + x^2}\,\mathrm{d}x\,,$$

where $\alpha \in (-1, 1)$.

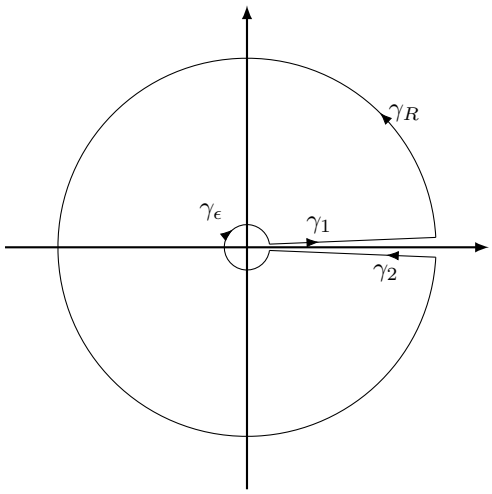
Consider the contour integral

$$\oint_\gamma \frac{z^\alpha}{1 + \sqrt{2}z + z^2}\,\mathrm{d}z\,.$$

The integrand has a branch point at $z = 0$, and simple poles at $z_1 = e^{\frac{3}{4}\pi i}$ and $z_2 = e^{\frac{5}{4}\pi i}$.

We usually find it is appropriate to choose a branch cut along the integration range, i.e. along the positive real axis in this case; we then define the branch by choosing $0 \le \theta < 2\pi$, where $z = re^{i\theta}$.

It is then necessary to use a 'keyhole contour', γ, in order to avoid the branch point and the branch cut. We consider the individual contributions from each part of the contour, $\gamma = \gamma_1 + \gamma_R + \gamma_2 + \gamma_\epsilon$ in turn.



The contribution from the contour γ_1 just above the branch cut is

$$\int_\epsilon^R \frac{x^\alpha}{1+\sqrt{2}x+x^2}\,\mathrm{d}x \to I$$

as $\epsilon \to 0$ and $R \to \infty$.

Substituting $z = re^{2\pi i}$, the contribution from the contour γ_2 just below the branch cut is

$$\int_R^\epsilon \frac{r^\alpha e^{2\pi\alpha i}}{1+\sqrt{2}r+r^2}\,\mathrm{d}r \to -e^{2\pi\alpha i} I$$

as $\epsilon \to 0$ and $R \to \infty$.

For γ_ϵ, substitute $z = \epsilon e^{i\theta}$, we obtain

$$\int_{\gamma_\epsilon} \frac{z^\alpha}{1+\sqrt{2}z+z^2}\,\mathrm{d}z = \epsilon^{\alpha+1}\int_{2\pi}^0 \frac{e^{i(\alpha+1)\theta}}{1+\sqrt{2}\epsilon e^{i\theta}+\epsilon^2 e^{2i\theta}} i\,\mathrm{d}\theta \to 0 \text{ as } \epsilon \to 0\,.$$

For γ_R, substitute $z = Re^{i\theta}$, we obtain

$$\begin{aligned}\int_{\gamma_R} \frac{z^\alpha}{1+\sqrt{2}z+z^2}\,\mathrm{d}z &= R^{\alpha+1}\int_0^{2\pi} \frac{e^{i(\alpha+1)\theta}}{1+\sqrt{2}Re^{i\theta}+R^2e^{2i\theta}} i\,\mathrm{d}\theta \\ &= R^{\alpha-1}\int_0^{2\pi} \frac{e^{i(\alpha+1)\theta}}{R^{-2}+\sqrt{2}R^{-1}e^{i\theta}+e^{2i\theta}} i\,\mathrm{d}\theta \to 0 \text{ as } R \to \infty\,.\end{aligned}$$

Therefore, as $\epsilon \to 0$ and $R \to \infty$,

$$\oint_\gamma \frac{z^\alpha}{1+\sqrt{2}z+z^2} = \oint_{\gamma_1+\gamma_R+\gamma_2+\gamma_\epsilon} \frac{z^\alpha}{(z-z_1)(z-z_2)}\,\mathrm{d}z \to (1-e^{2\pi\alpha i})I\,.$$

Because both poles are inside γ, the residue theorem gives

$$\operatorname*{res}_{z=z_1} \frac{z^\alpha}{(z-z_1)(z-z_2)} = \frac{z_1^\alpha}{z_1-z_2} = \frac{e^{\frac{3}{4}\pi\alpha i}}{e^{\frac{3}{4}\pi\alpha i}-e^{\frac{5}{4}\pi\alpha i}}\,,$$

$$\operatorname*{res}_{z=z_2} \frac{z^\alpha}{(z-z_1)(z-z_2)} = \frac{z_2^\alpha}{z_2-z_1} = \frac{e^{\frac{5}{4}\pi\alpha i}}{e^{\frac{5}{4}\pi\alpha i} - e^{\frac{3}{4}\pi\alpha i}},$$

$$(1-e^{2\pi\alpha i})I = 2\pi i\left(\frac{e^{\frac{3}{4}\pi\alpha i}}{i\sqrt{2}} - \frac{e^{\frac{5}{4}\pi\alpha i}}{i\sqrt{2}}\right),$$

and so

$$I = \sqrt{2}\pi\frac{\sin\left(\frac{\alpha\pi}{4}\right)}{\sin(\alpha\pi)}.$$

13 Transform Methods

13.1 Jordan's Lemma

Consider

$$\lim_{R\to\infty}\int_\gamma g(z)e^{i\lambda z}\,\mathrm{d}z\ ,$$

where

(i) λ is a real positive constant;

(ii) $g(z)$ is holomorphic in the upper half-plane except possibly at a finite number of poles;

(iii) the contour γ is a semicircle of radius R in the upper half-plane: $\gamma:[0,\pi]\to\mathbb{C}$, $\theta\mapsto Re^{i\theta}$.

Lemma 13.1. The upper bound for the contour integral is given by

$$\left|\int_\gamma g(z)e^{i\lambda z}\,\mathrm{d}z\right|\le\frac{\pi}{\lambda}M_R\,,$$

where

$$M_R=\sup_{\theta\in[0,\pi]}\left|g(Re^{i\theta})\right|,$$

with equality when g vanishes everywhere.

Proof. Since on γ,

$$\left|e^{i\lambda z}\right|=\left|e^{-\lambda R\sin\theta}\right|,$$

we have

$$\begin{aligned}\left|\int_\gamma g(z)e^{i\lambda z}\,\mathrm{d}z\right|&\le\int_\gamma\left|e^{i\lambda z}g(z)\right||\mathrm{d}z|\\&=M_RR\int_0^\pi e^{-\lambda R\sin\theta}\,\mathrm{d}\theta\\&=2M_RR\int_0^{\frac{\pi}{2}}e^{-\lambda R\sin\theta}\,\mathrm{d}\theta\ .\end{aligned}$$

Using the inequality that for $\theta\in[0,\pi/2]$,

$$1\ge\frac{\sin\theta}{\theta}\ge\frac{2}{\pi}\,,$$

$$\begin{aligned}\int_\gamma g(z)e^{i\lambda z}\,\mathrm{d}z&\le 2M_RR\int_0^{\frac{\pi}{2}}e^{-\lambda R\frac{2\theta}{\pi}}\,\mathrm{d}\theta\\&=\frac{\pi M_R}{\lambda}(1-e^{-\lambda R})\\&\le\frac{\pi M_R}{\lambda}\,,\end{aligned}$$

with equality when g is identically 0. □

Remark. An analogous statement for a semicircular clockwise contour in the lower half-plane holds when $\lambda<0$.

Remark. For the case $\lambda=0$, this reduces to the ML estimation lemma (Lemma 6.23).

Lemma 13.2 (Jordan's Lemma). If $g(z) \to 0$ uniformly on γ as $R \to \infty$, i.e. if

$$\lim_{R\to\infty} M_R = 0\,,$$

then

$$\lim_{R\to\infty} \int_\gamma g(z)e^{i\lambda z}\,\mathrm{d}z = 0\,.$$

Proof. It follows trivially from Lemma 13.1. □

13.1.1 Example of Jordan's Lemma

Consider

$$I = \int_{-\infty}^{\infty} \frac{\sin x}{x}\,\mathrm{d}x\ .$$

Note that the singularity at the origin is removable. Since $\sin z = \frac{1}{2i}(e^{iz} - e^{-iz})$, we can apply Jordan's lemma by splitting up the integral.

$$\begin{aligned} I &= \frac{1}{2i}\left(\int_{-\infty}^{\infty} \frac{e^{iz}}{z}\,\mathrm{d}z - \int_{-\infty}^{\infty} \frac{e^{-iz}}{z}\,\mathrm{d}z\right) \\ &= \mathrm{Im}\left(\int_{-\infty}^{\infty} \frac{e^{iz}}{z}\right). \end{aligned}$$

But now the contour passes through a pole, so instead consider the limit

$$I = \mathrm{Im}\left[\lim_{\epsilon\to 0}\lim_{R\to\infty}\left(\int_{-R}^{\epsilon} \frac{e^{iz}}{z}\,\mathrm{d}z + \int_{\epsilon}^{R} \frac{e^{iz}}{z}\,\mathrm{d}z\right)\right].$$

Define the contour $\Gamma = \gamma_R + \gamma_- + \gamma_\epsilon + \gamma_+$, where

$$\begin{cases} \gamma_R : \theta \mapsto Re^{i\theta} & ,\ \theta \in [0,\pi] \\ \gamma_- : t \mapsto t & ,\ t \in [-R,-\epsilon] \\ \gamma_\epsilon : \theta \mapsto \epsilon e^{i(\pi-\theta)} & ,\ \theta \in [0,\pi] \\ \gamma_+ : t \mapsto t & ,\ t \in [\epsilon, R]\,. \end{cases}$$

Then since Γ does not enclose any pole, from Cauchy's theorem (Theorem 6.26),

$$\begin{aligned} \oint_\Gamma \frac{e^{iz}}{z}\,\mathrm{d}z &= \int_{\gamma_R} \frac{e^{iz}}{z}\,\mathrm{d}z + \int_{\gamma_-} \frac{e^{iz}}{z}\,\mathrm{d}z + \int_{\gamma_\epsilon} \frac{e^{iz}}{z}\,\mathrm{d}z + \int_{\gamma_+} \frac{e^{iz}}{z}\,\mathrm{d}z \\ &= 0\,. \end{aligned}$$

On γ_ϵ, $z = \epsilon e^{i(\pi-\theta)}$, so

$$\begin{aligned} \int_{\gamma_\epsilon} \frac{e^{iz}}{z}\,\mathrm{d}z &= \int_\pi^0 \frac{\exp\big(i\epsilon e^{i\theta}\big)}{\epsilon e^{i\theta}} i\epsilon e^{i\theta}\,\mathrm{d}\theta \\ &= -i\int_0^\pi \sum_{r=0}^{\infty} \frac{i^r\epsilon^r e^{ir\theta}}{r!}\,\mathrm{d}\theta \\ &= -i\int_0^\pi 1 + O(\epsilon)\mathrm{d}\theta\,. \end{aligned}$$

Thence,

$$\lim_{\epsilon\to 0}\int_{\gamma_\epsilon} \frac{e^{iz}}{z}\,\mathrm{d}z = -i\pi\,.$$

Further, from Jordan's lemma, we know that

$$\lim_{R\to\infty}\int_{\gamma_R}\frac{e^{iz}}{z}\,\mathrm{d}z = 0\,.$$

Hence, taking the double limit $\epsilon \to 0$ and $R \to \infty$, we have

$$\begin{aligned} I &= \operatorname{Im}\lim_{\epsilon\to 0}\lim_{R\to\infty}\left(-\int_{\gamma_R}\frac{e^{iz}}{z}\,\mathrm{d}z - \int_{\gamma_\epsilon}\frac{e^{iz}}{z}\,\mathrm{d}z\right)\\ &= \operatorname{Im}(i\pi) = \pi\,. \end{aligned}$$

Remark. Similar methods can be used to evaluate

$$\int_{-\infty}^{\infty}\frac{\sin^2 x}{x^2}\,\mathrm{d}x\ .$$

13.2 Fourier Transform Methods

Here, we will state a more precise version of the Fourier transform that allows us to incorporate techniques of contour integrations.

Definition 13.3. The *Cauchy principal value* of the integral of an integrable function f in the range $(-\infty, \infty)$ is given by

$$\mathrm{PV}\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x \equiv \fint_{-\infty}^{\infty} f(x)\,\mathrm{d}x \coloneqq \lim_{R\to\infty}\int_{-R}^{R} f(x)\,\mathrm{d}x\ .$$

Remark. Some functions do not have integral from $-\infty$ to ∞ in the normal sense, but have Cauchy principal value. For example,

$$\fint_{-\infty}^{\infty}\frac{x}{1+x^2}\,\mathrm{d}x = 0$$

but

$$\int_{-\infty}^{\infty}\frac{x}{1+x^2}\,\mathrm{d}x$$

does not converge.

Definition 13.4. Let $f : \mathbb{R} \to \mathbb{C}$ be absolutely integrable, i.e. $\int_{-\infty}^{\infty}|f(x)|\,\mathrm{d}x$ exists, has bounded variation and a finite number of discontinuities. The *Fourier transform* of a function $f(x)$ is

$$\tilde{f}(k) \equiv \mathcal{F}[f(x)] \coloneqq \fint_{-\infty}^{\infty} f(x)e^{-ikx}\,\mathrm{d}x\ .$$

Remark. Sometimes we can take the Fourier transform of functions that do not satisfy the requirement of absolute integrability, like $f(x) = 1$. These can be handled using distributions.

$$\tilde{f}(k) = 2\pi\delta(k)\,.$$

Theorem 13.5 (Fourier inversion theorem). The inverse Fourier transform acting on $\tilde{f}(k)$,

$$\frac{1}{2}(f(x^+) + f(x^-)) = \frac{1}{2\pi}\fint_{-\infty}^{\infty}\tilde{f}(k)e^{ikx}\,\mathrm{d}k\ ,$$

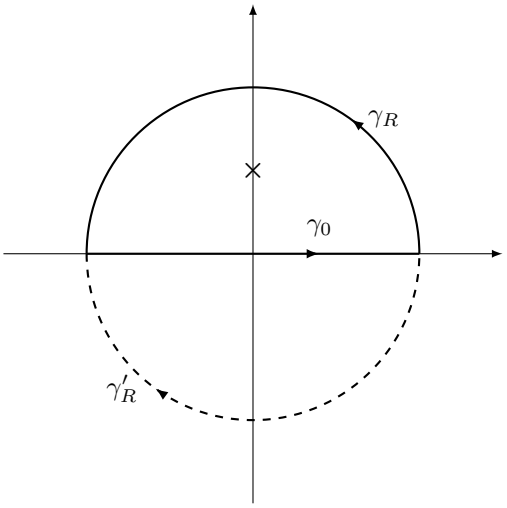
returns $f(x)$ when it is continuous while giving the average value of the left and right hand side limits at a discontinuity.

13.2.1 Fourier Transform using Contour Integration

Example. Consider the inverse Fourier transform of

$$f(x) = \frac{1}{a + ix}\,,$$

where $a > 0$ is a constant. Let γ_0 be the contour from $-R$ to R in the real axis, γ_R be the semicircle of radius R in the upper half-plane, γ'_R be the semicircle in the lower half-plane. We let $\gamma = \gamma_0 + \gamma_R$ and $\gamma' = \gamma_0 + \gamma'_R$.



We can see that $f(x)$ has only one single pole at $x = ia$, so we get

$$\oint_\gamma f(x)e^{-ikx}\,\mathrm{d}x = 2\pi i \operatorname*{res}_{x=ia} \frac{e^{-ikx}}{i(x - ia)} = 2\pi e^{ka}\,.$$

While we have

$$\oint_{\gamma'} f(x)e^{-ikx}\,\mathrm{d}x = 0\,.$$

Now, if $k < 0$, applying Jordan's lemma with $\lambda = -k$ to γ_R gives that

$$\int_{\gamma_R} f(x)e^{-ikx}\,\mathrm{d}x \to 0 \text{ as } R \to \infty\,.$$

Hence,

$$\begin{aligned}
\tilde{f}(k) &= \fint_{-\infty}^{\infty} f(x)e^{-ikx}\,\mathrm{d}x \\
&= \lim_{R\to\infty} \int_{\gamma_0} f(x)e^{-ikx}\,\mathrm{d}x \\
&= \lim_{R\to\infty} \left(\int_{\gamma} f(x)e^{-ikx}\,\mathrm{d}x - \int_{\gamma_R} f(x)e^{-ikx}\,\mathrm{d}x \right) \\
&= 2\pi e^{ka}\,.
\end{aligned}$$

For $k > 0$, we close the contour in the lower half plane. Since there is no singularity, we get

$$\int_{-\infty}^{\infty} f(x)e^{-ikx}\,\mathrm{d}x = 0\,.$$

Therefore,

$$\tilde{f}(k) = \begin{cases} 0 & k > 0 \\ 2\pi e^{ak} & k < 0 \,. \end{cases}$$

13.2.2 Damped Harmonic Oscillator

Consider the equation for the amplitude $x(t)$ of a driven damped harmonic oscillator,

$$\ddot{x}(t) + 2\gamma\dot{x}(t) + \omega_0^2 x(t) = f(t)\,, \tag{$\dagger$}$$

where $f(t)$ is the forcing function, $\omega_0 > 0$ is real, and $\gamma > 0$ is real and represents the effects of damping.

Assume that $x(t) \to 0$ as $|t| \to \infty$ so that we can define Fourier transform, and its inverse, of $x(t)$ as

$$\tilde{x}(\omega) = \int_{-\infty}^{\infty} x(t) e^{-i\omega t}\,\mathrm{d}t$$
$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{x}(\omega) e^{i\omega t}\,\mathrm{d}\omega\ ,$$

where we use ω as the Fourier variable when the function depends on t as is conventional.

By Proposition 3.5, we have

$$\mathcal{F}[\dot{x}(t)] = i\omega\tilde{x}(\omega)$$
$$\mathcal{F}[\ddot{x}(t)] = -\omega^2\tilde{x}(\omega)\,.$$

Take Fourier transform on both sides of (†), we have

$$(-\omega^2 + 2i\gamma\omega + \omega_0^2)\tilde{x}(\omega) = \tilde{f}(\omega)\,.$$

It follows that

$$\tilde{x}(\omega) = \tilde{f}(\omega)\tilde{g}(\omega)\,,$$

where

$$\begin{aligned} \tilde{g}(\omega) &= \frac{-1}{\omega^2 - 2i\gamma\omega - \omega_0^2} \\ &= \frac{-1}{(\omega-\omega_+)(\omega-\omega_-)} \qquad \text{and } \omega_\pm = i\gamma \pm \sqrt{\omega_0^2 - \gamma^2}\,. \end{aligned}$$

We can find $x(t)$ by taking the inverse Fourier transform

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{f}(\omega)\tilde{g}(\omega)e^{i\omega t}\,\mathrm{d}\omega\ .$$

Recall that the convolution theorem (Theorem 3.10) states

$$h(t) = \int_{-\infty}^{\infty} f(s)g(t-s)\,\mathrm{d}s \iff \tilde{h}(\omega) = \tilde{f}(\omega)\tilde{g}(\omega)\,.$$

Hence, we deduce that

$$x(t) = \int_{-\infty}^{\infty} f(s)g(t-s)\,\mathrm{d}s\ ,$$

where, with $\tau = t - s$,

$$g(\tau) = \mathcal{F}^{-1}[\tilde{g}(\omega)] = -\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{i\omega\tau}}{(\omega-\omega_+)(\omega-\omega_-)}\,\mathrm{d}\omega\ .$$

Remark. This is equivalent to a solution using the Green's function $G(t,s) = g(t-s)$ of

$$L = \frac{\mathrm{d}^2}{\mathrm{d}t^2} + 2\gamma\frac{\mathrm{d}}{\mathrm{d}t} + \omega_0^2 \,.$$

To complete the solution to the problem, we now have to determine $g(\tau)$ by integrating over ω. We will do this by employing a contour integral in the complex ω plane.

If $\tau < 0$, we choose a contour γ that goes along the real axis and is closed with a semicircle in the lower half-plane (γ_∞). If $\tau > 0$ we instead close the contour with a semicircle in the upper half plane. Since, by our assumption, $\tilde{g}(\omega) \to 0$ as $|\omega| \to \infty$, Jordan's lemma (Lemma 13.2) implies that in both cases the integral over the semicircle will vanish. It then follows that

$$\begin{aligned} g(\tau) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{g}(\omega) e^{i\omega\tau}\,\mathrm{d}\omega \\ &= \frac{1}{2\pi}\left(\int_{-\infty}^{\infty} \tilde{g}(\omega) e^{i\omega\tau}\,\mathrm{d}\omega + \int_{\gamma_\infty} \tilde{g}(\omega) e^{i\omega\tau}\,\mathrm{d}\omega\right) \\ &= \frac{1}{2\pi}\oint_{\gamma} \tilde{g}(\omega) e^{i\omega\tau}\,\mathrm{d}\omega \,. \end{aligned}$$

As long as ω_0 is real, so that $\omega_0^2 > 0$, poles of $\tilde{g}(\omega)$ at $\omega = \omega_\pm$ are both in the upper half plane. Therefore, from the residue theorem,

$$g(\tau) = 0 \quad \text{when} \quad \tau < 0 \,.$$

In other words, $g(t-s)$ is zero if $t < s$. Suppose that the forcing term is not switched on until $t = 0$, i.e. suppose that $f(t) = 0$ for $t < 0$, it follows that

$$x(t) = \int_0^{\infty} f(s) g(t-s)\,\mathrm{d}s = 0 \quad \text{for} \quad t < 0 \,.$$

Remark. This means that there is no response until the forcing term is switched on. This is a *causal behaviour*, i.e. effect follows cause and not the other way around. The Green's function, $G(t,s) = g(t-s)$ is said to be a *causal Green's function.*

For $\tau > 0$, there are two simple poles within γ. As we assume that $\gamma \neq \omega_0$, the residue at $\omega = \omega_\pm$ are given by

$$\operatorname*{res}_{\omega=\omega_\pm}\left(\frac{1}{2\pi}\tilde{g}e^{i\omega\tau}\right) = \frac{-e^{i\omega_\pm\tau}}{2\pi(\omega_\pm - \omega_\mp)} = \mp\frac{e^{-\gamma\tau}e^{\pm i\tau\sqrt{\omega_0^2-\gamma^2}}}{4\pi\sqrt{\omega_0^2-\gamma^2}} \,.$$

- *Underdamped Oscillator.* For $\gamma < \omega_0$, the oscillator is said to be *underdamped*. We can deduce that

$$g(\tau) = \frac{e^{-\gamma\tau}}{\sqrt{\omega_0^2-\gamma^2}}\sin\left(\tau\sqrt{\omega_0^2-\gamma^2}\right) \quad \text{for } \tau > 0 \,.$$

  Remark. Suppose that there is a unit impulse at $t = 0$, i.e. $f(t) = \delta(t)$. It follows that $x(t) = g(t)$, and hence the response to an impulsive force is oscillatory with an amplitude that dies away exponentially over a time of order $1/\gamma$.

  For $\gamma \ll \omega_0$, the main effect of the damping term is to cause the oscillation to slowly reduce in amplitude rather than change phase.

- *Overdamped Oscillator.* For $\gamma > \omega_0$, the oscillator is said to be *overdamped*. We can deduce that

$$g(\tau) = \frac{e^{-\gamma\tau}}{\sqrt{\gamma^2-\omega_0^2}}\sinh\left(\tau\sqrt{\gamma^2-\omega_0^2}\right) \quad \text{for } \tau > 0 \,.$$

- *Critically Damped Oscillator.* When $\gamma = \omega_0$, the oscillator is critically damped. For $\tau > 0$, there is a double pole at $\omega = i\gamma$ inside the contour γ. From Proposition 12.4, we have

$$\begin{aligned}\operatorname*{res}_{\omega=i\gamma}\left(\frac{1}{2\pi}\tilde{g}e^{i\omega\tau}\right) &= \operatorname*{res}_{\omega=i\gamma}\left(-\frac{e^{i\omega\tau}}{2\pi(\omega-i\gamma)^2}\right)\\ &= \lim_{\omega\to i\gamma}\left(\frac{\mathrm{d}}{\mathrm{d}\omega}\left(-\frac{e^{i\omega\tau}}{2\pi}\right)\right)\\ &= -\frac{i\tau e^{-\gamma\tau}}{2\pi}\,.\end{aligned}$$

  Hence, the residue theorem yields

$$g(\tau) = \tau e^{-\gamma\tau} \quad \text{for } \tau > 0\,.$$

13.2.3 Gaussian Integration Lemma

Lemma 13.6 (Gaussian integration lemma). For any constant $c \in \mathbb{C}$,

$$\int_{-\infty}^{\infty} e^{-(u+c)^2}\,\mathrm{d}u = \sqrt{\pi}\,.$$

Proof. Let $c = a + bi$. Extend the integral to the complex plane, and define a new complex variable $z = u + c$. Then

$$I = \int_{-\infty}^{\infty} e^{-(u+c)^2}\,\mathrm{d}u = \int_{\gamma_i} e^{-z^2}\,\mathrm{d}z\ ,$$

where the contour γ_i is the horizontal line in the complex z plane with $\operatorname{Im} z = \operatorname{Im} c = b$.

The integrand e^{-z^2} is analytic everywhere and so the integral of e^{-z^2} around any closed contour is zero.

Consider the rectangular counterclockwise contour with vertices at $\pm R$ and $\pm R + ib$, $R \in \mathbb{R}$.

Apply Cauchy's theorem to this contour to obtain

$$\begin{aligned}0 &= \lim_{R\to\infty}\oint_{\gamma_R} e^{-z^2}\,\mathrm{d}z\\ &= \lim_{R\to\infty}\left[\int_{-R}^{R} e^{-z^2}\,\mathrm{d}z + \int_0^b e^{-(R+iy)^2} i\,\mathrm{d}y + \int_{R+ib}^{-R+ib} e^{-z^2}\,\mathrm{d}z + \int_b^0 e^{-(-R+iy)^2} i\,\mathrm{d}y\right]\\ &= \sqrt{\pi} - I + \lim_{R\to\infty} 2e^{-R^2}\int_0^b e^{y^2}\sin(2Ry)\,\mathrm{d}y\ .\end{aligned}$$

In the limit $R \to \infty$ the final term tends to zero, and so we deduce that, for any $c \in \mathbb{C}$,

$$I = \int_{-\infty}^{\infty} e^{-(u+c)^2}\,\mathrm{d}u = \sqrt{\pi}\,.$$

□

13.2.4 Solutions to Partial Differential Equations

Consider the initial-boundary value problem of the heat distribution on an infinite bar:

$$\begin{cases}\dfrac{\partial\theta}{\partial t} = \lambda\dfrac{\partial^2\theta}{\partial x^2} & (x,t) \in \mathbb{R}\times(0,\infty)\\ \theta(x,0) = f(x) & x \in \mathbb{R}\\ \dfrac{\partial\theta}{\partial x} \to 0 & \text{as } x \to \pm\infty, t \in (0,\infty)\,.\end{cases}$$

Take the Fourier transform of the equation, we have

$$\frac{\partial\tilde{\theta}}{\partial t} = -\lambda k^2\tilde{\theta}\,.$$

The solution of this equation is

$$\tilde{\theta}(k,t) = \tilde{\theta}_0(k)e^{-\lambda k^2 t}\,.$$

Applying the initial condition, we have $\tilde{\theta}_0(k) = \tilde{f}(k)$. Rewrite the expression of the transformed solution as

$$\tilde{\theta}(k,t) = \tilde{f}(k)\tilde{G}(k,t)\,,$$

where

$$\tilde{G}(k,t) = e^{-\lambda k^2 t}\,.$$

By the convolution theorem, the solution is

$$\theta(x,t) = \theta(x) * G(x,t) = \int_{-\infty}^{\infty} \theta_0(y)G(x-y,t)\,\mathrm{d}y\ ,$$

where using the substitution $u = \sqrt{\lambda t}k$, $G(x,t)$ can be evaluated as

$$\begin{aligned} G(x,t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx-\lambda k t^2}\,\mathrm{d}k \\ &= \frac{e^{-\frac{x^2}{4\lambda t}}}{2\pi\sqrt{\lambda t}}\int_{-\infty}^{\infty}\exp\left(-\left(u-\frac{ix}{2\sqrt{\lambda t}}\right)^2\right)\mathrm{d}u \\ &= \frac{e^{-\frac{x^2}{4\lambda t}}}{\sqrt{4\pi\lambda t}} \end{aligned}$$

by the Gaussian integration lemma. Hence,

$$\theta(x,t) = \frac{1}{\sqrt{4\pi\lambda t}}\int_{-\infty}^{\infty} f(y)\exp\left(-\frac{(x-y)^2}{4\lambda t}\right)\mathrm{d}y\ .$$

Let us consider the specific case $f(x) = H(x)$, the Heaviside step function.

Recall the error function:

Definition 13.7. The *error function* is defined as

$$\operatorname{erf}(x) \coloneqq \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,\mathrm{d}t$$

such that $\operatorname{erf}(-\infty) = -1$ and $\operatorname{erf}(\infty) = 1$.

Then, using the substitution $v = \frac{y-x}{\sqrt{4\lambda t}}$,

$$\begin{aligned} \theta(x,t) &= \frac{1}{\sqrt{4\pi\lambda t}}\int_0^{\infty}\exp\left(-\frac{(x-y)^2}{4\lambda t}\right)\mathrm{d}y \\ &= \frac{1}{\sqrt{\pi}}\int_{-\frac{x}{\sqrt{4\lambda t}}}^{\infty} e^{-v^2}\,\mathrm{d}v \\ &= \frac{1}{\sqrt{\pi}}\left[\int_0^{\infty} e^{-v^2}\,\mathrm{d}v + \int_0^{\frac{x}{\sqrt{4\lambda t}}} e^{-v^2}\,\mathrm{d}v\right] \\ &= \frac{1}{2}\left[1+\operatorname{erf}\left(\frac{x}{\sqrt{4\lambda t}}\right)\right]. \end{aligned}$$

13.3 Laplace Transforms (Non-examinable)

The main shortcoming of Fourier transforms is the restriction to absolutely integrable functions. Many systems encountered in the real world involve growing functions, such as e^t, which we cannot manage with Fourier transforms, not even by resorting to distributional theory.

The Laplace transform provides a handle to treat such functions in a manner analogous to the Fourier domain. Furthermore, we will see that Laplace transforms have a natural way of incorporating boundary conditions, which makes them very suitable for solving ordinary differential equations.

13.3.1 Laplace Transform and Analytic Continuation

Definition 13.8. Let $f(t)$ be a function defined for all $t \geq 0$. The *Laplace transform* of $f(t)$ is given by

$$F(s) = \mathcal{L}\{f(t)\}(s) \coloneqq \int_0^\infty f(t)e^{-st}\,\mathrm{d}t\ ,$$

$s \in \mathbb{C}$ provided that the integral exists.

Remark. A sufficient condition for the existence of a Laplace transform is that $f(t)$ grows no more than exponential.

