MODERN PHYSICS
BUT MOSTLY
QUANTUM MECHANICS

Seeing the invisible, Proving the impossible, Thinking the unthinkable

Modern physics stretches your imagination beyond recognition, pushes the intellectual envelope to the limit, and takes you to the forefront of human knowledge.

MASAYASU AOTANI
(青谷正妥)
Kyoto University
Kyoto, Japan

ANAX PARTHENOPE (ギンヤンマ)

2013
No insects, no life!
Modern Physics

Masayasu AOTANI (青谷正妥)

Spring 2014
Copyright ©1991–2014
Masayasu AOTANI
Permission is granted for personal use only.
## Contents

1 Classical Physics vs. Modern Physics ........................................ 13  
1.1 Extreme Conditions, Precise Measurements, and New Physics .... 13  
1.2 Black-Body Radiation .................................................. 15  
1.3 Quantum Theory of Light .................................................. 15  
  1.3.1 Photoelectric Effect ................................................. 17  
  1.3.2 Einstein’s Idea of Quantization .................................. 17  
1.4 de Broglie-Einstein Postulates ............................................ 18  

2 Mathematical Preliminaries .................................................. 19  
2.1 Linear Vector Spaces .................................................... 19  
2.2 Inner Product Spaces .................................................... 23  
2.3 $L^2$-Space ................................................................. 29  
2.4 The Bra-ket Notation ..................................................... 31  
2.5 Vector Subspace ........................................................... 41  
2.6 Linear Operators ........................................................... 41  
2.7 Matrix Representation of Linear Operators ......................... 43  
  2.7.1 Matrix Representations of Operator Products ................. 47  
2.8 The Adjoint of an Operator .............................................. 48  
2.9 Eigenvalues and Eigenvectors .......................................... 50  
2.10 Special Types of Operators ............................................. 53  
  2.10.1 Hermitian Operators .............................................. 53  
  2.10.2 Simultaneous Diagonalization .................................. 65  
2.11 Active and Passive Transformations ................................ 66  

3 Fundamental Postulates and the Mathematical Framework of QUan-
tum Mechanics ................................................................. 71  
3.1 The Fundamental Postulates ............................................. 72  
3.2 Unitary Time Evolution ................................................... 75
11 Electron Spin

11.1 What is the Electron Spin? ........................................ 235
  11.1.1 Stern-Gerlach Experiment .................................. 236
  11.1.2 Fine Structure of the Hydrogen Spectrum: Spin-Orbit Interaction ........................................ 237
11.2 Spin and Pauli Matrices ........................................... 238
11.3 Sequential Measurements ......................................... 245
11.4 Spin States in an Arbitrary Direction ......................... 248
  11.4.1 Spin Matrix \( S_u \) and Its Eigenvectors ................... 248
  11.4.2 Sequential Measurements Revisited ....................... 251
  11.4.3 Comparison with the Classical Theory ..................... 251

12 Addition of Orbital and Spin Angular Momenta .................. 255

13 Molecular Rotation .................................................... 257

Appendix A Matrix Algebra ............................................. 267
  A.1 Invertible Matrices ............................................. 267
  A.2 Rank of a Matrix ............................................... 268
  A.3 Change of Basis .................................................. 271
  A.4 Multiplicity of an Eigenvalue .................................. 275
  A.5 Diagonalizability and Simultaneous Diagonalization ......... 276
  A.6 Some Useful Definitions, Facts, and Theorems ............... 286

Appendix B Holomorphic Functional Calculus ....................... 287

Appendix C Traveling Waves and Standing Waves .................... 289
  C.1 The Wave Equation ............................................. 289
    C.1.1 Traveling Wave Solutions ................................. 290
      C.1.1.1 Real and Complex Expressions for the Traveling Wave 290
    C.1.2 Standing Wave Solutions ................................ 291
      C.1.2.1 Standing Waves on a String ........................ 291
      C.1.2.2 Standing Waves in Quantum Mechanics ............... 292
Appendix L Laplacian Operator in Spherical Coordinates 329

Appendix M Legendre and Associated Legendre Polynomials 341
   M.1 Legendre Polynomials ........................................... 341
   M.2 Associated Legendre Polynomials ............................ 342

Appendix N Laguerre and Associated Laguerre Polynomials 349

Appendix O Fine Structure and Lamb Shift 351
   O.1 Relativistic Effects: The Dirac Theory .................... 351
      O.1.1 Variation of the Mass with the Velocity: $H_k$ ........ 353
      O.1.2 Spin-Orbit Interaction: $H_{SO}$ ....................... 354
      O.1.3 The Darwin Term: $H_D$ ............................... 355
   O.2 The Lamb Shift .................................................. 356

Bibliography 358

Answers to Exercises 359
   Chapter 2 ........................................................... 359
   Chapter 3 ........................................................... 364
   Chapter 5 ........................................................... 364
   Chapter 6 ........................................................... 368
   Chapter 7 ........................................................... 369
   Chapter 8 ........................................................... 376
   Chapter 9 ........................................................... 376
   Chapter 10 ........................................................... 377
   Chapter 11 ........................................................... 379
   Chapter 12 ........................................................... 379

Subject Index 381
List of Tables

3.1 Physical observables and corresponding quantum operators . . . . . . 73
3.2 Quantum operators in three dimensions . . . . . . . . . . . . . . . . 73

10.1 Spherical Harmonics for \( l = 0, 1, \) and \( 2 \) . . . . . . . . . . . . . . . . . 217
10.2 Some Normalized Radial Eigenfunctions for the One-Electron Atom . 226
10.3 Some Normalized Full Eigenfunctions for the One-Electron Atom . . 227
List of Figures

1.1 Black-Body Radiation (Source: Wikimedia Commons; Author: Darth Kule) .................................................. 16

10.1 Spherical Coordinate System (Source: Wikimedia Commons; Author: Andeggs) ...................................... 199

10.2 Spherical Coordinate System Used in Mathematics (Source: Wikimedia Commons; Author: Dmcq) ............ 200

10.3 Volume Element in Spherical Coordinates (Source: Victor J. Montemayor, Middle Tennessee State University) .......................................................... 203

11.1 Stern-Gerlach Experiment (Diagram drawn by en wikipedia Theresa Knott.) ........................................... 236

11.2 Results of Stern-Gerlach Experiment with and without a non-uniform external magnetic field (Source: Stern and Gerlach’s original paper [Gerlach and Stern, 1922, p.350]) .................................................. 237
Chapter 1

Classical Physics vs. Modern Physics

Very roughly speaking, Classical Physics consists of Newtonian Mechanics and the theory of Electricity and Magnetism by Maxwell, while Modern Physics typically means Relativity and Quantum Mechanics.

Classical Physics
- Newtonian Mechanics
- Maxwell’s Electricity and Magnetism

Modern Physics
- Relativity
- Quantum Mechanics

1.1 Extreme Conditions, Precise Measurements, and New Physics

Quantum Mechanics and the Theory of Relativity emerged early in the 20th century when technological advances were beginning to make laboratory experiments under extreme conditions possible. The same advances also made high-precision measurements readily available. Under these extreme conditions, physicists found that Newtonian Mechanics no longer held. One such condition is a microscopic phenomenon such as the behavior of an electron, and another phenomenon involves
objects moving at speeds comparable to that of light. The former led to the discovery of Quantum Mechanics, and the latter to Special Relativity\(^1\).

**Two Types of Extreme Circumstances**

1. Microscopic Phenomena (atomic and subatomic)
   - Tunneling: A moving electron can go through a potential barrier which requires more energy than the kinetic energy of the electron.
   - Wave-particle duality: Microscopic particles such as an electron sometimes behaves like a wave showing interference and superposition.

2. At High Speeds (near the speed of light denoted by \(c\))
   - Fast moving particles live longer.
   - Galilean transformation does not work.

In this book, we will only deal with Quantum Mechanics, and the following list shows some of the key events, discoveries, and theories that contributed to the birth of quantum mechanics.

**Historical Events, Discoveries, and Theories**

1. Black-body Radiation: (Quantization, Planck’s Postulate)
2. Particle-Like Properties of Radiation: (Wave-Particle Duality)
3. Einstein’s Quantum Theory of Light: A bundle of energy is localized in a small volume of space.
4. Atomic spectra indicating discrete energy levels \(E_1, E_2, E_3, \ldots\) as opposed to continuous distribution of energy levels \(E \in [0, \infty)\)\(^2\)

---

\(^1\)There are two distinct theories of relativity; the Special Theory of Relativity and the General Theory of Relativity. It is the Special Theory of Relativity that deals with high-speed phenomena. The General Theory of Relativity is an attempt to explain gravity geometrically. We only concern ourselves with the Special Theory, which is a theory of space and time that successfully accounts for the deviations from Newtonian Mechanics for fast-moving objects.

\(^2\)We will later see in Section 7.6.1 that \(E = 0\) is impossible in some quantum mechanical systems such as a simple harmonic oscillator.
1.2 Black-Body Radiation

A black body is an idealization of “real-life” black objects in that it does not reflect any light; i.e. perfectly black. When an object is perfectly black, all the radiation coming from the object is due to emission, and the pattern of emission depends only on its temperature. The thermal emission patterns are shown for 3000 K (red), 4000 K (green), and 5000 K (blue) in Figure 1.1. The figure also shows the pattern predicted by the classical theory (dark brown). As you can see, the classical theory predicts that the emitted radiation increases rapidly without bound as the frequency increases. This phenomenon is called ultraviolet catastrophe, which does not agree with the actual observation.

