A Mathematical Approach to Quantum Mechanics

Jose Alfredo Alanis[1,a] and Ali Behzadan[2,a]

[a]Sacramento State University, Sacramento CA 95819, United States

June 14, 2025

josefreddyalanis@gmail.com

Abstract

This document is intended for any mathematician who does not have a strong background in physics but has an understanding of mathematical analysis. In this document, we will discuss mathematical quantum mechanics. These notes are motivated by Thomas A Garrity's book *Electricity and Magnetism for Mathematicians* [1].

Contents

1 Hilbert Spaces

1.1 Historical context

As with everything that has a strong foundation, it begins with a powerful mathematical basis. Quantum mechanics is no different. Indeed, it is possible to understand the basics of this phenomenon through experiments conducted in the lab. But just like any other subject, you reach a barrier and ultimately need to rely on mathematics. The universe doesn't behave according to our classical intuitions. It follows the laws of physics, and those rules are expressed mathematically. This boils down to two general areas in mathematics: functional analysis and Lie theory.

There's an ongoing debate in physics. Did the mathematics predict the physics, or did experiments force us to develop the mathematics? This question is particularly relevant to quantum mechanics. A notorious example of this is the Schrödinger's equation which was initially derived from mathematical principles. However, it was not until later that the Schrödinger's equation was confirmed by experiments. And thus, we arrive at a chicken or egg debate with your guess being just as good as mines.

In this text, we will develop our theory through logic. We will start from the ground up, as done in traditional mathematics classes, in the hope that we can accurately reflect reality. Of course, what better way to begin than by reviewing Analysis.

1.2 Mathematical Analysis

Before we discuss Hilbert spaces, we will first go over some critical definitions. These definitions can be found in any analysis textbook. We will start with a metric space.

Definition

Let $\mathbb{X}$ be a non-empty set. A metric on $\mathbb{X}$ is a function $d : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ is said to be a distance function or a metric if it satisfies the following.

- **Non-Negativity:** For all $x, y \in \mathbb{X}$,

$$d(x, y) \geq 0.$$

- **Identity of Indiscernibles:** For all $x, y \in \mathbb{X}$,

$$d(x, y) = 0 \iff x = y.$$

- **Symmetry:** For all $x, y \in \mathbb{X}$,

$$d(x, y) = d(y, x).$$

- **Triangle Inequality:** For all $x, y, z \in \mathbb{X}$,

$$d(x, y) \leq d(x, z) + d(z, y).$$

A set $\mathbb{X}$ equipped with a metric d is called a metric space written as $(\mathbb{X}, d)$

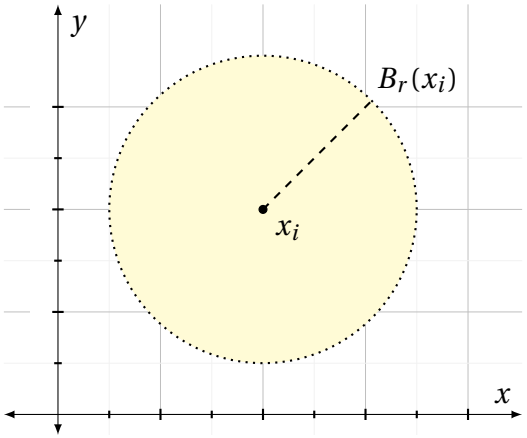
Metric spaces allow us to utilize distances. Our first notation will be that of an open ball.

Definition

Let $(\mathbb{X}, d)$ be a metric space and $x \in \mathbb{X}$. The ball $B_r(x)$ of radius r centered at x_i is the set of all points y_i such that the absolute value of the difference between x_i and y_i is less than r. Formally,

$$\begin{aligned} B_r(x_i) &= \{y_i \in \mathbb{X} \mid |x_i - y_i| < r\} \\ &= \{y_i \in \mathbb{X} : d(x_i, y_i) < r\} \end{aligned}$$

Let's look at a picture of this.



Indeed, this definition ensures that the distance between any two points inside the open ball is always less than the distance from the center of the ball, in our case x_i, to the edge of the ball. We denote this distance by ϵ.

Now, we will introduce the definition of convergence. The definition of convergence is fundamental in mathematical analysis.

Before we define what it means to converge, we must first define what a sequence is.

Definition

A sequence is a mapping from $f : \mathbb{N} \to S$. That is, it assignments each number from $\mathbb{N}$ to a corresponding element of the set S.

Example 1.1. Determine by definition which of the following are sequences.

1. $(x_n)_n = \left(\frac{1}{n}\right)_n = \left(1, \frac{1}{2}, \frac{1}{3}, \ldots\right)$ is a sequence. Notice that this is defined as a mapping $f : \mathbb{N} \to S$ by $f(n) = \frac{1}{n}$, where S is the codomain of the sequence.

2. Likewise, we consider $(1,1,1,1,1,\ldots)$ to be a constant sequence. Indeed, we can define a mapping $f : \mathbb{N} \to S$ by $f(n) = 1$ for all $n \in \mathbb{N}$.

3. However, $(1,2,3,4,5,\ldots,10000000)$ is not a sequence. Why? The list of numbers is finite. A sequence requires a mapping from $\mathbb{N}$ to S, meaning it must be defined for all $n \in \mathbb{N}$, which is infinite.

Definition

Being more general, let $(x_n)_n$ be a sequence of real numbers. The expression of the form

$$\sum_{n=1}^{\infty} x_n = x_1 + x_2 + x_3 + \ldots$$

is called an infinite sum.

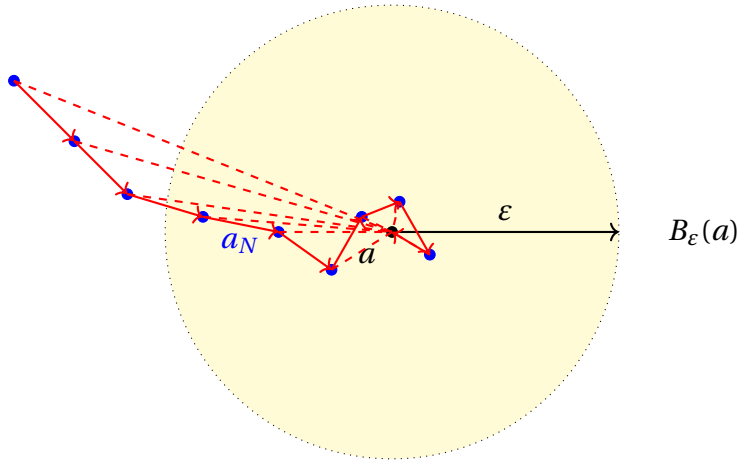
Being more general, let $(x_n)_n$ be a sequence of real numbers.

Definition

Let $(\mathbb{X}, d)$ be a metric space. A sequence (a_n) in $\mathbb{X}$ converges to $a \in \mathbb{X}$ if for all ϵ, there exist and N such that for all $n > N$ then $d(a_n, a) < \epsilon$. If a sequence $(a_n) \to a$ then we say

$$\lim_{n\to\infty} a_n = a.$$

It is important to note that here, we are interested in the distances between a_n and the limit a. Ultimately, all but a finite number of elements in the sequence must be inside the ball. This means that there can be a finite number of terms outside the ball. These finite numbers can be as big as you'd like. For example 10, 100, or even 100,000,000 all work. However, as long as the remaining (infinitely many) terms of the sequence eventually stay inside the ball, we say the sequence converges. The reason is that no matter how large the finite number of terms outside the ball is, it does not affect the long-term behavior of the sequence.



After a_N, all points satisfy $|a_n - a_m| < \varepsilon$ for $n, m \geq N$.

Definition

Let $(\mathbb{X}, d)$ be a metric space. A sequence (a_n) is cauchy if for all ϵ, there exist and N such that for all $n, m > N$ then $d(a_n, a_m) < \epsilon$.

In other words, all of a_n's past a_N distances from each other are less than ε. The definition of Cauchy sequences does not focus on the limit point. Instead, we are

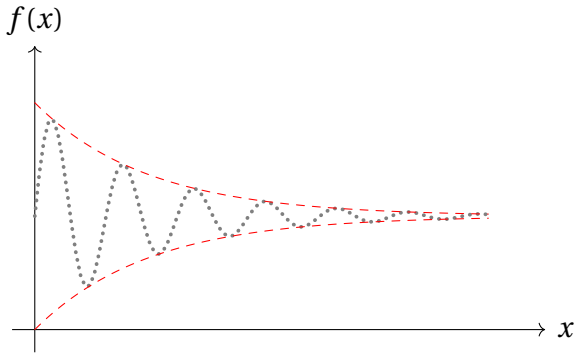


Figure 1: Cauchy

interested in the distances between the terms of the sequence a_n and a_m. These terms must become arbitrarily small as $n, m \to \infty$.

Traditionally when we are faced with a difficult equation with limited information, a common approach is to study simpler, related equations. Let us call the original hard equation E.

We first look at a simpler equation E_1 that resembles E. If E_1 is still too difficult, we consider an even simpler equation E_2. By continuing this process we construct a sequence of equations $E_1, E_2, E_3, \ldots$. Indeed, each simpler equation is a better approximation of E.

We do this by finding the corresponding solutions x_n for each of these simpler equations E_n. This allows us to generate a sequence (x_n) of solutions.

The goal is to show that the sequence (x_n) converges to a solution of the original hard equation E. To show this we utilize the definition of being Cauchy. If we can show that (x_n) is Cauchy in a complete metric space, then the sequence must converge to some limit x. If this limit exists and satisfies the original equation E, we have effectively solved the problem by approximating it through simpler equations.

The power of this method lies in the completeness of the space we are working in. In a complete metric space, every Cauchy sequence converges. If the limit satisfies the original equation, then it is the desired solution to E. This Cauchy approach is widely used in mathematical application.

This approach is widely used in mathematical analysis, numerical methods, and functional analysis, particularly in solving differential equations, integral equations, and other problems where direct solutions are challenging to obtain.

Definition

A *vector space* over a field $\mathbb{F}$ is a set $\mathbb{V}$ equipped with two operations:

1. *Vector addition*: a binary operation $+ : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$, and
2. *Scalar multiplication*: a binary operation $\cdot : \mathbb{F} \times \mathbb{V} \to \mathbb{V}$,

such that the following axioms hold for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{V}$ and $c, d \in \mathbb{F}$:

- Closure under addition: $\mathbf{u} + \mathbf{v} \in \mathbb{V}$.

such that the following axioms hold for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{V}$ and $c, d \in \mathbb{F}$:

- Commutativity of addition: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.
- Associativity of addition: $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$.
- Existence of additive identity: There exists an element $\mathbf{0} \in \mathbb{V}$ such that $\mathbf{u} + \mathbf{0} = \mathbf{u}$.
- Existence of additive inverses: For every $\mathbf{u} \in \mathbb{V}$, there exists $-\mathbf{u} \in \mathbb{V}$ such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$.
- Closure under scalar multiplication: $c \cdot \mathbf{u} \in \mathbb{V}$.
- Distributivity of scalar multiplication over vector addition: $c \cdot (\mathbf{u} + \mathbf{v}) = c \cdot \mathbf{u} + c \cdot \mathbf{v}$.
- Distributivity of scalar multiplication over field addition: $(c + d) \cdot \mathbf{u} = c \cdot \mathbf{u} + d \cdot \mathbf{u}$.
- Associativity of scalar multiplication: $c \cdot (d \cdot \mathbf{u}) = (c \cdot d) \cdot \mathbf{u}$.
- Identity element of scalar multiplication: $1 \cdot \mathbf{u} = \mathbf{u}$, where 1 is the multiplicative identity in $\mathbb{F}$.

Definition

A subspace H of a vector space V over a field $\mathbb{F}$ (typically $\mathbb{R}$ or $\mathbb{C}$ is a subset of V such that the following three conditions hold.

Non-emptiness: H contains the zero vector 0.

$$0 \in H$$

- Closed under vector addition: For any vectors $u, v \in H$, their sum $u+v$ is also in H.

$$u, v \in H \implies u+v \in H$$

- Closed under scalar multiplication: For any vector $v \in H$ and any scalar $c \in \mathbb{F}$, the scalar multiple cv is also in H.

$$v \in H \text{ and } c \in \mathbb{F} \implies cv \in H$$

A subspace is itself a vector space, as it is closed under addition and scalar multiplication and includes the zero vector, thus satisfying all the axioms of a vector space.