Let us consider a very important example. By direct integration, we can find

$$\mathcal{L}\{1\}(s) = \int_0^\infty e^{-st}\,\mathrm{d}t = \frac{1}{s}\,,$$

which may look completely trivial. However, notice that this integral only converges for $\mathrm{Re}(s) > 0$. Despite this, we may still extend the domain of the Laplace transform function $F(s)$ to the entire range where it is defined analytically. This process is the analytic continuation. In this case, we extended the domain of the Laplace transform from $U' = \{z \in \mathbb{C} \mid \mathrm{Re}(z) > 0\}$ to $U = \mathbb{C} \setminus \{0\}$, where the uniqueness of the extended function is guaranteed by the analyticity of the transformed function.

Here are some further examples.

- We can integrate by parts to find

$$\mathcal{L}\{t\}(s) = \int_0^\infty te^{-st}\,\mathrm{d}t = \left[-\frac{t}{s}e^{-ts}\right]_{t=0}^\infty + \int_0^\infty \frac{1}{s}e^{-st}\,\mathrm{d}t = \frac{1}{s^2}\,.$$

- For a constant λ, we can directly integrate

$$\mathcal{L}\{e^{\lambda t}\}(s) = \int_0^\infty e^{(\lambda - s)t}\,\mathrm{d}t = \frac{1}{s-\lambda}\,.$$

  Again, this integral is only defined if $\mathrm{Re}(s) > \mathrm{Re}(\lambda)$, but we can analytically continue the function for all $s \in \mathbb{C}$ except $s = \lambda$.

- Using the previous result, we find that

$$\mathcal{L}\{\sin t\} = \mathcal{L}\left\{\frac{1}{2i}(e^{it} - e^{-it})\right\}(s) = \frac{1}{2i}\left(\frac{1}{s-i} - \frac{1}{s+i}\right) = \frac{1}{s^2+1}\,.$$

13.3.2 Properties of the Laplace Transform

Proposition 13.9. The Laplace transform has the following properties.

(i) *Linearity.* For constants $\alpha, \beta \in \mathbb{C}$,

$$\mathcal{L}\{\alpha f + \beta g\} = \alpha \mathcal{L}\{f\} + \beta \mathcal{L}\{g\}\,.$$

(ii) *Translation.* For real constant $t_0 \in \mathbb{R}$,

$$\mathcal{L}\{f(t-t_0)H(t-t_0)\}(s) = e^{-st_0}F(s)\,.$$

(iii) *Scaling.* For constant $\lambda > 0$,

$$\mathcal{L}\{f(\lambda t)\}(s) = \frac{1}{\lambda}F\Big(\frac{s}{\lambda}\Big)\,.$$

(iv) *Shifting.* For a constant $s_0 \in \mathbb{C}$,

$$\mathcal{L}\big\{e^{s_o t}f(t)\big\}(s) = F(s-s_0)\,.$$

(v) *Transform of a derivative.*

$$\mathcal{L}\{f'(t)\}(s) = sF(s) - f(0)\,.$$

By repeatedly applying this formula, we find

$$\mathcal{L}\{f''(t)\}(s) = s\mathcal{L}\{f'(t)\}(s) - f'(0) = s^2F(s) - sf(0) - f'(0)$$

and so forth.

(vi) *Derivative of a transform.*

$$F'(s) = \mathcal{L}\{-tf(t)\}(s)\,.$$

More generally,

$$F^{(n)}(s) = \mathcal{L}\{(-t)^n f(t)\}(s)\,.$$

(vii) *Asymptotic limit.*

$$\lim_{s\to\infty} sF(s) = f(0)\,,$$
$$\lim_{s\to 0} sF(s) = f(\infty)\,,$$

provided that the limit $\lim_{t\to\infty} f(t)$ exists.

Proof.

(i) Follows from the linearity of integrals.

(ii) Setting $\tilde{t} = t - t_0$, we have

$$\begin{aligned}\int_0^\infty f(t-t_0)e^{-st}\,\mathrm{d}t &= \int_{-t_0}^\infty f(\tilde{t})e^{-s(\tilde{t}+t_0)}\,\mathrm{d}\tilde{t}\\ &= e^{-st_0}\int_{-t_0}^\infty f(\tilde{t})e^{-s\tilde{t}}\,\mathrm{d}\tilde{t}\,.\end{aligned}$$

So

$$\begin{aligned}\int_0^\infty f(t-t_0)H(t-t_0)e^{-st}\,\mathrm{d}t &= e^{-st_0}\int_{-t_0}^\infty f(\tilde{t})H(\tilde{t})e^{-s\tilde{t}}\,\mathrm{d}\tilde{t}\\ &= e^{-st_0}\int_0^\infty f(\tilde{t})e^{-s\tilde{t}}\,\mathrm{d}\tilde{t} = e^{-st_0}F(s)\,.\end{aligned}$$

(iii) Define $\tilde{t} = \lambda t$, we find

$$\int_0^\infty f(\lambda t)e^{-st}\,\mathrm{d}t = \int_0^\infty f(\tilde{t})e^{-\frac{s}{\lambda}\tilde{t}}\frac{1}{\lambda}\,\mathrm{d}\tilde{t} = \frac{1}{\lambda}F\Big(\frac{s}{\lambda}\Big)\,.$$

(iv)
$$\int_0^\infty e^{s_0 t} f(t) e^{-st} \,\mathrm{d}t = F(s - s_0)\,.$$

(v)
$$\int_0^\infty f'(t) e^{-st} \,\mathrm{d}t = [f(t)e^{-st}]_{t=0}^\infty - \int_0^\infty -sf(t)e^{-st} \,\mathrm{d}t = sF(s) - f(0)\,.$$
Once again, we encounter the subtlety of analytic continuation. The proof breaks down if the integral $\int_0^\infty f(t)e^{-st} \,\mathrm{d}t$ does not exist. The relation still holds for analytically continued Laplace transforms, though.

(vi) Differentiating the definition of the Laplace transform gives
$$F'(s) = \int_0^\infty -tf(t)e^{-st} \,\mathrm{d}t = \mathcal{L}\{-tf(t)\}\,.$$

(vii) We have
$$sF(s) = f(0) + \int_0^\infty f'(t)e^{-st} \,\mathrm{d}t\ .$$
By requirement, the limit $\lim_{t\to\infty} f(t)$ exists, so $f(t)$ and $f'(t)$ do not grow faster than exponential. For $s \to \infty$, the integral on the right-hand side therefore vanishes and we obtain
$$\lim_{s\to\infty} sF(s) = f(0)\,.$$
In the limit $s \to 0$, on the other hand, we have $e^{-st} \to 1$, the integral just becomes $\int_0^\infty f'(t) \,\mathrm{d}t$, and we recover
$$\lim_{s\to 0} sF(s) = f(\infty)\,.$$

□

Example. Previously, we have
$$\mathcal{L}\{1\} = \frac{1}{s}\,.$$
From Proposition 13.9 (vi), we find
$$\mathcal{L}\{t^n\}(s) = (-1)^n \frac{\mathrm{d}^n}{\mathrm{d}s^n}\frac{1}{s} = \frac{n!}{s^{n+1}}\,.$$

Remark. We can generalise this formula to obtain the generalisation of factorials into $\mathbb{C}$.

Definition 13.10. The *Euler's gamma function* is defined for complex number $n \in \mathbb{C} \setminus \mathbb{Z}_{\leq 0}$ as
$$\Gamma(n) \coloneqq \int_0^\infty e^{-t} t^{n-1} \,\mathrm{d}t$$
such that
$$\Gamma(n) = \mathcal{L}\left\{t^{n-1}\right\}(1) =: (n-1)!\,.$$

13.3.3 The Inverse Laplace Transform

Theorem 13.11 (Inverse Laplace transform). For a given function $F(s)$, its inverse Laplace transform that recovers $f(t)$ is given by the *Bromwich inversion formula*
$$f(t) = \frac{1}{2\pi i}\int_{\alpha-i\infty}^{\alpha+i\infty} F(s)e^{st} \,\mathrm{d}s\ ,$$
where α is a real constant chosen such that the *Bromwich inversion contour* $\gamma = \{s \in \mathbb{C} \mid \mathrm{Re}(s) = \alpha\}$ lies to the right of all singularities of $F(s)$.

Proof. Since $f(t)$ has a Laplace transform, we have $f(t) = 0$ for $t < 0$ and f does not grow faster than exponential. We can therefore choose an $\alpha \in \mathbb{R}$ such that

$$g(t) = f(t)e^{-\alpha t}$$

decays exponentially as $t \to \infty$ and therefore has a Fourier transform

$$\tilde{g}(\omega) = \int_{-\infty}^{\infty} f(t)e^{-\alpha t}e^{-i\omega t} = F(\alpha + i\omega)\,.$$

This enables us to apply the inverse Fourier transform, so

$$g(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\alpha + i\omega)e^{i\omega t}\,\mathrm{d}\omega\ .$$

Substitute $s = \alpha + i\omega$, we obtain

$$f(t)e^{-\alpha t} = \frac{1}{2\pi i}\int_{\alpha+i\infty}^{\alpha-i\infty} F(s)e^{(s-\alpha)t}\,\mathrm{d}s$$

$$f(t) = \frac{1}{2\pi i}\int_{\alpha-i\infty}^{\alpha+i\infty} F(s)e^{st}\,\mathrm{d}s\ .$$

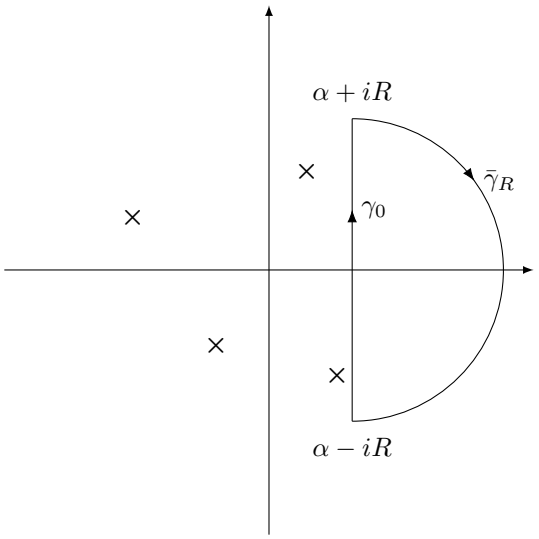
The additional requirement that the contour $\mathrm{Re}(s) > \alpha$ lies to the right of all singularities of $F(s)$ fixes a constant of integration and thus ensures that $f(t) = 0$ for $t < 0$. □

In practice, the Laplace transform and its inverse are often applied to functions with a finite number of singularities. This simplifies the inverse Laplace transform considerably.

Theorem 13.12 (Inverse Laplace transform). Let $F(s)$ be the Laplace transform of a function $f(t)$ and have only a finite number of isolated singularities $s_k \in \mathbb{C}$. Let $F(s) \to 0$ as $|s| \to \infty$. Then $f(t) = 0$ for $t < 0$ and for $t > 0$,

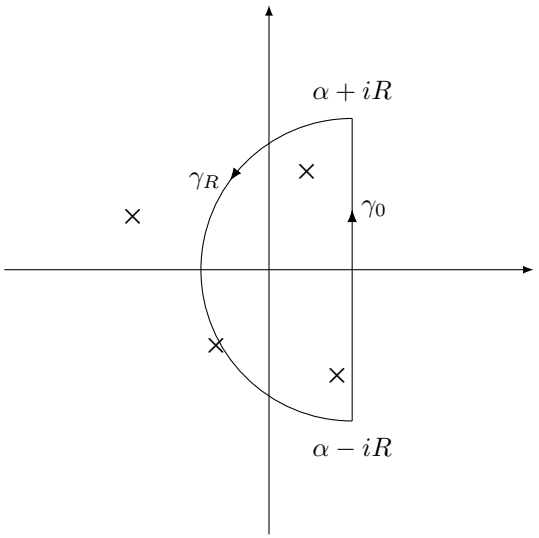
$$f(t) = \sum_{k=1}^{n} \operatorname*{res}_{s=s_k}\left(F(s)e^{st}\right).$$

Proof. First, let us consider the case $t < 0$ and construct the contour $\bar{\gamma} = \gamma_0 + \bar{\gamma}_R$ as shown below.



By Jordan's lemma,

$$\left|\int_{\bar{\gamma}_R} F(s)e^{st}\,\mathrm{d}s\right| \to 0 \text{ as } R \to \infty\,.$$



Therefore,

$$\int_{\gamma_0} F(s)e^{st}\,\mathrm{d}s = \int_{\bar{\gamma}} F(s)e^{st}\,\mathrm{d}s$$

in the limit $R \to \infty$. The contour does not enclose any singularities of the integrand, so by Cauchy's theorem, we get $f(t) = 0$ for $t < 0$.

For $t > 0$, let us consider the contour $\gamma = \gamma_0 + \gamma_R$ enclosed on the left of the vertical line as shown below.

Since $F(s)$ only has a finite number of isolated singularities, we will enclose all of them as $R \to \infty$. As before, the contribution of γ_R vanishes as $R \to \infty$ by Jordan's lemma. We can use the residue theorem to compute $f(t)$ using the Bromwich inversion formula

$$\begin{aligned} f(t) &= \frac{1}{2\pi i} \lim_{R\to\infty} \int_{\gamma_0} F(s)e^{st}\,\mathrm{d}s \\ &= \frac{1}{2\pi i} \lim_{R\to\infty} \int_{\gamma} F(s)e^{st}\,\mathrm{d}s \\ &= \sum_{k=1}^{n} \operatorname*{res}_{s=s_k} \left(F(s)e^{st}\right). \end{aligned}$$

□

Example. Consider

$$F(s) = \frac{1}{s-1}\,,$$

which has a simple pole at $s = 1$. For $t > 0$, we have

$$f(t) = \operatorname*{res}_{s=1} \left(\frac{e^{st}}{s-1}\right) = e^t\,.$$

Example. Consider

$$F(s) = s^{-n}$$

$n \in \mathbb{N}$, we have a pole of order n at $s = 0$. We have

$$f(t) = \operatorname*{res}_{s=0} \left(\frac{e^{st}}{s^n}\right) = \lim_{s\to 0} \left[\frac{1}{(n-1)!} \frac{\mathrm{d}^{n-1}}{\mathrm{d}s^{n-1}} e^{st}\right] = \frac{t^{n-1}}{(n-1)!}\,.$$

13.3.4 Solving Differential Equations using Laplace Transform

Example. Consider the ODE

$$t\ddot{f}(t) - t\dot{f}(t) + f(t) = 2\,,\ f(0) = 2\,,\ \dot{f}(0) = -1\,.$$

The transform of this ODE gives

$$-s^2F'(s) - 2sF(s) + f(0) + sF'(s) + F(s) + F(s) = \frac{2}{s}\,,$$

which organises to

$$sF'(s) + 2F(s) = \frac{2}{s}\,.$$

This equation can be easily solved, with a general solution

$$F = \frac{2}{s} + \frac{A}{s^2}\,.$$

Inverse Laplace transform gives

$$f(t) = 2 + At\,,$$

where the boundary condition determines $A = -1$.

Similar methods can be applied to partial differential equations as well.

13.3.5 The Convolution Theorem

Recall that the convolution of two functions $f, g : \mathbb{R} \to \mathbb{R}$ is defined as

$$(f * g)(t) = \int_{-\infty}^{\infty} f(t-u)g(u)\,\mathrm{d}u\ .$$

Note that if $f(t)$ and $g(t)$ vanishes for $t < 0$, then

$$(f * g)(t) = \int_{0}^{t} f(t-u)g(u)\,\mathrm{d}u\ .$$

Theorem 13.13 (The convolution theorem). The Laplace transform of a convolution is given by

$$\mathcal{L}\{f * g\}(s) = \mathcal{L}\{f\}(s) \cdot \mathcal{L}\{g\}(s) = F(s)G(s)\,.$$

Proof.

$$\begin{aligned}\mathcal{L}\{f * g\}(s) &= \int_0^\infty \left[\int_0^t f(t-u)g(u)\,\mathrm{d}u\right] e^{-st}\,\mathrm{d}t \\ &= \int_0^\infty \left[\int_u^\infty f(t-u)g(u)e^{-st}\,\mathrm{d}t\right]\mathrm{d}u\end{aligned}$$

by swapping the order of integration. Defining $x = t - u$, we obtain

$$\begin{aligned}\mathcal{L}\{f * g\}(s) &= \int_0^\infty \left[\int_0^\infty f(x)g(u)e^{-sx}e^{-su}\,\mathrm{d}x\right]\mathrm{d}u \\ &= \int_0^\infty \left[\int_0^\infty f(x)e^{-sx}\,\mathrm{d}x\right] g(u)e^{-su}\,\mathrm{d}u = F(s)G(s)\,.\end{aligned}$$

□

Example. Suppose that we wish to find the inverse of

$$H(s) = \frac{1}{s(s^2+1)}\,.$$

Set $F(s) = s^{-1}$ and $G(s) = (s^2+1)^{-1}$, so that $f(t) = 1$ and $g(t) = \sin t$. We have

$$\mathcal{L}^{-1}\left\{\frac{1}{s(s^2+1)}\right\}(t) = 1 * \sin t = \int_0^t \sin u \,\mathrm{d}u = 1 - \cos t\,.$$

Example. Consider the ODE

$$4\ddot{f}(t) + f(t) = h(t)\,,\ f(0) = 3\,,\ \dot{f}(0) = -7\,.$$

We can transform this equation to

$$4\Big[s^2F(s) - sf(0) - \dot{f}(0)\Big] + F(s) = H(s)\,,$$

which simplifies to

$$F(s) = \frac{3s}{s^2+\frac{1}{4}} - \frac{7}{s^2+\frac{1}{4}} + \frac{H(s)}{4}\frac{1}{s^2+\frac{1}{4}}\,.$$

We can directly inverse transform the first two terms, and the inverse transform of the third term can be solved using the convolution theorem. We have

$$\begin{aligned} f(t) &= 3\cos\frac{t}{2} - 14\sin\frac{t}{2} + \frac{1}{4}h(t) * \left(2\sin\frac{t}{2}\right) \\ &= 3\cos\frac{t}{2} - 14\sin\frac{t}{2} + \frac{1}{2}\int_0^t \sin\frac{u}{2}h(t-u)\,\mathrm{d}u\ . \end{aligned}$$

14 Partial Differential Equations on Unbounded Domains

14.1 Well-posedness (Non-examinable)

Claim 14.1. Hadamard famously declared that a problem is well-posed if

(i) a solution exists;

(ii) the solution is unique;

(iii) the solution depends continuously on the initial and boundary data.

Points (i)-(ii) are self-explanatory, but (iii) is far more subtle. If the initial and/or boundary data for a problem lie in a space X and the solution lies in Y, then the solution to a generic initial-boundary value problem on $\Omega \times (0, \infty)$ describes an abstract map

$$S_t : X \to Y \,.$$

We will offer some examples.

Example. Consider the differential equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\kappa x \,,\ x(0) = x_0 \,.$$

It has solution

$$X_0(t) = x_0 e^{-\kappa t} \,.$$

The solution clearly exists and is unique. If we consider the same problem but with initial data $x(0) = x_1$, then

$$|X_1(t) - X_0(t)| = e^{-\kappa t}|x_1 - x_0| \,.$$

Therefore, if we measure the 'closeness' of solutions in terms of *uniform norm*:

$$\|f\|_\infty \coloneqq \sup_x \|f(x)\| \,,$$

we clearly have

$$\|X_1 - X_0\|_\infty \leq |x_1 - x_0| \,.$$

So if two solutions start close, then they will always remain "close". The problem is well-posed.

Example. Consider the initial-boundary value problem for the backward heat equation on $\Omega = (0, \pi)$.

$$\begin{cases} \dfrac{\partial \varphi}{\partial t} + \kappa \nabla^2 \varphi = 0 & (\mathbf{x}, t) \in \Omega \times (0, \infty) \\ \varphi = f & (\mathbf{x}, t) \in \Omega \times \{t = 0\} \\ \varphi = 0 & (\mathbf{x}, t) \in \partial\Omega \times (0, \infty) \,. \end{cases}$$

When $f(x) = 0$, we have the solution $\varphi = 0$. For $f(x) = f_n(x) = \frac{1}{n}\sin(nx)$, we have the solution $\varphi_n(x, t) = \frac{1}{n}\sin(nx)e^{\kappa n^2 t}$. Note that

$$\|f - f_n\|_\infty = \frac{1}{n} \to 0 \text{ as } n \to \infty \,.$$

So the initial data for these solutions gets arbitrarily close. However, for arbitrary $t = T$,

$$\|\varphi(x, T) - \varphi_n(x, T)\|_\infty = \frac{1}{n} e^{\kappa n^2 T} \to \infty \text{ as } n \to \infty \,.$$

This problem is not well-posed, even locally in time.

14.2 Method of Characteristics (Non-examinable)

In this section, we will study problems of the form

$$\begin{cases} a(x,y)\frac{\partial u}{\partial x} + b(x,y)\frac{\partial u}{\partial y} = c(x,y,u) & (x,y) \in \mathbb{R}^2 \\ u = f & \text{on curve } \mathcal{C}\,. \end{cases}$$

These are called *quasi-linear* partial differential equations, since c might be non-linear in u. Consider the vector field

$$\mathbf{v} = \begin{pmatrix} a(x,y) \\ b(x,y) \end{pmatrix}.$$

Then we have

$$a(x,y)\frac{\partial u}{\partial x} + b(x,y)\frac{\partial u}{\partial y} = \mathbf{v}\cdot\boldsymbol{\nabla} u\,,$$

the directional derivative of u along the vector field $\mathbf{v}$ at the point.

Definition 14.2. The *integral curves* of a vector field $\mathbf{v}$ are defined as

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{v}\,.$$

Remark. The integral curve goes along the vector field $\mathbf{v}$, with a parameter t along the curve.

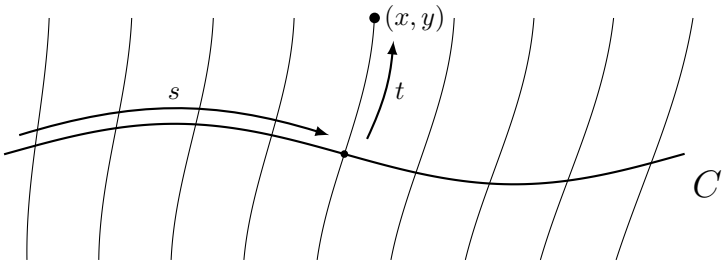
We call these the *characteristic curves* for the PDE. Suppose that we have solved these equations for $x(t)$ and $y(t)$ with the initial conditions $(x(0), y(0)) = (x_0, y_0) \in \mathcal{C}$. This will give us a family of characteristic curves crossing $\mathcal{C}$ at $t = 0$. Define $z(t)$ to be the value of $u(x,y)$ evaluated on the characteristic curve:

$$z(t) = u(x(t), y(t))\,,$$

by chain rule, we have

$$\begin{aligned} \frac{\mathrm{d}z}{\mathrm{d}t} &= \frac{\mathrm{d}x}{\mathrm{d}t}\frac{\partial u}{\partial x} + \frac{\mathrm{d}y}{\mathrm{d}t}\frac{\partial u}{\partial y} \\ &= a\frac{\partial u}{\partial x} + b\frac{\partial u}{\partial y} \\ &= c(x(t), y(t), z)\,. \end{aligned}$$

This gives us an ordinary differential equation for z, with initial value $z_0 = u(x_0, y_0) = f(x_0, y_0)$.



To get the solution at an arbitrary point (x,y), we find the characteristic curve that goes through it and trace the solution back to the curve $\mathcal{C}$. This will mean inverting a relationship between (x,y) and (s,t), where s is the coordinate along $\mathcal{C}$ and t is the coordinate along a characteristic curve.

We will want to make sure the vector field $\mathbf{v} = (a,b)$ is not tangent to the curve $\mathcal{C}$ on which the initial data is on, otherwise we would not be able to carry the initial data off the curve $\mathcal{C}$ along characteristic curves to an arbitrary point (x,y). If $\mathbf{n}$ denotes the normal to the curve $\mathcal{C}$, this condition is $\mathbf{n}\cdot\mathbf{v} \neq 0$ for $(x,y) \in \mathcal{C}$. This is called the *non-characteristic condition*.

Example. Consider the problem

$$\begin{cases} (x^2+1)\dfrac{\partial u}{\partial x} + \dfrac{\partial u}{\partial y} = u + 1 \\ u(0,y) = f(y)\,. \end{cases}$$

The characteristic curves are given by

$$\frac{\mathrm{d}x}{\mathrm{d}t} = x^2 + 1 \quad \frac{\mathrm{d}y}{\mathrm{d}t} = 1\,,$$

with initial data on the curve

$$(x(0), y(0)) = (0, s)\,.$$

This gives the solution

$$\begin{cases} x = \tan t \\ y = s + t\,. \end{cases}$$

The z-equation is

$$\frac{\mathrm{d}z}{\mathrm{d}t} = z + 1\,,$$

which gives the solution

$$z = (1 + z_0)e^t - 1\,,$$

where z_0 is the initial value

$$z_0 = u(x_0, y_0) = u(0, s) = f(s)\,.$$

This gives the solution

$$z(t, s) = (1 + f(s))e^t - 1\,.$$

Invert $(t, s) \to (x, y)$:

$$\begin{cases} t = \arctan x \\ s = y - \arctan x \end{cases}$$

to obtain the solution

$$u(x, y) = [1 + f(y - \arctan x)]e^{\arctan x}\,.$$

14.3 Higher-dimensional Fourier Transform (Non-examinable)

Definition 14.3. For $f : \mathbb{R}^n \to \mathbb{C}$, its *Fourier transform* is defined as

$$\tilde{f}(\boldsymbol{\lambda}) \coloneqq \int e^{-i\boldsymbol{\lambda}\cdot\mathbf{x}} f(\mathbf{x})\,\mathrm{d}^n\mathbf{x}\,.$$

Theorem 14.4. The *inverse Fourier transform* is given by

$$f(\mathbf{x}) = \frac{1}{(2\pi)^n} \int e^{i\boldsymbol{\lambda}\cdot\mathbf{x}} \tilde{f}(\boldsymbol{\lambda})\,\mathrm{d}^n\boldsymbol{\lambda}\,.$$

Definition 14.5. The convolution of two functions $f, g : \mathbb{R}^n \to \mathbb{C}$ is

$$(f * g)(\mathbf{x}) \coloneqq \int f(\mathbf{x} - \mathbf{y}) g(\mathbf{y})\,\mathrm{d}^n\mathbf{y}\,.$$

Theorem 14.6 (The convolution theorem).

$$\mathcal{F}_{\mathbf{x}\to\boldsymbol{\lambda}}[(f * g)(\mathbf{x})] = \tilde{f}(\boldsymbol{\lambda})\tilde{g}(\boldsymbol{\lambda})\,.$$

Proposition 14.7.

$$\mathcal{F}_{\mathbf{x}\to\boldsymbol{\lambda}}\left[\frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}}\frac{\partial^{\alpha_2}}{\partial x_2^{\alpha_2}}\cdots\frac{\partial^{\alpha_n}}{\partial x_n^{\alpha_n}} f(\mathbf{x})\right] = (i\lambda_1)^{\alpha_1}\ldots(i\lambda_n)^{\alpha_n}\tilde{f}(\boldsymbol{\lambda})\,.$$

14.4 Green's Function for the Heat Equation (Non-examinable)

Definition 14.8. If the domain of a problem is all of $\mathbb{R}^n$ space, then the Green's function is known as the *fundamental solution*.

We will solve the forced heat equation on $\mathbb{R}^n$:

$$\begin{cases} \dfrac{\partial u}{\partial t} - \kappa \nabla^2 u = F(\mathbf{x}, t) & (\mathbf{x}, t) \in \mathbb{R}^n \times (0, \infty) \\ u(\mathbf{x}, 0) = f(\mathbf{x}) & x \in \mathbb{R}^n \,. \end{cases} \tag{$\dagger$}$$

We will split this problem into two parts.

(a) Zero forcing and non-zero initial data.

$$\begin{cases} \dfrac{\partial u}{\partial t} - \kappa \nabla^2 u = 0 & (\mathbf{x}, t) \in \mathbb{R}^n \times (0, \infty) \\ u(\mathbf{x}, 0) = f(\mathbf{x}) & x \in \mathbb{R}^n \,. \end{cases} \tag{$\dagger_a$}$$

(b) Zero initial data and non-zero forcing.

$$\begin{cases} \dfrac{\partial u}{\partial t} - \kappa \nabla^2 u = F(\mathbf{x}, t) & (\mathbf{x}, t) \in \mathbb{R}^n \times (0, \infty) \\ u(\mathbf{x}, 0) = 0 & x \in \mathbb{R}^n \,. \end{cases} \tag{$\dagger_b$}$$

Definition 14.9. The *heat kernel* in n dimensions is defined as

$$K(\mathbf{x}; t) \coloneqq \frac{1}{(4\pi\kappa t)^{\frac{n}{2}}} \exp\left[-\frac{|x|^2}{4\kappa t}\right].$$

Proposition 14.10. For each $t > 0$, the Fourier transform of the heat kernel is

$$\tilde{K}(\boldsymbol{\lambda}; t) = e^{-\kappa t |\boldsymbol{\lambda}|^2} \,.$$

Proof. If

$$g(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{|vbx|^2}{2}} \,,$$

then

$$\begin{aligned} \tilde{g}(\boldsymbol{\lambda}) &= \int e^{-i\boldsymbol{\lambda}\cdot\mathbf{x}} g(\mathbf{x}) \, \mathrm{d}^n \mathbf{x} \\ &= \int \mathrm{d}x_1 \ldots \int \mathrm{d}x_n \left[\frac{1}{\sqrt{2\pi}} e^{-i\lambda_1 x_1 - x_1^2}\right] \ldots \left[\frac{1}{\sqrt{2\pi}} e^{-i\lambda_n x_n - x_n^2}\right] \\ &= e^{-\frac{\lambda_1^2}{2}} \ldots e^{-\frac{\lambda_n^2}{2}} \\ &= e^{-\frac{|\boldsymbol{\lambda}|^2}{2}} \,. \end{aligned}$$

Then

$$\begin{aligned} \tilde{K}(\boldsymbol{\lambda}, t) &= \frac{1}{(4\pi\kappa t)^{\frac{n}{2}}} \int e^{-i\boldsymbol{\lambda}\cdot\mathbf{x} - \frac{|\mathbf{x}|^2}{4\kappa t}} \, \mathrm{d}^n \mathbf{x} \\ &= \frac{(2\kappa t)^{\frac{n}{2}}}{(4\pi\kappa t)^{\frac{n}{2}}} \int e^{-i(\boldsymbol{\lambda}\sqrt{2\kappa t})\cdot\mathbf{x} - \frac{|\mathbf{x}|^2}{2}} \, \mathrm{d}^n \mathbf{x} \qquad \mathbf{x}' = \frac{\mathbf{x}}{\sqrt{2\kappa t}} \,, \text{ then drop the primes} \\ &= \int e^{-i(\boldsymbol{\lambda}\sqrt{2\kappa t})\cdot\mathbf{x}} g(\mathbf{x}) \, \mathrm{d}^n \mathbf{x} \\ &= \tilde{g}(\boldsymbol{\lambda}\sqrt{2\kappa t}) \\ &= e^{-\kappa t |\boldsymbol{\lambda}|^2} \,. \end{aligned}$$

□

Using this result, we can find the solution to equation $(\dagger_a)$.

