Mathematical Methods, Michaelmas 2006

Prof. F.H.L. Essler

December 6, 2006

Please report errors and typos to fab@thphys.ox.ac.uk

I thank K. Drescher and W. Goetze for their help in preparing some of these notes.

Contents

1 Introduction

In the first part of the course we will learn about **linear vector spaces** and operations on them.

A lot of this material will be rather abstract and many of you may consider it boring. However, this is the framework that underlies some of the most fascinating theories such as

- Quantum Mechanics and
- Relativity.

Furthermore it is very useful for just about anything that involves any mathematics: Finance, economics, engineering, computer science, etc.

1.1 Things you (should) already know about vectors and matrices

Let us consider vectors in the plane. We can **add** vectors, and we can **multiply** vectors

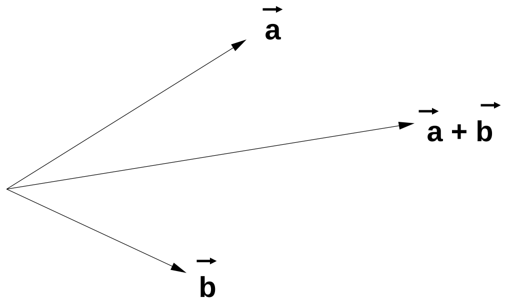


Figure 1: Adding two vectors in the plane.

by real numbers as is shown in Figs 1 and 2. Given two vectors $\vec{\mathbf{a}}$ and $\vec{\mathbf{b}}$, we can define

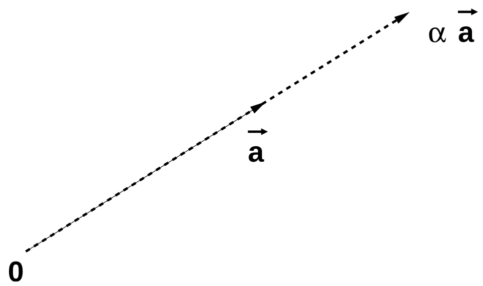


Figure 2: Multiplication of the vector $\vec{a}$ by the real number α amounts to elongating $\vec{a}$ by a factor α.

their **scalar product**, which is a real number, as

$$|\vec{\mathbf{a}}|\ \ |\vec{\mathbf{b}}| \cos\varphi, \tag{1}$$

where φ is the angle between the two vectors.

Let us now choose a **coordinate system**, see Fig. 3. Here $\vec{\mathbf{e}}_1$ and $\vec{\mathbf{e}}_2$ are vectors of unit length, which are orthogonal (this means that the angle between them is 90 degrees). In this coordinate system, we can represent each vector in the form

$$\vec{\mathbf{a}} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}. \tag{2}$$

This is a shorthand notation expressing the fact that $\vec{\mathbf{a}}$ can be written as a **linear combination** of $\vec{\mathbf{e}}_1$ and $\vec{\mathbf{e}}_2$

$$\vec{\mathbf{a}} = a_1\ \vec{\mathbf{e}}_1 + a_2\ \vec{\mathbf{e}}_2\ . \tag{3}$$

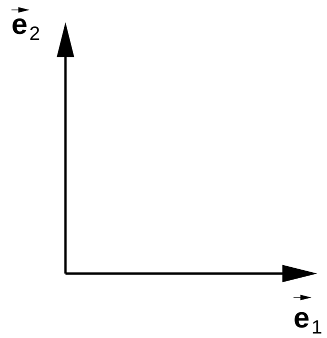


Figure 3: A coordinate system. $\vec{\mathbf{e}}_{1,2}$ are orthogonal vectors of unit length.

Using the coordinate representation the scalar product is simply

$$a_1 b_1 + a_2 b_2 = \begin{pmatrix} a_1, & a_2 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \vec{\mathbf{a}}^{\mathrm{T}} \cdot \vec{\mathbf{b}} \tag{4}$$

The length of the vector **a** is defined in terms of the scalar product as

$$|\vec{\mathbf{a}}| = \sqrt{\vec{\mathbf{a}}^T \cdot \vec{\mathbf{a}}} = \sqrt{a_1^2 + a_2^2}, \tag{5}$$

where $\vec{\mathbf{a}}^T = \begin{pmatrix} a_1, & a_2 \end{pmatrix}$ is the transpose vector to $\vec{\mathbf{a}}$. It is a row vector, whereas $\vec{\mathbf{a}}$ is a column vector. Last, but not least we can multiply vectors by matrices, consider e.g.

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \tag{6}$$

and multiply M by the vector $\vec{\mathbf{a}}$ to give

$$M\ \vec{\mathbf{a}} = \begin{pmatrix} M_{11}a_1 + M_{12}a_2 \\ M_{21}a_1 + M_{22}a_2 \end{pmatrix}, \tag{7}$$

which clearly is a vector. Multiplication by a matrix is a **linear** operation because

$$M(\alpha \vec{\mathbf{a}} + \beta \vec{\mathbf{b}}) = \alpha(M\vec{\mathbf{a}}) + \beta(M\vec{\mathbf{b}}) \tag{8}$$

We can also multiply two matrices and obtain another matrix.

It is very important to realize that the **vectors** exist independently from any specific coordinate representation. As a matter of fact there are many different coordinate representations of the same vector. For example, we could choose a coordinate system Here $\vec{\mathbf{e}}'_1$ and $\vec{\mathbf{e}}'_2$ are orthogonal vectors of unit length and any vector can be written in the form

$$\vec{\mathbf{a}} = \begin{pmatrix} a'_1 \\ a'_2 \end{pmatrix}, \tag{9}$$

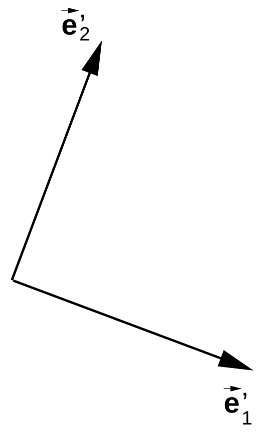


Figure 4: A second coordinate system. $\vec{\mathbf{e}}'_{1,2}$ are orthogonal vectors of unit length.

which means that

$$\vec{\mathbf{a}} = a'_1\ \vec{\mathbf{e}}'_1 + a'_2\ \vec{\mathbf{e}}'_2\ . \tag{10}$$

What we have just reviewed is a special example of a **MUCH** more general concept, the **linear vector space (LVS)**. In particular, quantum mechanical wave functions form a LVS and physical "observables" such as energy (E) and momentum (p) are represented as "matrices" on this LVS.

2 Definition of Linear Vector Spaces

Let's start with a set V of objects, which we represent by the symbol $|\rangle$ (a Dirac "ket" vector). This notation is called **Dirac notation** and is very widely used in Quantum Mechanics. We define two operations on V

1. **"addition"** of vectors: $|\mathbf{a}\rangle + |\mathbf{b}\rangle$ and
2. **"multiplication"** of vectors by a number α: $\alpha\,|\mathbf{a}\rangle$

By definition "addition" is

$$\text{commutative}: \ |\mathbf{a}\rangle + |\mathbf{b}\rangle = |\mathbf{b}\rangle + |\mathbf{a}\rangle \tag{11}$$

$$\text{associative}: (|\mathbf{a}\rangle + |\mathbf{b}\rangle) + |\mathbf{c}\rangle = |\mathbf{a}\rangle + (|\mathbf{b}\rangle + |\mathbf{c}\rangle) \tag{12}$$

These basically mean that it does not matter in which order we add vectors. Multiplication fulfils

$$1\ |\mathbf{a}\rangle = |\mathbf{a}\rangle \tag{13}$$

$$\alpha(\beta\,|\mathbf{a}\rangle) = (\alpha\beta)\,|\mathbf{a}\rangle \tag{14}$$

$$(\alpha+\beta)\,|\mathbf{a}\rangle = \alpha\,|\mathbf{a}\rangle + \beta\,|\mathbf{a}\rangle \tag{15}$$

$$\alpha(|\mathbf{a}\rangle + |\mathbf{b}\rangle) = \alpha\,|\mathbf{a}\rangle + \alpha\,|\mathbf{b}\rangle \tag{16}$$

The last two are called *distributive laws.*
The set V is called a **linear vector space** if

1. *V is closed under addition:*

$$|\mathbf{a}\rangle\,, |\mathbf{b}\rangle \in V \Longrightarrow |\mathbf{a}\rangle + |\mathbf{b}\rangle \in V. \tag{17}$$

   This means that the sum of two vectors belonging to V is a vector that also belongs to V.

2. *V is closed under multiplication:*

$$|\mathbf{a}\rangle \in V,\ \alpha \text{ a number } \Longrightarrow \alpha\,|\mathbf{a}\rangle \in V. \tag{18}$$

   This means that if we multiply a vector belonging to V by a number, we get another vector that belongs to V.

3. There is a *null element* $|\mathbf{0}\rangle \in V$, such that for any vector in V

$$|\mathbf{a}\rangle + |\mathbf{0}\rangle = |\mathbf{a}\rangle \ \ . \tag{19}$$

4. For any $|\mathbf{a}\rangle \in V$ there is an element $|\mathbf{a}'\rangle \in V$ such that

$$|\mathbf{a}\rangle + |\mathbf{a}'\rangle = |\mathbf{0}\rangle \ \ . \tag{20}$$

If we consider multiplication by real numbers "$\alpha \in \mathbb{R}$", V is called a **real vector space**. If α is complex "$\alpha \in \mathbb{C}$", V is called a **complex vector space**.

Remark: One can consider more general vectors spaces, where multiplication is defined not for real or complex numbers, but for general "fields", e.g. rational numbers.

Examples of linear vector spaces:

A. $\mathbb{R}$ (the real numbers) is a real vector space. The "vectors" in this space are simply the real numbers. Addition and multiplication (by real numbers) fulfil the requirements on a real vector space. The null element is the number zero. $\alpha + (-\alpha) = 0$ as in the 4^{th} condition above.

B. $\mathbb{C}$ (the complex numbers) is a complex vector space. The vectors in this space are simply the complex numbers themselves.

C. "Arrows in the plane:" Let us consider the space of arrows in the plane (which originate from a common point, the origin). Let us check whether this space is in fact a linear vector space.

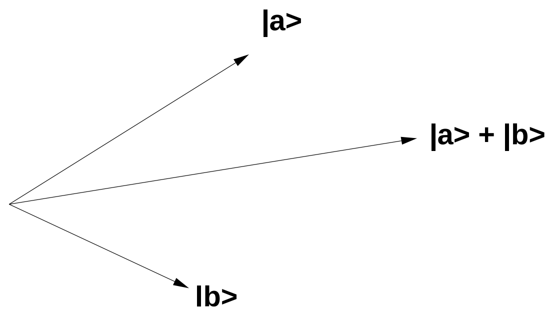


Figure 5: Arrows in the plane are a linear vector space.

1. Law of addition: we define the sum of two arrows $|\mathbf{a}\rangle + |\mathbf{b}\rangle$ as the arrow obtained by the geometrical addition shown in Fig.5. This law obviously fulfils
   - $|\mathbf{a}\rangle + |\mathbf{b}\rangle = |\mathbf{b}\rangle + |\mathbf{a}\rangle \ \surd$
   - $(|\mathbf{a}\rangle + |\mathbf{b}\rangle) + |\mathbf{c}\rangle = |\mathbf{a}\rangle + (|\mathbf{b}\rangle + |\mathbf{c}\rangle) \ \surd$
2. Multiplication by **real** numbers: Changes the length of the arrow $|\mathbf{a}\rangle$ by the factor α. Clearly $\alpha\,|\mathbf{a}\rangle$ is also an arrow in the plane. Furthermore, we have
   - $1 \ \ |\mathbf{a}\rangle = |\mathbf{a}\rangle \quad \surd$
   - $\alpha(\beta\,|\mathbf{a}\rangle) = (\alpha\beta)\,|\mathbf{a}\rangle \quad \surd$
   - $(\alpha + \beta)\,|\mathbf{a}\rangle = \alpha\,|\mathbf{a}\rangle + \beta\,|\mathbf{a}\rangle \quad \surd$
   - $\alpha(|\mathbf{a}\rangle + |\mathbf{b}\rangle = \alpha\,|\mathbf{a}\rangle + \alpha\,|\mathbf{b}\rangle \quad \surd$
3. The null element is the arrow of zero length $|\mathbf{0}\rangle$
   - $|\mathbf{a}\rangle + |\mathbf{0}\rangle = |\mathbf{a}\rangle \quad \surd$
   - $|\mathbf{0}\rangle \equiv 0\,|\mathbf{a}\rangle$ as a matter of fact.
4. For every $|\mathbf{a}\rangle$ we have $|\mathbf{a}\rangle + (-1)\,|\mathbf{a}\rangle = |\mathbf{0}\rangle \quad \surd$

We see that the space of "arrows in the plane" is a **real vector space**. Another familiar representation of this vector space is as two component vectors ("coordinate representation"):

$$|\mathbf{a}\rangle = \begin{pmatrix} x \\ y \end{pmatrix} = \vec{\mathbf{a}} \tag{21}$$

Then addition and multiplication by a number simply work like

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \end{pmatrix} \ , \qquad \alpha \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \alpha x \\ \alpha y \end{pmatrix} \ . \tag{22}$$

This vector space is called "$\mathbb{R}^2$".

D. An obvious generalization is to consider "arrows in a 3D space" or equivalently vectors

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{23}$$

with analogous laws of addition and multiplication. This vector space is called $\mathbb{R}^3$.

E. It is now just a small step to consider arrows in an n-dimensional space, or equivalently vectors

$$\vec{\mathbf{x}} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \tag{24}$$

with multiplication and addition defined as

$$\alpha\vec{\mathbf{x}} = \begin{pmatrix} \alpha x_1 \\ \vdots \\ \alpha x_n \end{pmatrix} \text{ and } \vec{\mathbf{x}} + \vec{\mathbf{y}} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}. \tag{25}$$

This real vector space is called $\mathbb{R}^n$.

F. What about other complex vector spaces? Well, consider again arrows in the plane with the usual geometrical addition. But now define multiplication by a **complex number** $re^{i\psi}$ in the way shown in Fig.6

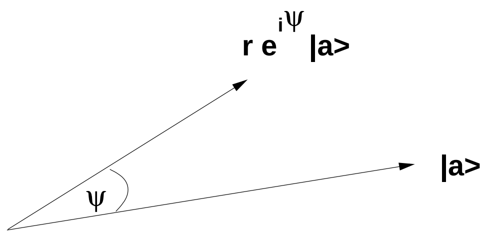


Figure 6: Multiplication of the vector $|a\rangle$ by the complex number $r\ e^{i\psi}$ is defined by elongating the vector by a factor r and rotating it by an angle ψ.

Let us consider a vector $|\mathbf{b}\rangle$ defined as

$$re^{i\psi}\ |\mathbf{a}\rangle = |\mathbf{b}\rangle\,. \tag{26}$$

We define multiplication such that

- the length of $|\mathbf{b}\rangle$ is r times the length of $|\mathbf{a}\rangle$

- $|\mathbf{b}\rangle$ is rotated by an angle of ψ with respect to $|\mathbf{a}\rangle$.

You can check that arrows in the plane with addition and multiplication defined in this way form a **complex** vector space.

Remark: This LVS is actually the same as the complex numbers $\mathbb{C}$.

G. Another complex vector space is obtained by considering vectors $\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$, where $z_{1,2}$ are now **complex** numbers. Addition and multiplication by a complex number α are defined as

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} + \begin{pmatrix} z_1' \\ z_2' \end{pmatrix} = \begin{pmatrix} z_1 + z_1' \\ z_2 + z_2' \end{pmatrix} , \qquad \alpha \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} \alpha z_1 \\ \alpha z_2 \end{pmatrix} . \tag{27}$$

This linear vector space is called $\mathbb{C}^2$. Similarly we get $\mathbb{C}^n$.

H. Last but not least consider the space $\mathrm{L}_2(a, b)$ of *square-integrable continuous complex valued functions* on some interval $[a, b]$, i.e. functions for which the integral

$$\int_a^b dx \, |f(x)|^2 \text{ exists, i.e. it is } < \infty . \tag{28}$$

Note that we take $f(x)$ to be a complex number here. $\mathrm{L}_2(a, b)$ is a LVS, because

(a) addition: $|\mathbf{f}\rangle + |\mathbf{g}\rangle$ is a square-integrable function defined by $f(x) + g(x)$

(b) multiplication: $\alpha \, |\mathbf{f}\rangle$ is the function $\alpha f(x)$

(c) there is a null element (the constant function zero)

(d) $|\mathbf{f}\rangle + (-1) \, |\mathbf{f}\rangle = |\mathbf{0}\rangle$, or equivalently $f(x) - f(x) = 0$ for all x.

3 Scalar Product and Dual Vectors

For arrows in the plane we know that the scalar product for two vectors $\vec{\mathbf{a}}$ and $\vec{\mathbf{b}}$ is simply

$$\vec{\mathbf{a}}^T \cdot \vec{\mathbf{b}} = a_1 b_1 + a_2 b_2, \tag{29}$$

where (as above)

$$\vec{\mathbf{a}} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \vec{\mathbf{a}}^T = \begin{pmatrix} a_1, & a_2 \end{pmatrix} . \tag{30}$$

The generalization to $\mathbb{R}^n$ is straightforward. Given two vectors

$$\vec{\mathbf{a}} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}, \vec{\mathbf{b}} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}, \tag{31}$$

we have $\vec{\mathbf{a}}^T = (a_1, a_2, \ldots, a_n)$ and

$$\vec{\mathbf{a}}^T \cdot \vec{\mathbf{b}} = \sum_{j=1}^{n} a_j b_j. \tag{32}$$

Note that

- the "length" of $\vec{\mathbf{a}}$ is

$$|\vec{\mathbf{a}}| = \sqrt{\vec{\mathbf{a}}^T \cdot \vec{\mathbf{a}}} = \sqrt{\sum_{j=1}^{n} a_j^2}. \tag{33}$$

- Two vectors $\vec{\mathbf{a}}$ and $\vec{\mathbf{b}}$ are called **orthogonal** if

$$\vec{\mathbf{a}}^T \cdot \vec{\mathbf{b}} = 0\ . \tag{34}$$

How does this work in complex vector spaces? Let us look at $\mathbb{C}^n$ first:

$$\vec{\mathbf{a}} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \tag{35}$$

but now a_j are complex numbers. We want the scalar product of $\vec{\mathbf{a}}$ with $\vec{\mathbf{a}}$ to be the square of the "length" of $\vec{\mathbf{a}}$, i.e. a **real and positive** number. Therefore we must define the scalar product as

$$\vec{\mathbf{b}}^\dagger \cdot \vec{\mathbf{a}} = \begin{pmatrix} b_1^*, & \cdots, & b_n^* \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \sum_{j=1}^{n} b_j^* a_j. \tag{36}$$

Using this definition we have

$$\vec{\mathbf{a}}^\dagger \cdot \vec{\mathbf{a}} = \sum_{j=1}^{n} a_j^* a_j = \sum_{j=1}^{n} |a_j|^2. \tag{37}$$

This is real and positive as desired. We call $\vec{\mathbf{b}}^\dagger$ the **dual vector** to $\vec{\mathbf{b}}$. It is defined by taking the transpose of the vector $\vec{\mathbf{b}}$ and then taking the complex conjugate of all elements. But **beware:**

- the dual vector to $\lambda\vec{\mathbf{b}}$ is $\lambda^*\vec{\mathbf{b}}^\dagger$!
- $\vec{\mathbf{b}}^\dagger \cdot \vec{\mathbf{a}} \equiv (\vec{\mathbf{a}}^\dagger \cdot \vec{\mathbf{b}})^* \neq \vec{\mathbf{a}}^\dagger \cdot \vec{\mathbf{b}}$ in general.

So for complex vector spaces the scalar product of $\vec{\mathbf{a}}$ with $\vec{\mathbf{b}}$ is generally not the same as the scalar product of $\vec{\mathbf{b}}$ with $\vec{\mathbf{a}}$.

3.1 General Definition

The scalar product is a rule that assigns a complex number to any given pair of vectors $|\mathrm{a}\rangle$ and $|\mathrm{b}\rangle$ belonging to a LVS. We denote this number by

$$\langle \mathbf{a} \mid \mathbf{b} \rangle \ , \qquad \text{called a bra} - \text{ket in Dirac notation.} \tag{38}$$

The following rules apply:

1. $\langle \mathbf{a} \mid \mathbf{b} \rangle = (\langle \mathbf{b} \mid \mathbf{a} \rangle)^*$.
2. $\langle \mathbf{a} \mid \mathbf{a} \rangle$ is real and ≥ 0 ($=0$ only if $|\mathbf{a}\rangle = |\mathbf{0}\rangle$).
   $\sqrt{\langle \mathbf{a} \mid \mathbf{a} \rangle}$ is the "length" of the vector $|\mathbf{a}\rangle$.
3. Let $|\mathbf{d}\rangle = \alpha\,|\mathbf{a}\rangle + \beta\,|\mathbf{b}\rangle$. Then the scalar product of $|\mathbf{d}\rangle$ with a vector $\langle \mathbf{c}|$ is defined as
   $$\langle \mathbf{c} \mid \mathbf{d} \rangle = \alpha \ \langle \mathbf{c} \mid \mathbf{a} \rangle + \beta \ \langle \mathbf{c} \mid \mathbf{b} \rangle \ . \tag{39}$$
4. $|\mathbf{a}\rangle$ and $|\mathbf{b}\rangle$ are **orthogonal** if $\langle \mathbf{a} \mid \mathbf{b} \rangle = \langle \mathbf{b} \mid \mathbf{a} \rangle = 0$.

Here $\langle \mathbf{a}|$ is the dual vector to $|\mathbf{a}\rangle$. It is also called a "bra" vector, whereas $|\mathbf{a}\rangle$ is called a "ket"-vector. The vector dual to the ket vector $|\mathbf{d}\rangle = \alpha\,|\mathbf{a}\rangle + \beta\,|\mathbf{b}\rangle$ is

$$\langle \mathbf{d}| = \alpha^* \,\langle \mathbf{a}| + \beta^* \,\langle \mathbf{b}| \ . \tag{40}$$

Note that you have to take the complex conjugate of α and β here! This rule is just like for the complex vector space case we discussd before.

Example: Scalar product for square integrable functions.

$$\langle \mathbf{f} \mid \mathbf{g} \rangle = \int_a^b dx\,[f^*(x)g(x)] \tag{41}$$

4 Linear Independence of Vectors

The vectors $|v_1\rangle$, $|v_2\rangle$,..., $|v_N\rangle$ are said to be **linearly independent** if the relation

$$\sum_{j=1}^{n} \alpha_j \,|\mathbf{v_j}\rangle = 0, \text{ where } \ (|v_j\rangle \neq 0) \tag{42}$$

necessarily implies that $\alpha_1 = \alpha_2 = ... = \alpha_N = 0$.

Examples:

A. $|\mathbf{v_1}\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $|\mathbf{v_2}\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ are linearly independent.

B. $|\mathbf{v_1}\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $|\mathbf{v_2}\rangle \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and $|\mathbf{v_3}\rangle = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ are linearly dependent because

$$|\mathbf{v_1}\rangle + |\mathbf{v_2}\rangle - |\mathbf{v_3}\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 0. \tag{43}$$

The maximal number of linearly independent vectors in a linear vectors space is called the **dimension** of the linear vector space. The set of vectors $|v_j\rangle$ that are linearly independent and have the property that **any** vector $|a\rangle$ in the LVS can be expressed as a linear combination, i.e. as

$$|\mathbf{a}\rangle = \sum_{j=1}^{n} a_j \, |\mathbf{v_j}\rangle \,, \tag{44}$$

is called a **basis** of the LVS. The numbers a_j are called the **components** of $|\mathbf{a}\rangle$ with respect to the basis vectors $|\mathbf{v_j}\rangle$.

Examples:

A. $\mathbb{R}^2$ has dimension 2 and a basis is

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} \ .$$

Any vector $\begin{pmatrix} a \\ b \end{pmatrix}$ can be written as a linear combination of the two basis vectors

$$\begin{pmatrix} a \\ b \end{pmatrix} = a \begin{pmatrix} 1 \\ 0 \end{pmatrix} + b \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Another basis for $\mathbb{R}^2$ is

$$\mathcal{B}' = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$$

Any vector can be written as a linear combination of the two basis vectors

$$\begin{pmatrix} a \\ b \end{pmatrix} = (a-b) \begin{pmatrix} 1 \\ 0 \end{pmatrix} + b \begin{pmatrix} 1 \\ 1 \end{pmatrix} \ . \tag{45}$$

B. $\mathbb{R}^n$ has dimension n and a basis is

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \right\}$$

C. $\mathbb{C}^2$ has dimension 2, a basis is

$$\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

D. Arrows in the plane with multiplication by a complex number is a LVS of dimension 1. This is because you can obtain any arrow from a given one by multiplication with a complex number (i.e. elongation and rotation).

5 Different Bases and Orthogonality

There are many different bases for a given LVS. For example for $\mathbb{R}^2$ $\underbrace{\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}}_{=\{|\mathbf{e_1}\rangle, |\mathbf{e_2}\rangle\}}$ is a basis and $\underbrace{\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}}_{=\{|\mathbf{e'_1}\rangle, |\mathbf{e'_2}\rangle\}}$ is another.

The first basis is our favourite, because it is **orthonormal** [1].

$$\left.\begin{aligned} \langle \mathbf{e_1} \mid \mathbf{e_1} \rangle &= \begin{pmatrix} 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 1 \\ \langle \mathbf{e_2} \mid \mathbf{e_2} \rangle &= \begin{pmatrix} 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 1 \\ \langle \mathbf{e_1} \mid \mathbf{e_2} \rangle &= \langle \mathbf{e_2} \mid \mathbf{e_1} \rangle = 0 \end{aligned}\right\} \quad \text{in shorthand} \quad \boxed{\langle \mathbf{e_i} \mid \mathbf{e_j} \rangle = \delta_{ij}}$$

<u>Remark:</u> Recall that the *Kronecker* symbol δ_{ij} is equal to one if $i = j$ and zero otherwise.

