Contents

Chapter 1

Vector Spaces

1.1 Linear Vector Space: definition

Let $\mathbb{F}$ be a 'Field' (real numbers $\mathbb{R}$ or complex numbers $\mathbb{C}$). A set of objects V is called a 'vector space over field $\mathbb{F}$' if $\forall \left|\alpha\right\rangle, \left|\beta\right\rangle, \left|\gamma\right\rangle \in V$ (called vectors) and $\forall a, b \in \mathbb{F}$ (called scalars), the following hold (with an operation of 'addition' of vectors and 'multiplication' of a vector by a scalar defined)

1. $\left|\alpha\right\rangle + \left|\beta\right\rangle \in V$ (Closure)
2. $\left|\alpha\right\rangle + \left|\beta\right\rangle = \left|\beta\right\rangle + \left|\alpha\right\rangle$ (Addition is Commutative)
3. $\left|\alpha\right\rangle + (\left|\beta\right\rangle + \left|\gamma\right\rangle) = (\left|\alpha\right\rangle + \left|\beta\right\rangle) + \left|\gamma\right\rangle$ (Addition is Associative)
4. Existence of a *null* or 'zero' vector $\left|\right\rangle_0 \in V$: $\left|\alpha\right\rangle + \left|\right\rangle_0 = \left|\right\rangle_0 + \left|\alpha\right\rangle = \left|\alpha\right\rangle$
5. Existence of additive inverse: For each $\left|\alpha\right\rangle \in V$, $\exists\ \left|-\alpha\right\rangle \in V$: $\left|\alpha\right\rangle + \left|-\alpha\right\rangle = \left|\right\rangle_0$
6. $a\left|\alpha\right\rangle \in V$
7. $a(\left|\alpha\right\rangle + \left|\beta\right\rangle) = a\left|\alpha\right\rangle + a\left|\beta\right\rangle$
8. $(a+b)\left|\alpha\right\rangle = a\left|\alpha\right\rangle + b\left|\alpha\right\rangle$
9. $a(b\left|\alpha\right\rangle) = (ab)\left|\alpha\right\rangle$
10. $1.\left|\alpha\right\rangle = \left|\alpha\right\rangle$

Examples:

- Set of all directed arrows in a plane, with addition defined by the parallelogram law and multipliation with a real number defined as scaling the length of the arrow.
- Set of all $n \times m$ real or complex matrices.
- Set of all continuous functions of a real variable defined over some interval.
- Solutions to homogeneous linear differential equations.
- Set of n-tuples $(x_1, x_2, .., x_n)$ of real or complex numbers with addition and multiplication by a number defined intuitively. This set is called $\mathbb{R}^n$ if the numbers are real, and $\mathbb{C}^n$ if they are complex.
- Set of all 'square-integrable' complex functions of a real variable over an interval (a, b)

$$\int_a^b dx\ \ |\psi(x)|^2 < \infty \tag{1.1}$$

This set is called $L^2(a, b)$.

- Infinite set of complex numbers $\{x_k\}$; $k = 1, 2, 3, ...$ with the condition

$$\sum_{k=1}^{\infty} |x_k|^2 < \infty \tag{1.2}$$

This set is called l^2.

Exercise 1.1.1. Do the following form a vector space?
(a) Set of all directed arrows in three-dimensional space, such that the z-component is positive.
(b) Functions $f(x)$ over interval $x \in (0, L)$ such that $f(0) = f(L)$.
(c) Functions $f(x)$ over interval $x \in (0, L)$ such that $f(0) = 2$.
(d) Set of all 2×2 invertible real matrices.
(e) Set of all polynomials with degree less than or equal to n.
(f) Set of all polynomials with degree greater than or equal to n.

Theorem 1.1.1. *The null vector in a vector space is unique.*

Proof Let $|\rangle_0$ and $|\rangle_0'$ be two null vectors. By definition, $|\rangle_0 + |\rangle_0' = |\rangle_0$ (visualizing $|\rangle_0'$ as the null vector). Similarly, $|\rangle_0 + |\rangle_0' = |\rangle_0'$ (visualizing $|\rangle_0$ as the null vector). Then, $|\rangle_0 = |\rangle_0'$.

Theorem 1.1.2. $0.\,|\alpha\rangle = |\rangle_0 \;\; \forall \;\; |\alpha\rangle \in V$.

Proof

$$\begin{aligned}
|\rangle_0 &= |\alpha\rangle + |-\alpha\rangle \\
&= (1+0)\,|\alpha\rangle + |-\alpha\rangle \\
&= (|\alpha\rangle + 0.\,|\alpha\rangle) + |-\alpha\rangle \\
&= (|\alpha\rangle + |-\alpha\rangle) + 0.\,|\alpha\rangle \\
&= |\rangle_0 + 0.\,|\alpha\rangle \\
&= 0.\,|\alpha\rangle
\end{aligned}$$

Theorem 1.1.3. $(-1)\,|\alpha\rangle = |-\alpha\rangle \;\; \forall \;\; |\alpha\rangle \in V$

Proof $|\alpha\rangle + (-1)\,|\alpha\rangle = (1-1)\,|\alpha\rangle = 0.\,|\alpha\rangle = |\rangle_0$.

Exercise 1.1.2. Prove that the additive inverse of a vector is unique.

Exercise 1.1.3. Prove the following:
(a)$c\,|\rangle_0 = |\rangle_0 \;\; \forall \; c \in \mathbb{F}$
(b) $c\,|\alpha\rangle = |\rangle_0$ iff $c = 0$ or $|\alpha\rangle = |\rangle_0$

1.2 Linear independence

Definition A set of vectors $|\alpha_1\rangle, |\alpha_2\rangle, .., |\alpha_n\rangle$ is said to be *linearly independent* (LI) if an equation of the form

$$c_1\,|\alpha_1\rangle + c_2\,|\alpha_2\rangle + .. + c_n\,|\alpha_n\rangle = |\rangle_0$$

has a unique solution $c_1 = c_2 = .. = c_n = 0$. An infinite set of vectors is said to be linearly independent if any finite subset of the set of vectors is linearly independent.

Definition If a set of vectors is not linearly independent, it is said to be *linearly dependent*.

If a set of vectors is linearly dependent then at least one of them can be expressed as a linear combintion of the others. Say, the set $|\alpha_1\rangle, |\alpha_2\rangle, .., |\alpha_n\rangle$ is linearly dependent. Consider the equation

$$c_1 |\alpha_1\rangle + c_2 |\alpha_2\rangle + .. + c_n |\alpha_n\rangle = |\rangle_0 \tag{1.3}$$

Since the set is not linearly independent, at least two of the coefficients are non-zero (it is not possible for just one coefficient to be zero. Why??). Say, $c_1 \neq 0$. Then, it follows from (1.3)that

$$|\alpha_1\rangle = -\frac{c_2}{c_1} |\alpha_2\rangle - \frac{c_3}{c_1} |\alpha_3\rangle .. - \frac{c_n}{c_1} |\alpha_n\rangle \tag{1.4}$$

Example Consider the space $L^2(a, b)$. We need to determine if a set of functions $f_1(x), f_2(x), .., f_n(x)$ is linearly independent. Consider the equation

$$c_1 f_1(x) + c_2 f_2(x) + .. + c_n f_n(x) = 0 \tag{1.5}$$

The set is LI, if this equation has the unique solution $c_i = 0 \; \forall i$. If we differentiate this equation $n-1$ times, we get a system of n equations

$$\begin{aligned}
c_1 f_1(x) + c_2 f_2(x) + .. + c_n f_n(x) &= 0 \\
c_1 f_1^{(1)}(x) + c_2 f_2^{(1)}(x) + .. + c_n f_n^{(1)}(x) &= 0 \\
c_1 f_1^{(2)}(x) + c_2 f_2^{(2)}(x) + .. + c_n f_n^{(2)}(x) &= 0 \\
\dots\dots\dots\dots\dots\dots &= .. \\
c_1 f_1^{(n-1)}(x) + c_2 f_2^{(n-1)}(x) + .. + c_n f_n^{(n-1)}(x) &= 0
\end{aligned}$$

We can write these as a single matrix equation

$$\begin{pmatrix} f_1(x) & f_2(x) & . & f_n(x) \\ f_1^{(1)}(x) & f_2^{(1)}(x) & . & f_n^{(1)}(x) \\ . & . & . & . \\ f_1^{(n-1)}(x) & f_2^{(n-1)}(x) & . & f_n^{(n-1)}(x) \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ . \\ c_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ . \\ 0 \end{pmatrix} \tag{1.6}$$

If there is even one point in the interval $[a, b]$ where the determinant of the matrix consisting of the functions and their derivatives is non-zero, the matrix will be invertible at that point, implying that $c_i = 0 \; \forall i$. Then, a sufficient condition for a set of functions to be LI is that their *Wronskian* (detrminant of the matrix) is non-vanishing at one or more points in the interval

$$\begin{vmatrix} f_1(x) & f_2(x) & . & f_n(x) \\ f_1^{(1)}(x) & f_2^{(1)}(x) & . & f_n^{(1)}(x) \\ . & . & . & . \\ f_1^{(n-1)}(x) & f_2^{(n-1)}(x) & . & f_n^{(n-1)}(x) \end{vmatrix} \neq 0 \quad \text{at one or more } x \in [a, b] \tag{1.7}$$

Exercise 1.2.1. Consider the set of matrices

$$\begin{aligned}
A &= \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \\
B &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \\
C &= \begin{pmatrix} -2 & -1 \\ 0 & -2 \end{pmatrix}
\end{aligned} \tag{1.8}$$

Is this set linearly independent?

Exercise 1.2.2. Let $|\alpha\rangle, |\beta\rangle$ and $|\beta\rangle$ be LI. Are the following sets LI?
(a) $|\alpha\rangle + |\beta\rangle$, $|\beta\rangle + |\gamma\rangle$, $|\alpha\rangle + |\gamma\rangle$
(b) $|\alpha\rangle$, $|\alpha\rangle + |\beta\rangle$, $|\alpha\rangle + |\beta\rangle + |\gamma\rangle$
(c) $|\alpha\rangle - |\beta\rangle$, $|\beta\rangle - |\gamma\rangle$, $|\gamma\rangle - |\alpha\rangle$

Exercise 1.2.3. Is the following set of 3-tuples LI?
$\alpha = (1, -2, 3)$, $\beta = (5, -2, 3)$, $\gamma = (7, 2, -3)$

Exercise 1.2.4. Show that any set of vectors that contains the null vector is linearly dependent.

Exercise 1.2.5. Consider the vector space of continuous functions. Show that the set of elements $\{1, x, x^2, x^3, .., x^n\}$ is LI for arbitrary n.

Exercise 1.2.6. Show that the set $\{1, \cos(2\pi kx), \sin(2\pi kx)\}$ for $k = 1, 2, .., n$ is LI for arbitrary positive integer n.

Exercise 1.2.7. Determine which of the set of functions are LI. If not, express one as a linear combination of the others.
(a) $\sinh x, \ \cosh x$
(b) $e^x, \ \sinh x, \ \cosh x$
(c) $1, \ \sin x, \ \cos 2x$
(d) $x, \ \log x$ over the interval $[1, 2]$

1.3 Linear Subspace

Definition Let U be a non-empty subset of a vector space V over a field $\mathbb{F}$. U is a *linear subspace* of V if the following are true

1. If $|\alpha\rangle, |\beta\rangle \in U$ then $|\alpha\rangle + |\beta\rangle \in U$
2. If $|\alpha\rangle \in U$ and $c \in \mathbb{F}$ then $c\,|\alpha\rangle \in U$

Exercise 1.3.1. Show that a linear subspace is a vector space in its own right.

Theorem 1.3.1. *Let U be a subset of a vector space V over a field $\mathbb{F}$. U is a subspace of V iff the following conditions hold:*
(a) $|\rangle_0 \in U$
(b) If $|\alpha\rangle, |\beta\rangle \in U$ *and* $k, l \in \mathbb{F}$ *then* $k\,|\alpha\rangle + l\,|\beta\rangle \in U$

Proof Trivial to prove.

Examples:

- Subset $(x_1, x_2, x_3, .., x_n)$ of $\mathbb{R}^n$ such that $x_2 = x_3 = ... = x_n = 0$.
- Set of all C^k (k^{th} derivative exists and is continuous) functions of a real variable defined over an interval is a subspace of the set of all continuous functions defined over the interval.
- Set of all $n \times n$ symmetric matrices is a subspace of the set of all $n \times n$ matrices.

Example Consider the vector space $\mathbb{R}^n$. Now, consider only those elements of $\mathbb{R}^n$ which are solutions to the given set of homogeneous equations

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + .. + a_{1n}x_n &= 0 \\
a_{21}x_1 + a_{22}x_2 + .. + a_{2n}x_n &= 0 \\
&.......... \\
a_{m1}x_1 + a_{m2}x_2 + .. + a_{mn}x_n &= 0
\end{aligned}$$

This set of conditions can be written as a matrix equation

$$AX = 0$$

where A is an $m \times n$ matrix with elements a_{ij} and $X = (x_1\ x_2\ .\ .\ x_n)^T$. If X_1 and X_2 are solutions to this equation, so are $X_1 + X_2$ and cX_1. This set therefore is a subspace.

Example Consider the solutions to the differential equation

$$\frac{d^2\psi}{dx^2} = -\omega^2\psi$$

These solutions form a subspace of the vector space of all continuous functions.

Theorem 1.3.2. *Let V be a vector space over a field $\mathbb{F}$. Let $|\alpha_1\rangle, |\alpha_2\rangle, ..|\alpha_r\rangle \in V$. Then, the set of all linear combinations of $|\alpha_1\rangle, |\alpha_2\rangle, ..|\alpha_r\rangle$ is a subspace of V.*

Proof Trivial.

Definition The subspace consisting of all linear combinations of a set of vectors $A = |\alpha_1\rangle, |\alpha_2\rangle, .., |\alpha_r\rangle$, denoted as $[A]$, is called the *span* of these vectors.

Theorem 1.3.3. *A set of vectors is linearly independent when there is only one way to express any element of their span as a linear combination of these vectors.*

Proof Let this be true. Then, if

$$c_1|\alpha_1\rangle + c_2|\alpha_2\rangle + ..c_r|\alpha_r\rangle = c_1'|\alpha_1\rangle + c_2'|\alpha_2\rangle + ..c_r'|\alpha_r\rangle$$

it implies that $c_1 = c_1', c_2 = c_2', .., c_r = c_r'$. Now , consider the eqquation

$$c_1|\alpha_1\rangle + c_2|\alpha_2\rangle + ..c_r|\alpha_r\rangle = |\rangle_0$$

This can be rewritten as

$$c_1|\alpha_1\rangle + c_2|\alpha_2\rangle + ..c_r|\alpha_r\rangle = 0.|\alpha_1\rangle + 0.|\alpha_2\rangle + ..0.|\alpha_r\rangle$$

which gives $c_1 = 0, c_2 = 0, .., c_r = 0$. That is, the vectors are linearly independent.

Definition Let S and J be subspaces of a vector space V. The *sum* of S and J, denoted as $S + J$, is the set of vectors of the form $|\alpha\rangle + |\beta\rangle$, where $|\alpha\rangle \in S$ and $|\beta\rangle \in J$. The *intersection* of S and J, denoted as $S \cap J$, is the set of vectors common to both S and J.

Theorem 1.3.4. *If S and J are subspaces of V then $S + J$ and $S \cap J$ are subspaces of V.*

Proof Let $|\psi\rangle, |\phi\rangle \in S + J$. Then, $\exists\ |\alpha_1\rangle, |\alpha_2\rangle \in S$ and $\exists\ |\beta_1\rangle, |\beta_2\rangle \in J$, such that

$$\begin{aligned} |\psi\rangle &= |\alpha_1\rangle + |\beta_1\rangle \\ |\phi\rangle &= |\alpha_2\rangle + |\beta_2\rangle \end{aligned}$$

Then,

$$\begin{aligned} |\psi\rangle + |\phi\rangle &= (|\alpha_1\rangle + |\beta_1\rangle) + (|\alpha_2\rangle + |\beta_2\rangle) \\ &= (|\alpha_1\rangle + |\alpha_2\rangle) + (|\beta_1\rangle + |\beta_2\rangle) \end{aligned}$$

But, $|\alpha_1\rangle + |\alpha_2\rangle \in S$ and $|\beta_1\rangle + |\beta_2\rangle \in J$, since S and J are subspaces. Therefore, the vector $|\psi\rangle + |\phi\rangle \in S + J$. Similarly, easy to show that $c|\psi\rangle \in S + J$. Next, let $|\psi\rangle, |\phi\rangle \in S \cap J$. Then, $|\psi\rangle, |\phi\rangle \in S$ and $|\psi\rangle, |\phi\rangle \in J$. Therefore, $|\psi\rangle + |\phi\rangle \in S$ and $|\psi\rangle + |\phi\rangle \in J$. Therefore, $|\psi\rangle + |\phi\rangle \in S \cap J$. Similarly, $c|\psi\rangle \in S$ and $c|\psi\rangle \in J$. Therefore, $c|\psi\rangle \in S \cap J$.

Definition Let M and N be subspaces of a vector space V. Then, V is said to be a *direct sum* of M and N, symbolically written as $M \oplus N$, iff $V = M + N$ and $M \cap N = \{|\rangle_0\}$.

Theorem 1.3.5. *$V = M \oplus N$ iff every vector $|\psi\rangle \in V$ has a unique representation $|\psi\rangle = |\alpha\rangle + |\beta\rangle$ where $|\alpha\rangle \in M$ and $|\beta\rangle \in N$.*

Proof Let $V = M \oplus N$. Then, in particular, $V = M + N$. Given $|\psi\rangle$ *in* V, $\exists\ |\alpha\rangle \in M$ and $|\beta\rangle \in N$ such that $|\psi\rangle = |\alpha\rangle + |\beta\rangle$. If this combination is not unique, $\exists\ |\alpha'\rangle \in M$ and $|\beta'\rangle \in N$ such that $|\psi\rangle = |\alpha'\rangle + |\beta'\rangle$. Then, $|\alpha\rangle + |\beta\rangle = |\alpha'\rangle + |\beta'\rangle$, which implies $|\alpha\rangle - |\alpha'\rangle = |\beta'\rangle - |\beta\rangle$. But, $|\alpha\rangle - |\alpha'\rangle \in M$ and $|\beta'\rangle - |\beta\rangle \in N$. Since $M \cap N = \{|\rangle_0\}$, this implies that $|\alpha\rangle - |\alpha'\rangle = |\rangle_0 \Rightarrow |\alpha'\rangle = |\alpha\rangle$. Similarly, $|\beta'\rangle = |\beta\rangle$. Therefore, $|\psi\rangle = |\alpha\rangle + |\beta\rangle$ is unique. Conversely, $|\psi\rangle \in V$ be a unique combination $|\psi\rangle = |\alpha\rangle + |\beta\rangle$ with $|\alpha\rangle \in M$ and $|\beta\rangle \in N$. Clearly, this implies that $V = M + N$. Let $M \cap N$ have a non-trivial vector $|\alpha\rangle \neq |\rangle_0$. Then, it is true that $|\alpha\rangle = |\alpha\rangle + |\rangle_0$ where we can visualise $|\alpha\rangle \in M$ and $|\rangle_0 \in N$. But, we can also write $|\alpha\rangle = |\rangle_0 + |\alpha\rangle$ where we can visualise $|\alpha\rangle \in N$ and $|\rangle_0 \in M$. Since the combination is unique, it follows that $|\alpha\rangle = |\rangle_0$. Therefore, $M \cap N = \{|\rangle_0\}$, and therefore $V = M \oplus N$.

Theorem 1.3.6. *Let S be a linearly independent subset of a vector space V, and let $J = [S]$ (span of S). Then for any vector $|\psi\rangle \in V, S \cup \{|\psi\rangle\}$ is LI iff $|\psi\rangle \notin J$.*

Proof Let $|\psi\rangle \in J$. Then,

$$|\psi\rangle = c_1 |\alpha_1\rangle + c_2 |\alpha_2\rangle + .. + c_k |\alpha_k\rangle$$

for $|\alpha_i\rangle \in S$. If $c_i = 0\ \forall\ i$ then $|\psi\rangle = |\rangle_0$, and then $S \cup\ \{|\rangle_0\}$ is linearly dependent (any set containing the null vector is linearly dependent). If any one of a_i is non-zero, then clearly

$$c_1 |\alpha_1\rangle + c_2 |\alpha_2\rangle + .. + c_k |\alpha_k\rangle + (-1) |\psi\rangle = |\rangle_0$$

with not all coefficients vanishing. Then, the set is again linearly dependent. The condition then is necessary. Conversely, let $|\psi\rangle \notin J$. Consider the equation

$$c_1 |\alpha_1\rangle + c_2 |\alpha_2\rangle + .. + c_k |\alpha_k\rangle + c_{k+1} |\psi\rangle = |\rangle_0$$

where $|\alpha_i\rangle \in S$. If $c_{k+1} \neq 0$ then $|\psi\rangle \in J$ which contradicts out assumption. Then, $c_{k+1} = 0$ and the equation reduces to

$$c_1 |\alpha_1\rangle + c_2 |\alpha_2\rangle + .. + c_k |\alpha_k\rangle = |\rangle_0$$

Since $|\alpha_i\rangle$ are LI, all the coefficients are zero. Then, $c_1 = c_2 = ..c_k = c_{k+1}$, so that the set $S \cup \{|\psi\rangle\}$ is LI.

This theorem tells us how to make a set of LI vectors larger. Given a set of LI vectors, if we can add to it a vector which does not belong to the span of the set, the resulting (larger) set will also be LI.

Definition A subset of a vector space V is said to be a *maximally LI subset* of V if it is not possible to add more vectors to this set and still keep it LI.

Theorem 1.3.7. *Let $S = \{|\alpha_1\rangle, |\alpha_2\rangle, .., |\alpha_k\rangle\}$ be a finite set of non-zero vectors. Then, S is dependent iff*

$$|\alpha_m\rangle \in [|\alpha_1\rangle, ..., |\alpha_{m-1}\rangle]$$

for some $m \leq k$.

Proof If $|\alpha_m\rangle \in [|\alpha_1\rangle, ..., |\alpha_{m-1}\rangle]$ then the set $\{|\alpha_1\rangle, .., |\alpha_m\rangle\}$ is dependent. Since this is a subset of S, S is dependent. Conversely, let S be a dependent set. Let m be the least integer such that $\{|\alpha_1\rangle, |\alpha_2\rangle, .., |\alpha_m\rangle\}$ is dependent. Then, for $c_1, c_2, .., c_m$ not all zero,

$$c_1 |\alpha_1\rangle + .. + |\alpha_m\rangle = |\rangle_0$$

But, $c_m \neq 0$ (by definition of m). Therefore,

$$|\alpha_m\rangle = -\frac{1}{c_m} (c_1 |\alpha_1\rangle + c_2 |\alpha_2\rangle + .. + c_m |\alpha_{m-1}\rangle)$$

which implies that $|\alpha_m\rangle \in [|\alpha_1\rangle, ..., |\alpha_{m-1}\rangle]$.

Theorem 1.3.8. *If a subspace J is spanned by a set of vectors $S = \{|\alpha_1\rangle, |\alpha_2\rangle, .., |\alpha_k\rangle\}$, there exists a LI subset of S which also spans J.*

Proof If S is a LI set, then S itself is the subset. If S is dependent, then from the previous theorem, there exists a least integer m such that

$$|\alpha_m\rangle \in [|\alpha_1\rangle, ..., |\alpha_{m-1}\rangle]$$

Now, let $S_1 = S - \{|\alpha_m\rangle\}$. Then, $J = [S] = [S_1]$. Buy repeating the argument on S_1, we will get a set S_2, and so on, till we find a set S_p which is LI.

In the previous theorem, we have constructed a *maximal LI subset* of the set S which spans J, and which cannot be extended.

Definition Given a subspace S of a vector space V, a subspace W is said to be a *complement* of S in V if $V = U \oplus S$.

The complement of a subspace in a vector space is not unique.

Exercise 1.3.2. Consider three dimensional Euclidean space $\mathbb{R}^3$. Let $|\alpha\rangle = (2, 1, 0), |\beta\rangle = (1, 0, -1), |\gamma\rangle = (-1, 1, 1)$ and $|\delta\rangle = (0, -1, 1)$. Let $S = [|\alpha\rangle, |\beta\rangle]$ and $J = [|\gamma\rangle, |\delta\rangle]$. Interpret $S+J$ and $S \cap J$ geometrically.

Exercise 1.3.3. Let $|\alpha\rangle = (2, 1, 0), |\beta\rangle = (1, 0, -1), |\gamma\rangle = (-1, 1, 1)$ and $|\delta\rangle = (0, -1, 1)$. Find all the maximally LI subsets of this set of vectors.

Exercise 1.3.4. Show that $\{|\alpha\rangle = (0, 1, 2, 3), |\beta\rangle = (1, 2, 3, 4)\}$ and $\{|\gamma\rangle = (2, 5, 8, 11), |\delta\rangle = (3, 5, 7, 9)\}$ span the same subspace of $\mathbb{R}^4$.

Exercise 1.3.5. Consider the vector space $\mathbb{R}^2$. Construct an example to illustrate that the complement of a subspace of $\mathbb{R}^2$ is not unique.

1.4 Basis

Definition A maximal LI subset of a vector space V is called a *basis* of V.

Clearly, a basis spans the vector space.

Theorem 1.4.1. *Every vector of a vector space has a unique representation as a linear combination of the vectors of a basis of the vector space.*

Proof Since the basis spans the vector space, any vector can be written as a linear combination of basis vectors. All we need to show is that this combination is unique. Let the basis consist of vectors $|\alpha_1\rangle, |\alpha_2\rangle, .., |\alpha_n\rangle$. Let a vector $|\psi\rangle$ have two possible linear representations in this basis

$$|\psi\rangle = c_1 |\alpha_1\rangle + c_2 |\alpha_2\rangle + .. + c_n |\alpha_n\rangle$$
$$|\psi\rangle = c_1' |\alpha_1\rangle + c_2' |\alpha_2\rangle + .. + c_n' |\alpha_n\rangle$$

Subtracting the two equations gives

$$(c_1 - c_1') |\alpha_1\rangle + (c_2 - c_2') |\alpha_2\rangle + .. + (c_n - c_n') |\alpha_n\rangle = |\rangle_0 \tag{1.9}$$

Since, by definition the set is LI, it follows that $c_1 - c_1' = 0, c_2 - c_2' = 0, .., c_n - c_n' = 0$. Then, the linear representation is unique.

Theorem 1.4.2. *Every basis of a vector space has the same number of elements.*

Proof Let $A = \{|\alpha_1\rangle, |\alpha_2\rangle, .., |\alpha_m\rangle\}$ and $B = \{|\beta_1\rangle, |\beta_2\rangle, .., |\beta_n\rangle\}$ be two bases. By definition, the set B is maximally LI. Then, the set $B_1 = \{|\alpha_1\rangle, |\beta_1\rangle, |\beta_2\rangle, .., |\beta_n\rangle\}$ is LD. Then, $\exists\ i \leq n$ such that $|\beta_i\rangle\ in\ [|\alpha_1\rangle, |\beta_1\rangle, |\beta_2\rangle, .., |\beta_{i-1}\rangle]$. Then, there exists $B_1' \subseteq B_1$ which contains $|\alpha_1\rangle$ as the first vector and this subset forms a basis of the vector space. Let $B_2 = \{|\alpha_2, B_1'\rangle$. This set will be LD. Then, there exists a vector in B_2 other than $|\alpha_1\rangle$ and $|\alpha_2\rangle$ (why?) that will be a linear combination of its preceding vectors. Then, there exists $B_2' \subseteq B_2$ containing both $|\alpha_1\rangle$ and $|\alpha_2\rangle$ which is a basis of the vector space. This way, we can remove all the $|\beta_i\rangle$. If all the $|\beta_i\rangle$ are removed before m steps, we will end up with the set $\{|\alpha_1\rangle, |\alpha_2\rangle, .., |\alpha_j\rangle\}$ for $j < m$, and this should be a basis. But, this is not possible, since the larger set $\{|\alpha_1\rangle, |\alpha_2\rangle, .., |\alpha_m\rangle\}$ is a basis. Then, $m \leq n$. We can now interchange the role of A and B, and we will get $n \leq m$. The two conditions imply that $n = m$.

Definition The number of elements in a basis of a vector space is called the *dimension* of the vector space.

Theorem 1.4.3. *Any linearly independent set of vectors in an n-dimensional vector space can be extended to a basis.*

Proof Let $A_k = \{|\alpha_1\rangle, |\alpha_2\rangle, .., |\alpha_k\rangle\}$ be a linearly independent set, and let $\{|\beta_1\rangle, |\beta_2\rangle, .., |\beta_n\rangle\}$ be a basis. Let $J_k = [|\alpha_1\rangle, |\alpha_2\rangle, .., |\alpha_k\rangle]$. If $|\beta_i\rangle \in J_k\ \forall\ k$ then the set A_k already forms a basis. If not, $\exists\ j \leq n$ such that $|\beta_j\rangle \notin J_k$. Then, $\{|\alpha_1\rangle, |\alpha_2\rangle, .., |\alpha_k\rangle, |\beta_j\rangle\}$ is a LI set, larger than the set A_k. Now, consider $J_{k+1} = [|\alpha_1\rangle, |\alpha_2\rangle, .., |\alpha_k\rangle, |\beta_j\rangle]$, and repeat the argument, till the set forms a basis.

Example Consider $\mathbb{R}^n$. Clearly, the set $\alpha_1 = (1, 0, .., 0), \alpha_2 = (0, 1, .., 0), .., \alpha_n = (0, 0, .., 1)$ is LI. It also spans the space, since

$$(x_1, x_2, ..x_n) = x_1\alpha_1 + x_2\alpha_2 + .. + x_n\alpha_n$$

Then, this set forms a basis of $\mathbb{R}^n$ and therefore its dimension is n. This basis is called the *standard basis* of $\mathbb{R}^n$.

Example Consider the vector space of all polynomials of degree n. The set $\{1, x, x^2, .., x^n\}$ is a basis for this space.

Example Consider the vector space $L^2(0, 1)$. The infinite set $\{1, \cos(2\pi kx), \sin(2\pi kx)\}$ for $k = 1, 2, .., ..$ is LI (since any finite subset is LI). From Fourier's Theorem, any $f(x) \in L^2(0, 1)$ can be expanded as a series of these functions. Then, this set spans $L^2(0, 1)$. The vector space $L^2(0, 1)$ is then *infinite dimensional*.