Proposition 14.11. The solution to equation $(\dagger_a)$ is

$$u(\mathbf{x}, t) = (K * f)(\mathbf{x}) = \int K(\mathbf{x} - \mathbf{y}, t) f(\mathbf{y}) \, \mathrm{d}^n \mathbf{y} \ .$$

Proof. Take the Fourier transform of $(\dagger_a)$ to get

$$\begin{cases} \dfrac{\partial \tilde{u}}{\partial t} + \kappa |\boldsymbol{\lambda}|^2 \tilde{u} = 0 \\ \tilde{u}(\boldsymbol{\lambda}, 0) = \tilde{f}(\boldsymbol{\lambda}) \, . \end{cases}$$

Solving this differential equation gives

$$\begin{aligned} \tilde{u}(\boldsymbol{\lambda}, t) &= \tilde{u}(\boldsymbol{\lambda}, 0) e^{-\kappa |\boldsymbol{\lambda}|^2 t} \\ &= \tilde{f}(\boldsymbol{\lambda}) \tilde{K}(\boldsymbol{\lambda}, t) \, . \end{aligned}$$

Therefore, by the convolution theorem, $u(\mathbf{x}, t) = (K * f)(\mathbf{x})$. □

Proposition 14.12. The solution to equation $(\dagger_b)$ is

$$u(\mathbf{x}, t) = \int_0^t \left[\int K(\mathbf{x} - \mathbf{y}, t - s) F(\mathbf{y}, s) \, \mathrm{d}^n \mathbf{y} \right] \mathrm{d}s \ .$$

Proof. Take the Fourier transform of $(\dagger_b)$ to get

$$\begin{cases} \dfrac{\partial \tilde{u}}{\partial t} + \kappa |\boldsymbol{\lambda}|^2 \tilde{u} = \tilde{F}(\boldsymbol{\lambda}, t) \\ \tilde{u}(\boldsymbol{\lambda}, 0) = 0 \, . \end{cases}$$

This differential equation is equivalent to

$$\frac{\partial}{\partial t} \left[e^{\kappa |\boldsymbol{\lambda}|^2 t} \tilde{u}(\boldsymbol{\lambda}, t) \right] = e^{\kappa |\boldsymbol{\lambda}|^2 t} \tilde{F}(\boldsymbol{\lambda}, t) \, ,$$

and so

$$\begin{aligned} \tilde{u}(\boldsymbol{\lambda}, t) &= \int_0^t e^{-\kappa (t-s) |\boldsymbol{\lambda}|^2} \tilde{F}(\boldsymbol{\lambda}, s) \, \mathrm{d}s \\ &= \int_0^t \tilde{K}(\boldsymbol{\lambda}, t - s) \tilde{F}(\boldsymbol{\lambda}, s) \, \mathrm{d}s \ . \end{aligned}$$

The result follows from the convolution theorem. □

Theorem 14.13. The solution to the heat equation in $\mathbb{R}^n$ of the form $(\dagger)$ is

$$u(\mathbf{x}, t) = (K * f)(\mathbf{x}) + \int_0^t \left[\int K(\mathbf{x} - \mathbf{y}, t - s) F(\mathbf{y}, s) \, \mathrm{d}^n \mathbf{y} \right] \mathrm{d}s \ .$$

Proof. Principle of superposition. □

If we set

$$G(\mathbf{x}, t; \mathbf{y}, s) = K(\mathbf{x} - \mathbf{y}, t - s) H(t - s)$$

and assume that $F(\mathbf{x}, t) = 0$ for $t < 0$, then we can write solution to $(\dagger_b)$ as

$$u(\mathbf{x}, t) = \iint G(\mathbf{x}, t; \mathbf{y}, s) F(\mathbf{y}, s) \, \mathrm{d}^n \mathbf{x} \, \mathrm{d}s \ .$$

This is the Green's function we have seen before.

Theorem 14.14. The Green's function for the heat equation on $\mathbb{R}^n$ with vanishing initial data is

$$\begin{aligned} G(\mathbf{x},t;\mathbf{y},s) &= K(\mathbf{x}-\mathbf{y},t-s)H(t-s) \\ &= \begin{cases} \dfrac{1}{[4\pi\kappa(t-s)]^{\frac{n}{2}}} \exp\left[-\dfrac{|\mathbf{x}-\mathbf{y}|^2}{4\kappa(t-s)}\right] & t > s \\ 0 & t < s\,. \end{cases} \end{aligned}$$

Proof. The Green's function should satisfy the equation

$$\frac{\partial G}{\partial t} - \kappa\nabla^2 G = \delta(\mathbf{x}-\mathbf{y})\delta(t-s)\,,$$

or equivalently its Fourier transform

$$\frac{\partial \tilde{G}}{\partial t} + \kappa|\boldsymbol{\lambda}|^2\tilde{G} = e^{-i\boldsymbol{\lambda}\cdot\mathbf{y}}\delta(t-s)\,.$$

Our proposed Green's function has Fourier transform

$$\tilde{G}(\boldsymbol{\lambda},t;\mathbf{y},s) = H(t-s)\tilde{K}(\boldsymbol{\lambda};t-s)e^{-i\boldsymbol{\lambda}\cdot\mathbf{y}} = H(t-s)e^{-i\boldsymbol{\lambda}\cdot\mathbf{y}}e^{-\kappa(t-s)|\boldsymbol{\lambda}|^2}\,.$$

This function clearly solves the Fourier transformed equation for $t > s$ and $t < s$.

If we integrate the transformed equation over $(s-\epsilon, s+\epsilon)$ and take $\epsilon \to 0$, we should require

$$\tilde{G}(\boldsymbol{\lambda},t;\mathbf{y},s)\Big|_{t=s_-}^{t=s_+} = e^{-i\boldsymbol{\lambda}\cdot\mathbf{y}}\,.$$

This is satisfied by the proposed Green's function since

$$\tilde{G}(\boldsymbol{\lambda},s_+;\mathbf{y},s) = e^{-i\boldsymbol{\lambda}\cdot\mathbf{y}} \text{ and } \tilde{G}(\boldsymbol{\lambda},s_-;\mathbf{y},s) = 0\,.$$

□

14.4.1 Duhamel's Principle

The fact that the same function $K(\mathbf{x},t)$ appeared in both the solution to the homogeneous equation with inhomogeneous boundary conditions ($\dagger_a$), and the solution to the inhomogeneous equation with homogeneous boundary conditions ($\dagger_b$) is not a coincidence.

To see this, let us return to the problem ($\dagger_a$) but now suppose we impose the initial condition

$$u(\mathbf{x},s) = f(\mathbf{x})$$

at time $t = s$ rather than $t = 0$. A simple translation shows that for times $t > s$, the problem is solved by

$$u(\mathbf{x},t) = \int K(\mathbf{x}-\mathbf{y},t-s)f(\mathbf{y})\,\mathrm{d}^n\mathbf{y}\,. \qquad (*_a)$$

On the other hand, the solution for the forced problem ($\dagger_b$) takes the form

$$u(\mathbf{x},t) = \int_0^t \left[\int K(\mathbf{x}-\mathbf{y},t-s)F(\mathbf{y},s)\,\mathrm{d}^n\mathbf{y}\right]\mathrm{d}s\,. \qquad (*_b)$$

Now suppose that for each fixed time $t = s$, we view the forcing term $F(\mathbf{y},t)$ as an initial condition imposed at $t = s$. The integral in square brackets above represents the effect of this condition propagated to time t as in ($*_a$). Finally, the time integral in ($*_b$) expresses the solution to the forced problem as the accumulation (superposition) of the effects from all these conditions at times s earlier than t, each propagated for time interval $t-s$ up to time t. The upper limit t of this integral arose

from the step function $H(t-s)$ in the Green's function and expresses *causality*: the solution at time t depends only on the cumulative effects of conditions applied at earlier times $s < t$.

The relation between solutions to homogeneous equations with inhomogeneous boundary conditions and inhomogeneous equations with homogeneous boundary conditions is known as *Duhamel's principle*.

14.5 Green's Function for the Wave Equation (Non-examinable)

Consider the general initial value problem for the wave equation on $\mathbb{R}^n$.

$$\begin{cases} \dfrac{\partial^2 u}{\partial t^2} - c^2\nabla^2 u = F(\mathbf{x},t) & (\mathbf{x},t) \in \mathbb{R}^n \times (0,\infty) \\ u(\mathbf{x},0) = f(\mathbf{x}) & \mathbf{x} \in \mathbb{R}^n \\ \dfrac{\partial u}{\partial t} = g(\mathbf{x}) & \mathbf{x} \in \mathbb{R}^n\,. \end{cases} \tag{$\dagger$}$$

We again split the problem into two parts.

(a) Zero forcing and non-zero initial data.

$$\begin{cases} \dfrac{\partial^2 u}{\partial t^2} - c^2\nabla^2 u = 0 & (\mathbf{x},t) \in \mathbb{R}^n \times (0,\infty) \\ u(\mathbf{x},0) = f(\mathbf{x}) & \mathbf{x} \in \mathbb{R}^n \\ \dfrac{\partial u}{\partial t} = g(\mathbf{x}) & \mathbf{x} \in \mathbb{R}^n\,. \end{cases} \tag{$\dagger_a$}$$

(b) Zero initial data and non-zero forcing.

$$\begin{cases} \dfrac{\partial^2 u}{\partial t^2} - c^2\nabla^2 u = F(\mathbf{x},t) & (\mathbf{x},t) \in \mathbb{R}^n \times (0,\infty) \\ u(\mathbf{x},0) = 0 & \mathbf{x} \in \mathbb{R}^n \\ \dfrac{\partial u}{\partial t} = 0 & \mathbf{x} \in \mathbb{R}^n\,. \end{cases} \tag{$\dagger$}$$

We will use the function $\Phi(\mathbf{x},t)$ defined implicitly by

$$\tilde{\Phi}(\boldsymbol{\lambda},t) \coloneqq \frac{\sin c|\boldsymbol{\lambda}|t}{c|\boldsymbol{\lambda}|}\,,$$

and so by Fourier inversion theorem,

$$\Phi(\mathbf{x},t) = \mathcal{I}_{\boldsymbol{\lambda}\to\mathbf{x}}\left[\frac{\sin c|\boldsymbol{\lambda}|t}{c|\boldsymbol{\lambda}|}\right] = \frac{1}{(2\pi)^n}\int e^{i\boldsymbol{\lambda}\cdot\mathbf{x}}\frac{\sin c|\boldsymbol{\lambda}|t}{c|\boldsymbol{\lambda}|}\,\mathrm{d}^n\boldsymbol{\lambda}\,.$$

The computation of this function is difficult in arbitrary dimensions. It depends on whether n is even or odd. We will consider two important cases.

Proposition 14.15. In $\mathbb{R}$,

$$\Phi(\mathbf{x},t) = \frac{1}{2c}(H(x+ct) - H(x-ct))\,.$$

Proof. Note that

$$H(x+ct) - H(x-ct) = \begin{cases} 1 & |x| < ct \\ 0 & \text{otherwise}\,. \end{cases}$$

It is easy to check that its Fourier transform is $\tilde{\Phi}$. □

Proposition 14.16. In $\mathbb{R}^3$,

$$\Phi(\mathbf{x}, t) = \frac{1}{4\pi c}[\delta(|\mathbf{x}| - ct) - \delta(|\mathbf{x}| + ct)] .$$

Proof. We will compute the inverse Fourier transform using spherical polar coordinates

$$\boldsymbol{\lambda} = (r \sin\theta \cos\phi, r \sin\theta \sin\phi, r \cos\theta) .$$

We will align the λ_3 axis with the direction of $\mathbf{x}$, so $\mathbf{x} \cdot \boldsymbol{\lambda} = r|\mathbf{x}| \cos\theta$. Consider

$$\begin{aligned}
\int_{|\boldsymbol{\lambda}|<R} e^{i\boldsymbol{\lambda}\cdot\mathbf{x}} \tilde{\Phi}(\boldsymbol{\lambda}, t)\, \mathrm{d}^3\boldsymbol{\lambda} &= \int_0^{2\pi} \mathrm{d}\phi \int_0^{\pi} \mathrm{d}\theta \int_0^R \mathrm{d}r\, \frac{\sin rct}{rc} e^{-i|\mathbf{x}| r\cos\theta} r^2 \sin\theta \\
&= 2\pi \int_0^{\pi} \mathrm{d}\theta \int_0^R \mathrm{d}r\, \frac{\sin(rct)}{i|\mathbf{x}|c} \frac{\partial}{\partial\theta}\left(e^{-i|\mathbf{x}| r\cos\theta}\right) \\
&= 2\pi \int_0^R \mathrm{d}r\, \frac{\sin(rct)}{i|\mathbf{x}|c}\left(e^{i|\mathbf{x}|r} - e^{-i|\mathbf{x}|r}\right) \\
&= \frac{4\pi}{c|\mathbf{x}|} \int_0^R \mathrm{d}r \sin(rct)\sin(r|\mathbf{x}|) .
\end{aligned}$$

The integrand is even, so we can replace $\int_0^R$ with $\frac{1}{2}\int_{-R}^R$. Using

$$\sin A \sin B = \frac{1}{2}[\cos(A - B) - \cos(A + B)] = \frac{1}{2} \operatorname{Re}\left[e^{i(A-B)} - e^{i(A+B)}\right]$$

gives

$$\begin{aligned}
\Phi(\mathbf{x}, t) &= \lim_{R\to\infty} \frac{1}{(2\pi)^3} \int_{|\boldsymbol{\lambda}|} e^{i\boldsymbol{\lambda}\cdot\mathbf{x}} \tilde{\Phi}(\boldsymbol{\lambda}, t)\, \mathrm{d}^n\lambda \\
&= \lim_{R\to\infty} \frac{\pi}{(2\pi)^3 c|\mathbf{x}|} \int_{-R}^R \mathrm{d}r \left(e^{ir(|\mathbf{x}|-ct)} - e^{ir(|\mathbf{x}|+ct)}\right) \\
&= \frac{\pi}{(2\pi)^3 c|\mathbf{x}|}[2\pi\delta(|\mathbf{x}| - ct) - 2\pi\delta(|\mathbf{x}| + ct)] ,
\end{aligned}$$

and the result follows. □

In both cases, we can write

$$\Phi(\mathbf{x}, t) = \Phi^+(\mathbf{x}, t) - \Phi^-(\mathbf{x}, t) .$$

When $n = 3$,

$$\Phi^{\pm}(\mathbf{x}, t) = \frac{\delta(|\mathbf{x}| \mp ct)}{4\pi c|\mathbf{x}|} .$$

They are often referred to as the *advanced* and *retarded* parts of $\Phi(\mathbf{x}, t)$.

Note that when $t > 0$, $\Phi^-(\mathbf{x}, t) = 0$.

Proposition 14.17. The solution to $(\dagger_a)$ is given by

$$u(\mathbf{x}, t) = (\Phi * g)(\mathbf{x}, t) + \frac{\partial}{\partial t}(\Phi * f)(\mathbf{x}, t) .$$

Proof. Taking the Fourier transform of $(\dagger_a)$ gives

$$\begin{cases}
\dfrac{\partial^2 \tilde{u}}{\partial t^2} + c^2|\boldsymbol{\lambda}|^2 \tilde{u} = 0 \\
\tilde{u}(\boldsymbol{\lambda}, 0) = \tilde{f}(\boldsymbol{\lambda}) \\
\dfrac{\partial \tilde{u}}{\partial t} = \tilde{g}(\boldsymbol{\lambda}) .
\end{cases}$$

The differential equation has a general solution

$$\tilde{u}(\boldsymbol{\lambda}, t) = A(\boldsymbol{\lambda}) \sin(c|\boldsymbol{\lambda}|t) + B(\boldsymbol{\lambda}) \cos(c|\boldsymbol{\lambda}|t)\,,$$

and the initial condition give

$$B(\boldsymbol{\lambda}) = \tilde{f}(\boldsymbol{\lambda})\,,\ A(\boldsymbol{\lambda}) = \frac{1}{c|\boldsymbol{\lambda}|}\tilde{g}(\boldsymbol{\lambda})\,.$$

We have

$$\tilde{u}(\boldsymbol{\lambda}, t) = \tilde{\Phi}(\boldsymbol{\lambda}, t)\tilde{g}(\boldsymbol{\lambda}) + \tilde{f}(\boldsymbol{\lambda}) \cos(c|\boldsymbol{\lambda}|t) = \tilde{\Phi}(\boldsymbol{\lambda}, t)\tilde{g}(\boldsymbol{\lambda}) + \frac{\partial}{\partial t}\Big(\tilde{\Phi}(\boldsymbol{\lambda}, t)\tilde{f}(\boldsymbol{\lambda})\Big)\,.$$

The result follows from the convolution theorem. □

Let us focus on the term

$$(\Phi * g)(\mathbf{x}) = \int \Phi(\mathbf{y}, t) g(\mathbf{x} - \mathbf{y})\, \mathrm{d}^n \mathbf{y}$$

in the solution. The contribution from the advanced part, relevant for $t > 0$, in the case $n = 3$ is

$$(\Phi^+ * g)(\mathbf{x}) = \frac{1}{4\pi c} \int \frac{\delta(|y| - ct)}{\mathbf{y}} g(\mathbf{x} - \mathbf{y})\, \mathrm{d}^n \mathbf{y} = \frac{1}{4\pi c^2 t} g(\mathbf{x} - \mathbf{y})\, \mathrm{d}S\ .$$

The integral is equivalent to

$$\frac{t}{4\pi (ct)^2} \int_{|\mathbf{x}-\mathbf{y}|=ct} g(\mathbf{y})\, \mathrm{d}S = t \times \bar{g}\,,$$

where $\bar{g}$ is the average of g over the sphere of radius of ct centred at $\mathbf{x}$.

It turns out, in odd dimensions, this part of the solution always looks a bit like this. It always involves the spherical mean of g over a sphere of radius ct centred at $\mathbf{x}$. This can be interpreted as a spherical wavefront emanating from the point $\mathbf{x}$ and travelling at speed c.

In even dimensions, the solution at (x, t) involves integrals over the ball $\{\mathbf{y} : |\mathbf{x} - \mathbf{y}| \le ct\}$.

Remark. These observations allow us to conclude the following very different nature of waves in an even or odd dimensional universe. In odd dimensions, if it is pitch black outside and someone flashes a torch from $\mathbf{x}_0$ at time $t = 0$, you will see an instantaneous flash of light at time $t = \frac{|\mathbf{x}_0|}{c}$. However, in an even-dimensional world you would see a flash at $t = \frac{|\mathbf{x}_0|}{c}$ but it would not instantaneously disappear: instead, it would slowly fade away.

Proposition 14.18. The solution to $(\dagger_b)$ is

$$u(\mathbf{x}, t) = \int_0^t \left[\int \Phi(\mathbf{x} - \mathbf{y}, t - s) F(\mathbf{y}, s)\, \mathrm{d}^n \mathbf{y} \right] \mathrm{d}s\ .$$

Proof. Taking the Fourier transform of $(\dagger_b)$ gives

$$\begin{cases} \dfrac{\partial^2 \tilde{u}}{\partial t^2} + c^2 |\boldsymbol{\lambda}|^2 \tilde{u} = \tilde{F}(\boldsymbol{\lambda}, t) \\ \tilde{u}(\boldsymbol{\lambda}, 0) = 0 \\ \dfrac{\partial \tilde{u}}{\partial t} = 0\,. \end{cases}$$

So we want the Green's function $G = G(t; s)$ for the operator

$$L = \frac{\mathrm{d}^2}{\mathrm{d}t^2} + c^2 |\boldsymbol{\lambda}|^2\,.$$

This is

$$G(t; s) = \frac{\sin(c|\boldsymbol{\lambda}|(t - s))}{c|\boldsymbol{\lambda}|} = \tilde{\Phi}(\boldsymbol{\lambda}, t - s)\,.$$

Therefore,

$$\tilde{u}(\boldsymbol{\lambda}, t) = \int_0^t \tilde{\Phi}(\boldsymbol{\lambda}, t-s)\tilde{F}(\boldsymbol{\lambda}, s)\,\mathrm{d}s\ ,$$

and the result follows from the convolution theorem. □

Theorem 14.19. The solution to the wave equation in $\mathbb{R}^n$ of the form (†) is

$$u(\mathbf{x}, t) = (\Phi * g)(\mathbf{x}, t) + \frac{\partial}{\partial t}(\Phi * f)(\mathbf{x}, t) + \int_0^t \left[\int \Phi(\mathbf{x}-\mathbf{y}, t-s)F(\mathbf{y}, s)\,\mathrm{d}^n\mathbf{y}\right]\mathrm{d}s\ .$$

Proof. Principle of superposition. □

Like the heat equation, if we define $F(\mathbf{x}, t) = 0$ for $t < 0$ and introduce the function

$$G(\mathbf{x}, t; \mathbf{y}, s) = H(t-s)\Phi(\mathbf{x}-\mathbf{y}, t-s)\,,$$

then the solution to $(\dagger_b)$ can be written as

$$u(\mathbf{x}, t) = \iint G(\mathbf{x}, t; \mathbf{y}, s)F(\mathbf{y}, s)\,\mathrm{d}^n\mathbf{y}\,\mathrm{d}s\ .$$

We call G the Green's function for the Wave equation on $\mathbb{R}^n$ with vanishing initial data. It satisfies

$$\frac{\partial^2 G}{\partial t^2} - c^2\nabla^2 G = \delta(\mathbf{x}-\mathbf{y}, t-s)\,.$$

14.6 Green's Function for the Laplacian

Lemma 14.20. A 1D real Green's function is symmetric.

$$G(x; \xi) = G(\xi; x)\,.$$

Proof.

$$G(x; \xi) = \sum_{n=1}^{\infty} \frac{1}{\lambda_n} y_n(x) y_n^*(\xi)\,,$$

where λ_n and y_n are the eigenvalues and eigenfunctions of the differential operators. It is obvious that for real $y(x)$,

$$G(x; \xi) = G(\xi; x)\,.$$

□

Corollary. Real Green's function is symmetric.

$$G(\mathbf{x}; \mathbf{y}) = G(\mathbf{y}; \mathbf{x})\,.$$

To find the Green's function of the Laplacian on a bounded domain Ω with Dirichlet boundary conditions, we would naturally require

$$\begin{cases} \nabla^2 G(\mathbf{x}; \mathbf{y}) = \delta(\mathbf{x}-\mathbf{y}) & \mathbf{x} \in \Omega \\ G(\mathbf{x}; \mathbf{y}) = 0 & \mathbf{x} \in \partial\Omega\,. \end{cases}$$

However, if we are solving Poisson's equation with Neumann boundary conditions, then instead of requiring $\frac{\partial G}{\partial n} = 0$ on $\partial\Omega$, we can only require

$$\frac{\partial G}{\partial n} \equiv \mathbf{n}(r) \cdot \boldsymbol{\nabla} G = \frac{1}{A} \text{ on } \partial\Omega\,,$$

where $A = \oint_{\partial\Omega} \mathrm{d}S$ is the surface area. This is because a zero flux on the boundary is impossible.

Proof.

$$\oint_{\partial\Omega} \boldsymbol{\nabla} G \cdot \mathbf{n}\,\mathrm{d}S = \int_\Omega \nabla^2 G\,\mathrm{d}V$$
$$= \int_\Omega \delta(\mathbf{x}-\mathbf{y})\,\mathrm{d}V = 1\,.$$

□

Remark. However, when the domain Ω is the whole $\mathbb{R}^n$ space, we still require $\frac{\partial G}{\partial n} = 0$ since $A \to \infty$.

We are interested in the Green's function for the $\mathbb{R}^n$ space with Dirichlet boundary conditions, also known as the fundamental solutions, such that

$$\begin{cases} \nabla^2 G(\mathbf{x};\mathbf{y}) = \delta(\mathbf{x}-\mathbf{y}) \\ |G(\mathbf{x};\mathbf{y})| \to 0 \text{ as } |x| \to \infty\,. \end{cases}$$

Taking the Fourier transform of the equation gives

$$\tilde{G}(\boldsymbol{\lambda};\mathbf{y}) = -\frac{e^{-i\boldsymbol{\lambda}\cdot\mathbf{y}}}{|\boldsymbol{\lambda}|^2}\,.$$

We need the following result.

Proposition 14.21. For $\alpha > 0$ and $\mathbf{x} \in \mathbb{R}^n$, we have

$$\mathcal{F}_{\mathbf{x}\to\boldsymbol{\lambda}}[|\mathbf{x}|^{-\alpha}] = C_{n,\alpha}|\boldsymbol{\lambda}|^{\alpha-n}\,.$$

Proof. Let $f_\alpha(\mathbf{x}) = |\mathbf{x}|^{-\alpha}$, we have

$$\tilde{f}_\alpha(\boldsymbol{\lambda}) = \int |\mathbf{x}|^{-\alpha} e^{-i\boldsymbol{\lambda}\cdot\mathbf{x}}\,\mathrm{d}^n\mathbf{x}\,.$$

If we make a change of variables $\mathbf{x} = \mathsf{R}^{\mathrm{T}}\mathbf{x}'$, where R is a rotation matrix, then since $\det\{\mathsf{R}\} = 1$, after dropping the primes, we have

$$\tilde{f}_\alpha(\boldsymbol{\lambda}) = \int \left|\mathsf{R}^{\mathrm{T}}\mathbf{x}\right|^{-\alpha} e^{-i\boldsymbol{\lambda}\cdot(\mathsf{R}^{\mathrm{T}}\mathbf{x})}\,\mathrm{d}^n\mathbf{x}$$
$$= \int |\mathbf{x}|^{-\alpha} e^{-i(\mathsf{R}\boldsymbol{\lambda})\cdot\mathbf{x}}\,\mathrm{d}^n\mathbf{x} = \tilde{f}_\alpha(\mathsf{R}\boldsymbol{\lambda})\,.$$

Therefore, $\tilde{f}_\alpha$ is rotation invariant. We can just write $\boldsymbol{\lambda} = |\boldsymbol{\lambda}|\mathbf{u}$ for any unit vector $\mathbf{u}$, and the result will not depend on the direction of $\mathbf{u}$. Making the substitution $\mathbf{x}' = |\boldsymbol{\lambda}|\mathbf{x}$ and dropping primes we get

$$\tilde{f}_\alpha(\boldsymbol{\lambda}) = \int |\mathbf{x}|^{-\alpha} e^{-i|\boldsymbol{\lambda}|(\mathbf{m}\cdot\mathbf{x})}\,\mathrm{d}^n\mathbf{x}$$
$$= |\lambda|^{-n} \int |\boldsymbol{\lambda}|^\alpha |\mathbf{x}|^{-\alpha} e^{-i(\mathbf{m}\cdot\mathbf{x})}\,\mathrm{d}^n\mathbf{x} = C_{n,\alpha}|\boldsymbol{\lambda}|^{\alpha-n}\,,$$

where $C_{n,\alpha}$ is only dependent on n and α. □

If $n > 2$, set $\alpha = n-2$, there is a constant c_n such that

$$\frac{1}{|\boldsymbol{\lambda}|^2} = c_n \mathcal{F}_{\mathbf{x}\to\boldsymbol{\lambda}}\left[|\mathbf{x}|^{2-n}\right].$$

Theorem 14.22. The free space Green's function for the Laplacian on $\mathbb{R}^n$ with $n > 2$ is

$$G(\mathbf{x};\mathbf{y}) = -\frac{1}{(n-2)\|S^{n-1}\|} \times \frac{1}{|\mathbf{x}-\mathbf{y}|^{n-2}},$$

where $\left\|S^{n-1}\right\|$ is the surface area of an $n-1$ sphere, given by

$$\|S_{n-1}\| = \frac{2\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})}.$$

In particular, for $n = 3$,

$$G(\mathbf{x};\mathbf{y}) = -\frac{1}{4\pi}\frac{1}{|\mathbf{x}-\mathbf{y}|}.$$

Proof. Let $F(\mathbf{x}) = c_n|\mathbf{x}|^{2-n}$, then

$$\tilde{G}(\boldsymbol{\lambda};\mathbf{y}) = -e^{-i\boldsymbol{\lambda}\cdot\mathbf{y}}\tilde{F}(\lambda) = -\mathcal{F}_{\mathbf{x}\to\boldsymbol{\lambda}}[F(\mathbf{x}-\mathbf{y})].$$

Therefore,

$$G(\mathbf{x};\mathbf{y}) = -\frac{c_n}{|\mathbf{x}-\mathbf{y}|^{n-2}}.$$

To find c_n, note that

$$-\nabla^2\left(|\mathbf{x}|^{2-n}\right) = \frac{\delta(\mathbf{x})}{c_n},$$

so integrating over $|\mathbf{x}| \le 1$ and using the divergence theorem gives

$$\frac{1}{c_n} = -\int_{|\mathbf{x}|=1} \boldsymbol{\nabla}\left(|\mathbf{x}|^{2-n}\right)\cdot \mathrm{d}\mathbf{S} = (n-2)\int_{|\mathbf{x}|=1}\mathrm{d}S = (n-2)\times\left\|S^{n-1}\right\|.$$

□

We need to use a different method to obtain the Green's function in $\mathbb{R}^2$.

Proposition 14.23. For $\mathbb{R}^2$, the Green's function for the Laplacian is given by

$$G(\mathbf{x};\mathbf{y}) = \frac{1}{2\pi}\ln|\mathbf{x}-\mathbf{y}| + C.$$

Proof. By translation invariance of the Laplacian, it is enough to consider the case $\mathbf{y} = \mathbf{0}$. For $\mathbf{y} = \mathbf{0}$, we have $|\mathbf{x}| = r$, so

$$\begin{aligned}\nabla^2 G &= \frac{1}{2\pi}\nabla^2 \ln r\\ &= \frac{1}{2\pi}\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\ln r\right) = 0.\end{aligned}$$

We need to check that what happens at $\mathbf{x} = \mathbf{0}$. By the divergence theorem, for any $\epsilon > 0$,

$$\begin{aligned}\int_{|\mathbf{x}|<\epsilon}\nabla^2\ln r\,\mathrm{d}^2\mathbf{x} &= \int_{|\mathbf{x}|=\epsilon}\boldsymbol{\nabla}\ln r\cdot\mathrm{d}\mathbf{S}\\ &= \int_0^{2\pi}\frac{\mathbf{e}_r}{\epsilon}\cdot\mathbf{e}_r\epsilon\,\mathrm{d}\theta = 2\pi.\end{aligned}$$

We have $\nabla^2\ln|\mathbf{x}| = 0$ for all $|\mathbf{x}| \ne 0$ and

$$\int_{|\mathbf{x}|<\epsilon}\nabla^2\ln|\mathbf{x}|\,\mathrm{d}^2\mathbf{x} = 2\pi$$

for any $\epsilon > 0$. Therefore

$$\nabla^2\ln|\mathbf{x}| = 2\pi\delta(\mathbf{x})$$

and so

$$G(\mathbf{x};\mathbf{y}) = \frac{1}{2\pi}\ln|\mathbf{x}-\mathbf{y}| + C.$$

□

Remark. It is impossible for the 2D Green's function to vanish at $\mathbf{x} \to \infty$. We can control the finite $\mathbf{x}$ at which G vanishes by adjusting the arbitrary constant C.

14.7 The Method of Images

Now we are using the method of images to find the Green's function of laplacian on a domain Ω that is not the full $\mathbb{R}^n$ space.

14.7.1 3D Half Space

Suppose that we want to find the Green's function for a domain Ω with Dirichlet boundary conditions, where Ω is the half-space of $\mathbb{R}^3$ with $z > 0$.