In order to circumvent this problem, in 1901 Max Planck found a mathematical expression that fit the data, in which he assumed that energy levels are not continuously distributed but are discrete, with each level being an integer multiple of a quantity called Planck constant denoted by $h$. The Planck constant is a proportionality constant between the energy $E$ and the frequency $\nu$ such that

$$E = h\nu.$$ (1.1)

For curiosity, Planck’s formula for the spectral radiance $I(\nu, t)$, the energy per unit time (or the power) radiated per unit area of emitting surface in the normal direction per unit solid angle per unit frequency by a black body at temperature $T$, is given by

$$I(\nu, T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{\frac{kT}{\nu}} - 1};$$ (1.2)

where $\nu$ is the frequency of the radiation, $h \approx 6.62606957 \times 10^{-34}$ (Js: joule-seconds) is the Planck constant, $c \approx 299792458$ (ms$^{-1}$: meters per second) is the speed of light in vacuum, and $k \approx 1.3806488 \times 10^{-23}$ (m$^2$kg s$^{-2}$K$^{-1}$: meter$^2$ kilograms per second per second per Kelvin) is the Boltzmann constant.

1.3 Quantum Theory of Light

While it was undeniable that light behaved as a wave, a physical phenomenon called photoelectric effect was not consistent with this view.
Figure 1.1: Black-Body Radiation (Source: Wikimedia Commons; Author: Darth Kule)
1.3.1 Photoelectric Effect

Photoelectric effect is a phenomenon where electrons, called photoelectrons, are ejected from a metal surface when light shines on it. The following two observations were made.

- The intensity of the incident light had no effect on the maximum kinetic energy of the photoelectrons. The brightness or dimness of the light had no effect on the maximum kinetic energy if the frequencies are the same.

- No electron was ejected by light with frequencies below a certain cutoff value, called the threshold frequency.

These were not consistent with the wave nature of light. First, the classical theory predicted that light with higher intensity ejects electrons with greater energy. Secondly, according to the wave picture, light with lower frequency can also eject electrons except that it takes longer.

1.3.2 Einstein’s Idea of Quantization

These discrepancies were resolved by Einstein’s idea of light quantum, which claims that radiation energy is not continuously distributed over the wavefront, but is localized forming bundles of energy or particles. They were later called photons. Einstein postulated that a single photon interacts with a single electron transferring its energy to the electron instantaneously. The amount of energy \( E \) carried by the photon is given by the Planck’s formula

\[
E = h \nu. \tag{1.3}
\]

Another important relation, which is a direct consequence of (1.3) is

\[
p = \frac{h}{\lambda}; \tag{1.4}
\]

where \( p \) is the momentum and \( \lambda \) is the wavelength. \(^3\)
1.4 de Broglie-Einstein Postulates

Einstein showed that light, which had classically been regarded as an electromagnetic wave, can also be viewed as a collection of particles, later called photons. This is known as wave-particle duality, i.e. light possesses both a wave nature and a particle nature.

A French physicist named de Broglie tried to generalize this idea. As light, classically a wave, is also particle-like, he conjectured that things such as an electron, so far considered to be a particle, could also possess a wave-nature. The wave associated with a particle is called de Broglie wave or de Broglie’s matter wave.

In order to compute the frequency and the wavelength of the matter wave, de Broglie assumed that Einstein’s relations (1.3) and (1.4) also apply here.

\[ E = h\nu \]  \[ p = \frac{h}{\lambda} \]  \hfill (1.7)

The relations (1.7) are known as de Broglie-Einstein relations.

The correctness of these formulas were soon confirmed by two American physicists Davisson and Germer, who confirmed that the interference pattern for an electron beam reflected from a single crystal of nickel fitted perfectly to the de Broglie-Einstein relations. Note that an interference phenomenon is possible only if electrons exhibit a wave nature.

Development of Quantum Mechanics was an attempt to merge this wave-particle duality and classical Newtonian Mechanics in one equation as we will see in Chapter 5.

The relation (1.4) is derived from (1.3) and the relativistic energy formula

\[ E^2 = (mc^2)^2 + (pc)^2; \]  \hfill (1.5)

where \( m \) is the rest mass, and \( c \) is the speed of light. Because \( m = 0 \) for a photon, we have

\[ E^2 = (pc)^2 \implies E = pc \implies E = p\nu\lambda \implies h\nu = p\nu\lambda \implies p = \frac{h}{\lambda}. \]  \hfill (1.6)
Chapter 2
Mathematical Preliminaries

This chapter is a summary of the mathematical tools used in quantum mechanics. It covers various important aspects of basic linear algebra, defines linear operators, and sets convenient notations for the rest of the book.

2.1 Linear Vector Spaces

Consider the three dimensional Cartesian space $\mathbb{R}^3$ equipped with the $x$-, $y$-, and $z$-axes. We have three unit vectors, perpendicular/orthogonal to one another, denoted by $(\mathbf{i}, \mathbf{j}, \mathbf{k})$, $(\hat{x}, \hat{y}, \hat{z})$, or $(\hat{x}, \hat{y}, \hat{z})$. In the component notation, we have $\mathbf{i} = (1, 0, 0)$, $\mathbf{j} = (0, 1, 0)$, and $\mathbf{k} = (0, 0, 1)$. Any point $\mathbf{p}$ in this space identified by three coordinates $x$, $y$, and $z$ such that $\mathbf{p} = (x, y, z)$ can also be regarded as a vector $xi + yj + zk$ or $(x, y, z)$ in the component notation. Note that we use $(x, y, z)$ both for the coordinates and components by abuse of notation. You must be familiar with such arithmetic operations as addition, subtraction, and scalar multiplication for vectors; which are basically component-wise operations$^1$. This is a canonical example of a linear vector space, and you should keep this example in mind as you study the following definition of a linear vector space. If you are not so inclined, you can safely skip this definition and still can understand the rest of this book fully.

Definition 2.1 (Linear Vector Space)

A linear vector space $V$ is a set $\{v_1, v_2, v_3, \ldots\}$ of objects called vectors for which “addition” and scalar multiplication are defined, such that

$^1$To be more precise, these operations can be done component-wise though there are other equivalent definitions/formulations for these arithmetic operations.
1. Both addition and scalar multiplication generate another member of $V$. This property is referred to as “closure” under addition and scalar multiplication.

2. Addition and scalar multiplication obey the following axioms.

**Axioms for Addition:** Consider $v_i, v_j,$ and $v_k$ taken from $V$.

(i) $v_i + v_j = v_j + v_i$ (commutativity)
(ii) $v_i + (v_j + v_k) = (v_i + v_j) + v_k$ (associativity)
(iii) There exists a unique null vector, denoted by $0$, in $V$ such that $0 + v_i = v_i + 0 = v_i$. (existence of the identity element)
(iv) For each $v_i$, there exists a unique inverse $(-v_i)$ in $V$ such that $v_i + (-v_i) = 0$. (existence of the inverse)$^a$

**Axioms for Scalar Multiplication:** Consider arbitrary vectors $v_i, v_j$ and arbitrary scalars $\alpha, \beta$.

(v) $\alpha(v_i + v_j) = \alpha v_i + \alpha v_j$
(vi) $(\alpha + \beta)v_i = \alpha v_i + \beta v_i$
(vii) $\alpha(\beta v_i) = (\alpha \beta)v_i$

**Fact 2.1** The following facts can be proved using the axioms.

1. $0v = 0$
2. $\alpha 0 = 0$
3. $(-1)v = (-v)$

---

$^a$The axioms (i) through (iv) means that a linear vector space forms an abelian group under addition.

**Definition 2.2** If the allowed values for the scalars $\{\alpha, \beta, \gamma, \ldots\}$ come from some field $\mathbb{F}^2$, we say the linear vector space $V$ is defined over the field $\mathbb{F}$. In particular, if $\mathbb{F}$ is the field of real numbers $\mathbb{R}$, $V$ is a real vector space. Likewise, if $\mathbb{F}$ is the field of complex numbers $\mathbb{C}$, $V$ is a complex vector space.

Clearly, $\mathbb{R}^3$ is not a complex vector space, but a real vector space. Complex vector spaces are the canonical vector spaces in quantum mechanics.

$^2$There is an abstract mathematical definition of a field. However, it suffices for elementary quantum mechanics to know the real numbers and the complex numbers respectively form a field.
Definition 2.3 A set of vectors \( \{v_1, v_2, \ldots, v_n\} \) is linearly independent (LI) if
\[
\sum_{i=1}^{n} \alpha_i v_i = 0 \implies \alpha_1 = \alpha_2 = \ldots = \alpha_n = 0.
\] (2.1)

Definition 2.3 is equivalent to saying that no vector in \( \{v_1, v_2, \ldots, v_n\} \) can be expressed as a linear combination of the other vectors in the set.

Vectors \( i, j, \) and \( k \) are linearly independent because
\[
ai + bj + ck = 0 \quad \text{or} \quad (a, b, c) = (0, 0, 0) \implies a = b = c = 0.
\] (2.2)

Generally speaking, vectors in \( \mathbb{R}^3 \) that are perpendicular to one another are linearly independent. However, these are not the only examples of a linearly independent set of vectors. To give a trivial example, vectors \( i \) and \( i + j \) are linearly independent, though they are not perpendicular to each other. This is because
\[
ai + b(i + j) = 0 \implies (a + b, b) = (0, 0) \implies a = b = 0.
\] (2.3)

Definition 2.4 A vector space is \( n \)-dimensional if it has at most \( n \) vectors that are linearly independent.

Notation: An \( n \)-dimensional vector space over a field \( F \) is denoted by \( \mathbb{V}^n(F) \). The ones we encounter in this course are usually members of \( \mathbb{V}^n(\mathbb{C}) \), including the cases where \( n = \infty \).

It is often clear what the field \( F \) is. In this case, we simply write \( \mathbb{V}^n \).