Exercise 1.4.1. What is the dimension of the following vector spaces:
(a) $\mathbb{R}$
(b) $\mathbb{C}$ over field $\mathbb{R}$
(c) $\mathbb{C}$ over field $\mathbb{C}$
(d) $\mathbb{R}^n$ over field $\mathbb{R}$
(e) $\mathbb{C}^n$ over field $\mathbb{R}$
(f) $\mathbb{C}^n$ over field $\mathbb{C}$
(g) Space of 2×2 complex matrices over $\mathbb{R}$
(h) Space of 2×2 complex matrices over $\mathbb{C}$?

Exercise 1.4.2. Show that the vectors $\alpha = (1, 1, 1, 1), \beta = (0, 1, 1, 1), \gamma = (0, 0, 1, 1)$ and $\delta = (0, 0, 0, 1)$ form a basis of $\mathbb{R}^4$. Express the standard basis vactors as linear cobinations of these vectors.

Exercise 1.4.3. Consider the vector space of solutions to the differential equation

$$\frac{d^2\psi}{dx^2} = -\omega^2\psi$$

What is the dimension of this vector space? Find at least two bases.

Exercise 1.4.4. What is the dimension of the vector space spanned by the vectors $|\alpha\rangle = (1, 2, 2, 1), |\beta\rangle = (3, 4, 4, 3)$ and $|\gamma\rangle = (1, 0, 0, 1)$? Find a basis for this space.

Exercise 1.4.5. Consider $\mathbb{R}^4$ with vectors of the form (x_1, x_2, x_3, x_4). Let S be a subspace of $\mathbb{R}^4$ defined by the constraints

$$\begin{aligned} 2x_1 - 3x_2 - x_3 + x_4 &= 0 \\ x_1 + x_2 + 2x_3 - x_4 &= 0 \end{aligned}$$

What is the dimension of this subspace? Find a basis.

1.5 Tangent Vectors

In this section, we make the notion of vectors as 'directed arrows' in space more precise. Consider a plane, with points labelled by Cartesinan coordinates (x, y). In this plane, consider a curve, parametrized by some parameter t (for instance, this could be the trajectory of a particle with t being time). Let $\vec{T}$ be the tangent vector to the curve at some point P

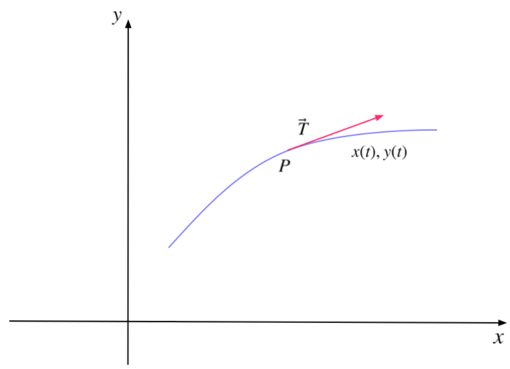

Figure 1.1: Tangent to a curve in a plane.

This tangent vector is an element of $\mathbb{R}^2$. To see this, we take two points on the curve at parameter t and $t+\Delta t$, with 'position vectors' $\vec{r}(t)$ and $\vec{r}(t+\Delta t)$. these position vectors can be visualised as elements of $\mathbb{R}^2$, such that $\vec{r}(t) = (x(t), t(t))$ and $\vec{r}(t+\Delta t) = (x(t+\Delta t), y(t+\Delta t))$. Then, their difference $\Delta\vec{r} = \vec{r}(t+\Delta t) - \vec{r}(t)$ is a vector belonging to $\mathbb{R}^2$. The tangent vector $\vec{T}$ is just

$$\vec{T} = \lim_{\Delta t \to \infty} \left(\frac{1}{\Delta t} \right) \Delta\vec{r} \tag{1.10}$$

which is an element of $\mathbb{R}^2$, being

$$\vec{T} = \left(\frac{dx}{dt}, \frac{dy}{dt} \right) \tag{1.11}$$

It should be noted that we can construct a tangent vector at the same point, parallel to $\vec{T}$, but of different length, by parametrizing the curve with a different parameter. Say, we parametrize the curve by a parameter t', such that $t' = t/\lambda$, where λ is a real number. Then, calculating the tangent vector as before, we will end up with a vector

$$\begin{aligned} \vec{T}' &= \lim_{\Delta t' \to \infty} \left(\frac{1}{\Delta t} \right) \Delta\vec{r} \\ &= \left(\frac{dx}{dt'}, \frac{dy}{dt'} \right) \\ &= \lambda \left(\frac{dx}{dt}, \frac{dy}{dt} \right) \\ &= \lambda\, \vec{T} \end{aligned}$$

Therefore, by parametrizing a curve differently, we can generate tangent vectors of all possible lengths at a point. Conversely, given a tangent vector of a given length at some point on a curve, there exists a parameter t, such that the vector is generated using the prescription (1.10). Further, given a directed arrow $\vec{T}$ in a plane at some point, we can always visualize it as being tangent to some curve $x(t), y(t)$, such that (1.11) is satisfied.

Now, we are ready to make a one to one correspondence between directed arrows in space and *directional derivatives*. The idea is briefly as follows. Consider some function $f(x, y)$ defined over the plane. In general, the function will vary from point to pint, and its variation along different directions will be different. For instance, the function $f(x)$ which is independent of y will only vary if we move parallel to the x-axis. There is no variation along the y-direction. Given a function, we show that there is a one to one correspondence between a directed arrow in the plae at some point, and the variation in the function along that direction. Given a directed arrow $\vec{T}$, there exists a curve parametrized by a parameter t such that (1.11) is satisfied. Let us ask the following question: how rapidly does the function vary along this curve *at this point*? Clearly, this is given by the derivative of the function with respect to the parameter t at this point. Using the chain rule, we know that

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} \tag{1.12}$$

which we rewrite as

$$\frac{df}{dt} = T^x \frac{\partial f}{\partial x} + T^y \frac{\partial f}{\partial y} \tag{1.13}$$

where T^x and T^y are the ‘components’ of the vector $\vec{T}$ in the (x, y) coordinate system. This equation is equivalent to the more familiar equation

$$\vec{T} = T^x \vec{i} + T^y \vec{j} \tag{1.14}$$

where $\vec{i}$ and $\vec{j}$ are directed arrows along the x and y directions respectively. We can write (1.13) more abstractly as

$$\frac{d}{dt} = T^x \frac{\partial}{\partial x} + T^y \frac{\partial}{\partial y} \tag{1.15}$$

where the ‘derivative operator’ is always assumed to act on some function. Then, there is a one-to one correspondence between such ‘directional derivatives’ and arrows attached to a point. It is also clear that the set of all directional derivatives *at a point* form a vector space. For instance, let us verify the closure property. Given two curves intersecting at a point parametrized by parameters t_1 and $t2$, the rate of change of a function f at this point along these curves will be df/dt_1 and df/dt_2 respectively. Question: does there exist a curve passing through this point, parametrized with some parameter t, such that

$$\frac{df}{dt} = \frac{df}{dt_1} + \frac{df}{dt_2} \tag{1.16}$$

It is easy to see that such a curve should exist. Using (1.13), we get

$$\begin{aligned}
\frac{df}{dt_1} + \frac{df}{dt_2} &= (T_1^x \frac{\partial f}{\partial x} + T_1^y \frac{\partial f}{\partial y}) + (T_2^x \frac{\partial f}{\partial x} + T_2^y \frac{\partial f}{\partial y}) \\
&= (T_1^x + T_2^x)\frac{\partial f}{\partial x} + (T_1^y + T_2^y)\frac{\partial f}{\partial y})
\end{aligned}$$

which is of the form (1.13). So, there clearly exists some curve parametrized by some parameter t, such that (1.16) is satisfied. It is geometrically easy to visualize this pictorially

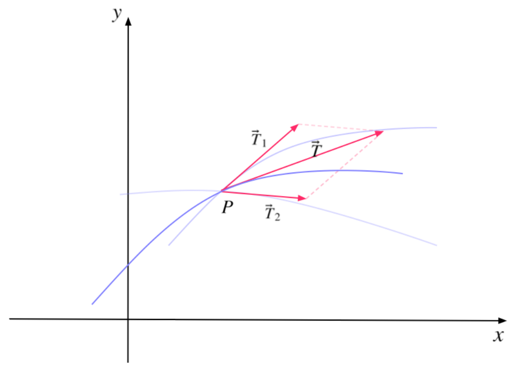

Figure 1.2: Closure of tangent vectors.

What is the geometrical interpretation of the partial derivatives with respect to x and y? They are just directional derivatives along the x and y directions respectively. To see this, consider the curve $x(t) = t, y(t) = c$ where c is a constant. This is just a corve parallel to the x-axis. The derivative of a function along this curve is

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} \tag{1.17}$$

$$= \frac{\partial f}{\partial x} \tag{1.18}$$

since $dx/dt = 1$ and $dy/dt = 0$. This completes our identification of (1.15) with (1.14). Note that this identification is strictly *local*. The derivative of a function depends on the point where it is taken. Therefore, this correspondence between directed arrows and derivatives is there at a single point. In other words, at each point, we have a different vector space.

Let us now label points on this plane using a different coordinate system, say a polar coordinate system

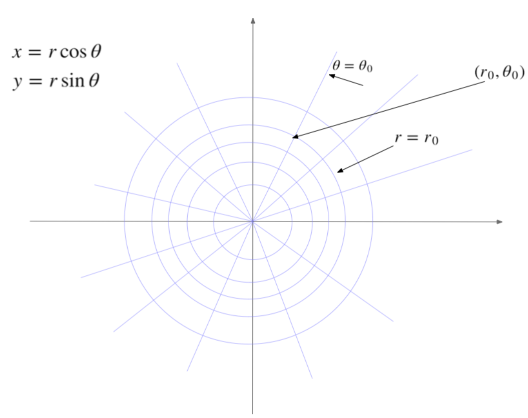

Figure 1.3: Polar coordinates in a plane.

Given a curve parametrized by a parameter t, we can describe it using polar cordinates, as $r(t), \theta(t)$. Then, the derivative of a function along the curve can be written as

$$\frac{df}{dt} = \frac{\partial f}{\partial r}\frac{dr}{dt} + \frac{\partial f}{\partial \theta}\frac{d\theta}{dt} \tag{1.19}$$

We can then abstractly write the derivative operator as

$$\frac{d}{dt} = T^r \frac{\partial}{\partial r} + T^\theta \frac{\partial}{\partial \theta} \tag{1.20}$$

We can interpret the partial derivatives with respect to r and θ as tangent vectors to the curves $\theta =$ constant and $r =$ constant trespectively. To see this, consider the curve $r(r) = t$, $\theta = c$ where c is constant. The directional derivative of a function along this curve will clearly be

$$\frac{df}{dt} = \frac{\partial f}{\partial r}$$

Then, $\partial/\partial r$ is tangent to the $\theta =$ constant curve. Similarly, $\partial/\partial\theta$ is tangent to the $r =$ constant curve. Then, equation (1.20) identifies T^r and T^θ as components of the tangent vector d/dt along tangent vectors $\partial/\partial r$ and $\partial/\partial\theta$

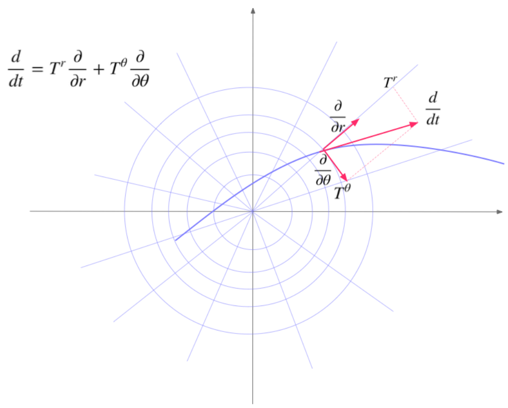

Figure 1.4: Tangent vectors along polar coordinate curves.

In general, we can use a completely arbitrary coordinate system. Consider two functions $u = u(x, y)$ and $v = v(x, y)$. For suitable choice of u and v, the curves $u =$ constant and $v =$ constant will fill the plane, and any point can be assigned coordinates (u, v). A curve in this plane can be described by $u = u(t), v = v(t)$. Then, the derivative of a function along this curve will be

$$\frac{df}{dt} = \frac{\partial f}{\partial u}\frac{du}{dt} + \frac{\partial f}{\partial v}\frac{dv}{dt} \tag{1.21}$$

As before, we can write this abstractly as

$$\frac{d}{dt} = T^u \frac{\partial}{\partial u} + T^v \frac{\partial}{\partial v} \tag{1.22}$$

where $\partial/\partial u$ and $\partial/\partial v$ are tangent vectors to curves $v =$ constant and $u =$ constant respectively, and $T^u = du/dt$, $T^v = dv/dt$ are components of the tangent vector d/dt along the 'basis' tangent vectors $\partial/\partial u$ and $\partial/\partial v$ respectively. These tangent vectors are referred to as 'coordinate-basis vectors'. It is useful to introduce the notation in which we write coordinate basis vectors as $\partial/\partial u = \underset{\sim}{e}_u, \partial/\partial v = \underset{\sim}{e}_v$, etc. An arbitrary tangent vector van be written as $\underset{\sim}{T}$. Then,

$$\underset{\sim}{T} = T^u \underset{\sim}{e}_u + T^v \underset{\sim}{e}_v \tag{1.23}$$

A given tangent vector can be 'resolved' in an arbitrary coordinate basis. Say, we are given two coordinat systems u, v and u', v'. If we know the components of a given tangent vector in one coordinate basis, we can compute it in the other coordinate basis easily, since these vectors are just derivatives, and follow the chain rule. First, we express one set of coordinate basis vectors in terms of the other set. Let f

be a function defined on the plane. Then, this can be visualized as either a function of (u, v) or equivalently as a function of (u', v'). Then, using the chain rule of partial derivatives, we get

$$\begin{aligned}
\frac{\partial f}{\partial u'} &= \frac{\partial u}{\partial u'}\frac{\partial f}{\partial u} + \frac{\partial v}{\partial u'}\frac{\partial f}{\partial v} \\
\frac{\partial f}{\partial v'} &= \frac{\partial u}{\partial v'}\frac{\partial f}{\partial u} + \frac{\partial v}{\partial v'}\frac{\partial f}{\partial v}
\end{aligned} \tag{1.24}$$

These can be written abstractly as

$$\begin{aligned}
\frac{\partial}{\partial u'} &= \frac{\partial u}{\partial u'}\frac{\partial}{\partial u} + \frac{\partial v}{\partial u'}\frac{\partial}{\partial v} \\
\frac{\partial}{\partial v'} &= \frac{\partial u}{\partial v'}\frac{\partial}{\partial u} + \frac{\partial v}{\partial v'}\frac{\partial}{\partial v}
\end{aligned} \tag{1.25}$$

Or, using the new notation for coordinate basis vectors,

$$\begin{aligned}
\utilde{e}_{u'} &= \frac{\partial u}{\partial u'}\,\utilde{e}_u + \frac{\partial v}{\partial u'}\,\utilde{e}_v \\
\utilde{e}_{v'} &= \frac{\partial u}{\partial v'}\,\utilde{e}_u + \frac{\partial v}{\partial v'}\,\utilde{e}_v
\end{aligned} \tag{1.26}$$

Given these relations, it is easy to check that the components of an arbitrary tangent vector $\utilde{T}$ evaluated in the two basis sets will be related as follows

$$\begin{aligned}
T^{u'} &= \frac{\partial u'}{\partial u}T^u + \frac{\partial u'}{\partial v}T^v \\
T^{v'} &= \frac{\partial v'}{\partial u}T^u + \frac{\partial v'}{\partial v}T^v
\end{aligned} \tag{1.27}$$

Exercise 1.5.1. In three dimensional Euclidean space, construct coordinate basis vectors associated with spherical polar and cylindrical coordinate systems, and relate them to Cartesian coordinate basis vectors. Further, find the relationships between components of an arbitrary tangent vector betwee these bases.

We now generalize these results to an arbitrary n-dimensional space, with each point characterized by a set of n coordinates. We label these coordinates as u^α. These are arbirary coordinates. Consider a curve in this space characterized by a parameter t. In u^α coordinates, the curve will be described by n equations of the form $u^\alpha = u^\alpha(t)$. The tangent vector to this curve at some point P will be given by

$$\utilde{T} = \sum_{\alpha=1}^{n} T^\alpha\,\utilde{e}_\alpha \tag{1.28}$$

where $\utilde{T} = d/dt$, $T^\alpha = du^\alpha/dt$ and $\utilde{e}_\alpha = \partial/\partial u^\alpha$. The partial derivative $\utilde{e}_\alpha$ is identified with the tangent vector to the curves along which coordinate u^α varies, with all other coordinates constant. It is clear that any tangent vector at point P can be expressed in this form. Then, the tangent vectors e_α form a basis for the vector space of tangent vectors at point P, denoted by V_P. This basis is called the *coordinate basis.* It should be noted that this vector space exists only at this point. Similar vector spaces will exist at other points as well, but in general will have no correlation with each other. The expansion coefficients T^α are components of $\utilde{T}$ in this coordinate basis. We can, of course, describe points in space using any other set of coordinates, say $u'^\mu; \mu = 1, 2, ..n$. We will then get an expression similar to (1.29)

$$\utilde{T} = \sum_{\mu=1}^{n} T'^\mu\,\utilde{e}'_\mu \tag{1.29}$$

where $\underset{\sim}{e}'_\mu = \partial/\partial u'^\mu$. The coordinate basis vectors $\underset{\sim}{e}'_\mu$ form another possible basis for V_P. Using the chain rule for partial derivatives, it is easy to express one set of basis vectors in terms of the other set

$$\underset{\sim}{e}_\alpha = \sum_{\mu=1}^{n} \left(\frac{\partial u'^\mu}{\partial u^\alpha} \right) \underset{\sim}{e}'_\mu \tag{1.30}$$

Then, if we know one set of coordinates as functions of the other set, we can relate the coordinate basis vectors associated with the two coordinate systems. Further, since the same vector $\underset{\sim}{T}$ is being expressed in the two bases, the components T^α and T'^μ are related. It is easy to check that the relationship is

$$T'^\mu = \sum_{\alpha=1}^{n} \left(\frac{\partial u'^\mu}{\partial u^\alpha} \right) T^\alpha \tag{1.31}$$

Exercise 1.5.2. Verify that the components of a tangent vector in two coordinate bases are related by (1.31).

Exercise 1.5.3. In $\mathbb{R}^3$, deduce the expression for the velocity of a particle in Spherical polar and cylindrical coordinates. Verify that the components transform according to (1.31)

1.6 Norm and Inner Product

We now extend the idea of 'magnitude' of a vector and 'dot product' of two vectors to an arbitrary vector space.

Definition A *norm* on a vector space V over a field $\mathbb{F}$ is a map $\|\|: V \longrightarrow \mathbb{R}$ such that

1. $\| \alpha \| \geq 0 \ \forall \left|\alpha\right\rangle \in V$ and $\| \alpha \| = 0 \Leftrightarrow \left|\alpha\right\rangle = \left|\right\rangle_0$.
2. $\forall \left|\alpha\right\rangle \in V$ and $\forall c \in \mathbb{F}$, $\| c\alpha \| = |c| \, \| \alpha \|$.
3. $\forall \left|\alpha\right\rangle, \left|\beta\right\rangle \in V$, $\| \alpha + \beta \| \leq \| \alpha \| + \| \beta \|$ (Triangle Inequality).

Example The set of all real/complex numbers is a normed vector space. The norm is just the modulus.

Example For the space $L^2(a, b)$, a possible definition of norm could be

$$\| \psi \|^2 = \int_a^b dx \, |\psi|^2 \tag{1.32}$$

Example For the space of real $n \times n$ matrices, a norm can be defined as

$$\| A \|^2 = \frac{1}{2} Tr \left(A^T A \right) \tag{1.33}$$

where 'Tr' refers to the sum of all diagonal elements.

If a vector space has a norm defined on it, we can easily extend the idea of 'distance' encountered in spaces such as $\mathbb{R}^n$ (distance between two points in space). This extension makes it a *metric space.*

Definition A metric space is a set V and a map $d : V \times V \longrightarrow \mathbb{R}$ such that

1. $\forall \alpha, \beta \in V$, $d(\alpha, \beta) \geq 0$ and $d(\alpha, \beta) = 0 \Leftrightarrow \alpha = \beta$.
2. $\forall \alpha, \beta \in V$, $d(\alpha, \beta) = d(\beta, \alpha)$.
3. $\forall \alpha, \beta, \gamma \in V$, $d(\alpha, \gamma) \leq d(\alpha, \beta) + d(\beta, \gamma)$.

Example $\mathbb{R}^2$ is a metric space if we define $d(\alpha, \beta) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$, where $\alpha = (x_1, x_2)$ and $\beta = (y_1, y_2)$.

Question: What is the interpretation of (3.) for $\mathbb{R}^2$?

Theorem 1.6.1. *A normed vector space is a metric space if we define* $d(|\alpha\rangle, |\beta\rangle) = \| \, |\alpha\rangle - |\beta\rangle \, \|$.

Proof Trivial.

Definition An *inner product* on a vector space V over a field $\mathbb{F}$ is an assignment to each pair of vectors $|\alpha\rangle$ and $|\beta\rangle$ a scalar $\langle\alpha|\,\beta\rangle \in \mathbb{F}$, satisfying the following conditions

1. $\langle\alpha|\, a\beta + b\gamma\rangle = a\,\langle\alpha|\,\beta\rangle + b\,\langle\alpha|\,\gamma\rangle \qquad \forall a, b \in \mathbb{F}$ (Linearity).
2. $\langle\alpha|\,\beta\rangle^* = \langle\beta|\,\alpha\rangle$.
3. $\langle\alpha|\,\alpha\rangle \geq 0$ and $\langle\alpha|\,\alpha\rangle = 0 \Leftrightarrow |\alpha\rangle = |\rangle_0$.

Note that the order in which the vectors appear in the inner product is important. In particular, though the inner product is linear in the 'second' vector, it is *antilinear* in the first vector

$$\langle a\beta + b\gamma|\,\alpha\rangle = a^*\,\langle\beta|\,\alpha\rangle + b^*\,\langle\gamma|\,\alpha\rangle \qquad \forall a, b \in \mathbb{F}$$

Example On vector space $\mathbb{C}^n$ with elements $|\alpha\rangle = (x_1, x_2, .., x_n)$ and $|\beta\rangle = (y_1, y_2, .., y_n)$, the inner product is defined to be

$$\langle\alpha|\,\beta\rangle = \sum_{i=1}^{n} x_i^* y_i \tag{1.34}$$

Example On space l^2 with elements $|\alpha\rangle = (x_1, x_2, .., x_k, ..)$ and $|\beta\rangle = (y_1, y_2, .., y_k, ..)$ satisfying constraints of the form (1.2), the inner product is defined to be

$$\langle\alpha|\,\beta\rangle = \sum_{i=1}^{\infty} x_i^* y_i \tag{1.35}$$

This is guaranteed to converge (will be proved later).

Example On space $L^2(a, b)$ of square integrable complex functions $f(x)$ and $g(x)$ defined over $[a, b]$, the inner product can be defined as

$$\langle f|\,g\rangle = \int_a^b dx \; f(x)^* g(x) \tag{1.36}$$

This is often be generalized to

$$\langle f|\,g\rangle = \int_a^b dx \; f(x)^* g(x) w(x) \tag{1.37}$$

where $w(x) \geq 0$ is called a *weight function.*

Theorem 1.6.2. *Schwarz Inequality: Any inner product on a complex vector space satisfies the inequality*

$$|\langle\alpha|\,\beta\rangle| \leq \sqrt{\langle\alpha|\,\alpha\rangle}\;\sqrt{\langle\beta|\,\beta\rangle} \tag{1.38}$$

Proof Let $|\psi\rangle = |\alpha\rangle + c\,|\beta\rangle$. Then, since $\langle\psi|\,\psi\rangle \geq 0$,

$$\begin{aligned} \langle\alpha + c\beta|\,\alpha + c\beta\rangle &\geq 0 \\ \Rightarrow \langle\alpha|\,\alpha\rangle + c^* c\,\langle\beta|\,\beta\rangle + c\,\langle\alpha|\,\beta\rangle + c^*\,\langle\beta|\,\alpha\rangle &\geq 0 \end{aligned}$$

Choosing $c = -\langle\beta|\,\alpha\rangle / \langle\beta|\,\beta\rangle$ gives the inequality.

Note: The equality holds only if $\langle\psi|\,\psi\rangle = 0$, which implies that $|\psi\rangle = |\alpha\rangle + c\,|\beta\rangle = 0$. That is, $|\alpha\rangle$ and $|\beta\rangle$ are linearly dependent.

Given an innerproduct defined on a vector space, a norm can be defined as

$$\parallel \alpha \parallel = \sqrt{\langle\alpha|\,\alpha\rangle} \tag{1.39}$$

Exercise 1.6.1. Prove the *Triangle Inequality* for vectors:

$$\parallel \alpha + \beta \parallel \leq \parallel \alpha \parallel + \parallel \beta \parallel \tag{1.40}$$

Exercise 1.6.2. Verify that the norm of a vector defined by (1.39) satisfies all the properties of a norm.

Exercise 1.6.3. Prove that the sum of square integrable functions is square integrable.

Exercise 1.6.4. Prove that the inner produce defined over l^2 converges.

Exercise 1.6.5. Show that the following satisfies properties of an inner product on the space of $n \times n$ complex matrices

$$\langle A|\,B\rangle = \frac{1}{2}Tr\left(A^\dagger B\right)$$

where $A^\dagger = (A^T)^*$ and Tr denotes the trace of a matrix (sum of diagonal elements).

Exercise 1.6.6. We can try to define an inner product over the space of functions in ways other than (1.36) or (1.37). For instance, consider the ector space of polynomials of degree less than n. Let us try to define an inner product as follows:

$$\langle P_n|\,P_m\rangle = \sum_{i=1}^{n} P_n(x_i)P_m(x_i) \tag{1.41}$$

where x_i are n distinct real numbers. Does this satisfy all the conditions for an inner product?

Definition A pair of vectors is said to be *orthogonal* if their inner product vanishes.

Definition A vector is said to be *normalized* if its norm is equal to unity.

Definition A pair of vectors $|\alpha\rangle$ and $|\beta\rangle$ is said to be an *orthonormal pair* if $\langle\alpha|\,\beta\rangle = 0$ and $\parallel \alpha \parallel = \parallel \beta \parallel = 1$.

Definition An orthonormal set of vectors is a set in which each vector is normalized and every pair of vectors is orthogonal to each other.

Definition In a finite dimensional vector space, an *orthonormal basis* is a basis set $|\alpha_1\rangle, |\alpha_2\rangle, .., |\alpha_n\rangle$ (where n is the dimension of the space) such that

$$\langle\alpha_i|\,\alpha_j\rangle = \delta_{ij} \tag{1.42}$$

Example In the space $L^2(0,1)$, the set of functions $\psi_n(x) = \sqrt{2}\cos(2\pi nx)$ for $n = 0, 1, ..$ and $\psi_n(x) = \sqrt{2}\sin(2\pi nx)$ for $n = 1, 2, ...$ form an infinite orthonormal set. In fact, Fourier's theorem tells us that they form a basis for $L^2(0,1)$.

Theorem 1.6.3. *Gram-Schmidt orthonormalization: Every finite dimensional vector space has an orthonormal basis.*

Proof Let $|\psi_1\rangle, |\psi_2\rangle, .., |\psi_n\rangle$ be a non-orthonormal basis for a vector spoace V. We pick any one vector, say $|\psi_1\rangle$, and produce a normalized vector,

$$|\alpha_1\rangle = \frac{1}{\parallel \psi_1 \parallel}\,|\psi_1\rangle$$

Next, let

$$|\beta_2\rangle = |\psi_2\rangle - |\alpha_1\rangle \langle\alpha_1|\, \psi_2\rangle$$

Geometrically, $|\alpha_1\rangle \langle\alpha_1|\, \psi_2\rangle$ is the 'projection' of $|\psi_2\rangle$ along the unit vector $|\alpha_1\rangle$. Then, $|\beta_2\rangle$ is $|\psi_2\rangle$ with its component along $|\alpha_1\rangle$ projected out. Then, as is easy to check, $\langle\alpha_1|\, \beta_2\rangle = 0$. We normalize the vector $|\beta_2\rangle$ to define $|\alpha_2\rangle$

$$|\alpha_2\rangle = \frac{1}{\parallel \beta_2 \parallel} |\beta_2\rangle$$

Then, we have constructed a pair of orthonormal vectors $|\alpha_1\rangle$ and $|\alpha_2\rangle$ out of the pair $|\psi_1\rangle$ and $|\psi_2\rangle$. Next, project out the components of $|\psi_3\rangle$ along the unit vectors $|\alpha_1\rangle$ and $|\alpha_2\rangle$ to get vector $|\beta_3\rangle$

$$|\beta_3\rangle = |\psi_3\rangle - |\alpha_1\rangle \langle\alpha_1|\, \psi_3\rangle - |\alpha_2\rangle \langle\alpha_2|\, \psi_3\rangle$$

Again, it is easy to check that $|\beta_3\rangle$ is orthogonal to $|\alpha_1\rangle$ and $|\alpha_2\rangle$. Next, we normalize it to get the unit vector $|\alpha_3\rangle$, and so on, till we end up with the orthonormal set $|\alpha_1\rangle, |\alpha_2\rangle, .., |\alpha_n\rangle$, which forms the orthonormal basis. Then, starting with any basis, we can always construct an orthonormal basis.