The Green's function satisfies

$$\begin{cases} \nabla^2 G = \delta(\mathbf{x} - \mathbf{y}) & \text{for } \mathbf{x} \in \Omega \\ G = 0 & \text{on } z = 0 \\ G = 0 & \text{as } |\mathbf{x}| \to \infty\,,\ \mathbf{x} \in \Omega\,. \end{cases}$$

The uniqueness of the solution allows us to solve this using a trick: remove the boundary at $z = 0$, consider all of space and add a point source of opposite sign, an *image source*, at the image point $\mathbf{y}' = (x', y', -z')$.

The Green's function satisfies

$$\nabla^2 G = \delta(\mathbf{x} - \mathbf{y}) - \delta(\mathbf{x} - \mathbf{y}')\,,$$

which, by superposition of the fundamental solutions, gives

$$G(\mathbf{x}; \mathbf{y}) = -\frac{1}{4\pi|\mathbf{x} - \mathbf{y}|} + \frac{1}{4\pi|\mathbf{x} - \mathbf{y}'|}\,.$$

We can check that this solution satisfies all the conditions, and therefore, by the uniqueness of the solution, Green's function is

$$G(\mathbf{x}; \mathbf{y}) = -\frac{1}{4\pi}\left(\frac{1}{|\mathbf{x} - \mathbf{y}|} - \frac{1}{|\mathbf{x} - \mathbf{y}'|}\right).$$

If instead we impose Neumann boundary conditions at $z = 0$:

$$\frac{\partial G}{\partial n} = \frac{\partial G}{\partial z} = 0 \text{ at } z = 0\,,$$

and still require $G \to 0$ as $|\mathbf{x}| \to \infty$, $\mathbf{x} \in \Omega$, then we need a point charge of the same sign at the image point. The Green's function is

$$G(\mathbf{x}; \mathbf{y}) = -\frac{1}{4\pi}\left(\frac{1}{|\mathbf{x} - \mathbf{y}|} + \frac{1}{|\mathbf{x} - \mathbf{y}'|}\right).$$

14.7.2 2D Quarter Plane

Suppose now we want to find Green's function for a domain Ω with Dirichlet boundary conditions, where Ω is the quarter plane of $\mathbb{R}^2$ with $x > 0$, $y > 0$.

The Green's function now satisfies

$$\begin{cases} \nabla^2 G = \delta(\mathbf{x}-\mathbf{y}_1) & \text{for } \mathbf{x}\in\Omega \\ G = 0 & \text{on } x = 0 \\ G = 0 & \text{on } y = 0 \\ G = 0 & \text{as } |\mathbf{x}|\to\infty\,,\ \mathbf{x}\in\Omega\,. \end{cases}$$

We then need three image charges: two of the opposite sign at $\mathbf{y}_2 = (-x_0, y_0)$, $\mathbf{y}_3 = (x_0, -y_0)$ and one of the same sign at $\mathbf{y}_4 = (-x_0, -y_0)$. By superposition of the fundamental solutions, we have

$$G(\mathbf{x};\mathbf{y}_0) = \frac{1}{2\pi}\ln\frac{|\mathbf{x}-\mathbf{y}_1||\mathbf{x}-\mathbf{y}_4|}{|\mathbf{x}-\mathbf{y}_2||\mathbf{x}-\mathbf{y}_3|}\,.$$

14.7.3 Sphere

Suppose now we want to find the Green's function for a domain $\Omega : |\mathbf{x}| < a$ in $\mathbb{R}^3$, under Dirichlet boundary conditions.

The Green's function satisfies

$$\begin{cases} \nabla^2 G = \delta(\mathbf{x}-\mathbf{y}) & \text{for } \mathbf{x}\in\Omega \\ G = 0 & \text{for } |\mathbf{x}| = a\,. \end{cases}$$

The image source would be of the strength $-\frac{a}{|\mathbf{y}|}$ at the *inverse point* of the source

$$\mathbf{y}' = \frac{a^2}{|\mathbf{y}|^2}\mathbf{y}\,.$$

The Green's function is therefore

$$G(\mathbf{x};\mathbf{y}) = -\frac{1}{4\pi}\left(\frac{1}{|\mathbf{x}-\mathbf{y}|} - \frac{a}{|\mathbf{y}||\mathbf{x}-\mathbf{y}'|}\right).$$

Proof. at $|\mathbf{x}| = a$,

$$\begin{aligned} |\mathbf{x}-\mathbf{y}'| &= \sqrt{|\mathbf{x}|^2 + |\mathbf{y}'|^2 - 2\mathbf{x}\cdot\mathbf{y}'} \\ &= \sqrt{a^2 + \frac{a^4}{|\mathbf{y}|^2} - 2\frac{a^2}{|\mathbf{y}|^2}\mathbf{x}\cdot\mathbf{y}} \\ &= \frac{a}{|\mathbf{y}|}\sqrt{|\mathbf{y}|^2 + a^2 - 2\mathbf{x}\cdot\mathbf{y}} \\ &= \frac{a}{|\mathbf{y}|}|\mathbf{x}-\mathbf{y}|\,. \end{aligned}$$

Therefore, $G(\mathbf{x};\mathbf{y}) = 0$ at the boundary $|\mathbf{x}| = a$. □

14.7.4 Circle

Suppose now we want to find the Green's function for a domain $\Omega : |\mathbf{x}| < a$ in $\mathbb{R}^2$, under Dirichlet boundary conditions.

The image point is the inverse point again, with $\mathbf{y}' = \frac{a^2}{|\mathbf{y}|^2}\mathbf{y}$, but now the image just needs to have the strength -1 as we are able to adjust the constant C. The Green's function is

$$G(\mathbf{x};\mathbf{y}) = \frac{1}{2\pi}\ln\frac{|\mathbf{x}-\mathbf{y}|}{|\mathbf{x}-\mathbf{y}'|} + C\,,$$

where the constant C is chosen to ensure that $G = 0$ on the circle $|\mathbf{x}| = a$.

14.8 The Integral Solution of Poisson's Equation

To find the solution of Poisson's equation with an arbitrary source distribution, we will need Green's identity.

Theorem 14.24 (Green's first identity). Suppose $\Omega \subset \mathbb{R}^n$ is a compact set with boundary $\partial\Omega$, and let $\phi, \psi : \Omega \to \mathbb{R}$ be a pair of functions on Ω that are regular throughout Ω.

$$\int_\Omega \phi\nabla^2\psi + (\boldsymbol{\nabla}\phi)\cdot(\boldsymbol{\nabla}\psi)\,\mathrm{d}V = \int_{\partial\Omega} \phi\boldsymbol{\nabla}\psi\cdot\mathbf{n}\,\mathrm{d}S\ ,$$

where $\mathbf{n}$ is the outward pointing normal to $\partial\Omega$.

Proof. Apply product rule and divergence theorem.

$$\begin{aligned}\int_\Omega \boldsymbol{\nabla}\cdot(\phi\boldsymbol{\nabla}\psi)\,\mathrm{d}V &= \int_\Omega \phi\nabla^2\psi + (\boldsymbol{\nabla}\phi)\cdot(\boldsymbol{\nabla}\psi)\,\mathrm{d}V\\ &= \int_{\partial\Omega}\phi\boldsymbol{\nabla}\psi\cdot\mathbf{n}\,\mathrm{d}S\ .\end{aligned}$$

□

Theorem 14.25 (Green's second identity). Suppose $\Omega \subset \mathbb{R}^n$ is a compact set with boundary $\partial\Omega$, and let $\phi, \psi : \Omega \to \mathbb{R}$ be a pair of functions on Ω that are regular throughout Ω.

$$\int_\Omega \phi\nabla^2\psi - \psi\nabla^2\phi\,\mathrm{d}V = \oint_{\partial\Omega}(\phi\boldsymbol{\nabla}\psi - \psi\boldsymbol{\nabla}\phi)\cdot\mathbf{n}\,\mathrm{d}S =: \oint_{\partial\Omega}\phi\frac{\partial\psi}{\partial n} - \psi\frac{\partial\phi}{\partial n}\,\mathrm{d}S\ .$$

Proof. Interchange ϕ and ψ in the Green's first identity to obtain

$$\int_\Omega \psi\nabla^2\phi + (\boldsymbol{\nabla}\psi)\cdot(\boldsymbol{\nabla}\phi)\,\mathrm{d}V = \int_{\partial\Omega}\psi\boldsymbol{\nabla}\phi\cdot\mathbf{n}\,\mathrm{d}S\ ,$$

then subtract from the Green's first identity. □

Remark. We will apply the Green's second identity to the fundamental solution. However, $G(\mathbf{x},\mathbf{y})$ is singular at $\mathbf{y}$ so it is not clear whether the identity is valid, because the divergence theorem usually requires functions to be regular throughout Ω.

Consider Poisson's equation in a domain Ω with inhomogeneous Dirichlet boundary conditions:

$$\begin{cases}\nabla^2\phi = \rho(\mathbf{x}) & \mathbf{x}\in\Omega\\ \phi(\mathbf{x}) = f(\mathbf{x}) & \mathbf{x}\in\partial\Omega\,.\end{cases}\tag{†}$$

We will first directly apply Green's identity with $\psi = G$ in a less rigorous way, which gives

$$\begin{aligned}&\int_\Omega(\phi\nabla^2 G - G\nabla^2\phi)\,\mathrm{d}V = \oint_{\partial\Omega}(\phi\boldsymbol{\nabla}G - G\boldsymbol{\nabla}\phi)\cdot\mathbf{n}\,\mathrm{d}S\\ \implies &\int_\Omega(\phi\delta(\mathbf{x}-\mathbf{y}) - G\rho(\mathbf{x}))\,\mathrm{d}V = \oint_{\partial\Omega}(f(\mathbf{x})\boldsymbol{\nabla}G - 0\times\boldsymbol{\nabla}\phi)\cdot\mathbf{n}\,\mathrm{d}S\\ \implies &\int_\Omega\phi(\mathbf{x})\delta(\mathbf{x}-\mathbf{y})\,\mathrm{d}V = \int_\Omega\rho(\mathbf{x})G(\mathbf{x};\mathbf{y})\,\mathrm{d}V + \oint_{\partial\Omega}f(\mathbf{x})(\boldsymbol{\nabla}G\cdot\mathbf{n})\,\mathrm{d}S\ .\end{aligned}$$

Theorem 14.26. The integral solution of the Poisson's equation in domain Ω with Dirichlet boundary condition

$$\begin{cases}\nabla^2\phi = \rho(\mathbf{x}) & \mathbf{x}\in\Omega\\ \phi(\mathbf{x}) = f(\mathbf{x}) & \mathbf{x}\in\partial\Omega\end{cases}\tag{†}$$

is given by

$$\phi(\mathbf{y}) = \int_\Omega\rho(\mathbf{x})G(\mathbf{x};\mathbf{y})\,\mathrm{d}V + \oint_{\partial\Omega}f(\mathbf{x})\frac{\partial G(\mathbf{x};\mathbf{y})}{\partial n}\,\mathrm{d}S\ .$$

Proof (Non-examinable). Here is the formal proof. To remove the singularity issues of the Green's function, consider a ball B_ϵ of radius $\epsilon \ll 1$ centred at the singular point $\mathbf{x} = \mathbf{y}$. Let Ω' be the region

$$\Omega' = \Omega - B_\epsilon \, .$$

Now the Green's function is perfectly regular everywhere in Ω', so we can apply Green's second identity with $\psi = G$ to get

$$\begin{aligned}\int_{\Omega'} \phi\nabla^2 G - G\nabla^2\phi \,\mathrm{d}V &= -\int_{\Omega'} G\rho(\mathbf{x})\,\mathrm{d}V \\ &= \int_{\partial\Omega'} \phi\boldsymbol{\nabla}G\cdot\mathbf{n} - G\boldsymbol{\nabla}\phi\cdot\mathbf{n}\,\mathrm{d}S \\ &= \int_{\partial\Omega} \phi\boldsymbol{\nabla}G\cdot\mathbf{n}\,\mathrm{d}S + \int_{S_\epsilon^{n-1}} \phi\boldsymbol{\nabla}G\cdot\mathbf{n} - G\boldsymbol{\nabla}\phi\cdot\mathbf{n}\,\mathrm{d}S \, ,\end{aligned}$$

where the first equality follows since $\nabla^2 G = 0$ in Ω'. Note that on the inner boundary, a sphere of radius ϵ, the outward-pointing unit normal is $\mathbf{n} = -\hat{\mathbf{r}}$. As $\epsilon \to 0$, the Green's function approaches the fundamental solution so

$$\begin{aligned} G|_{S_\epsilon^{n-1}} &= -\frac{1}{(n-2)\left\|S_\epsilon^{n-1}\right\|}\frac{1}{\epsilon^{n-2}} \, , \\ \mathbf{n}\cdot\boldsymbol{\nabla}G|_{S_\epsilon^{n-1}} &= -\frac{1}{\|S^{n-1}\|}\frac{1}{\epsilon^{n-1}} \, .\end{aligned}$$

The measure on an $(n-1)$-sphere of radius ϵ is $\mathrm{d}S = \epsilon^{n-1}\,\mathrm{d}\Omega_n$, where $\mathrm{d}\Omega_n$ is an integral over angles. Therefore, we have

$$-\int_{S_\epsilon^{n-1}} G\boldsymbol{\nabla}\phi\cdot\mathbf{n}\,\mathrm{d}S = \frac{\epsilon}{\left\|S_\epsilon^{n-2}\right\|}\int \mathbf{n}\cdot\boldsymbol{\nabla}\phi\,\mathrm{d}\Omega_n \, .$$

Since ϕ is regular at $\mathbf{y}$, the value of this integral is bounded, so this term vanishes as $\epsilon \to 0$. Also,

$$\int_{S_\epsilon^{n-1}} \phi\boldsymbol{\nabla}G\cdot\mathbf{n}\,\mathrm{d}S = -\frac{1}{\left\|S_\epsilon^{n-1}\right\|}\int \phi\,\mathrm{d}\Omega_n = -\bar{\phi} \, ,$$

where $\bar{\phi}$ is the average value of ϕ on the sphere surrounding $\mathbf{y}$. As $\epsilon \to 0$, this $\bar{\phi} \to \phi(\mathbf{y})$. Putting all these together we find that, as $\epsilon \to 0$ so that $\Omega' = \Omega$,

$$\phi(\mathbf{y}) = \int_\Omega G\rho(\mathbf{x})\,\mathrm{d}V + \int_{\partial\Omega} \phi\boldsymbol{\nabla}G\cdot\mathbf{n}\,\mathrm{d}S \, .$$

This is the result as claimed. □

If we want Ω to be all space, we can use the fundamental solution for G but we need to ensure that the surface integral $\to 0$.

Corollary. The solution to the Poisson's equation on all space:

$$\begin{cases} \nabla^2\phi = \rho(\mathbf{x}) & \mathbf{x} \in \mathbb{R}^n \\ \phi(\mathbf{x}) = 0 & |\mathbf{x}| \to \infty \end{cases}$$

is given by

$$\phi(\mathbf{y}) = \int_\Omega \rho(\mathbf{x})G(\mathbf{x};\mathbf{y})\,\mathrm{d}V \, .$$

Corollary. The solution to the Laplace's equation

$$\begin{cases} \nabla^2\phi = 0 & \mathbf{x} \in \Omega \\ \phi(\mathbf{x}) = f(\mathbf{x}) & \mathbf{x} \in \partial\Omega \end{cases}$$

is given by

$$\phi(\mathbf{y}) = \oint_{\partial\Omega} f(\mathbf{x})\frac{\partial G(\mathbf{x};\mathbf{y})}{\partial n}\,\mathrm{d}S \, .$$

An integral solution of Poisson's equation with Neumann boundary conditions

$$\begin{cases} \nabla^2\phi = \rho(\mathbf{x}) & \mathbf{x} \in \Omega \\ \dfrac{\partial\phi}{\partial n} = f(\mathbf{x}) & \mathbf{x} \in \partial\Omega \end{cases}$$

can also be derived. From Green's identity

$$\int_\Omega (\phi\nabla^2 G - G\nabla^2\phi)\,\mathrm{d}V = \oint_{\partial\Omega} (\phi\boldsymbol{\nabla}G - G\boldsymbol{\nabla}\phi)\cdot\mathbf{n}\,\mathrm{d}S\ ,$$

we have

$$\phi(\mathbf{y}) = \int_\Omega \rho(\mathbf{x})G(\mathbf{x},\mathbf{y})\,\mathrm{d}V + \frac{1}{A}\oint_{\partial\Omega}\phi(\mathbf{x})\,\mathrm{d}S - \oint_{\partial\Omega} f(\mathbf{x})G(\mathbf{x};\mathbf{y})\,\mathrm{d}S\ .$$

Theorem 14.27. The integral solution of the Poisson's equation over full $\Omega = \mathbb{R}^n$ domain with Neumann boundary condition

$$\begin{cases} \nabla^2\phi = \rho(\mathbf{x}) & \mathbf{x} \in \Omega \\ \dfrac{\partial\phi}{\partial n} = f(\mathbf{x}) & \mathbf{x} \in \partial\Omega \end{cases} \tag{$\dagger$}$$

is given by

$$\phi(\mathbf{y}) = \int_\Omega \rho(\mathbf{x})G(\mathbf{x};\mathbf{y})\,\mathrm{d}V - \oint_{\partial\Omega} f(\mathbf{x})G(\mathbf{x};\mathbf{y})\,\mathrm{d}S\ .$$

as long as the surface integral over ϕ is finite.

Remark. The solution also works with finite Ω and $\partial\Omega$ although it contains an integral of the unknown Φ because this is essentially a constant term. Since we are dealing with a problem with Neumann boundary conditions, the solution is determined up to an unknown constant and such an integral term will not affect our result. In a problem with the Neumann boundary condition, we only care about $\boldsymbol{\nabla}\phi$.

Example. Solve the electric potential ϕ of a charged wire of length $2L$ with charged density μ per unit length, lying along the z axis from $z = -L$ to $z = L$.

We have the Poisson's equation

$$\nabla^2\phi = \rho(\mathbf{x})$$

with

$$\rho(\mathbf{x}) = \begin{cases} \frac{\mu}{\epsilon_0}\delta(\mathbf{x}) & \text{for } -L \le z \le L \\ 0 & \text{otherwise}\,. \end{cases}$$

The Green's function is the fundamental solution, so the integral solution is

$$\begin{aligned} \phi(\mathbf{y}) &= \int_{\mathbb{R}^3} \frac{\rho(\mathbf{x})}{4\pi\epsilon_0|\mathbf{x}-\mathbf{y}|}\,\mathrm{d}^3\mathbf{x} \\ &= \frac{\mu}{4\pi\epsilon_0}\int_{-L}^{L}\mathrm{d}z \iint \frac{\delta(\mathbf{x})}{|\mathbf{x}-\mathbf{y}|}\,\mathrm{d}x\,\mathrm{d}y \\ &= \frac{\mu}{4\pi\epsilon_0}\int_{-L}^{L}\frac{\mathrm{d}z}{\sqrt{x'^2+y'^2+(z-z')^2}}\,, \end{aligned}$$

which, after a substitution of $z - z' = \sqrt{x'^2+y'^2}\sinh u$ and relabelling, gives

$$\phi = \frac{\mu}{4\pi\epsilon_0}\left[\sinh^{-1}\frac{L-z}{\sqrt{x^2+y^2}} + \sinh^{-1}\frac{L+z}{\sqrt{x^2+y^2}}\right].$$

Example. Solution of Laplace's equation in 3D half space.

Find the solution of Laplace's equation in the 3D half-space with $z > 0$ subject to the Dirichlet boundary condition $\phi = f(x, y)$ on $z = 0$.

Ω it the half-space, and the bounding surface $\partial\Omega$ is the $z = 0$ plus the hemisphere at ∞. We can neglect the hemisphere at ∞ since $\phi \to 0$ there. Hence, the integral solution is given by

$$\phi(\mathbf{y}) = \oint_{\partial\Omega} f(\mathbf{x}) \frac{\partial G}{\partial n} \, \mathrm{d}S \ .$$

The Green's function, derived previously using the method of image, is

$$G(\mathbf{x}, \mathbf{y}) = -\frac{1}{4\pi}\left(\frac{1}{|\mathbf{x} - \mathbf{y}|} - \frac{1}{|\mathbf{x} - \mathbf{z}|}\right).$$

We have

$$\begin{aligned}\left.\frac{\partial G}{\partial z}\right|_{z=0} &= -\frac{1}{4\pi}\frac{\partial}{\partial z}\left\{\frac{1}{\sqrt{(x-x')^2 + (y-y')^2 + (z-z')^2}}\right. \\ &\qquad \left. - \frac{1}{\sqrt{(x-x')^2 + (y-y')^2 + (z+z')^2}}\right\}\Bigg|_{z=0} \\ &= -\frac{z'}{2\pi[(x-x')^2 + (y-y')^2 + z'^2]^{\frac{3}{2}}} \, \mathrm{d}x \, \mathrm{d}y \ ,\end{aligned}$$

and so

$$\phi(\mathbf{y}) = \frac{z'}{2\pi} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{f(x, y)}{[(x-x')^2 + (y-y')^2 + z'^2]^{\frac{3}{2}}} \, \mathrm{d}x \, \mathrm{d}y \ .$$

15 Group Theory

15.1 Mappings

These alternative terminologies are common when talking about groups.

Definition 15.1. A map $f : X \to Y$ is *one-to-one* (injective) if $f(x_1) = f(x_2)$ implies $x_1 = x_2$. f maps distinct elements to distinct elements.

Definition 15.2. A map $f : X \to Y$ is *onto* (surjective) if for every $y \in Y$, there is at least one $x \in X$ such that $f(x) = y$.

15.2 Groups

Definition 15.3. A *group* is a triple $(G, \cdot, I)$ of a set G, a binary operation $\cdot$ on G, and an element $I \in G$ such that the following axioms are satisfied:

(G0) *Closure.* $\forall g_1, g_2 \in G$, $g_2 \cdot g_1 \in G$;

(G1) *Associativity.* $\forall g_1, g_2, g_3 \in G$, $g_3 \cdot (g_2 \cdot g_1) = (g_3 \cdot g_2) \cdot g_1$;

(G2) *Identity.* $\exists I \in G$ such that $\forall g \in G$, $g \cdot I = I \cdot g = g$;

(G3) *Inverse.* $\forall g \in G$, $\exists g^{-1} \in G$ such that $g \cdot g^{-1} = g^{-1} \cdot g = I$.

Remarks.

- Sometimes axiom G0 is not stated since it is implied in the definition of binary operations.
- It is common to abbreviate the group $(G, \cdot, I)$ as G, and group product $g_2 \cdot g_1$ as $g_2 g_1$.
- A set of elements with two binary operations can form more complicated algebraic structures, including *rings* and *fields*.

Definition 15.4. A set R together with two binary operations and two identities, one for each operation, $(R, +, e, \cdot, u)$, is a *ring* if elements in R satisfy the ring axioms:

(R0) *Closure.* $\forall a, b \in R$, $a + b$, $a \cdot b \in R$;

(R1) *Associativity.* $\forall a, b, c \in R$, $a + (b + c) = (a + b) + c$, $a \cdot (b \cdot c) = (a \cdot b) \cdot c$;

(R2) *Additive commutativity.* $\forall a, b \in R$, $a + b = b + a$;

(R3) *Distributivity.* $\forall a, b, c \in R$, $a \cdot (b + c) = a \cdot b + a \cdot c$, $(a + b) \cdot c = a \cdot c + b \cdot c$;

(R4) *Additive identity.* $\exists e \in R$ such that $\forall a \in R$, $e + a = a$;

(R5) *Multiplicative identity.* $\exists u \in R$, $u \neq e$ such that $\forall a \in R$, $u \cdot a = a \cdot u = a$;

(R6) *Additive inverse.* $\forall a \in R$, $\exists b \in R$ such that $a + b = e$.

Definition 15.5. A ring $\mathbb{F}$ is a *field* if it is commutative in multiplication and any non-zero element has a multiplicative inverse:

(F7) *Multiplicative commutativity.* $\forall a, b \in \mathbb{F}$, $a \cdot b = b \cdot a$.

(F8) *Multiplicative inverse.* $\forall a \in \mathbb{F}$, $a \neq e$, $\exists b \in \mathbb{F}$ such that $a \cdot b = u$.

Remark. Common fields include $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ under addition and multiplication.

Proposition 15.6. The identity of a group is unique.

Proof. Assume group G has two distinct identities I, I', then

$$II' = I = I' \,,$$

which contradicts our assumption. □

Proposition 15.7. A group element's inverse is unique.

Proof. Assume some $g \in G$ has two distinct inverses h, k, then

$$\begin{aligned} & gh = I \text{ and } kg = I \\ \implies \quad & (kg)h = k(gh) = kI = k \\ \implies \quad & Ih = k \text{ so } h = k \,, \end{aligned}$$

which contradicts our assumption. □

Proposition 15.8. The inverse of a product is given by

$$(g_2 g_1)^{-1} = g_1^{-1} g_2^{-1} \,.$$

Proof.

$$\begin{aligned} g_1^{-1} g_2^{-1} g_2 g_1 &= g_1^{-1} I g_1 \\ &= g_1^{-1} g_1 \\ &= I \,. \end{aligned}$$

□

15.2.1 Commutativity and Order

Definition 15.9. A group $(G, \cdot, I)$ is *Abelian* if for all $g_1, g_2 \in G$, we have $g_1 g_2 = g_2 g_1$. Otherwise, the group is *non-Abelian*.

Definition 15.10. A group is *finite* if it contains a finite number of elements. The *order* of a finite group is the number of elements it contains, denoted as $|G|$.

An *infinite group* has infinitely many elements.

Definition 15.11. The *order* of a group element $g \in G$ is the least integer q such that $g^q = I$. For any finite group,

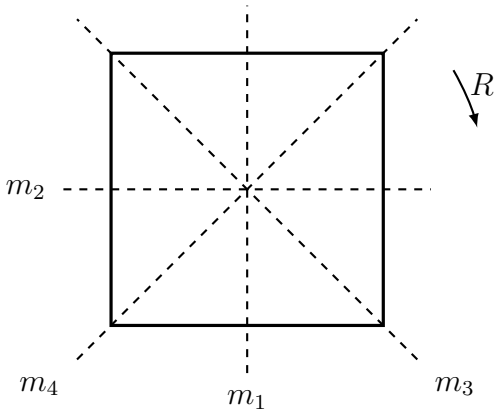
$$q \leq |G| \,.$$

15.3 Symmetry of the Square

Groups arise naturally in the study of symmetries. Let us first consider an n-gon lying on the complex plane $\mathbb{C}$.

Definition 15.12. An *isometry* is a rigid motion or transformation which preserves distances between points.

Let set D_n consist of the isometries of $\mathbb{C}$ that send the n-gon to itself. If $f, g : \mathbb{C} \to \mathbb{C}$ are isometries that send the n-gon to itself, then so is the composition $f \circ g$, so $\circ$ defines a binary operation on D_n. Let $I \in D_n$ be the isometry $I(x) = x$.



Theorem 15.13. $(D_n, \circ, I)$ is a group, called the n^{th} dihedral group. It has order $2n$.

Remark. Most pure mathematicians will confusingly denote this group as D_{2n}, based on its order.

Let us consider the group of transformations which represent the symmetries of the square. Let rotations of 0°, 90°, 180°, 270° anticlockwise be I, R, R^2, R^3 respectively. Let reflections under four axes of symmetry be m_1, m_2, m_3, m_4 respectively.

This is then the *4-fold dihedral group*, D_4. This is an example of a *point group*, for which the operations leave one point fixed.

Each reflection is its own inverse:

$$m_1^2 = m_2^2 = m_3^2 = m_4^2 = I\,,$$

and for rotations,

$$RR^3 = R^2R^2 = R^4 = I\,.$$

Definition 15.14. There is a minimal subset from which all other group elements can be obtained by composition. We say the group is *generated* by such a subset, and the members of the subset are called the *generators* of the group.

Example. For instance, the subset $\{R, m_1\}$ generates the group as follows:

$$\{I,\, R,\, R^2,\, R^3,\, m_1,\, m_2,\, m_3,\, m_4\} = \{R^4,\, R,\, R^2,\, R^3,\, m_1,\, R^2m_1,\, R^3m_1,\, Rm_1\}\,.$$

15.3.1 Group Table

We can construct a table for all g_2g_1, where g_1 are on the top row and g_2 are on the leftmost column. For the D_4 group, the group table is:

I	R^2	R	R^3	m_1	m_2	m_3	m_4
R^2	I	R^3	R	m_2	m_1	m_4	m_3
R	R^3	R^2	I	m_4	m_3	m_1	m_2
R^3	R	I	R^2	m_3	m_4	m_2	m_1
m_1	m_2	m_3	m_4	I	R^2	R	R^3
m_2	m_1	m_4	m_3	R^2	I	R^3	R
m_3	m_4	m_2	m_1	R^3	R	I	R^2
m_4	m_3	m_1	m_2	R	R^3	R^2	I

By observing the group table, we can notice that each row and column is a complete rearrangement of the group elements. This is the rearrangement theorem.

Theorem 15.15 (Rearrangement theorem). Let $G = \{g_1, \ldots, g_n\}$ be a group, then $\forall g \in G$,

$$gG = \{gg_1, \ldots, gg_n\}$$

contains each element of the group once and only once.

Proof. $\forall g_\alpha \in G$, $\exists g_\beta \in G$ such that $gg_\beta = g_\alpha$ (just let $g_\beta = g^{-1}g_\alpha$), so gG contains all elements in G.

Also, the set gG has no repeated element, since if $gg_\beta = gg'_\beta$, then

$$g^{-1}gg_\beta = g^{-1}gg'_\beta \implies g_\beta = g'_\beta \,.$$

□

Corollary. We also have that, $\forall g \in G$,

$$Gg = \{g_i g \mid g_i \in G\} = G \,.$$

15.3.2 Subgroups

Definition 15.16. Let $(G, \cdot, I)$ be a group and $H \subseteq G$, then $(H, \cdot, I)$ is a *subgroup* of G, denoted $H \leq G$ if

(i) $I \in H$;

(ii) for all $h_1, h_2 \in H$, $h_1 \cdot h_2 \in H$;

(iii) $(H, \cdot, I)$ is a group.

A subgroup $H \leq G$ is *proper* if $H \neq \{I\}$ and $H \neq G$.

Remark. To exclude the case that $H = G$, we use $H < G$.

Example. Proper subgroups of D_4.

By examining the group table of D_4, we can spot 5 order-2 subgroups:

$$\{I,\, R^2\},\, \{I,\, m_1\},\, \{I,\, m_2\},\, \{I,\, m_3\},\, \{I,\, m_4\}\,,$$

and an order-4 cyclic subgroup, C_4:

$$\{I,\, R,\, R^2,\, R^3\}\,.$$

In addition, there are two other order-4 subgroups:

$$\{I,\, R^2,\, m_1,\, m_2\} \text{ and } \{I,\, R^2,\, m_3,\, m_4\}\,,$$

which are called *Klein four-group* or *Vierergruppe*, denoted by K_4 or V_4.

15.3.3 Cyclic Groups

Definition 15.17. A *cyclic group* is a group G that can be generated by a single group element g. A finite cyclic group of order n is denoted as C_n.

A finite cyclic group of order n contains group members

$$\{I,\, g,\, g^2, \ldots,\, g^{n-1}\}\,,$$

and it follows that $I = g^n$.