An important issue is how to construct an orthogonal basis for a given LVS. Given a set of N linearly independent vectors $|\mathbf{v_1}\rangle, \dots, |\mathbf{v_N}\rangle$ one can always construct a set $|\mathbf{e_1}\rangle, \dots, |\mathbf{e_N}\rangle$ of orthonormal vectors by the so-called **Schmidt procedure** (also known as the Gram-Schmidt procedure in the English literature).

[1] Note however that there are infinitely many orthonormal bases of $\mathbb{R}^2$

5.1 Schmidt procedure

- **Step 1:** Normalize one of the initial vectors, e.g.

$$|\mathbf{e_1}\rangle \equiv \frac{1}{\sqrt{\langle \mathbf{v_1} \mid \mathbf{v_1}\rangle}} |\mathbf{v_1}\rangle \tag{46}$$

- **Step 2:** Now form a linear combination of $|\mathbf{v_2}\rangle$ and $|\mathbf{e_1}\rangle$ that is orthogonal to $|\mathbf{e_1}\rangle$:

$$|\mathbf{v_2}'\rangle = |\mathbf{v_2}\rangle - \langle \mathbf{e_1} \mid \mathbf{v_2}\rangle |\mathbf{e_1}\rangle . \tag{47}$$

This vector is orthogonal to $|\mathbf{e_1}\rangle$ because

$$\langle \mathbf{e_1}| \, \mathbf{v_2}'\rangle = \langle \mathbf{e_1}| \left\{ |\mathbf{v_2}\rangle - \langle \mathbf{e_1} \mid \mathbf{v_2}\rangle |\mathbf{e_1}\rangle \right\} = \langle \mathbf{e_1} \mid \mathbf{v_2}\rangle - \langle \mathbf{e_1} \mid \mathbf{v_2}\rangle \underbrace{\langle \mathbf{e_1} \mid \mathbf{e_1}\rangle}_{=1} = 0. \tag{48}$$

Now normalize the vector $|\mathbf{v_2'}\rangle$ so that we end up with a vector of unit length

$$|\mathbf{e_2}\rangle \equiv \frac{|\mathbf{v_2}'\rangle}{\sqrt{\langle \mathbf{v_2}' \mid \mathbf{v_2}'\rangle}}. \tag{49}$$

After Step 2 we have two normalized vectors $|\mathbf{e}_1\rangle$, $|\mathbf{e}_2\rangle$, which are orthogonal.

- **Step 3:** Now form a linear combination of $|\mathbf{v_3}\rangle$, $|\mathbf{e_1}\rangle$, $|\mathbf{e_2}\rangle$ that is orthogonal to $|\mathbf{e_1}\rangle$ and $|\mathbf{e_2}\rangle$:

$$|\mathbf{v_3}'\rangle = |\mathbf{v_3}\rangle - \langle \mathbf{e_1} \mid \mathbf{v_3}\rangle |\mathbf{e_1}\rangle - \langle \mathbf{e_2} \mid \mathbf{v_3}\rangle |\mathbf{e_2}\rangle . \tag{50}$$

We can check explicitly that this vector is indeed orthogonal to $|\mathbf{e}_1\rangle$ and $|\mathbf{e}_2\rangle$ by calculating the scalar products with these vectors

$$\langle \mathbf{e_1} \mid \mathbf{v_3}'\rangle = \langle \mathbf{e_1} \mid \mathbf{v_3}\rangle - \langle \mathbf{e_1} \mid \mathbf{v_3}\rangle = 0 \qquad \text{because } \langle \mathbf{e_i} \mid \mathbf{e_j}\rangle = \delta_{ij} \tag{51}$$

$$\langle \mathbf{e_2} \mid \mathbf{v_3}'\rangle = \langle \mathbf{e_2} \mid \mathbf{v_3}\rangle - \langle \mathbf{e_2} \mid \mathbf{v_3}\rangle = 0 \qquad \text{''} \tag{52}$$

Finally we normalize the vector

$$|\mathbf{e_3}\rangle = \frac{|\mathbf{v_3'}\rangle}{\sqrt{\langle \mathbf{v_3'} \mid \mathbf{v_3'}\rangle}}. \tag{53}$$

Now we have three normalized vectors $|\mathbf{e}_1\rangle$, $|\mathbf{e}_2\rangle$ and $|\mathbf{e}_3\rangle$, which are orthogonal.

Continue this procedure.

Example:

$$|\mathbf{v_1}\rangle = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} \quad , \quad |\mathbf{v_2}\rangle = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \quad , \quad |\mathbf{v_3}\rangle = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \tag{54}$$

Step 1: we take the first vector and normalize it

$$\boxed{|\mathbf{e_1}\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} .} \tag{55}$$

Step 2: now we construct a vector orthogonal to $|\mathbf{e_1}\rangle$ from $|\mathbf{v_2}\rangle$

$$|\mathbf{v_2}'\rangle = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} - \left[(1,0,0) \cdot \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \right] \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} . \tag{56}$$

This vector is already normalized, so that

$$\boxed{|\mathbf{e_2}\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} .} \tag{57}$$

Step 3: now we construct a vector orthogonal to $|\mathbf{e_1}\rangle$ and $|\mathbf{e_2}\rangle$ from $|\mathbf{v_3}\rangle$

$$|\mathbf{v_3'}\rangle \quad = \quad \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \left[(1,0,0) \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right] \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - \left[(0,1,0) \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right] \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} . \tag{58}$$

This vector is already normalized and hence

$$\boxed{|\mathbf{e_3}\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} .} \tag{59}$$

This completes the procedure.

6 Linear Operators and Matrices

You are all familiar with matrices acting on vectors. A matrix is something that acts on vectors and gives other vectors as results, e.g.

$$\begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} . \tag{60}$$

In our new language this is an example of a **linear operator**. A linear operator $\mathbf{A}$ acts like

$$\mathbf{A} \left| \mathbf{v} \right\rangle = \left| \mathbf{v}' \right\rangle \ , \tag{61}$$

where $\left|\mathbf{v}\right\rangle$ and $\left|\mathbf{v}'\right\rangle$ are vectors in our LVS. $\mathbf{A}$ is **linear** because we require that

$$\mathbf{A} \left\{ \alpha \left| \mathbf{a} \right\rangle + \beta \left| \mathbf{b} \right\rangle \right\} = \alpha \left\{ \mathbf{A} \left| \mathbf{a} \right\rangle \right\} + \beta \left\{ \mathbf{A} \left| \mathbf{b} \right\rangle \right\} . \tag{62}$$

Linear operators can be added and multiplied (just like matrices)

- $\mathbf{C} = \mathbf{A} + \mathbf{B}$ acts on **any** vector like

$$\mathbf{C} \left| \mathbf{a} \right\rangle = \mathbf{A} \left| \mathbf{a} \right\rangle + \mathbf{B} \left| \mathbf{a} \right\rangle \tag{63}$$

- $\mathbf{D} = \mathbf{A} \cdot \mathbf{B}$ is defined through its action on any vector as

$$\mathbf{D} \left| \mathbf{a} \right\rangle = \mathbf{A} \left(\mathbf{B} \left| \mathbf{a} \right\rangle \right) \tag{64}$$

$\mathbf{C}$ and $\mathbf{D}$ are linear operators as well.

6.1 Commutator

In general it matters in which order we multiply matrices, i.e.

$$\mathbf{AB} - \mathbf{BA} \neq 0 \ . \tag{65}$$

For example

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 2 \\ -2 & 0 \end{pmatrix} . \tag{66}$$

$\mathbf{AB} - \mathbf{BA}$ is called the **commutator** of the linear operators $\mathbf{A}$ and $\mathbf{B}$ and is denoted by

$$\boxed{[\mathbf{A}, \mathbf{B}] \equiv \mathbf{AB} - \mathbf{BA}.} \tag{67}$$

6.2 Functions of Operators

Having defined the product of operators, we can of course also define arbitrary powers of an operator

$$\mathbf{A}^m = \underbrace{\mathbf{A}\cdot\mathbf{A}\cdot\mathbf{A}\cdot\ldots\cdot\mathbf{A}}_{m\text{times}} . \tag{68}$$

We can then define functions of operators, for example

$$e^{\mathbf{A}} \equiv 1 + \mathbf{A} + \frac{1}{2}\mathbf{A}^2 + \frac{1}{3!}\mathbf{A}^3 + \cdots = \sum_{n=0}^{\infty} \frac{1}{n!}\mathbf{A}^n \ . \tag{69}$$

Example: Let us consider the so-called Pauli matrix σ^x, which is defined as

$$\sigma^x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \ . \tag{70}$$

It is straightforward to work out products of Pauli matrices

$$\begin{aligned} \sigma^x\sigma^x &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \equiv \mathbf{I} = (\sigma^x)^4 = (\sigma^x)^6 = \cdots = (\sigma^x)^{2n} \ , \\ \sigma^x &= (\sigma^x)^3 = (\sigma^x)^5 = \cdots = (\sigma^x)^{2n+1} \ . \end{aligned} \tag{71}$$

This allows us to calculate the complex exponential of σ^x

$$\exp(i\alpha\sigma^x) = \sum_{n=0}^{\infty} \frac{(i\alpha)^n}{n!}(\sigma^x)^n = \sum_{n=0}^{\infty} (-1)^n \frac{\alpha^{2n}}{(2n)!}(\sigma^x)^{2n} + i(-1)^n \frac{\alpha^{2n+1}}{(2n+1)!}(\sigma^x)^{2n+1} \tag{72}$$

$$= \left(\sum_{n=0}^{\infty} (-1)^n \frac{\alpha^{2n}}{(2n)!}\right)\mathbf{I} + \left(\sum_{n=0}^{\infty} (-1)^n \frac{\alpha^{2n+1}}{(2n+1)!}\right)\sigma^x \tag{73}$$

$$= \cos(\alpha)\ \mathbf{I} + i\sin(\alpha)\ \sigma^x \ . \tag{74}$$

Exercise: Work out the complex exponentials

$$\exp(i\alpha\sigma^y) \ , \qquad \exp(i\alpha\sigma^z) \ , \tag{75}$$

where σ^y and σ^z are the other two Pauli matrices

$$\sigma^y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \ , \qquad \sigma^z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \ . \tag{76}$$

Exercise: Express the product

$$\exp(i\alpha\sigma^z)\ \sigma^x \exp(-i\alpha\sigma^z) \tag{77}$$

in terms of multiple commutators of Pauli matrices and then evaluate it.

7 Linear Operators As Matrices

Let $B = \{|\mathbf{e_1}\rangle, \ldots, |\mathbf{e_N}\rangle\}$ be an orthonormal basis [2] of the LVS V. Then we have

$$\begin{aligned} \langle \mathbf{e_i} \mid \mathbf{e_j} \rangle &= \delta_{ij} \ , && (78) \\ |\mathbf{a}\rangle &= \sum_{j=i}^{N} a_j |\mathbf{e_j}\rangle \qquad \text{for any } |\mathbf{a}\rangle \, \text{in} V \ . && (79) \end{aligned}$$

Because $\langle \mathbf{e_k} \mid \mathbf{a} \rangle = \sum a_j \langle \mathbf{e_k} \mid \mathbf{e_j} \rangle = a_k$ we can **represent** $|\mathbf{a}\rangle$ by its components in the basis B:

$$|\mathbf{a}\rangle \mapsto \begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix} = \begin{pmatrix} \langle \mathbf{e_1} \mid \mathbf{a} \rangle \\ \vdots \\ \langle \mathbf{e_N} \mid \mathbf{a} \rangle \end{pmatrix} \tag{80}$$

Now consider a linear operator $\mathbf{A}$. It acts on a vector $|\mathbf{a}\rangle$ and produces another vector $|\mathbf{a}'\rangle$

$$|\mathbf{a}'\rangle = \mathbf{A} |\mathbf{a}\rangle \ . \tag{81}$$

By taking scalar products of equation (81) with the vectors $\langle \mathbf{e_j}|$ we obtain N equations

$$a'_j = \langle \mathbf{e_j} \mid \mathbf{a}' \rangle = \langle \mathbf{e_j}| \mathbf{A} |\mathbf{a}\rangle = \sum_{l=1}^{N} \langle \mathbf{e_j}| \mathbf{A} |\mathbf{e_l}\rangle a_l \ . \tag{82}$$

The second identity in (82) is obtained by expressing the vector $|\mathbf{a}\rangle$ as a linear combination of the basis vectors $|\mathbf{e_l}\rangle$

$$|\mathbf{a}\rangle = \sum_l a_l |\mathbf{e_l}\rangle \ . \tag{83}$$

If we define a matrix with "**matrix elements in the basis** B"

$$A_{jl} = \langle \mathbf{e_j}| \mathbf{A} |\mathbf{e_l}\rangle \tag{84}$$

then (81) can be written in form of a matrix equation

$$\boxed{\begin{pmatrix} a'_1 \\ \vdots \\ a'_N \end{pmatrix} = \begin{pmatrix} A_{11} & \ldots & A_{1N} \\ \vdots & \ddots & \\ A_{N1} & \ldots & A_{NN} \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix}} \tag{85}$$

What we have just shown is that by choosing a particular basis we can "**represent**" linear operators by matrices! The expression for the matrix representing a given linear operator in a particular basis of course depends on the basis.

[2] We can always orthogonalize any basis by the Schmidt procedure.

Given the relationship between linear operators and matrices we obtain the following rules for adding & multiplying linear operators

$$\mathbf{C} = \mathbf{A} + \mathbf{B} \Leftrightarrow \quad C_{ij} = A_{ij} + B_{ij} \tag{86}$$

$$\mathbf{C} = \mathbf{A} \cdot \mathbf{B} \Leftrightarrow \quad C_{ij} = \sum_l A_{il} B_{lj} \tag{87}$$

In (87) we recognize the usual rule of matrix multiplication.

Example: Let us return to the LVS V of arrows in the plane. The vectors in V are arrows, which we however may represent as two-component vectors with respect to the usual Cartesian basis

$$|\mathbf{e_1}\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \ , \quad |\mathbf{e_2}\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} . \tag{88}$$

A general vector $|\mathbf{a}\rangle$ is expressed as

$$|\mathbf{a}\rangle = a_1 |\mathbf{e_1}\rangle + a_2 |\mathbf{e_2}\rangle \Longrightarrow |\mathbf{a}\rangle = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} . \tag{89}$$

A linear operator $\mathbf{A}$ maps arrows to other arrows in the plane and hence "exists" independently of any representation in terms of a particular basis. In order to represent $\mathbf{A}$ in our basis, we merely need to determine how it acts on the basis vectors. Expressing the vectors $\mathbf{A}|\mathbf{e_j}\rangle$ as linear combinations of the basis vectors *defines* a matrix with elements A_{ij}, which are simply the coefficients of this expansion

$$\mathbf{A}|\mathbf{e_j}\rangle = \sum_i A_{ij} |\mathbf{e_i}\rangle . \tag{90}$$

Taking scalar products of these equations with $\langle \mathbf{e_i}|$ and using the orthonormality of the basis, we arrive at the general expression (84) for the matrix elements. It is clear that the procedure we have described depends on our choice of basis. If we were to use the orthonormal basis

$$|\mathbf{e'_1}\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \ , \quad |\mathbf{e'_2}\rangle = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \ , \tag{91}$$

then we would represent a given arrow $|\mathbf{a}\rangle$ by a *different* two-component vector

$$|\mathbf{a}\rangle = a'_1 |\mathbf{e'_1}\rangle + a'_2 |\mathbf{e'_2}\rangle \Longrightarrow |\mathbf{a}\rangle = \begin{pmatrix} a'_1 \\ a'_2 \end{pmatrix} . \tag{92}$$

Similarly we would represent the linear operator $\mathbf{A}$ by a different 2×2 matrix, defined by how $\mathbf{A}$ acts on the new basis vectors

$$\mathbf{A}|\mathbf{e'_j}\rangle = \sum_i A'_{ij} |\mathbf{e'_i}\rangle . \tag{93}$$

8 Operations on Square Matrices

Let us now define a number of useful operations on square matrices with in general complex matrix elements.

1. **Trace:**

$$\boxed{\operatorname{tr}\mathbf{A} \equiv \textstyle\sum_{j=1}^{N} A_{jj}}. \tag{94}$$

The trace is the sum of the diagonal matrix elements. It is a (complex) number.

The trace is "**cyclic**":

$$\operatorname{tr}(\mathbf{AB}) = \operatorname{tr}(\mathbf{BA}) \tag{95}$$

$$\operatorname{tr}(\mathbf{ABC}) = \operatorname{tr}(\mathbf{CAB}) \qquad \text{etc.} \tag{96}$$

Proof:

$$\operatorname{tr}\mathbf{AB} = \sum_{j=1}^{N} (\mathbf{AB})_{jj} = \sum_j \sum_k A_{jk} B_{kj} \tag{97}$$

$$\equiv \sum_k \sum_j B_{kj} A_{jk} = \sum_k (\mathbf{BA})_{kk} = \operatorname{tr}\mathbf{BA} \quad \text{QED.} \tag{98}$$

2. **Transpose:**

$\mathbf{A}^{\mathrm{T}}$ is defined as $A^{\mathrm{T}}_{ij} = A_{ji}$ [swaps rows & columns]

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} A_{11} & A_{21} \\ A_{12} & A_{22} \end{pmatrix} \tag{99}$$

3. **Hermitian Conjugate:**

$\mathbf{A}^{\dagger}$; $(A^{\dagger})_{ij} = A^{*}_{ji}$ [transpose & complex conjugate]

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{\dagger} = \begin{pmatrix} A^{*}_{11} & A^{*}_{21} \\ A^{*}_{12} & A^{*}_{22} \end{pmatrix}. \tag{100}$$

4. **Determinant:**

General definition:

Let $P = (P_1 \ldots P_N)$ be a **permutation** of $(1, 2, \ldots, N)$. There are $N!$ different permutations. The **sign** of P is $(-1)^{\alpha}$, where α is the number of pair exchanges needed to arrive at $(P_1, \ldots, P_N)$ starting with $(1, \ldots, N)$.

Example:

$$P_1 = (2,3,1) \qquad \mathrm{sgn}(P_1) = (-1)^2 = 1 \tag{101}$$
$$P_2 = (3,2,1) \qquad \mathrm{sgn}(P_2) = (-1)^1 = -1 \tag{102}$$

The **determinant** of the $N \times N$ matrix $\mathbf{A}$ is

$$\boxed{\det \mathbf{A} = \textstyle\sum_P \mathrm{sgn}(P) A_{1P_1}, \ldots A_{NP_N}} \tag{103}$$

Here the sum is over all permutations $P = (P_1, \ldots, P_N)$ of $(1, \ldots, N)$.

Let us verify that in the case $N = 2$ this agrees with the expression for the determinant we already know. There are two different permutations

$$P \ = \ (1,2) \qquad \mathrm{sgn}\, P = 1 \tag{104}$$
$$\text{or} \quad P' \ = \ (2,1) \qquad \mathrm{sgn}\, P' = -1 \tag{105}$$

Evaluating the general expression above for the case $N = 2$ then gives

$$\det A = \mathrm{sgn}\, P \ A_{1P_1} A_{2P_2} + \mathrm{sgn}\, P' \ A_{1P'_1} A_{2P'_2} = A_{11}A_{22} - A_{21}A_{12}. \tag{106}$$

This is as expected. Let us also prove that adding one row to another does not change the determinant. Consider

$$A = \begin{pmatrix} A_{11} & A_{12} & \ldots & A_{1n} \\ A_{21} & A_{22} & \ldots & A_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ A_{n1} & A_{n2} & \ldots & A_{nn} \end{pmatrix} \tag{107}$$

and

$$B = \begin{pmatrix} A_{11} + \alpha A_{21} & A_{12} + \alpha A_{22} & \ldots & A_{1n} + \alpha A_{2n} \\ A_{21} & A_{22} & \ldots & A_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ A_{n1} & A_{n2} & \ldots & A_{nn} \end{pmatrix}. \tag{108}$$

Then

$$\begin{aligned} \det(B) &= \sum_P \mathrm{sgn}\, P \ \left[A_{1P_1} + \alpha A_{2P_1}\right] A_{2P_2} \ldots A_{NP_N} \\ &= \sum_P \mathrm{sgn}\, P \ A_{1P_1} A_{2P_2} \ldots A_{NP_N} + \sum_P \mathrm{sgn}\, P \ \alpha A_{2P_1} A_{2P_2} \ldots A_{NP_N} \end{aligned} \tag{109}$$

The first sum is just $\det(A)$. The second sum is zero as terms cancel pairwise. Indeed, consider the permutation $P' = (P_2, P_1, \ldots P_n)$. We have $\mathrm{sgn}(P') = -\mathrm{sgn}(P)$

and its contribution to the sum over permutations is $-\operatorname{sgn} P\ \alpha A_{2P_2} A_{2P_1} \dots A_{NP_N}$, which precisely cancels against the contribution from P. This proves that $\det(A) = \det(B)$.

The determinant of the **product** of 2 matrices is

$$\boxed{\det \mathbf{A} \cdot \mathbf{B} = \det \mathbf{A} \det \mathbf{B}} \tag{110}$$

Furthermore

$$\boxed{\det \mathbf{A} \equiv \det \mathbf{A}^{\mathrm{T}}} \tag{111}$$

5. **Inverse of a Matrix**

The inverse of a matrix $\mathbf{A}$ is denoted by $\mathbf{A}^{-1}$ and fulfils

$$\mathbf{A} \cdot \mathbf{A}^{-1} = \mathbf{I} = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} = \mathbf{A}^{-1} \cdot \mathbf{A}$$

For $\mathbf{A}^{-1}$ to exist we need $\boxed{\det \mathbf{A} \neq 0}$.

How to construct the inverse?

$$(\mathbf{A}^{-1})_{ij} = \frac{C^T_{ij}}{\det \mathbf{A}} \tag{112}$$

Here C is the matrix of cofactors. It is related to the matrix M of **minors** by $C_{ij} = (-1)^{i+j} M_{ij}$. The matrix of minors is defined by

$$M_{ij} = \det \begin{pmatrix} A_{11} & \cdots & \cancel{A_{1j}} & \cdots & A_{1N} \\ A_{21} & \cdots & \cancel{A_{2j}} & \cdots & A_{2N} \\ \vdots & & & & \\ \cancel{A_{i1}} & \cdots & \cancel{A_{ij}} & \cdots & \cancel{A_{iN}} \\ \vdots & & & & \\ A_{N1} & \cdots & \cancel{A_{Nj}} & \cdots & A_{NN} \end{pmatrix} \tag{113}$$

(Cross out i^{th} row & j^{th} column).

Inverse of a Product of Matrices:

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \tag{114}$$

8.1 Dual vector to $\mathbf{A}\,|\mathbf{a}\rangle$

Later on we will need to know the dual vector to a vector

$$\mathbf{A}\,|\mathbf{a}\rangle\ . \tag{115}$$

Let us show that the answer is

$$\langle\mathbf{a}|\,\mathbf{A}^\dagger\ . \tag{116}$$

The calculation is easily done in a given orthonormal basis:

$$\begin{aligned}
\mathbf{A}\,|\mathbf{a}\rangle &= \mathbf{A}\Big[\sum_j \langle\mathbf{e_j}\mid\mathbf{a}\rangle\,|\mathbf{e_j}\rangle\Big] = \sum_j \langle\mathbf{e_j}\mid\mathbf{a}\rangle\ \ \mathbf{A}\,|\mathbf{e_j}\rangle \\
&= \sum_j \langle\mathbf{e_j}\mid\mathbf{a}\rangle \sum_l \langle\mathbf{e_l}|\,\mathbf{A}\,|\mathbf{e_j}\rangle\ \ |\mathbf{e_l}\rangle = \sum_j \langle\mathbf{e_j}\mid\mathbf{a}\rangle \sum_l A_{lj}\ \ |\mathbf{e_l}\rangle \\
&= \sum_l \Big[\sum_j A_{lj}\ a_j\Big]\,|\mathbf{e_l}\rangle\ .
\end{aligned} \tag{117}$$

Recalling the definition of dual vectors we now take the dual of the right hand side of (117)

$$\begin{aligned}
\sum_l \Big[\sum_j A_{lj}\ a_j\Big]^*\,\langle\mathbf{e_l}| &= \sum_l \Big[\sum_j A^*_{lj}\ a^*_j\Big]\,\langle\mathbf{e_l}| = \sum_l \Big[\sum_j \left(A^\dagger\right)_{jl}\ a^*_j\Big]\,\langle\mathbf{e_l}| \\
&= \sum_l \Big[\sum_j \langle\mathbf{e_j}|\,A^\dagger\,|\mathbf{e_l}\rangle\ \ \langle\mathbf{a}\mid\mathbf{e_j}\rangle\Big]\,\langle\mathbf{e_l}| = \sum_l \langle\mathbf{a}|\,A^\dagger\,|\mathbf{e_l}\rangle\ \ \langle\mathbf{e_l}| \qquad (118) \\
&= \langle\mathbf{a}|\,A^\dagger\ . \qquad (119)
\end{aligned}$$

The last identity (119) is easily seen to be true by taking scalar products with $|\mathbf{e_i}\rangle$ and then using the orthonormality of the basis, i.e. $\langle\mathbf{e_l}\mid\mathbf{e_i}\rangle = \delta_{il}$. The second to last identity (118) is proved analogously. The calculation we have just done tells us that the dual vector to $\mathbf{A}\,|\mathbf{a}\rangle$ is indeed $\langle\mathbf{a}|\,\mathbf{A}^\dagger$.