Given an orthonormal basis in a vector space, the expansion of an arbitrary vector in the basis vectors has a particularly simple form. Let $|\alpha_1\rangle, |\alpha_2\rangle, .., |\alpha_n\rangle$ be an orthonormal basis. Let the expansion of a vector $|\psi\rangle$ have the form

$$|\psi\rangle = c_1 |\alpha_1\rangle + c_2 |\alpha_2\rangle + .. + c_n |\alpha_n\rangle \tag{1.43}$$

Taking the inner product with $|\alpha_k\rangle$ and using the linearity of the inner product

$$\begin{aligned} \langle\alpha_k|\, \psi\rangle &= \sum_{i=1}^{n} c_i \langle\alpha_k|\, \alpha_i\rangle \\ &= \sum_{i=1}^{n} c_i \delta_{ki} \\ &= c_k \end{aligned}$$

Then, the expansion is

$$|\psi\rangle = \sum_{i=1}^{n} |\alpha_i\rangle \langle\alpha_i|\, \psi\rangle \tag{1.44}$$

Given two vectors $|\psi\rangle$ and $|\phi\rangle$, given their expansion in an orthonormal basis, it is a particularly easy to compute their norm and inner product. Say, these vectors have expansions

$$\begin{aligned} |\psi\rangle &= \sum_{i=1}^{n} a_i |\alpha_i\rangle \;\; ; \quad a_i = \langle\alpha_i|\, \psi\rangle \\ |\phi\rangle &= \sum_{i=1}^{n} b_i |\alpha_i\rangle \;\; ; \quad b_i = \langle\alpha_i|\, \phi\rangle \end{aligned}$$

Then,

$$\begin{aligned} \langle\psi|\, \phi\rangle &= \sum_i \sum_j a_i^* b_j \langle\alpha_i|\, \alpha_j\rangle \\ &= \sum_i \sum_j a_i^* b_j \delta_{ij} \\ &= \sum_{i=1}^{n} a_i^* b_j \end{aligned} \tag{1.45}$$

Similarly,

$$\begin{aligned}\| \psi \|^2 &= \langle\psi|\ \psi\rangle \\ &= \sum_i \sum_j a_i^* a_j \langle\alpha_i|\ \alpha_j\rangle \\ &= \sum_i \sum_j a_i^* a_j \delta_{ij} \\ &= \sum_{i=1}^{n} |a_i|^2 \end{aligned} \tag{1.46}$$

Exercise 1.6.7. $\alpha = (1,1,1), \beta = (-1,1,0)$ and $\gamma = (-1,0,1)$ are a basis for $\mathbb{R}^3$. Construct an orthonormal basis.

Exercise 1.6.8. Consider the space $\mathcal{P}_3(x)$ of polynomials of degree 3. The standard basis for this space is the set $\{1, x, x^2, x^3\}$. Using Gram-Schmidt orthonormalization, construct an orthonormal basis over the inerval $[-1, 1]$.

Exercise 1.6.9. Consider the vector space $L^2(\mathbb{R})$. A possible basis for this space is the set of functions $f_n(x) = x^n e^{-x^2/2}; n = 0, 1, 2, ...$ Consider the subspace spanned by the functions f_0, f_1, f_2 and f_3. Construct an orthonormal basis for this subspace.

Exercise 1.6.10. Consider the vector space $L^2(0,\infty)$. A possible basis for this space is the set of functions $f_n(x) = x^n e^{-x/2}; n = 0, 1, 2, ...$ Consider the subspace spanned by the functions f_0, f_1, f_2 and f_3. Construct an orthonormal basis for this subspace.

Definition Let S be a subspace of an inner product space V. The *orthogonal complement* of S is the set of all vectors in V that are orthogonal to every vector in S. It is denoted by $S^\perp$.

Example Let S be a subsapce of $\mathbb{R}^4$ consisting of vectors of the form $(x_1, x_2, 0, 0)$. Then, $S^\perp$ is the set of vectors of the form $(0, 0, x_3, x_4)$.

Theorem 1.6.4. *Let S be a subspace of an inner product space V. Then*
(a) $S^\perp$ is a subspace of V.
(b) $S \cap S^\perp = 0$.

Proof Let $|\psi\rangle, |\phi\rangle \in S^\perp$, and $|\alpha\rangle \in S$. Then,

$$\begin{aligned}\langle\alpha|\ \psi+\phi\rangle &= \langle\alpha|\ \psi\rangle + \langle\alpha|\ \phi\rangle \\ &= 0\end{aligned}$$

and

$$\begin{aligned}\langle\alpha|\ c\psi\rangle &= c\,\langle\alpha|\ \psi\rangle \\ &= 0\end{aligned}$$

Then, $S^\perp$ is clearly a subspace. Let there be some vector $|\alpha\rangle$ *in* S and $|\alpha\rangle$ *in* $S^\perp$. Then, $\langle\alpha|\ \alpha\rangle = 0$. But, this implies that $|\alpha\rangle = |\rangle_0$. Therefore, $S \cap S^\perp = 0$.

Theorem 1.6.5. *Let S be a subspace of a finite dimensional inner product space V. Then $V = S \oplus S^\perp$ and $dim(V) = dim(S) + dim(S^\perp)$.*

Proof We already know that $S \cap S^\perp = 0$. Therefore, we only need to prove that $V = S + S^\perp$. Let V be n-dimensional, and S m-dimensional. Let $\{|\alpha_1\rangle, |\alpha_2\rangle, .., |\alpha_m\rangle$ be an orthonormal basis of S. Then, we can extend this to an orthonormal basis of V by adding vectors $|\alpha_{m+1}\rangle, .., |\alpha_n\rangle$ which are orthogonal to the basis vectors of S. These $n - m$ orthonormal vectors clearly belong to $S^\perp$ since each is perpendicular to all vectors in S (since they are perpendicular to all the basis vectors of S). Furthur, these must be the maximal set of orthonormal vectors in $S^{\perp}$, for if there are any more, they together with the basis vectors of S will exceed to dimension of V. Then, these form an orthonormal basis for $S^{\perp}$. Since the two orthonormal sets together form a basis for V, it is easy to see that any vector in V is the sum of a vector in S and a vector in $S^{\perp}$. Further, $dim(S) = m$ and $dim(S^{\perp}) = n - m$, so that $dim(V) = dim(S) + dim(S^{\perp})$.

Any vector in V can be uniquely expressed as

$$|\psi\rangle = |\psi\rangle_S + |\psi\rangle_{S^{\perp}} \tag{1.47}$$

where $|\psi\rangle_S \in S$ and $|\psi\rangle_{S^{\perp}} \in S^{\perp}$. The vector $|\psi\rangle_S \in S$ is said to be the *projection* of $|\psi\rangle$ on the subspace S. Given any vector space (finite or infinite), let S be a subspace with orthonormal basis $\{|\alpha_1\rangle, |\alpha_2\rangle, .., |\alpha_k\rangle, ...\}$. Then, the projection of $|\psi\rangle \in V$ onto S can be obtained as

$$|\psi\rangle_S = \sum_i |\alpha_i\rangle \langle\alpha_i|\, \psi\rangle \tag{1.48}$$

where the sum runs only over the basis vectors of S.

The projection of a vector onto a subspace has the following interpretation: it is the vector in the subspace which best approximates the actual vector. To see this, let S be a subspace of the vector space V. Then, a vector $|\psi\rangle \in V$ can always be expressed as (1.47). Let $|\phi\rangle \in S$. Then, a reasonable measure of how 'different' $|\phi\rangle$ is from $|\psi\rangle$ is the norm of their difference, $\parallel \psi - \phi \parallel$. Given the decomposition (1.47), we have

$$\begin{aligned} \parallel \psi - \phi \parallel^2 &= \parallel (\psi_S - \phi) + \psi_{S^{\perp}} \parallel^2 \\ &= \parallel \psi_S - \phi \parallel^2 + \parallel \psi_{S^{\perp}} \parallel^2 \end{aligned}$$

since $\langle\psi_{S^{\perp}}|\, \phi\rangle = 0$. This is clearly least, when $\parallel \psi_S - \phi \parallel = 0$. That is, when $|\phi\rangle = |\psi\rangle_S$. Therefore, the vector in subspace S which is 'closest' to $|\psi\rangle \in V$ is the projection of $|\psi\rangle$ onto subspace S.

Exercise 1.6.11. In the inner product space $\mathcal{P}_3(x)$, with the inner product defined over $x \in [0, 1]$, find the orthogonal complement of the subspace generated by 1 and x.

Exercise 1.6.12. Prove that $(S^{\perp})^{\perp} = S$ for any subspace of a finite dimensional vector space.

Exercise 1.6.13. Vectors $\alpha_1 = (1/\sqrt{6})(1, -2, 1)$ and $\alpha_2 = (1/\sqrt{5})(2, 0, 1)$ form an orthonormal basis for a two-dimensional subspace of $\mathbb{R}^3$. Find the projection of $\alpha = (1, -1, 1)$ onto this subspace, and express α as a sum of a vector belonging to this subspace, and its orthogonal complement.

Exercise 1.6.14. Find the polynomial of degree three that best approximates e^x over the interval $x \in [-1, 1]$. What about the interval $x \in [0, 1]$?

Exercise 1.6.15. Consider the vector space $L^2(-\pi, \pi)$. A possible orthonormal basis for this space is the set of functions $\phi_0 = 1/\sqrt{2\pi}$, $\phi_n = 1/\sqrt{\pi}\cos(nx)$ and $\psi_n = 1/\sqrt{\pi}\sin(nx)$ for $n \in \mathbb{N}$. Given an arbitrary function $f(x)$, devise an algorithm to determine the linear superposition of ϕ_0, ϕ_n and ψ_n for $n = 1, 2, ..N$ that best approximates the function $f(x)$ over this interval.

1.7 Inner product of tangent vectors: the metric

We now define the inner product in the vector space of tangent vectors (section (1.5)). Consider an n-dimensional space, with each point characterized by a set of n coordinates u^{α}. Consider a curve in this space characterized by a parameter t. The tangent vector to this curve at some point P will be given by (1.29)

$$\underset{\sim}{T} = \sum_{\alpha=1}^{n} \frac{du^{\alpha}}{dt}\, \underset{\sim}{e}_{\alpha} \tag{1.49}$$

where $\underset{\sim}{T} = d/dt$ and $\underset{\sim}{e}_\alpha = \partial/\partial u^\alpha$. Let the distance between point P (with coordinates $u^\alpha(t)$) and a neighboring point Q (with coordinates $u^\alpha(t + dt)$ be ds. We define an inner product on the tangent vectors, such that apart from satisfying the general properties of the inner product (linearity, etc.), it is such that

$$ds^2 = \underset{\sim}{T} \cdot \underset{\sim}{T}\ dt^2 \tag{1.50}$$

Note that this is just a generalization of the dot product of tangent vectors (such as velocity of a particle along a curve parametrized by time) in $\mathbb{R}^3$ (three dimensional Euclidean space). Now, the general expression for this distance squared in the coordinates u^α will be quadratic in du^α (infinitesimal change in coordinates u^α from P to Q), and will be of the form

$$ds^2 = \sum_{\alpha,\beta=1}^{n} g_{\alpha\beta} du^\alpha du^\beta \tag{1.51}$$

This distance (squared) is called a *metric*, with the set of n^2 numbers called a 'components' of the metric in the coordinate system u^α. If we change the coordinate system to some other coordinates u'^μ, the *same* metric will have a different form

$$ds^2 = \sum_{\mu,\nu=1}^{n} g'_{\mu\nu} du'^\mu du'^\nu \tag{1.52}$$

Substituting (1.49) in (1.50) and using the linearity of the inner product, we get

$$ds^2 = \left(\sum_{\alpha,\beta=1}^{n} \frac{du^\alpha}{dt} \frac{du^\beta}{dt}\ \underset{\sim}{e}_\alpha \cdot \underset{\sim}{e}_\beta \right) dt^2 \tag{1.53}$$

Cancelling the dt's, we get

$$ds^2 = \sum_{\alpha,\beta=1}^{n} du^\alpha du^\beta \left(\underset{\sim}{e}_\alpha \cdot \underset{\sim}{e}_\beta \right) \tag{1.54}$$

Comparing with the expression for the metric components, we deduce that

$$\underset{\sim}{e}_\alpha \cdot \underset{\sim}{e}_\beta = g_{\alpha\beta} \tag{1.55}$$

Then, given any coordinate system, the associated tangent vectors have inner products related to the metric components in that coordinate system. As an example, consider three dimensional Euclidean space. This is defined as a space in which there exists a coordinate system (x, y, z), such that at *any* point in this space, the infinitesimal distance between two points with coordinates (x, y, z) and $(x + dx, y + dy, z + dz)$ has the form

$$ds^2 = dx^2 + dy^2 + dz^2 \tag{1.56}$$

The metric for this space in this coordinate system has the simple form

$$g_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{1.57}$$

We recognize this as the Kronecker delta. That is, $g_{ij} = \delta_{ij}$. From (1.55), we see that the associated coordinate basis tangent vectors $\underset{\sim}{e}_x$, $\underset{\sim}{e}_y$ and $\underset{\sim}{e}_z$ are othonormal. These are in fact to be identified with the traditional Cartesian unit vectors $\hat{i}$, $\hat{j}$ and $\hat{k}$ respectively. Now, let us express the same metric in spherical polar coordinates defined by

$$\begin{aligned} x &= r \sin\theta \cos\phi \\ y &= r \cos\theta \cos\phi \\ z &= r \cos\theta \end{aligned} \tag{1.58}$$

In these coordinates, the metric has the form

$$ds^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2 \tag{1.59}$$

Since the 'cross-terms' of coordinate differentials are not there in this expression, we immediately see that the coordinate basis vectors $\underset{\sim}{e}_r, \underset{\sim}{e}_\theta$ and $\underset{\sim}{e}_\phi$ are orthogonal to each other. However, they are not all normalized. In fact, we see from (1.55) that

$$\begin{aligned} \underset{\sim}{e}_r \cdot \underset{\sim}{e}_r &= g_{rr} \\ &= 1 \\ \underset{\sim}{e}_\theta \cdot \underset{\sim}{e}_\theta &= g_{\theta\theta} \\ &= r^2 \\ \underset{\sim}{e}_\phi \cdot \underset{\sim}{e}_\phi &= g_{\phi\phi} \\ &= r^2 \sin^2\theta \end{aligned} \tag{1.60}$$

We can then construct unit vectors along these tangent vectors. They will be

$$\begin{aligned} \hat{e}_r &= \underset{\sim}{e}_r \\ \hat{e}_\theta &= \frac{1}{r}\, \underset{\sim}{e}_\theta \\ \hat{e}_\phi &= \frac{1}{r\sin\theta}\, \underset{\sim}{e}_\phi \end{aligned} \tag{1.61}$$

Let us compute the expression for the velocity of a particle in spherical polar coordinates. Since velocity is tangent to the trajectory of the particle parametrized by time, we have (from (1.49))

$$\underset{\sim}{v} = \frac{dr}{dt}\, \underset{\sim}{e}_r + \frac{d\theta}{dt}\, \underset{\sim}{e}_\theta + \frac{d\phi}{dt}\, \underset{\sim}{e}_\phi \tag{1.62}$$

We now replace the coordinate basis vectors by unit vectors along the coordinate directions, using (1.61). We get

$$\underset{\sim}{v} = \frac{dr}{dt}\, \hat{e}_r + r\frac{d\theta}{dt}\, \hat{e}_\theta + r\sin\theta \frac{d\phi}{dt}\, \hat{e}_\phi \tag{1.63}$$

which is the usual expression for velocity.

The concept of a *metric* allows us to generalize the idea of an inner product so that the third axiom of inner product ($\langle\alpha|\ \alpha\rangle \geq 0$ and $\langle\alpha|\ \alpha\rangle = 0 \Leftrightarrow |\alpha\rangle = |\rangle_0$) is relaxed. For real vector spaces, this results in a rule $V \times V \to \mathbb{R}$ which is no longer called an inner product, but a *symmetric bilinear*. However, we will keep pretending it is an inner product, but with the third axiom relaxed. This allows us to explore inner product on tangent vectors defined on spaces which are more interesting, such as *Minkowski Space*. We define Minkowski space as a four dimensional space in which there exists a coordinate system (x^0, x^1, x^2, x^3) such that the expression for the infinitesimal 'distance squared' (metric) takes the form

$$ds^2 = -(dx^0)^2 + (dx^1)^2 + (dx^2)^2 + (dx^3)^2 \tag{1.64}$$

Physically, x^0 is the time coordinate and $x^i : i = 1, 2, 3$ the Cartesian spatial coordinates assigned to an event by an inertial observer. Why is this being called a 'distance'? Given another observer in relative motion, say in the x^1- direction, using coordinates (x'^0, x'^1, x'^2, x'^3), the coordinates are related by a Lorentz transformation (we are using units in which speed of light is unity)

$$\begin{aligned} x'^0 &= \gamma(v)\left(x^0 - vx^1\right) \\ x'^1 &= \gamma(v)\left(x^1 - vx^0\right) \\ x'^2 &= x^2 \\ x'^3 &= x^3 \end{aligned} \tag{1.65}$$

where $\gamma(v) = 1/\sqrt{1-v^2}$, v being the relative velocity (along the $x-$ direction, in this case)[1]. It is easy to check that under transformation from coordinates (x^0, x^1, x^2, x^3) to (x'^0, x'^1, x'^2, x'^3), the form of ds^2 remains unchanged. Then, all inertial observers, assigning different coordinates to events, measure ds^2 to be the same between a pair of infinitesimally close events. This prompts us to believe that perhaps we live in a four-dimensional 'space-time', which is a geometrical entity (such as three-dimensional Euclidean space) in its own right, with absolute distances between points (events) defined. Then, associated with any inertial coordinate system is a set of tangent vectors $\underset{\sim}{e}_0 = \partial/\partial x^0, \underset{\sim}{e}_1 = \partial/\partial x^1, \underset{\sim}{e}_2 = \partial/\partial x^1$ anbd $\underset{\sim}{e}_3 = \partial/\partial x^3$, and any vector tangent to a curve in this four-dimensional space can be expressed as a linear combination of these basis vectors. Of course, associated with any inertial observer is a similar set of tangent vectors, the sets between different inertial observers related through the chain rule indiced by Lorentz transformations. What is the significance of a vector tangent to a curve in space-time? To understand this, we first note that given a minus sign in the metric (1.64), the dot product of a vector with itself need not be positive definite. For instance, we can directly read off the dot products of the coordinate basis vectors associated with an inertial observer. These will be

$$\begin{aligned} \underset{\sim}{e}_0 \cdot \underset{\sim}{e}_0 &= -1 \\ \underset{\sim}{e}_i \cdot \underset{\sim}{e}_i &= \delta_{ij}; \quad i,j = 1,2,3 \\ \underset{\sim}{e}_0 \cdot \underset{\sim}{e}_i &= 0 \end{aligned} \tag{1.66}$$

These basis vectors are orthogonal to each other, and even orthonormal, in a general sense. Note that the tangent vector $\underset{\sim}{e}_0$ associated with the time coordinate x^0 has a negative inner product with itself. Clearly, this is an example of a situation where we need to relax the positivity rule of the traditional inner product. Also, note that it is possible in this space for a non-trivial vector to have a vanishing inner product with itself. For example, given $\underset{\sim}{u} = \underset{\sim}{e}_0 + \underset{\sim}{e}_1$, it is easy to check (using linearity of the inner product) that $\underset{\sim}{u} \cdot \underset{\sim}{u} = 0$. What is the significance of this tangent vector? To find out, we construct the curve to which it is tangent. Let this curve be parametrized by some parameter λ, and its form be given by $x^\mu = x^\mu(\lambda)$, where $\mu = 0,1,2,3$. Then, it follows that

$$\underset{\sim}{u} = \sum_{\mu=0}^{3} \frac{dx^\mu}{d\lambda}\, \underset{\sim}{e}_\mu \tag{1.67}$$

Given $\underset{\sim}{u} = \underset{\sim}{e}_0 + \underset{\sim}{e}_1$, it follows that the equation of the curve satisfies

$$\begin{aligned} \frac{dx^0}{d\lambda} &= 1 \\ \frac{dx^1}{d\lambda} &= 1 \\ \frac{dx^2}{d\lambda} &= \frac{dx^3}{d\lambda} = 0 \end{aligned}$$

The first two equations can be expressed as

$$\frac{dx^1}{dx^0} = 1$$

which is the equation for velocity of an entity, the velocity being equal to 1. Since we are using units in which the speed of light is unity, the curve in question is the trajectory of a light signal in space-time. What about trajectories of other objects? Consider a curve in space-time depicting the trajectory of a particle. A physical parameter for this curve is the so-called 'proper-time' of the particle, which is the time read by a clock carried by the particle. Each instant of time τ on this clock corresponds to some point in space-time, which is assigned coordinates $x^\mu = x^\mu(\tau)$ by an inertial observer. We can visualize this assignment as follows: imagine three-dimensional space filled by a rectangular grid of rigid rods (of infinitesimal thickness). Along each rod, a Cartesian coordinate varies, other two being constant.

[1] If the direction of relative motion is reversed, the sign of v reverses

The points of inersection of these rods form a three-dimensional lattice, each point on the lattice having Cartesian coordinates (x^1, x^2, x^2). At each lattice point is attached a clock, all such clocks assumed to be synchronized, ticking at the same rate, reading time x^0. As the particle moves across the lattice, it is assigned the three Cartesian coordinates and a time coordinate which is just the time shown by the clock attached to the lattice point. The tangent vector to the trajectory is just the derivative with respect to the parameter (the proper time here). We call this tangent vector the 'four-velocity' $\underset{\sim}{u} = d/d\tau$ of the particle (for reasons which will become clear in a moment). The four-velocity vector can be 'resolved' as usual in the coordinate basis

$$\underset{\sim}{u} = \frac{dx^0}{d\tau}\,\underset{\sim}{e}_0 + \frac{dx^1}{d\tau}\,\underset{\sim}{e}_1 + \frac{dx^2}{d\tau}\,\underset{\sim}{e}_2 + \frac{dx^3}{d\tau}\,\underset{\sim}{e}_3 \tag{1.68}$$

Let us calculate the inner product of the four-velocity with itself. Since it is a geometrical object (tangent to a curve), this inner product should be the same, whatever coordinates are used to evaluate it. In a general inertial coordinate system, this inner product can be easily calculated using linearity and the inner products of the basis vectors. Then,

$$\underset{\sim}{u} \cdot \underset{\sim}{u} = -\left(\frac{dx^0}{d\tau}\right)^2 + \left(\frac{dx^1}{d\tau}\right)^2 + \left(\frac{dx^2}{d\tau}\right)^2 + \left(\frac{dx^3}{d\tau}\right)^2 \tag{1.69}$$

Let us choose an inertial frame in which this expression is the simplest. Consider an inertial frame, in which the particle is momentarily at rest when its clock reads proper time τ. Let the coordinates used by this inertial observer be (x'^0, x'^1, x'^2, x'^3). since the particle is at rest in this coordinate system, $dx'^1/d\tau = dx'^2/d\tau = dx'^3/d\tau = 0$. Further, the infinitesimal time dx'^0 lapsing at the latticle clock at whose position the particle is (momentarily) at rest should be the same as $d\tau$, the infinitesimal time duration lapsing on the clock carried by the observer[2]. Then, in this coordinate system, we see that $\underset{\sim}{u} \cdot \underset{\sim}{u} = -1$. Since this is independent of coordinates, it follows that the four-velocty is a 'unit vector' (apart from the sign), whatever the motion of the particle. We can now use this fact and substitute it in (1.69), where it is evaluated in an arbitrary inertial coordinate system, in which the particle is not at rest, but moving with 'three velocty' with components $(dx^1/dt, dx^2/dt, dx^3/dt) = (dx^1/dx^0, dx^2/dx^0, dx^3/dx^0)$. The substitution gives

$$-\left(\frac{dx^0}{d\tau}\right)^2 + \left(\frac{dx^1}{d\tau}\right)^2 + \left(\frac{dx^2}{d\tau}\right)^2 + \left(\frac{dx^3}{d\tau}\right)^2 = -1 \tag{1.70}$$

which can be simplified to give

$$\frac{d\tau}{dt} = \sqrt{1 - v^2} \tag{1.71}$$

where $v^2 = (dx^1/dt)^2 + (dx^2/dt)^2 + (dx^3/dt)^2$. This the so-called time-dilation formula. It says that as the particle moves with speed v (which could change with time), between any two infinitesimally close positions along its trajectory, the time lapsed on its own clock is always less than the time that lapses on the inertial clocks (located at the imaginary lattice points) at these points. Then, time-dilation could be thought of as a consequence of the geometric fact that the four-velocity has a fixed 'magnitude' (equal to one). Let us compute the components $dx^\mu/d\tau$ of four-velocity in an inertial coordinate basis $\underset{\sim}{e}_\mu$. Using (1.71), we find

$$\begin{aligned} \frac{dx^0}{d\tau} &= \frac{dt}{d\tau} \\ &= \gamma(v) \end{aligned} \tag{1.72}$$

where v is the magnitude of the 'three-velocity' $\vec{v} = d\vec{x}/dt$. Similarly,

$$\begin{aligned} \frac{dx^i}{d\tau} &= \frac{dt}{d\tau}\frac{dx^i}{dt} \\ &= \gamma(v)v^i \end{aligned} \tag{1.73}$$

[2]This is the so-called clock hypothesis, which essentially implies that two clocks, irrespective of their motion (accelerated or otherwise), should tick at the same rate if they are relatively at rest. The idea is that any effects of acceleration can be negated by suitably adjusting the mechanism of the clock.

where $v^i = dx^i/dt$ is the i^{th} component of the three-velocity $\vec{v}$. Then, we finally have

$$\underset{\sim}{u} = \gamma(v)\ \underset{\sim}{e}_0 + \gamma(v) \sum_{i=1}^{3} v^i\ \underset{\sim}{e}_i \tag{1.74}$$

Of course, the last piece in this expression is just the three-velocity

$$\vec{v} = \sum_{i=1}^{3} v^i\ \underset{\sim}{e}_i \tag{1.75}$$

The *four-momentum* $\underset{\sim}{p}$ of a particle is defined to be a tangent vector in Minkowski space, equal to the 'rest mass' m of the particle times the four-velocity. Then,

$$\begin{aligned} \underset{\sim}{p} &= m\underset{\sim}{u} \\ &= m\gamma(v)\ \underset{\sim}{e}_0 + m\gamma(v)\vec{v} \end{aligned} \tag{1.76}$$

We compactly write this as

$$\underset{\sim}{p} = E\ \underset{\sim}{e}_0 + \vec{p} \tag{1.77}$$

where $\vec{p} = m\gamma(v)\vec{v}$. E is then identified by the energy of the particle, and $\vec{p}$ its 'three-momentum'. It is clear that these satisfy the relation

$$E = \sqrt{p^2 + m^2} \tag{1.78}$$

where $p^2 = \vec{p} \cdot \vec{p}$.

In any geometrical space, the inner product of two vectors has an absolute, geometrical significance, independent of any basis used to evaluate it. Let us evaluate a useful inner product in Minkowski space. Consider an (in general) accelerating observer with four-velocity $\underset{\sim}{u}_{obs}$. Say, there is a particle with four-momentum $\underset{\sim}{p}$ whose trajectory intersects with that of the observer at some point. At that point, there is a tangent space, with the tangent vectors $\underset{\sim}{u}_{obs}$ and $\underset{\sim}{p}$ belonging to this vector space. Let us interpret their inner product. To evaluate it, we choose a coordinate system attributed to an inertial observer who is momentarily at rest with respect to the accelerating observer at that event. Then, in this inertial system, the four-velocity of the accelerating observer has expression $\underset{\sim}{u}_{obs} = \underset{\sim}{e}_0$ where $\underset{\sim}{e}_0$ is the coordinate basis vector for the inertial observer along his time-axis. In this coordinate system, the four-momentum of the particle will have the form (1.77), where E is the energy of the particle in this frame of reference, and $\vec{p}$ its three-momentum. The inner product of $\underset{\sim}{u}_{obs}$ and $\underset{\sim}{p}$ gives

$$\begin{aligned} \underset{\sim}{u}_{obs} \cdot \underset{\sim}{p} &= \underset{\sim}{e}_0 \cdot (E\ \underset{\sim}{e}_0 + \vec{p}) \\ &= -E \end{aligned} \tag{1.79}$$

since $\underset{\sim}{e}_0 \cdot \underset{\sim}{e}_0 = -1$ and $\underset{\sim}{e}_0 \cdot \vec{p} = 0$. Now, E is the energy of the particle as measured in this particular frame (in which the accelerating observer is at rest). The, it is the same as the energy measured by the accelerating observer. Then, we can attach an invariant meaning to this inner product. The inner product of the four velocity of an observer and the four-momentum of a particle is related to the energy of the particle *measured by this observer*

$$\underset{\sim}{u}_{obs} \cdot \underset{\sim}{p} = -E_{obs} \tag{1.80}$$

where E_{obs} is the energy of the particle measured by the observer with four-velocity $\underset{\sim}{u}_{obs}$.

Exercise 1.7.1. Consider a two-dimensional Euclidean space with Cartesian coordinates defined on it. Introduce hyperbolic coordinates u, v defined by

$$\begin{aligned} u &= -\frac{1}{2}\log\left(\frac{y}{x}\right) \\ v &= \sqrt{xy} \end{aligned} \tag{1.81}$$

Plot curves of constant u and v and describe the region of space over which this is a 'good' coordinate system. Construct the coordinate basis vectors as linear superpositions of the Cartesian unit vectors at every point. Are these vectors orthonormal? find the velocity of a particle in this plane in terms of unit vectors along the u and v coordinate directions.

Exercise 1.7.2. In the same Euclidean space, introduce parabolic coordinates σ, τ defined by

$$\begin{aligned} x &= \sigma\tau \\ y &= \frac{1}{2}\left(\tau^2 - \sigma^2\right) \end{aligned} \tag{1.82}$$

Plot curvers of constant σ and τ and describe the region of space over which this is a 'good' coordinate system. Construct the coordinate basis vectors as linear superpositions of the Cartesian unit vectors at every point. Are these vectors orthonormal? Find the velocity of a particle in this plane in terms of unit vectors along the σ and τ coordinate directions.

Exercise 1.7.3. *Relativistic Doppler shift*: Assume that the energy of a photon is related to its frequency as $E = h\nu$. Furthur, assume that the magnitude of its momentum is related to its energy by $E = pc$. Let us use the inner product (1.80) to derive the expression for relativistic doppler effect. Assume that an accelerating observer is holding a source of light which emits photons with frequency ν_0. Analyze the inner product of the four-velocity of this observer with the four-momentum of the photon in two frames of reference: (a) the accelerating observer's own frame and (b) the frame of reference of an inertial observer who sees the accelerating observer move with instantaneous velocity $\vec{v}$. Since the dot product is the same (geometrically an absolute quantity), equating the result in the two frames will give the relationship between the frequency of the photon emitted by the emitter and measured by the inertial observer. Derive this expression. Do not assume anything special about the relative direction of motion of the two observers, or the direction in which the light is emitted.