15.4 Homomorphism

Definition 15.18. If $(H, \cdot_H, I_H)$ and $(G, \cdot_G, I_G)$ are groups then a function $\phi : H \to G$ is called a *homomorphism* if for all $a, b \in H$, we have

$$\phi(a \cdot_H b) = \phi(a) \cdot_G \phi(b)\,.$$

Remark. A homomorphism is a map between groups that preserves group operations but is not necessarily 1-1 or onto.

Definition 15.19. If $\phi : H \to G$ is a homomorphism, then

- the *image* of ϕ is
$$\mathrm{Im}(\phi) \coloneqq \{g \in G \mid g = \phi(h) \text{ for some } h \in H\}\,.$$
- the *kernel* of ϕ is
$$\ker(\phi) \coloneqq \{h \in H \mid \phi(h) = I_G\}\,.$$

Lemma 15.20. If $\phi : H \to G$ is a homomorphism, then

(i) $\phi(I_H) = I_G$,

(ii) for all $h \in H$, we have $\phi(h^{-1}) = \phi(h)^{-1}$.

Proof.

(i)
$$\begin{aligned} \phi(I_H) \cdot_G \phi(I_H) &= \phi(I_H \cdot_H I_H) = \phi(I_H) \\ \implies \quad \phi(I_H) &= I_G\,. \end{aligned}$$

(ii)
$$\begin{aligned} \phi(h) \cdot_G \phi(h^{-1}) &= \phi(h \cdot_H h^{-1}) = \phi(I_H) = I_G \\ \phi(h^{-1}) \cdot_G \phi(h) &= \phi(h^{-1} \cdot_H h) = \phi(I_H) = I_G \end{aligned}$$

These are the defining properties of $\phi(h)^{-1}$, so $\phi(h^{-1}) = \phi(h)^{-1}$. □

Example. Let $G = (\mathbb{R}, +, 0)$ and $G' = U(1)$, the multiplicative group of unit-magnitude complex numbers. Define a mapping $\phi : \mathbb{R} \to U(1)$ such that for $x \in \mathbb{R}$,

$$\phi(x) = e^{ix}\,.$$

This is a homomorphism because

$$\phi(x + y) = e^{i(x+y)} = e^{ix}e^{iy} = \phi(x)\phi(y)\,.$$

The kernel of ϕ is $\ker(\phi) = \{2N\pi \mid N \in \mathbb{Z}\} = \{\dots, -2\pi, 0, 2\pi, \dots\}$

Proposition 15.21. If $\phi : H \to G$ is a homomorphism, then $\mathrm{Im}(\phi) \leq G$ and $\ker(\phi) \leq H$.

Proof. We have $\phi(I_H) = I_G \in \mathrm{Im}(\phi)$. $\mathrm{Im}(\phi)$ is closed because for any $g_1 = \phi(h_1), g_2 = \phi(h_2) \in \mathrm{Im}(\phi)$, $g_1 g_2 = \phi(h_1 h_2) \in \mathrm{Im}(\phi)$. Also, the inverse exists for all elements in $\mathrm{Im}(\phi)$ by Lemma 15.20.

We have $I_H \in \ker(\phi)$. If h_1, $h_2 \in \ker(\phi)$, then $h_1 h_2 \in \ker(\phi)$ since it must be mapped to I_G. Also, if $h \in \ker(\phi)$, then $h^{-1} \in \ker(\phi)$ since $\phi(h^{-1}) = (\phi(h))^{-1} = I_G^{-1} = I_G$, so $\ker(\phi) \leq H$. □

Definition 15.22. If a homomorphism $\phi : H \to G$ is invertible then it is an *isomorphism*, written $G \cong H$.

Remark. Two groups are usually considered identical if there is an isomorphism between them.

The following theorem is useful for checking whether a homomorphism is an isomorphism.

Theorem 15.23. If $\phi : H \to G$ is a homomorphism, then it is an isomorphism if and only if $\mathrm{Im}(\phi) = G$ and $\ker(\phi) = I_H$.

15.5 Group Actions (Non-examinable)

We motivated our definition of a group G as the symmetries of an object X, but when defining a group we abstracted the properties that symmetries of an object have and there was no longer an X.

It is now the time to take the object that our groups acting on back.

Definition 15.24. An action of a group $(G, \cdot, I)$ on a set X is a function $* : G \times X \to X$ satisfying

(A1) *Identity.* For all $x \in X$ we have $I * x = x$.

(A2) *Associativity.* For all $a, b \in G$ and $x \in X$ we have $(a \cdot b) * x = a * (b * x)$.

Examples.

(i) Any group acts on any set by the *trivial action* $g * x = x$.

(ii) Any group acts on the set $X = G$ by the *left regular action* $g * g' = g \cdot g'$ and the *right regular action* $g * g' = g' \cdot g$.

(iii) The dihedral group acts on the set of vertices of the regular n-gon.

Definition 15.25. Let G act on X. The *orbit* of $x \in X$ is the set

$$G * x \coloneqq \{y \in X \mid y = g * x \text{ for some } g \in G\}.$$

The *stabiliser* of $x \in X$ is

$$G_x \coloneqq \{g \in G \mid g * x = x\}.$$

Theorem 15.26 (Orbit-stabiliser theorem). Let a finite group G act on a set X.

$$|G| = |G_x||G * x|.$$

Definition 15.27. The *symmetric group* of a set X, $\mathrm{Sym}(X)$ is the group whose elements are all the bijection functions from the set to itself $\phi : X \to X$, and whose group operation is the composition of functions.

Theorem 15.28. The action $*$ of a group G on a set X is the same as the homomorphism

$$\begin{aligned} \rho : G &\to \mathrm{Sym}(X) \\ g &\mapsto t_g, \end{aligned}$$

where t_g is the function

$$\begin{aligned} t_g : X &\to X \\ x &\mapsto g * x. \end{aligned}$$

Theorem 15.29 (Cayley's theorem). Any group is isomorphic to a subgroup of some symmetric group.

Proof. Consider the left regular action of G on the set $X = G$. By the construction in Theorem 15.28, this corresponds to a homomorphism $\rho : G \to \mathrm{Sym}(G)$. The image $\mathrm{Im}(\rho)$ of ρ is a subgroup of $\mathrm{Sym}(G)$, and we may consider as a homomorphism $\rho : G \to \mathrm{Im}(\rho)$. If $g \in \ker(\rho)$ then $g * h = h$ for all $h \in G$, but as $g * h = g \cdot h$ it then follows that $g = I$, so $\ker(\rho) = \{I\}$. It then follows that $\rho : G \to \mathrm{Im}(\rho)$ is an isomorphism, so G is isomorphic to $\mathrm{Im}(\rho)$, which is a subgroup of $\mathrm{Sym}(G)$. □

15.6 Cosets and Lagrange's Theorem

15.6.1 Cosets

Definition 15.30. A *left coset* of a subgroup $H \leq G$ is an orbit of the right regular action of H on G. We write G/H for the set of orbits of this action, and call it *the set of left cosets*. If $g \in G$ then its left coset is

$$gH \coloneqq \{g' \in G \mid g' = gh \text{ for some } h \in H\}\,.$$

A *right coset* of $H \leq G$ is an orbit of the left regular action of H on G, and we write $H\backslash G$ for *the set of right cosets*.

Proposition 15.31. For a subgroup H of an Abelian group G, the left coset gH and the right coset Hg are identical.

Proof. Trivial by the definition of Abelian groups. □

Proposition 15.32. A subgroup H of G and its left (or right) cosets partition G.

Proof.

(i) Two cosets are either disjoint or equal.

Suppose g_1H and g_2H have one element in common: $g_1h_1 = g_2h_2$. Then

$$g_1H = g_2h_2h_1^{-1}H = g_2H$$

since $h_2h_1^{-1} \in H$. So if there is one element in common, the cosets are identical.

(ii) Two cosets g_1H and g_2H are identical if and only if $g_1^{-1}g_2 \in H$.

If $g_1^{-1}g_2 = h \in H$, then

$$g_1H = g_1hH = g_1g_1^{-1}g_2H = g_2H\,.$$

Conversely, if $g_1H = g_2H$,

$$H = g_1^{-1}g_2H\,,$$

so $g_1^{-1}g_2h \in H\ \forall h \in H$, so $g_1^{-1}g_2 \in H$.

(iii) Every element of G is in some coset.

Since H contains I, then for any element $g \in G$, the coset gH contains g. □

15.6.2 Lagrange's Theorem

Theorem 15.33 (Lagrange's theorem). Let G be a finite group and let $H \leq G$, then

$$|G| = n|H|, \quad n \in \mathbb{Z}\,.$$

Proof. This follows immediately from Proposition 15.32. □

Corollary. The order of every element of G also divides $|G|$.

Proof. Any $g \in G$ generates a cyclic subgroup of G. □

Proposition 15.34. Any group of prime order is cyclic.

Proof. Let G be a group of order p, where p is a prime number, and $g \in G$, $g \neq I$. Consider the subgroup H of G generated by g, which must be a cyclic group with order greater than 1. By Lagrange's theorem, a subgroup of G can only have order 1 or p, so the group generated by g must be G itself, and so G is cyclic. □

15.7 Conjugacy Class

15.7.1 Conjugacy Class

Definition 15.35. Two group elements $g_1, g_2 \in G$ are *conjugate* to each other if there exists some group element g such that

$$g_2 = g g_1 g^{-1}, \text{ or equivalently } g_2 g = g g_1 .$$

The *conjugate action* of G on a subgroup $H \leq G$ is

$$g * h \coloneqq g h g^{-1} .$$

Definition 15.36. An *equivalence relation* is a binary relationship between elements of a set, written $g_1 \sim g_2$, which satisfies

(i) *Reflexivity*: $g_1 \sim g_1$.

(ii) *Symmetry*: $g_1 \sim g_2$ implies $g_2 \sim g_1$.

(iii) *Transitivity*: If $g_1 \sim g_2$ and $g_2 \sim g_3$, then $g_1 \sim g_3$.

Proposition 15.37. Conjugacy is an equivalence relation.

Proof.

(i) Choose $g = I$.

$$g_1 = I g_1 I^{-1} .$$

(ii) Suppose that $g_2 = g g_1 g^{-1}$, then for $g' = g^{-1}$,

$$g_1 = g' g_2 g'^{-1} .$$

(iii) If $g_2 = g_a g_1 g_a^{-1}$ and $g_3 = g_b g_2 g_b^{-1}$, then

$$g_3 = g g_1 g^{-1} ,$$

where $g = g_b g_a$. □

Definition 15.38. The conjugacy relation partitions any group G into disjoint equivalence classes called *conjugacy classes*. Elements in the same class are conjugate, and elements in different classes are not.

Proposition 15.39. For a group G,

(i) For any $g \in G$, $a \in G$, gag^{-1} is in the same conjugacy class as a.

(ii) The identity of any group is a conjugacy class by itself.

(iii) Each element of an Abelian group is in a class by itself.

Proof.

(i) Trivial by definition.

(ii) For any $g \in G$,

$$g I g^{-1} = g g^{-1} = I .$$

(iii) For any g_1, g_2 in an Abelian group G,

$$g_2 g_1 g_2^{-1} = g_2 g_2^{-1} g_1 = g_1 ,$$

so g_1 can only be a conjugacy class by itself. □

Example. Conjugacy classes of D_4.

I	R^2	R	R^3	m_1	m_2	m_3	m_4
R^2	I	R^3	R	m_2	m_1	m_4	m_3
R	R^3	R^2	I	m_4	m_3	m_1	m_2
R^3	R	I	R^2	m_3	m_4	m_2	m_1
m_1	m_2	m_3	m_4	I	R^2	R	R^3
m_2	m_1	m_4	m_3	R^2	I	R^3	R
m_3	m_4	m_2	m_1	R^3	R	I	R^2
m_4	m_3	m_1	m_2	R	R^3	R^2	I

The 5 conjugacy classes are:

$$\{I\},\ \{R^2\},\ \{R,\ R^3\},\ \{m_1,\ m_2\},\ \{m_3,\ m_4\}\,.$$

15.7.2 Normal Subgroup

Definition 15.40. A subgroup H of G is a *normal subgroup* if for every $h \in H$, $g \in G$,

$$ghg^{-1} \in H\,.$$

We write $H \trianglelefteq G$ to mean H is a normal subgroup of G, and $H \triangleleft G$ to exclude $H = G$.

The normal subgroup $H \trianglelefteq G$ is *proper* if it is not $\{I\}$ or G.

Remark. A normal subgroup consists of complete conjugacy classes of the group.

Proposition 15.41. Any subgroup of an Abelian group is normal.

Proof. Let H be a subgroup of an Abelian group G. For all $g \in G$ and $h \in H$ we have

$$ghg^{-1} = gg^{-1}h = h\,.$$

□

Proposition 15.42. The left and right cosets of H are identical if and only if $H \trianglelefteq G$.

Proof.

$$gHg^{-1} = H \quad \iff \quad gH = Hg\,.$$

□

Proposition 15.43. If $\phi : H \to G$ is a homomorphism, then $\ker(\phi) \trianglelefteq H$.

Proof. We have $\ker(\phi) \leq H$ by Proposition 15.21. $\ker(\phi)$ is normal in H since for $k \in \ker(\phi)$ and $h \in H$,

$$\phi(hkh^{-1}) = \phi(h)\phi(k)\phi(h^{-1}) = \phi(h)I_G\phi(h)^{-1} = I_G\,,$$

so $hkh^{-1} \in \ker(\phi)$. □

15.7.3 Quotient Groups

We can try to define a group operation on the cosets of $H \leq G$ by

$$(g_1 H) \cdot (g_2 H) \coloneqq (g_1 g_2) H \,,$$

but we must be worried about where this is well-defined: the same coset may be represented by many elements of G, and we must show that the answer obtained does not depend on which representative of each coset we choose.

If $g_1 H = g_1' H$ and $g_2 H = g_2' H$, then we must have $g_1 = g_1' h_1$ and $g_2' = g_2 h_2$. Then

$$g_1' g_2' H = g_1 h_1 g_2 h_2 H = g_1 h_1 g_2 H = g_1 g_2 (g_2^{-1} h_1 g_2) H \,,$$

which is equal to $g_1 g_2 H$ (for which the product is therefore well-defined) if and only if $g_2^{-1} h_1 g_2 \in H$ holds for all $g_2 \in G$ and $h_1 \in H$. Therefore, H must be normal in G.

Theorem 15.44. If H is a normal subgroup of G, then

$$(g_1 H) \cdot (g_2 H) \coloneqq g_1 g_2 H$$

is a well-defined binary operation on the set G/H of left cosets, and $(G/H, \cdot, IH)$ is a group.

Proof. The binary operation is well-defined, as shown above. The associativity is satisfied since

$$\begin{aligned} (g_1 H \cdot g_2 H) \cdot g_3 H &= (g_1 g_2 H) \cdot g_3 H \\ &= (g_1 g_2 g_3) H \\ &= g_1 H \cdot (g_2 H \cdot g_3 H) \,. \end{aligned}$$

The identity is IH since $gH \cdot IH = (g \cdot I)H = gH$, and we also have the inverse: $gH \cdot g^{-1} H = (g \cdot g^{-1})H = IH$. $(G/H, \cdot, IH)$ is a well-defined group. □

Definition 15.45. If H is a normal subgroup of G, then the quotient group is the set G/H of left cosets with the group structure described in Theorem 15.44.

Theorem 15.46 (The first isomorphism theorem). Let $\phi : G \to H$ be a homomorphism. Then the mapping

$$\begin{aligned} \rho : G/\ker(\phi) &\to \mathrm{Im}(\phi) \\ g \ker(\phi) &\mapsto \phi(g) \end{aligned}$$

is well-defined and is a group isomorphism.

Proof. From Proposition 15.43, $\ker(\phi) \trianglelefteq G$, so we have a quotient group $G/\ker(\phi)$.

If $g \ker(\phi) = g' \ker(\phi)$, then $g' = gk$ for some $k \in \ker(\phi)$. Thus

$$\begin{aligned} \phi(g') &= \phi(g \cdot k) \\ &= \phi(g) \cdot \phi(k) \\ &= \phi(g) \cdot I = \phi(g) \,, \end{aligned}$$

and hence $\rho(g \ker(\phi)) = \rho(g' \ker(\phi))$, which means that ρ is well-defined.

We have

$$\begin{aligned} \rho(a \ker(\phi) \cdot b \ker(\phi)) &= \rho(ab \ker(\phi)) \\ &= \phi(a) \cdot \phi(b) \\ &= \rho(a \ker(\phi)) \cdot \rho(b \ker(\phi)) \,, \end{aligned}$$

so ρ is a homomorphism,

The function ρ is surjective, as the set $\mathrm{Im}(\phi)$ consists, by definition, of the elements of the form $\phi(g)$. If $\rho(a \ker(\phi)) = I_H$ then $\phi(a) = I_H$, so $a \in \ker(\phi)$, but then $a \ker(\phi) = I_G \ker(\phi)$. Thus the kernel of ρ is $\{I_G \ker(\phi)\}$, so ρ is injective, and is therefore an isomorphism. □

Example. Consider a homomorphism $\phi : D_3 \to C_2$, where

$$\{I_D, R, R^2, m_1, m_2, m_3\} \mapsto \{I_C, I_C, I_C, a, a, a\}.$$

The kernel of ϕ is

$$K = \{I_D, R, R^2\} = C_3\,.$$

We can check that D_3/C_3 is isomorphic to C_2.

15.8 The Permutation Groups

15.8.1 Permutation Groups and Permutations

Definition 15.47. The *finite symmetric group*, or the *permutation group*, S_n, is the symmetric group of a finite set of n elements.

Proposition 15.48. The order of a permutation group S_n is

$$|S_n| = n!\,.$$

Proposition 15.49. For $n > 1$, S_{n-1} is a subgroup of S_n.

Remark. Just think of leaving the n^{th} element fixed and permuting the other $n-1$ elements.

Since a permutation is a bijection of a set to itself, it can be represented using the following convention.

Definition 15.50. *Cauchy's two-line notation* is a notation of permutations putting each element in the first row and its image below in the second row. If $\sigma \in S_n$ is a permutation of the set $X = \{x_1, x_2, ..., x_n\}$, then

$$\sigma = \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ \sigma(x_1) & \sigma(x_2) & \dots & \sigma(x_n) \end{pmatrix}.$$

Remark. The inverse permutation is easily found by swapping the two rows.

$$\sigma^{-1} = \begin{pmatrix} \sigma(x_1) & \sigma(x_2) & \dots & \sigma(x_n) \\ x_1 & x_2 & \dots & x_n \end{pmatrix}.$$

Remark. We can omit the columns referring to the unchanged objects, and swap the order of any columns.

Example.

$$\begin{pmatrix} 2 & 3 & 1 & 4 & 5 \\ 3 & 1 & 2 & 4 & 5 \end{pmatrix} \equiv \begin{pmatrix} 2 & 3 & 1 \\ 3 & 1 & 2 \end{pmatrix} \equiv \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}.$$

Definition 15.51. For a permutation group S_n acting on a set of objects $X = \{a_i\}_{i=1}^n$, a k-cycle

$$(a_1\ a_2\ \dots\ a_k) \in S_n\,,$$

on the elements $a_1, a_2, \dots, a_k \in \{x_1, x_2, \dots, x_n\}$ is the permutation given by

$$\begin{pmatrix} a_1 & a_2 & \dots & a_k \\ a_2 & a_3 & \dots & a_1 \end{pmatrix}.$$

Remark. Remember that $\sigma \in S_n$ is a bijection function acting on a set X.

For example, let $a_1, a_2, \dots, a_k \in \{1, 2, \dots, n\}$, then the k cycle

$$(a_1\ a_2\ \dots\ a_k)(i) = \begin{cases} a_{j+1} & \text{if } i = a_j \text{ for } j < k \\ a_1 & \text{if } i = a_k \\ i & \text{if } i \neq a_j \text{ for any } j\,. \end{cases}$$

Lemma 15.52.

(i) Cycles can be cycled. $(a_1\ a_2\ \dots\ a_k) = (a_k\ a_1\ \dots\ a_{k-1})$.

(ii) Disjoint cycles commute. If $\sigma = (\sigma_1\ \dots\ \sigma_m)$ and $\tau = (\tau_1\ \dots\ \tau_n)$ are disjoint cycles ($\{\sigma_i\}_{i=1}^m \cap \{\tau_j\}_{j=1}^n = \varnothing$), then

$$\sigma\tau = \tau\sigma\,.$$

Theorem 15.53. Any permutation can be uniquely decomposed into disjoint cycles up to

- cycling the terms in a cycle;
- reordering the cycles.

Proof. Consider a general permutation $\sigma \in S_n$:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & \dots & n \\ p_1 & p_2 & p_3 & \dots & p_n \end{pmatrix}.$$

In this notation, we can rearrange the n columns in any order without changing the meaning of the permutation. Move the second column corresponding to p_1, and the third column corresponding to the second row of the second column, etc., until there is a subset of columns on the left arranged in the k-cycle form:

$$\sigma = \begin{pmatrix} 1 & p_1 & p_r & \dots & p_t & \dots \\ p_1 & p_r & p_s & \dots & 1 & \dots \end{pmatrix}.$$

Repeat this process to the remaining columns until all the columns have been arranged in groups of k-cycle forms. The k-cycles constructed in this way are necessarily disjoint. □

Definition 15.54. The *cycle shape* of a permutation $\sigma \in S_n$ is the list of numbers $(n_2, n_3, \dots)$ specifying the number of 2-cycles, 3-cycles, etc. in the unique decomposition of σ into disjoint cycles.

Lemma 15.55. 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. If $\sigma^{-1}(i) \notin \{a_1, a_2, \dots, a_k\}$, then

$$\sigma(a_1\ a_2\ \dots\ a_k)(\sigma^{-1}(i)) = \sigma\sigma^{-1}(i) = i\,.$$

For the right hand side, as $i \notin \{\sigma(a_1), \sigma(a_2), \dots, \sigma(a_k)\}$, $(\sigma(a_1)\ \sigma(a_2)\ \dots\ \sigma(a_k))$ also fixes i.

If $\sigma^{-1}(i) = a_j$, then

$$\begin{aligned}(\sigma(a_1\ a_2\ \dots\ a_k)\sigma^{-1})(i) &= (\sigma(a_1\ a_2\ \dots\ a_k))(\sigma^{-1}(i)) \\ &= (\sigma(a_1\ a_2\ \dots\ a_k))(a_j) \\ &= \sigma(a_{j+1})\,,\end{aligned}$$

which is also the result of applying $(\sigma(a_1)\ \sigma(a_2)\ \dots\ \sigma(a_k))$ to $i = \sigma(a_j)$. □

Theorem 15.56. Elements $\tau, \tau' \in S_n$ are conjugate if and only if they have exactly the same cycle shape.

Proof. If

$$\tau = (a_1^1\ a_2^1\ \dots\ a_{k_1}^1)(a_1^2\ a_2^2\ \dots\ a_{k_2}^2)\dots(a_1^r\ a_2^r\ \dots\ a_{k_r}^r)$$

is the disjoint cycle decomposition of τ, then for $\sigma \in S_n$, by Lemma 15.55, we have

$$\sigma\tau\sigma^{-1} = (\sigma(a_1^1)\ \sigma(a_2^1)\ \dots\ \sigma(a_{k_1}^1))(\sigma(a_1^2)\ \sigma(a_2^2)\ \dots\ \sigma(a_{k_2}^2))\dots(\sigma(a_1^r)\ \sigma(a_2^r)\ \dots\ \sigma(a_{k_r}^r))\,,$$

which have an identical cycle shape.

Conversely, if τ and τ' have exactly the same cycle shape, write

$$\tau = (a_1^1\ a_2^1\ \dots\ a_{k_1}^1)(a_1^2\ a_2^2\ \dots\ a_{k_2}^2)\dots(a_1^r\ a_2^r\ \dots\ a_{k_r}^r)$$

$$\tau' = (b_1^1\ b_2^1\ \dots\ b_{k_1}^1)(b_1^2\ b_2^2\ \dots\ b_{k_2}^2)\dots(b_1^r\ b_2^r\ \dots\ b_{k_r}^r)\,.$$

Define a function $\sigma : \{1, 2, \dots, n\} \to \{1, 2, \dots, n\}$ by

$$\sigma(a_j^i) = b_j^i\,,$$

then by construction, $\tau' = \sigma\tau\sigma^{-1}$, so τ and τ' are conjugate.

15.8.2 Transpositions

Definition 15.57. A 2-cycle is called a *transposition*.

Proposition 15.58. An n-cycle can be decomposed into $(n-1)$ 2-cycles.

Proof.

$$\begin{aligned}(a_1\ a_2\ \dots\ a_n) &= (a_1\ a_n)(a_1\ a_2\ \dots\ a_{n-1})\\ &= (a_1\ a_n)(a_1\ a_{n-1})\dots(a_1\ a_3)(a_1\ a_2)\,.\end{aligned}$$

□

Remark. Note that these 2-cycles are not disjoint and therefore non-commutative.

Definition 15.59. A permutation $\sigma \in S_n$ is *odd* if it is a composition of an odd number of 2-cycles, and is *even* if it is a composition of an even number of 2-cycles. The evenness and oddness is called the *parity* of the permutation.

The *sign* of the permutation is

$$\operatorname{sign}(\sigma) \coloneqq (-1)^{\#\text{of transpositions in the composition of } \sigma}\,.$$

Lemma 15.60. Every permutation $\sigma \in S_n$ is a composition of transpositions, and the sign of a permutation is unique. That is, the function

$$\operatorname{sign} : S_n \to \{-1, 1\}$$

is well-defined.

Theorem 15.61. The function $\operatorname{sign} : S_n \to \{-1, 1\} = C_2$ is a homomorphism.

Proof. If σ can be written as a composition of a transpositions and τ can be written as a composition of b transpositions, then $\sigma\tau$ can be written as $a + b$ transpositions. Thus,

$$\operatorname{sign}(\sigma\tau) = (-1)^{a+b} = (-1)^a(-1)^b = \operatorname{sign}(\sigma)\cdot\operatorname{sign}(\tau)\,.$$

□

Definition 15.62. The *alternating group* A_n is the subgroup of S_n consisting of even permutations.

Corollary. A_n is the kernel of the homomorphism $\operatorname{sign} : S_n \to C_2$, so is a normal subgroup of S_n by Proposition 15.43.

16 Representation Theory

16.1 Group of Matrices

Definition 16.1. The n^{th} *general linear group* over $\mathbb{F}$ is

$$GL(n,\mathbb{F}) \coloneqq (\{X \in \mathrm{Mat}_{n\times n}(\mathbb{F}) \mid \det(X) \neq 0\}, \cdot, \mathsf{I}_{n\times n}),$$

where $\mathrm{Mat}_{n\times n}(\mathbb{F}) \coloneqq \{n \times n \text{ matrices with entries in } \mathbb{F}\}$ and $\mathbb{F}$ is usually $\mathbb{R}$ or $\mathbb{C}$.

It can be easily confirmed that $GL(n,\mathbb{F})$ satisfies the axioms of groups.

Lemma 16.2. The mapping

$$\det : (GL(n,\mathbb{F}), \cdot, \mathsf{I}_{n\times n}) \to (\mathbb{F}\setminus\{0\}, \times, 1)$$

is a group homomorphism.

Proof. This is because the determinant satisfies

$$\det(\mathsf{A}\cdot\mathsf{B}) = \det(\mathsf{A}) \times \det(\mathsf{B}).$$

□

Definition 16.3. The n^{th} *special linear group* over $\mathbb{F}$ is

$$SL(n,\mathbb{F}) \coloneqq (\{X \in \mathrm{Mat}_{n\times n}(\mathbb{F}) \mid \det(X) = 1\}, \cdot, \mathsf{I}_{n\times n}).$$

Proposition 16.4. $SL(n,\mathbb{F})$ is a normal subgroup of $GL(n,\mathbb{F})$.

Proof. $SL(n,\mathbb{F})$ is the kernel of the map $\det : GL(n,\mathbb{F}) \to \mathbb{F}\setminus\{0\}$. □

Remark. $\det : GL(n,\mathbb{F}) \to \mathbb{F}\setminus\{0\}$ is surjective, so by the isomorphism theorem, we have

$$\frac{GL(n,\mathbb{F})}{SL(n,\mathbb{F})} \cong \mathbb{F}\setminus\{0\}.$$

Remark. The group $GL(n,\mathbb{F})$ acts on $\mathbb{F}^n$, where we think of $\mathbb{F}^n$ as column vectors, via

$$\begin{aligned} GL(n,\mathbb{F}) \times \mathbb{F}^n &\to \mathbb{F}^n \\ (\mathsf{A}, \mathbf{v}) &\mapsto \mathsf{A}\mathbf{v}. \end{aligned}$$

This corresponds to a homomorphism

$$\rho : GL(n,\mathbb{F}) \to \mathrm{Sym}(\mathbb{F}^n),$$

which is injective, and whose image consists of those bijections $\mathbb{F}^n \to \mathbb{F}^n$ which are linear maps. This gives an isomorphism from $GL(n,\mathbb{F})$ to the group of invertible linear maps from $\mathbb{F}^n$ to itself.

Definition 16.5. The n^{th} *orthogonal group* is

$$O(n) \coloneqq \{X \in GL(n,\mathbb{R}) \mid X^{\mathrm{T}}X = \mathsf{I}_{n\times n}\},$$

and the n^{th} *special orthogonal group* is

$$SO(n) \coloneqq \{X \in SL(n,\mathbb{R}) \mid X^{\mathrm{T}}X = \mathsf{I}_{n\times n}\}.$$

Definition 16.6. The n^{th} *unitary group* is

$$U(n) \coloneqq \{X \in GL(n,\mathbb{C}) \mid X^{\dagger}X = \mathsf{I}_{n\times n}\},$$

and the n^{th} *special unitary group* is

$$SU(n) \coloneqq \{X \in SL(n,\mathbb{C}) \mid X^{\dagger}X = \mathsf{I}_{n\times n}\}.$$

16.2 Representation

Definition 16.7. A *representation* of a group G acting on a vector space V over a field $\mathbb{F}$ is a group homomorphism from G to $GL(V)$. It is a map

$$\rho : G \to GL(V)$$

such that

$$\rho(g_1 g_2) = \rho(g_1)\rho(g_2)$$

for all $g_1, g_2 \in G$. We will denote the representation as (ρ, V), or simply ρ.

16.2.1 Faithful Representation

Definition 16.8. A representation of a group G is *faithful* when the kernel of the mapping from G to $GL(V)$ is $\{I\}$.

Remark. This is equivalent to saying that the homomorphism from G to $GL(V)$ is injective (one-to-one). Under such constraint, the group G will be isomorphic to a subgroup of $GL(V)$.

Example. A faithful representation of D_4 is

$$\rho(I) = \mathsf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \rho(R) = \mathsf{R} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

$$\rho(R^2) = \mathsf{R}^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \qquad \rho(R^3) = \mathsf{R}^3 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

$$\rho(m_1) = \mathsf{m}_1 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \rho(m_2) = \mathsf{m}_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

$$\rho(m_3) = \mathsf{m}_3 = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \qquad \rho(m_4) = \mathsf{m}_4 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Remark. Underlying vector space.

Imagine an arbitrary vector $\mathbf{x}$ in the plane of the square with its tail at the centre. A left-multiplication by the matrices in the faithful representations of D_4 groups performs exactly the corresponding symmetry operations to $\mathbf{x}$.