9 Change of Basis

For simplicity, consider two orthonormal bases

$$B = \{|\mathbf{e_1}\rangle, \ldots, |\mathbf{e_N}\rangle\}\ , \qquad B' = \{|\mathbf{e'_1}\rangle, \ldots, |\mathbf{e'_N}\rangle\} \tag{120}$$

of the same LVS V. By the definition of a basis, we can express any vector in V as a linear combination of the basis vectors. In particular, we may express the basis vectors $|\mathbf{e_j}\rangle$ as linear combinations of the basis vectors $\{|\mathbf{e'_k}\rangle, k = 1, \ldots N\}$

$$\boxed{|\mathbf{e_j}\rangle = \textstyle\sum_k U_{kj}\,|\mathbf{e'_k}\rangle\ .} \tag{121}$$

The coefficients (which are complex numbers) of these expansions define a $N \times N$ matrix

$$\mathbf{U} = \begin{pmatrix} U_{11} & \cdots & U_{1N} \\ \vdots & & \\ U_{N1} & \cdots & U_{NN} \end{pmatrix} , \tag{122}$$

where the matrix elements are equal to

$$U_{kj} \equiv \langle \mathbf{e'_k} \mid \mathbf{e_j} \rangle \ . \tag{123}$$

The latter expression is obtained by taking the the scalar product of (121) with $\langle \mathbf{e'_k}|$.

Taking the scalar product of (121) with $\langle \mathbf{e_l}|$ we obtain the following set of N^2 equations

$$\delta_{jl} = \sum_k U_{kj} \langle \mathbf{e_l} \mid \mathbf{e'_k} \rangle = \sum_k U_{kj} \left(\langle \mathbf{e'_k} \mid \mathbf{e_l} \rangle \right)^* = \sum_k U_{kj} U^*_{kl} = \sum_k \left(U^\dagger \right)_{lk} U_{kj}. \tag{124}$$

In matrix form these equations read

$$\boxed{\mathbf{U}^\dagger \mathbf{U} = \mathbf{U}\mathbf{U}^\dagger = \mathbf{I} = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix} .}$$

We see that the matrix U has the property, that its Hermitian conjugate is equal to its inverse! Matrices with this property are called **unitary** matrices. The determinant of a unitary matrix is a complex number of unit magnitude as can be shown as follows

$$\det(\mathbf{U}^\dagger \mathbf{U}) = \det \mathbf{U}^\dagger \det \mathbf{U} = |\det \mathbf{U}|^2 = 1 \Rightarrow \boxed{\det \mathbf{U} = e^{i\varphi}} . \tag{125}$$

Here we have used that the determinant of the product of two matrices is equal to the product of determinants and that the determinant of the Hermitian conjugate of a matrix is equal to the complex conjugate of the determinant. The latter can be seen quite easily by using the general definition of the determinant (103).

Given our two bases B and B' we may express a general vector $|\mathbf{a}\rangle$ as a linear combination of the basis vectors in B or as a linear combination of the basis vectors in B'

$$|\mathbf{a}\rangle = \sum_l a_l |\mathbf{e_l}\rangle = \sum_k a'_k |\mathbf{e'_k}\rangle \ . \tag{126}$$

The components a_l and a'_k of the vector $|\mathbf{a}\rangle$ in the two bases are related by the unitary matrix. In order to see this, let us use (121) to rewrite the expression after the first equality sign in equation (126)

$$\sum_l a_l |\mathbf{e_l}\rangle = \sum_{l,k} a_l U_{kl} |\mathbf{e'_k}\rangle \ . \tag{127}$$

If we now take scalar products of (127) and (126) with the basis vectors $\langle \mathbf{e'_k}|$ and use that the basis vectors in B' are orthonormal, we find the following set of N equations

$$\boxed{a'_k = \sum_{l=1}^{N} U_{kl} a_l} \; . \tag{128}$$

These equations may be written in matrix form as

$$\begin{pmatrix} a'_1 \\ \vdots \\ a'_N \end{pmatrix} = \mathbf{U} \begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix} \; . \tag{129}$$

We say that under the change of basis from B to B', vectors transform according to the transformation law (129).

Now we can turn things around and consider the change of basis as a linear transformation on vectors:

$$\vec{a}' = \mathbf{U}\vec{a} \qquad \vec{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix} \; . \tag{130}$$

This transformation has the important property that it leaves the scalar product **invariant**:

$$\boxed{\vec{b}'^{\dagger} \cdot \vec{a}' \equiv \vec{b}^{\dagger} \cdot \vec{a} \; .} \tag{131}$$

You can verify this by direct calculation, but it is clear from the fact that both are equal to $\langle \mathbf{b} \mid \mathbf{a} \rangle$, just expressed in different bases.

Above we have assumed implicitly that we are dealing with a **complex** LVS. If we have a **real** LVS instead, then the matrix elements

$$\langle \mathbf{e'_k} \mid \mathbf{e_j} \rangle \equiv R_{kj} \tag{132}$$

are real and hence

$$\boxed{\mathbf{R}^T\mathbf{R} = \mathbf{I}} \; . \tag{133}$$

Such matrices are called **orthogonal matrices**. The determinant of orthogonal matrices is equal to ± 1. If $\det(\mathbf{R}) = 1$, then $\mathbf{R}$ describes a **rotation**. If $\det(\mathbf{R}) = -1$, then $\mathbf{R}$ describes a **rotation** followed by a **reflection**.

9.1 Transformation of Matrices

Recall that the matrix elements of the linear operator $\mathbf{A}$ in the basis B are

$$A_{ij} = \langle \mathbf{e_i} | \, \mathbf{A} \, | \mathbf{e_j} \rangle \; . \tag{134}$$

In order to work out what the matrix elements of $\mathbf{A}$ in the basis B' are, we express the basis vectors in B in terms of those in B'

$$|\mathbf{e_j}\rangle = \sum_k U_{kj}\,|\mathbf{e'_k}\rangle \ , \tag{135}$$

$$\langle\mathbf{e_i}| = \sum_l (U_{li}\,|\mathbf{e'_l}\rangle)^\dagger = \sum_l \langle\mathbf{e'_l}|\,U^*_{li} \tag{136}$$

Using (135) and (136) in (134) we obtain

$$A_{ij} = \sum_{l,k} \underbrace{\langle\mathbf{e'_l}|\,\mathbf{A}\,|\mathbf{e'_k}\rangle}_{\mathbf{A}'_{lk}} U^*_{li}U_{kj} = \sum_{l,k} U^\dagger_{il}A'_{lk}U_{kj} \tag{137}$$

In matrix form these equations read

$$\boxed{\mathbf{A} = \mathbf{U}^\dagger\mathbf{A}'\mathbf{U}} \tag{138}$$

$$\boxed{\mathbf{A}' = \mathbf{U}\mathbf{A}\mathbf{U}^\dagger} \tag{139}$$

This tells us how the operator $\mathbf{A}$ looks like in the basis B'.

9.2 Change of Basis: Example in $\mathbb{R}^2$

Consider the two orthonormal bases of $\mathbb{R}^2$ shown in Fig7. Let us represent the first basis, which we call B, as

$$|\mathbf{e_1}\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \ , \quad |\mathbf{e_2}\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \ . \tag{140}$$

The second basis, denoted by B', is obtained by a rotation by an angle θ and hence may be represented as

$$|\mathbf{e'_1}\rangle = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \ , \quad |\mathbf{e'_2}\rangle = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix} \ . \tag{141}$$

The orthogonal matrix relating the basis can be calculated from our general formula (132)

$$R_{kj} = \langle\mathbf{e'_k} \mid \mathbf{e_j}\rangle \quad \Rightarrow \quad \mathbf{R} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \ . \tag{142}$$

This is obviously **a rotation matrix**! Its transpose is equal to

$$\mathbf{R}^T = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \ , \tag{143}$$

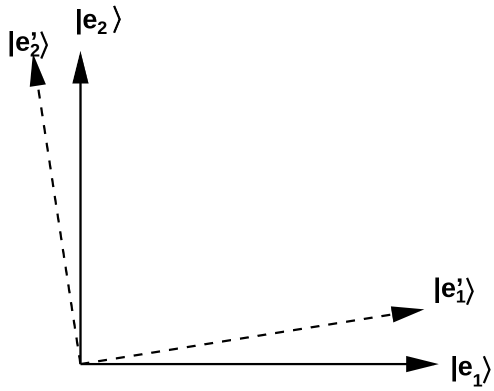


Figure 7: Two orthornormal bases of $\mathbb{R}^2$ related by a rotation.

and you can easily verify that $\mathbf{R}\mathbf{R}^T = I$. Furthermore $\det(\mathbf{R}) = 1$. A general vector transform as

$$\begin{pmatrix} a_1' \\ a_2' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} . \tag{144}$$

This agrees with what we already know. Let us now see what happens with matrices. Let's consider the linear operator $\mathbf{A}$, which acts on vectors by reflecting them along the horizontal axis on the plane, see Fig. 8 In the basis B the linear operator $\mathbf{A}$ is represented

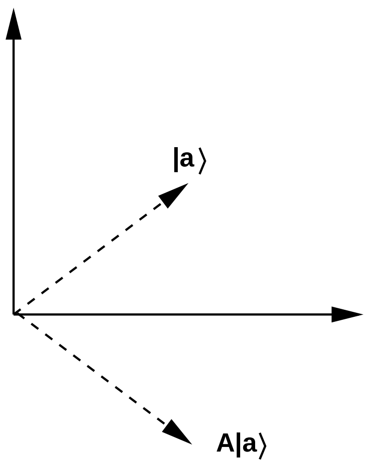


Figure 8: Action of the linear operator $\mathbf{A}$ on vectors in the plane.

as the matrix

$$A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} . \tag{145}$$

In the basis B' is it clearly more difficult to represent $\mathbf{A}$. From our general discussion we know that

$$A' = R \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} R^T \ . \tag{146}$$

It is now a simple matter to interpret this formula. The action of the linear operator is always the same, irrespective of the basis we represent it in. If we are given a vector represented in the basis B' and want to determine how $\mathbf{A}$ acts on it, we first rotate it by an angle $-\theta$, which maps the basis B' onto the basis B. This is achieved by multiplying with the matrix R^T. In the basis B we already know how $\mathbf{A}$ acts, namely via multiplication by the matrix A. Finally we have to rotate the resulting vector back to the basis B', which is achieved by matrix multiplication with the rotation matrix R.

The general case works in the same way. The only additional complication is that in a complex vector space, a "rotation" looks a bit different and is represented by a unitary matrix.

The above discussion for $\mathbb{R}^2$ is readily generalised to $\mathbb{R}^3$. In particular, a rotation around the z-axis by an angle θ is represented by the 3×3 matrix

$$\mathbf{R} = \begin{pmatrix} \cos\theta & \sin\theta & \\ -\sin\theta & \cos\theta & \\ & & 1 \end{pmatrix} . \tag{147}$$

9.3 Generalization to Arbitrary Basis Transformations

If $B = \{|\mathbf{v_1}\rangle, \ldots, |\mathbf{v_N}\rangle\}$, $B' = \{|\mathbf{v'_1}\rangle, \ldots, |\mathbf{v'_N}\rangle\}$ are bases of the LVS V but are not orthonormal, then

$$|\mathbf{a}\rangle = \sum_{j=1}^{N} a_j \, |\mathbf{v_j}\rangle = \sum_{j=1}^{N} a'_j \, |\mathbf{v'_j}\rangle \tag{148}$$

and

$$\begin{pmatrix} a'_1 \\ \vdots \\ a'_N \end{pmatrix} = \begin{pmatrix} S_{11} & \ldots & S_{1N} \\ \vdots & & \\ S_{N1} & \ldots & S_{NN} \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix} \tag{149}$$

where $\mathbf{S}$ is defined by $|\mathbf{v_j}\rangle = \sum_{i=1}^{N} S_{ij} \, |\mathbf{v'_i}\rangle$. It is easy to show that $\det \mathbf{S} \neq 0$. This is a consequence of B and B' both being bases, i.e. all vectors in B (and similarly in B') being linearly independent. Matrices transform according to

$$\boxed{\mathbf{A'} = \mathbf{S}\mathbf{A}\mathbf{S}^{-1}} \tag{150}$$

At this point we will "digress" a bit from the natural flow of topics and talk about

9.4 Co-ordinate Transformations

Consider the real LVS made from "**4-vectors**"

$$|\mathbf{x}\rangle = \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} . \tag{151}$$

4-vectors can be added and multiplied by a real number and form a linear vector space. We now **define** the scalar product between two 4-vectors as

$$\begin{aligned} \langle \tilde{\mathbf{x}} \mid \mathbf{x} \rangle &= c^2 t\tilde{t} - x\tilde{x} - y\tilde{y} - z\tilde{z} \\ &= (\, c\tilde{t}, \quad \tilde{x}, \quad \tilde{y}, \quad \tilde{z}\,) \underbrace{\begin{pmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix}}_{\text{metric}} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} . \end{aligned} \tag{152}$$

Note that this is not **positive definite**, i.e. we don't always have $\langle \mathbf{x} \mid \mathbf{x} \rangle \geq 0$. The 4×4 matrix that appear in the definition of the scalar product is called a **metric**.

Going from one reference frame to another corresponds to a basis transformation in our linear vector space. In the particular case where one frames moves with respect to the other along the x-direction with a velocity v, we have

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \gamma & -\beta\gamma & & \\ -\beta\gamma & \gamma & & \\ & & 1 & \\ & & & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} . \tag{153}$$

Here $\beta = \frac{v}{c}$ and $\gamma = \frac{1}{\sqrt{1-\beta^2}}$. You have already encountered these four equations last year in the Special Relativity course. They represent a **Lorentz transformation**. A very important feature Lorentz transformations is that they leave the scalar product invariant

$$\langle \mathbf{y}' \mid \mathbf{x}' \rangle \equiv \langle \mathbf{y} \mid \mathbf{x} \rangle \ . \tag{154}$$

This shows that Lorentz transformation are similar to rotations!

10 Eigenvalues and Eigenvectors

Let $\mathbf{A}$ be a linear operator acting on a linear vector space V. The vector $|\mathbf{v}\rangle$ is called an **eigenvector** of $\mathbf{A}$ if

$$\mathbf{A}\,|\mathbf{v}\rangle = \lambda\,|\mathbf{v}\rangle \tag{155}$$

for some complex number λ. λ is called the **eigenvalue** corresponding to the eigenvector $|\mathbf{v}\rangle$. How to determine the eigenvalues for a given operator $\mathbf{A}$? Let us rewrite (155) as

$$[\mathbf{A} - \lambda I]\,|\mathbf{v}\rangle = 0\ . \tag{156}$$

Equation (156) can have a non-trivial solution only if

$$\det(\mathbf{A} - \lambda I) = 0. \tag{157}$$

If $\mathbf{A}$ is an $N \times N$ matrix, then (157) is a N^{th} order polynomial equation in λ. The polynomial (in λ) $\det(\mathbf{A} - \lambda)$ is called **characteristic polynomial** of the matrix $\mathbf{A}$.

Example:

$$\mathbf{A} = \sigma^x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\ . \tag{158}$$

The eigenvalues are determined from the equation

$$\det(\mathbf{A} - \lambda I) = \det\begin{pmatrix} -\lambda & 1 \\ 1 & -\lambda \end{pmatrix} = \lambda^2 - 1\ . \tag{159}$$

Hence we find that $\lambda_{1,2} = \pm 1$. In order to find the corresponding eigenvectors we insert our results for the eigenvalues back into (155) and then solve the resulting linear system of equations.

- Eigenvector for $\lambda_1 = 1$.

  For λ_1 we need to solve the matrix equation

  $$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}\ . \tag{160}$$

  This implies that $u_1 = u_2$ and the corresponding **normalized** eigenvector is

  $$|\mathbf{v_1}\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}\ . \tag{161}$$

  We have

  $$\mathbf{A}\,|\mathbf{v_1}\rangle = |\mathbf{v_1}\rangle\ , \qquad \langle \mathbf{v_1} \mid \mathbf{v_1}\rangle = 1. \tag{162}$$

- Eigenvector for $\lambda_2 = -1$.

  For λ_- we find
$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = - \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \ . \tag{163}$$
  This implies that $w_1 = -w_2$ and the corresponding normalized eigenvector is
$$|\mathbf{v_2}\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \ . \tag{164}$$
  It fulfils
$$\mathbf{A} \, |\mathbf{v_2}\rangle = - \, |\mathbf{v_2}\rangle \ , \qquad \langle \mathbf{v_2} \mid \mathbf{v_2} \rangle = 1. \tag{165}$$

Interestingly we find that
$$\langle \mathbf{v_1} \mid \mathbf{v_2} \rangle = 0 \ . \tag{166}$$
This means that the two eigenvectors of σ^x form an orthonormal basis [3] of the linear vector space $\mathbb{C}^2$. Let us calculate $\mathbf{A}$ in this basis:
$$A'_{ij} = \langle \mathbf{v_i} | \, \mathbf{A} \, | \mathbf{v_j} \rangle \quad \Rightarrow A' = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \ . \tag{167}$$

Observation: in the basis of its eigenvectors $\mathbf{A}$ is a diagonal matrix.

Using our results on basis transformations we can work out the unitary transformation that maps the usual Cartesian basis onto the eigenbasis of $\mathbf{A}$. We find
$$A' = RAR^T \ , \qquad R = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \ , \quad RR^T = 1 \ . \tag{168}$$
We say that $\mathbf{A}$ *can be diagonalized* (by an orthogonal transformation. Interestingly, the columns of the matrix R are equal to the eigenvectors of $\mathbf{A}$, i.e.
$$R = \left(\, |\mathbf{v_2}\rangle \quad |\mathbf{v_1}\rangle \, \right) \ . \tag{169}$$

10.1 Normal, Hermitian and Symmetric Matrices

Our example raises the immediate question, whether **any** matrix can be diagonalized. The answer is no. However, so-called **normal matrices** can be diagonalized. These matrices have the defining property that
$$AA^\dagger = A^\dagger A \ . \tag{170}$$
We will focus on two particular types of normal matrices as they are especially important.

[3] They are orthogonal and hence linearly independent and as the dimension of the underlying vector space is two, they also form a basis.

- Hermitian Matrices

  These are defined by the property

$$A = A^\dagger \ , \tag{171}$$

  i.e. a Hermitian matrix is identical to its Hermitian conjugate matrix.

- Real Symmetric Matrices

  These are defined by the property

$$A = A^T \ , \tag{172}$$

  i.e. a symmetric matrix is identical to its transposed matrix. Clearly real symmetric matrices are a special case of Hermitian matrices: a Hermitian matrix with only real entries is symmetric. Because of this fact we will concentrate on Hermitian matrices in what follows.

11 Hermitian Matrices

Hermitian matrices are defined by the property

$$A = A^\dagger \ , \qquad \Leftrightarrow \qquad \boxed{A_{ij} = A^*_{ji}} \ . \tag{173}$$

For example, the 2×2 matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is Hermitian, if a, d are real and $b = c^*$. The following theorem is of great importance in Quantum Mechanics:

Theorem: (i) All eigenvalues of a Hermitian matrix are real.
(ii) Eigenvectors corresponding to different eigenvalues are orthogonal.

Proof using Dirac notation: (i) The eigenvalue equation for the Hermitian matrix $\mathbf{A}$ reads

$$\mathbf{A} \left| \mathbf{a} \right\rangle = \lambda \left| \mathbf{a} \right\rangle . \tag{174}$$

Taking the scalar product of this equation with $\langle \mathbf{a} |$, we obtain

$$\langle \mathbf{a} | \, \mathbf{A} \, | \mathbf{a} \rangle = \lambda \, \langle \mathbf{a} \mid \mathbf{a} \rangle . \tag{175}$$

Now we take the complex conjugate of equation (175)

$$\langle \mathbf{a} | \, \mathbf{A}^\dagger \, | \mathbf{a} \rangle = \lambda^* \, \langle \mathbf{a} \mid \mathbf{a} \rangle . \tag{176}$$

However, $\mathbf{A} = \mathbf{A}^\dagger$ and hence :

$$\langle \mathbf{a} | \, \mathbf{A} \, | \mathbf{a} \rangle = \lambda^* \, \langle \mathbf{a} \mid \mathbf{a} \rangle \, . \tag{177}$$

Subtracting (177) from (175) we obtain

$$(\lambda - \lambda^*) \, \langle \mathbf{a} \mid \mathbf{a} \rangle = 0. \tag{178}$$

As $\langle \mathbf{a} \mid \mathbf{a} \rangle \neq 0$ by assumption we must have that $\lambda = \lambda^*$ and hence the eigenvalues must all be real.

Proof using matrix notation: If you are uncomfortable with the Dirac notation, let's go through exactly the same steps in matrix notations. We start with the eigenvalue equation, which reads

$$A\vec{a} = \lambda \vec{a} \ . \tag{179}$$

The j^{th} component of this equation is

$$\sum_k A_{jk} \ a_k = \lambda a_j \ . \tag{180}$$

Taking the scalar product with $\vec{a}^\dagger$ we have

$$\sum_j \sum_k a_j^* \ A_{jk} \ a_k = \lambda \sum_j |a_j|^2 \ . \tag{181}$$

The complex conjugate of (181) is

$$\sum_j \sum_k a_j \ A_{jk}^* \ a_k^* = \lambda^* \sum_j |a_j|^2 \ . \tag{182}$$

The RHS of this equation can be rewritten by simply changing the summation indices

$$\sum_j \sum_k a_j \ A_{jk}^* \ a_k^* = \sum_k \sum_j a_k \ A_{kj}^* \ a_j^* = \sum_k \sum_j a_j^* \ A_{kj}^* \ a_k \ . \tag{183}$$

Subtracting (182) from (181) and using (183) we obtain

$$\sum_{j,k} a_j^* \left[A_{jk} - A_{kj}^* \right] a_k = (\lambda - \lambda^*) \sum_j |a_j|^2 \ . \tag{184}$$

However, the LHS of the equation vanishes, because $A = A^\dagger$ and hence

$$A_{jk} = A_{kj}^* \ . \tag{185}$$

This then implies that $\lambda = \lambda^*$ and hence the eigenvalue is real.

(ii) Let us now consider two eigenvectors for different eigenvalues $\lambda \neq \lambda'$

$$\mathbf{A}\,|\mathbf{a}\rangle = \lambda\,|\mathbf{a}\rangle\;, \tag{186}$$

$$\mathbf{A}\,|\mathbf{a}'\rangle = \lambda'\,|\mathbf{a}'\rangle\;. \tag{187}$$

Taking the scalar product of (186) with $\langle\mathbf{a}'|$ and of (187) with $\langle\mathbf{a}|$ we find

$$\langle\mathbf{a}'|\,\mathbf{A}\,|\mathbf{a}\rangle = \lambda\,\langle\mathbf{a}' \mid \mathbf{a}\rangle\;, \tag{188}$$

$$\langle\mathbf{a}|\,\mathbf{A}\,|\mathbf{a}'\rangle = \lambda'\,\langle\mathbf{a} \mid \mathbf{a}'\rangle\;. \tag{189}$$

The complex conjugate of equation (189) is

$$\langle\mathbf{a}'|\,\mathbf{A}^\dagger\,|\mathbf{a}\rangle = \lambda'\,\langle\mathbf{a}' \mid \mathbf{a}\rangle\;, \tag{190}$$

where we have used that λ' is real according to part (i) of the theorem, which we have already proved. Now we use that $\mathbf{A}$ is Hermitian and hence $\mathbf{A}^\dagger = \mathbf{A}$, so that

$$\langle\mathbf{a}'|\,\mathbf{A}\,|\mathbf{a}\rangle = \lambda'\,\langle\mathbf{a}' \mid \mathbf{a}\rangle\;. \tag{191}$$

Subtracting (191) from (188) we thus find

$$(\lambda - \lambda')\,\langle\mathbf{a}' \mid \mathbf{a}\rangle = 0. \tag{192}$$

Given that $\lambda \neq \lambda'$ this proves that $|\mathbf{a}\rangle$ and $|\mathbf{a}'\rangle$ are indeed orthogonal

$$\langle\mathbf{a}' \mid \mathbf{a}\rangle = 0. \tag{193}$$

This completes the proof of the theorem.

If we have several eigenvalues that are equal, we can always choose the corresponding eigenvectors to be orthogonal as we can always achieve this by means of the Schmidt procedure.

Our theorem tells us two remarkable facts about Hermitian matrices: all their eigenvalues are real, which is crucial importance for Quantum Mechanics. Many other matrices are diagonalizable as well (normal matrices), but their eigenvalues are generally complex. Secondly, we have proved that the eigenvectors of a Hermitian matrix can be chosen such that they form an orthonormal **basis** of the vector space.