Exercise 1.7.4. *Geometry of Minkowski Space*: The metric for two-dimensional Minkowski space in inertial coordinates is given by

$$ds^2 = -dt^2 + dx^2 \tag{1.83}$$

Deduce orthogonal properties of tangent vectors $\underset{\sim}{e}_t$ and $\underset{\sim}{e}_x$. Explore the following coordinate systems and their assiociated tangent vectors:
(a) Light-cone coordinates:

$$\begin{aligned} x^+ &= \frac{1}{\sqrt{2}}(t + x) \\ x^- &= \frac{1}{\sqrt{2}}(t - x) \end{aligned} \tag{1.84}$$

(b) Coordinates ρ, τ used by a uniformly accelerated observer:

$$\begin{aligned} t &= \rho \sinh \tau \\ x &= \rho \cosh \tau \end{aligned} \tag{1.85}$$

For each of these coordinate systems, plot the associated coordinate curves in the $t - x$ plane, and analyze the orthogonality of the associated tangent vectors with each other, and tangent vectors of the other coordinate systems. Are curves that appear orthogonal in this plane necessarily orthogonal? Is the converse true?

Exercise 1.7.5. *More Minkowski Space*: A few definitions: A vector $\underset{\sim}{u}$ in M.S. is said to be *timelike*, if $\underset{\sim}{u} \cdot \underset{\sim}{u} < 0$, *spacelike* if $\underset{\sim}{u} \cdot \underset{\sim}{u} > 0$ and *lightlike* if $\underset{\sim}{u} \cdot \underset{\sim}{u} = 0$. A timelike vector is said to be *future pointing* if its component along $\underset{\sim}{e}_t = \partial/\partial t$ is positive. In a two-dimensional Minkowski space, sketch examples of lightlike, future-pointing timelike, past-pointing timelike and spacelike vectors at an arbitrary point.

In a four-dimensional Minkowski space, prove that:
(a) A vector orthogonal to a timelike vector has to be spacelike.
(b) A vector orthogonal to a null vector can be spacelike or null, but nor timelike.
(c) If two future pointing lightlike vectors have unit length, then their difference is spacelike.
(d) *Anti-Schwarz Inequality*: Let $\underset{\sim}{u}$ and $\underset{\sim}{v}$ be timelike future pointing vectors. Then

$$\underset{\sim}{u} \cdot \underset{\sim}{v} \leq -\sqrt{-\underset{\sim}{u} \cdot \underset{\sim}{u}}\ \sqrt{-\underset{\sim}{v} \cdot \underset{\sim}{v}} \tag{1.86}$$

(e) *Anti-Triangle Inequality (Twin-Paradox)*: Let $\underset{\sim}{u}$ and $\underset{\sim}{v}$ be timelike future pointing vectors. Let $\underset{\sim}{w} = \underset{\sim}{u} + \underset{\sim}{v}$. Then

$$|\underset{\sim}{w}| \geq |\underset{\sim}{u}| + |\underset{\sim}{v}| \tag{1.87}$$

Chapter 2

Linear Transformations

2.1 Linear Transformations

Definition Let U and V be vector spaces over a field $\mathbb{F}$. A map $f : U \longrightarrow V$ is a *linear transformation*, if
(a) $\forall \left|\alpha\right\rangle \in U$ and $\forall a \in \mathbb{F}$, $f(a\alpha) = af(\alpha)$
(b) $\forall \left|\alpha\right\rangle, \left|\beta\right\rangle \in U$, $f(\alpha+\beta) = f(\alpha) + f(\beta)$

Note: A map $f : U \longrightarrow V$ is *one-to-one* or *injective* if two distinct elements $\left|\alpha_1\right\rangle$ and $\left|\alpha_2\right\rangle$ in U are mapped to distinct elements in V. This is the same as saying that $f(\alpha) = f(\beta) \Rightarrow \left|\alpha\right\rangle = \left|\beta\right\rangle$. The map is *onto* or *surjective* if the *image* of the map (the set of elements $f(\alpha)\ \forall\ \left|\alpha\right\rangle \in U$) is the entire set V. This is the same as saying that for any vector $\left|\psi\right\rangle \in V$, there exists at least one vector $\left|\alpha\right\rangle \in U$ such that $f(\alpha) = \left|\psi\right\rangle$. A map is *bijective* (an *isomorphism*) if it is one-to-one and onto.

Definition Two vector spaces are said to be isomorphic, if there exists a linear isomorphism connecting them.

Theorem 2.1.1. *Any n-dimensional vector space V over a field $\mathbb{F}$ is isomorphic to $\mathbb{F}^n$.*

Proof Let $\{\left|\alpha_1\right\rangle, \left|\alpha_2\right\rangle, .., \left|\alpha_n\right\rangle$ be a basis for V. Any vector $\left|\psi\right\rangle \in V$ has a unique expansion

$$\left|\psi\right\rangle = c_1\left|\alpha_1\right\rangle + c_2\left|\alpha_2\right\rangle + .. + c_n\left|\alpha_n\right\rangle$$

where $c_i \in \mathbb{F}$. Define a transformation $T : V \longrightarrow \mathbb{F}^n$ as

$$T(\psi) = (c_1, c_2, .., c_n)$$

Because of the uniqueness of the expansion, the map is one-to-one. Conversely, let $(b_1, b_2, .., b_n)$ be an element of $\mathbb{F}^n$. Since the basis of V spans V, there must exist a vector $\left|\phi\right\rangle \in V$, such that

$$\left|\phi\right\rangle = b_1\left|\alpha_1\right\rangle + b_2\left|\alpha_2\right\rangle + .. + b_n\left|\alpha_n\right\rangle$$

Then, the map is onto. It is easy to check that the map is linear. Then, it is an isomorphism and the spaces V and $\mathbb{F}^n$ are isomorphic.

Definition Two linear transformations T_1 and T_2 from vector space U to V are said to be equal iff $T_1(\alpha) = T_2(\alpha)\ \forall \left|\alpha\right\rangle \in U$.

Definition The sum $T_1 + T_2$ and scalar multiple cT of linear transformations $U \to V$ are defined by

$$\begin{aligned}(T_1 + T_2)(\alpha) &= T_1(\alpha) + T_2(\alpha) \qquad (2.1)\\ (cT)(\alpha) &= cT(\alpha)\end{aligned}$$

Theorem 2.1.2. *Let U and V be vector spaces over a field $\mathbb{F}$ and let $\mathcal{T}$ be the set of all linear transformations from U to V. Then, the set $\mathcal{T}$ is a vector space, denoted by $\mathcal{L}(U, V)$.*

Proof The proof is trivial. All we need to do is to identify an additive inverse and a null element. The additive inverse of T is clearly $(-1).T$. The null transformation Z, by definition, is one that maps a vector in U to a null vector in V.

Definition *Special transformations:* The *identity* transformation $I : U \to V$ is defined by $I(\alpha) = |\alpha\rangle \, \forall \, |\alpha\rangle$. The *null* transformation $Z : U \to V$ is defined by $Z(\alpha) = |\rangle_0 \in V, \ \forall \, |\alpha\rangle \in U$.

Definition Let U, V, W be vector spaces over a field $\mathbb{F}$. Let $T_1 : U \to V$ and $T_2 : V \to W$ be linear transformations. Then, the product transformation $T_1 \cdot T_2 : U \to W$ is defined as

$$T_1 \cdot T_2(\alpha) = T_1(T_2(\alpha)) \tag{2.2}$$

It is easy to check that $T_1 \cdot T_2$ is linear.

Definition A linear transformation $T : V \to V$ from a vector space into itself is called a *Linear Operator*. We will denote it as $\hat{T}$.

Definition A linear operator $\hat{T}$ is said to be *idempotent*, if $\hat{T}^2 = \hat{T} \cdot \hat{T} = \hat{T}$.

Definition A non-null linear operator $\hat{T}$ is said to be *nilpotent* of degree k if $\hat{T}^k = Z$, but $\hat{T}^{k-1} \neq \hat{Z}$, where $\hat{Z}$ is the null operator.

Let us consider a few examples of linear transformations, in particular, linear operators.

Example Rotation in $\mathbb{R}^2$: $\hat{R} : \mathbb{R}^2 \to \mathbb{R}^2$

$$\hat{R}(x_1, x_2) = (x_1 \cos\theta - x_2 \sin\theta, x_1 \sin\theta + x_2 \cos\theta) \tag{2.3}$$

Note that this linear transformation preserves the norm in $\mathbb{R}^2$.

Example Dilation in $\mathbb{R}^n$: $\hat{D} : \mathbb{R}^n \to \mathbb{R}^n$.

$$\hat{D}(x_1, x_2, .., x_n) = (cx_1, cx_2, ..., cx_n) \qquad c \geq 0 \tag{2.4}$$

Note that this linear transformation does not preserve the norm.

Example Projection $\mathbb{R}^2$: $P : \mathbb{R}^3 \to \mathbb{R}^2$

$$P(x_1, x_2, x_3) = (x_1, x_2) \tag{2.5}$$

Example Let $C^\infty[a, b]$ be the space of infinitely differentiable functions defined over $x \in [a, b]$. Let $a_0(x), a_1(x), .., a_n(x)$ be functions in $C^\infty[a, b]$. Then,$\forall f \in C^\infty[a, b]$, the following defines a linear transformation $T : C^\infty[a, b] \to C^\infty[a, b]$

$$T(f) = a_0(x)f(x) + a_1(x)\frac{df}{dx} + a_2(x)\frac{d^2 f}{dx^2} + ..a_n(x)\frac{d^n f}{dx^n} \tag{2.6}$$

The transformation T is then a differential operator of order n.

Example Let A be an $m \times n$ matrix. The map $T : \mathbb{R}^n \to \mathbb{R}^m$ defined by

$$T(\psi) = A\psi \tag{2.7}$$

where ψ is an n-dimensional column vector $\in \mathbb{R}^n$ is a linear transformation.

Example Define a mapping $T : M_{m\times n} \to M_{n\times m}$ by

$$T(A) = A^T$$

This is a linear transformation.

Exercise 2.1.1. Show that any linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ maps straight lines to straight lines and the origin to itself.

Exercise 2.1.2. Is the following (Translation in $\mathbb{R}^2$ a linear transformation?

$$\hat{T}(x_1, x_2) = (x_1 + a_1, x_2 + a_2) \tag{2.8}$$

where a_1, a_2 are fixed real numbers.

Exercise 2.1.3. Geometrical interpretation of linear transformations: In an arbitrary vector space V over field $\mathbb{F}$, define the null vector as 'origin' and a *ray* $c\,|\alpha\rangle$ as 'straight line', where $|\alpha\rangle \in V$ and $c \in \mathbb{F}$. Show that a linear transformation maps the origin to itself and a straight line to another straight line

Exercise 2.1.4. Consider the following linear transformations on $\mathbb{R}^2$

$$\hat{T}_1(x, y) = (x, 0)$$
$$\hat{T}_2(x, y) = (0, y)$$
$$\hat{T}_3(x, y) = (y, x)$$

Using this example, demonstrate that the product of two non-null transformations can be a null transformation, the order in which linear transformations are composed matters, and there can exist transformations other than identity and null transformations such that $\hat{T}^2 = \hat{T} \cdot \hat{T} = \hat{T}$ (such transformations are called *idempotent.*)

Exercise 2.1.5. Consider the vector space of polynomials of degree less than equal to n, $\mathcal{P}_n(x)$. Show that $\hat{D}$ is a linear operator on this vector space, where

$$\hat{D}P(x) = \frac{dP(x)}{dx}$$

Furthur, show that $\hat{D}^{n+1} = \hat{Z}$, where $\hat{Z}$ is the null operator.

Definition Let $T : U \to V$ be a linear transformation. Then, the set U is said to be the *domain* of T. The *range* or *image* of T is the set of all elements $\mathcal{R}_T = \{|\beta\rangle = T(\alpha) \in V,\ \forall\,|\alpha\rangle \in U\}$

Clearly, $\mathcal{R}_T \subseteq V$.

Theorem 2.1.3. *Let $T : U \to V$ be a linear transformation. The range $\mathcal{R}_T$ is a subspace of V.*

Proof Let $|\psi\rangle, |\phi\rangle \in \mathcal{R}_T$. Then, $\exists\,|\alpha\rangle, |\beta\rangle \in U$, such that $|\psi\rangle = T(\alpha)$ and $|\phi\rangle = T(\beta)$. Then,

$$\begin{aligned} |\psi\rangle + |\phi\rangle &= T(\alpha) + T(\beta) \\ &= T(\alpha + \beta) \end{aligned}$$

Since $|\alpha\rangle + |\beta\rangle \in U$, therefore $|\psi\rangle + |\phi\rangle \in \mathcal{R}_T$. It is easy to show that $c\,|\psi\rangle \in \mathcal{R}_T$ as well.

$\mathcal{R}_T$ is called the *range space* of T.

Definition Let $T : U \to V$ be a linear transformation. Then, the set $\mathcal{N}_T \subseteq U$ of all vectors $|\alpha\rangle \in U$ such that $T(\alpha) = |\rangle_0 \in V$ is said to be the *null space* of T.

Theorem 2.1.4. *Let $T : U \to V$ be a linear transformation. The null space $\mathcal{N}_T$ is a subspace of U.*

Proof Let $|\alpha\rangle, |\beta\rangle \in \mathcal{N}_T$. Then,

$$\begin{aligned} T(\alpha+\beta) &= T(\alpha)+T(\beta) \\ &= |\rangle_0 + |\rangle_0 \\ &= |\rangle_0 \in V \end{aligned}$$

Similarly,

$$\begin{aligned} T(c\alpha) &= cT(\alpha) \\ &= c\,|\rangle_0 \\ &= |\rangle_0 \in V \end{aligned}$$

Exercise 2.1.6. Let $T : U \to V$ be a linear transformation. Prove that the inverse image of a subspace of V is a subspace of U.

Exercise 2.1.7. Describe the image and null spaces of a projection from $\mathbb{R}^3$ to $\mathbb{R}^2$.

Exercise 2.1.8. Define a mapping T defined on the space of polynomials $\mathcal{P}_n$ as follows

$$T(p(x)) = \frac{dp(x)}{dx}$$

Describe the image and null space of T.

Exercise 2.1.9. Let $T : \mathbb{R}^4 \to \mathbb{R}^3$ be defined by

$$T(a,b,c,d) = (a+b, b-c, a+d)$$

(a) Describe the null space and construct a basis for it.
(b) Describe the range space and construct a basis.

Exercise 2.1.10. Let $T : \mathcal{P}_4 \to \mathcal{P}_2$, such that

$$T(p) = \frac{d^2p}{dx^2}$$

Determine the image and null spaces.

Definition The *rank* $\rho(T)$ of a linear transformation T is the dimension of its range space $\mathcal{R}_T$.

Definition The *nullity* $\nu(T)$ of a linear transformation T is the dimension of its null space $\mathcal{N}_T$.

Theorem 2.1.5. *If T is a LT from $U \to V$ and S is a LT from $V \to W$ then*
(a) $\mathcal{R}_{TS} \subseteq \mathcal{R}_S$ and $\rho(TS) \leq \rho(S)$
(b) $\mathcal{N}_T \subseteq \mathcal{N}_{TS}$ and $\nu(TS) \geq \nu(T)$

Proof Easy to see.

Theorem 2.1.6. *Rank-Nullity Theorem: Let $T : U \to V$ be a linear transformation from a vector space U of dimension n to a vector space V of dimension m. Then,*

$$\rho(T) + \nu(T) = n$$

Proof Let $\{|\alpha_1\rangle, |\alpha_2\rangle, .., |\alpha_{\nu(T)}\rangle\}$ be a basis for the null space $\mathcal{N}_T$ of T. Let us extend this to a basis of U by adding $n - \nu_T$ linearly independent vectors $\{|\alpha_{\nu(T)+1}\rangle, .. |\alpha_n\rangle\}$. Let $T(\psi) = |\phi\rangle$, where $|\psi\rangle \in U$ and $|\phi\rangle \in V$. Then, we can expand $|\psi\rangle$ as

$$|\psi\rangle = c_1 |\alpha_1\rangle + .. + c_{\nu(T)} |\alpha_{\nu(T)}\rangle + c_{\nu(T)+1} |\alpha_{\nu(T)+1}\rangle + .. + c_n |\alpha_n\rangle$$

Then,

$$T(\psi) = c_{\nu(T)+1} T(\alpha_{\nu(T)+1}) + .. + c_n T(\alpha_n)$$

since $T(\alpha_1) = .. = T(\alpha_{\nu(T)}) = 0$. Therefore, the set of $n - \nu(T)$ vectors $T(\alpha_{\nu(T)+1}), .., T(\alpha_n)$ span $\mathcal{R}_T$. Furthur, since T is linear and the vectors $|\alpha_{\nu(T)+1}\rangle, .., |\alpha_n\rangle$ are linearly independent (by construction), the vectors $T(\alpha_{\nu(T)+1}), .., T(\alpha_n)$ are LI (prove it!) and therefore form a basis for $\mathcal{R}_T$. Then, the dimension of $\mathcal{R}_T$ is $n - \nu(T)$ and the theorem is proved.

Corollaries of the rank-nullity theorem:

Corollary 2.1.7 (Rank-Nullity Theorem)**.** *Let $T : U \to V$ be a linear transformation from a vector space U of dimension n to a vector space V of dimension m. If $m < n$, then $T(\alpha) = |\rangle_0$ for some $|\alpha\rangle \neq |\rangle_0 \in U$.*

Corollary 2.1.8 (Rank-Nullity Theorem)**.** *Let $T : U \to V$ be a linear transformation from a vector space U of dimension n to a vector space V of dimension m. If $m = n$, then the only vector satisfying $T(\alpha) = |\rangle_0$ is $|\alpha\rangle = |\rangle_0 \in U$.*

Example Consider the system of equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n &= 0 \\ .... &= 0 \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n &= 0 \end{aligned}$$

If $m < n$, it is guaranteed that this system has a non-trivial solution. To see this, we cast his equation in matrix form

$$A\psi = 0$$

where $A_{ij} = a_{ij}$ and

$$\psi = \begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ x_n \end{pmatrix}$$

Define a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ as

$$T(\psi) = A\psi$$

If $m < n$, the dimension of the image space is less than the dimension of the domain space $\mathbb{R}^n$. Then, there exists a non-null vector $\in \mathbb{R}^n$ which is mapped to the null vector in $\mathbb{R}^m$.

Example Consider an inhomogeneous system of equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n &= y_1 \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n &= y_2 \\ .... &= . \\ a_{n1}x_1 + a_{n2}x_2 + ... + a_{nn}x_n &= y_n \end{aligned}$$

If the homogeneous system of equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n &= 0 \\ .... &= 0 \\ a_{n1}x_1 + a_{n2}x_2 + ... + a_{nn}x_n &= 0 \end{aligned}$$

has only the trivial solution $x_1 = x_2 = .. = x_n = 0$, then the inhomogeneous system has a solution for all $y_1, y_2, ..y_n$. To see this, consider the linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$T(\psi) = A\psi$$

where $A_{ij} = a_{ij}$. If the homogenoeus system has only a trivial solution, then the dimension of the null space is zero. Therefore, the dimension of the image space is n, and so the transformation is one-to-one, and therefore there is a unique ψ which is mapped to a given $\phi = (y_1, y_2, .., y_n) \in \mathbb{R}^n$.

Exercise 2.1.11. Let

$$\sum_{i=1}^{n} a_{ij}x_j = y_i : \quad i = 1, 2, ..m$$

be an overdetermined system of linear equations-that is, the number m of equations is greater than the number of unknowns $x_1, x_2, .., x_n$. Assume that this system has a *unique* solution. Show that it is possible to select a subset of n of these these equations which uniquely determine the solution.

Definition A linear transformation $T : U \to V$ is said to be *non-singular* iff $\exists$ a transformation T^{-1} from $\mathcal{R}_T$ onto U such that $T^{-1}T = I$, where I is the identity transformation.

Theorem 2.1.9. *The transformation $T^{-1} : \mathcal{R}_T \to U$ is linear.*

Proof Let $T(\psi) = |\alpha\rangle \in V$ and $T(\phi) = |\beta\rangle \in V$, where $|\phi\rangle, |\psi\rangle \in U$. Then

$$\begin{aligned} T^{-1}(c_1\alpha + c_2\beta) &= T^{-1}(c_1T(\psi) + c_2T(\phi)) \\ &= T^{-1}T(c_1\psi + c_2\phi) \\ &= \left(T^{-1}T\right)(c_1\psi + c_2\phi) \\ &= c_1|\psi\rangle + c_2|\phi\rangle \\ &= c_1T^{-1}(\alpha) + c_2T^{-1}(\beta) \end{aligned}$$

since $T^{-1}T = I$. Then, T^{-1} is linear.

Theorem 2.1.10. *Let $T : U \to V$ be a linear transformation. Then, the following statements are equivalent*
(a) T is non-singular
(b) T is one to one
(c) $\mathcal{N}_T = \{|\rangle_0\}$
(d) $\nu(T) = 0$
(e) $\rho(T) = dim(U)$
(f) T maps any basis of U onto a basis for $\mathcal{R}_T$

Proof (a) $\Rightarrow$ (b):
Let $T(\alpha) = T(\beta)$. Then, $T^{-1}(T(\alpha)) = T^{-1}(T(\beta)) \implies |\alpha\rangle = |\beta\rangle$.
(b) $\Rightarrow$ (c):
Let $|\psi\rangle \in \mathcal{N}_T$. Then, $T(\psi) = |\rangle_0 \in V$. But, since T is linear, $T(0) = |\rangle_0$ (T maps a null vector in U to a null vector in V). Since T is one to one, it follows that $|\psi\rangle = |\rangle_0$.

(c) $\Rightarrow$ (d): obvious.
(d) $\Rightarrow$ (e): obvious.
(e) $\Rightarrow$ (f):
From the rank-nullity theorem, it follows that $\rho(T) = dim(U)$. Let $\{|\alpha_1\rangle, |\alpha_2\rangle, .., |\alpha_n\rangle\}$ be a basis for U. Since T is linear, the set $T(\alpha_1), .., T(\alpha_n)$ is LI. Also, since this set has n vectors (which is equal to the dimension of the range space), it follows that this set is a basis for $\mathcal{R}_T$.
(f)$\Rightarrow$ (a): Let $\{|\alpha_1\rangle, |\alpha_2\rangle, .., |\alpha_n\rangle\}$ be a basis for U. Then, $T(\alpha_1), .., T(\alpha_n)$ is a basis for $\mathcal{R}_T$. Then, any vector $|\psi\rangle \in V$ has a unique expansion

$$\begin{aligned} |\psi\rangle &= \sum_{i=1}^{n} c_i T(\alpha_i) \\ &= T\left(\sum_{i=1}^{n} c_i \alpha_i\right) \end{aligned}$$

since T is linear. Then, $\exists\, |\phi\rangle = \sum_{i=1}^{n} c_i |\alpha_i\rangle \in U$, such that $|\psi\rangle = T(\phi)$. Define $T^{-1}(\psi) = \phi$. Then, $(T^{-1}T)(\phi) = T^{-1}(\psi) = |\phi\rangle$. Therefore, $T^{-1}T = I$.

Theorem 2.1.11. *If T is a linear transformation from $U \to V$ and T^{-1} is a linear transformation from $\mathcal{R}_T \to U$, then*

$$T^{-1}T = I \text{ on } U \iff TT^{-1} = I \text{ on } \mathcal{R}_T$$

Theorem 2.1.12. *The inverse of a linear transformation is unique.*

Proof Simple.

Proof Since $\exists T^{-1}$ such that $T^{-1}T = I$ on U, then any $|\psi\rangle \in \mathcal{R}_T$ can be uniquely represented as $|\psi\rangle = T(\phi)$ where $|\phi\rangle \in U$. Then,

$$\begin{aligned} T(T^{-1}(\psi)) &= T(T^{-1}T(\phi)) \\ &= T(\phi) \\ &= |\psi\rangle \end{aligned}$$

Therefore, $TT^{-1} = I$ on $\mathcal{R}_T$.

Theorem 2.1.13. *Let T be a linear transformation from $U \to V$ and S a linear transformation from $\mathcal{R}_T \to W$. Then, ST is non-singular iff both S and T are non-singular. If ST is non-singular, then $(ST)^{-1} = T^{-1}S^{-1}$.*

Proof Let T and S be non-singular. Then, both T and S are one-to-one. This implies that the product transformation ST is one-to-one. Therefore, it is invertible. Conversely, let ST be invertible. Then, we show that this implies both T and S are one-to-one. Let $|\alpha\rangle, |\beta\rangle \in U$ such that $T(\alpha) = T(\beta)$. Then, $ST(\alpha) = ST(\beta)$. Since ST is invertible, this implies that $|\alpha\rangle = |\beta\rangle$. Therefore, T is one-to-one and so invertible. Similarly, it is easy to show that S is invertible. Next, assuming that ST is non-singular, it is easy to check that $T^{-1}S^{-1}$ is an inverse. Since the inverse is unique, this is the only inverse.

Definition A linear bijection between vector spaces over the same field is called an *Isomorphism*.

Isomorphic vector spaces are 'equivalent'.

Theorem 2.1.14. *Isomorphic vector spaces have the same dimension.*

Proof Let $T : U \to V$ be an isomorphism. Then, T is one-to-one and onto, and therefore invertible. Then, $\mathcal{R}_T = V$ and since $\rho(T) = dim(U)$, it follows that $dim(V) = dim(U)$.

Exercise 2.1.12. Prove that the composition of isomorphisms is an isomorphism.

Theorem 2.1.15. *Let $\hat{T} : V \to V$ be a linear operator. Then, the following statements are equivalent: (a) $\hat{T}$ has a trivial Kernel. (b) $\hat{T}$ is one-to-one. (c) $\hat{T}$ is onto. (d) $\hat{T}$ is an isomorphism.*

Proof Simple, follows from previous theorems.

Example $\mathbb{C}$ is isomorphic to $\mathbb{R}^2$.

Example Consider the set $\mathcal{M}$ of all 2×2 matrices of the form

$$M = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

Define a transformation $T : \mathbb{C} \to \mathcal{M}$ as:

$$T(a+ib) = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

It is easily checked that this is an isomorphism. this is an interesting isomorphism, since the algebra of these matrices replicates the algebra of complex numbers under multiplication.

Example Consider the vector space $L^2(0, a)$ of square integrable complex functions over $x \in [0, a]$. This is an infinite dimensional vector space with basis $\{\sin(2\pi nx/a), \cos(2\pi nx/a), 1\}$ with $n = 1, 2, 3, ...$ We can define an alternative *complex* basis

$$\phi_n(x) = \frac{1}{\sqrt{a}} e^{ik_n x} \;; \quad k_n = \frac{2\pi n}{a}, \; n \in \mathbb{Z} \tag{2.9}$$

A function $f(x) \in L^2(0, 1)$ can be expressed as

$$f(x) = \sum_{n=-\infty}^{\infty} c_n \phi_n(x)$$

Since $f \in L^2$,

$$\| f \|^2 = \int_0^a |f|^2 < \infty$$

Then, it follows from orthonormality of ϕ_n that

$$\| f \|^2 = \sum_{n=-\infty}^{\infty} |c_n|^2 < \infty$$

Define a transformation $T : L^2 \to l^2$ as follows:

$$T(f) = (..c_{-k}, .., c_{-}1, c_0, c_1, .., c_k, ..)$$

This is an isomorphism. Then, spaces L^2 and l^2 are isomorphic.

Exercise 2.1.13. Consider linear transformations $L : \mathcal{P}_3 \to \mathbb{R}^3$ and $T : \mathcal{P}_2 \to \mathbb{R}^3$

$$\begin{aligned} L(f) &= \begin{pmatrix} f(1) \\ f(2) \\ f(3) \end{pmatrix} \\ T(f) &= \begin{pmatrix} f(1) \\ f(2) \\ f(3) \end{pmatrix} \end{aligned}$$

Are these isomorphisms?

Exercise 2.1.14. Consider the transformation $T : l^2 \to l^2$

$$T(x_1, x_2, x_3, ...) = (x_1, x_3, x_5, ...)$$

Is this an isomorphism?

Exercise 2.1.15. Consider the transformation $T : l^2 \to l^2$

$$T(x_1, x_2, x_3, ...) = (0, x_1, x_2, x_3, ...)$$

Is this an isomorphism?

Exercise 2.1.16. Consider the transformation $T : \mathcal{P} \to l^2$ where $\mathcal{P}$ is the space of all polynomials

$$T(f) = (f(0), f(1), f(2), ...)$$

Is this an isomorphism?

Exercise 2.1.17. Define an isomorphism between $\mathcal{P}_3$ and $\mathbb{R}_{2\times2}$ (space of 2×2 real matrices), if you can.