It can be seen that the two Vierergruppen leave some subspaces of $\mathbb{R}^2$ unchanged. The Vierergruppe $\{I, R^2, m_1, m_2\}$ leaves invariant the two subspaces of $\mathbb{R}^2$ consisting of the scalar multiples of $(1,0)^{\mathrm{T}}$ and $(0,1)^{\mathrm{T}}$. The Vierergruppe $\{I, R^2, m_3, m_4\}$ leaves invariant the two subspaces of $\mathbb{R}^2$ consisting of the scalar multiples of $(1,1)^{\mathrm{T}}$ and $(-1,1)^{\mathrm{T}}$.

16.2.2 Regular Representation

Claim 16.9. Any finite group can be represented faithfully by matrices.

Remark. This is a consequence of Cayley's theorem (Theorem 15.29). Since the symmetric group S_n has a natural faithful permutation representation as the group of $n \times n$ matrices with entries only 0 and 1, and exactly one 1 in each row and column (see the regular representation below), it follows that every finite group is a matrix group.

We will demonstrate this by constructing the *regular representation* of a group.

Consider a group G of order n. Map the identity in the group $I \in G$ to the $n \times n$ identity matrix $\mathsf{I}_{n\times n}$. We are then able to form a set of $(n-1)$ other matrices by rearranging the rows of the identity matrix in such a way as to correspond to the group table.

Example. The Regular Representation of D_3.

We can work out the regular representation of D_3 from the group table.

I	R	R^2	m_1	m_2	m_3
R	R^2	I	m_3	m_1	m_2
R^2	I	R	m_2	m_3	m_1
m_1	m_2	m_3	I	R	R^2
m_2	m_3	m_1	R^2	I	R
m_3	m_1	m_2	R	R^2	I

$$\rho(I) = \begin{pmatrix} 1&0&0&0&0&0\\ 0&1&0&0&0&0\\ 0&0&1&0&0&0\\ 0&0&0&1&0&0\\ 0&0&0&0&1&0\\ 0&0&0&0&0&1 \end{pmatrix} \qquad \rho(R) = \begin{pmatrix} 0&0&1&0&0&0\\ 1&0&0&0&0&0\\ 0&1&0&0&0&0\\ 0&0&0&0&1&0\\ 0&0&0&0&0&1\\ 0&0&0&1&0&0 \end{pmatrix}$$

$$\rho(R^2) = \begin{pmatrix} 0&1&0&0&0&0\\ 0&0&1&0&0&0\\ 1&0&0&0&0&0\\ 0&0&0&0&0&1\\ 0&0&0&1&0&0\\ 0&0&0&0&1&0 \end{pmatrix} \qquad \rho(m_1) = \begin{pmatrix} 0&0&0&1&0&0\\ 0&0&0&0&1&0\\ 0&0&0&0&0&1\\ 1&0&0&0&0&0\\ 0&1&0&0&0&0\\ 0&0&1&0&0&0 \end{pmatrix}$$

$$\rho(m_2) = \begin{pmatrix} 0&0&0&0&1&0\\ 0&0&0&0&0&1\\ 0&0&0&1&0&0\\ 0&0&1&0&0&0\\ 1&0&0&0&0&0\\ 0&1&0&0&0&0 \end{pmatrix} \qquad \rho(m_3) = \begin{pmatrix} 0&0&0&0&0&1\\ 0&0&0&1&0&0\\ 0&0&0&0&1&0\\ 0&1&0&0&0&0\\ 0&0&1&0&0&0\\ 1&0&0&0&0&0 \end{pmatrix}$$

The row of g in the group table is obtained by multiplying the row vector $(I, R, R^2, m_1, m_2, m_3)$ on the left of $\rho(g)$. For example,

$$(I, R, R^2, m_1, m_2, m_3)\rho(R) = (R, R^2, I, m_3, m_1, m_2)\,.$$

If we multiply the column vector $(I, R, R^2, m_1, m_2, m_3)^{\mathrm{T}}$ by $\rho(g)$ on the right, then we get the row corresponding to g^{-1}. For example,

$$\rho(R)\begin{pmatrix} I\\ R\\ R^2\\ m_1\\ m_2\\ m_3 \end{pmatrix} = \begin{pmatrix} R^2\\ I\\ R\\ m_2\\ m_3\\ m_1 \end{pmatrix}\,.$$

Remark. Note that the traces of all the matrices above vanish, except for the identity matrix $\rho(I) = \mathsf{I}$.

$$\mathrm{tr}(\rho(g)) = \begin{cases} |G| & \text{if } g = I\\ 0 & \text{otherwise}\,. \end{cases}$$

Note also that the determinants are all ± 1, with a positive sign for the identity and the rotations and a negative sign for the reflections. This corresponds to the sign of the permutations if we think of rotations as permuting the 3 vertices.

Remark. For the regular representation of any $g \in G$, we have

$$(\rho(g))^n = \mathsf{I}\,,$$

where n is the order of G. This follows from the isomorphism $g \mapsto \rho(g)$ and that $g^n = I$ for every $g \in G$. This restricts the determinants of the regular representation

$$\det(\rho(g))^n = 1\,,$$

i.e. $\det(\rho(g))$ is an n^{th} root of unity.

16.3 Equivalence and Inequivalence

Definition 16.10. If (ρ, V) and (ρ', W) are representations of G, we say a linear map $\phi : V \to W$ is a *G-linear map* if

$$\phi \circ \rho(g) = \rho'(g) \circ \phi$$

for all $g \in G$. The vector space of G-linear maps between two representations (ρ, V) and (ρ', W) of G is denoted as $\mathrm{Hom}_G(V, W)$.

Definition 16.11. If the G-linear map ϕ between two representations (ρ, V) and (ρ', W) is an isomorphism (and therefore is invertible), then

$$\rho'(g) = \phi \circ \rho(g) \circ \phi^{-1}$$

for all $g \in G$. We then say that ϕ *intertwines* ρ and ρ'.

Remark. $\phi \in \mathrm{Hom}_G(V, W)$ is an intertwining map precisely if ϕ is a bijection.

We often write the intertwining maps between representations as matrices.

Definition 16.12. The two sets of matrices $\{\rho(g_i)\}_{i=1}^n$ and $\{\rho'(g_i)\}_{i=1}^n$ representing a group G are called *equivalent* if there exists a invertible matrix S such that, for all i,

$$\rho'(g_i) = \mathsf{S}\rho(g_i)\mathsf{S}^{-1}\,.$$

Such transformation from ρ to ρ' is called a *similarity transformation*. If no such S exists, then the two representations are inequivalent.

Remark. This corresponds to a change of basis in the underlying vector space.

Example. Two equivalent representations of C_4.

Consider the two representations of the C_4 group:

$$\rho(I) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \rho(g) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \rho(g^2) = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \quad \rho(g^3) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

and

$$\phi(I) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \phi(g) = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \quad \phi(g^2) = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \quad \phi(g^3) = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}.$$

These two representations are equivalent because they are related by

$$\mathsf{S} = \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}, \quad \mathsf{S}^{-1} = \frac{1}{2}\begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}.$$

Example. Two inequivalent representations of C_4.

Consider the two representations of the C_4 group:

$$\rho(I) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \rho(g) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \rho(g^2) = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \quad \rho(g^3) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

and

$$\psi(I) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \psi(g) = \begin{pmatrix} i & 0 \\ 0 & i \end{pmatrix} \quad \psi(g^2) = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \quad \psi(g^3) = \begin{pmatrix} -i & 0 \\ 0 & -i \end{pmatrix}.$$

The two representations are now inequivalent. The only solution of S for $\rho(g_2)\mathsf{S} = \mathsf{S}\psi(g_2)$ has $\det \mathsf{S} = 0$, and so any choice of S would be non-invertible.

Example. One-dimensional representations of C_n.

Consider the cyclic group of order n, C_n:

$$C_n = \{I, g, \dots, g^{n-1}\}.$$

Let

$$\omega = \exp\left(\frac{2\pi i}{n}\right),$$

then C_n has a faithful representation defined by

$$\rho(I) = 1 \quad \rho(g) = \omega \quad \rho(g^2) = \omega^2 \quad \dots \quad \rho(g^{n-1}) = \omega^{n-1}.$$

Now consider the cyclic group of prime order, C_p, where p is a prime number. Then we can find $p-1$ one-dimensional faithful representations of C_p, given by

$$\rho_q(I) = 1 \quad \rho_q(g) = \omega^q \quad \rho_q(g^2) = \omega^{2q} \quad \dots \quad \rho_q(g^{p-1}) = \omega^{q(p-1)},$$

for $q \in \{1, 2, \dots, p-1\}$. All these representations are faithful since

$$\rho_q(g^r) \neq \rho(I)$$

for all $q, r \in \{1, 2, \dots, p-1\}$, and they are all clearly inequivalent.

Example. Quaternions.

The quaternions form an order-8 group Q_8. It can be faithfully represented by 2×2 matrices. The elements of Q_8 can be denoted as $\{\pm 1, \pm\mathcal{I}, \pm\mathcal{J}, \pm\mathcal{K}\}$, where

$$1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \mathcal{I} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \qquad \mathcal{J} = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} \qquad \mathcal{K} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}.$$

There are some properties of Q:

(i) Q_8 has three order-4 subgroups, all isomorphic to C_4.

(ii) Q_8 has only one order-2 subgroup.

(iii) All the subgroups of Q_8 are normal

(iv) Q_8 is not Abelian.

We can observe the *Hamilton's relations*:

$$\mathcal{I}^2 = \mathcal{J}^2 = \mathcal{K}^2 = \mathcal{I}\mathcal{J}\mathcal{K} = -1,$$

and further we have

$$\mathcal{I}\mathcal{J} = \mathcal{K} = -\mathcal{J}\mathcal{I}.$$

16.4 Characters

Definition 16.13. For a representation of a group G, $\rho : G \to GL(V)$, where V is a finite-dimensional vector space over $\mathbb{F}$, the *character* of ρ is the function $\chi_\rho : G \to \mathbb{F}$ given by

$$\chi_\rho(g) = \mathrm{tr}(\rho(g)) \,.$$

Example. Character of regular representations.

In regular representations, a group G of order n is represented by n matrices of dimensions $n \times n$. Only the representation of the identity is $\mathrm{tr}(\mathsf{I}) = n$, while others are traceless. Therefore, the character of the regular representation of G is

$$\{n, 0, 0, ..., 0\} \,.$$

Theorem 16.14. Two finite-dimensional complex representations of a finite group have the same character if and only if they are equivalent.

Proof. We will prove the theorem in one direction only. Traces are invariant under cyclic permutations of matrices.

$$\mathrm{tr}(\mathsf{ABC}\dots\mathsf{MN}) = A_{ij}B_{jk}\dots M_{mn}N_{ni} = \mathrm{tr}(\mathsf{BC}\dots\mathsf{MNA}) \,.$$

Therefore,

$$\mathrm{tr}\big(\mathsf{S}\rho(g)\mathsf{S}^{-1}\big) = \mathrm{tr}\big(\rho(g)\mathsf{S}^{-1}\mathsf{S}\big) = \mathrm{tr}(\rho(g)) \,.$$

□

Example. Representations of the cyclic group C_4.

We have three representations of C_4:

$$\rho(I) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \rho(g) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \rho(g^2) = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \quad \rho(g^3) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} ,$$

$$\phi(I) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \phi(g) = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \quad \phi(g^2) = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \quad \phi(g^3) = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix} ,$$

$$\psi(I) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \psi(g) = \begin{pmatrix} i & 0 \\ 0 & i \end{pmatrix} \quad \psi(g^2) = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \quad \psi(g^3) = \begin{pmatrix} -i & 0 \\ 0 & -i \end{pmatrix} .$$

The representations ρ and ϕ are equivalent, while ψ is inequivalent to them. The characters of these representations are

$$\chi_\rho = \{2, 0, -2, 0\} \,,$$
$$\chi_\phi = \{2, 0, -2, 0\} \,,$$
$$\chi_\psi = \{2, 2i, -2, -2i\} \,.$$

We can see that, as stated in Theorem 16.14, the characters are the same for equivalent representations and different for inequivalent representations.

Proposition 16.15. Characters are the same within a conjugacy class.

Proof. If $g_1, g_2 \in G$ are conjugate, then there exists some $g \in G$ such that

$$g_2 = g g_1 g^{-1} \,.$$

Consider the faithful representation ρ of G,

$$\rho(g_2) = \rho(g)\rho(g_1)(\rho(g))^{-1} \,.$$

$$\implies \mathrm{tr}(\rho(g_2)) = \mathrm{tr}(\rho(g_1)) \,,$$

so g_2 and g_1 have the same character. □

16.5 Reducibility

Definition 16.16. Given two matrices M and N, their *direct sum* is given by

$$\mathsf{M} \oplus \mathsf{N} = \begin{pmatrix} & & & 0 & \cdots & 0 \\ & \mathsf{M} & & \vdots & \ddots & \vdots \\ & & & 0 & \cdots & 0 \\ 0 & \cdots & 0 & & & \\ \vdots & \ddots & \vdots & & \mathsf{N} & \\ 0 & \cdots & 0 & & & \end{pmatrix}.$$

Definition 16.17. The *direct sum* of two representations $\rho : G \to GL(U)$ and $\phi : G \to GL(V)$ is given by

$$\begin{aligned} \rho \oplus \phi : G &\to GL(U \oplus V) \\ g &\mapsto \rho(g) \oplus \phi(g) \end{aligned}$$

Proposition 16.18. If $\rho : G \to GL(U)$ and $\phi : G \to GL(V)$ are two finite-dimensional representations of a group G, then the direct sum of the two representations $\psi = \rho \oplus \phi$ is also a representation of G.

Proof.

$$\begin{aligned} \psi(g_i)\psi(g_j) &= (\rho(g_i) \oplus \phi(g_i))(\rho(g_j) \oplus \phi(g_j)) \\ &= (\rho(g_i)\rho(g_j)) \oplus (\phi(g_i)\phi(g_j)) \\ &= \rho(g_i g_j) \oplus \phi(g_i g_j) \\ &= \psi(g_i g_j)\,. \end{aligned}$$

□

Corollary. The direct sum of multiple representations of G,

$$\begin{aligned} \bigoplus_{i=1}^{n} \rho_i : G &\to GL\left(\bigoplus_{i=1}^{n} V_i\right) \\ g &\mapsto \bigoplus_{i=1}^{n} \rho_i(g)\,, \end{aligned}$$

is also a representation of G.

We know that smaller dimensional representations can combine to form larger ones. How can we know if a representation can be broken down into smaller representations or not (reducibility)?

Definition 16.19. Let $\rho : G \to GL(V)$ be a representation of G. A linear subspace $W \subseteq V$ is called *G-invariant* if for all $g \in G$ and $\mathbf{w} \in W$,

$$\rho(g)\mathbf{w} \in W\,.$$

Definition 16.20. If $\rho : G \to GL(V)$ is a representation of G and the subspace W of V is G-invariant, then we may define a representation $\rho_W : G \to GL(W)$ by

$$\rho_W(g)\mathbf{w} = \rho(g)\mathbf{w} \text{ for } \mathbf{w} \in W\,.$$

We call ρ_W a *subrepresentation* of ρ.

Remark. All representations can form a subrepresentation with the *trivial G-invariant subspaces*, including the whole vector space V, and $\{\mathbf{0}\}$.

Definition 16.21. A representation

$$\rho : G \to GL(V)$$

is said to be *irreducible* if it has only trivial subrepresentations in V and $\{\mathbf{0}\}$. If there is a proper (nontrivial) invariant subspace, then ρ is said to be *reducible*.

Proposition 16.22. Suppose $\rho : G \to GL(V)$ is a representation, and $V = U \oplus W$. Then ρ is a direct sum of two subrepresentations (ρ_U, U) and (ρ_W, W) if and only if there is a basis $\mathbf{v}_1, \dots, \mathbf{v}_d$ of V such that $\mathbf{v}_1, \dots, \mathbf{v}_r$ is a basis of U and $\mathbf{v}_{r+1}, \dots, \mathbf{v}_d$ is a basis for W and the matrices $\rho(g)$ are all block diagonal.

Example. Representations of Vierergruppe.

Consider the following two representations of V_4.

$$\rho(I) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \rho(g_1) = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \quad \rho(g_2) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \quad \rho(g_3) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

$$\phi(I) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \phi(g_1) = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \quad \phi(g_2) = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \quad \phi(g_3) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Note that the ρ representation is diagonal, so ρ is reducible to two 1-dimensional representations. ϕ is an equivalent representation of ρ, related by a similarity transformation:

$$\rho(g) = \mathsf{S}\phi(g)\mathsf{S}^{-1}, \text{ where } \mathsf{S} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

This similarity transformation corresponds to a rotation of 45°. Since the representation ρ has invariant subspaces $U_\rho = \text{span}((1,0))$, $W_\rho = \text{span}((0,1))$, the representation ϕ should have invariant subspaces $U_\phi = \text{span}((1,1))$, $W_\phi = \text{span}((1,-1))$.

We have seen that the direct sum of representations of G forms block diagonal matrices, which is a reducible representation. However, the inverse statement is not necessarily true. The matrices of reducible representation need not be fully block-diagonal.

For example, consider the product of two block-triangular matrices

$$\begin{pmatrix} a_1 & b_1 & e_1 \\ c_1 & d_1 & f_1 \\ 0 & 0 & h_1 \end{pmatrix} \begin{pmatrix} a_2 & b_2 & e_2 \\ c_2 & d_2 & f_2 \\ 0 & 0 & h_2 \end{pmatrix} = \begin{pmatrix} a_1a_2 + b_1c_2 & a_1b_2 + b_1d_2 & a_1e_2 + b_1f_2 + e_1h_2 \\ c_1a_2 + d_1c_2 & c_1b_2 + d_1d_2 & c_1e_2 + d_1f_2 + f_1h_2 \\ 0 & 0 & h_1h_2 \end{pmatrix}.$$

We can see that the matrix multiplication of the upper-left 2×2 block and lower-right 1×1 block is preserved. Therefore, this 3-dimensional representation can be reduced to a two-dimensional and a one-dimensional representation:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ and } (h).$$

However, note that there is only one invariant subspace, which is spanned by $(1,0,0)$ and $(0,1,0)$. Vectors proportional to $(0,0,x_3)$ will not remain in the space spanned by $(0,0,1)$.

Proposition 16.23. Suppose $\rho : G \to GL(V)$ is a representation and $W \leq V$, then (ρ_W, W) is a subrepresentation, and ρ is therefore reducible, if and only if there is a basis $\mathbf{v}_1, \dots, \mathbf{v}_d$ of V such that $\mathbf{v}_1, \dots, \mathbf{v}_r$ is a basis of W and the matrices $\rho(g)$ are all upper-block-triangular.

$$\rho(g) = \begin{pmatrix} \boxed{\rho_W(g)} & \boxed{\mathsf{A}(g)} \\ \begin{matrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{matrix} & \boxed{\rho_U(g)} \end{pmatrix}.$$

Remark. In this case, (ρ, V) is not a direct sum of (ρ_W, W) and (ρ_U, U), and (ρ, V) is said *not fully reducible*.

Definition 16.24. A representation is *fully reducible* if it is block diagonal and can be written as a direct sum of its subrepresentations.

Remark. This is the case when we include infinite groups. If we only consider the finite groups, we can state something further.

However, before we proceed, here is a quick caveat. Throughout this chapter, we will only consider fields $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, which are of characteristic zero. This means that in their algebraic structures,

$$1 + 1 + \cdots + 1 \neq 0\,,$$

where 1 is the multiplicative identity and 0 is the additive identity of the field.

Here is a quick glance of how we will proceed.

Theorem 16.25 (Maschke's theorem). Let G be a finite group and let (ρ, V) be a representation of G over a field $\mathbb{F}$ of characteristic zero. Suppose $W \leq V$ is a G-invariant subspace. Then there is a G-invariant complement to W, i.e. a G-invariant subspace U of V such that $V = U \oplus W$.

Corollary (fully reducibility). If G is a finite group, and (ρ, V) is a reducible representation of G over a field of characteristic zero. Then $V = \bigoplus_{i=1}^{r} W_i$ is a direct sum of representations with each W_i irreducible.

Remark. For a finite group, reducible is equivalent to fully reducible. This is because, for finite groups, every finite-dimensional representation is equivalent to a unitary representation.

16.5.1 Group-invariant Inner Product and Unitarity

Definition 16.26. Recall that if V is a complex vector space then a *Hermitian inner product* is a map $(-,-) : V \times V \to \mathbb{C}$ satisfying

(i) *Sesquilinear.* For $a, b \in \mathbb{C}$, $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$,

$$(a\mathbf{x} + b\mathbf{y}, \mathbf{z}) = a^*(\mathbf{x}, \mathbf{z}) + b^*(\mathbf{y}, \mathbf{z})\,,$$

$$(\mathbf{x}, a\mathbf{x} + b\mathbf{z}) = a(\mathbf{x}, \mathbf{y}) + b(\mathbf{x}, \mathbf{z})\,.$$

(ii) *Hermitian.*

$$(\mathbf{x}, \mathbf{y}) = (\mathbf{y}, \mathbf{x})^*\,.$$

(iii) *Positive definite.* For all $\mathbf{x} \in V \setminus \{\mathbf{0}\}$,

$$(\mathbf{x}, \mathbf{x}) > 0\,.$$

Remark. The standard inner product on $\mathbb{C}^n$ is given by

$$\langle \mathbf{x} | \mathbf{y} \rangle = \sum_{i=1}^{n} x_i^* y_i\,.$$

Definition 16.27. Let (ρ, V) be a representation of G. A Hermitian inner product on a V is *G-invariant* if

$$(\rho(g)\mathbf{x}, \rho(g)\mathbf{y}) = (x, y)$$

for all $g \in G$ and $\mathbf{x}, \mathbf{y} \in V$.

Definition 16.28. We say that a representation (ρ, V) of a group G is *unitary* if there is a basis of V such that the corresponding map $G \to GL(n, \mathbb{C})$ has an image inside $U(n)$.

Lemma 16.29. A representation (ρ, V) of G is unitary if and only if V has a G-invariant inner product.

Proof. If (ρ, V) is unitary then let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be a basis for V such that $\rho(g) \in U(n)$ for all $g \in G$. It is easy to check that

$$\left(\sum_{i=1}^{n} \lambda_i \mathbf{e}_i, \sum_{j=1}^{n} \mu_j \mathbf{e}_j \right) = \sum_{i=1}^{n} \lambda_i^* \mu_i$$

defines a G-invariant inner product on V.

Conversely, if V has a G-invariant inner product $(-,-)$, we can find an orthonormal basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ for V. Then $(-,-)$ corresponds to the standard inner product with respect to this basis. So if the inner product is G-invariant, then

$$(\mathbf{v}_1, \mathbf{v}_2) = (\rho(g)\mathbf{v}_1, \rho(g)\mathbf{v}_2) = (\rho(g)^\dagger \rho(g)\mathbf{v}_1, \mathbf{v}_2)$$

for all $g \in G$ and $\mathbf{v}_1, \mathbf{v}_2 \in V$, and so

$$\rho(g)^\dagger \rho(g) = \mathsf{I}\,,$$

which implies that ρ is unitary. □

Theorem 16.30 (Weyl's unitary trick). If (ρ, V) is a complex representation of a finite group G, then there is a G-invariant Hermitian inner product on V, and (ρ, V) is unitary.

Proof. Pick any Hermitian inner product (e.g. choose a basis $\{\mathbf{e}_i\}$ for V and use the standard inner product). Define the new inner product $[-,-]$ by

$$[\mathbf{x}, \mathbf{y}] = \sum_{g \in G} \langle \rho(g)\mathbf{x} | \rho(g)\mathbf{y} \rangle \ .$$

It is easy to see that $[-,-]$ is Hermitian because $\langle -|- \rangle$ is defined so.

The Hermitian inner product defined so is G-invariant because for any $h \in G$ and $\mathbf{x}, \mathbf{y} \in V$,

$$\begin{aligned} [\rho(h)\mathbf{x}, \rho(h)\mathbf{y}] &= \sum_{g \in G} \langle \rho(h)\rho(g)\mathbf{x} | \rho(h)\rho(g)\mathbf{y} \rangle \\ &= \sum_{g' \in G} \langle \rho(g')\mathbf{x} | \rho(g')\mathbf{y} \rangle \\ &= [\mathbf{x}, \mathbf{y}] \end{aligned}$$

by Cayley's theorem (Theorem 15.29). Then by Lemma 16.29, the representation (ρ, V) is unitary.□

Remark. The restriction of G being finite comes from the summation and the application of Cayley's theorem.

Corollary. Every finite subgroup G of $GL(n, \mathbb{C})$ is conjugate to a subgroup of $U(n)$.

Proof. If $G \leq GL(n, \mathbb{C})$, then the inclusion map $\rho : G \to GL(n, \mathbb{C})$ is a representation. By the unitary trick, ρ is a unitary representation. There is $\mathsf{P} \in GL(n, \mathbb{C})$ such that $\mathsf{P}\rho(g)\mathsf{P}^{-1} \in U(n)$ for all $g \in G$. □

Remark. All finite groups can be seen as groups of generalised rotations and reflections within a complex vector space, representable by unitary matrices. Here, generalised rotations and reflections mean group actions in the underlying vector space which preserve the length (metric induced by inner product, defined as $[\mathbf{x}, \mathbf{x}]^{1/2}$) and orthogonality.

Finally, combining all the statements we made before in this chapter, we can prove the Maschke's theorem.

Theorem 16.31 (Maschke's theorem). Let (ρ, V) be a representation of a finite group G in V equipped with an inner product $(-,-)$. For any G-invariant subspace $W < V$, its *orthogonal complement*, defined as

$$W^\perp \coloneqq \{\mathbf{v} \in V \mid (\mathbf{v}, \mathbf{w}) = 0 \text{ for all } \mathbf{w} \in W\},$$

is also an invariant subspace.

Proof. First, $W^\perp$ is indeed a subspace of V over $\mathbb{F}$. Suppose that $\mathbf{w} \in W$ and $\mathbf{w}_1, \mathbf{w}_2 \in W^\perp$. For $a, b \in \mathbb{F}$,

$$(\mathbf{w}, a\mathbf{w}_1 + b\mathbf{w}_2) = a(\mathbf{w}, \mathbf{w}_1) + b(\mathbf{w}, \mathbf{w}_2) = 0\,,$$

so $a\mathbf{w}_1 + b\mathbf{w}_2 \in W^\perp$.

Then, for $g \in G$, $\mathbf{w} \in W$ and $\mathbf{w}' \in W^\perp$, and by the unitarity of representations of a finite group (Theorem 16.30), we have

$$\begin{aligned}(\mathbf{w}, \rho(g)\mathbf{w}') &= (\rho(g)\rho(g^{-1})\mathbf{w}, \rho(g)\mathbf{w}') \\ &= (\rho(g^{-1})\mathbf{w}, \mathbf{w}') \\ &= 0\,,\end{aligned}$$

since $\rho(g^{-1})\mathbf{w} \in W$.

Hence, $\rho(g)\mathbf{w}' \in W^\perp$ for all $g \in G$, and so $W^\perp$ is a G-invariant subspace of V. □

Corollary. The following statements of the representation (ρ, V) of a finite group G are equivalent:

(i) (ρ, V) is reducible.

(ii) (ρ, V) is fully reducible (reduction to block diagonal form).

(iii) V has mutually orthogonal G-invariant subspaces.

16.5.2 Schur's Lemma

Lemma 16.32 (Schur's lemma). Suppose (ρ_V, V) and (ρ_W, W) are irreducible representations of G, where both V and W are vector spaces over an algebraically closed field $\mathbb{F}$ (e.g. $\mathbb{C}$). Then

(i) every element of $\mathrm{Hom}_G(V, W)$ is either 0 or an isomorphism;

(ii) the only nontrivial G-linear maps are the scalar multiples of the identity.

Proof (Non-examinable).

(i) Let ϕ be a non-zero G-linear map from V to W. Let $\mathbf{x} \in \ker\phi < V$, then for any $g \in G$,

$$\phi(\rho_V(g)\mathbf{x}) = \rho_W(g)\phi(\mathbf{x}) = \mathbf{0}\,.$$

Therefore, $\rho_V(g)\mathbf{x}$ is in the null space of ϕ, so $\ker\phi$ is a invariant subspace of $\rho_V(g)$. Since ρ_V is irreducible, the invariant subspace $\ker\phi$ must be $\{\mathbf{0}\}$, so ϕ is injective.

Similarly $0 \neq \mathrm{Im}\,\phi \leq W$ so $\mathrm{Im}\,\phi = W$ since ρ_W is irreducible. Thus ϕ is both injective and surjective, so an isomorphism.

(ii) Suppose $\phi \in \mathrm{Hom}_G(V, W)$ is non-zero. Then by (i) ϕ is an isomorphism so $V = W$. Since $\mathbb{F}$ is algebraically closed we may find λ an eigenvalue of ϕ. Then $\phi - \lambda\mathsf{I}_V$ has non-zero and Ginvariant kernel and so the map is zero. Thus $\phi = \lambda\mathsf{I}_V$ as required. □

Remark. This means that if the two irreducible representations ρ_α and ρ_β are inequivalent, then only a zero map $\mathsf{X} = \mathsf{0}$ satisfies

$$\mathsf{X}\rho_\alpha = \rho_\beta \mathsf{X}\,.$$

If the two irreducible representations are equivalent, then X can only be a multiple of the identity map ($\mathsf{X} = \lambda \mathsf{I}$).

Corollary. Every irreducible complex representation of an Abelian group G is one-dimensional.

Proof. Let (ρ, V) be a complex irreducible representation of G. For each $g \in G$, $\rho(G) \in \mathrm{Hom}_G(V, V)$ as G is Abelian. So by Schur's lemma, $\rho(g) = \lambda_g \mathsf{I}_V$ for some $\lambda_g \in \mathbb{C}$. This is irreducible only if V is one-dimensional. □

Theorem 16.33 (The great orthogonality theorem). Let G be a finite group with $|G|$ elements. Let $\{\rho_i(g)\}_{i=1}^m$ be the set of the inequivalent irreducible representations of G, with dimensions $\{n_i\}_{i=1}^m$. For any two of these representations, ρ_α and ρ_β, the matrix elements satisfy

$$\sum_{g\in G} (\rho_\alpha(g))_{ij} (\rho_\beta(g^{-1}))_{kl} = \frac{|G|}{n_\alpha} \delta_{\alpha\beta}\delta_{il}\delta_{jk}\,.$$

Proof (Non-examinable). This follows from Schur's lemma. Let ρ_α and ρ_β be two irreducible unitary representations with dimensions n_α and n_β. Consider an arbitrary $n_\alpha \times n_\beta$ matrix X, and let

$$\mathsf{A} = \sum_{g\in G} \rho_\alpha(g)\mathsf{X}\rho_\beta(g^{-1}) = \sum_{g\in G} \rho_\alpha(g)\mathsf{X}\rho_\beta^\dagger(g)\,.$$

Now multiply by $\rho_\alpha(h)$ on the left of A for any $h \in G$ to get

$$\begin{aligned}
\rho_\alpha(h)\mathsf{A} &= \sum_{g\in G} \rho_\alpha(h)\rho_\alpha(g)\mathsf{X}\rho_\beta(g^{-1}) \\
&= \sum_{g\in G} \rho_\alpha(h)\rho_\alpha(g)\mathsf{X}\rho_\beta(g^{-1})\rho_\beta(h^{-1})\rho_\beta(h) \\
&= \sum_{g\in G} \rho_\alpha(hg)\mathsf{X}\rho_\beta^\dagger(g)\rho_\beta^\dagger(h)\rho_\beta(h) \\
&= \sum_{g\in G} \rho_\alpha(hg)\mathsf{X}(\rho_\beta(h)\rho_\beta(g))^\dagger\rho_\beta(h) \\
&= \left(\sum_{g\in G} \rho_\alpha(hg)\mathsf{X}\rho_\beta((hg)^{-1})\right)\rho_\beta(h) \\
&= \mathsf{A}\rho_\beta(h)
\end{aligned}$$

by the rearrangement theorem (Theorem 15.15).