11.1 Diagonalization of Hermitian Matrices

Let us now discuss how to diagonalize a Hermitian $N \times N$ matrix. We begin with the eigenvalue equations

$$\begin{aligned} A\,|\mathbf{e_1}\rangle &= \lambda_1\;|\mathbf{e_1}\rangle\;, \\ A\,|\mathbf{e_2}\rangle &= \lambda_2\;|\mathbf{e_2}\rangle\;, \\ &\vdots \\ A\,|\mathbf{e_N}\rangle &= \lambda_N\;|\mathbf{e_N}\rangle\;. \end{aligned} \tag{194}$$

As we have just proved, we can always arrange things in such a way that the eigenvectors of A form an orthonormal set

$$\langle \mathbf{e_i} \mid \mathbf{e_j} \rangle = \delta_{ij} \ . \tag{195}$$

Let us now form the matrix

$$U^\dagger = \begin{pmatrix} \vdots & \vdots & & \vdots \\ |\mathbf{e_1}\rangle & |\mathbf{e_2}\rangle & \dots & |\mathbf{e_N}\rangle \\ \vdots & \vdots & & \vdots \end{pmatrix} \ , \tag{196}$$

i.e. the columns of $U^\dagger$ are the eigenvectors of A. Then

$$UAU^\dagger = \begin{pmatrix} \lambda_1 & & & & \\ & \lambda_2 & & & \\ & & \lambda_3 & & \\ & & & \ddots & \\ & & & & \lambda_N \end{pmatrix} \ . \tag{197}$$

Proof: Let us first calculate $AU^\dagger$. Using that the vectors $\vec{e}_j$ are eigenvectors of A we obtain

$$AU^\dagger = \begin{pmatrix} \vdots & \vdots & & \vdots \\ \lambda_1 |\mathbf{e_1}\rangle & \lambda_2 |\mathbf{e_2}\rangle & \dots & \lambda_N |\mathbf{e_N}\rangle \\ \vdots & \vdots & & \vdots \end{pmatrix} \ , \tag{198}$$

i.e. the columns of $AU^\dagger$ are the eigenvectors of A times their corresponding eigenvalues. Then

$$\begin{aligned} UAU^\dagger &= \begin{pmatrix} \dots & \langle \mathbf{e_1}| & \dots \\ \dots & \langle \mathbf{e_2}| & \dots \\ & \vdots & \\ \dots & \langle \mathbf{e_N}| & \dots \end{pmatrix} \begin{pmatrix} \vdots & \vdots & & \vdots \\ \lambda_1 |\mathbf{e_1}\rangle & \lambda_2 |\mathbf{e_2}\rangle & \dots & \lambda_N |\mathbf{e_N}\rangle \\ \vdots & \vdots & & \vdots \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1 \langle \mathbf{e_1} \mid \mathbf{e_1}\rangle & \lambda_2 \langle \mathbf{e_1} \mid \mathbf{e_2}\rangle & \dots & \lambda_N \langle \mathbf{e_1} \mid \mathbf{e_N}\rangle \\ & & \vdots & \\ \lambda_1 \langle \mathbf{e_N} \mid \mathbf{e_1}\rangle & \lambda_2 \langle \mathbf{e_N} \mid \mathbf{e_2}\rangle & \dots & \lambda_N \langle \mathbf{e_N} \mid \mathbf{e_N}\rangle \end{pmatrix} \ . \end{aligned} \tag{199}$$

Using orthogonality $\langle \mathbf{e_j} \mid \mathbf{e_k} \rangle = \delta_{jk}$ we arrive at (197). **Q.E.D.**

We can understand the above formulas in terms of a basis transformation as follows. Let us say that we are given a representation of the linear operator $\mathbf{A}$ in the "usual" Cartesian basis

$$B = \left\{ |\mathbf{c_1}\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, |\mathbf{c_2}\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, |\mathbf{c_N}\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \right\} \tag{200}$$

We denote the matrix elements with respect to B by

$$A_{ij} \ . \tag{201}$$

On the other hand, with respect to its eigenbasis $E = \{|\mathbf{e_1}\rangle, \ldots, |\mathbf{e_N}\rangle\}$ the matrix elements are

$$A'_{ij} = \lambda_i \ \delta_{ij} \ . \tag{202}$$

As we have seen previously, there is a unitary transformation described by a matrix U that maps B onto E. The matrix elements of U are defined by

$$U_{kj} \equiv \langle \mathbf{e_k} \mid \mathbf{c_j} \rangle \tag{203}$$

Using the explicit form of the vectors $|\mathbf{c_j}\rangle$ the matrix U can be written as

$$U = \begin{pmatrix} \ldots & \langle \mathbf{e_1}| & \ldots \\ \ldots & \langle \mathbf{e_2}| & \ldots \\ & \vdots & \\ \ldots & \langle \mathbf{e_N}| & \ldots \end{pmatrix} \ . \tag{204}$$

According to our general formula, the matrix A transforms under the basis transformation as

$$A' = UAU^\dagger \ . \tag{205}$$

This is identical to (197). Hence diagonalizing a matrix is equivalent to carrying out a basis transformation to the basis of eigenvectors of A!
Exercise: Show that if H is Hermitian, then

$$e^{i\alpha H} \ , \quad \alpha \in \mathbb{R} \tag{206}$$

is unitary.

11.2 Trace and Determinant of Hermitian Matrices

Let A be a Hermitian matrix $A = A^\dagger$ and let U be the unitary matrix that diagonalizes it

$$UAU^\dagger = \begin{pmatrix} \lambda_1 & & & & \\ & \lambda_2 & & & \\ & & \lambda_3 & & \\ & & & \ddots & \\ & & & & \lambda_N \end{pmatrix} \ . \tag{207}$$

The trace of a Hermitian matrix is equal to the sum of its eigenvalues

$$\mathrm{tr}(A) = \sum_{j=1}^{N} \lambda_j \tag{208}$$

Proof: Clearly we have

$$\mathrm{tr}(UAU^\dagger) = \sum_{j=1}^{N} \lambda_j \ . \tag{209}$$

On the other hand, using the cyclicity of the trace and the fact that U is unitary and hence $UU^\dagger = I$ we have

$$\mathrm{tr}(UAU^\dagger) = \mathrm{tr}(U^\dagger UA) = \mathrm{tr}(A) \ . \tag{210}$$

Q.E.D.

The determinant of a Hermitian matrix is equal to the product of its eigenvalues

$$\det(A) = \prod_{j=1}^{N} \lambda_j \ . \tag{211}$$

Proof: Using that the determinant of a product of matrices is equal to the product of determinants we find that

$$\det(UAU^\dagger) = \det(U)\det(A)\det(U^\dagger) = \det(A)\det(U^\dagger U) = \det(A). \tag{212}$$

Q.E.D.

12 Jordan Normal Form

As we have discussed, not all matrices are diagonalizable. The best one can do for a general, non-normal matrix is to bring it to **Jordan normal form**. This is a "block-diagonal" structure

$$A = \begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_k \end{pmatrix} , \tag{213}$$

where A_j are **Jordan blocks**

$$A_j = \begin{pmatrix} \lambda_j & 1 & & & \\ & \lambda_j & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda_j & 1 \\ & & & & \lambda_j \end{pmatrix} . \tag{214}$$

It is easy to see that the diagonal entries λ_j are eigenvalues of A.

13 Simultaneous Diagonalization of Hermitian Matrices

In Quantum Mechanics one usually wants to diagonalize several Hermitian operators **simultaneously**, which corresponds to e.g. measuring energy and momentum at the same time. This cannot always be achieved. By simultaneous diagonalization we mean to find a basis transformation to a basis in which both operators are diagonal.

The necessary and sufficient condition for two Hermitian operators $\mathbf{A} = \mathbf{A}^\dagger$ and $\mathbf{B} = \mathbf{B}^\dagger$ to be simultaneously diagonlizable is that they **commute**

$$[\mathbf{A}, \mathbf{B}] = AB - BA = 0. \tag{215}$$

Proof: (i) $\mathbf{A}$, $\mathbf{B}$ simultaneously diagonalizable $\Rightarrow$ $[\mathbf{A}, \mathbf{B}] = 0$. This is easy: by assumption there exists an orthonormal basis in which both $\mathbf{A}$ and $\mathbf{B}$ are diagonal, i.e.

$$A = \begin{pmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_N \end{pmatrix}, \qquad B = \begin{pmatrix} \beta_1 & & \\ & \ddots & \\ & & \beta_N \end{pmatrix}. \tag{216}$$

In this basis we obviously have $AB = BA$ and hence $\mathbf{A}$ and $\mathbf{B}$ commute. Note that the property of two operators commuting is independent of the choice of basis. Let $[A, B] = 0$ and consider the commutator in a different orthonormal basis

$$\left(UAU^\dagger\right)\left(UBU^\dagger\right) - \left(UBU^\dagger\right)\left(UAU^\dagger\right) = U[A, B]U^\dagger = 0. \tag{217}$$

(ii) $[\mathbf{A}, \mathbf{B}] = 0 \ \Rightarrow \mathbf{A}$, $\mathbf{B}$ simultaneously diagonalizable.

As $\mathbf{A}$ is Hermitian it is diagonalizable, i.e. in a basis of eigenstates of $\mathbf{A}$ we have the following matrix representation

$$A = \begin{pmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_N \end{pmatrix}. \tag{218}$$

In this basis, $\mathbf{B}$ has the form

$$B = \begin{pmatrix} B_{11} & \dots & B_{1N} \\ \vdots & \vdots & \vdots \\ B_{N1} & \dots & B_{NN} \end{pmatrix}. \tag{219}$$

We now may calculate the commutator $[A, B]$ and then use that by assumption it is zero:

$$AB - BA = \begin{pmatrix} 0 & (\alpha_1 - \alpha_2)B_{12} & (\alpha_1 - \alpha_3)B_{13} & \dots \\ (\alpha_2 - \alpha_1)B_{21} & 0 & (\alpha_2 - \alpha_3)B_{23} & \dots \\ (\alpha_3 - \alpha_1)B_{31} & (\alpha_3 - \alpha_2)B_{32} & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}. \tag{220}$$

Now if all eigenvalues of $\mathbf{A}$ are different, $\alpha_i \neq \alpha_j$, then it follows that $B_{jk} = 0$ for $j \neq k$ and hence B is diagonal, proving (ii). However, if one or several eigenvalues of $\mathbf{A}$ are the same, some of the off-diagonal elements of B can be nonzero. Let us discuss the case where only two eigenvalues are the same, i.e.

$$\alpha_1 = \alpha_2 \ . \tag{221}$$

Then B has the form

$$B = \begin{pmatrix} B_{11} & B_{12} & 0 & 0 & \dots \\ B_{12}^* & B_{22} & 0 & 0 & \dots \\ 0 & 0 & B_{33} & 0 & \dots \\ 0 & 0 & 0 & B_{44} & \dots \\ \vdots & \vdots & \vdots & & \ddots \end{pmatrix} = B^\dagger \ , \tag{222}$$

i.e. B is diagonal except for a 2×2 block corresponding to the eigenvalues $\alpha_1 = \alpha_2$. Let us now consider a basis transformation from the orthonormal eigenbasis of A $\{|\mathbf{e_1}\rangle, \dots, |\mathbf{e_N}\rangle\}$ to a new orthonormal basis

$$\{|\mathbf{e_1'}\rangle, |\mathbf{e_2'}\rangle, |\mathbf{e_3}\rangle, \dots, |\mathbf{e_N}\rangle\}. \tag{223}$$

The unitary matrix representing this basis transforamtion is of the form

$$U = \begin{pmatrix} U_{11} & U_{12} & 0 & 0 & \dots \\ U_{21} & U_{22} & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & \dots \\ & & & \ddots & \end{pmatrix} = \begin{pmatrix} \tilde{U} & 0 \\ 0 & I \end{pmatrix}, \tag{224}$$

where $\tilde{U}$ is a 2×2 matrix and I is the $N-2$ dimensional unit matrix. Clearly, U acts non-trivially only on the 2×2 blocks in the upper left-hand corners or B and A. Let us denote these blocks by $\tilde{B}$ and $\tilde{A}$ respectively. We have

$$\tilde{B} = \begin{pmatrix} B_{11} & B_{12} \\ B_{12}^* & B_{22} \end{pmatrix} \ , \qquad \tilde{A} = \alpha_1 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \ . \tag{225}$$

The basis transformation leaves $\tilde{A}$ unchanged, because it is proportional to the unit matrix

$$\tilde{U}\tilde{A}\tilde{U}^\dagger = \tilde{A}. \tag{226}$$

On the other hand, $\tilde{B}$ transforms as

$$\tilde{U}\tilde{B}\tilde{U}^\dagger \ . \tag{227}$$

However, $\tilde{B}$ is by construction a Hermitian matrix and hence can be diagonalized by a unitary transformation. Hence, by making an appropriate choice for $\tilde{U}$ we can achieve that

$$\tilde{U}\tilde{B}\tilde{U}^\dagger = \begin{pmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{pmatrix}. \tag{228}$$

In this basis $\tilde{A}$ and $\tilde{B}$ are now both diagonal and then so are A and B!

It is obvious that the case where several eigenvalues of A are equal can be treated in a completely analogous way.

This completes the proof. **Q.E.D**.

14 Tensor Product of Vector Spaces

The tensor product is a simple way of constructing a new vector space from two (or more) given ones. Let V and W two linear vector spaces with bases

$$B = \{|\mathbf{v_1}\rangle, |\mathbf{v_2}\rangle, \ldots, |\mathbf{v_N}\rangle\}, \qquad B' = \{|\mathbf{w_1}\rangle, |\mathbf{w_2}\rangle, \ldots, |\mathbf{w_M}\rangle\}. \tag{229}$$

The **tensor product** $V \otimes W$ is a linear vector space of dimension

$$\dim(V \otimes W) = NM\ , \tag{230}$$

with a basis given by

$$\{|\mathbf{v_j}\rangle \otimes |\mathbf{w_k}\rangle\ ,\ j = 1, \ldots, N;\ k = 1, \ldots, M\}. \tag{231}$$

If $\mathbf{A}$ and $\mathbf{C}$ are linear operators on V and W respectively, then $\mathbf{A} \otimes \mathbf{C}$ is a linear operator on $V \otimes W$, acting like

$$(\mathbf{A} \otimes \mathbf{C})\ \ (|\mathbf{v}\rangle \otimes |\mathbf{w}\rangle) \equiv (\mathbf{A}\,|\mathbf{v}\rangle) \otimes (\mathbf{C}\,|\mathbf{w}\rangle)\,. \tag{232}$$

This is the mathematical structure underlying the treatment of "spin" in the Schrödinger equation.

Example: Let us take as our vector spaces two copies of $\mathbb{C}^2$ with bases

$$B = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}\ , \qquad B' = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}. \tag{233}$$

In QM this corresponds to considering two independent spins $\frac{1}{2}$. A basis of $\mathbb{C}^2 \otimes \mathbb{C}^2$ is then given by the *four* vectors

$$\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}. \tag{234}$$

Let us now consider two linear operators [4]

$$\sigma^x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} , \qquad \sigma^z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} , \tag{235}$$

which we take to act on the first an second copy of $\mathbb{C}^2$ respectively. Then we have

$$[\sigma^x \otimes \sigma^z] \left[\begin{pmatrix} a \\ b \end{pmatrix} \otimes \begin{pmatrix} c \\ d \end{pmatrix} \right] \equiv \left[\sigma^x \begin{pmatrix} a \\ b \end{pmatrix} \right] \otimes \left[\sigma^z \begin{pmatrix} c \\ d \end{pmatrix} \right] = \begin{pmatrix} b \\ a \end{pmatrix} \otimes \begin{pmatrix} c \\ -d \end{pmatrix} . \tag{236}$$

15 Difference Equations

Consider the differential equation

$$\frac{d^2 u}{dx^2} = f(x) , \tag{237}$$

where $a \leq x \leq b$. Let us divide the interval $[a, b]$ into N pieces where we imagine N to be quite large

$$x_j = a + j\ h , \quad j = 0, 1, \ldots, N . \tag{238}$$

The size of each sgment is

$$h = \frac{b-a}{N} . \tag{239}$$

Let us further denote

$$u(x_j) = u_j . \tag{240}$$

For large N we may then approximate

$$\left. \frac{du}{dx} \right|_{x=x_j} \approx \frac{u(x_{j+1}) - u(x_j)}{x_{j+1} - x_j} = \frac{u_{j+1} - u_j}{h}. \tag{241}$$

Similarly we obtain

$$\left. \frac{d^2 u}{dx^2} \right|_{x=x_j} \approx \frac{\frac{du}{dx}(x_{j+1}) - \frac{du}{dx}(x_j)}{x_{j+1} - x_j} = \frac{u_{j+2} - 2u_{j+1} + u_j}{h^2}. \tag{242}$$

For large N we can approximate the differential equation (237) by a set of algebraic equations

$$u_{k+2} - 2u_{k+1} + u_k = h^2 f(x_k) , \quad k = 0, 1, \ldots, N-2. \tag{243}$$

[4] In QM these correspond to the operators of spin angular momentum, e.g. $S^x = \frac{\hbar}{2}\sigma^x$.

This can be written in matrix form

$$\begin{pmatrix} -2 & 1 & 0 & 0 & \dots & & \\ 1 & -2 & 1 & 0 & \dots & & \\ 0 & 1 & -2 & 1 & \dots & & \\ & & \dots & & & & \\ 0 & \dots & & & 1 & -2 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_{N-2} \\ u_{N-1} \end{pmatrix} = h^2 \begin{pmatrix} f(0) \\ f(h) \\ \vdots \\ f(b-3h) \\ f(b-2h) \end{pmatrix} - \begin{pmatrix} u_0 \\ 0 \\ \vdots \\ 0 \\ -u_n \end{pmatrix} \tag{244}$$

Imposing boundary conditions such as $u(a) = u(b) = 0$ fixes $u_0 = u_N = 0$ and then we are left with a matrix problem! The function $u(x)$ is represented as a vector and the differential operator as a matrix! We conclude that we can think of differential operators as infinite dimensional matrices.

16 Second Order ODEs

Let us begin with a quick review of second order ordinary differential equations. The most general form is

$$\frac{d^2y}{dx^2} + P(x)\frac{dy}{dx} + Q(x)y = F(x)\ . \tag{245}$$

The general solution to (245) is of the form

$$y(x) = c_1 y_1(x) + c_2 y_2(x) + y_p(x)\ . \tag{246}$$

Here $y_{1,2}(x)$ are **linearly independent** solutions of the **homogeneous equation** (i.e. equation (245) with $F(x) = 0$) and $y_p(x)$ is a **particular** solution of the inhomogeneous equation (245). Here linear independence means that the equation

$$\alpha_1 y_1(x) + \alpha_2 y_2(x) = 0 \tag{247}$$

has the only solution $\alpha_1 = \alpha_2 = 0$.

17 Series Solutions to ODEs

In our forthcoming study of partial differential equations we will encounter several examples of second order ODEs, which don't have solutions in terms of familiar functions. In fact, we may take the ODEs to **define** new functions, which are their solutions.

We will now discuss a general method for constructing series solutions to second order linear ODEs. Note that this topic is not part of the syllabus.

17.1 Warmup Exercise: simple harmonic oscillator

As a warmup exercise consider the ODE for the simple (classical) harmonic oscillator (SHO)

$$\frac{d^2y}{dx^2} + m^2 y = 0 \ . \tag{248}$$

Let us assume that we do not know that the solution to (248) is

$$A \exp(imx) + B \exp(-imx) \ . \tag{249}$$

Let us make the following power-series ansatz for y:

$$y(x) = \sum_{n=0}^{\infty} a_n \ x^n \ . \tag{250}$$

Substituting (250) back into the differential equation (248) we obtain

$$\sum_{n=0}^{\infty} n(n-1)a_n x^{n-2} + m^2 \sum_{n=0}^{\infty} a_n x^n = 0 \ . \tag{251}$$

In order for (251) to hold for all values of x, the coefficient of each power of x must vanish separately. It is useful to collect the coefficients of the various powers of x in a table

	x^0	x^1	x^2	x^3		x^n
$\frac{d^2y}{dx^2}$	$2a_2$	$6a_3$	$12a_4$	$20a_5$	$\ldots$	$(n+2)(n+1)a_{n+2}$
m^2y	m^2a_0	m^2a_1	m^2a_2	m^2a_3	$\ldots$	m^2a_n

We see that we must have

$$a_2 = -\frac{m^2 a_0}{2} \ , \quad a_4 = -\frac{m^2 a_2}{12} \ , \quad \ldots \ , \quad a_{2n+2} = -\frac{m^2 a_{2n}}{(2n+2)(2n+1)} \ , \tag{252}$$

and independently

$$a_3 = -\frac{m^2 a_1}{6} \ , \quad a_5 = -\frac{m^2 a_3}{20} \ , \quad , \ldots \ , \quad a_{2n+1} = -\frac{m^2 a_{2n-1}}{(2n+1)2n} . \tag{253}$$

The relations are called **recurrence relations**. Clearly we can express all even coefficients a_{2n} in terms of a_0

$$a_{2n} = \frac{(-1)^n m^{2n}}{(2n)!} \ a_0 \ . \tag{254}$$

Similarly we can express all odd coefficients a_{2n+1} in terms of a_1

$$a_{2n+1} = \frac{(-1)^n m^{2n}}{(2n+1)!} \ a_1 \ . \tag{255}$$

Substituting these back into our ansatz for $y(x)$ we obtain

$$\begin{aligned} y(x) &= a_0 \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} (mx)^{2n} + \frac{a_1}{m} \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} (mx)^{2n+1} \\ &= a_0 \cos(mx) + \frac{a_1}{m} \sin(mx) , \end{aligned} \tag{256}$$

where we define the factorial of zero to be equal to 1. This is indeed the general solution to our ODE, which we have found in a rather complicated way here. The point is that the series solution method works also in cases, where the ODE does not have a solution in terms of familiar functions. Let us turn to such an example next.

17.2 Bessel's equation

Bessel's equation arises is the study of partial differential equations in problems with cylidrical symmetry. It is given by

$$x^2 \frac{d^2y}{dx^2} + x\frac{dy}{dx} + (x^2 - m^2)y = 0 , \tag{257}$$

where m is an integer. We will see that late how Bessel's equation emerges in physical problems with cylindrical symmetry. Let us again try to find a solution in terms of a power series

$$y(x) = \sum_{n=0}^{\infty} a_n \, x^n . \tag{258}$$

Substituting (258) back into the differential equation (257) we obtain

$$\sum_{n=0}^{\infty} n(n-1)a_n x^n + \sum_{n=0}^{\infty} n a_n x^n + \sum_{n=0}^{\infty} a_n x^{n+2} - m^2 a_n x^n = 0. \tag{259}$$

Let us consider the case $m = 0$ first

	x^0	x^1	x^2	x^3	x^4	x^n
$x^2\frac{d^2y}{dx^2}$	0	0	$2a_2$	$6a_3$	$12a_4$	$n(n-1)a_n$
$x\frac{dy}{dx}$	0	a_1	$2a_2$	$3a_3$	$4a_4$	na_n
x^2y	0	0	a_0	a_1	a_2	a_{n-2}

We see that for the x^1 term to vanish, we must have $a_1 = 0$. However, the we must have $a_3 = 0$ for the x^3 term to vanish and by the same mechanism all odd a_{2n+1} must actually be zero. This leaves us with the even a_{2n}'s.

From the conditions that the coefficient of the x^2 term must be zero we find

$$a_2 = -\frac{a_0}{2^2} . \tag{260}$$

From the conditions that the coefficient of the x^4 term must be zero we find

$$a_4 = -\frac{a_2}{4^2} = \frac{a_0}{4^2 2^2} \ . \tag{261}$$

In general we have

$$\begin{aligned} a_{2n} &= -\frac{a_{2n-2}}{(2n)^2} = \frac{a_{2n-4}}{(2n)^2(2n-2)^2} = -\frac{a_{2n-6}}{(2n)^2(2n-2)^2(2n-4)^2} \\ &= (-1)^n \frac{a_0}{(2n)^2(2n-2)^2(2n-4)^2 \cdots 4^2 2^2} \\ &= (-1)^n \frac{a_0}{2^{2n}(n!)^2} \ . \end{aligned} \tag{262}$$

This gives us the series representation for the Bessel function $J_0(x) = \frac{y(x)}{a_0}$

$$J_0(x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{2^{2n}(n!)^2} \ . \tag{263}$$

Carrying out the analogous analysis for $m = 1, 2, 3, \dots$ leads to series representation for the Bessel functions $J_m(x)$

$$J_m(x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+m}}{2^{2n+m} n! \ (n+m)!} \ . \tag{264}$$

At this point we should realize that we have run into a problem. Bessel's equation is a second order ODE, so for every value of m we should find **two** linearly independent solutions. Why have we missed the second solution? Well, basically our ansatz for $y(x)$ has been too restrictive. The easiest way around the problem is to solve Bessel's equation (257) for generic noninteger m by an ansatz of the form

$$y_{\pm}(x) = z^{\pm m} \sum_{n=0}^{\infty} a_n x^n \ . \tag{265}$$

The recurrence relation is

$$a_n = -\frac{a_{n-2}}{n(n \pm 2m)} \ , \tag{266}$$

and in addition we have $a_1 = 0$. This then leads to **two** solutions $J_m(x)$ and $J_{-m}(x)$ of Bessel's equation. The trouble for integer m is that these solutions become linearly dependent:

$$J_{-m}(x) = (-1)^m J_m(x) \ , \qquad \text{m integer.} \tag{267}$$

The way around this problem is to define for integer m

$$N_m(x) = \lim_{\nu \to m} \frac{\cos(\pi\nu) \ J_\nu(x) - J_{-\nu}(x)}{\sin(\pi\nu)}. \tag{268}$$

These **Neumann** functions provide the second linearly independent solution to Bessel's equation for integer m.

17.3 Associated Legendre Equation

The associated Legendre equation arises is the study of partial differential equations in problems with spherical symmetry. It is of the form

$$(1-x^2)\frac{d^2y}{dx^2} - 2x\frac{dy}{dx} + \left[l(l+1) - \frac{m^2}{1-x^2}\right] y = 0 \ . \tag{269}$$

Here l and m are integers and m ranges between $-l$ and l. Let us first simplify this ODE by separating off a factor $(1-x^2)^{\frac{m}{2}}$, i.e.

$$y = (1-x^2)^{\frac{m}{2}} \ z \ . \tag{270}$$

Substituting this back into (269) we obtain

$$(1-x^2)\frac{d^2z}{dx^2} - 2x(m+1)\frac{dz}{dx} - [m(m+1) - l(l+1)]z = 0 \ . \tag{271}$$

The latter ODE can be solved by making the ansatz

$$z = \sum_{k=0}^{\infty} a_k \ x^k \ . \tag{272}$$

The coefficients a_k fulfil the recurrence relation

$$(k+2)(k+1)a_{k+2} = [(m+k+\frac{1}{2})^2 - \frac{1}{4} - l(l+1)]a_k \ . \tag{273}$$

The special case $m = 0$ of the associated Legendre equation is known as **Legendre's equation**. The recurrence relation in this case is

$$(k+2)(k+1)a_{k+2} = [k(k+1) - l(l+1)]a_k \ . \tag{274}$$

The corresponding solutions are polynomials, the **Legendre polynomials**. The first few of them are

$$P_0(x) = 1 \ , \quad P_1(x) = x \ , \quad P_2(x) = \frac{1}{2}(3x^2-1) \ , \quad P_3(x) = \frac{1}{2}(5x^3-3x) \ . \tag{275}$$

18 Eigenvalues/Eigenfunctions of Differential Operators

As we have seen some time ago, there is a close relationship between differential operators and matrices/linear operators. We now want to explore this connection further.