Theorem 2.1 Suppose \( \{v_1, v_2, \ldots, v_n\} \) are linearly independent vectors in an \( n \)-dimensional vector space \( \mathbb{V}^n \). Then, any vector \( v \) in \( \mathbb{V}^n \) can be written as a linear combination of \( \{v_1, v_2, \ldots, v_n\} \).

Proof
As \( \mathbb{V}^n \) is \( n \)-dimensional, you can find a set of \( n + 1 \) scalars \( \{\alpha_1, \alpha_2, \ldots, \alpha_n, \alpha_{n+1}\} \), not all zero, such that
\[
\left( \sum_{i=1}^{n} \alpha_i v_i \right) + \alpha_{n+1} v = 0.
\] (2.4)

Else, \( \{v_1, v_2, \ldots, v_n, v\} \) are linearly independent, and \( \mathbb{V}^n \) is at least \( n+1 \)-dimensional, which is a clear contradiction. Furthermore, \( \alpha_{n+1} \neq 0 \). If not, at least one of
\{\alpha_1, \alpha_2, \ldots, \alpha_n\} \text{ is not } 0, \text{ and yet } \sum_{i=1}^n \alpha_i v_i = 0, \text{ contradicting the assumption that } \{v_1, v_2, \ldots, v_n\} \text{ are linearly independent. We now have}

\[ v = \sum_{i=1}^n \frac{-\alpha_i}{\alpha_{n+1}} v_i. \quad (2.5) \]

We will next show that there is only one way to express \(v\) as a linear combination of \(\{v_1, v_2, \ldots, v_n\}\) in Theorem 2.1.

**Theorem 2.2** The coefficients of Equation 2.5 are unique.

**Proof**

Consider any linear expression of \(v\) in terms of \(\{v_1, v_2, \ldots, v_n\}\)

\[ v = \sum_{i=1}^n \beta_i v_i. \quad (2.6) \]

Subtracting Equation 2.6 from Equation 2.5,

\[ 0 = \sum_{i=1}^n \left( \frac{-\alpha_i}{\alpha_{n+1}} - \beta_i \right) v_i. \quad (2.7) \]

However, we know \(\{v_1, v_2, \ldots, v_n\}\) are linearly independent. Hence,

\[ \sum_{i=1}^n \left( \frac{-\alpha_i}{\alpha_{n+1}} - \beta_i \right) v_i = 0 \implies \frac{-\alpha_i}{\alpha_{n+1}} - \beta_i = 0 \text{ or } \beta_i = \frac{-\alpha_i}{\alpha_{n+1}} \text{ for all } i = 1, 2, \ldots, n. \quad (2.8) \]

In order to understand Definition 2.5 below, just imagine how any vector in \(\mathbb{R}^3\) can be expressed as a unique linear combination of \(i, j, \) and \(k\).

**Definition 2.5** A set of linearly independent vectors \(\{v_1, v_2, \ldots, v_n\}\), which can express any vector \(v\) in \(\mathbb{V}^n\) as a linear combination (Equation 2.6), is called a basis that spans \(\mathbb{V}^n\). The coefficients of the linear combination \(\beta_i\) are called the components of \(v\) in the basis \(\{v_1, v_2, \ldots, v_n\}\). We also say that \(\mathbb{V}^n\) is the vector space spanned by the basis \(\{v_1, v_2, \ldots, v_n\}\)\(^3\).

\(^3\)This means that we can define \(\mathbb{V}^n\) as the collection of all objects of the form \(\sum_{i=1}^n \alpha_i v_i\).
Once we pick a basis \( \{v_1, v_2, \ldots, v_n\} \), and express \( v \) as a linear combination of the basis vectors \( v = \sum_{i=1}^{n} \beta_i v_i \), we get a component notation of \( v \) as follows.

\[
v = \sum_{i=1}^{n} \beta_i v_i = (\beta_1, \beta_2, \ldots, \beta_n).
\]

With this notation, vector additions and scalar multiplications can be done component by component. For vectors \( v = (\beta_1, \beta_2, \ldots, \beta_n) \), \( w = (\gamma_1, \gamma_2, \ldots, \gamma_n) \) and a scalar \( \alpha \), we simply have

\[
v + w = (\beta_1 + \gamma_1, \beta_2 + \gamma_2, \ldots, \beta_n + \gamma_n) \quad \text{and} \quad \alpha v = (\alpha \beta_1, \alpha \beta_2, \ldots, \alpha \beta_n).
\]

### 2.2 Inner Product Spaces

This is nothing but a generalization of the familiar scalar product for \( \mathbb{V}^3(\mathbb{R}) \). Namely, it is an extension of the inner product defined on \( \mathbb{R}^3 \) to complex vectors. Readers should be familiar with the component-wise definition of the usual inner product on \( \mathbb{R}^3 \), also called dot product, such that

\[
a \cdot b = (a_1, a_2, a_3) \cdot (b_1, b_2, b_3) = a_1 b_1 + a_2 b_2 + a_3 b_3.
\]

We can express this as a matrix product of a row vector \( a \) and a column vector \( b \).

\[
\begin{bmatrix}
a_1 & a_2 & a_3
\end{bmatrix}
\begin{bmatrix}
b_1 \\
b_2 \\
b_3
\end{bmatrix} = 
\begin{bmatrix}
a_1 b_1 + a_2 b_2 + a_3 b_3
\end{bmatrix};
\]

where the one-by-one matrix on the right-hand side is a scalar by definition. This definition of inner product is extended to include complex components by taking the complex conjugate of the components of \( a \). So, the extended inner product, which we simply call an inner product in the rest of the book, is given by

\[
\begin{bmatrix}
a_1 & a_2 & a_3
\end{bmatrix}^* 
\begin{bmatrix}
b_1 \\
b_2 \\
b_3
\end{bmatrix} = 
\begin{bmatrix}
a_1^* & a_2^* & a_3^*
\end{bmatrix}
\begin{bmatrix}
b_1 \\
b_2 \\
b_3
\end{bmatrix} = 
\begin{bmatrix}
a_1^* b_1 + a_2^* b_2 + a_3^* b_3
\end{bmatrix}.
\]

Definition 2.6 below is an abstract version of the familiar inner product defined on \( \mathbb{R}^3 \). This definition contains the inner product defined on \( \mathbb{R}^3 \), but not all inner products are of that type. Nevertheless, you should keep the inner product on \( \mathbb{R}^3 \) in mind as you read the following general definition of inner product. In Definition 2.6, we are writing \( a^* \) as \( \langle a \vert \) and \( b \) as \( \vert b \rangle \) so that \( a^* \cdot b = \langle a \vert b \rangle \) in anticipation of Dirac's braket notation to be introduced on p.31.

---

4 This is the notation suggested on p.21, which is the same as \( \mathbb{R}^3 \).

5 Naively, \( a^* \cdot b = \langle a \vert b \rangle \), but we use only one vertical line and write \( \langle a \vert b \rangle \).
Definition 2.6 (Inner Product) An inner product of \( \mathbf{v}_i \) and \( \mathbf{v}_j \), denoted by \( \langle \mathbf{v}_i \mid \mathbf{v}_j \rangle \), is a function of two vectors mapping onto \( \mathbb{C} \) \((\langle \mid \rangle : \mathbb{C}^n \times \mathbb{C}^n \rightarrow \mathbb{C})\) satisfying the axioms below.

(i) \( \langle \mathbf{v}_i \mid \mathbf{v}_i \rangle \geq 0 \) with the equality holding if and only if \( \mathbf{v}_i = \mathbf{0} \)

(ii) \( \langle \mathbf{v}_i \mid \mathbf{v}_j \rangle = \langle \mathbf{v}_j \mid \mathbf{v}_i \rangle^* \), where \( ^* \) denotes the complex conjugate

(iii) \( \langle \mathbf{v}_i \mid \alpha \mathbf{v}_j + \beta \mathbf{v}_k \rangle = \alpha \langle \mathbf{v}_i \mid \mathbf{v}_j \rangle + \beta \langle \mathbf{v}_i \mid \mathbf{v}_k \rangle \) (linearity in the second argument)

It is convenient to add the fourth property to this list, though it is not an axiom but a consequence of Axioms (i) and (iii).

(iv) \( \langle \alpha \mathbf{v}_i + \beta \mathbf{v}_j \mid \mathbf{v}_k \rangle = \alpha^* \langle \mathbf{v}_i \mid \mathbf{v}_k \rangle + \beta^* \langle \mathbf{v}_j \mid \mathbf{v}_k \rangle \) (anti-linearity in the first argument)

We say the inner product is linear in the second vector and antilinear in the first vector.

In order to see how (iv) follows from (ii) and (iii), see the derivation below.

\[
\langle \alpha \mathbf{v}_i + \beta \mathbf{v}_j \mid \mathbf{v}_k \rangle = \langle \mathbf{v}_k \mid \alpha \mathbf{v}_i + \beta \mathbf{v}_j \rangle^* \quad \text{by (ii)} \\
= \langle \alpha \langle \mathbf{v}_k \mid \mathbf{v}_i \rangle + \beta \langle \mathbf{v}_k \mid \mathbf{v}_j \rangle \rangle^* \quad \text{by (iii)} \\
= \alpha^* \langle \mathbf{v}_i \mid \mathbf{v}_k \rangle + \beta^* \langle \mathbf{v}_j \mid \mathbf{v}_k \rangle \quad \text{by (ii)}
\]

(2.13)

A moment’s thought would tell us that (iii) and (iv) can be extended to the following property.

\[
\left\langle \sum_{i=1}^{n} \alpha_i \mathbf{v}_i \mid \sum_{j=1}^{n} \beta_j \mathbf{v}_j \right\rangle = \sum_{i,j=1}^{n} \alpha_i^* \beta_j \langle \mathbf{v}_i \mid \mathbf{v}_j \rangle
\]

(2.14)

Definition 2.7 (Inner Product Space) A vector space on which an inner product is defined is called an inner product space.