Definition

Let $\mathbb{V}$ be a vector space. An inner product on $\mathbb{V}$ is a function $\langle \cdot, \cdot \rangle : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ such that.

- **Linearity (or Sesquilinearity in the complex case):**
  - For all $u, v, w \in \mathbb{V}$ and $a, b \in \mathbb{R}$:

$$\langle au + bv, w \rangle = a\langle u, w \rangle + b\langle v, w \rangle.$$

  - For all $u, v, w \in \mathbb{V}$ and $a, b \in \mathbb{C}$:

$$\langle u, av + bw \rangle = \overline{a}\langle u, v \rangle + \overline{b}\langle u, w \rangle,$$

where $\overline{a}$ denotes the complex conjugate of a.

- **Symmetry (or Hermitian Symmetry in the complex case):**
  - For all $u, v \in \mathbb{V}$:
$$\langle u, v\rangle = \overline{\langle v, u\rangle}.$$
  - In the real case, this reduces to:
$$\langle u, v\rangle = \langle v, u\rangle \quad \text{(symmetry)}.$$
- **Positive Definiteness:**
  - For all $u \in \mathbb{V}$:
$$\langle u, u\rangle \geq 0,$$
    and $\langle u, u\rangle = 0$ if and only if $u = 0$.
- **Non-Degeneracy:**
  - The inner product separates points in the sense that:
$$\langle u, u\rangle = 0 \quad \text{for all } u \in \mathbb{V} \implies u = 0.$$

A vector space with an inner product is classed an inner product space. Every inner product space is a metric space. Suppose that $\langle \cdot, \cdot \rangle$ is an inner product on the vector space $\mathbb{V}$.Then for all $u, v \in \mathbb{V}$

$$d(u, v) = \sqrt{\langle (u - v), (u - v)\rangle}.$$

That is, the distance function is induced by an inner product.

Theorem 1.2. *Cauchy-Schwarz Inequality*

$$|\langle y, x\rangle|^2 \leq \langle x, x\rangle\langle y, y\rangle \quad \textit{for all } x, y \in \mathbb{X}$$

Proof. For $y \neq 0$:

$$\begin{aligned} 0 &\leq \left\langle x - \frac{\langle y, x\rangle}{\langle y, y\rangle} y, x - \frac{\langle y, x\rangle}{\langle y, y\rangle} y \right\rangle \\ &= \langle x, x\rangle - \frac{\overline{\langle y, x\rangle}}{\langle y, y\rangle}\langle y, x\rangle - \frac{\langle y, x\rangle}{\langle y, y\rangle}\langle x, y\rangle + \frac{\overline{\langle y, x\rangle}}{\langle y, y\rangle}\frac{\langle y, x\rangle}{\langle y, y\rangle}\langle y, y\rangle \\ &= \langle x, x\rangle - \frac{\langle y, x\rangle^2}{\langle y, y\rangle}. \end{aligned}$$

Rearranging.

$$|\langle y, x\rangle|^2 \leq \langle x, x\rangle\langle y, y\rangle.$$

□

Definition

A norm $\|\cdot\|$ on a vector space $\mathbb{V}$ is said to be induced by an inner product $\langle\cdot,\cdot\rangle$ if it satisfies the formula:

$$\|x\| = \sqrt{\langle x, x\rangle},$$

for all $x \in \mathbb{V}$.

It is important to know that $\|x\| = \sqrt{\langle x, x\rangle}$, is just a number. That is, consider the vector space $\mathbb{R}^2$ with the standard dot product. Indeed, $v_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $v_2 = \begin{bmatrix} -3 \\ 5 \end{bmatrix} \in \mathbb{R}^2$. Notice that

$$\begin{aligned} \langle v_1, v_2\rangle &= \left\langle \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -3 \\ 5 \end{bmatrix} \right\rangle \\ &= (1)(-3) + (2)(5) \\ &= 7. \end{aligned}$$

Now, of course,

$$\begin{aligned}\langle v_1, v_1\rangle &= \left\langle \begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}1\\2\end{bmatrix} \right\rangle \\ &= (1)(1) + (2)(2) \\ &= 1^2 + 2^2 \\ &= 5\end{aligned}$$

Since $\|x\| = \sqrt{\langle x, x\rangle}$, then $\|v_1\| = \sqrt{\langle v_1, v_1\rangle}$. Raising both sides by two gives us, $\|v_1\|^2 = \langle v_1, v_1\rangle$.

1.3 Hilbert Spaces

Definition

A metric space (X, d) in which every cauchy sequence converges is a **complete** metric space. Likewise, An inner product space $(V, \langle\cdot\rangle)$ is called **complete** if every Cauchy sequence in V converges to a limit in V. This complete inner product space is called a **Hilbert space** denoted as $\mathscr{H}$.

Let's proceed with some examples.

Example 1.3. Determine by definition which of the following are Hilbert spaces.

1. $\mathbb{R}^n$ with the Euclidean inner product is a Hilbert space. Let $x = \begin{bmatrix}x_1\\\vdots\\x_n\end{bmatrix} \in \mathbb{R}^n$. One can see that $(\langle x, x\rangle) = \left\langle \begin{bmatrix}x_1\\\vdots\\x_n\end{bmatrix}, \begin{bmatrix}x_1\\\vdots\\x_n\end{bmatrix} \right\rangle = x_1^2 + \ldots + x_n^2$. Notice that the induced norm

(length) is

$$\begin{aligned}\|\cdot\|_2 &= \sqrt{x_1^2 + ...x_n^2} \\ &= \sqrt{\sum_{i=1}^{n} x_i^2} \\ &= (\langle x, x\rangle)^{\frac{1}{2}}\end{aligned}$$

Additionally, notice that $(\mathbb{R}^n, \|\cdot\|_2)$ is complete.

2. $C[a,b] = \{f : [a,b] \to \mathbb{R} : f \text{ is continuous}\}$ with the L^2 inner product $\langle f, g\rangle = \int_a^b f(x)g(x)dx$ is not a Hilbert space. Although the induced norm is

$$\begin{aligned}\|f\|_2 &= \sqrt{\int_a^b |f|^2} \\ &= (\langle f, f\rangle).\end{aligned}$$

Notice that $(C[a,b], \|\cdot\|_2)$ is not complete. The distance is defined as $d(f,g) = ||f - g||$. One may ask, how do we find distance with this information. Consider the function $f : [0,1] \to \mathbb{R}$ by $f(x) = x^3$. Using the above we can find the distance of f as follows,

$$\begin{aligned}|f| &= \sqrt{\int_0^1 f^2 dx} \\ &= \sqrt{\int_0^1 x^3 dx} \\ &= \sqrt{\frac{1}{7}(1)^6 + \frac{1}{7}(0)^6} \\ &= \sqrt{\frac{1}{7}}.\end{aligned}$$

Moreover, one can easily calculate the distance between two functions. Let $f = x^2$ and $g = x^2$. Then

$$\begin{aligned} d(x^2, x^3) &= ||f - g|| \\ &= \sqrt{\int_0^1 |x^2 - x^3| dx}. \end{aligned}$$

3. Define $\ell^2 = \{(x_1, x_2, ...) : \sum_{i=1}^{\infty} |x_i|^2 < \infty\}$, let $(x_n)_n, (y_n)_n \in \ell^2$ defined by $\langle x_n, y_n \rangle = \sum_{i=1}^{\infty} x_k y_k$. Indeed $(\ell^2, \langle \cdot \rangle)$ is a Hilbert space. First, notice that $(\ell^2, \langle \cdot \rangle)$ is complete and the induced norm is

$$\begin{aligned} \| \cdot \|_2 &= \sqrt{\sum_{i=1}^{\infty} x_i^2} \\ &= (\langle x_n, x_n \rangle)^{\frac{1}{2}} \end{aligned}$$

In the second example, we said that the L^2 inner product with respect to the continuous functions, $\langle f, g \rangle = \int_a^b f(x) g(x) dx$, does not define a Hilbert space. We know that the vector space of continuous functions is not complete. Therefore, we would like to create a space of functions that includes continuous functions and is also complete. We will now define the Lebesgue inner product.

Theorem 1.4. *Let $I = [a, b]$ be any interval. Define $L_2(I)$ to be the set of all square Lebesgue integrable functions on I. The inner product on $L^2(I)$ as*

$$\langle f, g \rangle = \int_I f(x) \overline{g(x)} dx,$$

is an inner product space. Notice that this is a mapping from the interval $I = [a, b]$ to $\mathbb{C}$. Note that our interval can be $\mathbb{R}^n$ such that $n \in \mathbb{N}$.

Proof. Firstly, notice that each element of this inner product space are functions. We define the inner product on $L^2(I)$ for complex-valued functions as

$$\langle f, g \rangle = \int_I f(x) \overline{g(x)} dx.$$

We verify that this satisfies the properties of an inner product. We begin by showing linearity. For any $f, g, h \in L^2(I)$ and any scalar $\alpha \in \mathbb{C}$, we have

$$\begin{aligned}\langle \alpha f + g, h\rangle &= \int_I (\alpha f(x) + g(x))\overline{h(x)}\,dx \\ &= \int_I \alpha f(x)\overline{h(x)}\,dx + \int_I g(x)\overline{h(x)}\,dx \\ &= \alpha \int_I f(x)\overline{h(x)}\,dx + \int_I g(x)\overline{h(x)}\,dx \\ &= \alpha\langle f, h\rangle + \langle g, h\rangle.\end{aligned}$$

Similarly, we have $\langle f + \beta g, h\rangle = \langle f, h\rangle + \beta\langle g, h\rangle$ for $\beta \in \mathbb{C}$.

Next, we show symmetry,

$$\begin{aligned}\langle f, g\rangle &= \int_I f(x)\overline{g(x)}\,dx \\ &= \int_I g(x)\overline{f(x)}\,dx = \langle g, f\rangle.\end{aligned}$$

Now, we check positive definiteness,

$$\begin{aligned}\langle f, f\rangle &= \int_I f(x)\overline{f(x)}\,dx \\ &= \int_I |f(x)|^2\,dx \geq 0.\end{aligned}$$

Since $|f(x)|^2 \geq 0$, we have $\langle f, f\rangle \geq 0$.

Lastly, we verify non-degeneracy. If $\langle f, f\rangle = 0$, then:

$$\int_I |f(x)|^2\,dx = 0,$$

which implies $f(x) = 0$ almost everywhere. Thus, $f = 0$ in $L^2(I)$.

Since we have verified all four properties (linearity, symmetry, positive definiteness, and non-degeneracy), we conclude that $(L^2(I), \langle\cdot,\cdot\rangle)$ is an inner product space when the scalars are complex.

□

Let's now define the norm (length) in $L^2(I)$.

The norm on $L^2(I)$ is called the L^2-norm and is induced by the inner product

$$\|f\|_2 = \sqrt{\langle f, f\rangle}.$$

For complex-valued functions, the inner product in $L^2(I)$ is defined as

$$\langle f, g\rangle = \int_I f(x)\overline{g(x)}\,dx.$$

Thus, the L^2-norm is explicitly given by

$$\|f\|_2 = \left(\int_I |f(x)|^2\,dx\right)^{1/2}.$$

Let's give an example of a function in $L^2(I)$.

Computing the L^2-Norm of $f(x) = x$ on $[0,1]$. Notice that this example is in $\mathbb{R}$.

We compute the L^2-norm of the function $f(x) = x$ over $[0,1]$.

$$\begin{aligned}
\|f\|_2 &= \left(\int_0^1 x \cdot x\,dx\right)^{1/2} \\
&= \left(\int_0^1 x^2\,dx\right)^{1/2} \\
&= \left(\frac{x^3}{3}\Big|_0^1\right)^{1/2} \\
&= \left(\frac{1^3}{3} - \frac{0^3}{3}\right)^{1/2} \\
&= \sqrt{\frac{1}{3}} \\
&= \frac{1}{\sqrt{3}}
\end{aligned}$$

Computing the L^2-Norm of $f(x) = e^{ix}$ on $[0,1]$. Notice that this example is in $\mathbb{C}$.