Consider the two cases:

- If ρ_α and ρ_β are inequivalent, then by Schur's lemma, $\mathsf{A} = \mathsf{0}$. Therefore,

$$\sum_{g\in G} \rho_\alpha(g)\mathsf{X}\rho_\beta^\dagger(g) = \mathsf{0}\,.$$

  Choose X to be the matrix such that $X_{mn} = \delta_{mj}\delta_{nk}$, then the expression reduces to

$$\sum_{g\in G} (\rho_\alpha(g))_{ij} (\rho_\beta(g^{-1}))_{kl} = 0\,.$$

- If ρ_α and ρ_β are equivalent, then by Schur's lemma, $\mathsf{A} = \lambda \mathsf{I}_{n\times n}$. Taking traces on both sides gives

$$\begin{aligned} n_\alpha \lambda &= \sum_{g\in G} \operatorname{tr}\big(\rho(g)\mathsf{X}\rho(g^{-1})\big) \\ &= \sum_{g\in G} \operatorname{tr}\big(\mathsf{X}\rho(g)\rho(g^{-1})\big) \\ &= |G| \operatorname{tr}\mathsf{X}\,. \end{aligned}$$

  Therefore, we have

$$\sum_{g\in G} \rho(g)\mathsf{X}\rho(g^{-1}) = \frac{|G|}{n_\alpha}\mathsf{I}_{n\times n} \operatorname{tr}\mathsf{X}\,.$$

  Take $X_{mn} = \delta_{mj}\delta_{nk}$ and write in index notation, we get

$$\sum_{g\in G} (\rho(g))_{ij}(\rho(g^{-1}))_{kl} = \frac{|G|}{n_\alpha}\delta_{il}\delta_{jk}\,.$$

Combining the two cases gives the theorem. □

16.6 Unfaithful Representations

Although our focus will be mainly drawn on faithful representations, we still need to consider unfaithful representations that are merely homomorphic (and not isomorphic) to G.

Proposition 16.34. The mapping from any group G to the trivial group $C_1 = \{I\}$:

$$\begin{aligned} \phi : G &\to C_1\,, \\ g &\mapsto I \quad \forall\, g \in G \end{aligned}$$

is a homomorphism.

Proof. Trivially,

$$\phi(g_2 g_1) = \phi(g_2)\phi(g_1) = I \quad \forall\, g_2, g_1 \in G\,.$$

□

The kernel of this homomorphism is the set of all the group elements G, and all the elements in G can have the same trivial representation

$$\phi(g) = \big(1\big)\ .$$

Remark. In the trivial representation, the representations of the group elements can also be the identity matrix of any dimension.

Example. There are also some non-trivial unfaithful mapping of D_4. Consider the mapping $D_4 \to C_2$, where the faithful representation of C_2 is taken to be $\big(1\big)$ and $\big(-1\big)$. We have the unfaithful representations of D_4:

$$\{I, R, R^2, R^3, m_1, m_2, m_3, m_4\} \mapsto \{(1),(1),(1),(1),(-1),(-1),(-1),(-1)\}$$

$$\{I, R, R^2, R^3, m_1, m_2, m_3, m_4\} \mapsto \{(1),(-1),(1),(-1),(1),(1),(-1),(-1)\}$$

$$\{I, R, R^2, R^3, m_1, m_2, m_3, m_4\} \mapsto \{(1),(-1),(1),(-1),(-1),(-1),(1),(1)\}\,.$$

The kernels of the three representations are the normal subgroups of D_4: C_4 and two Vierergruppen respectively.

Remark. Examining the 1-dimensional representations of a group is instructive since

- All 1D representations are irreducible.
- Any two distinct 1-dimensional representations are inequivalent.

  Proof. We always have $\mathsf{D} = \mathsf{SDS}^{-1}$ if all these matrices are 1D, since 1D matrices are commutative in multiplication. □
- The trace of the representation is the entry in the 1×1 matrix.

16.7 Character Table

Definition 16.35. A *character table* is a list of the characters of all the inequivalent irreducible representations.

	I	g_1	$\cdots$	g_i	$\cdots$	g_r
χ_1	1	1	$\cdots$	1	$\cdots$	1
$\vdots$				$\vdots$		
χ_j	$\cdots$	$\cdots$	$\cdots$	$\chi_j(g_i)$	$\cdots$	$\cdots$
$\vdots$				$\vdots$		
χ_r				$\vdots$		

Note that it is common to put the trivial representation in the first row.

Before we continue stating the properties of the character table, we need some definitions and lemmas.

Definition 16.36. We say a function $f : G \to \mathbb{F}$ is a *class function* if $f(hgh^{-1}) = f(g)$ for all $g, h \in G$, i.e. the function outputs the same result for the group elements in the same conjugacy class.

We will write $\mathcal{C}_G$ for the $\mathbb{F}$-vector space of class functions on G.

Remark. The character, $\chi : G \to \mathbb{F}$ is a class function by Proposition 16.15.

We can make $\mathcal{C}_G$, the space of class functions into a Hermitian inner product space.

Definition 16.37. Define the *Hermitian inner product in* $\mathcal{C}_G$ to be

$$\langle f_1 | f_2 \rangle_G \coloneqq \frac{1}{|G|} \sum_{g \in G} f_1(g) f_2(g)^* .$$

Remark. Let $|c_g|$ be the number of elements in the same conjugacy class of $g \in G$. Select one element from each conjugacy class to form the set $\{g_1, g_2, \ldots, g_r\}$, then the inner product simplifies to

$$\langle f_1 | f_2 \rangle_G \coloneqq \sum_{i=1}^{r} \frac{|c_g|}{|G|} f_1(g_i) f_2(g_i)^* .$$

Theorem 16.38 (Character orthogonality). If χ_α, χ_β are the characters of the irreducible representations $(\rho_\alpha, V_\alpha), (\rho_\beta, V_\beta)$ of G, then

$$\langle \chi_\alpha | \chi_\beta \rangle = \begin{cases} 1 & \text{if } \rho_\alpha \cong \rho_\beta \\ 0 & \text{if } \chi_\alpha \text{ and } \chi_\beta \text{ are inequivalent} \end{cases}.$$

Proof. Set $i = j$ and $k = l$ and sum over repeated indices in the great orthogonality theorem (Theorem 16.33) using

$$\delta_{jl}\delta_{jl} = \delta_{jj} = n_\alpha \,,$$

we find

$$\sum_{g \in G} \chi_\alpha(g)\chi_\beta(g)^* = |G|\delta_{\alpha\beta} \,.$$

□

Finally, we can state the following properties of the character table.

Theorem 16.39. Let $\{\chi_i\}_{i=1}^m$ be the characters of the inequivalent irreducible representations, $\{(\rho_i, V_i)\}_{i=1}^m$, of the group G with order n.

(i) The sum of characters for all the group elements of a representation, excluding the trivial representation, is zero.

$$\sum_{g \in G} \chi_i(g) = \begin{cases} n & \text{for } i = 1 \\ 0 & \text{for } i \neq 1 \end{cases}.$$

(ii) Characters are the same within the same conjugacy class.

$$\chi_i(g) = \chi(hgh^{-1}) \text{ for } g, h \in G \,.$$

(iii) The rows of the character table form orthogonal vectors. For $i, j \in \{1, 2, ..., m\}$,

$$\sum_{g \in G} \chi_i(g)^* \chi_j(g) = \begin{cases} 0 & \text{if } i \neq j \\ n & \text{if } i = j \end{cases}.$$

(iv) The number of inequivalent irreducible representations is equal to the number of conjugacy classes.

(v) The columns of the character table of different conjugacy classes form orthogonal vectors. Let $|c_g|$ denote the number of elements conjugate with $g \in G$. For $g, h \in G$,

$$\sum_{i=1}^m \chi_i(g)^* \chi_i(h) = \begin{cases} \frac{n}{|c_g|} & \text{if } g \sim h \\ 0 & \text{otherwise} \end{cases}.$$

(vi) The sum of the squares of the dimensions of all the representations is equal to the order of the group.

$$\sum_{i=1}^m (\dim_{\mathbb{F}} V_i)^2 = n \,.$$

(vii) The dimension of each representation divides n.

$$n = k \dim_{\mathbb{F}} V_i \,, \; k \in \mathbb{N} \,.$$

We can prove some parts of the above theorem.

Proof.

(i) This follows from the orthogonality of the characters (Theorem 16.38). Let ρ_β be the trivial irreducible representation such that $\chi_\beta(g) = 1$ for all $g \in G$, then the equation above simplifies to

$$\sum_{g\in G} \chi_\alpha(g) = |G|\delta_{\alpha\beta} = \begin{cases} |G| & \text{if } \rho^{(\alpha)} \text{ is trivial} \\ 0 & \text{otherwise}\,. \end{cases}$$

(ii) A direct result of Proposition 16.15.

(iii) Equivalent to Theorem 16.38.

(iv) A full proof is beyond the scope of the course. Here is the outline. (iii) already shows that $\{\chi_i\}$ is an orthonormal set in $\mathcal{C}_G$. By Schur's lemma, we can also show that χ_i spans $\mathcal{C}_G$. Therefore, $\{\chi_i\}$ forms an orthonormal basis for the space of class functions $\mathcal{C}_G$. A direct corollary is this theorem.

(v) Let X be the character table (keeping only one column in a conjugacy class) thought of as a matrix: $X_{ij} = \chi_i(g_j)$ and let D be the diagonal matrix whose diagonal entries are $|G|/|c_G|$. Orthogonality of the characters (Theorem 16.38) tells us that

$$\sum_k \frac{|c_g|}{|G|} X^*_{ik} X_{jk} = \delta_{ij}\,,$$

$$\mathsf{X}^*\mathsf{D}^{-1}\mathsf{X}^T = \mathsf{I}\,.$$

Since X is square (by (iv)), we can write the above equation as

$$\mathsf{D}^{-1}\mathsf{X}^\dagger = \mathsf{X}^{-1} \implies \mathsf{X}^\dagger\mathsf{X} = \mathsf{D}\,.$$

This is equivalent to

$$\sum_k \chi_k(g_i)^* \chi_k(g_j) = \delta_{ij} \frac{|G|}{|c_{g_i}|}\,.$$

(vi) Let n_i be the dimension of the irreducible representation (ρ_i, V_i). Let ρ_{reg} be the regular representation of G with character χ_{reg}. Then $\chi_{\text{reg}}(g) = |G|$ when $g = I$ and $\chi_{\text{reg}}(g) = 0$ otherwise. Therefore we have

$$\langle \chi_{\text{reg}} | \chi_i \rangle = n_i$$

for all i. Since

$$\chi_{\text{reg}} = \sum_{i=1}^m n_i \chi_i\,,$$

(see decomposition formula later), we have, for the identity element,

$$\chi_{\text{reg}}(I) = \sum_{i=1}^m n_i^2\,,$$

$$\sum_{i=1}^m n_i^2 = n\,.$$

Alternatively, take the inner product of the first column with itself. By (v),

$$\sum_{i=1}^m (\dim V_i)^2 = \sum_{i=1}^m \chi_i(I)^2 = |G|\,.$$

(vii) We will not prove this. □

Corollary. All groups of orders less than or equal to 4 are Abelian.

Proof. From (iii), every nontrivial irreducible representation of G has a dimension less than or equal to $\sqrt{|G|-1}$, so $|G| < 5$ means that every irreducible representation has a dimension of 1. If all representations of G commute, G can only be Abelian. □

Remark. The only group of order 5 is C_5 up to isomorphism, which is also Abelian. Therefore, the smallest non-Abelian group is actually D_3 (or its isomorph S_3) of order 6.

Example. Character table of D_4.

For D_4, we have 5 inequivalent irreducible representations: a trivial representation, 3 one-dimensional representations and a two-dimensional representation. The character table is given as follows.

D_4	I	R^2	R	R^3	m_1	m_2	m_3	m_4
χ_1	1	1	1	1	1	1	1	1
χ_2	1	1	1	1	−1	−1	−1	−1
χ_3	1	1	−1	−1	1	1	−1	−1
χ_4	1	1	−1	−1	−1	−1	1	1
χ_5	2	2	0	0	0	0	0	0

All six statements of Theorem 16.39 can be checked.

Example. Cyclic group of prime order, C_p.

For the cyclic group of prime order p, we have p one-dimensional inequivalent representations. Writing the p^{th} root of unity

$$\omega = e^{\frac{2\pi i}{p}},$$

we have the character table

C_p	I	g	$\dots$	g^{p-1}
χ_1	1	1	$\dots$	1
χ_2	1	ω	$\dots$	ω^{p-1}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
χ_q	1	ω^{q-1}	$\dots$	$\omega^{(p-1)(q-1)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
χ_p	1	ω^{p-1}	$\dots$	$\omega^{(p-1)^2}$

16.8 Decomposition of a Reducible Representation

Definition 16.40. The *tensor product* of two matrices

$$\mathsf{A} = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1q} \\ A_{21} & A_{22} & \cdots & A_{2q} \\ \cdots & \cdots & \ddots & \vdots \\ A_{p1} & A_{p2} & \cdots & A_{pq} \end{pmatrix}, \quad \mathsf{B} = \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1n} \\ B_{21} & B_{22} & \cdots & B_{2n} \\ \cdots & \cdots & \ddots & \vdots \\ B_{m1} & B_{m2} & \cdots & B_{mn} \end{pmatrix}$$

is defined as the $(pm) \times (qn)$ matrix

$$\mathsf{A} \otimes \mathsf{B} \coloneqq \begin{pmatrix} A_{11}\mathsf{B} & A_{12}\mathsf{B} & \cdots & A_{1q}\mathsf{B} \\ A_{21}\mathsf{B} & A_{22}\mathsf{B} & \cdots & A_{2q}\mathsf{B} \\ \cdots & \cdots & \ddots & \vdots \\ A_{p1}\mathsf{B} & A_{p2}\mathsf{B} & \cdots & A_{pq}\mathsf{B} \end{pmatrix},$$

where each $A_{ij}\mathsf{B}$ is a $m \times n$ block given by

$$A_{ij}\mathsf{B} \coloneqq \begin{pmatrix} A_{ij}B_{11} & A_{ij}B_{12} & \cdots & A_{ij}B_{1n} \\ A_{ij}B_{21} & A_{ij}B_{22} & \cdots & A_{ij}B_{2n} \\ \cdots & \cdots & \ddots & \vdots \\ A_{ij}B_{m1} & A_{ij}B_{m2} & \cdots & A_{ij}B_{mn} \end{pmatrix}.$$

Denote the irreducible representations of G by $\rho_1, \ldots, \rho_k$. Let P be a reducible representation of G, then P can be decomposed into irreducible representations as the direct sum

$$P = m_1\rho_1 \oplus m_2\rho_2 \oplus \cdots \oplus m_k\rho_k = \bigoplus_{i=1}^{k} m_i\rho_i\,,$$

where $m_i \in \mathbb{Z}_{>0}$ are the *multiplicities*. Therefore, the matrix $P(g)$ can be written, after a similarity transformation by a g-independent matrix S, as

$$\begin{aligned} \mathsf{S}P(g)\mathsf{S}^{-1} &= \mathsf{I}_{m_1} \otimes \rho_1(g) \oplus \mathsf{I}_{m_2} \otimes \rho_2(g) \oplus \cdots \oplus \mathsf{I}_{m_k} \otimes \rho_k(g) \\ &= \bigoplus_{i=1}^{k} \left(\mathsf{I}m_i \otimes \rho_i(g)\right), \end{aligned}$$

where I_{m_i} is the $m_i \times m_i$ identity matrix. This produces m_i copies of each $\rho_i(g)$ along the diagonal after similarity transformations. Since the character is invariant under similarity transformation, by taking the trace of the above decomposition, we have

$$\chi_P(g) = \sum_{i=1}^{k} m_i\chi_i(g)\,. \tag{$\dagger$}$$

We can use this to find $m_1, \ldots, m_k$ without finding the appropriate similarity transformations.

Theorem 16.41. The multiplicity m_i of the decomposition of P into irreducible representations ρ_i with characters χ_i is given by

$$m_i = \frac{1}{|G|} \sum_{c_{g_j}} \left|c_{g_j}\right| \chi_P(g_j)\chi_i(g_j)^*\,.$$

Proof. Sum over all the g for equation ($\dagger$), multiply by $\chi_j(g^{-1})$, then use the orthogonality of the characters (Theorem 16.38).

$$\begin{aligned} \sum_{c_{g_j}} \left|c_{g_j}\right| \chi_P(g_j)\chi_i(g_j)^* &= \sum_{g} \chi_P(g)\chi_j(g^{-1}) \\ &= \sum_{g} \sum_{i=1}^{k} m_i\chi_i(g)\chi_j(g^{-1}) \\ &= \sum_{i=1}^{k} m_i|G| \left\langle \chi_i | \chi_j \right\rangle \\ &= m_i|G|\,. \end{aligned}$$

□

17 Small Oscillations

17.1 Pendulum

17.1.1 Simple Pendulum

Consider a simple pendulum of a string of length l with a mass m hanging at the end. The string makes an angle θ with the vertical. The equation of motion, obtained from Newton's second law or conservation of energy, is

$$ml\ddot{\theta} = -mg\sin\theta\,.$$

For small θ, by the approximation $\sin\theta \approx \theta$, we have

$$\ddot{\theta} = -\frac{g}{l}\theta\,,$$

a simple harmonic oscillator equation. Defining $\omega^2 = g/l$, we have the solution

$$\theta = C\sin\omega t + D\cos\omega t\,,$$

or alternatively

$$\theta = A\sin\omega(t - t_0)\,,$$

where A is the amplitude and $\theta_0 = \omega t_0$ is the phase.

17.1.2 Coupled Pendula

Consider two pendula of length l and mass m connected by a massless spring with spring constant k. The two pendula make angles θ_1 and θ_2 with vertical respectively, and they are of distance b apart in equilibrium.

For small oscillations, the extension or compression of the spring away from its equilibrium length b is approximated to be

$$x = l(\theta_2 - \theta_1)$$

for sufficient accuracy. This can be shown by setting the origin O at the pivot of the first pendulum. Then the positions of the two pendula are

$$\mathbf{x}_1 = \begin{pmatrix} l\sin\theta_1 \\ -l\cos\theta_1 \end{pmatrix} \quad \text{and} \quad \mathbf{x}_2 = \begin{pmatrix} b + l\sin\theta_2 \\ -l\cos\theta_2 \end{pmatrix}.$$

Using the small angle approximations

$$\sin\theta \approx \theta\,,\ \cos\theta \approx 1 - \frac{\theta^2}{2}\,,$$

we can find

$$\begin{aligned} |\mathbf{x}_2 - \mathbf{x}_1| - b &= \sqrt{(b + l\sin\theta_2 - l\sin\theta_1)^2 + (l\cos\theta_1 - l\cos\theta_2)^2} - b \\ &\approx \sqrt{(b + l\theta_2 - l\theta_1)^2 + O(\theta^4)} - b \\ &\approx l(\theta_2 - \theta_1)\,. \end{aligned}$$

The equations of motion, from Newton's second law, are

$$\begin{cases} ml\ddot{\theta_1} = -mg\theta_1 + kl(\theta_2 - \theta_1) \\ ml\ddot{\theta_2} = -mg\theta_2 + kl(\theta_1 - \theta_2)\,, \end{cases}$$

where the small angle approximation has been used for the gravitational terms as well.

In general, solutions for θ_1 and θ_2 are complicated. However, special solutions to the system can be found.

Definition 17.1. There are periodic solutions to the coupled equations, described by a single frequency, called the *harmonic solutions*. These are called the *normal modes* of the oscillation. The frequencies corresponding to the normal modes of the oscillation are called the *normal frequencies*.

In the case of the coupled pendula, there are two normal modes.

(i) *In phase solution,* $\theta_1 = \theta_2$. In this case, the spring exerts no force since its length does not change. Each equation reduces to the same form:

$$\begin{cases} ml\ddot{\theta}_1 = -mg\theta_1 \\ ml\ddot{\theta}_2 = -mg\theta_2 \, , \end{cases}$$

as if the two pendula were uncoupled. Therefore, we have

$$\theta_1 = \theta_2 = A\sin\omega(t - t_0) \quad \text{with} \quad \omega^2 = \frac{g}{l} \, ,$$

where A and t_0 are arbitrary.

(ii) 180° *out of phase solution,* $\theta_1 = -\theta_2$. Each equation is again reduced to the same form:

$$\begin{cases} ml\ddot{\theta}_1 = -(mg + 2kl)\theta_1 \\ ml\ddot{\theta}_2 = -(mg + 2kl)\theta_2 \, , \end{cases}$$

with solutions

$$\theta_1 = -\theta_2 = B\sin\Omega(t - t_1) \quad \text{with} \quad \Omega^2 = \frac{g}{l} + \frac{2k}{m} \, .$$

Here B and t_1 are also arbitrary.

Each of the two special cases describes a harmonic motion with a single, pure frequency:

(i)

$$\omega = \sqrt{\frac{g}{l}} \, .$$

(ii)

$$\Omega = \sqrt{\frac{g}{l} + \frac{2k}{m}} \, .$$

Theorem 17.2. The general solution of a coupled oscillation can be written as a linear combination of the normal modes.

For example, in this case, the solutions can be written as

$$\begin{cases} \theta_1 = A\sin\omega(t - t_0) + B\sin\Omega(t - t_1) \\ \theta_2 = A\sin\omega(t - t_0) - B\sin\Omega(t - t_1) \, . \end{cases}$$

This works because the system of differential equations is linear, so the principle of superposition applies, and we have four arbitrary constants that we can specify.

Remark. If Ω/ω is an irrational number, then the general solution is not periodic. Special linear combinations, like $\theta_1 \pm \theta_2$, are periodic. These are called the *normal coordinates* for the system.

17.1.3 Lagrangian Dynamics and Coupled Pendula

Modify the double pendula problem slightly so that $m_1 \neq m_2$. We will use Lagrange's equations to work out the general theory. We take our generalised coordinates to be θ_1 and θ_2. The Lagrangian of the system is

$$L = T - V\,,$$

which satisfies Lagrange's equations (Theorem 9.14)

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{\theta}_i}\right) - \frac{\partial L}{\partial \theta_i} = 0$$

for $i \in \{1, 2\}$.

For the coupled pendula, the kinetic energy is

$$\begin{aligned} T &= \frac{1}{2}m_1 v_1^2 + \frac{1}{2}m_2 v_2^2 \\ &= \frac{1}{2}m_1 l^2 \dot{\theta}_1^2 + \frac{1}{2}m_2 l^2 \dot{\theta}_2^2\,. \end{aligned}$$

For small oscillation angle θ_i, the gravitational potential energy can be approximated to be

$$V_{gi} = m_i g l(1 - \cos\theta_i) \approx \frac{1}{2}m_i g l \theta_i^2\,,$$

and the elastic potential energy stored in the spring is approximately

$$V_e = \frac{1}{2}kl^2(\theta_2 - \theta_1)^2\,.$$

Therefore, we have

$$V = \frac{1}{2}m_1 g l \theta_1^2 + \frac{1}{2}m_2 g l \theta_2^2 + \frac{1}{2}kl^2(\theta_2 - \theta_1)^2$$

and

$$L = \frac{1}{2}m_1 l^2 \dot{\theta}_1^2 + \frac{1}{2}m_2 l^2 \dot{\theta}_2^2 - \frac{1}{2}m_1 g l \theta_1^2 - \frac{1}{2}m_2 g l \theta_2^2 - \frac{1}{2}kl^2(\theta_2 - \theta_1)^2\,.$$

Substitute the Lagrangian into Lagrange's equations, we have

$$\begin{cases} m_1 l\ddot{\theta}_1 = -m_1 g\theta_1 - kl(\theta_1 - \theta_2) \\ m_2 l\ddot{\theta}_2 = -m_2 g\theta_2 + kl(\theta_1 - \theta_2)\,, \end{cases}$$

or, in matrix form,

$$\begin{pmatrix} m_1 l & 0 \\ 0 & m_2 l \end{pmatrix}\begin{pmatrix} \ddot{\theta}_1 \\ \ddot{\theta}_2 \end{pmatrix} = \begin{pmatrix} -m_1 g - kl & kl \\ kl & -m_2 g - kl \end{pmatrix}\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}\,. \tag{$\dagger$}$$

This matrix equation can be rewritten as

$$\mathsf{T}\ddot{\mathsf{q}} = -\mathsf{V}\mathsf{q}\,,$$

where

$$\mathsf{q} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}\,,\ \mathsf{T} = \begin{pmatrix} m_1 l^2 & 0 \\ 0 & m_2 l^2 \end{pmatrix} \text{ and } \mathsf{V} = \begin{pmatrix} m_1 g l + kl^2 & -kl^2 \\ -kl^2 & m_2 g l + kl^2 \end{pmatrix}\,.$$

Note that we have multiplied both sides by l so that we can write the Lagrangian as a matrix equation with the degrees of freedom expressed as a column vector:

$$L = \frac{1}{2}T_{ij}\dot{\theta}_i\dot{\theta}_j - \frac{1}{2}V_{ij}\theta_i\theta_j = \frac{1}{2}\dot{\mathsf{q}}^{\mathrm{T}}\mathsf{T}\dot{\mathsf{q}} - \frac{1}{2}\mathsf{q}^{\mathrm{T}}\mathsf{V}\mathsf{q}\,.$$

We want to find the normal modes of this system, so we try solutions with a single, pure frequency of the form

$$\begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix} = \begin{pmatrix}a_1\\ a_2\end{pmatrix} e^{i\omega t}\,,$$

with constants a_1, a_2. Substitute this into the equation (†), we have

$$-\omega^2 \begin{pmatrix}m_1 l & 0\\ 0 & m_2 l\end{pmatrix}\begin{pmatrix}a_1\\ a_2\end{pmatrix} = \begin{pmatrix}-m_1 g - kl & kl\\ kl & -m_2 g - kl\end{pmatrix}\begin{pmatrix}a_1\\ a_2\end{pmatrix}\,,$$

or equivalently,

$$\begin{pmatrix}\omega^2 m_1 l - m_1 g - kl & kl\\ kl & \omega^2 m_2 l - m_2 g - kl\end{pmatrix}\begin{pmatrix}a_1\\ a_2\end{pmatrix} = \mathbf{0}\,.$$

This equation has non-trivial solutions only if the matrix determinant vanishes, i.e.

$$\begin{aligned}&\begin{vmatrix}\omega^2 m_1 l - m_1 g - kl & kl\\ kl & \omega^2 m_2 l - m_2 g - kl\end{vmatrix}\\ &= (\omega^2 l - g)[m_1 m_2(\omega^2 l - g) - (m_1 + m_2)kl]\\ &= 0\,,\end{aligned}$$

which has solutions

$$\omega_1^2 = \frac{g}{l} \text{ and } \omega_2^2 = \frac{g}{l} + \frac{k(m_1+m_2)}{m_1 m_2}\,.$$

Corresponding to each of the solutions, there is a pair of amplitudes

$$\begin{pmatrix}a_1^{(1)}\\ a_2^{(1)}\end{pmatrix} \text{ and } \begin{pmatrix}a_1^{(2)}\\ a_2^{(2)}\end{pmatrix}\,,$$

which can be determined up to an overall normalisation. For ω_1,

$$\begin{pmatrix}-kl & kl\\ kl & -kl\end{pmatrix}\begin{pmatrix}a_1^{(1)}\\ a_2^{(1)}\end{pmatrix} = \mathbf{0}\,,$$

which solves to be $a_1^{(1)} = a_2^{(1)}$. This corresponds to $\theta_1 = \theta_2$: the masses are swinging in phase. For ω_2,

$$\begin{pmatrix}\frac{m_1}{m_2}kl & kl\\ kl & \frac{m_2}{m_1}kl\end{pmatrix}\begin{pmatrix}a_1^{(2)}\\ a_2^{(2)}\end{pmatrix} = \mathbf{0}\,,$$

which implies $m_1 a_1^{(2)} = -m_2 a_2^{(2)}$. Here the two masses are swinging 180° out of phase, with scaled amplitudes.

The normal coordinates can be found by taking linear combinations of the rows of the equation (†) to isolate one of the two normal modes. We can get

$$m_1\ddot{\theta}_1 + m_2\ddot{\theta}_2 = -\frac{g}{l}(m_1\theta_1 + m_2\theta_2)\,,$$

with oscillating frequency $\omega_1 = \frac{g}{l}$, and

$$\ddot{\theta}_1 - \ddot{\theta}_2 = -\left[\frac{g}{l} + k\left(\frac{1}{m_1} + \frac{1}{m_2}\right)\right](\theta_1 - \theta_2)\,,$$

with oscillating frequency $\omega_2 = \frac{g}{l} + k\left(\frac{1}{m_1} + \frac{1}{m_2}\right)$. Therefore, the normal coordinates are $m_1\theta_1 + m_2\theta_2$ and $\theta_1 - \theta_2$, corresponding to the normal frequencies ω_1 and ω_2 respectively.

17.2 General Theory of Small Oscillations

Consider a system with N degrees of freedom, represented by N generalised coordinates

$$\{q_i\} = \{q_1, q_2, \ldots, q_N\}\,.$$

We can represent this as an N-component vector $\mathbf{q}$.

Let $V(\mathbf{q})$ be the potential energy of the system, and assume, without loss of generality, that the coordinates have been chosen so that $\mathbf{q} = \mathbf{0}$ is a position of stable equilibrium. Expand $V(\mathbf{q})$ in Taylor series

$$\begin{aligned} V(\mathbf{q}) &= V(\mathbf{0}) + \frac{1}{2}\left.\frac{\partial^2 V}{\partial q_i \partial q_j}\right|_{\mathbf{q}=\mathbf{0}} q_i q_j + O(q_i^3) \\ &= V(\mathbf{0}) + \frac{1}{2} V_{ij} q_i q_j + \ldots \end{aligned}$$

We have therefore implicitly defined V_{ij} as the components of the constant, symmetric matrix V of second derivatives evaluated at $\mathbf{q} = 0$. Since this is an equilibrium point, the first derivatives of V vanish at this point and V is positive definite.

Similarly, let us assume we can write kinetic energy as

$$T = \frac{1}{2} T_{ij} \dot{q}_i \dot{q}_j\,,$$

where T_{ij} are components of a constant, symmetric matrix. We assume T is taken to be positive definite, that is all modes of oscillation contribute to kinetic energy at the lowest order.