Consider for example Legendre's equation

$$\underbrace{\left[(1-x^2)\frac{d^2}{dx^2} - 2x\frac{d}{dx}\right]}_{\hat{\mathcal{L}}} y = \underbrace{-l(l+1)}_{\lambda} y \ . \tag{276}$$

This has the form of an **eigenvalue equation**

$$\hat{\mathcal{L}} y = \lambda \ y \tag{277}$$

for the **differential operator** $\hat{\mathcal{L}}$. Here λ is the **eigenvalue** and y is the **eigenfunction** and plays the role of the eigenvector in the linear vector space on which $\hat{\mathcal{L}}$ acts. Note that we can cast the differential equation for the simple harmonic oscilaltor (248) in the same form

$$\hat{\mathcal{L}}_{\text{SHO}} \ = \ \frac{d^2}{dx^2}, \qquad \lambda_{\text{SHO}} = -m^2 \ . \tag{278}$$

You'll see plenty of other examples of this structure in the Quantum Mechanics course. There you have for example

$$\underbrace{\left[-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + V(x)\right]}_{\hat{\mathcal{H}}} \psi = E\psi \ , \tag{279}$$

where the energy E is the eigenvalue of the linear operator $\hat{\mathcal{H}}$, the **Hamiltonian** of the system.

An important role is played by the boundary conditions of the differential equation. In particular, the boundary conditions lead to constraints on the possible eigenvalues: they lead to the **quantization** of the eigenvalues. Consider for example the Schrödinger equation for a free particle in a one-dimensional box

$$-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\psi = E\psi \ . \tag{280}$$

The general solution to this equation is of course

$$\psi(x) = A\exp(ikx) + B\exp(-ikx) \ , \tag{281}$$

where $k^2 = \frac{2mE}{\hbar^2}$. At this point the energy E is completely arbitrary. However, once we impose boundary conditions

$$\psi(0) = 0 = \psi(L) \ , \tag{282}$$

we find that $A = -B$ and

$$\exp(2ikL) = 1. \tag{283}$$

The latter condition leads to the quantization of the energy eigenvalues

$$E_n = \frac{\hbar^2}{2m}\left[\frac{\pi n}{L}\right]^2 , \qquad n = 1, 2, \ldots \tag{284}$$

This relationship holds quite generally, i.e.

boundary conditions $\longrightarrow$ **quantization of eigenvalues**.

19 The Sturm-Liouville Problem

Let us consider a rather general second order ODE of the form

$$\hat{\mathcal{L}}\ u(x) = p_0(x)\frac{d^2u}{dx^2} + p_1(x)\frac{du}{dx} + p_2(x)u(x) = 0 , \tag{285}$$

where $p_j(x)$ are at least $2-j$ times differentiable functions of some interval

$$a \leq x \leq b . \tag{286}$$

In the linear algebra part of the course we have seen that **Hermitian operators** have particularly nice properties. Let us therefore investigate under what circumstances the differential operator in (285) is Hermitian. Recall that a Hermitian operator has the property that $A = A^\dagger$, or equivalently

$$\langle \mathbf{v}|\,\mathbf{A}\,|\mathbf{u}\rangle = \langle \mathbf{u}|\,\mathbf{A}\,|\mathbf{v}\rangle^* . \tag{287}$$

So for our differential operator to be Hermitian, we require that

$$\langle \mathbf{v}|\,\hat{\mathcal{L}}\,|\mathbf{u}\rangle = \langle \mathbf{u}|\,\hat{\mathcal{L}}\,|\mathbf{v}\rangle^* , \tag{288}$$

where

$$\langle \mathbf{v} \mid \mathbf{u}\rangle = \int_a^b dx\ v^*(x)\ u(x) . \tag{289}$$

Hence

$$\langle \mathbf{v}|\,\hat{\mathcal{L}}\,|\mathbf{u}\rangle = \int_a^b dx\ v^*(x)[p_0(x)\frac{d^2u}{dx^2} + p_1(x)\frac{du}{dx} + p_2(x)u(x)]. \tag{290}$$

Integrating by parts we can bring this to the form

$$\begin{aligned}\langle \mathbf{v}|\,\hat{\mathcal{L}}\,|\mathbf{u}\rangle &= \int_a^b dx\ u(x)\{p_0(v^*)'' + [2p_0' - p_1]\,(v^*)' + v^*(p_2 - p_1' + p_0'')\}\\ &\quad + p_0[v^*u' - u(v^*)']\Big|_a^b\end{aligned} \tag{291}$$

In order for this to be equal to $\langle \mathbf{u}|\,\hat{\mathcal{L}}\,|\mathbf{v}\rangle^*$, we require that

1. $p_0'(x) = p_1(x)$

2. The boundary contributions in (291) cancel, i.e.

$$\begin{aligned} 0 &= p_0(x)[v^*(x)u'(x) - u(x)(v^*(x))']\Big|_a^b \\ &= p_0(x)v^*(x)u'(x)\Big|_a^b - p_0(x)u(x)(v^*(x))'\Big|_a^b . \end{aligned} \tag{292}$$

   This is a requirement on the boundary conditions imposed on the solutions of the differential equation. We emphasize that hermiticity involves the boundary conditions imposed on the differential equations.

   Note that as the boundary condition (292) have to hold for any functions $v(x)$ and $u(x)$, it is sufficient that

$$\boxed{0 = p_0(x)v^*(x)u'(x)\Big|_a^b .} \tag{293}$$

Under these conditions we can write our Hermitian differential operator as

$$\hat{\mathcal{L}}_{\rm SL} = \frac{d}{dx}\left[p_0(x)\frac{du}{dx}\right] + p_2(x) \ . \tag{294}$$

The associated eigenvalue equation

$$\hat{\mathcal{L}}_{\rm SL}\ u(x) = \lambda\ w(x)\ u(x) \tag{295}$$

is called a **Sturm-Liouville problem**. The function $w(x) > 0$ is called the **weight function**. All the differential equations we have encountered so far can be written in the form (295)

Equation	$p_0(x)$	$p_2(x)$	λ	$w(x)$	$[a, b]$
SHO	1	0	n^2	1	$[-\pi, \pi]$
Legendre	$1 - x^2$	0	$-l(l+1)$	1	$[-1, 1]$
Associated Legendre	$1 - x^2$	$-\frac{m^2}{1-x^2}$	$-l(l+1)$	1	$[-1, 1]$
Bessel	x	$-\frac{n^2}{x}$	-1	x	$[0, \infty]$

The SHO problem is also compatible with the boundary conditions (292) as we may impose $u(\pi) = u(-\pi)$. Similarly the Legendre and associated Legendre equations fulfil appropriate boundary conditions for integer l (and $|m| \leq l$ for the associated Legendre equations). The case of Bessel's equation is more involved.

19.1 Orthogonality and Completeness

In analogy to the case of Hermitian matrices in Linear Algebra, the eigenvalues and eigenfunctions of Hermitian Sturm-Liousville operators have a number of very important properties.

1. The eigenvalues of a Hermitian operator are real.
2. The eigenfunctions of a Hermitian operator are orthogonal.
3. The eigenfunctions of a Hermitian operator form a **complete set**.

These properties are completely analogous to what we have seen for matrices in Linear Algebra. On the other hand, as the setting is slightly different, let us repeat parts of the proofs. Let

$$\begin{aligned} \hat{\mathcal{L}}\ u_i - \lambda_i\ w\ u_i &= 0\ , \quad (296) \\ \hat{\mathcal{L}}\ u_j - \lambda_j\ w\ u_j &= 0\ . \quad (297) \end{aligned}$$

Taking the complex conjugate of the second equation, we get

$$\hat{\mathcal{L}}\ u_j^* - \lambda_j^*\ w\ u_j^* = 0\ . \quad (298)$$

Here we have used the the functions entering the definition of $\hat{\mathcal{L}}$ as well as the weight function $w(x)$ are real. Subtracting (298) from (296) and then integrating over $[a,b]$ we have

$$\int_a^b dx\ u_j^* \hat{\mathcal{L}} u_i - \int_a^b dx\ u_i \hat{\mathcal{L}} u_j^* = (\lambda_i - \lambda_j^*) \int_a^b dx u_i u_j^* w\ . \quad (299)$$

As $\hat{\mathcal{L}}$ is Hermitian, the left hand side of this equation is zero. Hence

$$(\lambda_i - \lambda_j^*) \int_a^b dx u_i u_j^* w = 0. \quad (300)$$

Now, setting $i = j$ we obtain

$$(\lambda_j - \lambda_j^*) \int_a^b dx |u_j|^2 w = 0. \quad (301)$$

However, by definition we have $w(x) > 0$ and hence the integral is positive. This leaves no choice other than

$$\lambda_j = \lambda_j^*. \quad (302)$$

Let us now turn to property 2. Let us return to equation (300) but now consider the case $\lambda_i \neq \lambda_j$. Then (300) implies that

$$\int_a^b dx\ u_i(x)\ u_j^*(x)\ w(x) = 0\ . \tag{303}$$

This is actually the definition of orthogonality of the eigenfunctions of a Strum-Liouville operator with weight function $w(x)$. Note the presence of $w(x)$! Just like for matrices, we have to consider the case where two or several eigenvalues are equal separately. Again orthogonality can be achieved through application of the Schmidt method.

Examples:

1. Sines and Cosines: Let's return to our SHO problem on the interval $[-\pi, \pi]$. As we have seen above, the solutions are $\cos(mx)$ and $\sin(mx)$. They fulfil the orthogonality relations

$$\begin{aligned}\int_{-\pi}^{\pi} dx\ \sin(nx)\ \sin(mx) &= \delta_{nm}\ ,\ n, m \geq 1\ ,\\ \int_{-\pi}^{\pi} dx\ \cos(nx)\ \cos(mx) &= \pi(\delta_{nm} + \delta_{n0})\ ,\\ \int_{-\pi}^{\pi} dx\ \sin(nx)\ \cos(mx) &= 0\ .\end{aligned} \tag{304}$$

   These will play an important role in the theory of Fourier series.

2. Legendre polynomials: The Legendre polynomials form an orthogonal set on the interval $[-1, 1]$

$$\int_{-1}^{1} dx\ P_n(x)\ P_m(x) = \frac{2}{2n+1}\delta_{nm}\ . \tag{305}$$

   Note that the standard normalization used in the literature is not to unity!

3. Associated Legendre functions: Using that the eigenvalue is $-l(l+1)$ whereas m enters the definition of $p_2(x)$ we conclude that

$$\int_{-1}^{1} dx\ P_l^m(x)\ P_{l'}^m(x) = \delta_{l,l'}\frac{2}{2l+1}\frac{(l+m)!}{(l-m)!}\ . \tag{306}$$

   Note that the standard normalization is once again not to unity.

A particularly important property of a Hermitian operator is that its eigenfunctions form a **complete set**. We have already encountered this in the Linear Algebra part of the course, where we showed that the eigenstates of a Hermitian matrix form a **basis** of the linear vector space on which the matrix acts. For differential operators $\hat{\mathcal{L}}_{\rm SL}$ of the Sturm-Liouville type on an interval $[a, b]$ this means that any well-behaved function $f(x)$ can be approximated by a series

$$f(x) = \sum_{n=0}^{\infty} a_n \ u_n(x) \tag{307}$$

to any degree of accuracy. Here $u_n(x)$ are the eigenfunctions of $\hat{\mathcal{L}}_{\rm SL}$ and

$$a_m = \int_a^b dx \ w(x) \ f(x) \ u_m(x) \ . \tag{308}$$

More precisely we have

$$\lim_{k\to\infty} \int_a^b dx \ w(x) \left[f(x) - \sum_{n=0}^{m} a_n \ u_n(x) \right]^2 = 0. \tag{309}$$

A proof of this remarkable property of Sturm-Liouville differential operators can be found *e.g.* in R. Courant and D. Hilbert, *Methods of Mathematical Physics*, Chapter 6 section 3.

Examples:

1. Simple Harmonic Oscillator:

   The completeness property of the eigenfunctions of the SHO Sturm - Liouville operator leads to the theory of **Fourier series**. As we have already seen, we may expand any reasonably well-behaved periodic function $f(x)$ as

   $$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \ \cos(nx) + b_n \ \sin(nx), \tag{310}$$

   where

   $$\begin{aligned} a_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} dx \ f(x) \ \cos(nx) \ , \\ b_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} dx \ f(x) \ \sin(nx). \end{aligned} \tag{311}$$

2. Legendre's Equation:

The completeness property of the eigenfunctions of the Sturm-Liouville operator of Legendre's equation allows us to expand a function $f(x)$ on the interval $[-1, 1]$ in a **Legendre series**

$$\begin{aligned} f(x) &= \sum_{n=1}^{\infty} a_n \; P_n(x) \;, \\ a_n &= \frac{2n+1}{2} \int_{-1}^{1} dx \; f(x) \; P_n(x) \;. \end{aligned} \tag{312}$$

Here the extra factor of $\frac{2n+1}{2}$ is due to the normalization of the Legendre polynomials.

20 Use of Complete Function Systems

20.1 Electrostatics: Linear Multipole Expansions

The **generating function** of the Legendre polynomials is

$$\boxed{[1 - 2xt + t^2]^{-\frac{1}{2}} = \sum_{n=0}^{\infty} P_n(x) \; t^n \;, \qquad |t| < 1.} \tag{313}$$

Indeed, for $|x|, |t| < 1$ we may Taylor expand the LHS around $t = 0$ to obtain

$$\begin{aligned} \frac{1}{\sqrt{1 - 2xt + t^2}} &= 1 + tx + \frac{t^2}{2}(3x^2 - 1) + \frac{t^3}{3!}(15x^3 - 9x) + \ldots \\ &= \sum_{n=0}^{\infty} \left[\sum_{k=0}^{[n/2]} (-1)^k \frac{(2n-2k)!}{2^{2n-2k} k!(n-k)!(n-2k)!} (2x)^{n-2k} \right] t^n. \end{aligned}$$

Let us use the generating function to describe the electrostatic potential of a **dipole**. Let us imagine we have a charge $+q$ at position $z = a$ and a charge $-q$ at position $z = -a$ as shown in Fig.9. The potential at position $\mathbf{r}$ is simply

$$V(\mathbf{r}) = \frac{q}{4\pi\epsilon_0} \left[\frac{1}{|\mathbf{r_1}|} - \frac{1}{|\mathbf{r_2}|} \right], \tag{314}$$

where $\mathbf{r_{1,2}}$ are the vectors shown in Fig.9. It is a simple matter to rewrite this as

$$V(\mathbf{r}) = \frac{q}{4\pi\epsilon_0} \left\{ \left[1 - 2\frac{a}{r} \cos(\theta) + \frac{a^2}{r^2} \right]^{-\frac{1}{2}} - \left[1 + 2\frac{a}{r} \cos(\theta) + \frac{a^2}{r^2} \right]^{-\frac{1}{2}} \right\}, \tag{315}$$

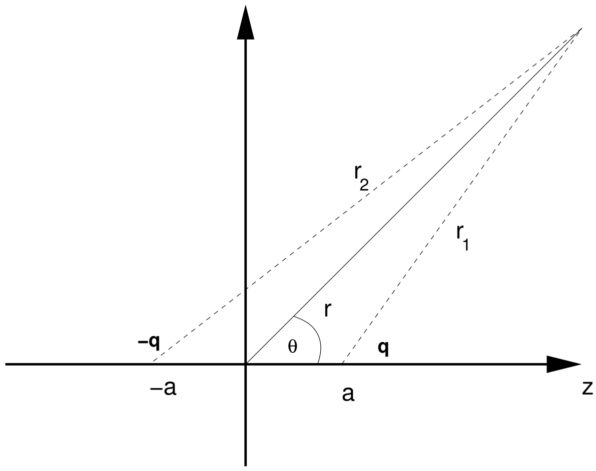


Figure 9: Electric Dipole.

where $r = |\mathbf{r}|$. Now we may use that the functions occurring in (315) are generating functions of Legendre polynomials. We obtain

$$V(\mathbf{r}) = \frac{2q}{4\pi\epsilon_0\ r} \sum_{n=0}^{\infty} P_{2n+1}(\cos(\theta)) \ \left(\frac{a}{r}\right)^{2n+1} \ , \qquad r > a \ . \tag{316}$$

The leading term in this series for $r \gg a$

$$\frac{2aq}{4\pi\epsilon_0} \ \frac{P_1(\cos\theta)}{r^2} \tag{317}$$

is called **electric dipole potential**. The potential due to two point charges of q at $\pm a$ and a point charge $-2q$ at $z = 0$ can be analyzed analogously. The expansion now starts with the term proportional to $P_2(\cos\theta)$ and describes a **linear electric quadrupole**.

20.2 Quantum Mechanics: Resolution of the Identity

In Quantum Mechanics observables such as energy or momentum are represented by Hermitian operators. The eigenfunctions of these operators then form complete function systems, i.e. **bases** of the linear vector spaces ("Hilbert spaces") of quantum mechanical states. Let $\hat{H}$ be the Hamiltonian of a quantum mechanical system. Then the energy eigenstates fulfil

$$\hat{H}\,|\psi_{\mathbf{n}}\rangle = E_n\,|\psi_{\mathbf{n}}\rangle \ , \tag{318}$$

where E_n are the quantized energies. Because the states $|\psi_{\mathbf{n}}\rangle$ form a complete set, we can expand **any** quantum mechanical state in the energy eigenbasis

$$|\mathbf{\Phi}\rangle = \sum_n \langle\psi_{\mathbf{n}} \mid \mathbf{\Phi}\rangle \ |\psi_{\mathbf{n}}\rangle \ . \tag{319}$$

The **identity operator** $\mathbb{I}$ is defined as the operator that when acting on **any** quantum mechanical state gives back the state

$$\mathbb{I}\,|\mathbf{\Phi}\rangle = |\mathbf{\Phi}\rangle \ . \tag{320}$$

Using that the eigenstates of the Hamiltonian form a complete set of states we can represent $\mathbb{I}$ as

$$\boxed{\mathbb{I} = \textstyle\sum_n |\psi_\mathbf{n}\rangle \langle\psi_\mathbf{n}| \ .} \tag{321}$$

Indeed acting with (321) on a state $|\mathbf{\Phi}\rangle$ we find

$$\mathbb{I}\,|\mathbf{\Psi}\rangle = \sum_n \{|\psi_\mathbf{n}\rangle \langle\psi_\mathbf{n}|\}\,|\mathbf{\Phi}\rangle = \sum_n |\psi_\mathbf{n}\rangle \langle\psi_\mathbf{n} \mid \mathbf{\Phi}\rangle \ . \tag{322}$$

This is indeed equal to $|\mathbf{\Phi}\rangle$ by equation (319). Equation (321) is sometimes referred to as **resolution of the identity**. This is something we use very frequently in Quantum Mechanics.

21 Partial Differential Equations

A partial differential equation is a differential equation for a function of several variables.
Examples:

- Time-dependent Schrödinger Equation

$$i\hbar\frac{\partial\Psi(\mathbf{r},t)}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\Psi(\mathbf{r},t) + V(\mathbf{r})\Psi(\mathbf{r},t). \tag{323}$$

  Here the "Laplacian" is given by

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \tag{324}$$

  You'll deal a lot with this PDE in the QM course. Ψ is the quantum mechanical wave function.

- Wave Equation

$$\left[\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2\right]\Psi(\mathbf{r},t) = 0. \tag{325}$$

  You have already met this PDE in the "waves" course last year.

- Diffusion Equation

$$\frac{\partial\Psi(\mathbf{r},t)}{\partial t} = D\nabla^2\Psi(\mathbf{r},t) \ . \tag{326}$$

  The diffusion equation describes the temperature Ψ in a region containing no heat sources or sinks. D is called *diffusion constant* .

- Laplace's Equation
$$\nabla^2\Psi(\mathbf{r}) = 0. \tag{327}$$
- Poisson's Equation
$$\nabla^2\Psi(\mathbf{r}) = -\frac{\rho(\mathbf{r})}{\epsilon_0}. \tag{328}$$
You deal with Laplace's equation and Poisson's equation in the Electromagnetism course.

All of these equations are of the form

$$\hat{\mathcal{D}}\Psi + f\Psi = g, \tag{329}$$

where $\hat{\mathcal{D}}$ is a differential operator and f, g are known functions.

- Such equations are **linear** PDEs as Ψ enters linearly. **Nonlinear** PDEs such as the sine-Gordon equation
$$\left[\frac{1}{c^2}\frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial x^2}\right]\Psi(t,x) + \mu\sin(\Psi(t,x)) = 0 \tag{330}$$
are **MUCH** harder to solve and we won't deal with them here.
- If $g = 0$ the PDE is called **homogeneous**. Homogeneus equations have the **wonderful** property of **linear superposition** of solutions: if Ψ_1 and Ψ_2 are solutions of the PDE
$$\hat{\mathcal{D}}\Psi_1 + f\Psi_1 = 0\ , \qquad \hat{\mathcal{D}}\Psi_2 + f\Psi_2 = 0\ , \tag{331}$$
then so is $\alpha\Psi_1 + \beta\Psi_2$ where α, β are numbers
$$\hat{\mathcal{D}}(\alpha\Psi_1 + \beta\Psi_2) + f(\alpha\Psi_1 + \beta\Psi_2) = 0\ . \tag{332}$$
- If $g \neq 0$ the PDE is called **inhomogeneous**.
- PDEs must always be supplemented by **boundary conditions** in order to make the solution unique. In general this is a tricky isssue, although not for the cases we consider here.

21.1 Classification of 2^{nd} order linear homogeneous PDEs

21.2 Boundary Conditions

In Physics problems there are three main types of boundary conditions. Let us denote by u the function subject to the PDE.

- Dirichlet Boundary Conditions

  u is specified at each point along a boundary of a region.

- Neumann Boundary Conditions

  The values of the normal derivative $\frac{\partial u}{\partial n}$ of the function [5] are specified at the boundary.

- Cauchy Boundary Conditions

  Both u and $\frac{\partial u}{\partial n}$ are specified at the boundary.

  An example is the one dimensional wave equation describing the displacement of a string at time t and position x

  $$\left[\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2}\right] u(x,t) = 0. \tag{333}$$

  Here we would specify the initial displacement $u(x, t = 0)$ as well as the initial velocity $\frac{\partial u}{\partial t}(x, t = 0)$ in order to obtain a unique solution. The boundary is the line $t = 0$ in the two-dimensional (x, t) plane.

22 Separation of Variables in Cartesian Coordinates

Let us consider heat flow in a semi-infinite rectangular plate, see Fig. 10.

The symmetry of the problem is such that cartesian coordinates are most convenient. Let us denote the temperature by $\Theta(t, x, y)$. It obeys the **heat equation**

$$\frac{\partial \Theta}{\partial t} = D\nabla^2\Theta. \tag{334}$$

If we wait long enough the system reaches a stationary state in which the temperature becomes time independent. Then the heat equation reduces to **Laplace's equation**

$$\boxed{\nabla^2\Theta = 0.} \tag{335}$$

The boundary conditions are as follows:

- The temperature at the sides of the plate at $x = 0$ and $x = L$ is zero, i.e.

  $$\Theta(x = 0, y) = 0 \ , \qquad \Theta(x = L, y) = 0 \ . \tag{336}$$

- The temperature at the bottom of the plate is T_0

  $$\Theta(x, y = 0) = T_0 \ , \quad 0 \leq x \leq L. \tag{337}$$

[5] $\frac{\partial u}{\partial n} = \nabla u \cdot \mathbf{n}$, where $\mathbf{n}$ is the normal to the boundary at each point.

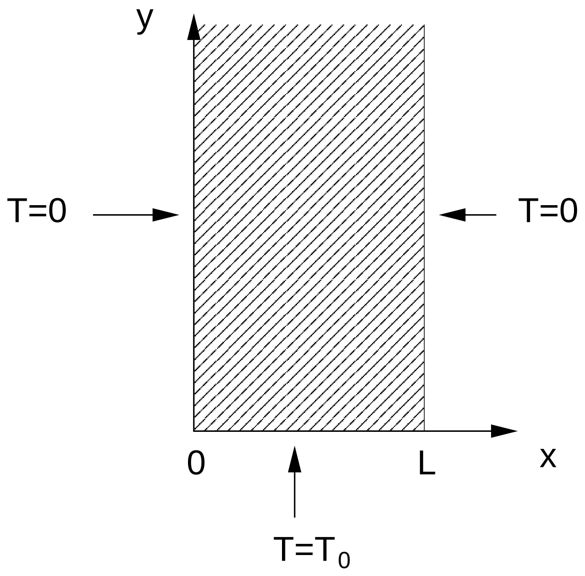


Figure 10: Heat flow in a semi-infinite rectangular plate.

- The temperature at infinity is zero

$$\Theta(x, y = \infty) = 0. \tag{338}$$

  This condition is physically motivated: in order to have a nonzero temperature at $y = 0$ an infinite amount of heat would have to be supplied to the plate.

Let us now solve Laplace's equation for the specified boundary conditions. This is done by a method called **separation of variables**. We make an ansatz

$$\Theta(x, y) = X(x)Y(y) \ , \tag{339}$$

and substitute this back into Laplace's equation. This gives

$$\frac{d^2X(x)}{dx^2}Y(y) + \frac{d^2Y(y)}{dy^2}X(x) = 0. \tag{340}$$

Note that we have total derivatives here as e.g. X is a function of x only. Dividing by XY we arrive at

$$\frac{1}{X(x)}\frac{d^2X(x)}{dx^2} = -\frac{1}{Y(y)}\frac{d^2Y(y)}{dy^2}. \tag{341}$$

Now comes the crucial observation: the LHS of this equation depends only on x, whereas the RHS depends only on y. The only way this is possible is if both sides are equal to the same x and y independent constant. Let us call this constant $-k^2$. Then

$$\begin{aligned} \frac{1}{X(x)}\frac{d^2X(x)}{dx^2} &= -k^2 , \\ \frac{1}{Y(y)}\frac{d^2Y(y)}{dy^2} &= k^2. \end{aligned} \tag{342}$$

We observe that the separation of variables ansatz has reduces the PDE to two ODEs, which are coupled through the separation constant.