We compute the L^2-norm of the function $f(x) = e^{ix}$ over $[0,1]$.

$$\|f\|_2 = \left(\int_0^1 |e^{ix}|^2 \, dx \right)^{1/2}.$$

Since $e^{ix}\overline{e^{ix}} = e^{ix}e^{-ix} = 1$, we obtain

$$\begin{aligned} \|f\|_2 &= \left(\int_0^1 1 \, dx \right)^{1/2} \\ &= \left(x \Big|_0^1 \right)^{1/2} \\ &= (1-0)^{1/2} \\ &= \sqrt{1} \\ &= 1. \end{aligned}$$

Computing the L^2-Norm of $f(x) = (1+i)x^2$ on $[0,1]$.

We compute the L^2-norm of the function $f(x) = (1+i)x^2$ over $[0,1]$.

$$\|f\|_2 = \left(\int_0^1 |(1+i)x^2|^2 \, dx \right)^{1/2}.$$

First, we compute the squared modulus:

$$\begin{aligned} |(1+i)x^2|^2 &= (1+i)x^2 \cdot \overline{(1+i)x^2} \\ &= (1+i)x^2 \cdot (1-i)x^2 \\ &= (1-i^2)x^4 \\ &= (1+1)x^4 \\ &= 2x^4. \end{aligned}$$

Thus, the norm calculation is,

$$\begin{aligned}
\|f\|_2 &= \left(\int_0^1 2x^4\,dx\right)^{1/2} \\
&= \left(2\int_0^1 x^4\,dx\right)^{1/2} \\
&= \left(2\cdot\frac{x^5}{5}\Big|_0^1\right)^{1/2} \\
&= \left(2\cdot\frac{1}{5}\right)^{1/2} \\
&= \left(\frac{2}{5}\right)^{1/2} \\
&= \frac{\sqrt{2}}{\sqrt{5}} \\
&= \frac{\sqrt{10}}{5}.
\end{aligned}$$

Lastly, let's define the distance in $L^2(I)$. The distance between two functions $f, g \in L^2(I)$ is given by

$$\begin{aligned}
d(f,g) &= \|f-g\|_2 \\
&= \sqrt{\langle f-g, f-g\rangle}.
\end{aligned}$$

Using the definition of the inner product for complex-valued functions in $L^2(I)$,

$$\langle f,g\rangle = \int_I f(x)\overline{g(x)}\,dx,$$

we obtain

$$d(f,g) = \left(\int_I (f(x)-g(x))\overline{(f(x)-g(x))}\,dx\right)^{1/2}.$$

Among all this, there is a silver lining. $L^2(I)$ is a complete inner product space. That is, every Cauchy sequence in $L^2(I)$ converges to some function in $L^2(I)$. Recall that there is a special name for complete inner product spaces. Yes, of course, $L^2(I)$ is a Hilbert space and, in fact, one of the more popular ones when it comes to quantum mechanics.

Theorem 1.5. *$L^2(I)$ is a Hilbert space.*

Theorem 1.6. *Define*

$$l_2 = \left\{ (a_1, a_2, a_3, \ldots) : a_k \in \mathbb{C} \text{ and } \sum_{k=1}^{\infty} |a_k|^2 < \infty \right\}.$$

For $(a_k), (b_k) \in l_2$, *and define*

$$\begin{aligned} \langle (a_k), (b_k) \rangle &= \sum_{k=1}^{\infty} \overline{a_k} b_k \\ &= \| \cdot \|_2 \end{aligned}$$

With this inner product, l_2 *is a Hilbert space. That is,* $(l_2, \| \cdot \|_2)$

1.4 Linear Algebra

Definition

Recall that any two vectors are *orthogonal* if they are perpendicular to each other. In other words, the dot product of the two vectors is zero.

give example dot product of two fn is 0 .....

Example 1.7. The vectors

$$e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

are orthonormal in $\mathbb{R}^2$

Indeed

$$\langle e_1, e_2 \rangle = 1(0) + 0(1) = 0.$$

We can generalize the following. The set of vectors $v_1, v_2, \ldots v_n$ are *mutually orthogonal* if every pair of vectors is orthogonal. That is, for all $i \neq j$, $\langle v_i, v_i \rangle = 0$

Definition

A set of vectors $\{v_1, v_2, \ldots, v_n\}$ in a vector space is called **linearly independent** if

the only solution to the equation

$$c_1 v_1 + c_2 v_2 + \cdots + c_n v_n = 0$$

is

$$c_1 = c_2 = \cdots = c_n = 0.$$

If there exists a nontrivial (nonzero) solution where some $c_i \neq 0$, then the vectors are **linearly dependent**.

Example 1.8. The two vectors

$$v_1 = \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

are linearly dependent.

v_1 and v_2 are linearly dependent because one is a scalar multiple of the other

$$v_1 = 2v_2.$$

This means there exists a nonzero solution to the equation

$$c_1 v_1 + c_2 v_2 = 0.$$

If we set $c_1 = 1$ and $c_2 = -2$, we get

$$(1)\begin{bmatrix} 2 \\ 4 \end{bmatrix} + (-2)\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2-2 \\ 4-4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Since $c_1 = 1$ and $c_2 = -2$ is a nonzero solution, the vectors are linearly dependent.

Definition

A set of vectors $\{v_1, v_2, \ldots, v_n\}$ is called a **basis** for an n-dimensional vector space if every vector in the space can be written as a **linear combination** of these basis vectors. This means that for any vector v in the space, there exist unique scalars

$c_1, c_2, \ldots, c_n$ such that

$$v = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n.$$

A **linear combination** of vectors $v_1, v_2, \ldots, v_n$ is any expression of the form:

$$c_1 v_1 + c_2 v_2 + \cdots + c_n v_n,$$

where $c_1, c_2, \ldots, c_n$ are scalars (real or complex numbers, depending on the vector space).

Example 1.9. The vector $\begin{bmatrix} -1 \\ 5 \end{bmatrix}$ is a linear combination of $v_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $v_2 = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$.

Let $v_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $v_2 = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$. A linear combination of these vectors is determined by,

$$c_1 v_1 + c_2 v_2 = c_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} 3 \\ -1 \end{bmatrix}.$$

If we let $c_1 = 2$ and $c_2 = -1$, we get

$$2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + (-1) \begin{bmatrix} 3 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix} + \begin{bmatrix} -3 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 5 \end{bmatrix}.$$

So the vector $\begin{bmatrix} -1 \\ 5 \end{bmatrix}$ is a linear combination of v_1 and v_2.

Example 1.10. The two vectors

$$v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

form a basis in $\mathbb{R}^2$.

That is, we can express any vector in $\mathbb{R}^2$ as a combination of these basis vectors.

Notice that:

$$\left\{ \begin{bmatrix} a \\ b \end{bmatrix} : a, b \in \mathbb{R} \right\} = \mathbb{R}^2.$$

Any vector $v = \begin{bmatrix} a \\ b \end{bmatrix}$ in $\mathbb{R}^2$ can be written as

$$v = av_1 + bv_2 = a \begin{bmatrix} 1 \\ 0 \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}.$$

The vectors v_1 and v_2 are linearly independent (no vector can be written as a multiple of the other). Since every vector in $\mathbb{R}^2$ can be written as a unique combination of v_1 and v_2, they form a basis.

Theorem 1.11. *Orthonormal Basis Expansion in a Hilbert Space Let $v_1, v_2, v_3, \ldots$ be an orthonormal basis for a Hilbert space $\mathscr{H}$. Then, for any vector $v \in \mathscr{H}$, we have*

$$v = \sum_{k=1}^{\infty} \langle v_k, v \rangle v_k.$$

1.5 Hermitian Operators

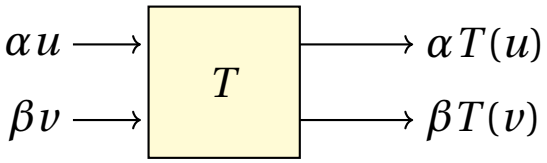
Since Hilbert spaces are vector spaces with an inner product, the self-mapping from the vector space $\mathbb{V}$ to itself is defined as a linear operator.

Definition

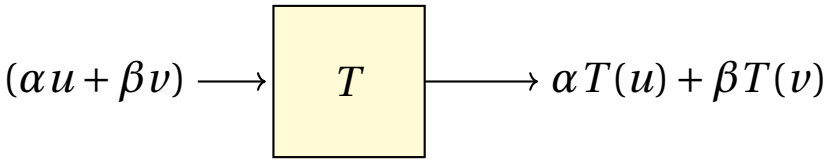
Since Hilbert spaces are vector spaces with an inner product, a self-mapping $T : \mathbb{V} \to \mathbb{V}$ on a vector space $\mathbb{V}$ is called a *linear operator* if it satisfies the following property for all $u, v \in \mathbb{V}$ and $\alpha, \beta \in \mathbb{C}$ (or $\mathbb{R}$ if $\mathbb{V}$ is over the real numbers):

$$T(\alpha u + \beta v) = \alpha T(u) + \beta T(v).$$

That is, if you have



where the arrows on the left show the scaled vectors αu and βv entering the operator T. Here, T can be viewed as a linear transformation. The arrows on the right denote the output. We say that the self mapping vector space is a linear operator if $\alpha u + \beta v$ enters the operator T and outputs a $\alpha T(u) + \beta T(v)$. That is,



Example 1.12. A Linear Operator on $L^2([0,1])$.

Consider the operator $T : L^2([0,1]) \to L^2([0,1])$ defined by,

$$(Tf)(x) = xf(x).$$

Where $(Tf)(x)$ means first apply T to f then evaluate at x. We want to check if T is linear, meaning,

$$T(\alpha f + \beta g) = \alpha T(f) + \beta T(g).$$

To check whether it's true, we apply both sides to an arbitrary input x and see if they match. So we compute,

$$(T(\alpha f + \beta g))(x).$$

Using the definition of T, we substitute f with $\alpha f + \beta g$,

$$(T(\alpha f + \beta g))(x) = x(\alpha f(x) + \beta g(x)).$$

Now we use the distributive property,

$$x(\alpha f(x) + \beta g(x)) = \alpha x f(x) + \beta x g(x).$$

By definition of T, we recognize that,

$$(Tf)(x) = xf(x), \quad (Tg)(x) = xg(x).$$

So we can rewrite the expression as,

$$\alpha(Tf)(x) + \beta(Tg)(x).$$

Since this is true for all x, we conclude,

$$T(\alpha f + \beta g) = \alpha T(f) + \beta T(g).$$

Definition

Let $\mathscr{H}$ be a Hilbert space. A linear operator $T : \mathscr{H} \to \mathscr{H}$ is **continuous** if whenever a sequence of vectors v_n in H converges to a vector v in H, then $T(v_n)$ converges to $T(v)$. That is, if

$$\lim_{n\to\infty} v_n = v,$$

then

$$\lim_{n\to\infty} T(v_n) = T(v).$$

The following theorem states that an operator is continuous if and only if it is bounded.

Theorem 1.13. *Let $\mathscr{H}$ be a Hilbert space. A linear operator $T : \mathscr{H} \to \mathscr{H}$ is **continuous** if and only if there exists a constant M such that for all vectors $v \in \mathscr{H}$, we have*

$$\|T(v)\| \leq M\|v\|.$$

It is important to understand that every matrix can be viewed as a function. Indeed, if we had a matrix $A^{n\times m}$ then we can define a function $A : \mathbb{R}^n \to \mathbb{R}^m$ by $A(\vec{x}) = A\vec{x}$. Likewise, for the complex space $A : \mathbb{C}^n \to \mathbb{C}^m$ by $A(\vec{x}) = A\vec{x}$.

Example 1.14. Consider the matrix $A = \begin{bmatrix} 2 & 3 \\ 5 & 6 \end{bmatrix}$.