Exercise 2.1.18. For which constants k is the linear transformation an isomorphism?

$$T(M) = \begin{pmatrix} 2 & 3 \\ 0 & 4 \end{pmatrix} M - M \begin{pmatrix} 3 & 0 \\ 0 & k \end{pmatrix}$$

Exercise 2.1.19. For which real numbers $c_0, c_1, ..c_n$ is the following linear transformation an isomorphism from $\mathcal{P}_n$ to $\mathbb{R}^{n+1}$?

$$T(f) = \begin{pmatrix} f(c_0) \\ f(c_1) \\ .. \\ f(c_n) \end{pmatrix}$$

2.2 Matrix representation of linear transformations

We have seen that any vector space V of dimension n over a field $\mathbb{F}$ is isomorphic to the space $\mathbb{F}^n$, with the isomorphism basis dependent. Given an *ordered* basis $|\alpha_1\rangle, |\alpha_2\rangle, .., |\alpha_n\rangle$, any vector $|\psi\rangle \in V$ has a unique representation

$$|\psi\rangle = c_1 |\alpha_1\rangle + c_2 |\alpha_2\rangle + .. + c_n |\alpha_n\rangle \tag{2.10}$$

We can use this representation to identify $|\psi\rangle \in V$ with $\psi \in \mathbb{F}^n$ as

$$\psi = \begin{pmatrix} c_1 \\ c_2 \\ . \\ . \\ c_n \end{pmatrix} \tag{2.11}$$

where ψ is a column vector. This association enables us to represent linear transformations as matrices. Let $T : U \to V$ be a linear transformation. Let $dim(U) = n$ and $dim(V) = m$. Let $|\alpha_1\rangle, |\alpha_2\rangle, .., |\alpha_n\rangle$ be a basis for U and $|\beta_1\rangle, |\beta_2\rangle, .., |\beta_m\rangle$ a basis for V. Then, given $|\psi\rangle \in U$ with representation $|\psi\rangle = a_1 |\alpha_1\rangle + a_2 |\alpha_2\rangle + .. + a_n |\alpha_n\rangle$ and $|\psi\rangle \in V$ with representation $|\phi\rangle = b_1 |\alpha_1\rangle + b_2 |\alpha_2\rangle + .. + b_m |\alpha_m\rangle$ such that $T(\psi) = |\phi\rangle$, we have

$$\begin{aligned} |\phi\rangle &= T(\psi) \\ &= a_1 T(\alpha_1) + a_2 T(\alpha_2) + .. + a_n T(\alpha_n) \\ &= \sum_{i=1}^{n} a_i T(\alpha_i) \end{aligned} \tag{2.12}$$

Since $T(\alpha_i) \in V$, we can express it as a linear superposition of the basis vectors of V. Then,

$$T(\alpha_i) = \sum_{j=1}^{m} T_{ji} \; |\beta_j\rangle \tag{2.13}$$

which defines a set of $m \times n$ expansion coefficients T_{ij}. Substituting (2.13) in (2.12), and expanding $|\phi\rangle$ in the $\{|\beta_i\rangle\}$ basis, we get

$$\begin{aligned} \sum_{j=1}^{m} b_j \, |\beta_j\rangle &= \sum_{i=1}^{n} a_i \sum_{j=1}^{m} T_{ji} \; |\beta_j\rangle \\ &= \sum_{j=1}^{m} \left(\sum_{i=1}^{n} T_{ji} a_i \right) |\beta_j\rangle \end{aligned} \tag{2.14}$$

Since the basis vectors $\{|\beta_i\rangle\}$ are LI, it follows that

$$b_j = \sum_{i=1}^{n} T_{ji} a_i \tag{2.15}$$

This can be expressed as a matrix equation

$$\phi = T\psi \tag{2.16}$$

where ϕ and ψ are the column vector $\mathbb{F}^n$ representations of $|\phi\rangle$ and $|\psi\rangle$ respectively. Let us summarize: A choice of basis in a vector space induces an isopmorphism between the vector space and the space $\mathbb{F}^n$. In this isomorphism, vectors are represented as column vectors and linear transformations from one space to another as matrices. If the linear transfromation takes an n dimensional space into an m dimensional space (each with a choice of basis, not necessarily orthonormal), in the isomorphism, it is represented as an $m \times n$ matrix, with matrix elements depending on both basis, and defined by (2.13).

Of particular interest are linear transformations $\hat{T} : V \to V$ (linear operators). Then, $|\psi\rangle, |\phi\rangle \in V$ and the same basis is used to represent them as elements of $\mathbb{F}^n$. In this case, the transformation matrix is defined as

$$\hat{T} \, |\alpha_i\rangle = \sum_{j=1}^{n} T_{ji} \; |\alpha_j\rangle \tag{2.17}$$

The matrix is $n \times n$, and as before, a transformation $\hat{T} \, |\psi\rangle = |\phi\rangle$ is represented by (2.16).

Example Consider the transformation $T : \mathcal{P}_2 \to \mathcal{P}_1$ given by

$$T(f) = \frac{df}{dx} + \frac{d^2 f}{dx^2}$$

Let us choose the standard bases ($\alpha_1 = 1, \alpha_2 = x, \alpha_3 = x^2$) for $\mathcal{P}_2$ and ($\beta_1 = 1, \beta_2 = x$) for $\mathcal{P}_1$. Then, the action of T on the $\{\alpha\}$ basis is

$$\begin{aligned} T(\alpha_1) &= 0 \\ T(\alpha_2) &= 1 \\ T(\alpha_3) &= 2x + 2 \end{aligned} \tag{2.18}$$

These can be rewritten in the form (2.13)

$$\begin{aligned} T(\alpha_1) &= 0.\beta_1 + 0.\beta_2 \\ T(\alpha_2) &= 1.\beta_1 + 0.\beta_2 \\ T(\alpha_3) &= 2.\beta_1 + 2.\beta_2 \end{aligned} \tag{2.19}$$

from which the transformation matrix can be easily read off. It will be

$$T = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 0 & 2 \end{pmatrix}$$

Exercise 2.2.1. Consider the linear transformation considered above $T : \mathcal{P}_2 \to \mathcal{P}_1$, and its representation in the standard basis. Determine the action of T on the polynomial $f(x) = 2 + 3x + 5x^2$ using the matrix representation.

Exercise 2.2.2. Let V be the span of $\cos x$ and $\sin x$ in C^∞. Consider the transformation

$$T(f) = 3f + 2f' - f''$$

Find the matrix representing this transformation in the given basis. Is this transformation an isomorphism (one-to-one and onto)?

Exercise 2.2.3. Consider the linear transsformation $T : M_{2\times 2} \to M_{2\times 2}$ given by

$$T(M) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} M - M \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

where M is a real 2×2 matrix.
(a) Determine the transformation matrix representing T in the standard basis. (b) Find the image space and kernel for the transformation, and bases for these spaces.

Exercise 2.2.4. Let T be a linear transformation from the space spanned by $\cos t$ and $\sin t$ into itself. In each case, determine the matrix representing the transformation, and determine if it is an isomorphism.
(a) $T(f) = f'$
(b) $T(f) = f'' + af' + bf$
(c) $T(f(t)) = f(t - \pi/2)$
(d) $T(f(t)) = f(t - \theta)$ for arbitrary real θ

Exercise 2.2.5. Consider $T : \mathbb{C} \to \mathbb{C}$ over the field $\mathbb{R}$. In each case, determine the matrix corresponding to T and if T is an isomorphism:
(a) $T(z) = z^*$
(b) $T(z) = (p + iq)\, z$ where p and q are real.

Exercise 2.2.6. Let $T : \mathcal{P}_2 \to \mathcal{P}_2$ be a linear transformation. In each case, determine the matrix corresponding to T in the standard basis and determine if T is an isomorphism:
(a) $T(f(t)) = f(3)$
(b) $T(f(t)) = f(-t)$

Consider a vector space V with finite dimension n. Let us take two sets of bases in this space, $\{|\alpha_i\rangle\}$ and $\{|\beta_i\rangle\}$. A vector in one basis can be expressed as a linear superposition of the vectors of the other basis. Then,

$$|\alpha_i\rangle = \sum_{j=1}^{n} T_{ji}\, |\beta_j\rangle \tag{2.20}$$

where T_{ji} are the expansion coefficients. These coefficients can be arranged as an $n \times n$ matrix T. Since there is a one-to-one correspondence between $n \times n$ matrices and linear operators from V to V, we can define a linear operator $\hat{T}$ such that

$$|\alpha_i\rangle = \hat{T}\, |\beta_i\rangle \tag{2.21}$$

Then, it is clear that T will be the matrix representation of $\hat{T}$ in the $\{|\beta_i\rangle\}$ basis

$$\hat{T}\, |\beta_i\rangle = \sum_{j=1}^{n} T_{ji}\, |\beta_j\rangle \tag{2.22}$$

Then, a change of basis can be visualized as a linear transformation. We can pose the following question: what is the matrix representation of $\hat{T}$ in the $|\alpha_i\rangle$ basis? Let $\hat{T}$ be represented by magtrix S in the $|\alpha_i\rangle$ basis. By definition, this implies that

$$\hat{T}\,|\alpha_i\rangle = \sum_{j=1}^{n} S_{ji}\,|\alpha_j\rangle \tag{2.23}$$

Using (2.20), we can write this as

$$\sum_{j=1}^{n} T_{ji}\;\hat{T}\,|\beta_j\rangle = \sum_{j=1}^{n} S_{ji}\,|\alpha_j\rangle \tag{2.24}$$

But, $\hat{T}\,|\beta_j\rangle = |\alpha_j\rangle$. Therefore, it follows that

$$\sum_{j=1}^{n} T_{ji}\;\;|\alpha_j\rangle = \sum_{j=1}^{n} S_{ji}\,|\alpha_j\rangle \tag{2.25}$$

which immediately gives $S_{ji} = T_{ji}$. That is, the linear transformation from one basis to the other is represented by the *same* matrix in either basis. This matrix is called the *Transfer matrix.* Of course, the inverse transformation from the second basis to the first will be represented by the inverse of the transfer matrix.

Consider now a general linear transformation $\hat{A} : V \to V$. In a given basis $\{|\alpha_i\rangle\}$, say it has a matrix representation A^{α}:

$$\hat{A}\,|\alpha_i\rangle = \sum_{j=1}^{n} A^{\alpha}_{ji}\;\;|\alpha_j\rangle \tag{2.26}$$

The same transformation will have a different matrix representation in another basis $\{|\beta_i\rangle\}$. In this basis, the matrix A^{β} representing it will satisfy:

$$\hat{A}\,|\beta_i\rangle = \sum_{j=1}^{n} A^{\beta}_{ji}\;\;|\beta_j\rangle \tag{2.27}$$

The two bases can be expressed in terms of each other through the transfer matrix (2.20). Substituting (2.20) in (2.26), we get

$$\sum_{j=1}^{n} T_{ji}\;\hat{A}\,|\beta_j\rangle = \sum_{j=1}^{n} A^{\alpha}_{ji}\;\sum_{k=1}^{n} T_{kj}\,|\beta_k\rangle \tag{2.28}$$

Using (2.27), we get

$$\sum_{j=1}^{n} T_{ji}\;\sum_{k=1}^{n} A^{\beta}_{kj}\;\;|\beta_k\rangle = \sum_{j=1}^{n} A^{\alpha}_{ji}\;\sum_{k=1}^{n} T_{kj}\,|\beta_k\rangle \tag{2.29}$$

which can be rearranged into

$$\sum_{k=1}^{n}\left(\sum_{j=1}^{n} T_{ji}\;A^{\beta}_{kj}\right)|\beta_k\rangle = \sum_{k=1}^{n}\left(\sum_{j=1}^{n} A^{\alpha}_{ji}\;T_{kj}\right)|\beta_k\rangle \tag{2.30}$$

Since $\{|\beta_i\rangle\}$ are LI, it follows that

$$\sum_{j=1}^{n} T_{ji}\;A^{\beta}_{kj} = \sum_{j=1}^{n} A^{\alpha}_{ji}\;T_{kj} \tag{2.31}$$

which can be recognized as the matrix equation $A^{\beta}T = TA^{\alpha}$, which we write as

$$A^{\beta} = T^{-1}A^{\alpha}T \tag{2.32}$$

This relates the matrix representations of the linear operator $\hat{A}$ in the two basis, through the transfer matrix. A matrix equation of the form (2.32) is said to be a *Similarity transformation*, and essentially implies that matrices A^α and A^β are 'similar', meaning they are different representations of the same linear transformation.

Note: Given a linear operator $\hat{T} : V \to V$, the column vectors of the matrix representing it in some basis $\{|\alpha_i\rangle\}$ have a simple interpretation: they are just the $\mathbb{R}^n$ representations of the vectors $\hat{T}(\alpha_i)$ in the $\{|\alpha_i\rangle\}$ basis.

Exercise 2.2.7. Let V be the subspace of C^∞ spanned by the functions e^x and e^{-x}, with bases $\mathcal{B}_1 = \{e^x, e^{-x}\}$ and $\mathcal{B}_2 = \{\cosh x, \sinh x\}$. Find the transfer matrix. Now consider the transformation $\hat{T} : V \to V$ defined by

$$T(f) = \frac{df}{dx} + \frac{d^2 f}{dx^2}$$

Determine the matrix representing this transformation in the two bases, and check that they are similar.

Exercise 2.2.8. Let $\hat{T} : \mathbb{R}^2 \to \mathbb{R}^2$ have the matrix representation:

$$A = \begin{pmatrix} 5 & -6 \\ 2 & -2 \end{pmatrix}$$

in the standard basis of $\mathbb{R}^2$. Find its representation in the basis $\beta_1 = (2,1), \beta_2 = (3,2)$.

Exercise 2.2.9. Consider the space $\mathcal{P}_2$ defined over $x \in [-1,1]$. Determine the transfer matrix between the standard basis $\{1, x, x^2\}$ and the orthonormal basis of Legendre polynomials $\{P_0, P_1, P_2\}$. Consider the transformation $\hat{T} : P_2 \to \mathcal{P}_2$ defined by $\hat{T}(f(x)) = f(-x)$. Is this a linear transformation? If so, determine its matrix representation in the two bases, and check that they are similar.

Given a linear operator $\hat{T} : V \to V$, its matrix representation in an orthonormal basis $\{|\alpha_i\rangle\}$ has a particularly simple form. The matrix elements T_{ij} are given by

$$\hat{T}\,|\alpha_i\rangle = \sum_{j=1}^{n} T_{ji}\,|\alpha_j\rangle \tag{2.33}$$

Taking the inner product of both sides with $|\alpha_k\rangle$ and using orthonormality, we get

$$\begin{aligned} T_{ij} &= \langle\alpha_i|\left(\hat{T}\,|\alpha_j\rangle\right) \\ &= \langle\alpha_i|\,\hat{T}\,|\alpha_j\rangle \end{aligned} \tag{2.34}$$

where the notation $\langle\psi|\left(\hat{T}\,|\phi\rangle\right) = \langle\psi|\,\hat{T}\,|\phi\rangle$ is used. In quantum mechanics, the expression $\langle\psi|\,\hat{T}\,|\phi\rangle$ is called 'matrix element of $\hat{T}$ with respect to $|\psi\rangle$ and $|\phi\rangle$'.

2.3 Linear functionals: Dual vectors

Consider linear transformations $f : V \to \mathbb{F}$ where $\mathbb{F}$ is the field over which V is defined. Then,

$$\begin{aligned} f(\alpha + \beta) &= f(\alpha) + f(\beta) \\ f(c\alpha) &= cf(\alpha) \end{aligned} \tag{2.35}$$

$\forall\,|\alpha\rangle \in V$ and $c \in \mathbb{F}$. Such a linear transformation is called a *linear functional.* Let V^* denote the set of all linear functionals defined on V. We can define addition of functionals and multiplication of functionals by elements of $\mathbb{F}$ as follows

$$\begin{aligned} (f+g)(\alpha) &= f(\alpha) + g(\alpha) \\ cf(\alpha) &= cf(\alpha) \end{aligned} \tag{2.36}$$

for all functionals in V^*. Then, it is easy to see that V^* is a vector space. This vector space is called the *dual space to* V.

Theorem 2.3.1. *Let V be a finite dimensional vector space, with dimension n. Then, the dual space V^* is also n-dimensional.*

Proof Let $\{|\alpha_i\rangle\}$ be an orthonormal basis. Then, any vector $|\psi\rangle \in V$ has a unique representation

$$|\psi\rangle = a_1 |\alpha_1\rangle + ... + a_n |\alpha_n\rangle$$

Define a linear functional

$$f(\psi) = c_1 a_1 + c_2 a_2 + .. + c_n a_n$$

for $c_i \in \mathbb{F}$.

It is easy to check that f is a linear functional. Conversely, any linear functional can be defined by such a relationship, since

$$\begin{aligned} f(\psi) &= f(a_1\alpha_1 + ... + a_n\alpha_n) \\ &= a_1 f(\alpha_1) + ... + a_n f(\alpha_n) \\ &= c_1 a_1 + ... + c_n a_n \end{aligned} \tag{2.37}$$

where $c_i = f(\alpha_i)$. Define linear functionals $f_{\alpha_i}, i = 1, 2, ..n$ as follows

$$f_{\alpha_i}(\alpha_j) = \delta_{ij} \tag{2.38}$$

Then, any linear functional can be expressed as

$$f = c_1 f_{\alpha_1} + c_2 f_{\alpha_2} + ... + c_n f_{\alpha_n} \tag{2.39}$$

It is easy to check that the functionals f_{α_i} are linearly independent, and so form a basis for the dual space V^*. The basis functionals $\{f_{\alpha_i}\}$ are said to be the *dual basis* to $\{|\alpha_i\rangle\}$. Then, corresponding to every basis of V is a dual basis of V^*.

Example Consider the space $C(0, 1)$, the space of real continuous functions $f(x)$ over $x \in [0, 1]$. Define a linear functional $f : C(0, 1) \to \mathbb{R}$ as

$$l(f) = f(x_0), \qquad x_0 \in [0, 1]$$

It is easy to check this is a linear functional. Another example would be

$$l(f) = \int_0^1 dx \ f(x)$$

Example Consider the space $C^1(0, 1)$ of differentiable functions over $[0, 1]$. Given a set of real numbers $(a_1, a_2, ...a_n)$, we can define a linear functional

$$l(f) = \sum_{i=1}^{n} a_n \frac{d^n f}{dx^n}$$

Example Cotangent vectors (Covectors): Consider the tangent space at point p of an n-dimensional space. Associated with every coordinate system $\{x^i\}$ is a coordinate basis of tangent vectors, $\underset{\sim}{e}_i = \partial/\partial x^i$. Then, any tangent vector at point p can be expressed as $\underset{\sim}{T} = \sum_i c_i \ \underset{\sim}{e}_i$. Let us define n linear functionals dx^i as

$$dx^i(\underset{\sim}{e}_j) = \delta_{ij} \tag{2.40}$$

Note that dx^i is *not* to be confused with the differential of coordinate x^i. This is just a notation (though the two are related). The set $\{dx^i\}$ forms the dual basis to The coordinate basis $\underset{\sim}{e}_i$. Consider now a function $f(x^1, x^2, .., x^n)$ defined on the space. We can difine a linear functional df at point p associated with this function as follows

$$df(\underset{\sim}{T}) = T^i \frac{\partial f}{\partial x^i} \tag{2.41}$$

where T^i are components of $\underset{\sim}{T}$ in the coordinate basis $\underset{\sim}{e_i} = \partial/\partial x^i$ at point p and the partial derivatives are evaluated at p. It is easy to see that this is a measure of the rate of change of the function along the curve to which $\underset{\sim}{T}$ is tangent. The linear functional df is called the *gradient functional.* It is a recognition of the fact that the gradient of a function is technically not a vector, but a covector. Let us express the functional df in terms of the dual basis

$$df = \sum_i c_i dx^i$$

The expanion coefficients are given by $c_i = f(\underset{\sim}{e_i} = \partial f/\partial x^i$. Then, the expansion for df is

$$df = \sum_i \frac{\partial f}{\partial x^i}\, dx^i \tag{2.42}$$

where all quantities are evaluated at point p.

Example Let $I = [a, b]$ be an interval on the real line, and $x_1, x_2, ..x_n$ be n distinct points in the interval. Let us prove that there exists a set of n numbers $m_1, m_2, ..m_n$, such that the *quadrature formula* holds for all polynomials $p(x)$ of degree less than n

$$\int_a^b dx\ p(x) = m_1 p(x_1) + m_2 p(x_2) + .. + m_n p(x_n)$$

Let us prove this assertion. Let $\mathcal{P}_{n-1}$ be the vector space of all polynomials of degree less than n. This is an n dimensional vector space. Let us define n linear functionals $f_j : \mathcal{P}_{n-1} \to \mathbb{R}$ as: $f_j(p(x)) = p(x_j)$. These are elements of the dual space $\mathcal{P}^*_{n-1}$. Let us show that they are LI. Let

$$c_1 f_1 + c_2 f_2 + ... + c_n f_n = 0$$

Operating on $p(x)$, we get

$$c_1 p(x_1) + c_2 p(x_2) + ... + c_n p(x_n) = 0 \tag{2.43}$$

for all polynomials of degree less than n. Now, define the polynomial $q_k(x)$ as

$$q_k(x) = \Pi_{i \neq k}(x - x_i)$$

$q_k(x)$ is of degree $n-1$, and is zero at all points $x_j; j \neq k$. In particular, $q_k(x_k) \neq 0$. Substituting $p(x) = q_k(x)$ in (2.43), we get $c_k = 0$. Then, the functionals $f_1, .., f_n$ are linearly independent. Since the dual space is n dimensional, these functionals form a basis. The integral over $p(x)$ is a linear functional, so can be expressed as above.

Exercise 2.3.1. Let $\mathcal{P}_2$ be the space of real polynomials $p(x)$ of degree less than or equal to 2. Let x_1, x_2, x_3 be three distinct real numbers. Define linear functionals $f_j(p) = p(x_j)$ for $j = 1, 2, 3$. Show that these are LI and form a basis for $\mathcal{P}^*_2$. find the polynomials to which these are dual, and show that they form a basis for $\mathcal{P}_2$.

In inner product spaces, there is a natural connection between linear functionals and inner product.

Theorem 2.3.2. *For every linear functional f on a finite dimensional vector space V, there exists a unique vector $|\psi_f\rangle$ such that*

$$f(\phi) = \langle \psi_f |\, \phi\rangle \tag{2.44}$$

Proof Choose an orthonormal basis $\{|\alpha_i\rangle\}$. Define a vector $|\psi_f\rangle$ as follows

$$|\psi_f\rangle = \sum_i f^*(\alpha_i) |\alpha_i\rangle \tag{2.45}$$

Then,

$$\langle\psi_f|\, \phi\rangle = \sum_i f(\alpha_i) \langle\alpha_i|\, \psi\rangle \tag{2.46}$$

But,

$$\begin{aligned} f(\phi) &= f(\sum_i \alpha_i \langle\alpha_i|\, \phi\rangle) \\ &= \sum_i \langle\alpha_i|\, \phi\rangle f(\alpha_i) \end{aligned} \tag{2.47}$$

which is the same as $\langle\psi_f|\, \phi\rangle$. To show that $|\psi_f\rangle$ is unique, let there be two vectors $|\psi_f\rangle$ and $\left|\psi_f'\right\rangle$ such that $f(\phi) = \langle\psi_f|\, \phi\rangle = \left\langle \psi_f' \middle|\, \phi \right\rangle \; \forall\, |\phi\rangle \in V$. Then, $\left\langle \psi_f - \psi_f' \middle|\, \phi \right\rangle = 0 \; \forall\, |\phi\rangle \in V$. Therefore, from property of inner product, it follows that $|\psi_f\rangle = \left|\psi_f'\right\rangle$.

If the vector space is infinite dimensional, then it is not necessary that for every linear functional we can associate a unique vector in the vector space. However, it can be shown that if the linear functional is *continuous*, then again the association exists.

2.4 Adjoint of a linear transformation

Let $T : U \to V$ be a linear transformation. Let U be n dimensional and V be m dimensional. Then, there exist dual spaces U^* and V^*. The linear transformation T naturally induces a linear transformation $T^* : V^* \to U^*$ as follows: given $f \in V^*$, we can define $T^* f = fT$, where fT is a linear functional in U^* with action

$$\begin{aligned} T^* f(\alpha) &= fT(\alpha) \\ &= f(T(\alpha)) \end{aligned} \tag{2.48}$$

where $|\alpha\rangle \in U$. The map T^* is called the *transpose* of T. Note that the linear transformation $T : U \to V$ does *not* induce a transformation from $U^* \to V^*$, but does so naturally from $V^* \to U^*$. Let us now choose bases in vector spaces U and V, and the corresponding dual bases in U^* and V^*. Let $\{|\alpha_i\rangle$ be a basis for U, with dual basis $\{f_{\alpha_i}\}$ for U^*, and $\{|\psi_i\rangle$ be a basis for V, with dual basis $\{f_{\psi_i}\}$ for U^*. The matrix representation for T with this choice of bases is given by

$$T(\alpha_i) = \sum_{j=1}^{m} T_{ji} |\psi_j\rangle \tag{2.49}$$

where the matrix is $m \times n$. Let us ask: what is the matrix representation of T^* in the dual bases? It will clearly be an $n \times m$ matrix, such that

$$T^*(f_{\psi_i}) = \sum_{j=1}^{n} T^*_{ji} f_{\alpha_j} \tag{2.50}$$

Now, we use the fact that $T^* f = fT$. Then,

$$f_{\psi_i} T = \sum_{j=1}^{n} T^*_{ji} f_{\alpha_j} \tag{2.51}$$

Let us act this functional on a basis vector $|\alpha_k\rangle \in U$. Then,

$$\begin{aligned} f_{\psi_i}T(|\alpha_k\rangle) &= \sum_{j=1}^{n} T^*_{ji} f_{\alpha_j}(|\alpha_k\rangle) \\ \Rightarrow f_{\psi_i}\left(\sum_{j=1}^{n} T_{jk}\,|\alpha_j\rangle\right) &= \sum_{j=1}^{n} T^*_{ji}\delta_{jk} \\ \Rightarrow \sum_{j=1}^{n} T_{jk} f_{\psi_i}(|\alpha_j\rangle) &= T^*_{ki} \\ \Rightarrow \sum_{j=1}^{n} T_{jk}\delta_{ij} &= T^*_{ki} \\ \Rightarrow T_{ik} &= T^*_{ki} \end{aligned} \tag{2.52}$$

Therefore, it follows that the matrix elements of T^* in the dual bases are related to the matrix elements of T (in the corresponding bases) as

$$T^*_{ij} = T_{ji} \tag{2.53}$$

Due to the isomorphism between a vector space and its dual space induced by the inner product, the transpose allows us to construct a linear map $T^\dagger : V \to U$, given the linear transformation $T : U \to V$. This linear map is called the *adjoint* of T. It can be thought of as the map 'dual' to T^*. We define this map as follows: say, we wish to map $|\psi\rangle \in V$ to a suitable vector $|\alpha\rangle \in U$, such that $T^\dagger(\psi) = |\alpha\rangle$. We first identify the dual of $|\psi\rangle \in V^*$, determine its image through T^* in U^*, and define the vector $|\alpha\rangle$ as dual to this, in U. The dual to $|\psi\rangle$ is $f_\psi \in V^*$. Then, its image in U^* is $T^* f_\psi$. Then, if $T^\dagger(\psi) = |\alpha\rangle$, by definition of $T^\dagger$, it follows that given $|\beta\rangle \in U$,

$$T^* f_\psi(\beta) = \langle\alpha|\,\beta\rangle \tag{2.54}$$

But, $T^* f_\psi(\beta) = f_\psi T(\beta) = f_\psi(T\beta) = \langle\psi|\,T\beta\rangle$. Since $|\alpha\rangle = T^\dagger(\psi)$, it follows that

$$\langle\psi|\,T\beta\rangle = \left\langle T^\dagger\psi \middle|\, \beta\right\rangle \tag{2.55}$$

Note that the inner product on the left is defined over V and that on the right is defined on U.

Let us now focus on linear operators $\hat{T} : V \to V$. Then, just as for every vector there is a dual vector, for every operator $\hat{T}$ is an adjoint, defined by the above relation, where now $|\psi\rangle, |\alpha\rangle \in V$. We can write the above relation in another, more practically useful way, where we recall that $\langle\phi|\,\psi\rangle = \langle\psi|\,\phi\rangle^*$, and the notation $\left\langle\phi\middle|\,\hat{T}\psi\right\rangle = \langle\phi|\,\hat{T}\,|\psi\rangle$

$$\langle\psi|\,\hat{T}\,|\alpha\rangle = \langle\alpha|\,\hat{T}^\dagger\,|\psi\rangle^* \tag{2.56}$$

What kind of a matrix is $\hat{T}^\dagger$ represented by? To see this, let us choose an orthonormal basis $\{|\alpha_i\rangle\}$ in V. Then, choosing $|\psi\rangle = |\alpha_i\rangle$ and $|\alpha\rangle = |\alpha_j\rangle$ in the above equation, we find that (after a complex conjugation of both sides)

$$T^\dagger_{ij} = T^*_{ji} \tag{2.57}$$

That is, the matrix $T^\dagger$ is obtained by taking the transpose of T, followed by complex conjugation.

The adjoint of a linear transformation satisfies the following properties (verify them!):

(a) $(A + B)^\dagger = A^\dagger + B^\dagger$
(b) $(AB)^\dagger = B^\dagger A^\dagger$
(c) If A is invertible, then $\left(A^{-1}\right)^\dagger = \left(A^\dagger\right)^{-1}$
(d) $\left(A^\dagger\right)^\dagger = A$

A linear operator $\hat{T} : V \to V$ is said to be *self-adjoint* or *Hermitian*, if $\hat{T}^\dagger = \hat{T}$. In this case, it follows from (2.55) that

$$\left\langle \psi \middle| \hat{T}\phi \right\rangle = \left\langle \hat{T}\psi \middle| \phi \right\rangle \tag{2.58}$$

Equivalently, it follows from (2.56) that

$$\langle \psi | \hat{T} | \alpha \rangle = \langle \alpha | \hat{T} | \psi \rangle^* \tag{2.59}$$

In a finite dimensional vector space, the matrix representing a Hermitian operator $\hat{T}$ will satisfy

$$T^\dagger = T \tag{2.60}$$

Hermitian linear operators are central to quantum physics. All physically measurable quantities such as energy, momentum, angular momentum, etc are represented by Hermitian operators in quantum theory.

Example Consider the differential operator $\hat{L} = -d^2/dx^2$ acting on $L^2(a, b)$, where we assume it is defined over the field $\mathbb{R}$. Let $\mathcal{D}(L)$ be the set of functions in $L^2(a, b)$, such that $\forall f \in \mathcal{D}(L)$, $\hat{D}(f) \in L^2(a, b)$ (these will be twice differentiable functions). The set $\mathcal{D}(L)$ is called the *domain* of $\hat{D}$. In infinite dimensional vector spaces, the domain of a linear operator will in general not be the same as the entire vector space, but will be a subspace of it. Let us check if this operator is Hermitian. Let $f, g \in \mathcal{D}(L)$. Then,

$$\left\langle f \middle| \hat{L}g \right\rangle = \int_a^b dx f(x) \left(-\frac{d^2 g}{dx^2} \right) \tag{2.61}$$

Integrating by parts twice, we get

$$\begin{aligned} \left\langle f \middle| \hat{L}g \right\rangle &= \left(gf' - fg' \right)|_a^b - \int_a^b dx g(x) \left(-\frac{d^2 f}{dx^2} \right) \\ &= \left(gf' - fg' \right)|_a^b + \left\langle g \middle| \hat{L}f \right\rangle \end{aligned} \tag{2.62}$$

In general, $\hat{L}$ is not Hermitian over its entire domain $\mathcal{D}(L)$, because of the boundary terms. However, it will be Hermitian over a subset $\mathcal{H}(L)$ of $\mathcal{D}(L)$, such that $\forall f, g, \in \mathcal{H}(L)$, the boundary term is zero. This imples that the functions in $\mathcal{H}(L)$ should satisfy either Dirichlet or Neumann boundary conditions. That is, either $f(x) = 0$ or $f'(x) = 0$ at $x = a,\ b$.