Our canonical example \( \mathbb{R}^3 \) is naturally an inner product space.

Definition 2.8 (Norm) The norm of a vector \( \mathbf{v} \), denoted by either \( |v| \) or \( \|v\| \), is given by \( \langle v \mid v \rangle^{1/2} \). A vector is normalized if its norm is unity. We call such a vector a unit vector.
2.2. INNER PRODUCT SPACES

**Definition 2.9 (Metric Space)** A metric space is a set where a notion of distance (called a "metric") between elements of the set is defined. In particular, a norm is a metric.

**Definition 2.10 (Cauchy Sequence)** A Cauchy sequence \( \{x_n\} \) is a sequence for which the terms become arbitrarily close to each other if \( n \) becomes sufficiently large. For example, a sequence of real numbers \( \{x_n\} \) is Cauchy if for any \( \varepsilon > 0 \), there exists a positive integer \( N_\varepsilon \) such that \( m,n > N_\varepsilon \implies |x_m - x_n| < \varepsilon \). More generally, a sequence \( \{y_n\} \) in a metric space is Cauchy if \( m,n > N_\varepsilon \implies d(y_m, y_n) < \varepsilon \).

**Definition 2.11 (Complete Metric Space)** A metric space \( M \) is called complete if every Cauchy sequence of points in \( M \) converges in \( M \); that is, if every Cauchy sequence in \( M \) has a limit that is also in \( M \).

There is a precise notion of a metric which is a real-valued function with two elements of a set as the arguments. Let us use the canonical notation \( d(x, y) \). Then, a metric \( d \) should satisfy the following conditions.

1. \( d(x, y) \geq 0 \)
2. \( d(x, y) = 0 \) if and only if \( x = y \)
3. \( d(x, y) = d(y, x) \)
4. \( d(x, z) \leq d(x, y) + d(y, z) \)

The first condition is redundant as \( d(x, y) = \frac{1}{2}[d(x, y) + d(y, x)] \geq \frac{1}{2}d(x, x) = 0 \).

**Definition 2.12 (Hilbert Space)** An inner product space which is complete with respect to the norm induced by the inner product is called a Hilbert space.

The inner product space \( \mathbb{R}^3 \) is a Hilbert space.

Let us backtrack here and review the relations among a vector space, an inner product space, and a Hilbert space.

**Vector Space** A (linear) vector space is defined by Definition 2.1. Two trivial examples are \( \mathbb{R}^2 \) the \( x,y \)-plane and \( \mathbb{R}^3 \) the \( x,y,z \)-space.

**Inner Product Space** An inner product space is a vector space on which an inner product is defined according to Definition 2.6. With the usual dot product \( u \cdot w \), \( \mathbb{R}^2 \) and \( \mathbb{R}^3 \) become inner product spaces.

---

\(^a\)There is a precise notion of a metric which is a real-valued function with two elements of a set as the arguments. Let us use the canonical notation \( d(x, y) \). Then, a metric \( d \) should satisfy the following conditions.

\(^6\)A normed vector space is automatically complete if it is finite dimensional.

\(^7\)Parenthetically, a normed complete vector space is called a Banach space.
Hilbert Space  A Hilbert space is an inner product space for which all Cauchy sequences converge to an element of the space. In particular, all finite dimensional inner product spaces are “automatically” Hilbert spaces. Hence, both \( \mathbb{R}^2 \) and \( \mathbb{R}^3 \) are Hilbert spaces.

Some clarifications are in order about the convergence (inside the Hilbert space \( \mathcal{H} \)). This means that if \( \| v_m - v_n \| \to 0 \) as both \( m \) and \( n \) tend to \( \infty \), then there is a vector \( v \in \mathcal{H} \) such that \( \| v_i - v \| \to 0 \). Forgetting about being a vector space for now, the sequence \( \{ \frac{1}{i} \}_{i=1}^{\infty} \) in the open interval \( (0, 1) \) and closed interval \( [0, 1] \) give an example of converging inside the set and not doing so. Noting that \( \frac{1}{i} \to 0 \), we can see that \( \{ \frac{1}{i} \}_{i=1}^{\infty} \) converges inside the set only for \( [0, 1] \). But, as stated above, \( (0, 1) \) and \( [0, 1] \) are not vector spaces.

Now, a good example of an inner product space which is not complete, and hence is not a Hilbert space, is a set of functions defined on an interval \( [a, b] \) with \( k \) continuous derivatives; denoted by \( \mathcal{C}^k([a, b]) \).

**Example 2.1 (\( \mathcal{C}^k([a, b]) \) is not complete.)** Consider a set \( \mathcal{C}^k([a, b]) \) of continuous functions with \( k \) continuous derivatives defined on \( [a, b] \). It is straightforward to show that \( \mathcal{C}^k([a, b]) \) is a linear vector space. Furthermore, it is an inner product space with the following inner product.

\[
\langle f | g \rangle = \sum_{i=0}^{k} \int_{a}^{b} f^i(x)g^i(x) \, dx \quad (2.15)
\]

Here, \( f^i \) and \( g^i \) are the \( i \)-th derivatives of \( f \) and \( g \), respectively. The norm given by this inner product is

\[
\| f \| = \left( \sum_{i=0}^{k} \int_{a}^{b} |f^i(x)|^2 \right)^{\frac{1}{2}}. \quad (2.16)
\]

The corresponding Hilbert space, which is the completion of \( \mathcal{C}^k([a, b]) \) with respect to the norm (2.16), is called a Sobolev space denoted by \( H^k((a, b)) \) or \( W^{k,2}((a, b)) \). It can be shown that \( H^k((a, b)) \) is strictly larger than \( \mathcal{C}^k([a, b]) \).

**Definition 2.13 (Orthogonality)** Two vectors \( u \) and \( w \) are orthogonal to each other if \( \langle u | w \rangle = 0 \).

Note here that in the familiar 2-dimensional and 3-dimensional cases, namely \( \mathbb{V}^2(\mathbb{R}) \) and \( \mathbb{V}^3(\mathbb{R}) \), the norm is nothing but the length of the vector, and orthogonality means the vectors are geometrically perpendicular to each other.
Definition 2.14 (Orthonormal Set) A set of vectors \( \{v_1, v_2, \ldots, v_n\} \) form an orthonormal set if

\[
\langle v_i | v_j \rangle = \delta_{ij};
\]

where the symbol \( \delta_{ij} \) is known as Kronecker’s delta with the following straightforward definition.

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

(2.18)

Definition 2.15 (Completeness) An orthonormal set of vectors \( \{v_1, v_2, \ldots, v_n\} \) is complete if

\[
\sum_{i=1}^{n} |v_i \rangle \langle v_i| = I.
\]

(2.19)

Alternatively, \( \{v_1, v_2, \ldots, v_n\} \) is complete if any vector \( v \) in the Hilbert space can be expressed as a linear combination of \( v_1, v_2, \ldots, v_n \). That is, any \( v = \sum_{i=1}^{n} a_i |v_i \rangle \) if and only if the set \( \{v_1, v_2, \ldots, v_n\} \) satisfies \( \sum_{i=1}^{n} |v_i \rangle \langle v_i| = I \).

Proof
Suppose \( \sum_{i=1}^{n} |v_i \rangle \langle v_i| = I \). Then,

\[
v = I v = \left( \sum_{i=1}^{n} |v_i \rangle \langle v_i| \right) |v\rangle = \sum_{i=1}^{n} \langle v_i|v \rangle |v_i \rangle.
\]

(2.20)

Now, suppose \( v = \sum_{j=1}^{n} a_j |v_j \rangle \) for any \( v \) in the Hilbert space. Then,

\[
\left( \sum_{i=1}^{n} |v_i \rangle \langle v_i| \right) v = \left( \sum_{i=1}^{n} |v_i \rangle \langle v_i| \right) \left( \sum_{j=1}^{n} a_j |v_j \rangle \langle v_j| \right) = \sum_{i,j=1}^{n} |v_i \rangle \langle v_i| a_j |v_j \rangle = \sum_{i,j=1}^{n} a_j |v_i \rangle \langle v_i| \delta_{i,j} = \sum_{j=1}^{n} a_j |v_j \rangle = v.
\]

(2.21)

Because (2.21) holds for an arbitrary vector \( v \), we have \( \sum_{i=1}^{n} |v_i \rangle \langle v_i| = I \) by definition.

Definition 2.16 (Orthonormal Basis) If a basis of a vector space \( \mathbb{V}^n \) forms an orthonormal set, it is called an orthonormal basis.\(^9\)

\(^8\)This is also known as the closure condition.

\(^9\)Note that a set of vectors which is orthonormal and complete forms an orthonormal basis.
The unit vectors along the $x$-, $y$-, and $z$-axes, denoted by $i$, $j$, and $k$, form an orthonormal basis of $\mathbb{R}^3$.