We can define the mapping $A:\mathbb{R}^2 \to \mathbb{R}^2$ by

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \to A \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 + 3x_2 \\ 5x_1 + 6x_2 \end{bmatrix}.$$

For the transpose, $A^T = \begin{bmatrix} 2 & 5 \\ 3 & 6 \end{bmatrix}$, and the function is defined as

$A^T:\mathbb{R}^2 \to \mathbb{R}^2$ by

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \to A^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 + 5x_2 \\ 3x_1 + 6x_2 \end{bmatrix}.$$

In quantum mechanics, we aim to analyze whether we can interchange the operator A in the inner product $\langle Ax, y\rangle$. Specifically, we are interested in whether we can move A to act on the vector y instead of x. In other words, can we express this as $\langle x, Ay\rangle$? Note that $\langle Ax, y\rangle$ means that we first apply the function to x and then do the dot product with y.

Theorem 1.15. *Let A be an $n \times n$ real matrix and let x, y be vectors in $\mathbb{R}^n$. Then, $\langle Ax, y\rangle = \langle x, A^T y\rangle$, where A^T is the transpose of A.*

Before we tackle the proof, we first notice that $z \cdot w = z^T \cdot w$ since

$$\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} z_1 & z_2 & \cdots & z_n \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}.$$

Also, recall that for any two matrix $(AB)^T = B^T A^T$. Shoes and socks, or so my professor once called it.

Proof.

$$\begin{aligned}\langle Ax, y\rangle &= (Ax)\cdot y \\ &= (Ax)^T y \\ &= x^T A^T y \\ &= (x^T)(A^T y) \\ &= (x^T)(Ay) \\ &= x(Ay) \\ &= \langle x, Ay\rangle\end{aligned}$$

□

It can be demonstrated that if B is a matrix satisfying $\langle Ax, y\rangle = \langle x, By\rangle$ for all vectors x and y, then B must equal A^T. This property establishes the uniqueness of the matrix that fulfills this condition.

Specifically, our interest lies in finding the matrix B that meets the equation $\langle Ax, y\rangle = \langle x, A^T y\rangle$. In the context of the vector space $\mathbb{R}^n$, the matrix B that satisfies this equality is indeed A^T. However, when the vector space is extended to $\mathbb{C}^n$, the appropriate matrix B is the conjugate transpose of A, denoted as $A^* = \overline{A}^T$. This concept will be elaborated upon in the subsequent theorem. In general, this B is called the adjoint of A.

The correct expression for the dot product in complex vector spaces is given by

$$z \cdot w = \overline{z}^T \cdot w.$$

Here, $\overline{z}$ denotes the complex conjugate of each component in the vector z. This conjugate is necessary because squaring a complex number (as occurs implicitly in dot product calculations) can result in a negative real part, which is not suitable when calculating lengths or angles.

For example, consider a vector $z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$ in a complex space. The squared length of

z, which should be a non-negative real number, is calculated as:

$$\begin{aligned}\|z\|^2 &= \overline{z}^T z \\ &= \overline{z_1} z_1 + \overline{z_2} z_2 + \ldots + \overline{z_n} z_n \\ &= |z_1|^2 + |z_2|^2 + \ldots + |z_n|^2.\end{aligned}$$

Theorem 1.16. *Let $\mathscr{H}$ be a Hilbert space. Given any continuous linear operator $T : \mathscr{H} \to \mathscr{H}$, there exists an* ***adjoint operator***

$$T^* : \mathscr{H} \to \mathscr{H}$$

such that for all $v, w \in \mathscr{H}$, we have

$$\langle w, T(v) \rangle = \langle T^*(w), v \rangle.$$

In functional analysis, the adjoint operator of a linear operator $T : \mathscr{H} \to \mathscr{H}$ in a Hilbert space $\mathscr{H}$ is another operator T^* that transfers T from one side of an inner product to the other. This equation essentially means that applying T to v and taking the inner product with w is the same as applying T^* to w and taking the inner product with v.

Recall that the transpose of a matrix

$$A = \begin{bmatrix} a & b-i \\ c+i & d \end{bmatrix}$$

is

$$A^T = \begin{bmatrix} a & c+i \\ b-i & d \end{bmatrix}.$$

A^* is the conjugate transpose. That is we first conjugate our matrix A,

$$\overline{A} = \begin{bmatrix} a & b+i \\ c-i & d \end{bmatrix}$$

and then take the transpose of this

$$\overline{A}^T = \begin{bmatrix} a & c-i \\ b+i & d \end{bmatrix}.$$

Example 1.17. Consider $\mathbb{C}^n$ with the standard inner product,

$$\langle x, y \rangle = \sum_{i=1}^{n} \overline{x_i} y_i.$$

The adjoint operator of a matrix A is the conjugate transpose A^*, meaning,

$$\langle x, Ay \rangle = \langle A^* x, y \rangle.$$

For example, if

$$A = \begin{bmatrix} 1 & i \\ 2 & 3 \end{bmatrix},$$

then the adjoint (conjugate transpose) is

$$A^* = \begin{bmatrix} 1 & 2 \\ -i & 3 \end{bmatrix}.$$

If you noticed, $A^* = A^T$ which is the transpose.

Definition

A linear operator $T : \mathscr{H} \to \mathscr{H}$ is **Hermitian** if

$$T = T^*.$$

Essentially, we want our matrix A to be the same as first conjugating A and then transposing it.

Example 1.18. Consider $A : \mathbb{C}^n \to \mathbb{C}^n$ where $A = \begin{bmatrix} 4 & 2+i \\ 2-i & 3 \end{bmatrix}$.

One can easily check that A is Hermitian.

$$A = \begin{bmatrix} 4 & 2+i \\ 2-i & 3 \end{bmatrix}$$

First, take the complex conjugate of each entry.

$$\overline{A} = \begin{bmatrix} 4 & 2-i \\ 2+i & 3 \end{bmatrix}$$

Next, transpose the matrix.

$$A^* = \begin{bmatrix} 4 & 2+i \\ 2-i & 3 \end{bmatrix}$$

Now compare A^* with A and notice that they are equal.

$$A^* = \begin{bmatrix} 4 & 2+i \\ 2-i & 3 \end{bmatrix}, \quad A = \begin{bmatrix} 4 & 2+i \\ 2-i & 3 \end{bmatrix}$$

Since $A^* = A$, we conclude that A is Hermitian.
We can define this concept across three different scenarios:

1. In the finite-dimensional space $\mathbb{R}^n$, if $A = A^T$, we say that A is *symmetric*.

2. In the finite-dimensional space $\mathbb{C}^n$, if $A = A^*$, we say that A is *Hermitian*. The notation A^* represents the conjugate transpose of A.

3. In an infinite-dimensional space, if an operator T satisfies $T = T^*$ for all points in its domain, where T^* is the adjoint of T, then T is described as *self-adjoint*.

Eventually we will see that a basic assumption in quantum mechanics is that anything that can be measured must correspond to a Hermitian operator. Additionally, the measurements which are numbers obtained in a lab, must be an eigenvalue of the operator. These numbers measured are in the spectrum of the operator. However, for now, we will strictly focus on eigenvalues.

In order for our measurements to have physical meaning, these measurements should yield real numbers. Note that the eigenvalues of a linear operator acting on

a complex vector space could be complex numbers, but the eigenvalues of a Hermitian operator must always be real. Indeed this very property is crucial in quantum mechanics because observable quantities, such as energy, momentum, and angular momentum, must have real values in any physical experiment.

To see why this must be the case, recall that in a Hilbert space $\mathscr{H}$ with inner product $\langle \cdot, \cdot \rangle$, an operator $T : \mathscr{H} \to \mathscr{H}$ is Hermitian if it satisfies the condition,

$$\langle w, Tv \rangle = \langle Tw, v \rangle, \quad \text{for all } v, w \in \mathscr{H}.$$

Now, suppose v is an eigenvector of a Hermitian operator T with eigenvalue λ, meaning,

$$Tv = \lambda v.$$

Taking the inner product of both sides with v, we obtain,

$$\langle v, Tv \rangle = \lambda \langle v, v \rangle.$$

Since T is Hermitian. Likewise, by the linearity of the inner product (and noting that λ is a scalar), we can pull λ out. I.e $\langle v, v\lambda \rangle = \lambda \langle v, v \rangle$. Now, notice,

$$\langle Tv, v \rangle = \langle v, Tv \rangle.$$

Thus,

$$\lambda \langle v, v \rangle = \langle v, Tv \rangle = \overline{\lambda} \langle v, v \rangle.$$

We can concluded that $\lambda = \overline{\lambda}$. Indeed the inner product $\langle v, v \rangle$ is always nonzero for nonzero vectors v. This means that λ is a real number. That said, this ensures that all of the measurements in quantum mechanics corresponds to a real number. We will elaborate on this and provide a more detail proof shortly.

In applications, operators such as the Hamiltonian H (which represents energy), the momentum operator P, and the angular momentum operator L are all Hermitian. This guarantees that energy levels, momenta, and angular momenta in quantum systems have real number. Therefore it makes it measurable unlike a complex number.

The requirement that observable quantities correspond to Hermitian operators is not just a mathematical convenience but a necessity for the consistency of quantum mechanics with real-world experiments.

In the Hilbert space $\mathscr{H}$, not all elements necessarily work with certain operator properties. To fix this, we can shrink our domain of an operator T to all the values that do comply with these properties. Let's denote these as values as $D(T)$. Similarly, we examine $D(T^*)$, which are both subsets of $\mathscr{H}$. Here, $T : D(T) \subseteq \mathscr{H} \to \mathscr{H}$ and $T^* : D(T^*) \subseteq \mathscr{H} \to \mathscr{H}$.

If $T : D(T) \subseteq \mathscr{H} \to \mathscr{H}$ satisfies $D(T) = D(T^*)$ and $T = T^*$, then T is called *self-adjoint*.

Likewise, if $T : D(T) \subseteq \mathscr{H} \to \mathscr{H}$ satisfies $D(T) \subseteq D(T^*)$ and $T = T^*$, then we referred it as *symmetric*. Do note that "symmetric" is used interchangeably with "self-adjoint" for bounded operators where $D(T) = \mathscr{H}$.

Lets consider an example using L^2. Choose the operator T on the space $L^2[0,1]$ by $T(x) = \frac{1}{x-1}$. Do you see the issue? We encounter problems when defining its domain due to $x = 1$. That is, we could specify $D(T)$ as $C[0,1]$ which if you recall is the space of continuous functions on $[0,1]$. Our restriction allows us to avoid the singularity at $x = 1$. Additionally, we considering an operator like $T = \frac{d}{dx}$, one would need to carefully define $D(T)$. One must make sure that all functions within it are differentiable. In this case it is $C^1[0,1]$ which is set of continuously differentiable functions on $[0,1]$.

Definition

Let A be an $n \times n$ matrix. A **nonzero** vector v is called an **eigenvector** of A if there exists a scalar λ such that,

$$Av = \lambda v.$$

The scalar λ is called an **eigenvalue** corresponding to the eigenvector v.

Notice that this equation means that when A acts on v. The set of all eigenvalues of A is called the **spectrum** of A.

To find eigenvalues of a matrix A, we solve

$$\det(A - \lambda I) = 0.$$

where I is the identity matrix of the same size as A, λ is the unknown eigenvalue, and $\det(A-\lambda I)$ is the determinant of $A-\lambda I$. Once we find an eigenvalue λ, we find the corresponding eigenvectors by solving for

$$(A-\lambda I)v = 0.$$

This gives a system of linear equations. Any nonzero solution v is an eigenvector corresponding to λ.

Example 1.19. Give the eigenvalues and vectors of

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

$$\det(A-\lambda I) = 0.$$

$$\det \begin{bmatrix} 4-\lambda & 2 \\ 1 & 3-\lambda \end{bmatrix} = 0.$$

Now we expand the determinant.

$$(4-\lambda)(3-\lambda) - (2\cdot 1) = 0.$$

$$12 - 4\lambda - 3\lambda + \lambda^2 - 2 = 0.$$

$$\lambda^2 - 7\lambda + 10 = 0.$$

Factoring the left-hand side.

$$(\lambda - 5)(\lambda - 2) = 0.$$

Thus, the eigenvalues are $\lambda_1 = 5$ and $\lambda_2 = 2$.
Now we find the the Eigenvectors. For $\lambda = 5$,

$$(A-5I)v=0.$$

$$\left(\begin{bmatrix}4 & 2\\ 1 & 3\end{bmatrix}-5\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\right)v=0.$$

$$\begin{bmatrix}-1 & 2\\ 1 & -2\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix}=\begin{bmatrix}0\\ 0\end{bmatrix}.$$

Solving $-x+2y=0$, we get $x=2y$. Choosing $y=1$, we get $x=2$, so one eigenvector is.

$$v_1=\begin{bmatrix}2\\ 1\end{bmatrix}.$$

For $\lambda=2$,

$$(A-2I)v=0.$$

$$\left(\begin{bmatrix}4 & 2\\ 1 & 3\end{bmatrix}-2\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\right)v=0.$$

$$\begin{bmatrix}2 & 2\\ 1 & 1\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix}=\begin{bmatrix}0\\ 0\end{bmatrix}.$$

Solving $2x+2y=0$, we get $x=-y$. Choosing $y=1$, we get $x=-1$, so another eigenvector is.

$$v_2=\begin{bmatrix}-1\\ 1\end{bmatrix}.$$

In general, a matrix being real does not necessarily mean that its eigenvalues are are real. Consider the the following matrix.