Exercise 2.4.1. Consider the space $L^2(-1, 1)$ with the following inner product

$$\langle f | \, g \rangle = \int_{-1}^1 dx \; e^{2x} f(x) g(x)$$

Let $\hat{L}$ be an operator defined by

$$\hat{L}(f) = -f'' - 2f'$$

over a suitable domain. What boundary conditions must f satisfy so that $\hat{L}$ is Hermitian?

Exercise 2.4.2. Show that the set of $n \times n$ Hermitian matrices form a vector space. Note that Hermitian matrices in general have complex entries. Over what field is this vector space defined? What is the dimension of this vector space?

Consider the vector space of 2×2 Hermitian matrices defined over field $\mathbb{R}$. Show that the following matrices form an orthonormal basis (with inner product $\langle A | \, B \rangle = (1/2) Tr(A^\dagger B)$:

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

The matrices $\sigma_i;\ i = 1, 2, 3$ are known as *Pauli sigma matrices.* Verify a few interesting properties of these matrices:

(a) $\sigma_i^\dagger = \sigma_i$: Hermitian
(b) $\sigma_i \sigma_i = I$: Unitary
(c) $\sigma_i \sigma_j + \sigma_j \sigma_i = 2\delta_{ij} I$
(d) $[\sigma_i, \sigma_j] = 2i \sum_k \epsilon_{ijk} \sigma_k$
where $[\sigma_i, \sigma_j] = \sigma_i \sigma_j - \sigma_j \sigma_i$ is the *commutator* of σ_i and σ_j, ϵ_{ijk} is a completely antisymmetric set of numbers, that is, it changes sign on interchanging any two indices, and $\epsilon_{123} = 1$. Determine all the 'components' of ϵ_{ijk}, and the explicit commutators $[\sigma_i, \sigma_j]$ for all pairs of σ matrices.

2.5 Unitary transformations

A linear operator $\hat{U} : v \to V$ is said to be *unitary*, if (a) it is invertible and (b) it preserves the norm

$$\| \hat{U}\psi \| = \| \psi \| \quad \forall \left|\psi\right\rangle \in V \tag{2.63}$$

If the vector space is finite, then it is enough to demand that the operator preserves the norm. It follows from this that the operator has an inverse, since its null space is trivial (prove it!). In an infinite dimensional vector space, though, it need not follow that it has an inverse, if it preserves the norm. As an example, consider the space l^2. Given $\psi = (x_1, x_2, ...x_k...)$ such that $\sum_k |x_k|^2 < \infty$, define a linear operator $\hat{U}$ as follows

$$\hat{U}(x_1, x_2, .., x_k, ..) = (0, x_1, x_2, .., x_k, ...) \tag{2.64}$$

Clearly, this preserves norms, but is not invertible.

Theorem 2.5.1. *If $\hat{U}$ is a unitary operator defined over an inner product space, it follows that*

$$\left\langle \hat{U}\psi \middle| \hat{U}\phi \right\rangle = \left\langle \psi \middle| \phi \right\rangle \tag{2.65}$$

That is, a unitary transformation preserves inner products.

Proof Outline of the proof: Let $|\alpha\rangle = |\psi\rangle + |\phi\rangle$. Then,

$$\begin{aligned} \| \alpha \|^2 &= \langle \alpha | \alpha \rangle \\ &= \| \psi \|^2 + \| \phi \|^2 + 2Re \langle \psi | \phi \rangle \end{aligned} \tag{2.66}$$

and

$$\| \hat{U}\alpha \|^2 = \| \hat{U}\psi \|^2 + \| \hat{U}\phi \|^2 + 2Re \left\langle \hat{U}\psi \middle| \hat{U}\phi \right\rangle \tag{2.67}$$

from which it follows that $Re \langle \psi | \phi \rangle = Re \left\langle \hat{U}\psi \middle| \hat{U}\phi \right\rangle$. Now, taking $|\alpha\rangle = |\psi\rangle - i\,|\phi\rangle$ gives $Im \langle \psi | \phi \rangle = Im \left\langle \hat{U}\psi \middle| \hat{U}\phi \right\rangle$, which completes the proof.

We had observed that a change of basis in a vector space can be viewed as a linear transformation (see (2.20) and (2.21)). What happens if the two bases are orthonormal?

Theorem 2.5.2. *A unitary transformation maps an orthonormal basis to another orthonormal basis.*

Proof Simple, at least for finite dimensional vector spaces.

Theorem 2.5.3. *Let $\{|\alpha_i\rangle\}$ be an orthonormal basis for a finite dimensional vector space. If the set $\{\hat{U}\,|\alpha_i\rangle\}$ is an orthonormal basis, then $\hat{U}$ is unitary.*

Proof Prove it!

Theorem 2.5.4. *A linear operator $\hat{U}$ is unitary iff*

$$\hat{U}^\dagger \hat{U} = \hat{U}\hat{U}^\dagger = \hat{I} \tag{2.68}$$

That is, its adjoint is its inverse.

Proof Let $\hat{U}^\dagger\hat{U} = \hat{I}$. Then,

$$\begin{aligned} \| \hat{U}\psi \| &= \left\langle \hat{U}\psi \middle| \hat{U}\psi \right\rangle \\ &= \left\langle \psi \middle| \hat{U}^\dagger\hat{U}\psi \right\rangle \\ &= \langle \psi | \psi \rangle \\ &= \| \psi \|^2 \end{aligned} \tag{2.69}$$

Conversely, let $\hat{U}$ be unitary. Then, $\left\langle \hat{U}\psi \middle| \hat{U}\phi \right\rangle = \langle \psi | \phi \rangle \;\; \forall \, |\psi\rangle, |\phi\rangle$. Then,

$$\begin{aligned} \left\langle \psi \middle| \hat{U}^\dagger\hat{U}\phi \right\rangle &= \left\langle \hat{U}\psi \middle| \hat{U}\phi \right\rangle \\ &= \langle \psi | \phi \rangle \end{aligned} \tag{2.70}$$

for all $|\psi\rangle, |\phi\rangle$. Then, $\hat{U}^\dagger\hat{U} = \hat{I}$.

Given a unitary operator $\hat{U}$, its matrix representation U in a given basis will satisfy a matrix equation corresponding to (2.68)

$$U^\dagger U = UU^\dagger = I \tag{2.71}$$

Example A rotation in three dimensional Euclidean space is a unitary transformation.

Example Consider the space $L^2[0,1]$ with orthonormal basis $\alpha_n = e^{i2\pi nx}$. The linear tranformation

$$\hat{U}(f(x)) = e^{i\omega x} f(x) \tag{2.72}$$

is a unitary transformation (check it!). The set of vectors $\beta_n = e^{i\omega x} e^{i2\pi nx}$ is an orthonormal basis. In general, consider the transformation

$$\hat{U}(f(x)) = m(x) f(x) \tag{2.73}$$

where $m(x)$ is a complex function of x such that $|m(x)| = 1$. It is easy to check that this is a unitary transformation.

Given a linear operator $\hat{A}$ expressed as a matrix in two different bases $\{|\alpha_i\rangle\}$ and $\{|\beta_i\rangle\}$, we have seen that the matrix representations are related through a similarity transformation (2.32)

$$A^\beta = T^{-1} A^\alpha T \tag{2.74}$$

where T is the transfer matrix, representing the linear transformation $|\alpha_i\rangle = \hat{T}\,|\beta_i\rangle$ connecting the bases vectors. If the bases are orthonormal, we have seen that $\hat{T}$ will be a unitary operator. Then, the transfer matrix T will be unitary, with $T^{-1} = T^\dagger$. Then, the similarity transformation now reads

$$A^\beta = T^\dagger A^\alpha T \tag{2.75}$$

2.6 Spectral theory

We motivate this discusssion by stating a physical problem. Say, we have a dynamical system, whose configuration or 'state' at any instant of time is given by a set of n real numbers. For instance, the state of a particle in three dimensions is completely described by specifying three coordinates and three momentum components. We visualize this set of n numbers representing the state of the system as an element of $\mathbb{R}^n$, expressed as a column vector X. A set of dynamical equations determines how this state evolves with time. These equations need not be differential equations. They could be equations over discrete intervals of time. Let us assume that the equations do not explicitly depend on time. Further, assume that they are linear, relating the configuration of the system at a certain instant to the configuration at a later instant linearly. In general, we can cast these equations in the form

$$X(t+T) = AX(t) \tag{2.76}$$

where T is some fixed time interval, and $X(0) = X_0$ is known. A is a matrix containing information about the dynamics of the system. We need to figure out the configuration of the system at arbitrary t, in particular, for large t. Starting at $t = 0$ and recursively using this relationship, we get

$$X(NT) = A^N X(0) \tag{2.77}$$

Then, the problem of determining the configuration of the system at arbitrary $t = NT$ reduces to the problem of determining powers of a matrix. In general, given an $n \times n$ matrix, this is a formidable task, even on a computer, for large enough N. We need a clever way out. this brings us to the idea of *eigenvalues* and *eigenvectors* of a linear operator. To motivate the discussion, say the initial configuration X_0 has a very special relationship with the matrix A, such that

$$AX_0 = aX_0 \tag{2.78}$$

where a is in general a complex number. Then, it follows that

$$\begin{aligned} A^N X_0 &= A^{N-1}.AX_0 \\ &= aA^{N-1}X_0 \\ &= aA^{N-2}.AX_0 \\ &= a^2 A^{N-2} X_0 \\ .. & \quad .... \\ &= a^n X_0 \end{aligned} \tag{2.79}$$

Then, the configuration at time t is given by

$$X(t) = a^{t/T} X(0) \tag{2.80}$$

In general, however, the system could start in some arbitrary state, which will not be related in a special way with the evolution matrix A. How do we analyze the behaviour of such a system? We will see that a deeper analysis of equation (2.78) leads to the answer.

Definition A vector $|\psi\rangle$ is said to be an eigenvector of a linear operator $\hat{A}$ in a vector space, if

$$\hat{A}\,|\psi\rangle = a\,|\psi\rangle \tag{2.81}$$

where $a \in \mathbb{F}$.

Theorem 2.6.1. *A linear operator has at least one eigenvector in a finite dimensional vector space over the field of complex numbers.*

Proof Let the vector space be n dimensional. Let $\hat{A}$ be a linear operator. Start with a non-zero vector $|\psi\rangle$ and construct the following set of $n+1$ vectors: $|\psi\rangle, \hat{A}\,|\psi\rangle, \hat{A}^2\,|\psi\rangle, .., \hat{A}^n\,|\psi\rangle$. Since these vectors cannot be linearly independent, the equation

$$\sum_{m=0}^{n} c_m \hat{A}^m\,|\psi\rangle = 0$$

has a nontrivial solution, not all c_m zero. We write this equation as

$$p(\hat{A})\,|\psi\rangle = 0$$

where p is the polynomial $p(x) = \sum_{m=0}^{n} c_m x^m$. Every polynomial over complex numbers can be written as a product of linear factors

$$p(x) = c \prod_j (x - a_j), \quad c \neq 0$$

Then,

$$c \prod_j (\hat{A} - a_j) |\psi\rangle = 0$$

Then, the operator $\prod_j (\hat{A} - a_j)$ maps a non -zero vector to the null vector. therefore, it is not invertible. Since product of invertible maps is invertible, it follows that at least one of the operators $\hat{A} - a_j \hat{I}$ is not invertibel. Then, this operator has a non-trivial null space (kernel space). Let $|\phi\rangle$ be an element of this null space. Then, it follows that

$$(\hat{A} - a_j \hat{I}) |\phi\rangle = 0$$

from which it follows that there exists at least one non-zero vector $|\phi\rangle$ such that

$$\hat{A} |\phi\rangle = a_j |\phi\rangle$$

How do we determine the eigenvectors and the corresponding eigenvalues of an operator? Here, the matrix representation of linea roperators becomes useful. We choose a basis $|\alpha_i\rangle$ (which need not be orthonormal). In this basis, the eigenvalue equation will be a matrix equation

$$A\phi = a\phi \tag{2.82}$$

which can be written as $(A - aI)\phi = 0$. Since ϕ is not a null vector, the matrix $(A - aI)$ cannot possess an inverse. Therefore, its determinant must be zero. That is,

$$\begin{aligned} P(a) &= |A - aI| \\ &= 0 \end{aligned} \tag{2.83}$$

where $P(a)$ is a polynomial of degree n, called the *characterisic polynomial.* Solving this equation yields n roots a_i, all distinct or some (or all of them) same. Say a_i is a root, and therefore an eigenvalue. The (column vector representation of) associated eigenvector ϕ_i will satisfy

$$A\phi_i = a_i \phi \tag{2.84}$$

where phi_i is to be determined. This relationship will give n equations (not all independent), which will allow us to compute the eigenvector(s) corresponding to this eigenvalue. It should be noted that if $|\psi\rangle$ is an eigenvector of a linear operator $\hat{A}$ with eigenvalue a, so it $c|\psi\rangle$ where $c \in \mathbb{F}$. Therefore, the eigenvectors ϕ will be arbitrary upto a multiplicative factor.

Example Let us find the eigenvalues and eigenvectors of a linear operator with the following matrix representation

$$A = \begin{pmatrix} 3 & 2 \\ 1 & 4 \end{pmatrix}$$

The characteristic polynomial is

$$\begin{aligned} P(a) &= |A - aI| \\ &= a^2 - 7a + 10 \end{aligned}$$

which has roots $a_1 = 2, a_2 = 5$. These are the eigenvalues. The corresponding eigenvectors can be found by solving for $A\phi_i = a_i\phi$ for each eigenvalue. The corresponding eigenvectors are (upto an overall factor)

$$\begin{aligned} \phi_1 &= \begin{pmatrix} 2 \\ -1 \end{pmatrix} \\ \phi_2 &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} \end{aligned}$$

Note that these are LI, and therefore form a basis for this vector space.

Example The Fibonacci sequence $f_0, f_1, \ldots$ is generated by the recurrence relation

$$f_{n+1} = f_n + f_{n-1} \tag{2.85}$$

with starting data $f_0 = 0, f_1 = 1$. This relation can be written as a matrix equation

$$\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} f_{n-1} \\ f_n \end{pmatrix} = \begin{pmatrix} f_n \\ f_{n+1} \end{pmatrix} \tag{2.86}$$

Therefore, we deduce recursively that

$$\begin{pmatrix} f_n \\ f_{n+1} \end{pmatrix} = A^n \begin{pmatrix} f_0 \\ f_1 \end{pmatrix}, \quad A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \tag{2.87}$$

We will try to find a general form for f_n by calculating the eigenvalues and eigenvectors of the matrix A. The eigenvalues are calculated to be

$$a_1 = \frac{1+\sqrt{5}}{2}, \; a_2 = \frac{1-\sqrt{5}}{2}$$

Note that a_1 is positive and greater than 1, whereas a_2 is negative and in absolute value much smaller than 1. The corresponding eigenvectors are computed to be

$$\begin{aligned} \phi_1 &= \begin{pmatrix} 1 \\ a_1 \end{pmatrix} \\ \phi_2 &= \begin{pmatrix} 1 \\ a_2 \end{pmatrix} \end{aligned}$$

These are LI, which allows us to express $(f_0 f_1) = (01)$ as a linear combination of these vectors. This is easily done, to give

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{5}}\,\phi_1 - \frac{1}{\sqrt{5}}\,\phi_2$$

Then, it follows that

$$\begin{aligned} \begin{pmatrix} f_n \\ f_{n+1} \end{pmatrix} &= A^n \begin{pmatrix} f_0 \\ f_1 \end{pmatrix} \\ &= \frac{1}{\sqrt{5}}\,A^n\phi_1 - \frac{1}{\sqrt{5}}\,A^n\phi_2 \\ &= \frac{a_1^n}{\sqrt{5}}\,\phi_1 - \frac{a_2^n}{\sqrt{5}}\,\phi_2 \end{aligned}$$

Since $a_2^n/\sqrt{5}$ is less than 1/2 abd since f_n is an integer, we find that

$$f_n = \text{nearest integer to } \frac{a_1^n}{\sqrt{5}}$$

Let us observe a few important properties of the characteristic polynomial. Though we calculate it using the determinant of a matrix which represents the linear transformation in a particular basis, it is in fact basis independent and therefore characteristic of the linear transformation itself. This of course is just a reflection of the fact that eigenvalues and eigenvectors of a linear operator are an intrinsic properties of the operator. To see that the characteristic polynomial is basis independent, say we compute the matrix form of the linear operator in a different basis, and let this matrix be A'. Then, there exists an invertible transfer matrix T such that the two matrix representations are related through a similarity transformation

$$A' = TAT^{-1}$$

Writing the identity matrix as TT^{-1}, the characteristic polynomial computed using A' will be

$$\begin{aligned} P'(a) &= \left|A' - aI\right| \\ &= \left|TAT^{-1} - aTT^{-1}\right| \\ &= \left|T(A - aI)T^{-1}\right| \\ &= \left|T^{-1}T(A - aI)\right| \\ &= \left|(A - aI)\right| \\ &= P(a) \end{aligned} \tag{2.88}$$

where we have used the property of the determinant that $|AB| = |BA| = |A|\,|B|$. Next, we write $P(a)$ as

$$P(a) = b_0 a^n + b_1 a^{n-1} + .. + b_n \tag{2.89}$$

Putting $a = 0$, we get

$$|A| = b_n \tag{2.90}$$

Further, $b_0 = (-1)^n, b_1 = (-1)^{n-1}(A_{11} + A_{22} + ...A_{nn}) = Tr(A)$ which tells us that the trace of a linear transformation is an intrinsic property of the transformation, since the characteristic polynomial is basis independent. Next, let us write $P(a)$ in terms of the eigenvalues

$$\begin{aligned} P(a) &= (a_1 - a)(a_2 - a)...(a_n - a) \\ &= (-1)^n[(a - a_1)(a - a_2)...(a - a_n)] \\ &= b_0 a^n + b_1 a^{n-1} + .. + b_n \end{aligned} \tag{2.91}$$

Comparing the a^{n-1} term, we get

$$b_1 = (-1)^{n-1}(a_1 + a_2 + ... + a_n) \tag{2.92}$$

which gives

$$Tr(A) = a_1 + a_2 + ... + a_n \tag{2.93}$$

That is, the trace of a linear transformation equals the sum of its eigenvalues. Further, $b_n = a_1.a_2...a_n$ which tells us

$$|A| = a_1.a_2....a_n \tag{2.94}$$

The determinant of a linear transformation equals the product of its eigenvalues.

Theorem 2.6.2. *The eigenvectors of a linear operator corresponding to distinct eigenvalues are linearly independent.*

Proof Let $|a_1\rangle, |a_2\rangle, ... |a_n\rangle$ be eigenvectors of $\hat{A}$, such that $\hat{A}|a_i\rangle = a_i |a_i\rangle$. Let

$$c_1 |a_1\rangle + c_2 |a_2\rangle = 0$$

Operating with $\hat{A}$, we get

$$c_1 a_1 |a_1\rangle + c_2 a_2 |a_2\rangle = 0$$

These equations together give

$$c_2(a_2 - a_1)|a_2\rangle = 0$$

Since $a_1 \neq a_2, \Rightarrow c_2 = 0$ from which it also follows that $c_1 = 0$. Then, $|a_1\rangle$ and $|a_2\rangle$ are LI. Now, we take

$$c_1 a_1 |a_1\rangle + c_2 a_2 |a_2\rangle + c_3 |a_3\rangle = 0$$

Proceeding similarly, we get

$$c_1(a_1 - a_3)|a_1\rangle + c_2(a_2 - a_3)|a_2\rangle = 0$$

Since $|a_1\rangle$ and $|a_2\rangle$ are LI, it follows that $c_1 = c_2 = 0$, from where it follows that $c_3 = 0$, and so on, till we see that the entire set is LI.

Theorem 2.6.3. *If a linear operator in an n dimensional vector space has n distinct eigenvalues, then its eigenvectors form a basis for the vector space.*

Theorem 2.6.4. ***Spectral mapping theorem:*** *(a) Let q be a polynomial, $\hat{A}$ a linear operator, and a an eigenvalue of $\hat{A}$. Then $q(a)$ is an eigenvalue of $q(\hat{A})$.*
(b) Every eigenvalue of $q(\hat{A})$ is of the form $q(a)$, where a is an eigenvalue of $\hat{A}$.

Proof (a) Easy to prove.
(b) Let b be an eigenvalue of $q(\hat{A})$. Then, $q(A) - bI$ is not invertible, where A is the matrix representation of $\hat{A}$ in some basis. We now factor the polynomial $q(x) - b$:

$$q(x) - b = c\prod_i (x - r_i)$$

Replacing x by A:

$$q(A) - bI = c\prod_i (A - r_i I)$$

Since $q(A) - bI$ is not invertible, at least one factor $A - r_i I$ is not invertible. Then, some r_i is an eigenvalue of A. Since r_i is a root of $q(x) - b$,

$$q(r_i) = b$$

which completes the proof.

Theorem 2.6.5. ***Cayley-Hamilton theorem :*** *Every matrix (or the linear operator it represents) satisfies its own characteristic equation:*

$$P(A) = 0$$

where P is the characteristic polynomial for A.

Proof See any standard textbook.

In general, if the characteristic polynomial has multiple roots, the associated operator does not possess n linearly independent eigenvectors, n being the dimension of the vector space. To make up for this, we define *generalized eigenvectors.*

Definition $|\psi\rangle$ is a generalized eigenvector of $\hat{A}$ with eigenvalue a if

$$(\hat{A} - a\hat{I})^m |\psi\rangle = 0 \tag{2.95}$$

for some positive integer m.

Theorem 2.6.6. ***Spectral theorem*** *Let $\hat{A}$ be a linear operator in an n dimensional vector space V over $\mathbb{C}$. Then, any vector in V can be expressed as a sum of eigenvectors of $\hat{A}$, genuine or generalized.*

To see a possible application of the spectral theorem, we return to the question of stability of periodic motion of a dynamical system. The question we wished to address was as follows: if we start with the system in a general configuration, does it return to a configuration in which is executes periodic motion? We could reduce the problem to determining the action of A^N on the initial configuration vector X_0, where A is the matrix representing the dynamics of the system, for large N. For a general vector, this action is not easilty computed. However, we saw that if X_0 is an eigenvector of A, it is easy to determine the eventual dynamical fate of this vector. However, we now see that any general vector X_0 can be expressed as a sum of generalized eigenvectors of A. Then, if the action of A^N on a generalized eigenvector is simple, we can determine the evolution of the system at late times. It turns out that this indeed is true. For instance, it can be easily shown (using mathematical induction) that for a generalized eigenvector $|\psi\rangle$ such that $(\hat{A} - a\hat{I})^2 |\psi\rangle = 0$,

$$\hat{A}^N |\psi\rangle = a^N |\psi\rangle + N a^{N-1} |\phi\rangle$$

where $|\phi\rangle = (\hat{A} - a\hat{I}) |\psi\rangle$ is a genuine eigenvector of $\hat{A}$.

Exercise 2.6.1. A stretch of desert in northwestern Mexico is populated mainly by two species of animals: coyotes and roadrunners. The populations $c(t)$ and $r(t)$ satisfy the following relations:

$$c(t+1) = 0.86\ c(t) + 0.08\ r(t)$$

$$r(t+1) = -0.12\ c(t) + 1.14\ r(t)$$

Given that the populations at instant $t = 0$ are c_0 and r_0, determine the populations at instant t. Explore the long term behaviour of the populations for different initial populations.

Exercise 2.6.2. If $\hat{A}$ is an invertible linear operator, show that the eigenvalues of $\hat{A}^{-1}$ are reciprocals of those of $\hat{A}$, and the eigenvectors are the same as those of $\hat{A}$.

Exercise 2.6.3. Let A and B be similar matrices, such that $B = T^{-1}AT$. Show that they have the same eigenvalues, and if X is an eigenvector of A, then $T^{-1}A$ is an eigenvector of B.

Exercise 2.6.4. Find all the eigenvalues and eigenvectors of the following matrices:
(a)

$$A = \begin{pmatrix} 4 & 1 \\ -1 & 2 \end{pmatrix}$$

(b)

$$A = \begin{pmatrix} 1 & 3 & 1 \\ 0 & 2 & 0 \\ 0 & 1 & 4 \end{pmatrix}$$

Hermitian and unitary operators have very special spectral properties.

Theorem 2.6.7. *The eigenvalues of a Hermitian operator are real.*

Proof Let $\hat{H}$ be Hermitian. Then, $\hat{H}^\dagger = \hat{H}$. Let $|\psi\rangle$ be an eigenvector of $\hat{H}$ with eigenvalue a. Then,

$$\hat{H}\,|\psi\rangle = a\,|\psi\rangle \tag{2.96}$$

Taking inner product with $|\psi\rangle$ gives

$$\langle\psi|\,\hat{H}\,|\psi\rangle = a\,\langle\psi|\,\psi\rangle \tag{2.97}$$

Taking the complex conjugate, we get

$$\begin{aligned} \langle\psi|\,\hat{H}\,|\psi\rangle^* &= a^*\,\langle\psi|\,\psi\rangle * \\ \Rightarrow \langle\psi|\,\hat{H}^\dagger\,|\psi\rangle &= a^*\,\langle\psi|\,\psi\rangle \\ \Rightarrow \langle\psi|\,\hat{H}\,|\psi\rangle &= a^*\,\langle\psi|\,\psi\rangle \\ \Rightarrow a\,\langle\psi|\,\psi\rangle &= a^*\,\langle\psi|\,\psi\rangle \end{aligned}$$

from which it follows that since $\langle\psi|\,\psi\rangle \neq 0$, therefore $a = a^*$.

Theorem 2.6.8. *Eigenvectors of a Hermitian operator corresponding to distinct eigenvalues are orthogonal.*

Proof Let $\hat{H}\,|a_1\rangle = a_1\,|a_1\rangle$ and $\hat{H}\,|a_2\rangle = a_1\,|a_2\rangle$, with $a_1 \neq a_2$. Taking the inner product of the first equation with $|a_2\rangle$ and the second with $|a_1\rangle$ gives the following relations

$$\begin{aligned} \langle a_2|\,\hat{H}\,|a_1\rangle &= a_1\,\langle a_2|\,a_1\rangle \\ \langle a_1|\,\hat{H}\,|a_2\rangle &= a_2\,\langle a_1|\,a_2\rangle \end{aligned}$$

Taking the complex conjugate of the second equation, and subtracting it from the first (remembering that $\hat{H}^\dagger = \hat{H}$ and that $a_2^* = a_2$), we get

$$(a_2 - a_1)\,\langle a_2|\,a_1\rangle = 0 \tag{2.98}$$

Since $a_1 \neq a_2, \Rightarrow \langle a_2|\,a_1\rangle = 0$.

Theorem 2.6.9. *Eigenvalues of a unitary operator are complex numbers of unit modulus.*

Proof Let $\hat{U}^\dagger\hat{U} = \hat{I}$, and $\hat{U}\left|\psi\right\rangle = a\left|\psi\right\rangle$. Taking the inner product of this equation with itself, we get

$$\begin{aligned} \left\langle \hat{U}\psi \middle| \hat{U}\psi \right\rangle &= a^* a \left\langle \psi \middle| \psi \right\rangle \\ \Rightarrow \left\langle \psi \middle| \hat{U}^\dagger\hat{U}\psi \right\rangle &= a^* a \left\langle \psi \middle| \psi \right\rangle \\ \Rightarrow \left\langle \psi \middle| \psi \right\rangle &= a^* a \left\langle \psi \middle| \psi \right\rangle \end{aligned}$$

which can be written as

$$(1 - |a|^2)\left\langle \psi \middle| \psi \right\rangle = 0 \tag{2.99}$$

from which it follows that $|a|^2 = 1$ (since $\left\langle \psi \middle| \psi \right\rangle \neq 0$).

Theorem 2.6.10. *The eigenvalues of a unitary operator for distinct eigenvalues are orthogonal.*

Proof Let $\hat{U}^\dagger\hat{U} = \hat{I}$, and $\hat{U}\left|a_1\right\rangle = a_1\left|a_1\right\rangle, \hat{U}\left|a_2\right\rangle = a_2\left|a_2\right\rangle$. Taking the inner product of the first equation with $\hat{U}\left|a_2\right\rangle$ and using the definition of adjoint, we get

$$\begin{aligned} \left\langle \hat{U}a_2 \middle| \hat{U}a_1 \right\rangle &= a_2^* a_1 \left\langle a_2 \middle| a_1 \right\rangle \\ \Rightarrow \left\langle a_2 \middle| \hat{U}^\dagger\hat{U}a_1 \right\rangle &= a_2^* a_1 \left\langle a_2 \middle| a_1 \right\rangle \\ \Rightarrow \left\langle a_2 \middle| a_1 \right\rangle &= a_2^* a_1 \left\langle a_2 \middle| a_1 \right\rangle \end{aligned}$$

which gives

$$\begin{aligned} (1 - a_2^* a_1)\left\langle a_2 \middle| a_1 \right\rangle &= 0 \\ \Rightarrow (a_1^* - a_2^*)\left\langle a_2 \middle| a_1 \right\rangle &= 0 \quad \text{(multiplying with } a_1^* \text{ and using } a_1^* a_1 = 1) \end{aligned}$$

which, if $a_1 \neq a_2$, gives $\left\langle a_2 \middle| a_1 \right\rangle = 0$.