Here is a caveat. Note that the representation of a vector $\mathbf{v}$ as an $n \times 1$ matrix (column vector) or a $1 \times n$ matrix (row vector) can be in terms of any basis. However, an orthonormal basis simplifies computations of inner products significantly. Let $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{w} = (w_1, w_2, \ldots, w_n)$ in the component notation with respect to some orthonormal basis $\{e_1, e_2, \ldots, e_n\}$. Then,

$$
\langle \mathbf{u} | \mathbf{w} \rangle = \left\langle \sum_{i=1}^{n} u_i e_i | \sum_{j=1}^{n} w_j e_j \right\rangle = \sum_{i,j=1}^{n} u_i^* w_j \langle e_i | e_j \rangle = \sum_{i,j=1}^{n} u_i^* w_j \delta_{ij} = \sum_{i=1}^{n} u_i^* w_i \tag{2.22}
$$

$$
= u_1^* w_1 + u_2^* w_2 + \ldots + u_{n-1}^* w_{n-1} + u_n^* w_n. 
$$

**Theorem 2.3 (Schwarz Inequality)** Any inner product satisfies the following inequality known as Schwarz Inequality.

$$
|\langle \mathbf{v}_i | \mathbf{v}_j \rangle|^2 \leq |\mathbf{v}_i|^2 |\mathbf{v}_j|^2 \quad \text{or} \quad |\langle \mathbf{v}_i | \mathbf{v}_j \rangle| \leq |\mathbf{v}_i||\mathbf{v}_j| \tag{2.23}
$$

**Proof**

Consider the vector

$$
\mathbf{v} = \mathbf{v}_i - \langle \mathbf{v}_j | \mathbf{v}_i \rangle \frac{\mathbf{v}_j}{|\mathbf{v}_j|^2}.
$$

Then, as $\langle \mathbf{v} | \mathbf{v} \rangle \geq 0$, we get

$$
\langle \mathbf{v}_i - \langle \mathbf{v}_j | \mathbf{v}_i \rangle \frac{\mathbf{v}_j}{|\mathbf{v}_j|^2} | \mathbf{v}_i - \langle \mathbf{v}_j | \mathbf{v}_i \rangle \frac{\mathbf{v}_j}{|\mathbf{v}_j|^2} \rangle = \langle \mathbf{v}_i | \mathbf{v}_i \rangle - \frac{\langle \mathbf{v}_j | \mathbf{v}_i \rangle \langle \mathbf{v}_i | \mathbf{v}_j \rangle}{|\mathbf{v}_j|^2} - \frac{\langle \mathbf{v}_j | \mathbf{v}_i \rangle^* \langle \mathbf{v}_j | \mathbf{v}_i \rangle}{|\mathbf{v}_j|^2} + \frac{\langle \mathbf{v}_j | \mathbf{v}_i \rangle^* \langle \mathbf{v}_j | \mathbf{v}_i \rangle}{|\mathbf{v}_j|^2}
$$

$$
= |\mathbf{v}_i|^2 - \frac{\langle \mathbf{v}_j | \mathbf{v}_i \rangle \langle \mathbf{v}_i | \mathbf{v}_j \rangle}{|\mathbf{v}_j|^2} - \frac{\langle \mathbf{v}_j | \mathbf{v}_i \rangle^* \langle \mathbf{v}_j | \mathbf{v}_i \rangle}{|\mathbf{v}_j|^2} + \frac{\langle \mathbf{v}_j | \mathbf{v}_i \rangle^* \langle \mathbf{v}_j | \mathbf{v}_i \rangle}{|\mathbf{v}_j|^2}
$$

$$
= |\mathbf{v}_i|^2 - \frac{\langle \mathbf{v}_j | \mathbf{v}_i \rangle \langle \mathbf{v}_i | \mathbf{v}_j \rangle}{|\mathbf{v}_j|^2} \geq 0 \implies |\langle \mathbf{v}_i | \mathbf{v}_j \rangle|^2 \geq |\langle \mathbf{v}_i | \mathbf{v}_j \rangle|\langle \mathbf{v}_i | \mathbf{v}_j \rangle. \tag{2.24}
$$
We can also see that the equality holds only when
\[ \mathbf{v} = \mathbf{v}_i - \langle v_j | v_i \rangle \frac{\mathbf{v}_j}{|v_j|^2} = 0 \quad \text{or} \quad \mathbf{v}_i = \langle v_j | v_i \rangle \frac{\mathbf{v}_j}{|v_j|^2}. \tag{2.25} \]

Now, let \( \mathbf{v}_i = \alpha \mathbf{v}_j \) for some scalar \( \alpha \). Then,
\[ \langle v_j | v_i \rangle \frac{\mathbf{v}_j}{|v_j|^2} = \langle v_j | \alpha \mathbf{v}_j \rangle \frac{\mathbf{v}_j}{|v_j|^2} = \langle v_j | v_j \rangle \frac{\alpha \mathbf{v}_j}{|v_j|^2} = |v_j|^2 \frac{\alpha \mathbf{v}_j}{|v_j|^2} = \alpha \mathbf{v}_j - \mathbf{v}_i. \tag{2.26} \]
Therefore, the equality holds if and only if one vector is a scalar multiple of the other.\(^b\)

\(^a\)Don’t make a beginner’s mistake of applying this to vectors expressed in a basis that is not orthonormal. For example, consider the space spanned by \( e_1 \) and \( e_2 \); i.e. the \( x, y \)-plane if you may. While \( \{e_1, e_2\} \) form an orthonormal basis, \( \{e_1, e_{1+e_2}/\sqrt{2}\} \) is a basis composed of normalized vectors which are not mutually orthogonal. Let \( \mathbf{u} = e_1 \) and \( \mathbf{w} = e_{1+e_2}/\sqrt{2} \). Then, \( \mathbf{u} = (1, 0) \) and \( \mathbf{w} = (0, 1) \), and \( u_1^*w_1 + u_2^*w_2 = 1 \times 0 + 0 \times 1 = 0 \). But, \( \mathbf{u} \cdot \mathbf{w} = e_1 \cdot \left( e_{1+e_2}/\sqrt{2} \right) = e_1 \cdot e_{1+e_2}/\sqrt{2} = 1/\sqrt{2} \neq 0 \).

\(^b\)This is not surprising as \( \langle v_j | v_i \rangle \frac{\mathbf{v}_j}{|v_j|^2} \) is the projection of \( \mathbf{v}_i \) along \( \mathbf{v}_j \). Needless to say, if \( \mathbf{v}_i \) is parallel to \( \mathbf{v}_j \), its projection in the direction of \( \mathbf{v}_j \) is \( \mathbf{v}_i \) itself.

### 2.3 \( L^2 \)-Space

It was mentioned on p.23 that there are other inner products than the familiar one on \( \mathbb{R}^3 \). And, there is an associated Hilbert space for each of those inner products. One important class of Hilbert spaces for us is the \( L^2 \)-space. It is a collection of functions with a common domain, for which an inner product is defined as an integral. Most of the Hilbert spaces we encounter in this course are \( L^2 \)-spaces. In the following, we will specialize to \( \mathbb{R} \) as the domain of our functions, but the same definitions and descriptions apply to a general domain\(^{10}\). In particular, if the domain is \( \mathbb{R}^2 \) or \( \mathbb{R}^3 \), we have a double integral, \( \iint_{-\infty}^{+\infty} \), or triple integral, \( \iiint_{-\infty}^{+\infty} \), respectively, rather than a single integral shown here.

\(^{10}\)To be precise, the domain should be a measure space so that an integral can exist. However, such purely mathematical details can be omitted as all the domains we consider are measure spaces.
**Definition 2.17 (L²-Function)** An L²-function $f$ is a square integrable function; that is, $f$ satisfies $\int_{-\infty}^{\infty} |f(x)|^2 \, dx < \infty$. ¹¹

The next step is to define an inner product on this set of square integrable functions.

**Definition 2.18 (L²-Inner Product)** Given two square integrable functions $f$ and $g$, we define an inner product by

$$\langle f | g \rangle := \int_{-\infty}^{+\infty} f(x)^* g(x) \, dx^{12};$$

(2.27)

where $f(x)^*$ signifies the complex conjugate.

With this inner product defined on a set of L²-functions, we have a Hilbert space.

As stated in the beginning of this section, the common domain of the functions can be any other measurable set such as $\mathbb{R}^2$, $\mathbb{R}^3$, $[0, 1]$, and $(-\frac{a}{2}, +\frac{a}{2})$ for $a > 0$. Using the notation given in Footnote 11, we will denote these L²-spaces by $L^2(\mathbb{R}^2)$, $L^2(\mathbb{R}^3)$, $L^2[0, 1]$, and $L^2(-\frac{a}{2}, +\frac{a}{2})$.

The central equation of quantum mechanics, called Schrödinger Equation which we will encounter in Chapter 5, is a differential equation, and its solutions are complex valued functions on a common domain. Therefore, the L²-space is our canonical Hilbert space.

Some examples of orthonormal bases for L²-spaces are as follows.

1. The trigonometric system $\{e^{i2\pi nx}\}_{n=-\infty}^{+\infty}$ is an orthonormal basis for $L^2[0, 1]$. The expansion of a function in this basis is called the Fourier series of that function.

2. The Legendre polynomials, which are obtained by taking the sequence of monomials $\{1, x, x^2, x^3, \ldots\}$ and applying the Gram-Schmidt orthogonalization process to it, form an orthonormal basis for $L^2[-1, 1]$. The first few normalized polynomials are as follows.

$$p_1(x) = \sqrt{1/2}$$

---

¹¹We often have an interval of finite length as the domain of $f$. The closed interval $[0, 1]$ is a canonical example. In this case, we require $\int_0^1 |f(x)|^2 \, dx < \infty$. We use notations such as $L^2(-\infty, +\infty)$ and $L^2[0, 1]$ to make this distinction.