Consider the matrix $B=\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}$.

$$\det(A - \lambda I) = 0.$$

$$\det \begin{bmatrix} 0-\lambda & 1 \\ -1 & 0-\lambda \end{bmatrix} = 0.$$

Now we expand the determinant.

$$(-\lambda)(-\lambda) - (1 \cdot -1) = 0.$$

$$\lambda^2 + 1 = 0.$$

Solving for λ.

$$\lambda^2 = -1.$$

$$\lambda = \pm i.$$

Thus, the eigenvalues are $\lambda_1 = i$ and $\lambda_2 = -i$.

Therefore, both of our eigenvalues are non real numbers. This makes our next theorem that more important. That is, if we have the symmetric (Hermitian) property, then we can be certain that the eigenvalues are real.

Theorem 1.20. *Spectrum part* 1. *The eigenvalues of a Hermitian operator are real numbers.*

Proof. Let λ be an eigenvalue with eigenvector v for a Hermitian operator $T : \mathscr{H} \to \mathscr{H}$. By definition, this means that

$$T(v) = \lambda v.$$

Now, consider the inner product,

$$\begin{aligned}
\lambda\langle v, v\rangle &= \langle v, \lambda v\rangle \\
&= \langle v, T(v)\rangle \\
&= \langle T^*(v), v\rangle, \quad \text{by the definition of the adjoint} \\
&= \langle T(v), v\rangle, \quad \text{since } T \text{ is Hermitian, so } T^* = T \\
&= \langle \lambda v, v\rangle \\
&= \lambda\langle v, v\rangle.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\langle \lambda v, v\rangle &= \lambda\langle v, v\rangle, \\
\langle v, \lambda v\rangle &= \overline{\lambda}\langle v, v\rangle.
\end{aligned}$$

Since $\langle v, v\rangle \neq 0$ for a nonzero eigenvector v, we must have

$$\lambda = \overline{\lambda}.$$

This implies that λ is a real number, $\lambda \in \mathbb{R}$, as desired. □

Theorem 1.21. *Spectrum part 2. Let V be a finite-dimensional vector space, let T be a linear operator, then the eigenvectors of T can be used to form an orthonormal basis for V.*

Proof. Since v and w are eigenvectors of T, we have

$$T(v) = \lambda v, \text{ and } T(w) = \mu w.$$

Now, consider the inner product

$$\begin{aligned}
\lambda\langle w, v\rangle &= \langle w, \lambda v\rangle \\
&= \langle w, T(v)\rangle \\
&= \langle T^*(w), v\rangle, \quad \text{by the definition of the adjoint} \\
&= \langle T(w), v\rangle, \quad \text{since } T \text{ is Hermitian, so } T^* = T \\
&= \langle \mu w, v\rangle \\
&= \mu\langle w, v\rangle.
\end{aligned}$$

Thus, we obtain the equation,

$$\lambda\langle w, v\rangle = \mu\langle w, v\rangle$$

Rearranging,

$$(\lambda - \mu)\langle w, v\rangle = 0.$$

Since we are given that $\lambda \neq \mu$, it follows that,

$$\langle w, v\rangle = 0.$$

Thus, the eigenvectors v and w are orthogonal, as required. □

To see the importance, if we wanted to construct an orthogonal basis for $\mathbb{R}^n$, we have to do the following three steps. First, we consider a symmetric matrix A, and then we find the eigenvalues λ_1 and λ_2 of A. Lastly, we find one eigenvector, call it $\mathbf{e}$, for λ_1.

We will construct an example in $\mathbb{R}^2$, and you will see how difficult these things can get.

Example 1.22. Find the basis in $\mathbb{R}^2$ with respect to the matrix $A = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$.

$$\det(A-\lambda I)=0.$$

$$\det\begin{bmatrix}1-\lambda & 3\\ 3 & 1-\lambda\end{bmatrix}=0.$$

Now we expand the determinant.

$$(1-\lambda)(1-\lambda)-(3\cdot 3)=0.$$

$$(1-\lambda)(1-\lambda)-9=0.$$

$$1+\lambda^2-2\lambda-9=0.$$

$$\lambda^2-2\lambda-8=0.$$

Factoring.

$$(\lambda-4)(\lambda+2)=0$$

Thus, the eigenvalues are $\lambda_1=4$ and $\lambda_2=-2$.

To find the eigenvector of λ_2 we will solve the system

$$[A-\lambda_2 I|0]=\left(\begin{array}{cc|c}-3 & 3 & 0\\ 3 & -3 & 0\end{array}\right)$$

Therefore, the corresponding system of equations is $-3x_1+3x_2=0$ then $x_1=x_2$. Here

$$\begin{aligned}w_2&=\begin{bmatrix}x_1\\ x_2\end{bmatrix}\\ &=\begin{bmatrix}x_2\\ x_2\end{bmatrix}\\ &=x_2\begin{bmatrix}1\\ 1\end{bmatrix}\end{aligned}$$

Where x_2 can be any non zero real numbers.

Likewise, finding the eigenvector of λ_1 we solve the system

$$[A-\lambda_1 I|0] = \left(\begin{array}{cc|c} 3 & 3 & 0 \\ 3 & 3 & 0 \end{array}\right)$$

Therefore, the corresponding system of equations is $3x_1+3x_2=0$ then $x_1=-x_2$. Here

$$\begin{aligned} w_1 &= \begin{bmatrix} -x_2 \\ x_2 \end{bmatrix} \\ &= x_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix} \end{aligned}$$

Where x_2 can be any non zero real numbers. Thus, w_1, w_2 forms a basis for $\mathbb{R}^2$.

Definition

Let $T:\mathscr{H}\to\mathscr{H}$ be a linear operator. A complex number λ is in the spectrum of T if the operator is the set of all possible eigenvalues that the operator can produce when acting on a quantum state.

In other words the spectrum of T can be though of all the eigenvalues of T. The eigenvector cannot be the zero vector which implies that the equation $(T-\lambda I)(v)=0$ has a nontrivial solution. In other words, $T-\lambda I$ has a nontrivial kernel, i.e., there exists a vector $v\neq 0$ such that

$$(T-\lambda I)\,v = 0.$$

Recall that if you have two sets $f:A\to B$, the kernel is all of the elements in A such that $f(a)$ is mapped to the identity in B. A matrix (or operator) that has a nontrivial kernel is not invertible because an invertible operator has a unique solution for every input. This means that its kernel consists only of the zero vector. Since $T-\lambda I$ is not invertible, this tells us that

$$\lambda \in \text{Spectrum}(T).$$

We then conclude that

Eigenvalues of $T \subseteq$ Spectrum of T.

This means that every eigenvalue of T is a part of the spectrum of T, but there might be elements in the spectrum that are not eigenvalues. The measurements in quantum mechanics correspond to operators on a Hilbert space. The possible outcomes of a measurement have to lie in the spectrum of the corresponding operator.

2 Quantum Mechanics Intuition

2.1 Physics Review

Before we advance, I would like to list several elementary properties.

Definition

Kinetic energy is the energy that an object possesses due to its motion. It depends on the velocity of the object and its mass. The faster an object moves, or the more massive it is, the more kinetic energy it has. The mathematical formula for kinetic energy (KE) is $KE = \frac{1}{2}mv^2$ where m is the mass of the object, and v is the velocity of the object.

Definition

Potential energy is the energy an object possesses due to its position relative to a gravitational field. It is defined as the work done against gravity to bring an object to a given point in a gravitational field. The formula for gravitational potential energy (PE) near the Earth's surface is $PE = mgh$ where , m is the mass of the object, g is the acceleration due to gravity ($9.81\,\mathrm{m/s^2}$ on Earth) and h is the height above the reference point.

We can think of potential energy quite literally. If someone lifts a hammer and holds it up, preparing for a strike, then the hammer possesses potential energy. However, once the person swings the hammer, it transitions to kinetic energy.

Definition

The Planck constant represents the quantum of action or the smallest possible action in physical processes. The value of the Planck constant is approximately $6.62607015 \times 10^{-34}$ joule-seconds (Js)

Definition

It's important to recall that the frequency of light is related to its wavelength λ and the speed of light c by the equation $f = \frac{c}{\lambda}$. This relationship is fundamental in connecting the concepts of energy, frequency, and wavelength in the study of

electromagnetic radiation.

Definition

A newton is defined as the force required to accelerate one kilogram of mass at a rate of one meter per second squared. The formula to express this relationship is $F = ma$ where F is the force in newtons, m is the mass in kilograms, and a is the acceleration in meters per second squared. On Earth, the gravitational force exerted on a 1 kg object is approximately 9.8 newtons, because the acceleration due to Earth's gravity is approximately 9.8 meters per second squared.

Definition

Recall that one joule is defined as the amount of energy transferred when applying one newton of force over a distance of one meter. In formula terms, it is $1\,\text{J} = 1\,\text{N} \cdot 1\,\text{m}$.

Definition

Momentum is the product of the mass of a particle and its velocity, $p = mv$. Momentum is a vector; that is, it has both magnitude and direction.

Definition

Velocity is a vector quantity that refers to the rate at which an object changes its position. It has both magnitude and direction. If $\vec{s}$ represents displacement and Δt represents the time interval, then velocity $\vec{v}$ is defined as $\vec{v} = \frac{\vec{s}}{\Delta t}$.

2.2 Light as a Wave and a Particle

What is light? This is the very question that stumped scientists for thousands of years. You might think that light is the thing that illuminates from your screen as you are reading (if you are reading this on your laptop). And indeed, you are correct, but that is a specific light called a light source. Other such light sources include stars such as our sun. The sun emits light that travels millions of miles, arriving at Earth in roughly 8 minutes. On the other hand, everything that you see is reflecting light, however, it is

not a light source. For example, consider yourself. You are not a light source, but the light that travels from the sun hits you and reflects off you, and thus when someone sees you, they really see the light that you emit. Likewise, if you are reading this from a physical paper, then you are seeing the light that the paper reflects from the sun. An easy way to think of a light source is like this: If I were to turn off the light, does this very thing emit light? In terms of the physical paper and you yourself, the answer is no.

Throughout history, there has been an on going debate over whether light is a particle or a wave. In the 17th century, Isaac Newton theorized that light was made up of tiny particles. Newton's theory explained how light travels in straight lines and casts sharp shadows. Shortly after, Christiaan Huygens opposed Newton's ideas. Huygens suggested that light was a wave. However, his ideas did not gain as much attention compared to Newton's theories.