2.7 Eigenvalues and eigenfunctions of Laplacian operator

The Laplacian operator is defined over the vector space $L^2(M)$, which is the space of square integrable functions defined over a region of space M in d dimensions. For instance, in three dimensions, it is some region M enclosed by a boundary ∂M, with $\hat{\mathbf{n}}$ the outward pointing normal to the surface ∂M

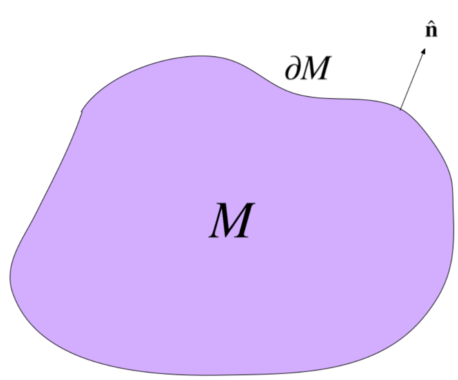

Figure 2.1: Region M with boundary ∂M

The elements of $L^2(M)$ are functions f defined on M such that

$$\int_M dv\, |f|^2 < \infty \tag{2.100}$$

which is a volume integral over M. For instance, if M is a one dimensional line segment $x \in [a, b]$, then the vector space is just $L^2(a, b)$. The Laplacian operator is $\hat{L} = -\nabla^2$, and acts on functions in L^2 which are in its domain. If M is a three dimensional region, the action of $\hat{L}$ in Cartesian coordinates is given by

$$\hat{L}(f) = -\left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}\right) \tag{2.101}$$

Clearly, the domain of $\hat{L}$ is the space $C^2(M)$, the set of twice differentiable functions on M. We wish to explore under what conditions the Laplacian operator is Hermitian. The Hermiticity condition for $\hat{L}$ is

$$\left\langle \hat{L}f \middle| g \right\rangle = \left\langle f \middle| \hat{L}g \right\rangle \tag{2.102}$$

where the inner product on $L^2(M)$ is defined as

$$\langle f|\, g\rangle = \int_M dv\ f^* g \tag{2.103}$$

Then, the Hermiticity condition is

$$\int_M dv\ \nabla^2 f^*\ g = \int_M dv\ f^*\ \nabla^2 g \tag{2.104}$$

Twice integration by parts of the left hand expression gives

$$\begin{aligned} \int_M dv\ \nabla^2 f^*\ g &= \int_{\partial M} \left(g\vec{\nabla} f^* - f^* \vec{\nabla} g\right) \cdot \hat{\mathbf{n}} + \int_M dv\ f^*\ \nabla^2 g \\ &= \int_{\partial M} \left(g\frac{\partial f^*}{\partial n} - f^* \frac{\partial g}{\partial n}\right) + \int_M dv\ f^*\ \nabla^2 g \end{aligned} \tag{2.105}$$

where $\partial f/\partial n$ is the normal derivative of function f evaluated at the boundary. Then, for operator $\hat{L}$ to be Hermitian on a set of functions, the functions belonging to the set should be such that either the function vanishes at the boundary (Dirichlet b.c.) or its normal derivative vanishes at the boundary (Neumann b.c.). It is also possible for the function to vanish on a part of the boundary, and its normal derivative to vanish on the rest of it (Mixed b.c.).

One dimensional Laplacian:

Consider $\hat{L} = -d^2/dx^2$ acting on $L^2(0, a)$ (it is straightforward to generalize the result to $L^2(a, b)$). The boundary conditions rendering $\hat{L}$ Hermitian are :
(a) Dirichlet: $f(a) = f(0) = 0$
(b) Neumann: $f'(a) = f'(0) = 0$
(c) Mixed: $f(a) = 0, f'(0) = 0$ or $f'(a) = 0, f(0) = 0$
Each of these conditions defines a subspace of $C^2(0, a) \subset L^2(0, a)$ over which $\hat{L}$ is Hermitian. We determine the eigenvalues and eigenvector of $\hat{L}$ for each of these subspaces.

Dirichlet: The eigenvalue equation is

$$-\frac{d^2 f}{dx^2} = \lambda\ f \tag{2.106}$$

whose solutions are either exponential functions or oscillatory functions, depending on the sign of λ. It is easy to ceck that for $\lambda \leq 0$, the Dirichlet boundary conditions are satisfied only by the trivial solution $f = 0$ (which implies that only the null vector is an eigenvector). For $\lambda > 0$, the general solution to (2.106) is

$$f(x) = c_1 \cos(\sqrt{\lambda}\ x) + c_2 \sin(\sqrt{\lambda}\ x) \tag{2.107}$$

Imposing Dirichlet conditions $f(a) = f(0) = 0$ gives eigenvalues

$$\lambda_n = \frac{n^2\pi^2}{a^2}; \qquad n = 1, 2, 3, 4, ... \tag{2.108}$$

with corresponding eigenvectors

$$f_n(x) = c_n \sin\left(\frac{n\pi x}{a}\right) \tag{2.109}$$

These eigenvectors are orthogonal, since

$$\begin{aligned}\langle f_n | \, f_m \rangle &= \int_0^a dx \; f_n^*(x) \; f_m(x) \\ &= c_n^* c_m \int_0^a dx \; \sin\left(\frac{n\pi x}{a}\right) \; \sin\left(\frac{m\pi x}{a}\right) \\ &= 0 \;\; \text{if } n \neq m \end{aligned} \tag{2.110}$$

The normalized eigenvectors are

$$f_n(x) = \sqrt{\frac{2}{a}} \; \sin\left(\frac{n\pi x}{a}\right) \tag{2.111}$$

Then, (2.111) are orthonormal

$$\langle f_n | \, f_m \rangle = \delta_{nm} \tag{2.112}$$

which is a verification of the general result for eigenvectors of Hermitian operators.
Neumann: As before, the eigenvalue equation is

$$-\frac{d^2 f}{dx^2} = \lambda \, f \tag{2.113}$$

whose solutions are either exponential functions or oscillatory functions, depending on the sign of λ. It is again easy to check that for $\lambda < 0$, the Neumann boundary conditions are satisfied only by the trivial solution $f = 0$. For $\lambda > 0$, the general solution to (2.106) is once again

$$f(x) = c_1 \cos(\sqrt{\lambda}\; x) + c_2 \sin(\sqrt{\lambda}\; x) \tag{2.114}$$

Then,

$$f'(x) = -\sqrt{\lambda} c_1 \sin(\sqrt{\lambda}\; x) + \sqrt{\lambda} c_2 \cos(\sqrt{\lambda}\; x) \tag{2.115}$$

Imposing Neumann conditions $f'(a) = f'(0) = 0$ gives eigenvalues

$$\lambda_n = \frac{n^2\pi^2}{a^2}; \qquad n = 0, 1, 2, 3, ... \tag{2.116}$$

and eigenvectors

$$f_n(x) = c_n \cos\left(\frac{n\pi x}{a}\right) \tag{2.117}$$

Note that now $n = 0$ is allowed, and gives rise to a constant function, whose derivative vanishes not just at the boundaries, but everywhere. These eigenvectors are orthogonal, since

$$\begin{aligned}\langle f_n | \, f_m \rangle &= \int_0^a dx \; f_n^*(x) \; f_m(x) \\ &= c_n^* c_m \int_0^a dx \; \cos\left(\frac{n\pi x}{a}\right) \; \cos\left(\frac{m\pi x}{a}\right) \\ &= 0 \;\; \text{if } n \neq m \end{aligned} \tag{2.118}$$

The normalized eigenvectors are

$$f_n(x) = \sqrt{\frac{2}{a}} \; \cos\left(\frac{n\pi x}{a}\right) \tag{2.119}$$

Then, (2.119) are orthonormal

$$\langle f_n | \, f_m \rangle = \delta_{nm} \tag{2.120}$$

Mixed I: The boundary conditions are $f(0) = 0, f'(a) = 0$. It is again easy to check that for $\lambda \leq 0$, only the trivial solution will satisfy the eigenvalue equation. For $\lambda > 0$, the solution is

$$f(x) = c_1 \cos(\sqrt{\lambda}\ x) + c_2 \sin(\sqrt{\lambda}\ x) \tag{2.121}$$

Then,

$$f'(x) = -\sqrt{\lambda} c_1 \sin(\sqrt{\lambda}\ x) + \sqrt{\lambda} c_2 \cos(\sqrt{\lambda}\ x) \tag{2.122}$$

Let us now impose the mixed conditions $f(0) =, f'(a) = 0$. $f(0) = 0$ gives $c_1 = 0$. Then,

$$f'(x) = \sqrt{\lambda} c_2 \cos(\sqrt{\lambda}\ x) \tag{2.123}$$

Imposing $f'(a) = 0$ gives

$$\sqrt{\lambda} c_2 \cos(\sqrt{\lambda}\ a) = 0 \tag{2.124}$$

which gives eigenvalues

$$\lambda_n = \frac{(n + 1/2)^2 \pi^2}{a^2}; \qquad n = 0, 1, 2, 3, 4, ... \tag{2.125}$$

with corresponding normalized eigenvectors

$$f_n(x) = \sqrt{\frac{2}{a}} \sin\left(\frac{n\pi x}{a}\right) \tag{2.126}$$

These eigenvectors are orthogonal, since

$$\begin{aligned} \langle f_n | \, f_m \rangle &= \int_0^a dx \ f_n^*(x) \ f_m(x) \\ &= c_n^* c_m \int_0^a dx \ \sin\left(\frac{n\pi x}{a}\right) \ \sin\left(\frac{m\pi x}{a}\right) \\ &= 0 \ \text{ if } n \neq m \end{aligned} \tag{2.127}$$

The normalized eigenvectors are

$$f_n(x) = \sqrt{\frac{2}{a}} \ \sin\left(\frac{(n + 1/2)\pi x}{a}\right) \tag{2.128}$$

As expected, (2.128) are orthonormal.

Another possible vector space over which the one dimensional Laplacian can operate is $L^2(C)$, the space of square-integrable functions defined on a circle. The circle can be simulated as a line segment $x \in [0, 2\pi]$ with the points $x = 0$ and $x = 2\pi$ 'identified'. That is, they are to be considered as the same point. this changes the boundary conditions rendering $\hat{L}$ Hermitian. As before, the Hermiticity condition is that

$$\left(f'^*(x) g(x) - g'(x) f^*(x)\right) \Big|_0^{2\pi} = 0 \tag{2.129}$$

We cannot however impose Dirichlet and Neumann conditions, since there is essentially nothing special about the points $x = 0$ and $x = 2\pi$, since they are on a circle, and could represent any point on it. The circle does not have a boundary! The way to impose these conditions is that $f(0) = f(2\pi)$ and $f'(0) = f'(2\pi)$ which on the circle implies that the functions and their derivatives should be *single valued*, that is, have a unique value at any point on the circle.

As before, negative eigenvalues will yied trivial solutions. For $\lambda \geq 0$, the general solution to the eigenvalue equation is, as before,

$$f(x) = c_1 \cos(\sqrt{\lambda}\ x) + c_2 \sin(\sqrt{\lambda}\ x) \tag{2.130}$$

with

$$f'(x) = -\sqrt{\lambda}c_1 \sin(\sqrt{\lambda}\ x) + \sqrt{\lambda}c_2 \cos(\sqrt{\lambda}\ x) \tag{2.131}$$

Let us now impose the conditions $f(0) = f(2\pi)$ and $f'(0) = f'(2\pi)$. These give two equations

$$\begin{aligned} c_1 &= c_1 \cos(\sqrt{2\pi\lambda}) + c_2 \sin(2\pi\sqrt{\lambda}) \\ c_2 &= -c_1 \sin(2\pi\sqrt{\lambda}) + c_2 \cos(2\pi\sqrt{\lambda}) \end{aligned} \tag{2.132}$$

which give a nontrivial solution only if

$$\begin{vmatrix} 1-\cos(2\pi\sqrt{\lambda}) & -\sin(2\pi\sqrt{\lambda}) \\ \sin(2\pi\sqrt{\lambda}) & 1-\cos(2\pi\sqrt{\lambda}) \end{vmatrix} = 0 \tag{2.133}$$

which gives

$$1 - \cos(2\pi\sqrt{\lambda}) = 0 \tag{2.134}$$

This gives eigenvalues

$$\lambda_n = n^2; \qquad n = 0, 1, 2, 3, ... \tag{2.135}$$

Substituting this in the equations for c_1 and c_2 gives

$$\begin{aligned} c_1 &= c_1 \cos(\sqrt{2\pi n}) + c_2 \sin(2\pi n) \\ c_2 &= -c_1 \sin(2\pi n) + c_2 \cos(2\pi n) \end{aligned} \tag{2.136}$$

which result in trivial relations $c_1 = c_1$ and $c_2 = c_2$. Therefore, the most general eigenvector is of the form

$$f_n(x) = a_n \cos(nx) + b_n \sin(nx) \tag{2.137}$$

Two dimensional Laplacian The two dimensional Laplacian $\hat{L}$ is defined on $L^(M)$m where M is a planar region bound by a curve ∂M. The Hermiticity of $\hat{L}$ and its eigenvectors and eigenvalues will depend on the precise shape of region M. We consider two shapes with symmetry: rectangular and circular. **Rectangular region** A generic rectangular region is illustrated

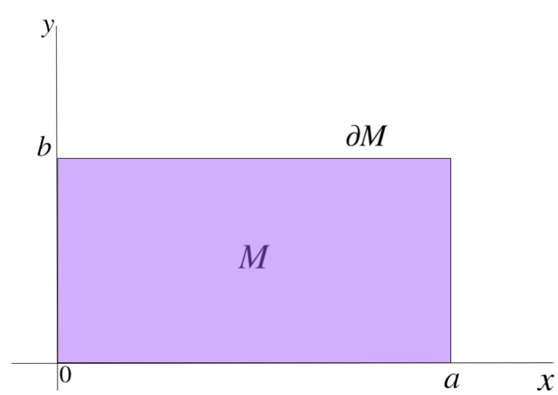

Figure 2.2: Rectangular region M with boundary ∂M

The space $L^2(M)$ is the space of all functions $f(x, y)$ satisfying

$$\int_0^a dx \int_0^b dy\, |f(x,y)|^2 < \infty \tag{2.138}$$

The eigenvalue equation for $\hat{L}$ is

$$-\left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}\right) = \lambda f \tag{2.139}$$

We need to sove this equation subject to Dirichlet, Neumann or mixed boundary conditions. The first step is to use separation of variables. We assume a solution of the form

$$f(x,y) = X(x)Y(y) \tag{2.140}$$

Substituting this in (2.139) and dividing throughout with XY gives

$$-\frac{X''(x)}{X(x)} - \frac{Y''(y)}{Y(y)} = \lambda \tag{2.141}$$

which we rewrite as

$$-\frac{X''(x)}{X(x)} = \frac{Y''(y)}{Y(y)} + \lambda \tag{2.142}$$

Since the left side is a function of x only and the right side a function of y only, the only way this is satisfied is for each to be constant, say γ

$$-\frac{X''(x)}{X(x)} = \frac{Y''(y)}{Y(y)} + \lambda = \gamma \tag{2.143}$$

which gives

$$\begin{aligned} X''(x) &= -\gamma\, X(x) \\ Y''(y) &= -(\lambda - \gamma)\, Y(y) \\ &= -\kappa\, Y(y) \end{aligned} \tag{2.144}$$

where $\kappa = \lambda - \gamma$. Let us now discuss boundary conditions. Say, we impose Dirichlet b.c. on the boundary segments $x = 0, 0 \leq y \leq b$ and $x = a, 0 \leq y \leq b$. This implies that $X(0)Y(y) = 0\ \forall y \in [0,b]$ and $X(a)Y(y) = 0\ \forall y \in [0,b]$. This is only possible if $X(0) = X(a) = 0$. Therefore, this reduces to Dirichlet b.c. on the function $X(x)$. Similarly, applying Neumann b.c to the segments $y = 0, 0 \leq x \leq a$ and $y = b, 0 \leq x \leq a$ implies that

$$\frac{\partial f}{\partial y} = 0; \qquad y = 0,\ 0 \leq x \leq a \text{ and } y = b,\ 0 \leq x \leq a \tag{2.145}$$

which translates to $X(x)Y'(0) = 0$ and $X(x)Y'(b) = 0$. this then implies that $Y'(0) = Y'(b) = 0$, which are just Neumann b.c for the functions $Y(y)$. Since we have already solved for eigenvalues and eigenvectors of one dimensional Laplacian with Dirichlet and Neumann b.c, we find the following normalized eigenfunctions

$$\begin{aligned} X_n(x) &= \sqrt{\frac{2}{a}}\ \sin\left(\frac{n\pi x}{a}\right) \\ \gamma_n &= \frac{n^2\pi^2}{a^2}; \qquad n = 1,2,3,... \end{aligned} \tag{2.146}$$

and

$$\begin{aligned} Y_m(x) &= \sqrt{\frac{2}{b}}\ \cos\left(\frac{m\pi y}{b}\right) \\ \kappa_m &= \frac{m^2\pi^2}{b^2}; \qquad m = 0,1,2,3,... \end{aligned} \tag{2.147}$$

Given that $\lambda = \gamma + \kappa$, we find that the eigenvalues of the two dimensional Laplacian defined over the rectangular region are characterized by two integers n and m, and given by

$$\lambda_{n,m} = \frac{n^2\pi^2}{a^2} + \frac{m^2\pi^2}{b^2}; \quad n = 1,2,3,...; \quad m = 0,1,2,3,... \tag{2.148}$$

The normalized eigenvector $f_{n,m}(x,y)$ corresponding to eigenvalue $\lambda_{n,m}$ is

$$\begin{aligned} f_{n,m}(x,y) &= X_n(x)Y_m(y) \\ &= \frac{2}{\sqrt{ab}} \sin\left(\frac{n\pi x}{a}\right) \cos\left(\frac{m\pi y}{b}\right) \end{aligned} \tag{2.149}$$

Circular region:

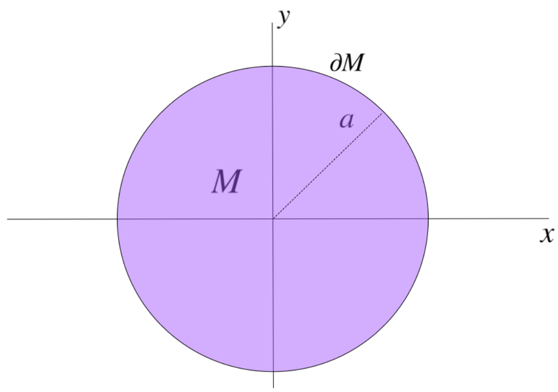

Figure 2.3: Circular region M with boundary ∂M

We express the eigenvalue equation (2.139) in polar coordinates r, θ

$$-\frac{\partial^2 f}{\partial r^2} - \frac{1}{r}\frac{\partial f}{\partial r} - \frac{1}{r^2}\frac{\partial f}{\partial \theta^2} = \lambda f \tag{2.150}$$

Note that in these coordinates, the inner product of two functions $f(r,\theta)$ and $g(r,\theta)$ is given as

$$\langle f|\, g\rangle = \int_0^a dr\ r \int_0^{2\pi} d\theta\ f^*(r,\theta) g(r,\theta) \tag{2.151}$$

and the $L^2(M)$ condition is

$$\int_0^a dr\ r \int_0^{2\pi} d\theta\ \ |f(r,\theta)|^2 < \infty \tag{2.152}$$

We assume a separated solution of the form $f(r,\theta) = R(r)\Theta(\theta)$. Note that with this separation, the $L^2(M)$ condition is equivalent to the following conditions

$$\int_0^a dr\ r\, |R(r)|^2 < \infty \tag{2.153}$$

for the radial function, and

$$\int_0^{2\pi} d\theta\ \ |\Theta(\theta)|^2 < \infty \tag{2.154}$$

for the angular function. The angular function constraint is just the vector space $L^2(C)$, and the radial constraint is the space $L^2(0,a,r)$ where r is a weight function in the inner product

$$\langle R_1|\, R_2\rangle = \int_0^a dr\ r\ R_1^*(r) R_2(r) \tag{2.155}$$

Substituting in (2.150) and dividing throughout by $R(r)\Theta(\theta)$, we get

$$\frac{1}{R(r)}\left(\frac{d^2R}{dr^2} + \frac{1}{r}\frac{dR}{dr}\right) + \frac{1}{r^2}\frac{1}{\Theta(\theta)}\frac{d^2\Theta}{d\theta^2} = -\lambda \tag{2.156}$$

which can be rewritten as

$$\frac{r^2}{R}\left(\frac{d^2R}{dr^2}+\frac{1}{r}\frac{dR}{dr}\right)+\lambda r^2=-\frac{1}{\Theta(\theta)}\frac{d^2\Theta}{d\theta^2} \tag{2.157}$$

since the left side is a function of r and the right side a function of θ, they must individually be constant

$$\frac{r^2}{R}\left(\frac{d^2R}{dr^2}+\frac{1}{r}\frac{dR}{dr}\right)+\lambda r^2=-\frac{1}{\Theta(\theta)}\frac{d^2\Theta}{d\theta^2}=\gamma \text{ (a constant)} \tag{2.158}$$

The θ equation is

$$\frac{d^2\Theta}{d\theta^2}=-\gamma\ \Theta \tag{2.159}$$

which is just the eigenvalue equation for the Laplacian on a circle, which we have solved for. The eigenvalues are

$$\gamma_n=n^2;\qquad n=0,1,2,3,.... \tag{2.160}$$

and eigenfunctions

$$\Theta_n(\theta)=\{\cos(n\theta),\sin(n\theta)\} \tag{2.161}$$

With $\gamma=n^2$, the radial equation is

$$\frac{r^2}{R}\left(\frac{d^2R}{dr^2}+\frac{1}{r}\frac{dR}{dr}\right)+\lambda_n r^2=n^2;\quad n=0,1,2,3,.... \tag{2.162}$$

which we rewrite as

$$-\left(\frac{d^2R}{dr^2}+\frac{1}{r}\frac{dR}{dr}\right)+\frac{n^2}{r^2}R=\lambda_n R \tag{2.163}$$

which is an eigenvalue equation for the differential operator $\hat{D}_n$

$$\hat{D}_n=-\frac{d^2}{dr^2}-\frac{1}{r}\frac{d}{dr}+\frac{n^2}{r^2} \tag{2.164}$$

such that

$$\hat{D}_n(R)=\lambda_n\ R \tag{2.165}$$

We now determine the boundary conditions under which operator $\hat{D}_n$ is Hermitian. $\hat{D}$ acts on the vector space $L^2(0,a,r)$ (actually, it acts on a subspace which is in its domain), which has inner product defined by (2.155). Then, $\hat{D}_n$ will be Hermitian on a set of functions which satisfy

$$\left\langle \hat{D}_n f \middle|\ g\right\rangle=\left\langle f\middle|\ \hat{D}_n g\right\rangle \tag{2.166}$$

Performing integration by parts, it is easy to check that

$$\left\langle \hat{D}_n f \middle|\ g\right\rangle=r\left(f^*\frac{dg}{dr}-g\frac{df^*}{dr}\right)\Bigg|_0^a+\left\langle f\middle|\ \hat{D}_n g\right\rangle \tag{2.167}$$

For $\hat{D}_n$ to be Hermitian, the condition at $r=0$ is that $\lim_{r\to 0} f(r)$ should exists and be finite. At $r=a$, the condition $f(a)=0$ ensures Hermiticity. with this boundary condition, the operator $\hat{D}_n$ is called a *Bessel Operator* for reasons which will become clear. Then, we determine the eigenvalues and eigenvectors of $\hat{D}_n$ with these boundary conditions. The eigenvalue equation can be writtes as

$$r^2\frac{d^2R}{dr^2}+r\frac{dR}{dr}+(r^2\lambda_n-n^2)R=0 \tag{2.168}$$

Define $x=\sqrt{\lambda_n}r$. Then, we get

$$x^2\frac{d^2R}{dx^2}+x\frac{dR}{dx}+(x^2-n^2)R=0 \tag{2.169}$$

This is *Bessel equation*, with general solution

$$R(x) = c_0 J_n(x) + c_1 Y_n(x) \tag{2.170}$$

where $J_n(s)$ and $Y_n(x)$ are Bessel and Neumann functions respectively of order n. In terms of coordinate r, the solution is

$$R(r) = c_0 J_n(\sqrt{\lambda_n} r) + c_1 Y_n(\sqrt{\lambda_n} r) \tag{2.171}$$

We now impose the boundary condition that $\lim_{r\to 0} R(r)$ should be finite. Since the Neumann functions diverge at $r = 0$, this boundary condition gives $c_1 = 0$. The condition $R(a) = 0$ gives

$$J_n(\sqrt{\lambda_n} a) = 0 \tag{2.172}$$

which determines the eigenvalues λ_n

$$\lambda_{n,m} = \frac{\alpha_{n,m}^2}{a^2}; \quad n = 0, 1, 2, 3, .. \;\; m = 1, 2, 3, 4, ... \tag{2.173}$$

where $\alpha_{n,m}$ is the m^{th} zero of $J_n(x)$

$$J_n(\alpha_{n,m}) = 0 \tag{2.174}$$

Then, the eigenvectors of the Laplacian defined over a circular region are given by

$$f_{n,m}(r, \theta) = \{\cos(n\theta)\; J_n\left(\alpha_{n,m}\; \frac{r}{a}\right),\; \sin(n\theta)\; J_n\left(\alpha_{n,m}\; \frac{r}{a}\right)\} \tag{2.175}$$

and eigenvalues given by (2.173). These are orthogonal to each other. However, they are not normalized. The normalized functions are

$$f_{n,m}(r, \theta) = \{f_{nm}^{(1)}(\theta, \phi)\;,\; f_{nm}^{(2)}(\theta, \phi)\} \tag{2.176}$$

where

$$\begin{aligned} f_{nm}^{(1)}(\theta, \phi) &= \sqrt{\frac{1}{\pi}}\; N_{nm}\; \cos(n\theta)\; J_n\left(\alpha_{n,m}\; \frac{r}{a}\right) \\ f_{nm}^{(2)}(\theta, \phi) &= \sqrt{\frac{1}{\pi}}\; N_{nm}\; \sin(n\theta)\; J_n\left(\alpha_{n,m}\; \frac{r}{a}\right) \end{aligned} \tag{2.177}$$

where N_{nm} are normalization constants such that

$$\int_0^a dr\; r\; J_n\left(\alpha_{n,m}\; \frac{r}{a}\right)\; J_n\left(\alpha_{n,m'}\; \frac{r}{a}\right) = \frac{1}{N_{nm}^2}\; \delta_{mm'} \tag{2.178}$$

Now, it is easy to check that $f_{nm}^{(1)}$ and $f_{nm}^{(2)}$ are orthonormal

$$\begin{aligned} \int_0^a dr\; r \int_0^{2\pi} d\theta\; f_{n'm'}^{(1)}\; f_{nm}^{(2)} &= 0 \\ \int_0^a dr\; r \int_0^{2\pi} d\theta\; f_{nm}^{(1)}\; f_{n'm'}^{(1)} &= \delta_{nn'}\; \delta_{mm'} \\ \int_0^a dr\; r \int_0^{2\pi} d\theta\; f_{nm}^{(2)}\; f_{n'm'}^{(2)} &= \delta_{nn'}\; \delta_{mm'} \end{aligned} \tag{2.179}$$

Laplacian on the surface of a sphere Consider the space of all square integrable functions defined on a sphere. These are functions that satisfy

$$\int_0^{2\pi} d\phi \int_0^{\pi} d\theta \sin\theta\; |f(\theta, \phi)|^2 < \infty \tag{2.180}$$

We consider the action of the Laplacian on these functions. Let the radius of the sphere be unity (the analysis does not really depend on the radius of the sphere). Then, the Laplacian operator in spherical polar coordinates is (without r dependence)

$$\hat{L}_S = -\frac{1}{\sin\theta}\left(\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin\theta}\frac{\partial^2}{\partial\phi^2}\right) \tag{2.181}$$

the eigenvalue equation for $\hat{L}$ becomes

$$\hat{L}_S = -\frac{1}{\sin\theta}\left(\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial f}{\partial\theta}\right) + \frac{1}{\sin\theta}\frac{\partial^2 f}{\partial\phi^2}\right) = \lambda\ f \tag{2.182}$$

We use separation of variables. Let $f(\theta,\phi) = \Theta(\theta)\Phi(\phi)$. Substituting in (2.182) gives

$$\frac{\sin\theta}{\Theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right) + \lambda\sin^2\theta = -\frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2} = \gamma \tag{2.183}$$

where γ is a constnat. The Φ equation is

$$\frac{d^2\Phi}{d\phi^2} = -\gamma\ \Phi \tag{2.184}$$

which is just the $L^2(C)$ eigenvalue equation. The eigenvalues and eigenfunctions are

$$\begin{aligned}\gamma_n &= n^2; \qquad n = 0,1,2,3,.... \\ \Phi_n(\phi) &= \{\cos(n\phi),\ \sin(n\phi)\}\end{aligned} \tag{2.185}$$

Note that the $n = 0$ function is constant, independent of ϕ. We now return to the Θ equation, which is

$$\frac{\sin\theta}{\Theta_n}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta_n}{d\theta}\right) + \lambda\sin^2\theta = n^2 \tag{2.186}$$

Let $x = \cos\theta$. Then, $x \in [-1,1]$. The equation reduces to

$$-\frac{d}{dx}\left[(1-x^2)\frac{d\Theta_n}{dx}\right] + \frac{n^2}{1-x^2}\Theta_n = \lambda\Theta_n \tag{2.187}$$

which is an eigenvalue equation for the differential operator

$$\hat{D}_n = -\frac{d}{dx}\left[(1-x^2)\frac{d}{dx}\right] + \frac{n^2}{1-x^2} \tag{2.188}$$

acting on the vector space $L^2(-1,1)$. Let us now return to our original vector space of functions defined on the surface of the unit sphere. It is easy to see that the functions which are independent of ϕ form a subspace of this vector space. LEt us focus only on this subspace for the moment. Then, we are only looking for ϕ independent functions, and so $n = 0$ in (2.187). Then, the eigenvalue equation reduces to

$$\hat{D}_0(\Theta) = \lambda\Theta \tag{2.189}$$

where

$$\hat{D}_0 = -\frac{d}{dx}\left[(1-x^2)\frac{d}{dx}\right] \tag{2.190}$$

We again determine the boundary conditions which render $\hat{D}_0$ Hermitian. The inner product on $L^2(-1,1)$ is

$$\langle f|\ g\rangle = \int_{-1}^{1} dx\ f^*(x)g(x) \tag{2.191}$$

Then, using integration by parts, we get

$$\begin{aligned}\left\langle \hat{D}_0 f \middle| g \right\rangle &= -\int_{-1}^{1} dx \, \frac{d}{dx}\left[(1-x^2)\frac{df^*}{dx}\right] g(x) \\ &= (1-x^2)\left(f^* g' - g f'^*\right)\Big|_{-1}^{1} + \left\langle f \middle| \hat{D}_0 g \right\rangle \end{aligned} \tag{2.192}$$

The boundary term vanishes for all functions which are finite at $x = \pm 1$. Then, $\hat{D}_0$ is Hermitian when acting on functions which are finite at $x = \pm 1$. With these boundary conditions, we determine the eigenvalues and eigenvectors of $\hat{D}_0$. The eigenvalue equation is

$$\frac{d}{dx}\left[(1-x^2)\frac{d\Theta}{dx}\right] + \lambda\Theta = 0 \tag{2.193}$$

This is just *Legendre equation*. If solved through the Frobenius series method, it will yield two independent solutions. However, the constraint that $\Theta(\pm 1)$ is finite forces the eigenvalues to be $\lambda = l(l+1)$ where $l = 0, 1, 2, ...$ and the corresponding eigenfunctions to be Legendre polynomials.

$$\begin{aligned}\lambda &= l(l+1); \quad l = 0, 1, 2, 3, ... \\ \Theta_l(\theta) &= c_l P_l(\cos\theta)\end{aligned} \tag{2.194}$$

where c_l is a normalization constant. Then, the eigenvectors of the Laplacian operator on a sphere with ϕ independence are Legendre polynomials as functions of $\cos\theta$. In absence of ϕ symmetry, we need to determine the eigenvalues and eigenvectors of the operator (2.188). The operator is once again Hermitian when acting on functions that are finite at $x = \pm 1$. The eigenvalue equation is

$$\frac{d}{dx}\left[(1-x^2)\frac{d\Theta}{dx}\right] + \left(\lambda - \frac{n^2}{1-x^2}\right)\Theta = 0 \tag{2.195}$$

A Frobenius analysis along with the condition that $\Theta(\pm 1)$ is finite gives the solution in terms of *Associated Legendre polynomials* P_l^n

$$\begin{aligned}\lambda &= l(l+1); \quad l = n, n+1, n+2, ... \\ \Theta_l^n &= c_l^n \, P_l^n(\cos\theta)\end{aligned} \tag{2.196}$$

where c_l^n is a normalization constant. The polynomials P_l^n are given in terms of the Legendre polynomials as follows:

$$P_l^n(x) = (-1)^n (1-x^2)^n \frac{d^n}{dx^n} P_l(x) \tag{2.197}$$

Then, finally, eigenvalues and eigenvectors of the Laplacian on a sphere are given as follows

$$\begin{aligned}\lambda_{l,n} &= l(l+1); \quad l = n, n+1, n+2, ...; \quad n = 0, 1, 2, 3, ... \\ f_{l,n}(\theta,\phi) &= \{\cos(n\phi)\ P_l^n(\cos\theta),\ \sin(n\phi)\ P_l^n(\cos\theta)\}\end{aligned} \tag{2.198}$$

If we focus on only those functions on the sphere which are independent of ϕ, then these functions form a subspace of the functions considered. The Laplacian is still Hermitian, with eigenvalues and eigenvectors determined from those above by setting $n = 0$

$$\begin{aligned}\lambda_l &= l(l+1); \quad l = 0, 1, 2, ... \\ f_l(\theta) &= P_l(\cos\theta)\end{aligned} \tag{2.199}$$

The *normalized* eigenvectors are

$$f_l(\theta) = \sqrt{\frac{2l+1}{2}}\ P_l(\cos\theta) \tag{2.200}$$

2.8 Application to differential equations

We now use the following result to solve linear partial differential equations: eigenvectors of a Hermitian operator defined on a vector space form a basis for the vector space. Te operator we will consider is the Laplacian operator. We could in priunciple choose any other Hermitian operator, but the choice of the Laplacian is motivated by its presence in the differential equations that will be considered. Choosing eigenvecors of the Laplacian as a basis will make solving the differential equations a trivially simple task. We will essentially focus on two differential equations: the homogeneous wave and heat-conduction equations

$$\nabla^2\phi - \frac{1}{v^2}\frac{\partial^\phi}{\partial t^2} = 0 \quad \text{Wave equation} \tag{2.201}$$

$$\frac{\partial \phi}{\partial t} - D\nabla^2\phi = 0 \quad \text{Heat conduction equation for temperature distribution } \phi \tag{2.202}$$

The key idea is that solutions to these equations with Dirichlet/Neumann/Mixed boundary conditions form a vector space, action of the Laplacian on which is Hermitian. For other boundary conditions, we will see that it is equally simple to aplly these techniques to exptract general solutions.