¹²We can easily check that this integral satisfies Definition 2.6.
2.4 The Braket Notation

This is a notation first used by Dirac. In fact, we have already been using it. You may have noticed that the inner product between \( v \) and \( w \) was denoted by \( \langle v \mid w \rangle \), and not by the more familiar \( \langle v, w \rangle \). We will see why shortly. Without much ado, let us define a “bra” and a “ket”. On p.23, we have already encountered \( \langle a \mid = a^\ast \) and \( \mid b \rangle = b \) such that the inner product between \( a \) and \( b \) is given by \( \langle a \mid b \rangle = a^\ast_1 b_1 + a^\ast_2 b_2 + a^\ast_3 b_3 \). Here, \( \langle a \mid \) is called “bra \( a \)”, and \( \mid b \rangle \) is called “ket \( b \)”. More generally, recall the component notation of a vector \( v = (v_1, v_2, \ldots, v_n) \). This usually means that we have picked a particular orthonormal basis\(^1\) \( \{e_1, e_2, \ldots, e_n\} \) and expressed \( v \) as a linear combination of \( e_1, e_2, \ldots, e_n \) such that \( v = v_1 e_1 + v_2 e_2 + \ldots + v_n e_n \). We can arrange the components either horizontally or vertically, but it is more convenient for us to arrange it vertically as an \( n \) by 1 matrix, often called a column vector. In quantum mechanics, we denote this column vector by \( \mid v \rangle \) called “ket \( v \)”.\(^2\)

\[
|v\rangle = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \quad (2.28)
\]

Now, what is “bra \( v \)” denoted by \( \langle v \mid \)? It is defined as the row vector whose components are the complex conjugates of the components of \( v \).

\[
\langle v \mid = [v^\ast_1, \ldots, v^\ast_n] \quad (2.29)
\]

\(^1\)Note that this is “bra + ket = braket”, and NOT “bracket”.

\(^2\)Actually, it does not have to be an orthonormal basis, but any basis would do. However, we will specialize to an orthonormal basis here in order to explain why we used the notation \( \langle u \mid |v\rangle \) for the inner product. Besides, orthonormal bases are by far the most common and useful bases.
In this notation, a basis vector $e_i$ has the following bra and ket. We sometimes denote $\langle e_i \rangle$ and $\ | e_i \rangle$ by $\langle i \rangle$ and $\ | i \rangle$, respectively.

$$\langle e_i \rangle = \langle i \rangle = [0, 0, \ldots, 1, \ldots, 0] \quad \text{and} \quad |e_i\rangle = |i\rangle = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

(2.30)

The only nonzero element 1 is in the $i$-th place. We can express $|v\rangle$ as a linear combination of $|i\rangle$'s.

$$|v\rangle = \sum_{i=1}^{n} v_i |i\rangle$$

(2.31)

Similarly,

$$\langle v | = \sum_{i=1}^{n} v_i^* \langle i | .$$

(2.32)

This suggests that we should define $|\alpha v\rangle$ and $\langle \alpha v |$ by

$$|\alpha v\rangle = \alpha |v\rangle \quad \text{and} \quad \langle \alpha v | = \alpha^* \langle v | = \langle v | \alpha^*.$$  

(2.33)

As shown above, it is customary, for conserving symmetry, to write $\langle v | \alpha^*$ rather than $\alpha^* \langle v |$. We say $\alpha |v\rangle$ and $\langle v | \alpha^*$ are adjoints of each other. More generally, the adjoint of

$$\sum_{i=1}^{n} \alpha_i |v_i\rangle$$

is

$$\sum_{i=1}^{n} \langle v_i | \alpha_i^* ,$$

(2.35)

and vice versa. In order to take the adjoint, we can simply replace each scalar by its complex conjugate and each bra/ket by the corresponding ket/bra.
Let us now try some manipulations with bras and kets in order to get used to their properties.

First, consider two vectors \( \mathbf{u} = u_1 \mathbf{e}_1 + u_2 \mathbf{e}_2 + \ldots + u_n \mathbf{e}_n \) and \( \mathbf{w} = w_1 \mathbf{e}_1 + w_2 \mathbf{e}_2 + \ldots + w_n \mathbf{e}_n \) in an \( n \)-dimensional vector space spanned by an orthonormal basis \( \{ \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n \} \). We have

\[
\langle \mathbf{u} \mid \mathbf{w} \rangle = [u_1^*, \ldots, u_n^*] \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = u_1^* w_1 + u_2^* w_2 + \ldots + u_n^* w_n.
\] (2.36)

Hence,

\[
\langle \mathbf{u} | \mathbf{w} \rangle = [u_1^*, \ldots, u_n^*] \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = u_1^* w_1 + u_2^* w_2 + \ldots + u_n^* w_n.
\] (2.37)

We drop one vertical bar form \( \langle \mathbf{u} | \mathbf{w} \rangle \) and write \( \langle \mathbf{u} | \mathbf{w} \rangle \), so that

\[
\langle \mathbf{u} \rangle = [u_1^*, \ldots, u_n^*] \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = u_1^* w_1 + u_2^* w_2 + \ldots + u_n^* w_n.
\] (2.38)

As you can see, the right-hand side is the inner product. Hence, the braket notation is consistent with the notation we have been using for the inner product. In passing, we will make a note of the obvious fact that

\[
\langle e_i \mid e_j \rangle = \langle i \mid j \rangle = \delta_{ij}.
\] (2.39)

Next, given an orthonormal basis \( \{|1\rangle, |2\rangle, \ldots, |n\rangle\} \), any vector \( |v\rangle \) can be expanded in this basis so that

\[
|v\rangle = \sum_{i=1}^{n} v_i |i\rangle,
\] (2.40)

where \( v_i \)'s are unique. Taking the inner product of both sides with \( \langle j \rangle \), we get

\[
\langle j \mid v \rangle = \sum_{i=1}^{n} v_i \langle i \mid j \rangle = \sum_{i=1}^{n} v_i \delta_{ij} = v_j.
\] (2.41)

Plugging this back into \( |v\rangle = \sum_{i=1}^{n} v_i |i\rangle \),

\[
|v\rangle = \sum_{i=1}^{n} \langle i \mid v \rangle |i\rangle = \sum_{i=1}^{n} |i\rangle \langle i \mid v \rangle.
\] (2.42)
The second equality in Equation 2.42 may not be clear. So, we will show it below explicitly by writing out each vector. The term on the left-hand side is as follows.

\[
\langle i | v \rangle | i \rangle = \begin{bmatrix} 0, 0, \ldots, 1, \ldots, 0 \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = v_i \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \end{bmatrix} = v_i \begin{bmatrix} 0 \\ 0 \\ \vdots \\ v_i \\ 0 \end{bmatrix} \tag{2.43}
\]

Note that the only nonzero element 1 is in the \( i \)-th place. On the other hand, the term on the right-hand side is

\[
|i \rangle \langle i| v \rangle = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \begin{bmatrix} 0, 0, \ldots, 1, \ldots, 0 \\ v_1 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ v_i \\ \vdots \\ 0 \end{bmatrix} = v_i \begin{bmatrix} 0 \\ 0 \\ \vdots \\ v_i \\ \vdots \\ 0 \end{bmatrix} \tag{2.44}
\]

Hence, the second equality in Equation 2.42 indeed holds term by term. In the case of (2.44), we have a product of an \( n \)-by-1 matrix \(|i\rangle\), a 1-by-\( n \) matrix \( \langle i|\), and an \( n \)-by-1 matrix \(|v\rangle\). By the associative law of matrix multiplication, it is possible to
compute $|i\rangle \langle i|$ first.

$$|i\rangle \langle i| v = \left( \begin{array}{c} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{array} \right) [0, 0, \ldots, 1, \ldots, 0] \left[ \begin{array}{c} v_1 \\ \vdots \\ v_n \end{array} \right]$$

(2.45)

$$= \left[ \begin{array}{cccc} 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \end{array} \right] \left[ \begin{array}{c} v_1 \\ \vdots \\ v_n \end{array} \right] = [\delta_{ij}] \left[ \begin{array}{c} v_1 \\ \vdots \\ v_n \end{array} \right] = \left[ \begin{array}{c} 0 \\ 0 \\ \vdots \\ v_i \\ \vdots \end{array} \right]$$

The symbol $\delta_{ij}$ in (2.45) means an $n$ by $n$ matrix whose only nonzero entry is the 1 at the intersection of the $i$-th row and the $j$-th column.

$$[\delta_{ij}] := \left[ \begin{array}{cccc} 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \end{array} \right]$$

(2.46)

It is easy to see that

$$|i\rangle \langle j| = [\delta_{ij}].$$

(2.47)

\footnote{As (2.45) shows, $|i\rangle \langle i|$ extracts the $i$-th component of vector $v$. So, $|i\rangle \langle i|$ is called the projection operator for the ket $|i\rangle$.}
And, it is also easy to see that
\[
\sum_{i=1}^{n} |i\rangle \langle i| = \sum_{i=1}^{n} [\delta_{ii}] = I_n;
\] (2.48)

where \(I_n\) is the \(n\)-dimensional identity operator.\(^{16}\) With this identity, Equation 2.42 reduces to triviality.

\[
|v\rangle = I|v\rangle = \left( \sum_{i=1}^{n} |i\rangle \langle i| \right) |v\rangle = \left( \sum_{i=1}^{n} |i\rangle \langle i| |v\rangle \right) = \sum_{i=1}^{n} |i\rangle \langle i| |v\rangle
\] (2.49)

Because the identity matrix can be inserted anywhere without changing the outcome of the computation, we can insert \(\sum_{i=1}^{n} |i\rangle \langle i|\) anywhere during our computation. This seemingly trivial observation combined with the associativity of matrix multiplication will prove useful later in this book.