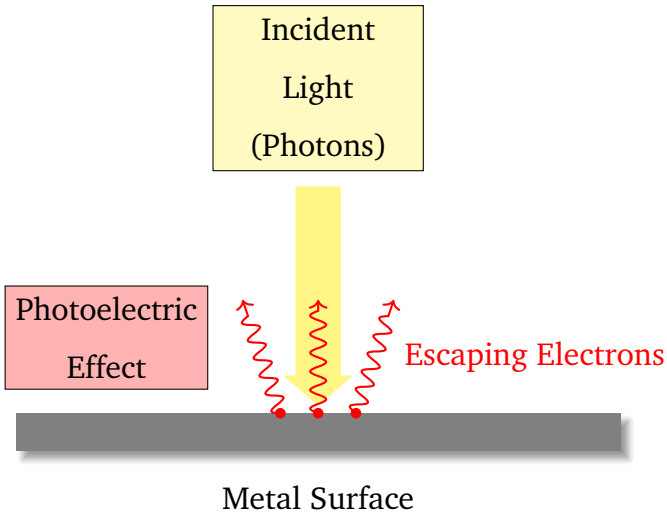
Jumping in time, in the 19th century, Thomas Young conducted the double-slit experiment. This experiment involved shining light or particles through two closely spaced slits onto a screen positioned on the other side of the slits. These particles then pass through the slits and hit a screen in which the collision can be recorded. What was observed was that instead of seeing two distinct lines or spots correlating to the slits, the screen displayed a series of bright and dark bands (known as interference fringes.) The bright bands occur where the waves from the two slits arrive in phase and constructively interfere. In other words, the peaks and troughs of the waves align. The dark bands occur where the waves arrive out of phase and destructively interfere, with the peaks of one wave aligning with the troughs of the other. One might expect particles to behave differently from waves when the same experiment is conducted with particles (electrons). Surely the particle will pass through one slit or the other, striking the screen in two places aligned with the slits. However, even when electrons are fired one at a time, an interference pattern that mimics that of waves gradually appears on the screen.

In the mid-19th century, Maxwell's equations described light as electromagnetic waves, electricity, magnetism, and optics. Not long after, Albert Einstein revisited the particle theory with a twist. He provided an explanation for the photoelectric effect which won him the Nobel Prize in Physics. He proposed that light could be thought of as energy, or 'photons,' which had particle like properties. Einstein introduced the

formula $E = h\nu$, where h is the Planck constant and f is the frequency of light.

Niels Bohr, Werner Heisenberg, and others further developed this concept. They introduced the idea that light and matter exhibit both wave-like and particle-like properties which is now a fundamental aspect of quantum theory.

Shining light on top of certain metal objects forces electrons in the metal to escape. This act is called the photoelectric effect and is a specific property of light. Since light is an electromagnetic wave and these waves transmit energy, it is indeed able to push electrons out of metal. However, at a quantum level, this model fails.

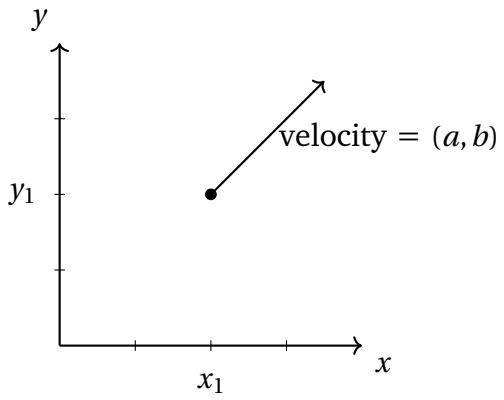


Our goal is to study the behaviors of particles. We then can apply mathematics to describe how a particle is influenced by its environment and to predict its changes over time and space. To illustrate this, consider a galaxy with mass m positioned at (x_1, y_1) and moving with velocity (a, b). We know both its position and momentum. Typically, the analysis requires knowledge of six parameters: three spatial coordinates and three corresponding momentum coordinates. Together, these form the *state* of the galaxy, represented as a point in $\mathbb{R}^6$:

$$\left(x, y, z, m\frac{dx}{dt}, m\frac{dy}{dt}, m\frac{dz}{dt}\right) \in \mathbb{R}^6.$$

Here, x, y, z are the position coordinates, and the other three components represent the momentum corresponding to each position.

To understand the galaxy's motion through time, we turn to Newton's second law, $F = ma$. Solving this involves identifying all the forces F acting on the galaxy and solving the corresponding differential equations.



We will start by stating six fundamental postulates.

Postulate 1

The state of any object in a Hilbert space is determined by a ray. That is, the state of any of these objects is specified by a non zero vector v in the Hilbert space. We say that two vectors specify the same state if they are products of each other and nonzero.

In general, a ray is a line that connects all the points in the Hilbert space. Unlike classical mechanics, where a point lives in $\mathbb{R}^6$, here we work in a Hilbert space $\mathscr{H}$. We normalize this ray, meaning that any two points on the same ray represent the same physical state.

We only consider the points of length one, those with norm equal to 1. Doing so lets us construct a unit circle (more generally a unit sphere in $\mathscr{H}$). This circle contains all the possible states. However, each physical state corresponds to exactly one point on this circle.

Each point is a square integrable function. The state of any system is given by one of these functions with norm 1, known as the wave function. While you could represent a state as any point on the ray, we always normalize it to have length one.

So in summary, the points on this unit circle represent the possible states of the system. In classical mechanics, a state is a point in $\mathbb{R}^6$, such as $\left(x, y, z, m\frac{dx}{dt}, m\frac{dy}{dt}, m\frac{dz}{dt}\right) \in \mathbb{R}^6$. In quantum mechanics, a state is a normalized vector in $\mathscr{H}$. This postulate tells us that the state of the system is described by a point of length 1 in the Hilbert space.

A ray is the set of all vectors in the Hilbert space that all represent the same state of a quantum system and differ only by a scalar. More specifically, a ray is a one dimensional

$1-D$ subspace of a Hilbert space. It is defined by taking all scalar multiples (including complex multiples) of a non-zero vector v.

The ray $\mathscr{R}$ that corresponds to any none zero $v \in \mathscr{H}$ is given by,

$$\mathscr{R} = \{cv : c \in \mathbb{C} \text{ and } c \neq 0\}.$$

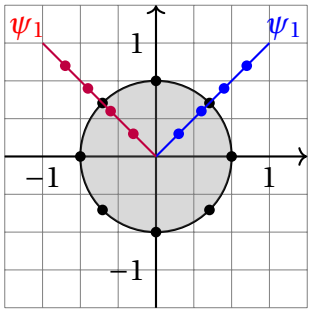
Note that the physical state of an object is not determined by the magnitude or the phase of the vector. Thus, all vectors in a ray describe the same physical state. Therefore we concluded that states are identified with rays in Hilbert space rather than with individual vectors.

To sum up this repetition, the state of a system is represented by a ray in a Hilbert space. That is a ray is a collection of vectors that differ only by a nonzero complex scalar. More importantly, they lie along the same line through the origin. This is quite different from classical mechanics, where the state of a particle is represented by a single point in $\mathbb{R}^6$.

To work with quantum states more easily, we usually normalize these vectors so that their norm is 1. Any vector on the same ray (any scalar multiple) represents the same physical state. To simplify calculations we pick the one with unit norm one. These normalized vectors form a unit sphere in the Hilbert space.

Each point on this unit sphere corresponds to a possible quantum state. Mathematically, these are square-integrable functions of norm 1, i.e $||\psi||_2 = \int |\psi|^2 = 1$. You can as think of them as wave functions. The full ray contains infinitely many such functions differing by scalar multiples. However, we treat them all the same.

So while a state could be described by any nonzero vector in the Hilbert space, we typically consider only those on the unit sphere because normalization doesn't change the physical state.



Theorem 2.1. *If ψ is any non zero vector in a Hilbert space $\mathscr{H}$, then $\frac{\psi}{\|\psi\|}$ is a unit vector.*

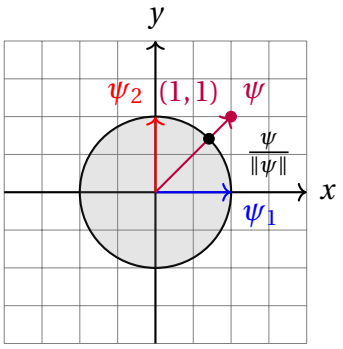
Proof.

$$\begin{aligned}\left\| \frac{\psi}{\|\psi\|} \right\| &= \left\| \frac{1}{\|\psi\|}\psi \right\| \\ &= \frac{1}{\|\psi\|}\|\psi\| \\ &= 1\end{aligned}$$

□

Assume for a moment that we are in the Hilbert space $\mathbb{R}^2$. We have a unit sphere, similar to the one above. However, what if we want to find the corresponding state to the point $(1,1)$, which is not on the unit circle? We do what was just said, that is, we normalize it. But how do we do this? Consider the two vectors $\psi_1 = \begin{bmatrix}1\\0\end{bmatrix}$ and $\psi_2 = \begin{bmatrix}0\\1\end{bmatrix}$. We define ψ as $\psi = 1\cdot\psi_1 + 1\cdot\psi_2$, and we compute $\frac{\psi}{\|\psi\|}$. Notice that $\|\psi\| = \sqrt{1^2+1^2} = \sqrt{2}$. That is,

$$\frac{\psi}{\|\psi\|} = \frac{1}{\sqrt{2}}\psi_1 + \frac{1}{\sqrt{2}}\psi_2 = \begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix}.$$



In quantum mechanics, a general state can be written as a linear combination: $\psi = \lambda_1\psi_1 + \cdots + \lambda_n\psi_n$, where each ψ_i is a basis state and λ_i is a (possibly complex) coefficient.

If the state is not normalized, the probability of measuring the system in the state ψ_1 is given by $\frac{|\lambda_1|^2}{|\lambda_1|^2+\cdots+|\lambda_n|^2}$.

If the state is normalized, that is, $|\lambda_1|^2 + \cdots + |\lambda_n|^2 = 1$, then the probability of measuring ψ_1 is simply $|\lambda_1|^2$.

Postulate 2

Any measurable quantity corresponds to a Hermitian operator $A : \mathscr{H} \to \mathscr{H}$

Therefore, if we wish to measure the velocity of a particle then we must have a velocity operator. Usually, we denote p as the momentum operator and m as mass. Together we $v = \frac{p}{m}$

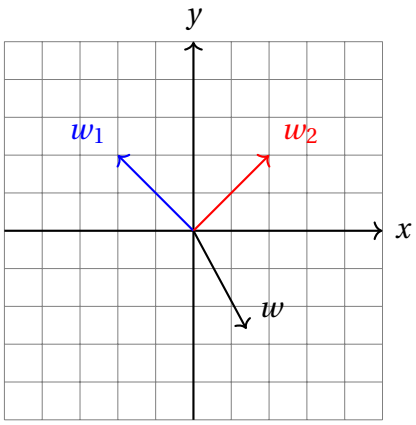
Postulate 3

An actual measurement will always be in the spectrum of the corresponding Hermitian operator.

Recall, that we will denote the set of all eigenvalues of A as the spectrum of A. Therefore, if λ is an eigenvalue of A, then $\lambda \in \text{Spectrum}(A)$. The spectrum of an operator A is not a subset of the set A itself. If A is a set of matrices then the Spectrum of A is just a set of scalars, which are the eigenvalues of those matrices.

Unlike classical physics, quantum mechanics utilizes a probabilistic approach. This framework uses linear operators on Hilbert spaces to represent physical quantities. Therefore, the next postulate allows us to leave classical mechanics behind and begin anew.

Let's bring back example 1.22. We found that the eigenvectors w_1 and w_2 were orthogonal with each other. That is $e_1 \cdot e_2 = 0$ (forms a 90 degree angle.)



Here w can be written as a linear combination of w_1 and w_2 where each of the coefficients of each eigenvector is a probability. That is $w = a_1 w_1 + a_2 w_2$.

If w represents a quantum state, then the squared magnitudes of the coefficients, $|a_1|^2$ and $|a_2|^2$, can be interpreted as probabilities.

Postulate 4

Let a particle have state $v \in \mathscr{H}$. Let A be a Hermitian operator with a eigenvalue λ and its corresponding eigenvector $w \in \mathscr{H}$. The probability for getting a number λ for a measurement on the particle corresponding to A is

$$\frac{|\langle w, v\rangle|^2}{\langle w, w\rangle\langle v, v\rangle}.$$

If $\langle w, w\rangle$ and $\langle v, v\rangle$ were both unit vectors, then the bottom normalizes to $\langle w, w\rangle = 1$ and $\langle v, v\rangle = 1$. That is, the simplified version is just $|\langle w, v\rangle|^2$.

Let's give a very simple example.

Example 2.2. Let's consider a Hermitian matrix $A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ and the Hilbert space $\mathbb{C}^2$.