You know from last year that the general solutions of these differential equations are

$$\begin{aligned} X_k(x) &= A_k \sin(kx) + B_k \cos(kx) , \\ Y_k(y) &= C_k e^{ky} + D_k e^{-ky}. \end{aligned} \tag{343}$$

Laplace's equation is a linear homogeneous equation, which means that the **superposition principle** holds. Hence the general solution is

$$\Theta(x,y) = \sum_k f_k \; X_k(x) \; Y_k(y). \tag{344}$$

Here f_k are arbitrary coefficients. Clearly they can be absorbed in a redefinition of A_k, B_k, C_k, D_k, so without loss of generality we can set $f_k = 1$. Hence

$$\boxed{\Theta(x,y) = \textstyle\sum_k X_k(x) \; Y_k(y).} \tag{345}$$

The sum is over all separation constants k.

Let us now impose the boundary conditions on $\Theta(x,y)$:

- $\Theta(x, y = \infty) = 0$

  This implies that for any value of x

  $$\lim_{y\to\infty} \sum_k X_k(x) Y_k(y) = 0 = \sum_k X_k(x) \lim_{y\to\infty} Y_k(y) = 0 \; . \tag{346}$$

  Now we use that the $X_k(x)$ form a **linearly independent** set of functions. Hence the sum over k can only be zero if every individual term is zero, i.e.

  $$\lim_{y\to\infty} Y_k(y) = 0 \quad \Rightarrow \qquad \boxed{C_k = 0.} \tag{347}$$

- $\Theta(0, y) = \Theta(L, y) = 0.$

For $x = 0$ we have

$$\sum_k X_k(0)\ Y_k(y) = 0. \tag{348}$$

Now we use that the $Y_k(y)$ form an independent set of functions and therefore the sum can only be zero if any individual term vanishes. Hence

$$X_k(0) = 0. \tag{349}$$

This implies that

$$\boxed{B_k = 0.} \tag{350}$$

By the same argument as above we obtain that $X_k(L) = 0$ as well, which implies that

$$\sin(kL) = 0 \quad \Rightarrow \quad \boxed{k_n = \frac{n\pi}{L}.} \tag{351}$$

At this point our solution looks like

$$\Theta(x, y) = \sum_{k_n} A_n \sin(k_n x)\ D_n e^{-k_n y} = \sum_{n=1}^{\infty} b_n \sin\left(\frac{n\pi x}{L}\right)\ e^{-n\pi y/L}. \tag{352}$$

Note that although we obtained the allowed values for k_n by considering a boundary condition on X_k, they enter into the form of Y_k as well! Don't get this wrong in the homework!

- The final boundary conditions is $\Theta(x, y = 0) = T_0$.

  This gives us

$$T_0 = \Theta(x, 0) = \sum_{n=1}^{\infty} b_n \sin\left(\frac{n\pi x}{L}\right). \tag{353}$$

This one condition fixes all coefficients b_n as it must hold for all $0 \le x \le L$! In order to pick out an individual term in the sum on the RHS we use a trick. Let us integrate both sides of the equation from 0 to L with $\sin(m\pi x/L)$:

$$\begin{aligned} T_0 \int_0^L dx\ \sin\Big(\frac{m\pi x}{L}\Big) &= \sum_{n=1}^{\infty} b_n \int_0^L dx\ \sin\Big(\frac{m\pi x}{L}\Big)\ \sin\Big(\frac{n\pi x}{L}\Big) \\ &= \sum_{n=1}^{\infty} b_n \int_0^L dx\ \frac{1}{2}\left[\cos\Big(\frac{(n-m)\pi x}{L}\Big) - \cos\Big(\frac{(n+m)\pi x}{L}\Big)\right] \\ &= \sum_{n=1}^{\infty} b_n \frac{L}{2}\delta_{nm} = \frac{L b_m}{2}. \end{aligned} \tag{354}$$

This gives us

$$b_m = \frac{2T_0}{L}\int_0^L dx\ \sin\Big(\frac{m\pi x}{L}\Big) = \begin{cases} \frac{4T_0}{m\pi} & \text{m odd} \\ 0 & \text{m even.} \end{cases} \tag{355}$$

Our final result in its full glory then reads

$$\boxed{\Theta(x,y) = \sum_{m=1}^{\infty} \frac{4T_0}{\pi(2m-1)} \sin\left(\frac{(2m-1)\pi x}{L}\right)\ \exp\left(-\frac{(2m-1)\pi y}{L}\right).} \tag{356}$$

22.1 Some Comments

- Choice of separation constant.

  Recall that we chose out separation constant to be $-k^2$, which we implicitly assumed to be real. This choice, perhaps unsurprisingly, turned out to be **very convenient**. However, a priori all we know is that the separation constant is a complex number. So if we had called our separation constant α and allowed it to be complex, we still would have arrived at exactly the same result, although the calculation would have been somewhat messier! How did we come up with our convenient choice then? Well, if you look at the boundary conditions you see that $X(x)$ must vanish both at $x = 0$ and $x + L$, which suggests that $X(x)$ exhibits oscillating behaviour. Inspection of the differential equation for $X(x)$ reveals that the solutions will be sines and cosines (with the desired oscillating behaviour) if the separation constant is real and negative.

- At what stage to impose boundary conditions?

  Note that we have imposed boundary conditions only **after** writing down the most general solution obtainable by separation of variables. Using the linear independence of the solutions with different separation constants, the boundary conditions (336), (338) on Θ translated to boundary conditions on X_k and Y_k. The answers are the same as imposing the boundary conditions directly on X_k and Y_k. However, as you can see from the treatment of the boundary conditions (337), you cannot always impose the boundary conditions directly on the solutions X_k, Y_k for fixed separation constant k. So the correct procedure is always to first find the general solution by superposition, and only then to impose the boundary conditions.

23 Separation of Variables in Cylindrical Polar Coordinates

Next we consider two examples of separation of variables in problems with cylindrical symmetry.

23.1 Infinitely long metal rod in a perpendicular electric field.

Consider an infinitely long earthed metal rod in an electric field. The axis of the rod is the z-axis and the electric field is taken to be perpendicular to it. Very far from the origin the electric field is given by

$$\vec{E} \longrightarrow (E, 0, 0). \tag{357}$$

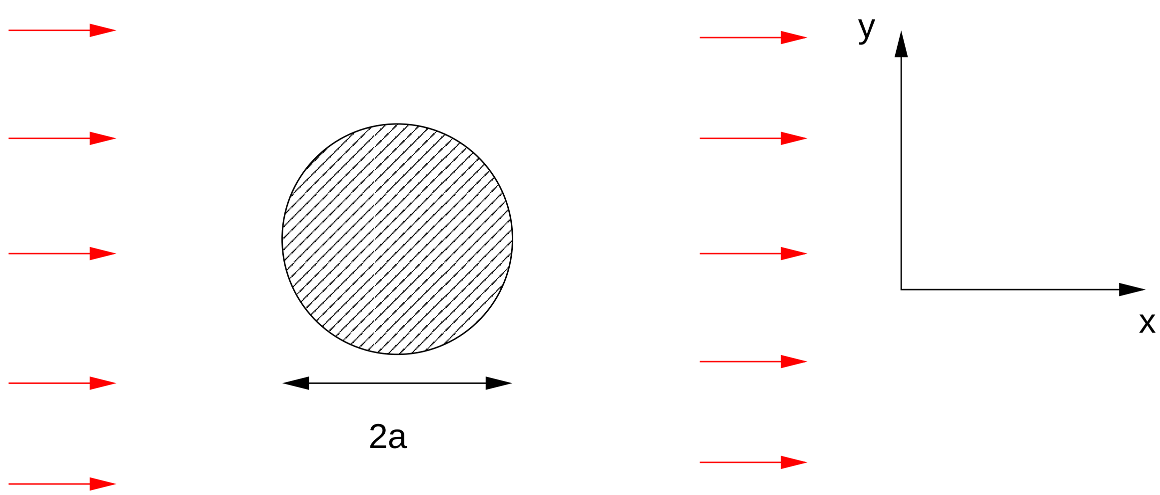


Figure 11: Metal rod in a perpendicular electric field.

The geometry of the problem is such that we can drop the z-coordinate as nothing depends on it. We want to determine the electric field close to the rod. In order to do so, we need to solve Laplace's equation for the scalar potential

$$\nabla^2 \Phi(x, y) = 0 \ . \tag{358}$$

This equation holds **outside** the rod, and the boundary conditions are

- $\Phi = 0$ on the surface of the rod (as it is earthed).
- $\vec{E} = -\vec{\nabla}\Phi$, which implies that far from the origin

$$\Phi \longrightarrow -Ex. \tag{359}$$

The problem clearly has cylindrical symmetry, implying that cylindrical coordinates are most convenient. As nothing depends on z these simple reduce to planar polar coordinates. We then have $\Phi = \Phi(r, \phi)$ and

$$\nabla^2 = \frac{1}{r}\frac{\partial}{\partial r}r\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \phi^2}. \tag{360}$$

The separation of variables ansatz reads

$$\Phi(r, \phi) = R(r)\ P(\phi). \tag{361}$$

Substituting this back into Laplace's equation we obtain

$$P(\phi)\frac{1}{r}\frac{d}{dr}\left[r\frac{dR}{dr}\right] + \frac{R}{r^2}\frac{d^2P}{d\phi^2} = 0. \tag{362}$$

Multiplying both sides by $r^2/(RP)$ we arrive at

$$\frac{r}{R}\frac{d}{dr}\left[r\frac{dR}{dr}\right] = -\frac{1}{P}\frac{d^2P}{d\phi^2}. \tag{363}$$

The LHS depends only on r whereas the RHS depends only on ϕ. This is only possible if both are equal to the same constant, which we denote by m^2. This gives us two ODEs

$$r\frac{d}{dr}\left[r\frac{dR}{dr}\right] = m^2R\ , \tag{364}$$

$$\frac{d^2P}{d\phi^2} = -m^2P. \tag{365}$$

The general solution to (365) is[6]

$$P_m(\phi) = a_m\cos(m\phi) + b_m\sin(m\phi). \tag{366}$$

As ϕ is an angular variable we must have

$$P_m(\phi) = P_m(\phi + 2\pi)\ , \tag{367}$$

which forces m to be **integer**.

The general solution of (364) is obtained by inspection [7]

$$R_m(r) = \begin{cases} c_m r^m + d_m r^{-m} & \text{if } m > 0 \\ c_0\ln(r) + d_0 & \text{if } m = 0. \end{cases} \tag{368}$$

[6]The P_m's are not to be confused with the Legendre polynomials!

[7]For fixed m (364) is a second order linear homogeneous ODE, which tells us that there must be two independent solutions.

Using the superposition principle we find that the general solution of (358) is given by

$$\Phi(r,\phi) = \sum_m R_m(r)\ P_m(\phi). \tag{369}$$

Next we impose the boundary conditions:

- $\Phi(a,\phi) = 0$.

  As the $P_m(\phi)$ are linearly independent, we must have

  $$R_m(a) = 0\ , \quad m = 0,1,2,\dots \tag{370}$$

  This implies that

  $$c_0 \ln(a) + d_0 = 0\ , \qquad c_m a^m + d_m a^{-m} = 0\ ,\ m = 1,2,\dots . \tag{371}$$

  Hence

  $$R_m(r) = c_m \left[r^m - \frac{a^{2m}}{r^m} \right]\ , \quad m = 1,2,\dots \tag{372}$$

- For $r \to \infty$ we must have $\Phi(r,\phi) \longrightarrow -Er\cos(\phi)$.

  This implies that

  $$\sum_{m=0}^{\infty} R_m(r) P_m(\phi) \longrightarrow -Er\cos(\phi). \tag{373}$$

  In order to isolate the individual summands we again use a trick. Note that

  $$\begin{aligned} \int_0^{2\pi} d\phi\ P_m(\phi)\ \cos(n\phi) &= \pi a_m\ \delta_{nm}\ , \\ \int_0^{2\pi} d\phi\ P_m(\phi)\ \sin(n\phi) &= \pi b_m\ \delta_{nm}. \end{aligned} \tag{374}$$

  Using these in (373) we find that

  $$\begin{aligned} R_n(r)\pi a_n &\longrightarrow \int_0^{2\pi} d\phi\ [-Er\cos(\phi)]\ \cos(n\phi) = -\pi Er\delta_{n,1}\ , \\ R_n(r)\pi b_n &\longrightarrow \int_0^{2\pi} d\phi\ [-Er\cos(\phi)]\ \sin(n\phi) = 0. \end{aligned} \tag{375}$$

  We conclude that

  $$\begin{aligned} b_n &= 0\ , \quad n = 1,2,\dots \\ a_n &= 0\ , \quad n = 0,2,3\dots \\ a_1 R_1(r) &= a_1 c_1 (r - \frac{a^2}{r}) \longrightarrow -Er\ . \end{aligned} \tag{376}$$

This fixes $a_1 c_1 = -E$ and our final solution hence takes the form

$$\boxed{\Phi(r,\phi) = E\left[\frac{a^2}{r} - r\right]\cos(\phi).} \tag{377}$$

23.2 Vibrations of a circular drum.

As another example of separation of variables in cylindrical variables let us consider the vibrations of a circular drum of radius a.

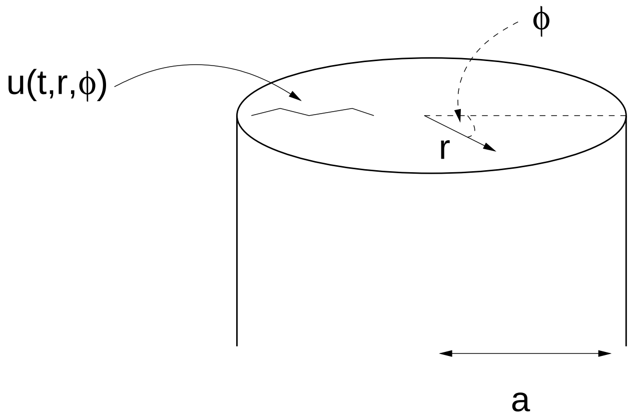


Figure 12: A circular drum.

The displacement $u(t,r,\phi)$ fulfils the wave equation

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = \nabla^2 u. \tag{378}$$

The boundary conditions reflect the fact that the displacement must vanish at the rim of the drum

$$u(t,a,\phi) = 0. \tag{379}$$

Let us try to find solutions by the following separation of variables ansatz

$$u(t,r,\phi) = T(t)R(r)P(\phi). \tag{380}$$

Substituting this back into the wave equation and dividing by TRP we arrive at

$$\frac{1}{c^2}\frac{1}{T}\frac{d^2T}{dt^2} = \frac{1}{rR}\frac{d}{dr}(r\frac{dR}{dr}) + \frac{1}{r^2}\frac{1}{P}\frac{d^2P}{d\phi^2}. \tag{381}$$

Now we proceed in two stages. We first realize that the LHS only depends on t, whereas the RHS only depends on r and ϕ. This is only possible if they are both equal to the same constant, which we denote by $-\omega^2/c^2$

$$\frac{1}{T}\frac{d^2T}{dt^2} = -\omega^2 \ , \tag{382}$$

$$\frac{1}{rR}\frac{d}{dr}(r\frac{dR}{dr}) + \frac{1}{r^2}\frac{1}{P}\frac{d^2P}{d\phi^2} = -\frac{\omega^2}{c^2}. \tag{383}$$

Equation (383) can be brought to the form

$$\frac{r}{R}\frac{d}{dr}(r\frac{dR}{dr}) + \frac{\omega^2}{c^2}r^2 = -\frac{1}{P}\frac{d^2P}{d\phi^2}. \tag{384}$$

Now the RHS only depends on ϕ and the LHS only on r, so they must both be equal to the same constant, which we denote by k^2. This leaves us with three ODEs coupled through the separation constants ω and k

$$\frac{d^2T}{dt^2} + \omega^2 T = 0 \ , \tag{385}$$

$$\frac{d^2P}{d\phi^2} + k^2 P = 0 \ , \tag{386}$$

$$r\frac{d}{dr}(r\frac{dR}{dr}) + \left[\frac{\omega^2}{c^2}r^2 - k^2\right] R = 0 \ . \tag{387}$$

The equations for T and P are easily solved

$$P(\phi) = A_m e^{im\phi} + B_m e^{-im\phi} \ , \tag{388}$$

$$T(t) = C_\omega e^{i\omega t} + D_\omega e^{-i\omega t} \ . \tag{389}$$

Here m must be an integer because ϕ is an angular variable $P(\phi) = P(\phi + 2\pi)$. Let us rewrite rhe equation for R using a dimensionless variable

$$\kappa = \frac{\omega r}{c}, \tag{390}$$

$$\boxed{\kappa^2\frac{d^2R_m}{d\kappa^2} + \kappa\frac{dR_m}{d\kappa} + (\kappa^2 - m^2)R_m = 0.} \tag{391}$$

This differential equation is known as **Bessel**'s equation and we have already encountered it in the ODE section. Its general solution is

$$R_m(\kappa) = e_m J_m(\kappa) + f_m N_m(\kappa) \ , \tag{392}$$

where J_m and N_m are Bessel and Neumann functions respectively. Recalling that the Neumann functions are singular at the origin

$$\lim_{\kappa \to 0} N_m(\kappa) = -\infty \ , \tag{393}$$

we conclude that on physical grounds

$$f_m = 0. \tag{394}$$

Superposing solutions we arrive at

$$u(t,r,\phi) = \sum_\omega \left\{ C_\omega e^{i\omega t} + D_\omega e^{-i\omega t} \right\} \sum_{m=0}^{\infty} J_m\left(\frac{\omega r}{c}\right) \left[A_m e^{im\phi} + B_m e^{-im\phi} \right] . \tag{395}$$

Next we impose the boundary condition $u(t,a,\phi) = 0$. Using the linear independence of the $T_\omega(t)$ and the $P_m(\phi)$ we conclude that we must have

$$J_m\left(\frac{\omega a}{c}\right) = 0. \tag{396}$$

This means that $\omega a/c$ must be a zero of $J_m(\kappa)$! Let us denote the zeroes of $J_m(\kappa)$ for fixed m by

$$\alpha_{m,p} \ , \qquad p = 1, 2, \ldots \tag{397}$$

The values of $\alpha_{m,p}$ are known from numerical tables. Condition (396) then fixes the allowed values of the normal mode frequencies

$$\boxed{\omega_{m,p} = \frac{c}{a} \alpha_{m,p} \ .} \tag{398}$$

Interestingly the allowed frequencies are not simple multiples of a fundamental frequency as is the case e.g. for string. This is the reason for the interesting sound of a drum.

The final form of our solution for the displacement is

$$u(t,r,\phi) = \sum_{m=0}^{\infty} \left[A_m e^{im\phi} + B_m e^{-im\phi} \right] \sum_{p=1}^{\infty} \left\{ C_{m,p} e^{i\omega_{m,p} t} + D_{m,p} e^{-i\omega_{m,p} t} \right\} J_m\left(\frac{\alpha_{m,p} r}{a}\right). \tag{399}$$

As expected, the boundary condition do not fix a unique solution. In order to do that we would need to specify initial conditions, i.e. say how we have hit the drum.

23.2.1 Some Low-Order Normal Modes

Let us look in more detail at the normal modes (i.e. terms in (399) with fixed m and p) for small m, p. The nodes (values of r, ϕ where $u(t,r,\phi) = 0$) are determined by e.g.

$$A_m e^{im\phi} + B_m e^{-im\phi} \equiv a_m \sin(m\phi + \delta_m) = 0. \tag{400}$$

So for $m = 1$ there is a node at $\phi = -\delta_1 \bmod \pi$.

In Fig.13 we show the lines of zeroes of several normal modes (we set $\delta_m = 0$ in these figures). We also indicate the regions in which $u(t,r,\phi)$ is positive (+) and negative (−).

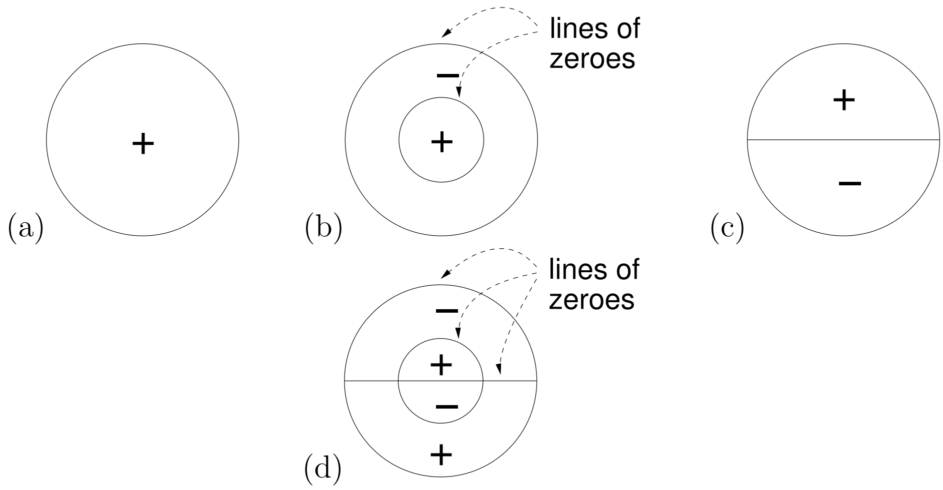


Figure 13: Normal modes of a circular drum: (a) $m = 0$, $p = 1$ (b) $m = 0$, $p = 2$ (c) $m = 1$, $p = 1$ (d) $m = 1$, $p = 2$

24 Separation of Variables in Spherical Polar Coordinates

Let us consider the problem of the near field of an earthed metal sphere or radius a in an electric field. We want to solve Laplace's equation for the potential

$$\nabla^2 V(r, \theta, \varphi) = 0 \ , \tag{401}$$

subject to the boundary conditions

$$V(r \to \infty, \theta, \varphi) = -Ez = -Er\cos(\theta) \ , \tag{402}$$
$$V(r = a, \theta, \varphi) = 0 \ . \tag{403}$$

The problem clearly has spherical symmetry and the most convenient choice of coordinate system is spherical polar coordinates. Recall that the Laplacian ∇^2 is spherical polar coordinates is

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left[r\frac{\partial}{\partial r}\right] + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2} + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left[\sin\theta\frac{\partial}{\partial\theta}\right] \tag{404}$$

Exercise: Derive this expression starting from $\nabla^2 = \frac{\partial^2}{\partial_x^2} + \frac{\partial^2}{\partial_y^2} + \frac{\partial^2}{\partial_z^2}$.