**Example 2.2** (\(|i\rangle, |i\rangle, \text{and } |i\rangle \langle i| \text{ in Two Dimensions}) In two dimensions, we have

\[
|1\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, |2\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \langle 1| = \begin{bmatrix} 1 & 0 \end{bmatrix}, \text{ and } |2\rangle = \begin{bmatrix} 0 & 1 \end{bmatrix}
\] (2.50)

Then,

\[
\sum_{i=1}^{2} |i\rangle \langle i| = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.
\] (2.51)

Now, consider any vector \(\mathbf{v} = (v_1, v_2)\) and compute \(|1\rangle \langle 1| \mathbf{v} \) and \(|2\rangle \langle 2| \mathbf{v} \).

\[
|1\rangle \langle 1| \mathbf{v} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} v_1 \\ 0 \end{bmatrix}
\] (2.52)

\[
|2\rangle \langle 2| \mathbf{v} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ v_2 \end{bmatrix}
\] (2.53)

---

\(^{16}\)When there is no possibility of confusion, we will simply use \(I\) instead of \(I_n\).
Let us consider the adjoint of (2.40).

\[ \langle v \rangle = \sum_{i=1}^{n} \langle i \rangle v_{i}^{*} \quad (2.54) \]

From our previous computation and due to a property of the inner product, we get

\[ \langle v \rangle = \sum_{i=1}^{n} \langle i \rangle (\langle i \rangle |v\rangle)^{*} = \sum_{i=1}^{n} \langle i \rangle |v\rangle \langle i \rangle = \sum_{i=1}^{n} |v\rangle \langle i \rangle . \quad (2.55) \]

Again, recall the identity \( \sum_{i=1}^{n} |i\rangle \langle i \rangle = \sum_{i=1}^{n} [\delta_{ii}] = I \). So,

\[ \sum_{i=1}^{n} |v\rangle \langle i \rangle = \langle v \rangle \left( \sum_{i=1}^{n} |i\rangle \langle i \rangle \right) = \langle v \rangle I = \langle v \rangle . \quad (2.56) \]

It is convenient at this point to have a rule by which we can mechanically arrive at the adjoint of an expression.

**Taking the Adjoint:**

1. *Reverse the order of all the factors in each term, including numerical coefficients, bras, and kets.*

2. *Take complex conjugate of the numerical coefficients.*

3. *Turn bras into kets and kets into bras.*

The following example demonstrates what one should do.

Consider

\[ \alpha |v\rangle = a |u\rangle + b |w\rangle \langle s|t \rangle + \ldots . \quad (2.57) \]

What is the adjoint?

**After Step 1**

\[ |v\rangle \alpha = |u\rangle a + |t\rangle \langle s|w \rangle b + \ldots \quad (2.58) \]

**After Step 2**

\[ |v\rangle \alpha^{*} = |u\rangle a^{*} + |t\rangle \langle s|w \rangle b^{*} + \ldots \quad (2.59) \]
After Step 3

\[ \langle v | \alpha^* = \langle u | a^* + \langle t | s \rangle \langle w | b^* + \ldots \) \] (2.60)

Therefore, the adjoint of

\[ \alpha | v \rangle = a | u \rangle + b | w \rangle \langle s | t \rangle + \ldots \] (2.61)

is

\[ \langle v | \alpha^* = \langle u | a^* + \langle t | s \rangle \langle w | b^* + \ldots \) \] (2.62)

**Theorem 2.4 (Gram-Schmidt Theorem)** Given \( n \) linearly independent vectors \( \{|v_1\rangle, |v_2\rangle, \ldots, |v_n\rangle\} \), we can construct \( n \) orthonormal vectors \( \{|1\rangle, |2\rangle, \ldots, |n\rangle\} \), each of which is a linear combination of \( |v_1\rangle, |v_2\rangle, \ldots, |v_n\rangle \). In particular, we can construct \( \{|1\rangle, |2\rangle, \ldots, |n\rangle\} \) such that

\[ |1\rangle = \frac{|v_1\rangle}{\sqrt{\langle v_1 | v_1 \rangle}}. \]

**Proof**

We will construct \( n \) mutually orthogonal vectors \( \{|1'\rangle, |2'\rangle, \ldots, |n'\rangle\} \) first. Once it is done, we will only need to normalize each vector to make them orthonormal. First let

\[ |1'\rangle = |v_1\rangle. \] (2.63)

Next, let

\[ |2'\rangle = |v_2\rangle - \frac{|1'\rangle \langle 1' | v_2 \rangle}{\langle 1' | 1' \rangle}, \] (2.64)

where

\[ \frac{|1'\rangle \langle 1' | v_2 \rangle}{\langle 1' | 1' \rangle} \]

is nothing but the projection of \( |v_2\rangle \) on \( |1'\rangle \). So, in (2.64), the component of \( |v_2\rangle \) in the direction of \( |1'\rangle \) is subtracted from \( |v_2\rangle \), making it orthogonal to \( |1'\rangle \). Indeed,

\[ \langle 1' | 2' \rangle = \langle 1' | v_2 \rangle - \frac{\langle 1' | 1' \rangle \langle 1' | v_2 \rangle}{\langle 1' | 1' \rangle} = 0. \] (2.66)

The third ket is defined proceeding in the same manner.

\[ |3'\rangle = |v_3\rangle - \frac{|1'\rangle \langle 1' | v_3 \rangle}{\langle 1' | 1' \rangle} - \frac{|2'\rangle \langle 2' | v_3 \rangle}{\langle 2' | 2' \rangle}. \] (2.67)
Then,
\[
\langle 1' | 3' \rangle = \langle 1' | v_3 \rangle - \frac{\langle 1' | 1' \rangle \langle 1' | v_3 \rangle}{\langle 1' | 1' \rangle} - \frac{\langle 1' | 2' \rangle \langle 2' | v_3 \rangle}{\langle 2' | 2' \rangle} = \langle 1' | v_3 \rangle - \langle 1' | v_3 \rangle - 0 = 0
\]
(2.68)
and
\[
\langle 2' | 3' \rangle = \langle 2' | v_3 \rangle - \frac{\langle 2' | 1' \rangle \langle 1' | v_3 \rangle}{\langle 2' | 2' \rangle} = \langle 2' | v_3 \rangle - \langle 2' | v_3 \rangle - 0 = 0.
\]
(2.69)

If \( n \leq 3 \), we are done at this point. So, assume \( n \geq 4 \), and suppose we have found \( i \) mutually orthogonal vectors \( |1'\rangle, |2'\rangle, \ldots, |i'\rangle \) for \( i \leq n - 1 \). Now, define \( |(i+1)\rangle \) as follows.
\[
|((i+1))\rangle = |v_{i+1}\rangle - \frac{|1'\rangle \langle 1' | v_{i+1} \rangle}{\langle 1' | 1' \rangle} - \frac{|2'\rangle \langle 2' | v_{i+1} \rangle}{\langle 2' | 2' \rangle} - \ldots - \frac{|i'\rangle \langle i' | v_{i+1} \rangle}{\langle i' | i' \rangle}
\]
(2.70)

For any \( 1 \leq j \leq i \),
\[
\langle j' | (i+1) \rangle = \langle j' | v_{i+1} \rangle - \sum_{k=1}^{i} \frac{\langle j' | k' \rangle \langle k' | v_{i+1} \rangle}{\langle k' | k' \rangle} = \langle j' | v_{i+1} \rangle - \frac{\langle j' | j' \rangle \langle j' | v_{i+1} \rangle}{\langle j' | j' \rangle} = 0.
\]
(2.71)

Hence, by mathematical induction, we can find \( n \) mutually orthogonal vectors \( |1\rangle, |2\rangle, \ldots, |n\rangle \). Finally define \( |i\rangle \) for \( 1 \leq i \leq n \) by
\[
|i\rangle = \frac{|i'\rangle}{\langle i' | i' \rangle^{1/2}},
\]
(2.72)
so that
\[
\langle i | i \rangle = \frac{\langle i' | i' \rangle}{\langle i' | i' \rangle^{1/2} \langle i' | i' \rangle^{1/2}} = 1.
\]
(2.73)
We have now constructed \( n \) orthonormal vectors \( |1\rangle, |2\rangle, \ldots, |n\rangle \).

**Question: Linear Independence**

If you are a careful reader, you may have noticed that we did not, at least explicitly,
use the fact that the vectors \( \{ |v_1 \rangle, |v_2 \rangle, \ldots, |v_n \rangle \} \) are linearly independent. So, why is this condition necessary? In order to see this, suppose

\[
|i+1\rangle' = |v_{i+1}\rangle - \frac{|1\rangle \langle 1' | v_{i+1}\rangle}{\langle 1'| 1' \rangle} - \frac{|2\rangle \langle 2' | v_{i+1}\rangle}{\langle 2'| 2' \rangle} - \cdots - \frac{|i'\rangle \langle i' | v_{i+1}\rangle}{\langle i'| i' \rangle} = 0 \tag{2.74}
\]

for some \( i \leq n - 1 \). Because our construction of \( |i\rangle \) requires \( |v_i\rangle \), this means we have at most \( n - 1 \) orthonormal vectors, and our construction fails. However, we know that \( |j\rangle \) is a linear combination of \( |v_1\rangle, |v_2\rangle, \ldots, |v_j\rangle \) for any \( j \) by construction. Therefore,

\[
-\frac{|1\rangle \langle 1' | v_{i+1}\rangle}{\langle 1'| 1' \rangle} - \frac{|2\rangle \langle 2' | v_{i+1}\rangle}{\langle 2'| 2' \rangle} - \cdots - \frac{|i'\rangle \langle i' | v_{i+1}\rangle}{\langle i'| i' \rangle} = 0 \tag{2.75}
\]

in (2.74) is a linear combination of \( |v_1\rangle, |v_2\rangle, \ldots, |v_i\rangle \). So, \( |(i+1)\rangle' \) is a linear combination of \( |v_1\rangle, |v_2\rangle, \ldots, |v_i\rangle \) where the coefficient of \( |v_{i+1}\rangle \) is 1. But, this means that there is a linear combination of \( |v_1\rangle, |v_2\rangle, \ldots, |v_{i+1}\rangle \) which equals \( 0 \) though not all the coefficients are zero, violating the linear independence assumption for \( \{ |v_1\rangle, |v_2\rangle, \ldots, |v_n\rangle \} \). Turning this inside out, we can conclude that \( |(i+1)\rangle' \neq 0 \) for all \( i \leq n - 1 \) if \( \{ |v_1\rangle, |v_2\rangle, \ldots, |v_n\rangle \} \) are linearly independent. This is why we needed the linear independence of \( \{ |v_1\rangle, |v_2\rangle, \ldots, |v_n\rangle \} \) in Theorem 2.4.