Let's compute the eigenvalues.

$$\det(A - \lambda I) = 0.$$

$$\det \begin{bmatrix} 1-\lambda & 0 \\ 0 & 2-\lambda \end{bmatrix} = 0.$$

Now we expand the determinant.

$$(1-\lambda)(2-\lambda) - (0 \cdot 0) = 0.$$

$$(1-\lambda)(2-\lambda) = 0.$$

Factoring.

$$\lambda = 1 \quad \text{or} \quad \lambda = 2$$

Thus, the eigenvalues are $\lambda_1 = 1$ and $\lambda_2 = 2$. Thus the eigenvector that corresponds to λ_1

is $w = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Suppose that our system was in the state $v = \begin{bmatrix} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix}$. We can now calculate

$$\frac{|\langle w, v \rangle|^2}{\langle w, w \rangle \langle v, v \rangle}$$

.

Indeed,

$$\begin{aligned} |\langle w, v \rangle| &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix} \\ &= 1 \cdot \frac{1}{2} + 0 \cdot \frac{\sqrt{3}}{2} \\ &= \frac{1}{2} \end{aligned}$$

Therefore $|\langle w, v \rangle|^2 = \frac{1}{4}$. Now,

$$\begin{aligned} \langle w, w \rangle &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ &= 1 \cdot 1 + 0 \cdot 0 \\ &= 1 \end{aligned}$$

and

$$\begin{aligned} \langle v, v \rangle &= \begin{bmatrix} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix} \\ &= \frac{1}{2} \cdot \frac{1}{2} + \frac{\sqrt{3}}{2} \cdot \frac{\sqrt{3}}{2} \\ &= \frac{1}{4} + \frac{3}{4} \\ &= 1 \end{aligned}$$

Together,

$$\begin{aligned}\frac{|\langle w, v\rangle|^2}{\langle w, w\rangle\langle v, v\rangle} &= \frac{\frac{1}{4}}{1\cdot 1}\\ &= \frac{1}{4}.\end{aligned}$$

Thus, if the system is in the state $v = \begin{bmatrix}\frac{1}{2}\\ \frac{\sqrt{3}}{2}\end{bmatrix}$, there is a 25% chance of measuring the eigenvalue $\lambda_1 = 1$ and a 75% of getting the other eigenvector. It is important to know that these are all elementary examples. We had to create certain conditions and assume that our Hilbert space was finite dimensional.

There are two types of bases that we prefer. The first type is when you can find a basis where the elements are pairwise perpendicular (i.e., their inner product is 0). However, the preferred case is when, in addition to being pairwise perpendicular, the elements of the basis also have unit length. We call this an orthonormal basis.

Suppose that $\{\psi_1,\ldots,\psi_n\}$ is an orthogonal basis for $\mathbb{R}^n$. Since $\{\psi_1,\ldots,\psi_n\}$ is a basis, any vector $x \in \mathbb{R}^n$ can be written as a linear combination $x = \alpha_1\psi_1 + \cdots + \alpha_n\psi_n$ for some scalars α_i.

Given a vector x, how can we find $\alpha_1,\ldots,\alpha_n$? If the basis $\{\psi_1,\ldots,\psi_n\}$ is orthogonal, then each α_i can be found by $\alpha_i = \frac{\langle\psi,\psi_i\rangle}{\langle\psi_i,\psi_i\rangle}$.

$$\begin{aligned}\alpha_1 &= \frac{\langle x, \psi_1\rangle}{\langle\psi_1,\psi_1\rangle}\\ &= \frac{\langle x, \psi_1\rangle}{\|\psi_1\|^2}\\ &\vdots\\ \alpha_n &= \frac{\langle x, \psi_n\rangle}{\langle\psi_n,\psi_n\rangle}\\ &= \frac{\langle x, \psi_n\rangle}{\|\psi_n\|^2}.\end{aligned}$$

Let's prove this.

Proof. Let $x = \alpha_1\psi_1 + \cdots + \alpha_n\psi_n$. Take the inner product of both sides with ψ_1:

$$\begin{aligned}\langle x, \psi_1\rangle &= \langle \alpha_1\psi_1 + \cdots + \alpha_n\psi_n, \psi_1\rangle \\ &= \alpha_1\langle \psi_1, \psi_1\rangle + \cdots + \alpha_n\langle \psi_n, \psi_1\rangle.\end{aligned}$$

Since the basis is orthogonal, all terms $\langle \psi_i, \psi_1\rangle$ are zero for $i \neq 1$. So we are left with:

$$\langle x, \psi_1\rangle = \alpha_1\langle \psi_1, \psi_1\rangle.$$

Therefore,

$$\alpha_1 = \frac{\langle x, \psi_1\rangle}{\langle \psi_1, \psi_1\rangle}.$$

□

If we had an orthonormal basis, then the bottom would normalize and thus $\alpha_1 = \langle x, \psi_1\rangle, \ldots \alpha_n = \langle x, \psi_n\rangle$.

Every time you are given an orthogonal basis $\{\psi_1, \ldots, \psi_n\}$, an orthonormal basis can be created. Let

$$\begin{aligned}\upsilon_1 &= \frac{\psi_1}{\|\psi_1\|} \\ \upsilon_2 &= \frac{\psi_2}{\|\psi_2\|} \\ &\vdots \\ \upsilon_n &= \frac{\psi_n}{\|\psi_n\|},\end{aligned}$$

where $\upsilon_1, \ldots \upsilon_n$ will be an orthonormal basis. Roughly speaking, it was previously discussed that if $T : \mathbb{V} \to \mathbb{V}$ is a symmetric operator, then the eigenvectors of T can be used to construct an orthogonal basis for $\mathbb{V}$.

The concept of a basis is important in this topic. You can think of a basis using an analogy with colors. Indeed, every single color that exists can be created using the three primary colors: red, blue, and yellow (r, b, y). Moreover, every other color can be written as a linear combination of (r, b, y). Therefore, we can think of the scalars and the vectors as the colors. This means we can construct all quantities of interest as

combinations of basis elements, though these combinations are not necessarily unique.

Since we will rely on probability, the number associated with $\frac{|\langle w,v\rangle|^2}{\langle w,w\rangle\langle v,v\rangle}$ will be a real number in the intervale $[0,1]$. That is, $\frac{|\langle w,v\rangle|^2}{\langle w,w\rangle\langle v,v\rangle} \in [0,1]$. Let's define this with a theorem.

Theorem 2.3. ***For any non-zero vector $v \in \mathscr{H}$,* $0 \leq \frac{|\langle w,v\rangle|^2}{\langle w,w\rangle\langle v,v\rangle} \leq 1$.

We will list two more theorems and prove the second one. The first one follows from the previous result.

Theorem 2.4. *If the state is an eigenvector w corresponding to the eigenvalue λ of a Hermitian operator A, then the probability of measuring the value λ with respect to A is one.*

Theorem 2.5. *If A is a Hermitian operator with a eigenvalue λ and corresponding eigenvector $w \in H$, and if v and μv (where $\mu \neq 0$ is any complex number) are two states, then the probability of obtaining the value λ when measuring the state v is precisely equal to the probability of obtaining λ when measuring the state μv.*

Proof. The probability of measuring λ for the state μv is given.

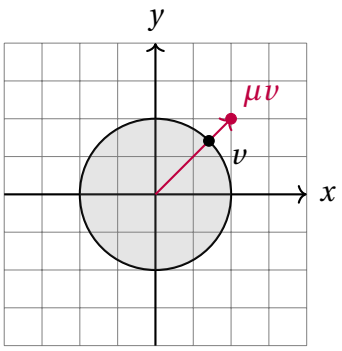
$$\frac{|\langle w,\mu v\rangle|^2}{\langle w,w\rangle\langle \mu v,\mu v\rangle} = \frac{|\mu\langle w,v\rangle|^2}{|\mu|^2\langle w,w\rangle\langle v,v\rangle}$$

Expanding the inner products and simplifying, we have.

$$= \frac{|\langle w,v\rangle|^2}{\langle w,w\rangle\langle v,v\rangle}$$

This shows that the probability that, in measuring v, we get the number λ is exactly the same as when measuring μv. Thus demonstrating the scalar invariance in measurement probabilities. □

This essentially the same thing as normalizing the vector μv so that their norm is 1, that is,



We can normalize μv to have the same direction and unit norm, effectively making it the same as v on the unit circle.

Indeed, multiplying by μ won't change much. Consider $v = a_1\phi_1 + \cdots + a_n\phi_n$ and a complex number $\mu \neq 0$ such that $\mu v = (\mu a_1)\phi_1 + \cdots + (\mu a_n)\phi_n$, then the probability of measuring λ_1 in each case is the same. Notice that

$$\frac{a_1^2}{a_1^2 + \cdots + a_n^2} = \frac{(\mu a_1)^2}{(\mu a_1)^2 + \cdots + (\mu a_n)^2}.$$

Postulate 5

Let A be a Hermitian operator with a simple eigenvalue λ. A simple eigenvalue is the eigenspace associated with λ which is one-dimensional.

Consider a state $v \in \mathscr{H}$ and take a measurement corresponding to the observable A. If the measurement outcome is λ, then the system collapses to the eigenspace associated with λ. Since this eigenspace is one-dimensional, then the post-measurement state must be a multiple of w. Therefore, the new state lies in the ray spanned by w.

After the measurement, the new state is

$$\frac{\langle w, v\rangle}{\|\langle w, v\rangle\|} w$$

Here $\langle w, v\rangle \neq 0$, ensuring a valid projection onto w.

For a moment, to make this more visual, consider a finite-dimensional space. Here, the position operator is T, the eigen functions are $\phi_1, \ldots, \phi_n$, and the eigenvalues are $\lambda_1, \ldots, \lambda_n$.

To describe the state ψ of a particle, we write

$$\psi = a_1\phi_1 + a_2\phi_2 + \cdots + a_n\phi_n.$$

The probability of measuring the value λ_1 is

$$\frac{|a_1|^2}{|a_1|^2 + |a_2|^2 + \cdots + |a_n|^2}.$$

Likewise, the probability of measuring λ_2 is

$$\frac{|a_2|^2}{|a_1|^2 + |a_2|^2 + \cdots + |a_n|^2}.$$

We would need to do this for all the eigenvalues of the operator. In fact, we can do this for any measurable quantity. For example, if we wish to do this for momentum, we consider the momentum operator M, whose eigenfunctions are $\zeta_1, \ldots, \zeta_n$, with corresponding eigenvalues $\lambda_1^*, \ldots, \lambda_n^*$. If the state of the system is

$$\psi = b_1\zeta_1 + b_2\zeta_2 + \cdots + b_n\zeta_n,$$

then the probability of measuring λ_1^* is

$$\frac{|b_1|^2}{|b_1|^2 + |b_2|^2 + \cdots + |b_n|^2}.$$

The main idea of this postulate is as follows: measurement changes the state of the particle. That is, it changes the state of the system. Whenever we perform a measurement, we affect the system's state by collapsing it onto one of the eigenstates of the operator being measured.

For example, suppose we are in a special case where the particle's state is

$$\psi = 1 \cdot \phi_1 + 0 \cdot \phi_2 + \cdots + 0 \cdot \phi_n.$$

Then the original state

$$\psi = a_1\phi_1 + a_2\phi_2 + \cdots + a_n\phi_n$$

has collapsed to $\psi = \phi_1$.

Likewise, if we have

$$\psi = a_1\phi_1 + a_2\phi_2 + \cdots + a_n\phi_n$$

and after a measurement we find the system in the state

$$\psi = \frac{1}{2}\phi_1 + \frac{1}{2}\phi_2 + 0\cdot\phi_3 + \cdots + 0\cdot\phi_n,$$

then the system has collapsed into the subspace spanned by ϕ_1 and ϕ_2. If the measurement outcome is ϕ_2, then the system collapses to the eigenstate ϕ_2.

2.3 The Schrödinger equation

Postulate 6

Let $v(t) \in \mathscr{H}$ represent the state of an evolving system as a function of time t. There exists a Hermitian operator $H(t)$ such that $v(t)$ satisfies the following partial differential equation,

$$i\hbar\frac{\partial v(t)}{\partial t} = H(t)v(t).$$

This is the Schrödinger's equation. The constant $\hbar$ is Planck's constant divided by 2π. More importantly, the Hermitian operator $H(t)$ is called the *Hamiltonian* and corresponds to the classical energy. Hence its eigenvalues will give the energy of the object.