One dimensional problems

Consider a string vibrating such that one end of the string is tied, and the other end is connected to a (effectively 'massless') ring, which is free to slide along the vertical direction.

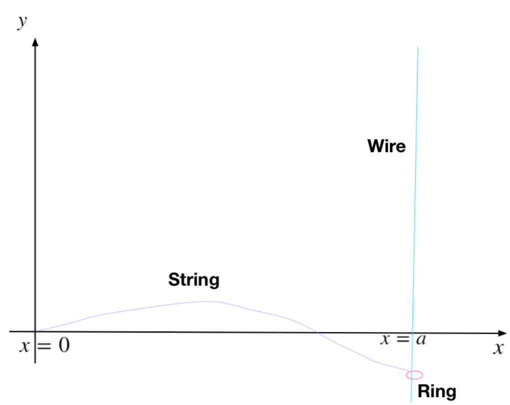

Figure 2.4: Vibrating string with Dirichlet b.c. at one end and Neumann b.c. at other end.

The vibrating string satisfies the wave equation with Dirichlet b.c. at one end (at $x = 0$) and Neumann b.c. at the other end (at $x = a$)

$$\frac{\partial^2 y}{\partial x^2} - \frac{1}{v^2}\frac{\partial^2 y}{\partial t^2} = 0; \quad y(0) = 0, \ \left.\frac{\partial y}{\partial x}\right|_{x=a} = 0 \tag{2.203}$$

The Neumann b.c. arises as follows: the y-component of the forcr acting on the ring due to the string tension is given by

$$F_y = T\sin\theta \tag{2.204}$$

where $\tan\theta$ is the slope of the tangent to the string near $x = a$. In the limit of small displacement, $\sin\theta \sim \tan\theta = \partial y/\partial x$. Since the mass of the ring is negligible, this force must be negligible, else it would produce a very large, unphysical acceleration. Then, $\partial y/\partial x = 0$ at the location of the ring ($x = a$). We now exploit the fact that at any instant, the diplacement of the string is a function $y(x, t)$ of x such that this function vanishes at $x = 0$ and its derivative vanishes at $x = a$. The set of all such functions forms a subspace of the vector space $L^2(0, a)$, such that the action of the Laplacian $\hat{L} = -d^2/dx^2$ is Hermitian over these functions. We have solved for the eigenvalues and eigenvectors of the Laplacian over this subspace of $L^2(0,a)$ ((2.125) and (2.128))

$$\begin{aligned} \lambda_n &= \frac{(n+1/2)^2\pi^2}{a^2}; \qquad n = 0,1,2,3,4,... \text{ Eigenvalues} \\ f_n(x) &= \sqrt{\frac{2}{a}}\,\sin\left(\frac{(n+1/2)\pi x}{a}\right) \qquad \text{Eigenvectors} \end{aligned} \tag{2.205}$$

Then, we can expand the function $y(x,t)$ at any instant in a series of these eigenvectors

$$y(x,t) = \sum_{n=0}^{\infty} a_n(t)\; f_n(x) \tag{2.206}$$

Substituting in the wave equation, we get

$$\sum_{n=0}^{\infty} a_n(t)\;\left(-\frac{d^2}{dx^2}\right) f_n(x) + \sum_{n=0}^{\infty} \frac{1}{v^2}\frac{d^2 a_n(t)}{dt^2}\; f_n(x) = 0 \tag{2.207}$$

Using the fact that f_n are eigenvectors of $\hat{L} = -d^2/dx^2$ with eigenvalues λ_n, we get

$$\begin{aligned} \sum_{n=0}^{\infty} \lambda_n\; a_n(t)\; f_n(x) + \sum_{n=0}^{\infty} \frac{1}{v^2}\frac{d^2 a_n(t)}{dt^2}\; f_n(x) &= 0 \\ \Rightarrow \sum_{n=0}^{\infty}\left(\lambda_n\; a_n(t)\; + \frac{1}{v^2}\frac{d^2 a_n(t)}{dt^2}\right) f_n(x) &= 0 \end{aligned} \tag{2.208}$$

Since the functions (vectors) f_n are orthonormal, they are linearly independent. Then, every coefficient of f_n in the above equation should be zero, which gives

$$\lambda_n\; a_n(t)\; + \frac{1}{v^2}\frac{d^2 a_n(t)}{dt^2} = 0 \tag{2.209}$$

which can be rewritten as

$$\frac{d^2 a_n(t)}{dt^2} = -\omega_n^2\; a_n(t); \quad \omega_n = v\;\frac{(n+1/2)\pi}{a} \tag{2.210}$$

which is solved to give

$$a_n(t) = c_n \cos\omega_n t + d_n \sin\omega_n t \tag{2.211}$$

Then, the most general solution to the wave equation with the given boundary conditions is

$$y(x,t) = \sum_{n=0}^{\infty} (c_n \cos\omega_n t + d_n \sin\omega_n t)\;\; f_n(x) \tag{2.212}$$

The expansion coefficients c_n and d_n are determined by the initial conditions

$$\begin{aligned} y(x,0) &= y_0(x) \\ \left.\frac{\partial y}{\partial t}\right|_{t=0} &= v_0(x) \end{aligned} \tag{2.213}$$

These conditions give

$$\begin{aligned} y_0(x) &= \sum_{n=0}^{\infty} c_n\; f_n(x) \\ v_0(x) &= \sum_{n=0}^{\infty} \omega_n\; d_n\; f_n(x) \end{aligned} \tag{2.214}$$

This are just expansions of vectors $y_0(x)$ and $v_0(x)$ in our vector space over orthonormal basis vectors. Then, the expansion coefficients are just inner products of the vector with the basis vectors

$$\begin{aligned} c_n &= \langle f_n | \; y_0 \rangle \\ &= \int_0^a dx \; f_n(x) \; y_0(x) \\ &= \sqrt{\frac{2}{a}} \int_0^a dx \; \sin\left(\frac{(n+1/2)\pi x}{a}\right) \; y_0(x) \end{aligned} \tag{2.215}$$

and similarly

$$d_n = \frac{1}{\omega_n} \sqrt{\frac{2}{a}} \int_0^a dx \; \sin\left(\frac{(n+1/2)\pi x}{a}\right) \; v_0(x) \tag{2.216}$$

Two-dimensional problems

Next, we solve the wave equation for a circular membrane constrained such that the displacement at the boundary is zero

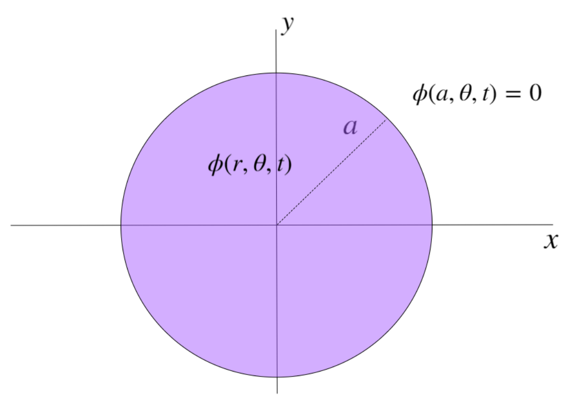

Figure 2.5: Vibrating circular membrane with Dirichlet b.c.

Again, we note that at any instant of time, the displacement is a function $\phi(r, \theta, t)$ deined on a circle, which is non-singular at all points and vanishes at the boundary of the circle. We have seen that such functions form a subspace of the vector space $L^2(M)$ (M being the circular region) over which the operator $\hat{L} = -\nabla^2$ is Hermitian. We have seen that the eigenvalues and eigenvectors of $\hat{L}$ are given by (2.173) and (2.176)

$$\begin{aligned} \lambda_{n,m} &= \frac{\alpha_{n,m}^2}{a^2}; \quad n = 0, 1, 2, 3, .. \;\; m = 1, 2, 3, 4, ... \quad \text{Eigenvalues} \\ f_{n,m}(r, \theta) &= \{ f_{nm}^{(1)}(\theta, \phi) \;, \; f_{nm}^{(2)}(\theta, \phi) \} \\ &= \{ \sqrt{\frac{1}{\pi}} \; N_{nm} \; \cos(n\theta) \; J_n\left(\alpha_{n,m} \; \frac{r}{a}\right), \; \sqrt{\frac{1}{\pi}} \; N_{nm} \; \sin(n\theta) \; J_n\left(\alpha_{n,m} \; \frac{r}{a}\right) \} \end{aligned} \tag{2.217}$$

Then, at any instant, we can expand the displacement of the membrane in terms of these eigenvectors

$$\phi(r, \theta, t) = \sum_{n,m} \left(a_{nm}^{(1)}(t) \; f_{nm}^{(1)}(\theta, \phi) + a_{nm}^{(2)}(t) \; f_{nm}^{(2)}(\theta, \phi) \right) \tag{2.218}$$

Substituting in the wave equation (2.201), we get

$$\sum_{n,m} \left(a_{nm}^{(1)}(t) \; \nabla^2 f_{nm}^{(1)} + a_{nm}^{(2)}(t) \; \nabla^2 f_{nm}^{(2)} \right) - \frac{1}{v^2} \sum_{n,m} \left(\frac{d^2 a_{nm}^{(1)}}{dt^2} \; f_{nm}^{(1)} + \frac{d^2 a_{nm}^{(2)}}{dt^2} \; f_{nm}^{(2)} \right) = 0 \tag{2.219}$$

Using the fact that

$$\begin{aligned} \nabla^2 f_{nm}^{(1)} &= -\lambda_{nm}\ f_{nm}^{(1)} \\ \nabla^2 f_{nm}^{(2)} &= -\lambda_{nm}\ f_{nm}^{(2)} \end{aligned} \tag{2.220}$$

and the fact that the functions $f_{nm}^{(1)}$ and $f_{nm}^{(2)}$ are LI, we get

$$\begin{aligned} \frac{d^2 a_{nm}^{(1)}}{dt^2} &= -\omega_{nm}^2\ a_{nm}^{(1)} \\ \frac{d^2 a_{nm}^{(2)}}{dt^2} &= -\omega_{nm}^2\ a_{nm}^{(2)} \end{aligned} \tag{2.221}$$

where $\omega_{nm} = (v\ \alpha_{nm})/a$. These are equations of SHM, solved to get

$$\begin{aligned} a_{nm}^{(1)}(t) &= c_{nm}^{(1)} \cos\omega_{nm}t + d_{nm}^{(1)} \sin\omega_{nm}t \\ a_{nm}^{(2)}(t) &= c_{nm}^{(2)} \cos\omega_{nm}t + d_{nm}^{(2)} \sin\omega_{nm}t \end{aligned} \tag{2.222}$$

which finally gives

$$\phi(r,\theta,t) = \sum_{n,m} \left[\left(c_{nm}^{(1)} \cos\omega_{nm}t + d_{nm}^{(1)} \sin\omega_{nm}t \right)\ f_{nm}^{(1)}(\theta,\phi) + \left(c_{nm}^{(2)} \cos\omega_{nm}t + d_{nm}^{(2)} \sin\omega_{nm}t \right)\ f_{nm}^{(2)}(\theta,\phi) \right] \tag{2.223}$$

The expansion coefficients $c_{nm}^{(1)}, c_{nm}^{(2)}, d_{nm}^{(1)}$ and $d_{nm}^{(2)}$ are determined from initial conditions on the displacement and velocity of the membrane

$$\begin{aligned} \phi(r,\theta,0) &= \phi_0(r,\theta) \\ \left.\frac{\partial\phi}{\partial t}\right|_{t=0} &= v_0(\theta,\phi) \end{aligned} \tag{2.224}$$

The first condition gives

$$\phi_0(r,\theta) = \sum_{n,m} \left[c_{nm}^{(1)}\ f_{nm}^{(1)}(\theta,\phi) + c_{nm}^{(2)}\ f_{nm}^{(2)}(\theta,\phi) \right] \tag{2.225}$$

The coefficients $c_{nm}^{(1)}$ and $c_{nm}^{(2)}$ are determined by taking the inner product of the vector $\phi_0(r,\theta)$ with the basis vectors $f_{nm}^{(1)}$ and $f_{nm}^{(2)}$

$$\begin{aligned} c_{nm}^{(1)} &= \left\langle f_{nm}^{(1)} \middle|\ \phi_0 \right\rangle \\ &= \int_0^a dr\ r \int_0^{2\pi} d\theta\ f_{nm}^{(1)}\ \phi_0(r,\theta) \\ c_{nm}^{(2)} &= \left\langle f_{nm}^{(2)} \middle|\ \phi_0 \right\rangle \\ &= \int_0^a dr\ r \int_0^{2\pi} d\theta\ f_{nm}^{(2)}\ \phi_0(r,\theta) \end{aligned} \tag{2.226}$$

Similarly, coefficients $d_{nm}^{(1)}$ and $d_{nm}^{(2)}$ are determined in terms of the velocity function $v_0(r,\theta)$

$$\begin{aligned} d_{nm}^{(1)} &= \frac{1}{\omega_{nm}} \left\langle f_{nm}^{(1)} \middle|\ v_0 \right\rangle \\ &= \frac{1}{\omega_{nm}} \int_0^a dr\ r \int_0^{2\pi} d\theta\ f_{nm}^{(1)}\ v_0(r,\theta) \\ d_{nm}^{(2)} &= \frac{1}{\omega_{nm}} \left\langle f_{nm}^{(2)} \middle|\ \phi_0 \right\rangle \\ &= \frac{1}{\omega_{nm}} \int_0^a dr\ r \int_0^{2\pi} d\theta\ f_{nm}^{(2)}\ v_0(r,\theta) \end{aligned} \tag{2.227}$$

As an example, consider a membrane which is struck at $t = 0$, such that the initial displacement is zero, and the initial velocity has circular symmetry. Then, $\phi_0 = 0$ and $v_0 = v_0(r)$. Then, it follows that $c^{(1)}_{nm} = c^{(2)}_{nm} = 0$. Further,

$$\begin{aligned} d^{(1)}_{nm} &= \frac{1}{\omega_{nm}} \int_0^a dr\ r \int_0^{2\pi} d\theta\ f^{(1)}_{nm}\ v_0(r) \\ &= \frac{1}{\omega_{nm}} \sqrt{\frac{1}{\pi}}\ N_{nm} \int_0^{2\pi} d\theta \cos(n\theta) \int_0^a dr\ r\ J_n\left(\alpha_{n,m}\ \frac{r}{a}\right)\ v_0(r) \end{aligned} \tag{2.228}$$

The θ integral is zero, unless $n = 0$, in which case it equals 2π. Then, the integral equals $2\pi\delta_{n0}$. Therefore,

$$d^{(1)}_{nm} = \frac{2\pi\delta_{n0}}{\omega_{0m}} \sqrt{\frac{1}{\pi}}\ N_{0m} \int_0^a dr\ r\ J_0\left(\alpha_{0,m}\ \frac{r}{a}\right)\ v_0(r) \tag{2.229}$$

where we have replaced n by zero everywhere. Here, the normalization constant N_{0m} is given by (2.178) (with $n = 0, m' = m$)

$$\begin{aligned} \frac{1}{N^2_{0m}} &= \int_0^a dr\ r\ J_0^2\left(\alpha_{0,m}\ \frac{r}{a}\right) \\ &= \frac{a^2}{2} J_1(\alpha_{0,m}) \end{aligned} \tag{2.230}$$

where $\alpha_{0,m}$ is the m^{th} zero of J_0

$$J_0(\alpha_{0,m}) = 0 \tag{2.231}$$

Similarly, coefficient $d^{(1)}_{nm}$ is given by

$$\begin{aligned} d^{(2)}_{nm} &= \frac{1}{\omega_{nm}} \int_0^a dr\ r \int_0^{2\pi} d\theta\ f^{(2)}_{nm}\ v_0(r) \\ &= \frac{1}{\omega_{nm}} \sqrt{\frac{1}{\pi}}\ N_{nm} \int_0^{2\pi} d\theta \sin(n\theta) \int_0^a dr\ r\ J_n\left(\alpha_{n,m}\ \frac{r}{a}\right)\ v_0(r) \\ &= 0 \quad \because \int_0^{2\pi} d\theta \sin(n\theta) = 0\ \forall\ n \end{aligned} \tag{2.232}$$

Then, finally, the displacement as a function of time is given by

$$\begin{aligned} \phi(r,\theta,t) &= \sum_{n,m} d^{(1)}_{nm}\ \sin\omega_{nm}t\ f^{(1)}_{nm}(\theta,\phi) \\ &= \sum_m \frac{2}{\omega_{0m}}\ N^2_{0m} \left(\int_0^a dr'\ r'\ J_0\left(\alpha_{0,m}\ \frac{r'}{a}\right)\ v_0(r')\right)\ \sin\omega_{0m}t\ \ J_0\left(\alpha_{0,m}\ \frac{r}{a}\right) \end{aligned} \tag{2.233}$$

Here, $\omega_{0m} = (v\ \alpha_{0,m})/a$.

Let us now consider the heat conduction equation. Consider a circular plate whose boundary is maintained at temperature $\phi = 0$, and which has an initial temperature distribution

$$\phi(r,\theta,0) = \phi_0(r,\theta) \tag{2.234}$$

As before, we can expand the instantaneous temperature on the disk in a series of eigenvectors of the Laplacian

$$\phi(r,\theta,t) = \sum_{n,m} \left(a^{(1)}_{nm}(t)\ f^{(1)}_{nm}(\theta,\phi) + a^{(2)}_{nm}(t)\ f^{(2)}_{nm}(\theta,\phi)\right) \tag{2.235}$$

Substituting this in the heat conduction equation (2.202), and using the fact that functions f_{nm} are eigenvectors of $-\nabla^2$ with eigenvalues (2.173)and form a LI set, we get

$$\begin{aligned}\frac{da^{(1)}_{nm}}{dt} &= -D\frac{\alpha^2_{nm}}{a^2}\ a^{(1)}_{nm}\\ \frac{da^{(2)}_{nm}}{dt} &= -D\frac{\alpha^2_{nm}}{a^2}\ a^{(2)}_{nm}\end{aligned} \tag{2.236}$$

which are integrated to give

$$\begin{aligned}a^{(1)}_{nm}(t) &= a^{(1)}_{nm}(0)\ e^{-(D\alpha^2_{nm}/a^2)\ t}\\ a^{(2)}_{nm}(t) &= a^{(2)}_{nm}(0)\ e^{-(D\alpha^2_{nm}/a^2)\ t}\end{aligned} \tag{2.237}$$

Then, the instantaneous temperature distribution is given by

$$\phi(r,\theta,t) = \sum_{n,m}\left(a^{(1)}_{nm}(0)\ f^{(1)}_{nm}(\theta,\phi) + a^{(2)}_{nm}(0)\ f^{(2)}_{nm}(\theta,\phi)\right)\ e^{-(D\alpha^2_{nm}/a^2)\ t} \tag{2.238}$$

It is clear that the temperature changes exponentially fast to reach the equilibriu configuration $\phi(r,\theta,t) = 0$. The expansion constants are determined by the initial temperature distribution

$$\phi_0(r,\theta) = \sum_{n,m}\left(a^{(1)}_{nm}(0)\ f^{(1)}_{nm}(\theta,\phi) + a^{(2)}_{nm}(0)\ f^{(2)}_{nm}(\theta,\phi)\right) \tag{2.239}$$

Using orthonormality of the eigenvectors, we get

$$\begin{aligned}a^{(1)}_{nm}(0) &= \left\langle f^{(1)}_{nm}\middle|\ \phi_0\right\rangle\\ &= \int_0^a dr\ r\int_0^{2\pi} d\theta\ f^{(1)}_{nm}\ \phi_0(r,\theta)\\ a^{(2)}_{nm}(0) &= \left\langle f^{(2)}_{nm}\middle|\ \phi_0\right\rangle\\ &= \int_0^a dr\ r\int_0^{2\pi} d\theta\ f^{(2)}_{nm}\ \phi_0(r,\theta)\end{aligned} \tag{2.240}$$

2.9 Inhomogeneous boundary conditions: Laplace's equation

Consider the heat conduction equation, with inhomogeneous boundary conditions: that is, the temperature at the boundary is non-zero, and in general an arbitrary (but constant) function ϕ_B defined on the boundary

$$\frac{\partial\phi}{\partial t} - D\nabla^2\phi = 0; \quad \phi = \phi_B \text{ on the boundary} \tag{2.241}$$

In such a situation, we can define a new function

$$\tilde{\phi}(x,y,z,t) = \phi(x,y,z,t) - \phi_P(x,y,z) \tag{2.242}$$

where the function $\phi_P(x,y,z)$ satisfies Laplace's equation

$$\nabla^2\phi_P = 0; \quad \phi_P = \phi_B \text{ on the boundary} \tag{2.243}$$

Since $\partial\phi_P/\partial t = 0$, the function $\tilde{\phi}$ also satisfies the heat conduction equation, but with homogeneous boundary condition

$$\frac{\partial\tilde{\phi}}{\partial t} - D\nabla^2\tilde{\phi} = 0; \quad \tilde{\phi} = 0 \text{ on the boundary} \tag{2.244}$$

The solution ϕ_P is physically the *steady state* solution, attained after the system has attained thermal equilibrium. Then, solutions to Laplace's equation can be visualized as steady state temperature distributions, given a fixed temperature at the boundary.

As an example, consider heat flow in a circular plate, with the temperature on the circular boundary maintained as a function $\phi_B(\theta)$

$$\phi(a,\theta,t) = \phi_B(\theta) \tag{2.245}$$

To solve the problem, all we need to do is find the steady state solution

$$\nabla^2\phi_P = 0; \quad \phi_P(a,\theta) = \phi_B(\theta) \tag{2.246}$$

The general solution will be given by $\phi(r,\theta,t) = \tilde{\phi}(r,\theta,t) + \phi_P(r,\theta)$, where we already know how to solve for $\tilde{\phi}(r,\theta,t)$ (solution with homogeneous b.c.). To solve for (2.246), we take the following approach. We can write the Laplacian on the disk as

$$\begin{aligned}\hat{L} &= -\frac{\partial^2}{\partial r^2} - \frac{1}{r}\frac{\partial}{\partial r} - \frac{1}{r^2}\frac{\partial^2}{\partial\theta^2} \\ &= -\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\hat{L}_C\end{aligned} \tag{2.247}$$

where $\hat{L}_C = -\partial^2/\partial\theta^2$ is the Laplacian defined on a circle. This suggests the following approach to solving (2.246). For a given r, $\phi_P(r,\theta)$ can be thought of as a function of θ, a function belonging to the vector space $L_2(C)$. We know that the operator $\hat{L}_C = -\partial^2/\partial\theta^2$ is Hermitian when acting on vectors belonging to $L_2(C)$. Therefore, any such function can be expanded interms of eigenvectors of $\hat{L}_C$. The eigenvalues and normalized eigenvectors of $\hat{L}_C$ are

$$\begin{aligned}\lambda_n &= n^2; \quad n = 0,1,2,3,... \\ f_n(\theta) &= \{\frac{1}{\sqrt{\pi}}\cos n\theta \ , \ \frac{1}{\sqrt{\pi}}\sin n\theta\}\end{aligned} \tag{2.248}$$

Then, we can expand $\phi_P(r,\theta)$ in a series of these functions, with r dependent coefficients

$$\phi_P(r,\theta) = \sum_{n=0}^{\infty}\left(a_n(r)\cos n\theta + b_n(r)\sin n\theta\right) \tag{2.249}$$

Then, Laplace's equation is just $\hat{L}\phi_P = 0$, which gives

$$\sum_{n=0}^{\infty}\left(-\frac{1}{r}\frac{d}{dr}\left(r\frac{da_n}{dr}\right)\cos n\theta - \frac{1}{r}\frac{d}{dr}\left(r\frac{db_n}{dr}\right)\sin n\theta\right) + \sum_{n=0}^{\infty}\left(\frac{n^2}{r^2}a_n(r)\cos n\theta + \frac{n^2}{r^2}b_n(r)\sin n\theta\right) = 0 \tag{2.250}$$

where the r derivatives in $\hat{L}$ act only on functions a_n and b_n, and $\hat{L}_C \sin n\theta = -n^2 \sin n\theta$, $\hat{L}_C\cos n\theta = -n^2\cos n\theta$. Using the LI of $\cos n\theta$ and $\sin n\theta$, we get

$$\begin{aligned}-\frac{1}{r}\frac{d}{dr}\left(r\frac{da_n}{dr}\right) + \frac{n^2}{r^2}a_n(r) &= 0 \\ -\frac{1}{r}\frac{d}{dr}\left(r\frac{db_n}{dr}\right) + \frac{n^2}{r^2}b_n(r) &= 0\end{aligned} \tag{2.251}$$

which reduce to simple equations

$$\begin{aligned}r^2\frac{d^2a_n}{dr^2} + r\frac{da_n}{dr} - n^2a_n &= 0 \\ r^2\frac{d^2b_n}{dr^2} + r\frac{db_n}{dr} - n^2a_n &= 0\end{aligned} \tag{2.252}$$

which have solutions

$$\begin{aligned}a_n(r) = c_1\ r^n + c_2\ r^{-n} \\ b_n(r) = d_1\ r^n + d_2\ r^{-n}\end{aligned} \tag{2.253}$$

substituting these in ϕ_P, we get

$$\phi_P(r,\theta) = \sum_{n=0}^{\infty} \left[\left(c_1\ r^n + c_2\ r^{-n}\right)\cos n\theta + \left(d_1\ r^n + d_2\ r^{-n}\right)\sin n\theta\right] \tag{2.254}$$

Since the temperature is non-singular at $r = 0$, $c - 2 = d_2 = 0$. This gives

$$\phi_P(r,\theta) = \sum_{n=0}^{\infty} [c_1\ r^n\ \cos n\theta + d_1\ r^n\ \sin n\theta] \tag{2.255}$$

The constants c_1 and d_1 are determined by the temperature distribution at $r = a$

$$\begin{aligned} \phi_B(\theta) &= \phi_P(a,\theta) \\ &= \sum_{n=0}^{\infty} a^n\, [c_1 \cos n\theta + d_1\ \sin n\theta] \end{aligned} \tag{2.256}$$

Using orthogonality of the sin and cos functions, we get

$$\begin{aligned} c_1 &= \frac{1}{\pi a^n}\int_0^{2\pi} d\theta\ \phi_B(\theta)\cos n\theta \\ d_1 &= \frac{1}{\pi a^n}\int_0^{2\pi} d\theta\ \phi_B(\theta)\sin n\theta \end{aligned} \tag{2.257}$$