Theorem 17.3. For a system with N degrees of freedom, represented by N generalised coordinates, $\{q_i\}$, if the potential energy and kinetic energy can be represented as

$$V(\mathbf{q}) = V(\mathbf{0}) + \frac{1}{2} V_{ij} q_i q_j\,,$$

$$T = \frac{1}{2} T_{ij} \dot{q}_i \dot{q}_j\,,$$

where T and V are symmetric, positive definite matrices, then the equations of motion are given by the N coupled second-order linear ODEs

$$T_{ij} \ddot{q}_j + V_{ij} q_j = 0\,. \tag{$\dagger$}$$

Proof. The Lagrangian of the system is

$$L = T - V = \frac{1}{2} T_{ij} \dot{q}_i \dot{q}_j - \frac{1}{2} V_{ij} q_i q_j\,.$$

Using Lagrange's equations (Theorem 9.14), the equations of motion are obtained. □

17.2.1 Normal Modes

Normal modes are special solutions of the equations of motion (†) which oscillate with a single, pure frequency. To find them, we assume solutions of the form

$$q_i(t) = Q_i \sin \omega (t - t_0)$$

or similarly with complex exponentials. Note that Q_i have to be independent of t. Substitute this into (†) yields

$$-\omega^2 T_{ij} Q_j + V_{ij} Q_j = 0\,,$$

or in matrix notation,

$$(-\omega^2 \mathsf{T} + \mathsf{V})\mathsf{Q} = \mathbf{0}\,, \tag{††}$$

where

$$\mathsf{Q} = \begin{pmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_N \end{pmatrix}.$$

Since we are interested in non-trivial solutions where Q is non-zero, we must have

$$\det\bigl(-\omega^2 \mathsf{T} + \mathsf{V}\bigr) = 0\,.$$

The left-hand side is a polynomial of degree N in ω^2, and the solutions are the squares of the normal frequencies. This is referred to as the *characteristic equation*.

Remark. We might find equation (††) resembles the eigenvector equation, so we can see ω^2 as some sort of generalised eigenvalue, with $\mathbf{Q}$ being the eigenvector.

Let the normal frequencies, defined as the positive square roots of the generalised eigenvalues, be $\{\omega_i\}$ and let the corresponding generalised eigenvectors be $\{\mathbf{Q}^{(i)}\}$, $i = 1, \dots, N$. The general solution is

$$\mathbf{q}(t) = \sum_{m=1}^{N} A^{(m)} \mathbf{Q}^{(m)} \sin\omega_m (t - t_0^{(m)})\,, \tag{∗}$$

where the constant $A^{(m)}$ is the amplitude of the m^{th} normal mode and $\omega_m t_0^{(m)}$ is the phase.

In the case where $\omega_m = 0$, the normal mode is called a *zero mode*. The corresponding term in the solution is obtained by taking the limit

$$\lim_{\omega_m \to 0} A^{(m)} \mathbf{Q}^{(m)} \sin\omega_m (t - t_0^{(m)}) = B^{(m)} \mathbf{Q}^{(\mathbf{m})} (t - t_0^{(m)})\,,$$

where we defined a new constant $B^{(m)} = \omega_m A^{(m)}$.

Remark. The constant $B^{(m)}$ is determined by the initial conditions, corresponding to an $A^{(m)}$ that diverges in the $\omega_m \to 0$ limit.

17.2.2 Orthogonality

Proposition 17.4. Two eigenvectors $\mathbf{Q}^{(m)}$ and $\mathbf{Q}^{(n)}$ with unequal normal frequencies are orthogonal.

Proof. From (††), we have

$$(-\omega_m^2 \mathsf{T} + \mathsf{V})\mathbf{Q}^{(m)} = \mathbf{0}\,,$$
$$(-\omega_n^2 \mathsf{T} + \mathsf{V})\mathbf{Q}^{(n)} = \mathbf{0}\,.$$

Left multiply the first and second equations by the row vectors $-(\mathbf{Q}^{(n)})^{\mathrm{T}}$ and $-(\mathbf{Q}^{(m)})^{\mathrm{T}}$ respectively. Note that T and V are symmetric matrices, so

$$Q_i^{(n)} V_{ij} Q_j^{(m)} = Q_i^{(m)} V_{ij} Q_j^{(n)}\,,$$

and therefore we have

$$(\omega_m^2 - \omega_n^2)(\mathbf{Q}^{(m)})^{\mathrm{T}} \mathsf{T} \mathbf{Q}^{(n)} = 0\,.$$

By supposition, $\omega_m \neq \omega_n$, so we must have $(\mathbf{Q}^{(m)})^{\mathrm{T}} \mathsf{T} \mathbf{Q}^{(n)} = 0$. Therefore, $\mathbf{Q}^{(m)}$ and $\mathbf{Q}^{(n)}$ are orthogonal with respect to T.

If $\omega_m = \omega_n$, then there exists linearly independent $\mathbf{Q}^{(m)}$ and $\mathbf{Q}^{(n)}$. An orthogonal pair can always be constructed using Gram-Schmidt orthogonalisation. □

Theorem 17.5. The generalised eigenvectors $\mathbf{Q}^{(m)}$ form an orthonormal set of generalised eigenvectors.

Proof. By scaling $\mathbf{Q}^{(m)}$ such that $(\mathbf{Q}^{(m)})^{\mathrm{T}}\mathsf{T}\mathbf{Q}^{(m)} = 1$, we have a set of generalised eigenvectors $\{\mathbf{Q}^{(m)}\}$ such that

$$(\mathbf{Q}^{(m)})^{\mathrm{T}}\mathsf{T}\mathbf{Q}^{(n)} = \delta_{mn}\,.$$

We then say that the $\{\mathbf{Q}^{(m)}\}$ form an orthonormal set of generalised eigenvectors. □

17.2.3 Normal Coordinates

Definition 17.6. *Normal coordinates* $\alpha^{(m)}(t)$ are linear combinations of the original generalised coordinates $q_j(t)$ which oscillates at a single, pure frequency ω_m and satisfy the simple harmonic equation.

Proposition 17.7. The normal coordinates are given by

$$\alpha^{(n)}(t) = Q_i^{(n)} T_{ij} q_j(t)\,.$$

Proof. Multiplying the j^{th} component of the general solution $(*)$ by $Q_i^{(n)}T_{ij}$, summing over j and using orthogonality, we find

$$\alpha^{(n)}(t) = Q_i^{(n)} T_{ij} q_j(t) = A^{(n)} \sin\omega_n(t - t_0^{(n)})\,.$$

□

Alternative proof. Since $\mathbf{Q}^{(m)}$ are linearly independent, we can write

$$q_i(t) = \sum_{m=1}^{N} \alpha^{(m)}(t) Q_i^{(m)}\,.$$

Substitute this into the equation (†) gives

$$\sum_{m=1}^{N} \left[\ddot{\alpha}^{(m)}(t) T_{ij} Q_j^{(m)} + \alpha^{(m)}(t) V_{ij} Q_j^{(m)}\right] = 0\,.$$

Using (††), we have

$$\sum_{m=1}^{N} \left[\ddot{\alpha}^{(m)}(t) + \omega_m^2 \alpha^{(m)}(t)\right] T_{ij} Q_j^{(m)} = 0\,.$$

Multiplying by $Q_i^{(n)}$ and using the orthogonality, we must have

$$\ddot{\alpha}^{(m)}(t) + \omega_m^2 \alpha^{(m)}(t) = 0$$

for each m. The solution to this differential equation is

$$\alpha^{(m)}(t) = A^{(m)} \sin\omega_m(t - t_0^{(m)})\,.$$

The $\alpha^{(m)}(t)$ are the normal coordinates. They can be written in terms of the $q_i(t)$ using orthonormality:

$$\begin{aligned} q_i(t) T_{ij} Q_j^{(n)} &= \sum_{m=1}^{N} \alpha^{(m)}(t) Q_i^{(m)} T_{ij} Q_j^{(n)} \\ &= \sum_{m=1}^{N} \alpha^{(m)}(t) \delta_{mn} \\ &= \alpha^{(n)}(t)\,. \end{aligned}$$

□

17.3 Examples

17.3.1 Vibrations in a CO_2 Molecule

Consider a carbon atom of mass M in the centre, connected with two oxygen atoms of mass m on opposite sides by springs of spring constant k. Treat the molecule as 1D and the motion as linear.

Denote the equilibrium positions of the two oxygen atoms by X_1 and X_3, and the position of the carbon atom as X_2. Denote displacements from these equilibrium positions by x_1, x_2 and x_3 respectively. Then the kinetic energy is

$$T = \frac{1}{2}m\dot{x}_1^2 + \frac{1}{2}M\dot{x}_2^2 + \frac{1}{2}m\dot{x}_3^2\,,$$

so

$$\mathsf{T} = \begin{pmatrix} m & 0 & 0 \\ 0 & M & 0 \\ 0 & 0 & m \end{pmatrix}.$$

The potential energy is

$$V = \frac{1}{2}k(x_2 - x_1)^2 + \frac{1}{2}k(x_3 - x_2)^2\,,$$

so

$$\mathsf{V} = k\begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix}.$$

By solving

$$\det\bigl(\mathsf{V} - \omega^2\mathsf{T}\bigr) = 0\,,$$

the normal frequencies are found to be

$$\omega_1^2 = 0\,,\ \omega_2^2 = \frac{k}{m}\,,\ \text{and}\ \omega_3^2 = \frac{k}{mM}(2m + M)\,.$$

We can then use

$$(\mathsf{V} - \omega_m^2\mathsf{T})\mathbf{Q}^{(m)} = \mathbf{0}$$

to find the three $\mathbf{Q}^{(m)}$, then use $(\mathbf{Q}^{(m)})^{\mathrm{T}}\mathsf{T}\mathbf{Q}^{(m)} = 1$ for normalisation. The results are as follows.

- $\omega_1^2 = 0$ implies

  $$\mathbf{Q}^{(1)} = \frac{1}{\sqrt{2m + M}}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

  This is a zero mode describing the rigid translation of the molecule.

- $\omega_2^2 = \frac{k}{m}$ implies

  $$\mathbf{Q}^{(2)} = \frac{1}{\sqrt{2m}}\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

  This corresponds to the reflection-symmetric mode where the oxygen atoms oscillate in opposing directions while the carbon atom is stationary.

- $\omega_3^2 = \frac{k}{mM}(2m + M)$ implies

  $$\mathbf{Q}^{(3)} = \left[2m\left(1 + \frac{2m}{M}\right)\right]^{-\frac{1}{2}}\begin{pmatrix} 1 \\ -\frac{2m}{M} \\ 1 \end{pmatrix}.$$

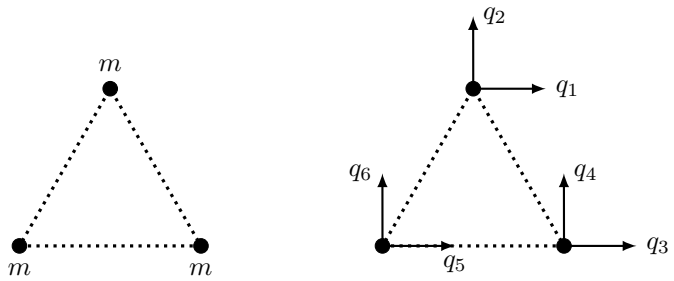
  Here the oxygen atoms oscillate in phase, and the carbon atom oscillates in the opposing direction.

The normal coordinates can be found from Proposition 17.7.

$$\begin{aligned}\alpha^{(1)}(t) &= \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} m & 0 & 0 \\ 0 & M & 0 \\ 0 & 0 & m \end{pmatrix} \frac{1}{\sqrt{2m+M}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\ &= \frac{mx_1 + Mx_2 + mx_3}{\sqrt{2m+M}} = B_1(t - t_0^{(1)})\,, \\ \alpha^{(2)}(t) &= \sqrt{\frac{m}{2}}(x_1 - x_3) = A_2 \sin\omega_2(t - t_0^{(2)})\,, \\ \alpha^{(3)}(t) &= \frac{x_1 - 2x_2 + x_3}{\sqrt{\frac{2}{m} + \frac{4}{M}}} = A_3 \sin\omega_3(t - t_0^{(3)})\,.\end{aligned}$$

17.3.2 Triangular Spring System

Consider three identical masses sitting on the vertices of an equilateral triangle connected by identical springs.



Taking the centre of mass as the origin, we have

$$\begin{aligned}\mathbf{x}_1 &= l\left(0, \frac{1}{\sqrt{3}}\right) + (q_1, q_2)\,, \\ \mathbf{x}_2 &= l\left(\frac{1}{2}, -\frac{1}{2\sqrt{3}}\right) + (q_3, q_4)\,, \\ \mathbf{x}_3 &= l\left(-\frac{1}{2}, -\frac{1}{2\sqrt{3}}\right) + (q_5, q_6)\,.\end{aligned}$$

The kinetic energy is

$$T = \frac{1}{2}m(\dot{\mathbf{x}}_1^2 + \dot{\mathbf{x}}_2^2 + \dot{\mathbf{x}}_3^2) = \frac{1}{2}m(\dot{q}_1^2 + \dot{q}_2^2 + \dot{q}_3^2 + \dot{q}_4^2 + \dot{q}_5^2 + \dot{q}_6^2)\,,$$

so the matrix T is diagonal:

$$T_{ij} = m\delta_{ij}\,.$$

The potential energy is

$$V = \frac{1}{2}k[(|\mathbf{x}_1 - \mathbf{x}_2| - l)^2 + (|\mathbf{x}_2 - \mathbf{x}_3| - l)^2 + (|\mathbf{x}_3 - \mathbf{x}_1| - l)^2]\,.$$

Assume small oscillations. We can write, say,

$$\begin{aligned}\mathbf{x}_1 - \mathbf{x}_2 &= l\left(-\frac{1}{2}, \frac{\sqrt{3}}{2}\right) + (q_1 - q_3, q_2 - q_4) \\ &= \mathbf{z} + \boldsymbol{\epsilon}\,,\end{aligned}$$

where $l = |\mathbf{z}| \gg |\boldsymbol{\epsilon}|$. Taylor expansion gives

$$|\mathbf{z} + \boldsymbol{\epsilon}| = \sqrt{\mathbf{z}\cdot\mathbf{z} + 2\mathbf{z}\cdot\boldsymbol{\epsilon} + \boldsymbol{\epsilon}\cdot\boldsymbol{\epsilon}} \approx l\left(1 + \frac{\mathbf{z}\cdot\boldsymbol{\epsilon}}{l^2}\right).$$

Hence, the potential energy is expanded to

$$V = \frac{k}{2}\left[-\frac{1}{2}(q_1 - q_3) + \frac{\sqrt{3}}{2}(q_2 - q_4)\right]^2 + \frac{k}{2}(q_3 - q_5)^2 + \frac{k}{2}\left[-\frac{1}{2}(q_5 - q_1) - \frac{\sqrt{3}}{2}(q_6 - q_2)\right]^2,$$

and so the matrix V is given by

$$\mathsf{V} = \frac{k}{4}\begin{pmatrix} 2 & 0 & -1 & \sqrt{3} & -1 & -\sqrt{3} \\ 0 & 6 & \sqrt{3} & -3 & -\sqrt{3} & -3 \\ -1 & \sqrt{3} & 5 & -\sqrt{3} & -4 & 0 \\ \sqrt{3} & -3 & -\sqrt{3} & 3 & 0 & 0 \\ -1 & -\sqrt{3} & -4 & 0 & 5 & \sqrt{3} \\ -\sqrt{3} & -3 & 0 & 0 & \sqrt{3} & 3 \end{pmatrix}.$$

Solving the characteristic equation

$$\det\left(\mathsf{V} - \omega^2\mathsf{T}\right) = 0$$

is too tedious in this case. However, we can guess most of the normal modes using the symmetries of the problem.

First of all, there are some zero modes, corresponding to the non-oscillatory rigid motions. The system can translate in two dimensions:

$$\mathbf{Q}^{(1)} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \text{ and } \mathbf{Q}^{(2)} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}.$$

Similarly, the system can rotate about its centre:

$$\mathbf{Q}^{(3)} = \begin{pmatrix} 1 \\ 0 \\ -\frac{1}{2} \\ -\frac{\sqrt{3}}{2} \\ -\frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{pmatrix}.$$

By symmetry, we expect there would be a mode where the masses expand and contract at the same frequency in the radial direction. This is expressed by the eigenvector

$$\mathbf{Q}^{(4)} = \begin{pmatrix} 0 \\ 1 \\ \frac{\sqrt{3}}{2} \\ -\frac{1}{2} \\ -\frac{\sqrt{3}}{2} \\ -\frac{1}{2} \end{pmatrix},$$

which has a normal frequency of $\omega^2 = \frac{3k}{m}$.

For a mode with reflection symmetry, we can try the form

$$\mathbf{Q}^{(5)} = \begin{pmatrix} 0 \\ -1 \\ \beta \\ \gamma \\ -\beta \\ \gamma \end{pmatrix},$$

which works with $\beta = \frac{\sqrt{3}}{2}$ and $\gamma = \frac{1}{2}$, and the normal frequency can be found out to be $\omega^2 = \frac{3k}{2m}$.

Using orthogonality, we can find out the last normal mode:

$$\mathbf{Q}^{(6)} = \begin{pmatrix} 1 \\ 0 \\ -\frac{1}{2} \\ \frac{\sqrt{3}}{2} \\ -\frac{1}{2} \\ -\frac{\sqrt{3}}{2} \end{pmatrix},$$

with $\omega^2 = \frac{3k}{2m}$.

17.4 Normal Modes and Group Representations

Consider an oscillating system with the Lagrangian of the form

$$L = \frac{1}{2}\dot{\mathsf{q}}^{\mathrm{T}}\mathsf{T}\dot{\mathsf{q}} - \frac{1}{2}\mathsf{q}^{\mathrm{T}}\mathsf{V}\mathsf{q}\,.$$

If the system has a symmetry group G, then the action of a symmetry transformation $g \in G$ on the vector of generalised coordinates is

$$\mathbf{q} \to \rho(g)\mathsf{q}\,.$$

If $\mathbf{q}$ has N components, then ρ is an N-dimensional representation of G. Since g is a symmetry operation on the system, the kinetic and potential energies must transform so that L is invariant. For all $g \in G$,

$$\begin{cases} \rho(g)^{\mathrm{T}}\mathsf{T}\rho(g) = \mathsf{T} \\ \rho(g)^{\mathrm{T}}\mathsf{V}\rho(g) = \mathsf{V}\,. \end{cases}$$

By Weyl's unitary trick (Theorem 16.30), we can always find a basis such that all the representations $\rho(g)$ are unitary. Since the coordinates are real, we can choose such a basis so that $\rho(g)$ are orthogonal, and

$$\begin{cases} \rho(g)^{-1}\mathsf{T}\rho(g) = \mathsf{T} \\ \rho(g)^{-1}\mathsf{V}\rho(g) = \mathsf{V}\,. \end{cases} \tag{$\dagger$}$$

Often the representation ρ is reducible, so the matrices $\rho(g)$ may be transformed into block-diagonal form by a similarity transformation

$$\mathsf{S}\rho(g)\mathsf{S}^{-1} = \begin{pmatrix} \phi^{(1)}(g) & \mathsf{0} & \cdots & \mathsf{0} \\ \mathsf{0} & \phi^{(2)}(g) & \cdots & \mathsf{0} \\ \vdots & \vdots & \ddots & \mathsf{0} \\ \mathsf{0} & \mathsf{0} & \cdots & \phi^{(k)}(g) \end{pmatrix}.$$

For simplicity, let us assume here that each of the irreducible representations $\phi^{(i)}$ are different from each other. Note that the matrices along the diagonal are square, of dimensions $\{n_\alpha\}_{\alpha=1}^k$. Transform equations (†), we have

$$\begin{aligned}&\begin{cases}\mathsf{S}\rho(g)^{-1}\mathsf{T}\rho(g)\mathsf{S}^{-1} = \mathsf{S}\mathsf{T}\mathsf{S}^{-1}\\ \mathsf{S}\rho(g)^{-1}\mathsf{V}\rho(g)\mathsf{S}^{-1} = \mathsf{S}\mathsf{V}\mathsf{S}^{-1}\end{cases}\\ \Longrightarrow\ &\begin{cases}(\mathsf{S}\mathsf{T}\mathsf{S}^{-1})(\mathsf{S}\rho(g)\mathsf{S}^{-1}) = (\mathsf{S}\rho(g)\mathsf{S}^{-1})(\mathsf{S}\mathsf{T}\mathsf{S}^{-1})\\ (\mathsf{S}\mathsf{V}\mathsf{S}^{-1})(\mathsf{S}\rho(g)\mathsf{S}^{-1}) = (\mathsf{S}\rho(g)\mathsf{S}^{-1})(\mathsf{S}\mathsf{V}\mathsf{S}^{-1})\,.\end{cases}\end{aligned}$$

Rewriting $\tilde{\mathsf{T}} = \mathsf{S}\mathsf{T}\mathsf{S}^{-1}$ and $\tilde{\mathsf{V}} = \mathsf{S}\mathsf{V}\mathsf{S}^{-1}$, and since $\mathsf{S}\rho(g)\mathsf{S}^{-1}$ is block diagonal, we have

$$\begin{aligned}\tilde{\mathsf{T}}(\mathsf{S}\rho(g)\mathsf{S}^{-1}) &= \begin{pmatrix}\mathsf{T}^{(11)} & \mathsf{T}^{(12)} & \cdots & \mathsf{T}^{(1k)}\\ \mathsf{T}^{(21)} & \mathsf{T}^{(22)} & \cdots & \mathsf{T}^{(2k)}\\ \vdots & \vdots & \ddots & \vdots\\ \mathsf{T}^{(k1)} & \mathsf{T}^{(k2)} & \cdots & \mathsf{T}^{(kk)}\end{pmatrix}\begin{pmatrix}\phi^{(1)}(g) & 0 & \cdots & 0\\ 0 & \phi^{(2)}(g) & \cdots & 0\\ \vdots & \vdots & \ddots & 0\\ 0 & 0 & \cdots & \phi^{(k)}(g)\end{pmatrix}\\ &= \begin{pmatrix}\mathsf{T}^{(11)}\phi^{(1)}(g) & \mathsf{T}^{(12)}\phi^{(2)}(g) & \cdots & \mathsf{T}^{(1k)}\phi^{(k)}(g)\\ \mathsf{T}^{(21)}\phi^{(1)}(g) & \mathsf{T}^{(22)}\phi^{(2)}(g) & \cdots & \mathsf{T}^{(2k)}\phi^{(k)}(g)\\ \vdots & \vdots & \ddots & \vdots\\ \mathsf{T}^{(k1)}\phi^{(1)}(g) & \mathsf{T}^{(k2)}\phi^{(2)}(g) & \cdots & \mathsf{T}^{(kk)}\phi^{(k)}(g)\end{pmatrix},\end{aligned}$$

$$\begin{aligned}(\mathsf{S}\rho(g)\mathsf{S}^{-1})\tilde{\mathsf{T}} &= \begin{pmatrix}\phi^{(1)}(g) & 0 & \cdots & 0\\ 0 & \phi^{(2)}(g) & \cdots & 0\\ \vdots & \vdots & \ddots & 0\\ 0 & 0 & \cdots & \phi^{(k)}(g)\end{pmatrix}\begin{pmatrix}\mathsf{T}^{(11)} & \mathsf{T}^{(12)} & \cdots & \mathsf{T}^{(1k)}\\ \mathsf{T}^{(21)} & \mathsf{T}^{(22)} & \cdots & \mathsf{T}^{(2k)}\\ \vdots & \vdots & \ddots & \vdots\\ \mathsf{T}^{(k1)} & \mathsf{T}^{(k2)} & \cdots & \mathsf{T}^{(kk)}\end{pmatrix}\\ &= \begin{pmatrix}\phi^{(1)}(g)\mathsf{T}^{(11)} & \phi^{(1)}(g)\mathsf{T}^{(12)} & \cdots & \phi^{(1)}(g)\mathsf{T}^{(1k)}\\ \phi^{(2)}(g)\mathsf{T}^{(21)} & \phi^{(2)}(g)\mathsf{T}^{(22)} & \cdots & \phi^{(2)}(g)\mathsf{T}^{(2k)}\\ \vdots & \vdots & \ddots & \vdots\\ \phi^{(k)}(g)\mathsf{T}^{(k1)} & \phi^{(k)}(g)\mathsf{T}^{(k2)} & \cdots & \phi^{(k)}(g)\mathsf{T}^{(kk)}\end{pmatrix}.\end{aligned}$$

Therefore, we must have

$$\mathsf{T}^{(\alpha\beta)}\phi^{(\beta)}(g) = \phi^{(\alpha)}(g)\mathsf{T}^{(\alpha\beta)}\,.$$

If the block diagonal elements of $\mathsf{S}\rho(g)\mathsf{S}^{-1}$ are different, then by Schur's Lemma (Lemma 16.32),

$$\begin{cases}\tilde{\mathsf{T}}^{(\alpha\beta)} = t_\alpha\delta_{\alpha\beta}\mathsf{I}_\alpha\\ \tilde{\mathsf{V}}^{(\alpha\beta)} = v_\alpha\delta_{\alpha\beta}\mathsf{I}_\alpha\,,\end{cases}$$

where $t_i, v_i \in \mathbb{R}$, and I_i is the $n_i \times n_i$ identity matrix. Therefore,

$$\mathsf{S}\mathsf{T}\mathsf{S}^{-1} = \begin{pmatrix}t_1\mathsf{I}_1 & 0 & \cdots & 0\\ 0 & t_2\mathsf{I}_2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & t_k\mathsf{I}_k\,,\end{pmatrix}$$

$$\mathsf{S}\mathsf{V}\mathsf{S}^{-1} = \begin{pmatrix}v_1\mathsf{I}_1 & 0 & \cdots & 0\\ 0 & v_2\mathsf{I}_2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & v_k\mathsf{I}_k\end{pmatrix}.$$

Once the kinetic and potential energy matrices are diagonalised in this way, finding the normal modes is straightforward. The normal frequencies are

$$\omega_\alpha^2 = \frac{v_\alpha}{t_\alpha}$$

with degeneracy n_α. The generalised eigenvectors of normal modes span the invariant subspace of the underlying vector space acted on by the corresponding irreducible representation of the symmetry group.

If an irreducible representation $\phi^{(\alpha)}(g)$ occurs with multiplicity $m_\alpha > 1$, then the corresponding diagonal block of t_α and v_α are replaced by an $m_\alpha \times m_\alpha$ block matrix. T and V are not completely diagonal, but finding the normal modes reduces to solving separate generalised eigenvalue problems for each α.

17.4.1 Example: CO_2 Molecule Revisited

We will, again, only consider the 1D motions of the atoms in a CO_2 molecule. Its symmetry group is therefore $G = \{I, m\}$. The action of the group on the coordinates $\mathbf{q} = (x_1, x_2, x_3)$ is given by the representation ϕ, where

$$\phi(I)\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} = \begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} \qquad \Longrightarrow \qquad \phi(I) = \begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix},$$
$$\phi(m)\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} = \begin{pmatrix}-x_3\\-x_2\\-x_1\end{pmatrix} \qquad \Longrightarrow \qquad \phi(m) = \begin{pmatrix}0&0&-1\\0&-1&0\\-1&0&0\end{pmatrix}.$$

The character of this representation is

$$\chi_\phi(I) = 3\,,\ \chi_\phi(m) = -1\,.$$

The character table of this group is

G	I	m
$\chi_{\rho^{(1)}}$	1	1
$\chi_{\rho^{(2)}}$	1	-1

so the representation is decomposed into

$$\phi = \rho^{(1)} \oplus 2\rho^{(2)}\,.$$

This corresponds to the single symmetric mode ($\mathbf{Q}^{(2)}$) and two anti-symmetric modes ($\mathbf{Q}^{(1)}$ and $\mathbf{Q}^{(3)}$) as seen before.

17.4.2 Example: Equilateral Triangle Revisited

Consider the 2D vibration of an equilateral triangle, which belongs to the symmetry group $D_3 = \{I, R, R^2, m_1, m_2, m_3\}$. Using the coordinate as before, we can find out the representation

$$\phi(I) = \begin{pmatrix}1&0&0&0&0&0\\0&1&0&0&0&0\\0&0&1&0&0&0\\0&0&0&1&0&0\\0&0&0&0&1&0\\0&0&0&0&0&1\end{pmatrix},$$

$$\phi(m_1) = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix},$$

$$\phi(R) = \begin{pmatrix} 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{\sqrt{3}}{2} \\ 0 & 0 & 0 & 0 & -\frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} & 0 & 0 & 0 & 0 \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{\sqrt{3}}{2} & 0 & 0 \\ 0 & 0 & -\frac{\sqrt{3}}{2} & -\frac{1}{2} & 0 & 0 \end{pmatrix},$$

with character

$$\chi_\phi(I) = 6\,,\ \chi_\phi(m_1) = 0\,,\ \chi_\phi(R) = 0\,.$$

We have the character table of D_3:

D_3	I	R	R^2	m_1	m_2	m_3
$\chi_{\rho^{(1)}}$	1	1	1	1	1	1
$\chi_{\rho^{(2)}}$	1	1	1	−1	−1	−1
$\chi_{\rho^{(3)}}$	2	−1	−1	0	0	0

and the representation is decomposed as

$$\phi = \rho^{(1)} \oplus \rho^{(2)} \oplus 2\rho^{(3)}\,.$$

The trivial representation $\rho^{(1)}$ is the symmetric breathing mode ($\mathbf{Q}^{(4)}$). $\rho^{(2)}$ corresponds to the rigid rotation ($\mathbf{Q}^{(3)}$) since it is unchanged by R, but reversed by m. The two translations ($\mathbf{Q}^{(1)}$ and $\mathbf{Q}^{(2)}$) corresponds to a two-dimensional irreducible representation $\rho^{(3)}$. The final $\rho^{(3)}$ corresponds to another pair of degenerate non-zero modes ($\mathbf{Q}^{(5)}$ and $\mathbf{Q}^{(6)}$).

This is demonstrated by the transformations of the normal mode eigenvectors. For the two translational modes ($\mathbf{Q}^{(1)}$ and $\mathbf{Q}^{(2)}$), we have

$$\rho(m_1)\begin{pmatrix} | & | \\ \mathbf{Q}^{(1)} & \mathbf{Q}^{(2)} \\ | & | \end{pmatrix} = \begin{pmatrix} | & | \\ \mathbf{Q}^{(1)} & \mathbf{Q}^{(2)} \\ | & | \end{pmatrix}\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$\rho(R)\begin{pmatrix} | & | \\ \mathbf{Q}^{(1)} & \mathbf{Q}^{(2)} \\ | & | \end{pmatrix} = \begin{pmatrix} | & | \\ \mathbf{Q}^{(1)} & \mathbf{Q}^{(2)} \\ | & | \end{pmatrix}\begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}.$$

The other pair of doubly degenerate modes $\begin{pmatrix}\mathbf{Q}^{(6)} & \mathbf{Q}^{(5)}\end{pmatrix}$ transforms just as $\begin{pmatrix}\mathbf{Q}^{(1)} & \mathbf{Q}^{(2)}\end{pmatrix}$. For the rotation and dilation modes, we have

$$\begin{aligned} \rho(m_1)\mathbf{Q}^{(3)} &= -\mathbf{Q}^{(3)}\,, \\ \rho(R)\mathbf{Q}^{(3)} &= \mathbf{Q}^{(3)}\,, \\ \rho(m_1)\mathbf{Q}^{(4)} &= \mathbf{Q}^{(4)}\,, \\ \rho(R)\mathbf{Q}^{(4)} &= \mathbf{Q}^{(4)}\,. \end{aligned}$$