Let's look at the Schrödinger's equation through an other lens. We will start off by stating, but not deriving three fundamental equations. First, we introduce the Laplace equation

$$\Delta u = 0$$

That is, if the unknowns were x, y, z then we denote it as $u(x, y, z)$. That is, if we had three differential equations then

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = \Delta u.$$

The heat equation depends on time and denoted as

$$\frac{\partial u}{\partial t} = k\Delta u,$$

for some constant k.

For example, consider buying a hot tea and bringing it inside. Eventually, the temperature of the tea will decrease with respect to the room temperature. That is, the heat equation describes how the temperature changes over time. However, after a long enough period, the temperature of the tea converges to that of the room. At that point, time becomes irrelevant, and the time derivative in the equation becomes 0. Once you have an item like tea sitting in a room for a long time, the process no longer depends on time, so the heat equation reduces to the Laplacian.

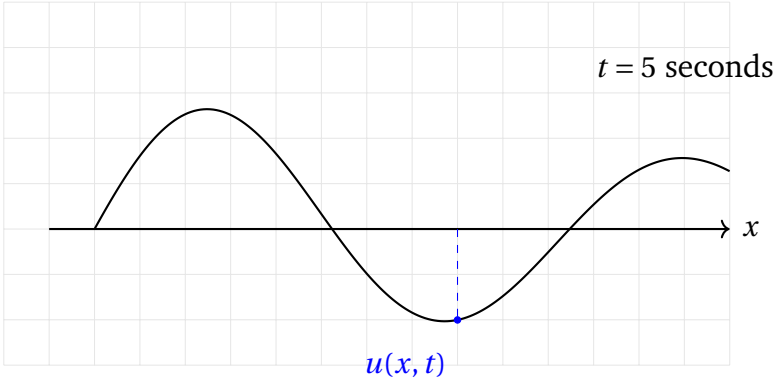
Moreover, the wave equation involves an unknown function $u(x, y, z, t)$ that depends on both space and time,

$$\frac{\partial^2 u}{\partial t^2} = k^2 \Delta u$$

Here, the differential equation is linear. That is, the unknown function u is not raised to any power greater than 1. Therefore the derivatives is linearly.

Note that $u(x, t)$ represents a one-dimensional function of space, with time t added as a parameter. The inclusion of t does not change the the function but allows us to observe how it changes over time. The same logic is applied to $u(x, y, t)$ and $u(x, y, z, t)$.

Consider a guitar being struck and its motion at time 5 seconds.



In the above diagram, $u(x, t)$ is negative. The vertical displacement $u(x, t)$ of the struck string is at time $t = 5$ seconds. The horizontal axis denoted as x is the length of the string. The dashed vertical line marks a specific location x, and the corresponding point on the curve is labeled $u(x, t)$.

In observation, there is no single, unique equation that defines all wave behavior. That is, we can always define a wave differently with respect to other types of waves.

We cannot categorize all waves into one general form. Perhaps a wave can be described as a shape or profile that moves through space, but this is not a comprehensive definition.

Now, let's introduce the term *separable solution*. Initially, we wanted to solve the heat equation. However, to make progress, we need to start somewhere. That is, we want to solve the equation for specific categories of solutions. One such type is the separable solution, which is a product of functions, each depending on a single variable, such as x or t. For example, $u(x,t) = xt$, $u(x,t) = \sin(x)\sin(t)$, and $u(x,t) = x^3\cos(t)$ are all separable solutions. In contrast, $u(x,t) = x + \sin(t + x^3 t - t)$ is not separable.

More generically, we define all the separable solutions as $u(x,t) = f(x)g(t)$. To show that this equation satisfies our wave equation we plug in $u(x,t)$ into our equation. i.e, let's find the solutions for $u(x,t) = f(x)g(t)$.

$$\frac{\partial^2 u}{\partial t^2} = k^2 \Delta u$$
$$\frac{\partial^2 u}{\partial t^2} = k\frac{\partial^2 u}{\partial x^2}$$

Substitute $u(x,t) = f(x)g(t)$

$$\frac{\partial^2}{\partial t^2}(f(x)g(t)) = k\frac{\partial^2}{\partial x^2}(f(x)g(t))$$

Since $f(x)$ and $g(t)$ are each constants we can pull them out.

$$f(x)\frac{\partial^2 g}{\partial t^2} = kg(t)\frac{\partial^2 f}{\partial x^2}$$

Divide both sides by

$$kf''(x)g''(t): \quad \frac{f(x)g''(t)}{kf''(x)g''(t)} = \frac{kg(t)f''(x)}{kf''(x)g''(t)}$$
$$\frac{f(x)}{kf''(x)} = \frac{g(t)}{g''(t)}$$

Since the above only holds true if the equations are constant, then we can concluded that

$$\frac{f(x)}{kf''(x)} = -\lambda$$
$$f(x) = -\lambda k f''(x)$$
$$f(x) + \lambda k f''(x) = 0$$

Dividing by λk

$$f''(x) + \frac{1}{\lambda k} f(x) = 0$$

If we let $\frac{1}{\lambda k} = k^2$ we have

$$\frac{\partial^2 f(x)}{\partial x^2} + k^2 f(x) = 0.$$

Schrödinger wanted to modify the classical wave equation

$$\frac{\partial^2 f(x)}{\partial x^2} + k^2 f(x) = 0.$$

First, the symbol Ψ was used to represent the matter wave

$$\frac{\partial^2 \Psi}{\partial x^2} + k^2 \Psi = 0.$$

Then, he rewrote the number $k = \frac{2\pi}{\lambda}$ in terms of wavelengths.

$$\frac{\partial^2 \Psi}{\partial x^2} + \frac{4\pi^2}{\lambda^2} \Psi = 0.$$

Schrödinger's then substituted de Broglie's related matter wave formula $\lambda = \frac{h}{p} = \frac{h}{mv}$ using it momentum. Here m is mass and v is velocity.

$$\frac{\partial^2 \Psi}{\partial x^2} + \frac{4\pi^2 m^2 v^2}{h^2} \Psi = 0.$$

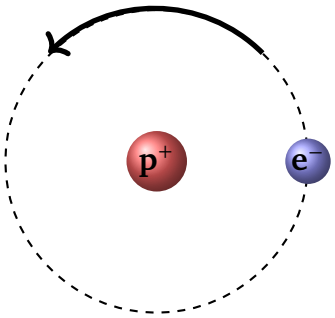
We know that kinetic energy is $KE = \frac{1}{2} m v^2$, thus

$$\frac{\partial^2 \Psi}{\partial x^2} + \frac{8\pi^2 m(KE)}{h^2}\Psi = 0.$$

The total energy $E = KE + V$, where V is the potential energy. Rearranging this equation gives us $KE = E - V$. Substituting that we have

$$\frac{\partial^2 \Psi}{\partial x^2} + \frac{8\pi^2 m(E-V)}{h^2}\Psi = 0.$$

Recall that the hydrogen atom has one proton and one electron. We assume that the proton is fixed. This creates an electric potential for the electron of the charge of the electron e^2. Below, we define r to be the radius from the electron to the proton,



Hydrogen Atom

$$\frac{\partial^2 \Psi}{\partial x^2} + \frac{8\pi^2 m\left(E - \frac{e^2}{r}\right)}{h^2}\Psi = 0.$$

Lastly, since atoms are three dimensional objects we add a spacial derivative and include x, y, z,

$$\frac{\partial^2 \Psi}{\partial x^2} + \frac{\partial^2 \Psi}{\partial y^2} + \frac{\partial^2 \Psi}{\partial z^2} + \frac{8\pi^2 m\left(E - \frac{e^2}{r}\right)}{h^2}\Psi = 0.$$

The initial equation, the classical wave equation only depended on position. To solve for the the vibrating string example, we first multiply the space solution $f(x)$ by the solutions in time $g(t)$ to compute the final postilion $y(x,t)$. Here, y is a function of both position and time. In the classical wave equation, the spacial and time components are the same. The only real difference is the constants. That is, the spacial time equation is

$$\frac{\partial^2 g(x)}{\partial x^2} + \omega^2 f(x) = 0.$$

Moreover, the solutions in time are just linear combinations of sine and cosine functions. Indeed,

$$g(t) = A\sin(\omega t) + B\cos(\omega t).$$

We can also express these solutions using complex exponentials. That is,

$$g(t) = A\sin(\omega t) + B\cos(\omega t)$$

By Euler's Formulas,

$$\begin{aligned}
&= A \cdot \frac{e^{i\omega t} - e^{-i\omega t}}{2i} + B \cdot \frac{e^{i\omega t} + e^{-i\omega t}}{2} \\
&= \frac{A}{2i} e^{i\omega t} - \frac{A}{2i} e^{-i\omega t} + \frac{B}{2} e^{i\omega t} + \frac{B}{2} e^{-i\omega t} \\
&= \left(\frac{B}{2} - \frac{A}{2i}\right) e^{i\omega t} + \left(\frac{B}{2} + \frac{A}{2i}\right) e^{-i\omega t}
\end{aligned}$$

Letting $C = \frac{B}{2} - \frac{A}{2i}$, we observe that the second term is the complex conjugate of the first, i.e., $\overline{C} = \frac{B}{2} + \frac{A}{2i}$. Thus,

$$g(t) = Ce^{i\omega t} + \overline{C}e^{-i\omega t} = 2\,\mathrm{Re}[Ce^{i\omega t}]$$

Therefore, up to a constant factor, we may write the general solution as,

$$g(t) = \mathrm{Re}[Ce^{i\omega t}]$$

This makes differentiation much simpler. Notice that the first derivative of $e^{i\omega t}$ is $\frac{\partial}{\partial t} e^{i\omega t} = i\omega e^{i\omega t}$ and the second derivative is $\frac{\partial^2}{\partial t^2} e^{i\omega t} = i^2\omega^2 e^{i\omega t^2} = -\omega^2 e^{i\omega t^2}$. Recall that the final answer is always the physical answer.

Let's start with the a complex representation of the wave function,

$$\Psi(x,t) = \Psi(x)e^{-i\omega t}$$

The energy of a matter wave is proportional to its frequency. By the Plank-Einstein relation we can rewrite the right as,

$$\Psi(x,t) = \Psi(x)e^{-2\pi iEt/h}$$

Differentiating.

$$\begin{aligned}\frac{\partial\Psi}{\partial t} &= -\left(\frac{2\pi iE}{h}\right)\Psi \\ E\Psi &= -\frac{h}{2\Psi i}\frac{\partial\Psi}{\partial t} \\ E\Psi &= -i\frac{h}{2\Psi}\frac{\partial\Psi}{\partial t}\end{aligned}$$

By differentiating, we see that the energy of the wave times the wave function is proportional to i times the derivative of the wave function.

$$E\Psi = -i\hbar\frac{\partial\Psi}{\partial t}.$$

Lets isolate $E\Psi$ from Schrödinger's time independent equation.

$$\begin{aligned}\frac{\partial^2\Psi}{\partial x^2} + \frac{8\pi^2 m}{h^2}(E-V)\Psi &= 0 \\ \frac{\partial^2\Psi}{\partial x^2} &= -\frac{8\pi^2 m}{h^2}(E-V)\Psi \\ \frac{h^2}{8\pi^2 m}\frac{\partial^2\Psi}{\partial x^2} &= -(E-V)\Psi \\ \frac{h^2}{8\pi^2 m}\frac{\partial^2\Psi}{\partial x^2} + E\Psi &= V\Psi \\ E\Psi &= -\frac{h^2}{8\pi^2 m}\frac{\partial^2\Psi}{\partial x^2} + V\Psi\end{aligned}$$

Plugging in the $E\Psi$ from above,

$$i\hbar\frac{\partial\Psi}{\partial t} = -\frac{h^2}{8\pi^2 m}\frac{\partial^2\Psi}{\partial x^2} + V\Psi.$$

Acknowledgements

I would like to thank Dr. Ali Behzadan, without whom I would have no idea what's going on.

References

[1] Thomas A Garrity. *Electricity and Magnetism for Mathematicians: A Guided Path from Maxwell's Equations to Yang-Mills*. Cambridge University Press, 2015.