Let us try to find solutions to Laplace's equation of the form

$$V(r, \theta, \varphi) = R(r)\ \Theta(\theta)\ \Phi(\varphi) \ . \tag{405}$$

Substituting this back into Laplace's equation and multiplying through by $r^2/(R\Theta\Phi)$ we arrive at

$$-\frac{1}{R}\frac{d}{dr}\left[r^2\frac{dR}{dr}\right] = \frac{1}{\sin^2\theta}\frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2} + \frac{1}{\sin\theta}\frac{1}{\Theta}\frac{d}{d\theta}\left[\sin\theta\frac{d\Theta}{d\theta}\right]. \tag{406}$$

The LHS depends only on r whereas the RHS depends only on θ and ϕ. Hence they both must be constant

$$\boxed{\frac{1}{R}\frac{d}{dr}\left[r^2\frac{dR}{dr}\right] = \text{const} = l(l+1),} \tag{407}$$

$$\frac{1}{\sin^2\theta}\frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2} + \frac{1}{\sin\theta}\frac{1}{\Theta}\frac{d}{d\theta}\left[\sin\theta\frac{d\Theta}{d\theta}\right] = -l(l+1). \tag{408}$$

Multiplying by $\sin^2\theta$ we can bring the second equation to the form

$$-\frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2} = \frac{\sin\theta}{\Theta}\frac{d}{d\theta}\left[\sin\theta\frac{d\Theta}{d\theta}\right] + l(l+1)\sin^2\theta. \tag{409}$$

Now the LHS depends only on ϕ and the RHS only on θ, so they must be equation to the same constant which we denote by m^2

$$\boxed{\frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2} = -m^2,} \tag{410}$$

$$\boxed{\frac{\sin\theta}{\Theta}\frac{d}{d\theta}\left[\sin\theta\frac{d\Theta}{d\theta}\right] + l(l+1)\sin^2\theta - m^2 = 0.} \tag{411}$$

Through separation of variables we have replaced Laplace's equation by three ODEs which are coupled through l and m. Let us now solve the ODEs. Equations (407) and (410) are solved by inspection

$$R_l(r) = a_l\, r^l + \frac{a_l}{r^{l+1}}\ , \tag{412}$$

$$\Phi_m(\phi) = C_m e^{im\phi} + D_m e^{-im\phi}. \tag{413}$$

As ϕ is an angular variable we must have $\Phi(\phi + 2\pi) = \Phi(\phi)$, which forces m to be an integer. For $m = 0$ we have

$$\Phi_0(\phi) = C_0 + D_0\phi. \tag{414}$$

This leaves us with (411). Changing variables to $x = \cos\theta$, $-1 \le x \le 1$, we find that Θ fulfils the **associated Legendre equation**

$$\boxed{\frac{d}{dx}\left[(1-x^2)\frac{d\Theta}{dx}\right] + \left[l(l+1) - \frac{m^2}{1-x^2}\right]\Theta(x) = 0.} \tag{415}$$

Non-singular (at $x = \pm 1$) solutions exist only for integer $l \geq 0$ and $|m| \leq l$. This is precisely the case we considered previously in the ODE section. The solutions are the associated Legendre polynomials $P_l^m(x)$. The general solution to Laplace's equation in spherical polar coordinates that is non-singular at $\theta = 0, \pi$ is therefore

$$\boxed{V(r,\theta,\phi) = \sum_{l=0}^{\infty} R_l(r) \sum_{m=0}^{l} \Phi_m(\phi)\ P_l^m(\cos\theta).} \tag{416}$$

Let us now return to our problem of a earthed sphere in an electric field. As nothing in our problem depends on ϕ we must have $\Phi(\phi) = const$. This fixes the coefficients C_m, D_m to be

$$D_m = C_m = 0\ , \qquad m = 1, 2, \dots\ , \qquad D_0 = 0. \tag{417}$$

Recalling that for $m = 0$ the associated Legendre polynomials reduce to the Legendre polynomials $P_l^0(\theta) = P_l(\theta)$, we conclude that

$$V(r,\theta) = \sum_{l=0}^{\infty} \left[a_l r^l + \frac{b_l}{r^{l+1}} \right] P_l(\cos(\theta))\ , \tag{418}$$

where P_l are Legendre polynomials. Let us now calculate

$$\begin{aligned} \int_0^\pi d\theta\ \sin(\theta)\ V(r,\theta)\ P_k(\cos(\theta) &= \sum_{l=0}^{\infty} \left[a_l r^l + \frac{b_l}{r^{l+1}} \right] \int_{-1}^{1} dx\ P_l(x)\ P_k(x), \\ &= \frac{2}{2k+1} \left[a_k r^k + \frac{b_k}{r^{k+1}} \right]. \end{aligned} \tag{419}$$

Here we have used the orthogonality relation for Legendre polynomials (305) to carry out the x-integral. On the other hand, for $r \to \infty$ we have by virtue of the boundary conditions (402)

$$\begin{aligned} \int_0^\pi d\theta\ \sin(\theta)\ V(r \to \infty,\theta)\ P_k(\cos(\theta) &= -Er \int_0^\pi d\theta\ \sin(\theta) \cos(\theta)\ P_k(\cos(\theta) \\ &= -Er \int_{-1}^{1} dx\ P_1(x)\ P_k(x) \\ &= -\frac{2}{3} Er \delta_{k,1}. \end{aligned} \tag{420}$$

Here we have used that

$$P_1(\cos(\theta)) = \cos(\theta). \tag{421}$$

Comparing equation (420) to (419) for $r \to \infty$ we conclude that

$$a_n = -E_0 \delta_{n,1}\ . \tag{422}$$

Similarly, by considering

$$\int_0^\pi d\theta \ \sin(\theta)\ V(r=a,\theta)\ P_k(\cos(\theta) \tag{423}$$

and then using the boundary conditions (403) we conclude that

$$b_n = E_0 a^3\ \delta_{n,1}\ . \tag{424}$$

The solution to our electrostatics problem is thus

$$V(r,\theta) = -Er\left(1-\frac{a^3}{r^3}\right) P_1(\cos(\theta)). \tag{425}$$

24.1 Spherical Harmonics

It is customary to combine the solutions to (410) and (411) to form the **spherical harmonics**. For $m \geq 0$ the latter are defined to be

$$Y_l^m(\theta,\phi) = (-1)^m \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}\ P_l^m(\cos\theta)\ e^{im\phi}. \tag{426}$$

For $m < 0$ they are dfined *via*

$$Y_l^{-|m|}(\theta,\phi) = (-1)^m \left(Y_l^{|m|}(\theta,\phi)\right)^*. \tag{427}$$

Using the spherical harmonics we can write the angular part of our solution (416) as

$$\sum_{m=0}^{l} \Phi_m(\phi) P_l^m(\cos\theta) = \sum_{m=-l}^{l} c_m Y_l^m(\theta,\phi), \tag{428}$$

where the c_m are linear combinations of C_m and D_m. The spherical harmonics form an orthonormal set of functions on the unit sphere (sphere of radius 1)

$$\boxed{\int_0^{2\pi} d\phi \int_0^\pi d\theta\ \sin\theta\ \left(Y_{l'}^{m'}(\theta,\phi)\right)^*\ Y_l^m(\theta,\phi) = \delta_{l,l'}\delta_{m,m'}.} \tag{429}$$

Furthermore, under very general conditions a function $f(\theta,\phi)$ of two angular variables can be expanded in terms of the spherical harmonics

$$\begin{aligned} f(\theta,\phi) &= \sum_{l=0}^\infty \sum_{m=-l}^{l} f_{lm}\ Y_l^m(\theta,\phi)\ , \\ f_{lm} &= \int_0^{2\pi} d\phi \int_0^\pi d\theta\ \sin\theta\ \left(Y_{l'}^{m'}(\theta,\phi)\right)^*\ f(\theta,\phi). \end{aligned} \tag{430}$$

25 Fourier Series

Fourier series and Fourier transforms are useful techniques for solving differential equations as well as integral equations. They are used for example in Quantum Mechanics and Electromagnetism.

Let us recall that one of our examples of linear vector spaces was a function space. One such function space familiar to you is the space of functions that can be expanded in a Taylor series around the point $x = 0$, which converges (to the value of the function) on an interval $[-a, a]$. A simple basis for this space is

$$\{1, x, x^2, x^3, \ldots\} \ . \tag{431}$$

We write

$$f(x) = f(0) + f'(0)x + \frac{1}{2}f''(0)x^2 + \ldots \tag{432}$$

In terms of our function space this is written in Dirac notation as

$$|f\rangle = f(0)|1\rangle + f'(0)|x\rangle + \frac{f''(0)}{2}|x^2\rangle + \ldots . \tag{433}$$

A very nice property about the Taylor expansion is that keeping the first few terms can give an excellent approximation to $f(x)$. This is of course of great practical importance as in many problems one needs to know a function only in close vicinity of a given point.

Many function do not have good Taylor expansions, but can be expanded in terms of other basis functions. An example are Fourier series and Fourier transforms. These are expansions in terms of plane waves or standing waves.

25.1 Periodic Functions with Period 2π

A function is periodic with period α if for every x

$$f(x + \alpha) = f(x). \tag{434}$$

An example of a 2π-periodic function is shown in Fig. 14. Clearly it is sufficient to know $f(x)$ in the interval $[-\pi, \pi]$ in order to determine its value for any x by repeated application of (434).

Note that there is nothing special about choosing this interval to be $[-\pi, \pi]$. As is shown in Fig.15, we could equally well choose the interval $[0, 2\pi]$ or in fact any interval of length 2π.

Our convention is to choose the interval to be $[-\pi, \pi]$. From now on we consider functions on this interval only, having in mind that for all other values of x the functions is defined by (434) with $\alpha = 2\pi$.

Hence the periodic function in our example takes the form shown in Fig. 16.

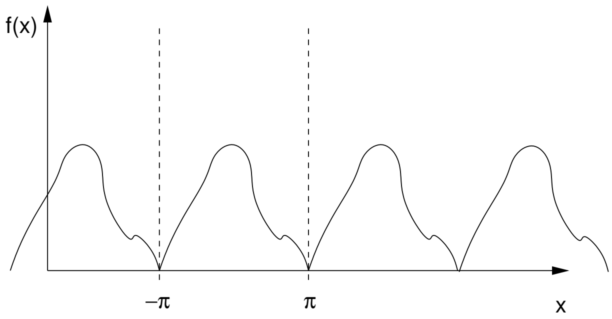


Figure 14: A periodic function with period 2π defined on the entire real axis.

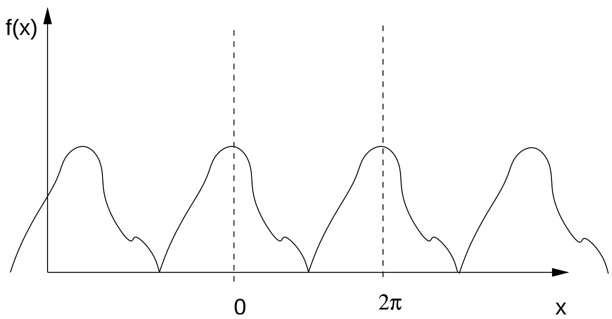


Figure 15:

Conversely, if we are given a function defined on the interval $[-\pi, \pi]$ we can always construct a 2π-periodic function on the entire real axis from it by using (434).

Let us for the time being require our periodic functions to be **continuous**, which implies that

$$f(\pi) = f(-\pi). \tag{435}$$

Continuous periodic functions with period 2π form a linear vector space. Indeed, if $f(x)$ and $g(x)$ are 2π-periodic functions, then so is $h(x) = \alpha f(x) + \beta g(x)$ (α, β are numbers):

$$h(x + 2\pi) = \alpha f(x + 2\pi) + \beta g(x + 2\pi) = \alpha f(x) + \beta g(x) = h(x). \tag{436}$$

The null element is the periodic constant function $z(x) = 0$.

Consider now the following set of functions on the interval $[-\pi, \pi]$:

$$\begin{aligned} c_n(x) &= \frac{1}{\sqrt{\pi}} \cos(nx) \;, \qquad n = 1, 2, \ldots \\ s_n(x) &= \frac{1}{\sqrt{\pi}} \sin(nx) \;, \qquad n = 1, 2, \ldots \end{aligned}$$

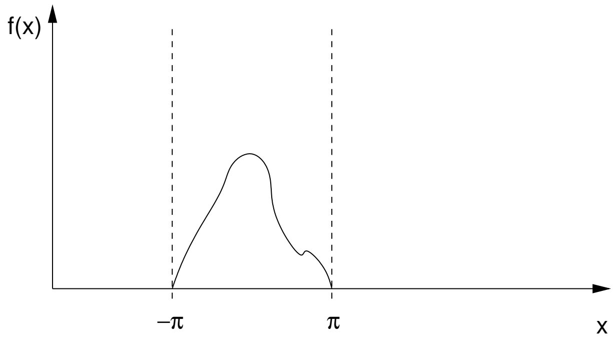


Figure 16: A periodic function on the interval $[-\pi, \pi]$.

$$c_0(x) \quad = \quad \frac{1}{\sqrt{2\pi}}. \tag{437}$$

Theorem: These functions form an orthonormal set with respect to the scalar product

$$\langle f|g\rangle = \int_{-\pi}^{\pi} dx f^*(x)g(x), \tag{438}$$

i.e. we have

$$\begin{aligned}
\langle c_n|c_m\rangle &= \int_{-\pi}^{\pi} dx\ c_n(x)\ c_m(x) = \delta_{nm}, \\
\langle s_n|s_m\rangle &= \int_{-\pi}^{\pi} dx\ s_n(x)\ s_m(x) = \delta_{nm}, \\
\langle c_n|s_m\rangle &= \int_{-\pi}^{\pi} dx\ c_n(x)\ s_m(x) = 0.
\end{aligned} \tag{439}$$

Proof:

$$\langle s_n|s_m\rangle \quad = \quad \int_{-\pi}^{\pi} \frac{dx}{\pi} \sin(nx)\ \sin(mx) = \int_{-\pi}^{\pi} \frac{dx}{2\pi}[\cos((n-m)x) - \cos((n+m)x)]. \tag{440}$$

For $n \neq m$ the integral of each individual term in the integrand is simply zero. For $n = m$ the integral over the second term is zero, but the first term is equal to 1 and hence the integral equals 1. All other orthonormality relations are proved analogously. QED.

Recalling that orthogonality implies linear independence an obvious question is then whether or not the set (437) of functions in fact forms a basis of the linear vector space of continuous periodic functions. It can be shown (see e.g. Dennery/Krzywicki) that they do. Hence we can expand any continuous periodic function $f(x)$ in a **Fourier series**

$$\begin{aligned} f(x) &= \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[a_n \cos(nx) + b_n \sin(nx) \right] , \\ a_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} dx\ f(x)\ \cos(nx)\ , \\ b_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} dx\ f(x)\ \sin(nx)\ . \end{aligned} \tag{441}$$

In Dirac notation this expansion reads

$$|f\rangle = \sum_{n=0}^{\infty} \langle c_n|f\rangle\ |c_n\rangle + \sum_{n=0}^{\infty} \langle s_n|f\rangle\ |s_n\rangle\ . \tag{442}$$

$$\begin{aligned} a_0 &= \sqrt{\frac{2}{\pi}} \langle c_0|f\rangle\ , \\ a_n &= \frac{1}{\sqrt{\pi}} \langle c_n|f\rangle\ , \quad n = 1, 2, \ldots\ , \\ b_n &= \frac{1}{\sqrt{\pi}} \langle s_n|f\rangle\ . \end{aligned} \tag{443}$$

It turns out that it is not really necessary for a periodic function to be continuous for it to be representable by a Fourier series. In particular, the function is allowed to have jumps. More precisely the so-called **Dirichlet conditions** must be fulfilled:

- $f(x)$ is periodic;
- $f(x)$ is continuous except possibly at a finite number of finite discontinuities;
- $f(x)$ must only have a finite number of maxima and minima within one period;
- the integral over one period of $|f(x)|$ must converge.

If the Dirichlet conditions are fulfilled, the Fourier series converges to

- $f(x)$ for all x at which $f(x)$ is continuous;
- $\frac{1}{2} \lim_{\epsilon \to 0} \left[f(\pi) + f(-\pi) \right]$ at $x = \pm\pi$.
- $\frac{1}{2} \lim_{\epsilon \to 0} \left[f(x + \epsilon) + f(x - \epsilon) \right]$ at all points x where f has discontinuities.

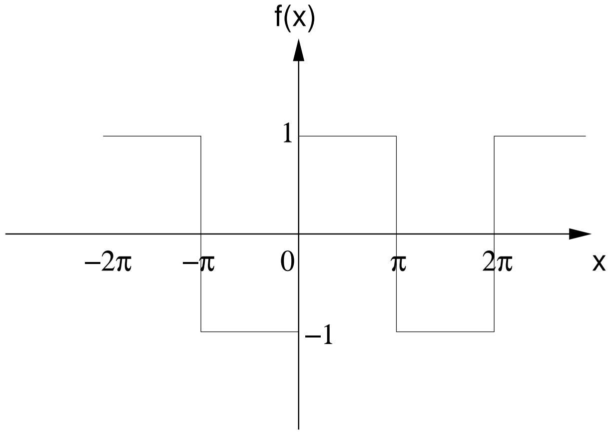


Figure 17: A periodic function on the interval $[-\pi, \pi]$.

Example:

$$f(x) = \begin{cases} 1 & 0 \leq x < \pi \\ -1 & -\pi \leq x < 0 \end{cases} \tag{444}$$

This function has discontinuities at $x = 0$ and $x = \pm\pi$, see Fig.17. Let us work out the Fourier coefficients a_n and b_n. The function $f(x)$ is **odd** ($f(-x) = -f(x)$) and therefore

$$a_n = 0. \tag{445}$$

The b_n's are given by

$$\begin{aligned} b_n &= \frac{1}{\pi}\int_{-\pi}^{0} dx(-1)\ \sin(nx) + \frac{1}{\pi}\int_{0}^{\pi} dx(+1)\ \sin(nx) \\ &= \begin{cases} \frac{4}{n\pi} & \text{n odd ,} \\ 0 & \text{n even.} \end{cases} \end{aligned} \tag{446}$$

Hence

$$f(x)'' ='' \sum_{n=1}^{\infty} \frac{4}{(2n-1)\pi}\ \sin((2n-1)x)\ . \tag{447}$$

We have put quotation marks in order to indicate that equality holds only for points x where $f(x)$ is continuous. Note that at the discontinuties we have e.g.

$$f(0) = 1\ , \qquad \lim_{\epsilon\to 0} \frac{f(\epsilon) + f(-\epsilon)}{2} = 0\ . \tag{448}$$

Evaluating both the function $f(x)$ and its Fourier series at $x = \frac{\pi}{2}$ (where the function is continuous), we obtain a very nice identity

$$f(\frac{\pi}{2}) = 1 = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{4}{(2n-1)\pi}. \tag{449}$$

Let us now see how well the Fourier series approximates the function if we keep only a finite number of modes. Let us define

$$f_k(x) = \sum_{n=1}^{k} \frac{4}{(2n-1)\pi} \sin((2n-1)x). \tag{450}$$

We plot $f_k(x)$ for several values of k in Fig.18. Clearly increasing k improves the approx-

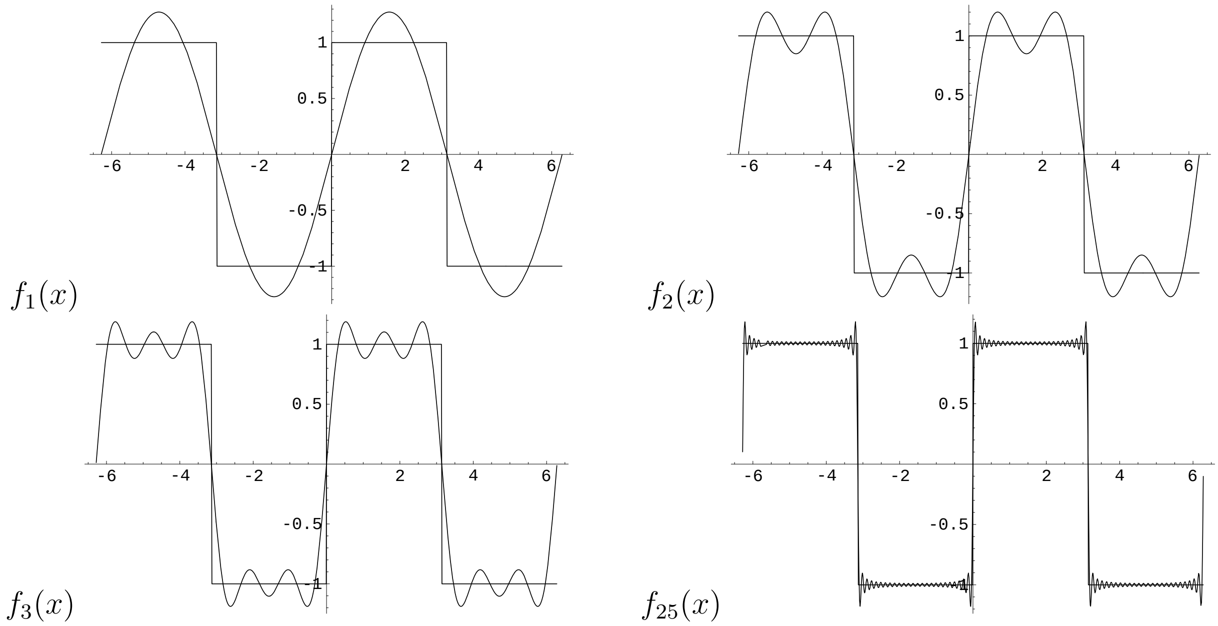


Figure 18: Approximations keeping a finite number of modes.

imation. Note the overshoot of the Fourier series in the vicinity of the discontinuities of $f(x)$. This overshoot moves arbitrarily close to discontinuity if we increase k, but never disappears.

25.2 Periodic Functions with Period L

Suppose that $f(x)$ is a periodic function with period L, i.e.

$$f(x+L) = f(x). \tag{451}$$

Such functions can be expanded in a Fourier series in terms of the following orthonormal set of basis functions

$$\begin{aligned} c'_n(x) &= \sqrt{\frac{2}{L}} \cos\Big(\frac{2\pi n}{L} x\Big) , \\ s'_n(x) &= \sqrt{\frac{2}{L}} \sin\Big(\frac{2\pi n}{L} x\Big) . \end{aligned} \tag{452}$$

We have

$$\begin{aligned}
\int_{-L/2}^{L/2} dx\ c_n'(x)\ c_m'(x) &= \delta_{nm}\ , \quad \int_{-L/2}^{L/2} dx\ s_n'(x)\ s_m'(x) = \delta_{nm}\ , \\
\int_{-L/2}^{L/2} dx\ c_n'(x)\ s_m'(x) &= 0\ .
\end{aligned} \tag{453}$$

The Fourier series is given by

$$\begin{aligned}
f(x) &= \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[a_n \cos\Big(\frac{2\pi n}{L}x\Big) + b_n \sin\Big(\frac{2\pi n}{L}x\Big) \right], \\
a_n &= \frac{2}{L} \int_{-L/2}^{L/2} dx\ f(x)\ \cos\Big(\frac{2\pi n}{L}x\Big)\ , \\
b_n &= \frac{2}{L} \int_{-L/2}^{L/2} dx\ f(x)\ \sin\Big(\frac{2\pi n}{L}x\Big)\ .
\end{aligned} \tag{454}$$

Proof: The change variables to $x = Lz/2\pi$ gives us a function that is 2π-periodic in z. Hence it can be expanded in a Fourier series in z

$$\begin{aligned}
f\Big(\frac{Lz}{2\pi}\Big) &= \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[a_n \cos(nz) + b_n \sin(nz) \right], \\
a_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} dz\ f\Big(\frac{Lz}{2\pi}\Big)\ \cos(nz)\ , \\
b_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} dz\ f\Big(\frac{Lz}{2\pi}\Big)\ \sin(nz)\ .
\end{aligned} \tag{455}$$

Changing variables back from z to x now gives the equations stated above. QED.

In Dirac notation we have

$$|f\rangle = \sum_{n=0}^{\infty} \langle c_n|f\rangle\ |c_n\rangle + \sum_{n=0}^{\infty} \langle s_n|f\rangle\ |s_n\rangle\ . \tag{456}$$

$$\begin{aligned}
a_0 &= \frac{2}{\sqrt{L}} \langle c_0|f\rangle\ , \\
a_n &= \sqrt{\frac{2}{L}} \langle c_n|f\rangle\ , \quad n = 1, 2, \dots\ , \\
b_n &= \sqrt{\frac{2}{L}} \langle s_n|f\rangle\ .
\end{aligned} \tag{457}$$

25.3 Parseval's Theorem

Parseval's theorem is a "sum rule" for the Fourier coefficients

$$\frac{1}{L}\int_{-L/2}^{L/2} dx\ |f(x)|^2 = \left(\frac{a_0}{2}\right)^2 + \frac{1}{2}\sum_{n=1}^{\infty}(a_n^2 + b_n^2)\ . \tag{458}$$

Equation (458) is a simple consequence of the orthonormality relations (453).
Exercise: Prove Parseval's theorem.

25.4 Fourier Series using Complex Exponentials

Clearly we may use complex exponentials instead of sines and cosines in order to expand a periodic function in a Fourier series. The appropriate orthonormal set on the interval $[-L/2, L/2]$ is

$$\begin{aligned} e_n(x) &= \frac{1}{\sqrt{L}}\exp\Big(\frac{2\pi inx}{L}\Big) = e_n(x+L)\ , \qquad n = 0, \pm 1, \pm 2, \dots \\ \langle e_n|e_m\rangle &= \frac{1}{L}\int_{-L/2}^{L/2} dx \exp\Big(-\frac{2\pi inx}{L}\Big)\ \exp\Big(\frac{2\pi imx}{L}\Big) = \delta_{nm}. \end{aligned} \tag{459}$$

The Fourier series for the function $f(x) = f(x+L)$ is

$$\begin{aligned} f(x) &= \sum_{n=-\infty}^{\infty} c_n \exp\Big(\frac{2\pi imx}{L}\Big)\ , \\ c_n &= \frac{1}{L}\int_{-L/2}^{L/2} dx \exp\Big(-\frac{2\pi inx}{L}\Big)\ f(x) \equiv \frac{1}{\sqrt{L}}\langle e_n|f\rangle. \end{aligned} \tag{460}$$

In Dirac notation

$$|f\rangle = \sum_{-\infty}^{\infty}\langle e_n|f\rangle\ |e_n\rangle\ . \tag{461}$$

25.5 Fourier Sine and Cosine Series

Non-periodic functions defined on an interval $[a, b]$ can be represented as Fourier series by *continuing* them outside $[a, b]$ in order to make them periodic. An example of what we mean by this is shown in Figs 19 and 20. In Fig. 20 two different periodic continuations of the function $f(x)$ defined on the interval $[0, a]$ are shown.

Both continuations coincide with $f(x)$ on the interval $[0, a]$ and both continuations can be expressed in terms of Fourier series. These Fourier series will be different, but both of them will coincide with $f(x)$ on the interval $[0, a]$! In the simplest case one has a function

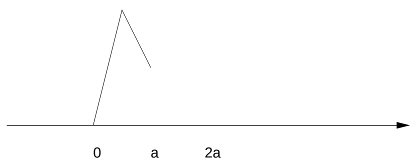


Figure 19: A function $f(x)$ defined on the interval $[0, a]$.

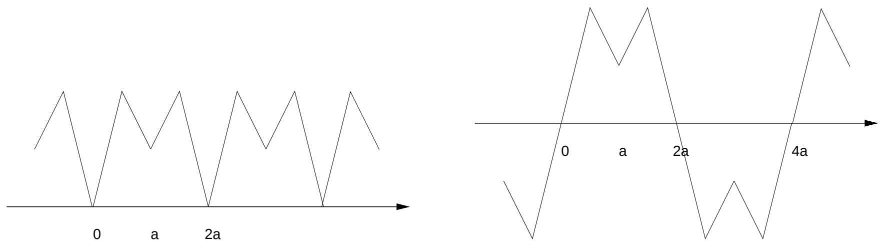


Figure 20: Two different periodic continuations of $f(x)$ with periods $2a$ and $4a$ respectively.

$g(x)$ defined on $[0, L]$ with $g(0) = g(L)$. Then there are two generic ways of continuing this function to the interval $[-L, L]$: a symmetric continuation $g_1(x) = g_1(-x) = g(x)$ and and antisymmetric one $g_2(x) = -g_2(-x) = g(x)$. The Fourier series for $g_1(x)$ will only contain cosines and is called the *Fourier cosine series*. The Fourier series for $g_2(x)$ will only contain sines and is called the *Fourier sine series*.

26 Fourier Transform

We can make the period L of the periodic functions we expand in Fourier series very large. An obvious questions is then whether we can take the limit $L \to \infty$, i.e. consider functions on $(-\infty, \infty)$, which then of course don't have to be periodic any longer. In order to carry out this limit, let us write the Fourier series as

$$f(x) = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} F(k_n)\, e^{ik_n x}\, (k_{n+1} - k_n) \ , \tag{462}$$

$$F(k_n) = \frac{1}{\sqrt{2\pi}} \int_{-L/2}^{L/2} dx\, e^{-ik_n x}\, f(x) \equiv \frac{L}{\sqrt{2\pi}} c_n \ , \tag{463}$$

where

$$k_n = \frac{2\pi n}{L} \ . \tag{464}$$

We can now formally take the limit $L \to \infty$, in which (462) turns into an integral

$$\begin{aligned} f(x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk\ F(k)\ e^{ikx}\ , \\ F(k) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx\ e^{-ikx}\ f(x)\ . \end{aligned} \tag{465}$$

In order for the limit $L \to \infty$ to exist, the function $f(x)$ must fulfil certain conditions ("Plancherel Theorem"). For square-integrable functions (as defined in (28)) the Fourier transform exists.

- $F(k)$ is called the Fourier transform of $f(x)$;
- $f(x)$ is called the inverse Fourier transform of $F(k)$,

The Fourier transform relates two functions $f(x)$ and $F(k)$. In the literature many different conventions for the Fourier transform are employed. For example you may find

$$\begin{aligned} f(x) &= \frac{\alpha}{2\pi} \int_{-\infty}^{\infty} dk\ \bar{F}(k)\ e^{\pm ik}\ , \\ \bar{F}(k) &= \frac{1}{\alpha} \int_{-\infty}^{\infty} dx\ e^{\mp ikx}\ f(x)\ , \end{aligned} \tag{466}$$

where α is taken to be 1, $\sqrt{2\pi}$ or 2π. The Fourier transform $\bar{F}(k)$ defined in this way is related to the one in our convention by

$$\bar{F}(k) = \frac{\sqrt{2\pi}}{\alpha} F(\pm k)\ . \tag{467}$$

Examples:

- $f(x) = \begin{cases} Ae^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$ and $\lambda > 0$.

  The Fourier transform is

$$F(k) = \frac{A}{\sqrt{2\pi}} \int_0^{\infty} dx\ e^{-ikx-\lambda x} = \frac{A}{\sqrt{2\pi}} \frac{e^{-ikx-\lambda x}}{-ik-\lambda}\bigg|_0^{\infty} = \frac{A}{\sqrt{2\pi}(\lambda + ik)}. \tag{468}$$

- $f(x) = \exp(-x^2/2a^2)$

  The Fourier transform is

$$F(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx \exp\left(-\frac{x^2}{2a^2} - ikx\right)$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-ik)^n}{n!} \int_{-\infty}^{\infty} dx\ x^n \exp\left(-\frac{x^2}{2a^2}\right) \\
&= \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-ik)^{2n}}{2n!} \int_{-\infty}^{\infty} dx\ x^{2n} \exp\left(-\frac{x^2}{2a^2}\right) \\
&= \sum_{n=0}^{\infty} \frac{(-ik)^{2n}}{2n!} a^{2n+1}(2n-1)!! = a \exp\left(-\frac{k^2a^2}{2}\right). \qquad (469)
\end{aligned}
$$

Here we have used that

$$
\begin{aligned}
I_n(a) &= \int_{-\infty}^{\infty} dx\ x^{2n} \exp\left(-\frac{x^2}{2a^2}\right) = -a^2 \int_{-\infty}^{\infty} dx\ x^{2n-1} \frac{d}{dx} \exp\left(-\frac{x^2}{2a^2}\right) \\
&= a^2(2n-1) \int_{-\infty}^{\infty} dx\ x^{2n-2} \exp\left(-\frac{x^2}{2a^2}\right) = (2n-1)a^2 I_{n-1}(a). \qquad (470)
\end{aligned}
$$

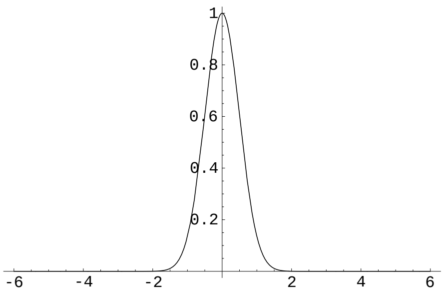
Needless to say that a **much simpler** way of doing this integral is to use complex analysis methods (so take that Short Option!)

$$
\begin{aligned}
F(k) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx \exp\left(-\frac{x^2}{2a^2} - ikx\right) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx \exp\left(-\frac{(x+ika^2)^2}{2a^2} - \frac{k^2a^2}{2}\right) \\
&= \exp\left(-\frac{k^2a^2}{2}\right) \frac{1}{\sqrt{2\pi}} \int_{-\infty+ika^2}^{\infty+ika^2} dx \exp\left(-\frac{z^2}{2a^2}\right) \\
&= \exp\left(-\frac{k^2a^2}{2}\right) \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx \exp\left(-\frac{z^2}{2a^2}\right) = a \exp\left(-\frac{k^2a^2}{2}\right). \qquad (471)
\end{aligned}
$$

Here we have shifted the path of integration back to the real axis and used the fact that we encounter no singularities of the integrand when doing this.

We have obtained an interesting result: the Fourier transform of a Gaussian is a Gaussian. However $f(x)$ had a standard deviation $a/\sqrt{2}$, whereas $F(k)$ has a standard deviation of $1/(a\sqrt{2})$.

Aside: This fact is related to the Heisenberg uncertainty relation in Quantum Mechanics. The ground state wave function of the simple harmonic oscillator is a Gaussian both in the position and the momentum representation and the two are related by Fourier transformation. We can calculate the mean square deviation of the position operator $\Delta x = \sqrt{\langle x^2 \rangle - \langle x \rangle^2} = a/\sqrt{2}$ in the position representation and the mean square deviation of the momentum operator $\Delta p = \sqrt{\langle p^2 \rangle - \langle p \rangle^2} = \hbar/a\sqrt{2}$ in the momentum representation. Because the width of the two Gaussians are inversely proportional we find that $\Delta x\ \Delta p = \hbar/2$: the ground state of the simple harmonic oscillator is a "state of minimum uncertainty".



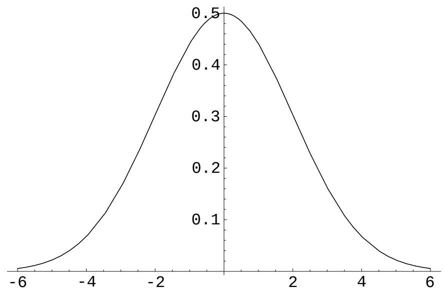


Figure 21: $f(x)$ and its Fourier transform $F(k)$

- $f(x) = \frac{1}{1+x^2}$
  The Fourier transform is $F(k) = \sqrt{\frac{\pi}{2}} e^{-|k|}$.
  Exercise: Prove this by using the residue theorem.

26.1 Properties of the Fourier Transform

Let $f(x)$ by a Fourier transformable function and $F(k)$ its Fourier transform. Then

1. The Fourier Transform of $f(ax)$ is $\frac{1}{a}F(k/a)$.

Proof: Substitute $y = ax$

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx \ f(ax) \ e^{-ikx} = \frac{1}{\sqrt{2\pi}a} \int_{-\infty}^{\infty} dy \ f(y) \ e^{-iky/a} \ . \tag{472}$$

2. The Fourier Transform of $f(a + x)$ is $e^{ika}F(k)$.

Proof: Substitute $y = x + a$.

3. The Fourier Transform of $e^{iqx}f(x)$ is $F(k - q)$.

Proof: By inspection.

4. The Fourier Transform of $\frac{d}{dx}f(x)$ is $ikF(k)$.

Proof: integrate by parts

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx \ f'(x) \ e^{-ikx} = \frac{f(x)e^{-ikx}}{\sqrt{2\pi}}\Big|_{-\infty}^{\infty} + \frac{ik}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dy \ f(x) \ e^{-ikx} \ . \tag{473}$$

The first term on the RHS must vanish in order for the Fourier transform to exist.

5. The Fourier Transform of $xf(x)$ is $i\frac{d}{dk}F(k)$.

Proof: differentiate under the integral (apparently one of Feynman's favourite tricks)

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dx\ xf(x)\ e^{-ikx} = i\frac{d}{dk}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dx\ f(x)\ e^{-ikx}. \tag{474}$$

26.2 Fourier Transform in Higher Dimensions

The Fourier transform is readily generalized to more than one dimension. In three dimensions it is defined as

$$F(k_x, k_y, k_z) = \frac{1}{(2\pi)^{3/2}}\int_{-\infty}^{\infty} dx\ dy\ dz\ e^{-ik_x x - ik_y y - k_z} f(x, y, z). \tag{475}$$

Using a vector notation $\mathbf{r} = (x, y, z)$, $\mathbf{k} = (k_x, k_y, k_z z)$ this can be written as

$$F(\mathbf{k}) = \frac{1}{(2\pi)^{3/2}}\int_{-\infty}^{\infty} d^3\mathbf{r} e^{-i\mathbf{k}\cdot\mathbf{r}} f(\mathbf{r}). \tag{476}$$

26.3 Convolutions

The function

$$f(x) = \int_{-\infty}^{\infty} dy\ h(x-y)\ g(y) \tag{477}$$

is called the convolution of f with h. It is denoted by $f = g * h$. Convolution is a commutative operation, i.e.

$$f * g = g * f. \tag{478}$$

This follows from substituting $y' = x - y$ in (477). A convolution has the important property that its Fourier transform is equal to $\sqrt{2\pi}$ times the product of the Fourier transforms

$$\begin{aligned} F(k) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy g(y)\ h(x-y) \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dy\ g(y)\ e^{-iky}\int_{-\infty}^{\infty} dx\ h(x-y)\ e^{-ik(x-y)} = \int_{-\infty}^{\infty} dy\ g(y)\ e^{-iky} H(k) \\ &= \sqrt{2\pi} G(k) H(k). \end{aligned} \tag{479}$$

26.4 Solving Integral Equations by Fourier Transform

An integral equation is an equation of the form

$$f(x) = f_0(x) + \int_{-\infty}^{\infty} dy\ G(x-y)\ f(y)\ , \tag{480}$$

where the "driving term" $f_0(x)$ and the "kernel" $G(x)$ are two known functions and we would like to find functions $f(x)$ that fulfil the integral equation. Integral equations are encountered quite frequently in Physics: examples are the Lippmann-Schwinger equation in QM scattering theory, the Dyson equations in diagrammatic perturbation theory etc. Equations of the type (480) are solvable by Fourier transform. Indeed, the integral equation involves a convolution

$$f(x) = f_0(x) + (G * f)(x). \tag{481}$$

Hence

$$F(k) = F_0(k) + \sqrt{2\pi}G(k)F(k) \ , \tag{482}$$

which gives us the Fourier transform $F(k)$ as

$$F(k) = \frac{F_0(k)}{1 - \sqrt{2\pi}G(k)}. \tag{483}$$

The inverse Fourier transform gives us the desired solution of the integral equation.

27 The Dirac Delta Function

The delta function is an example of a "generalized function". Unfortunately we do not have the time to study this important subject is any detail, so have a look at the books by Dennery and Krzywicki and by Lighthill.

Let us a start by considering the following set of very well-behaved functions

$$d_n(x) = \sqrt{\frac{n}{\pi}} e^{-nx^2} \ , \qquad n = 1, 2, \ldots \tag{484}$$

These functions have the properties that

- $\int_{-\infty}^{\infty} dx \ d_n(x) = 1.$
- $\frac{d^k}{dx^k} d_n(x)$ exists for any k and goes to zero faster than any power of $\frac{1}{|x|}$ when $x \to \pm\infty$.

We define the delta function formally as the limit

$$\delta(x) = \lim_{n\to\infty} d_n(x). \tag{485}$$

As is shown in Fig.22, when n increases $d_n(x)$ becomes more and more peaked around $x = 0$. In the limit $n \to \infty$ we end up with a "function" that is zero everywhere except at $x = 0$, where it is infinite in such a way that

$$\int_{-\infty}^{\infty} dx \ \delta(x) = 1. \tag{486}$$

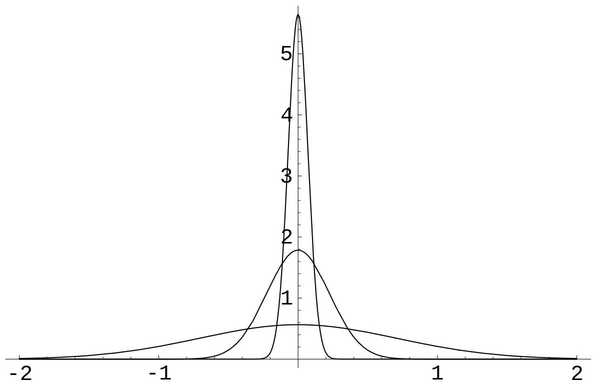


Figure 22: $d_n(x)$ for $n = 1, 10, 100$.

This clearly does not make sense for a function in the usual sense. All properties of the Dirac δ-function are to be understood in terms of the limit $n \to \infty$. For example, for any "well-behaved" [8] function $f(x)$ we have

$$\boxed{\int_{-\infty}^{\infty} dx\ f(x)\ \delta(x - x_0) = f(x_0).} \tag{487}$$

Proof: Let us Taylor-expand $f(x)$ around $x = x_0$

$$\begin{aligned}
\int_{-\infty}^{\infty} dx\ f(x)\ \delta(x - x_0) &\equiv \lim_{n\to\infty} \int_{-\infty}^{\infty} dx\ f(x)\ d_n(x - x_0) \\
&= \lim_{n\to\infty} \int_{-\infty}^{\infty} dx\ [f(x_0) + f'(x_0)(x - x_0) + \ldots]\ \ d_n(x - x_0) \\
&= \lim_{n\to\infty} \Big\{ f(x_0) \int_{-\infty}^{\infty} dx\ \ d_n(x - x_0) \\
&\qquad + \frac{f''(x_0)}{2} \int_{-\infty}^{\infty} dx\ (x - x_0)^2\ d_n(x - x_0) + \ldots \Big\} \\
&= \lim_{n\to\infty} \Big\{ f(x_0) + \frac{f''(x_0)}{2} \frac{1}{2n} + \ldots \Big\} = f(x_0).
\end{aligned} \tag{488}$$

Note that we have interchanged the limit $\lim_{n\to\infty}$ with the integral over x in the first step. This **defines** what we mean by the integral over the delta function (or in fact any other generalized function).

[8]By this we mean functions that are infinitely many times differentiable if its modulus and that of its derivatives does not increase faster than some power of x for $|x| \to \infty$. Polynomials are well behaved in this sense.

You often will find the equation

$$f(x)\ \delta(x - x_0) = f(x_0)\ \delta(x - x_0) \tag{489}$$

in the literature. This equations is to be understood **under an integral** and you may view it as a shorthand for (487).

Note that as $\delta(x - x_0)$ is zero everywhere except at $x = x_0$ we have $(\epsilon, \epsilon' > 0)$

$$\int_{\infty}^{\infty} dx\ f(x)\ \delta(x - x_0) = \int_{x_0-\epsilon}^{x_0+\epsilon'} dx\ f(x)\ \delta(x - x_0). \tag{490}$$

Another important identity invoving the delta function is

$$\boxed{\int_{-\infty}^{\infty} dx\ f(x)\ \delta(y(x)) = \sum_j \frac{f(x_j)}{|y'(x_j)|}.} \tag{491}$$

Here the sum is over the roots x_j of the equation $y(x) = 0$ and $y'(x)$ is the derivative of y with respect to x. **We assume that $y'(x_j) \neq 0$ for all x_j.** Let us see how (491) comes about. As $\delta(y(x))$ is zero everywhere except at the points x_j where $y(x_j) = 0$, we may use (490) to write

$$\int_{-\infty}^{\infty} dx\ f(x)\ \delta(y(x)) = \sum_j \int_{x_j-\epsilon}^{x_j+\epsilon} dx\ f(x)\ \delta(y(x)). \tag{492}$$

For each of these integrals we now change variables. We first note that sufficiently close to x_j we may always invert the function $y(x)$ because $y'(x_j) \neq 0$. We therefore can change variables from x to $y(x)$. Using that $dy = y'(x)dx$ we obtain for the j^{th} term of the sum

$$\int_{y(x_j-\epsilon)}^{y(x_j+\epsilon)} \frac{dx}{y'(x)}\ f(x(y))\ \delta(y) = \frac{f(x_j)}{|y'(x_j)|}, \tag{493}$$

where we have used (487), (490) to evaluate the integral. The absolute value on the RHS appears because if $y'(x_j) < 0$ then $y(x_j + \epsilon) < y(x_j - \epsilon)$ and we pick up an extra minus sign from $\int_a^b dx = -\int_b^a dx$.

Under an integral we therefore may write

$$\boxed{\delta(y(x)) = \sum_j \frac{\delta(x - x_j)}{|y'(x_j)|}\ ,} \tag{494}$$

where the sum is over all roots x_j of $y(x) = 0$. A special case of this equation is

$$\boxed{\delta(ax) = \frac{1}{|a|}\delta(x).} \tag{495}$$

Finally, setting $a = -1$ we find that the delta function is formally even

$$\boxed{\delta(-x) = \delta(x).} \tag{496}$$

The properties of the delta function are **independent** of which sequence of functions one uses to define it. Some other sequences are used to define the delta function are

$$\delta(x) = \lim_{\epsilon \to 0} \frac{1}{\pi} \frac{\epsilon}{x^2 + \epsilon^2} , \tag{497}$$

$$\delta(x) = \lim_{n \to \infty} \frac{\sin(nx)}{\pi x} . \tag{498}$$

27.1 Derivative of the Delta Function

The derivative of the delta function is defined as the limit

$$\delta'(x) = \lim_{n \to \infty} d'_n(x). \tag{499}$$

The defining property of the derivative of the delta function is that for any "well-behaved" function $f(x)$

$$\int_{-\infty}^{\infty} dx\ f(x)\ \delta'(x - x_0) = -f'(x_0). \tag{500}$$

Let us again see how this comes about:

$$\begin{aligned}
\int_{-\infty}^{\infty} dx\ f(x)\ \delta'(x - x_0) &\equiv \lim_{n \to \infty} \int_{-\infty}^{\infty} dx\ f(x)\ d'_n(x - x_0) \\
&= \lim_{n \to \infty} \left\{ f(x) d_n(x - x_0) \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} dx\ f'(x)\ d_n(x - x_0) \right\} \\
&= -\int_{-\infty}^{\infty} dx\ f'(x)\ \delta(x - x_0) = -f'(x_0).
\end{aligned} \tag{501}$$

The first line merely states the definition of the derivative of the delta function. In going from the first line to the second we have integrated by parts. We then have used that for any well-behaved function $\lim_{x \to \pm\infty} d'_n(x) f(x) = 0$.

Note that while the delta function is formally even, its derivative is odd

$$\delta'(-x) = -\delta'(x). \tag{502}$$

Exercise: Show that this is true.

27.2 Multidimensional Delta Function

The delta function is readily generalized to more than one dimension. The three dimensional delta function is defined as

$$\delta^{(3)}(\mathbf{r}) \equiv \delta(x)\ \delta(y)\ \delta(z). \tag{503}$$

It has the porperty that

$$\begin{aligned}\int d^3\mathbf{r}\ f(\mathbf{r})\ \delta^{(3)}(\mathbf{r}-\mathbf{r_0}) &= \int dx\ dy\ dz\ f(x,y,z)\ \delta(x-x_0)\ \delta(y-y_0)\ \delta(z-z_0) \\ &= f(x_0,y_0,z_0) = f(\mathbf{r_0}).\end{aligned} \tag{504}$$

27.3 Fourier Transform of the Delta Function

For any $d_n(x)$ we can calculate its Fourier transform $D_n(k)$:

$$D_n(k) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dx\ e^{-ikx}\sqrt{\frac{n}{\pi}}e^{-nx^2} = \frac{1}{\sqrt{2\pi}}e^{-k^2/4n}. \tag{505}$$

The inverse Fourier transform is

$$d_n(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dk\ e^{ikx}\ D_n(k). \tag{506}$$

Let us now take the limit $n \to \infty$ of (506) using that $\lim_{n\to\infty} d_n(x) = \delta(x) =$

$$\begin{aligned}\delta(x) &= \lim_{n\to\infty}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dk\ e^{ikx}\ D_n(k) \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dk\ e^{ikx}\ \lim_{n\to\infty} D_n(k) \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\ e^{ikx}\ .\end{aligned} \tag{507}$$

We conclude that

$$\boxed{\delta(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\ e^{ikx} = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\ e^{-ikx}\ .} \tag{508}$$

This is very strange! We know that the integral in (508) does not exist, so what does this mean? [9] We said before that the delta function makes sense only under an integral. So let us look at the equation

$$\int_{-\infty}^{\infty} dx\ f(x)\ \delta(x-x_0) = f(x_0). \tag{509}$$

[9]Note that for any finite n the integrals we encounter are perfectly well defined.

Using equation (508) to represent the delta function on the RHS we obtain

$$\begin{aligned}\int_{-\infty}^{\infty} dx\ f(x)\ \delta(x-x_0) &= \int_{-\infty}^{\infty} dx\ f(x)\ \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\ e^{-ik(x-x_0)} \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dk\ e^{ikx_0}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dx\ e^{-ikx}\ f(x) \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dk\ e^{ikx_0}\ F(k) = f(x_0). \qquad (510)\end{aligned}$$

In the second line we have interchanged the order of the x and k integration. In the third line we have used the definition of the Fourier transform $F(k)$ and finally the definition of the inverse Fourier transform (465). We see that under an integral our equation for the Fourier transform of the delta function makes perfect sense!

27.4 Normalization Condition in the Case of Continuous Spectra

Recall that for Fourier series we had a basis functions

$$e_n(x) = \frac{1}{\sqrt{2\pi}}e^{inx}\ , \qquad (511)$$

with

$$\langle \mathbf{e_n} \mid \mathbf{e_m}\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} dx\ e^{ix(m-n)} = \delta_{nm}. \qquad (512)$$

These basis functions are labelled by a discrete parameter, the integer n.

Four Fourier transforms we have a basis set labelled by a *continuous parameter*, the real number k

$$e_k(x) = \frac{1}{\sqrt{2\pi}}e^{ikx}\ . \qquad (513)$$

Now the normalization condition reads

$$\langle \mathbf{e_k} \mid \mathbf{e_p}\rangle = \frac{1}{2\pi}\int_{-\infty}^{\infty} dx\ e^{i(k-p)x} = \delta(k-p). \qquad (514)$$

We say that our basis functions are normalized to a delta function. This is always the case when we are dealing with a continuous spectrum, e.g. in QM in the infinite volume.

28 Green's Functions

So far we have considered only homogeneous PDEs. How to solve, e.g. Poisson's equation

$$\nabla^2\Psi = -\frac{\rho}{\epsilon_0}? \qquad (515)$$

Let us assume that we are given a solution to

$$\nabla^2 G(\mathbf{r}) = -\delta^{(3)}(\mathbf{r}) = -\delta(x)\delta(y)\delta(z). \tag{516}$$

Then

$$\Psi(\mathbf{r}) = \frac{1}{\epsilon_0}\int d^3\mathbf{r}'\; G(\mathbf{r}-\mathbf{r}')\;\rho(\mathbf{r}') \tag{517}$$

solves Poisson's equation. G is called the **Green's function**.
Proof:

$$\begin{aligned}\nabla^2\Psi(\mathbf{r}) &= \frac{1}{\epsilon_0}\int d^3\mathbf{r}'\;\nabla^2_{\mathbf{r}}G(\mathbf{r}-\mathbf{r}')\;\rho(\mathbf{r}')\\ &= \frac{1}{\epsilon_0}\int d^3\mathbf{r}'\;[-\delta^{(3)}(\mathbf{r}-\mathbf{r}')]\;\rho(\mathbf{r}')\\ &= -\frac{\rho(\mathbf{r})}{\epsilon_0}.\end{aligned} \tag{518}$$

So far so good, but we still have to solve the equation (516) for the Green's function. This is easily done by Fourier transform:

$$\begin{aligned}\frac{1}{(2\pi)^{3/2}}\int dx\; dy\; dz\; e^{-i(k_x x+k_y y+k_z z)}\; G(\mathbf{r}) &= g(\mathbf{k})\;,\\ \frac{1}{(2\pi)^{3/2}}\int d^3\mathbf{k}\; e^{i\mathbf{k}\cdot\mathbf{r}}\; g(\mathbf{k}) &= G(\mathbf{r})\;.\end{aligned} \tag{519}$$

Fourier transforming both sides of (516) and using that the Fourier transform of $\frac{d}{dx}G(\mathbf{r})$ is $ik_x g(\mathbf{k})$ and so on we find

$$-(k_x^2+k_y^2+k_z^2)g(\mathbf{k}) = \frac{1}{(2\pi)^{3/2}}. \tag{520}$$

Here we have used that the Fourier transform of the three dimensional delta function is $(2\pi)^{-3/2}$. Hence

$$g(\mathbf{k}) = -\frac{1}{(2\pi)^{3/2}|\mathbf{k}|^2}. \tag{521}$$

Employing the inverse Fourier transform we have

$$G(\mathbf{r}) = -\frac{1}{(2\pi)^3}\int d^3\mathbf{k}\frac{1}{|\mathbf{k}|^2}e^{i\mathbf{k}\cdot\mathbf{r}}. \tag{522}$$

This integral is most conveniently done in polar coordinates ($|\mathbf{k}| = k$)

$$\begin{aligned}G(\mathbf{r}) &= -\frac{1}{4\pi^2}\int_0^\infty dk\int_0^\pi d\theta\frac{k^2\sin\theta}{k^2}e^{ikr\cos\theta}\\ &= -\frac{1}{4\pi^2}\int_0^\infty dk\;\frac{2\sin(kr)}{kr} = \frac{1}{2\pi^2 r}\int_0^\infty dz\frac{\sin(z)}{z} = -\frac{1}{4\pi r}.\end{aligned} \tag{523}$$

The last integral is done by writing $z = \lim_{\epsilon \to 0} \frac{1}{2} \left[\frac{1}{z+i\epsilon} + \frac{1}{z-i\epsilon} \right]$ and then using the residue theorem. We have found that the Green's function of the Laplacian is equal to

$$\boxed{G(\mathbf{r}) = -\frac{1}{4\pi|\mathbf{r}|}.} \tag{524}$$

This allows us to express the solution to Poisson's equation as

$$\boxed{\Psi(\mathbf{r}) = -\frac{1}{4\pi\epsilon_0} \int d^3\mathbf{r}' \, \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}.} \tag{525}$$