---

\(^a\)In the familiar \( x, y \)-plane,

\[
\frac{|1\rangle \langle 1' | v_2\rangle}{\langle 1'| 1' \rangle} = \frac{\|1'\| \|v_2\| \cos \theta}{\|1'\|^2} = \|v_2\| \cos \theta \hat{1}',
\]

where \( \theta \) is the angel between \( 1' \) and \( v_2 \), and \( \hat{1}' \) is the unit vector in the direction of \( 1' \).

**Theorem 2.5** The maximum number of mutually orthogonal vectors in an inner product space equals the maximum number of linearly independent vectors.

**Proof**

Consider an \( n \)-dimensional space \( \mathbb{V}^n \), which admits a maximum of \( n \) linearly independent vectors by definition. Suppose we have \( k \) mutually orthogonal vectors \( |v_1\rangle, |v_2\rangle, \ldots, |v_k\rangle \), and assume

\[
\sum_{i=1}^{k} \alpha_i |v_i\rangle = 0. \tag{2.76}
\]
Then,
\[
\langle v_j \left| \sum_{i=1}^{k} \alpha_i |v_i\rangle \right. \rangle = \sum_{i=1}^{k} \alpha_i \langle v_j \left| v_i \right. \rangle = \alpha_j \langle v_j \left| v_j \right. \rangle = 0 \implies \alpha_j = 0 \quad \text{for} \quad j = 1, 2, \ldots, k.
\]

(2.77)

So, the vectors $|1\rangle, |2\rangle, \ldots, |k\rangle$ are linearly independent. This implies that \( k \leq n \) as \( V^n \) can only admit up to \( n \) linearly independent vectors. On the other hand, Theorem 2.4 (Gram-Schmidt Theorem) assures that we can explicitly construct \( n \) mutually orthogonal vectors. This proves the theorem.

### 2.5 Vector Subspace

**Definition 2.19 (Vector Subspaces)** A subset of a vector space \( V \) that form a vector space is called a subspace. Our notation for a subspace \( i \) of dimensionality \( n_i \) is \( V_{n_i}^i \).

**Definition 2.20 (Orthogonal Complement)** The orthogonal complement of a subspace \( W \) of a vector space \( V \) equipped with a bilinear form \( B \), of which an inner product is the most frequently encountered example, is the set \( W_\perp \) of all vectors in \( V \) that are orthogonal to every vector in \( W \). It is a subspace of \( V \).

**Definition 2.21 (Direct Sum)** The direct sum of two vector spaces \( V \) and \( W \) is the set \( V \oplus W \) of pairs of vectors \((v, w)\) in \( V \) and \( W \), with the operations:

\[
(v, w) + (v', w') = (v + v', w + w')
\]

\[
c(v, w) = (cv, cw)
\]

With these operations, the direct sum of two vector spaces is a vector space.

### 2.6 Linear Operators

An operator \( \Omega \) operates on a vector \( |u\rangle \) and transforms it to another vector \( |v\rangle \). We denote this as below.

\[
\Omega |u\rangle = |v\rangle
\]

(2.78)

The same operator \( \Omega \) can also act on \( \langle u| \) to generate \( \langle v| \). In this case, we will reverse the order and write

\[
\langle u| \Omega = \langle v|.
\]

(2.79)
For an operator $\Omega$ to be a linear operator, it should have the following set of properties.

\[ \Omega (\alpha | u \rangle) = \alpha (\Omega | u \rangle) \]  
\[ \Omega \{ \alpha | u \rangle + \beta | v \rangle \} = \alpha (\Omega | u \rangle) + \beta (\Omega | v \rangle) \]  
\[ (\langle u | \alpha) \Omega = (\langle u | \Omega) \alpha \]  
\[ (\langle u | \alpha + \langle v | \beta) \Omega = (\langle u | \Omega) \alpha + (\langle v | \Omega) \beta \]  

Recall that $| u \rangle$ is a column vector and $\langle u |$ is a row vector. So, the best way to understand a linear operator is to regard it as an $n \times n$ square matrix. Then, its action on $| u \rangle$ and $\langle u |$ as well as linearity follows naturally. The significance of linearity of an operator $\Omega$ is that its action on any vector is completely specified by how it operates on the basis vectors $|1\rangle, |2\rangle, \ldots, |n\rangle$. Namely, if

\[ |v\rangle = \sum_{i=1}^{n} \alpha_i |i\rangle , \]  
then,

\[ \Omega |v\rangle = \Omega \left( \sum_{i=1}^{n} \alpha_i |i\rangle \right) = \sum_{i=1}^{n} \alpha_i \Omega |i\rangle . \]  
The action of the product of two operators $\Lambda \Omega$ on a ket $|u\rangle$ is defined in an obvious way.

\[ (\Lambda \Omega) |u\rangle = \Lambda (\Omega |u\rangle) \]  
$\Omega$ acts on $|u\rangle$ first followed by $\Lambda$. This is an obvious associative law if the operators $\Lambda$ and $\Omega$ are regarded as matrices as suggested already. With this view, $\Lambda \Omega$ is nothing but a matrix multiplication, and as such, $\Lambda \Omega$ is not the same as $\Omega \Lambda$ generally speaking.

**Definition 2.22 (Commutator)** The commutator of operators $\Lambda$ and $\Omega$, denoted by $[\Lambda, \Omega]$ is defined as follows.

\[ [\Lambda, \Omega] := \Lambda \Omega - \Omega \Lambda \]  
When the commutator $[\Lambda, \Omega]$ is zero, we say $\Lambda$ and $\Omega$ commute, and the order of operation does not affect the outcome. We have the following commutator identities.

\[ [\Omega, \Lambda \Gamma] = \Lambda [\Omega, \Gamma] + [\Omega, \Lambda] \Gamma \]  
\[ [\Lambda \Omega, \Gamma] = \Lambda [\Omega, \Gamma] + [\Lambda, \Gamma] \Omega \]
Definition 2.23 (Inverse) The inverse $\Omega^{-1}$ of an operator $\Omega$ satisfies

$$\Omega\Omega^{-1} = \Omega^{-1}\Omega = I; \quad (2.90)$$

where $I$ is the identity operator such that $I|u\rangle = |u\rangle$ and $\langle u| I = \langle u|$. Needless to say, $I$ corresponds to the identity matrix. Note that not all operators have an inverse.

Note that the inverse satisfies $(\Omega\Lambda)^{-1} = \Lambda^{-1}\Omega^{-1}$, which is consistent with the matrix representations of $\Omega$ and $\Lambda$.

2.7 Matrix Representation of Linear Operators

Recall that the action of a linear operator $\Omega$ is completely specified once its action on a basis is specified. Pick a basis $\{|1\rangle, |2\rangle, \ldots, |n\rangle\}$ of normalized vectors which are not necessarily mutually orthogonal; that is, it is not necessarily an orthonormal basis. The action of $\Omega$ on any ket $|v\rangle$ is completely specified if we know the coefficients $\Omega_{ji}$ for $1 \leq i, j \leq n$ such that

$$\Omega |i\rangle = \sum_{j=1}^{n} \Omega_{ji} |j\rangle. \quad (2.91)$$

The coefficients in (2.91) can be computed as follows.

$$\langle k| \Omega |i\rangle = \langle k| \sum_{j=1}^{n} \Omega_{ji} |j\rangle = \sum_{j=1}^{n} \Omega_{ji} \langle k| |j\rangle = \sum_{j=1}^{n} \Omega_{ji} \delta_{kj} = \Omega_{ki} \quad (2.92)$$

So, $\Omega_{ji} = \langle j| \Omega |i\rangle$ in (2.91). Note that $\langle k| |j\rangle$ was used instead of $\langle k| |j\rangle$ to make it clear that this matrix product

$$\begin{pmatrix}
1 & k & n \\
0 & 0 & \ldots & 1 & \ldots & 0
\end{pmatrix}
\begin{pmatrix}
0 \\
0 \\
\vdots \\
1 \\
\vdots \\
0
\end{pmatrix} = \begin{pmatrix}
1 \\
0 \\
\vdots \\
j \\
\vdots \\
n
\end{pmatrix}$$
is not necessarily an inner product unless our basis happens to be orthonormal.\(^{17}\)

There is another way to find \(\Omega_{ji}\). From (2.48), we have

\[
\sum_{i=1}^{n} |i\rangle \langle i| = I,
\]

and so,

\[
\Omega |i\rangle = I \Omega |i\rangle = \sum_{j=1}^{n} \langle j| \Omega |i\rangle \langle i| = \sum_{j=1}^{n} \Omega_{ji} v_i |j\rangle.
\] (2.93)

Now, let

\[
|v\rangle = \sum_{i=1}^{n} v_i |i\rangle \quad \text{and} \quad |v'\rangle = \Omega |v\rangle = \sum_{j=1}^{n} v'_j |j\rangle.
\]

Then,

\[
|v'\rangle = \Omega |v\rangle = \Omega \left( \sum_{i=1}^{n} v_i |i\rangle \right) = \sum_{i=1}^{n} v_i \Omega |i\rangle = \sum_{i=1}^{n} v_i \sum_{j=1}^{n} \Omega_{ji} |j\rangle = \sum_{j=1}^{n} \left( \sum_{i=1}^{n} \Omega_{ji} v_i \right) |j\rangle,
\] (2.94)

implies

\[
v'_j = \sum_{i=1}^{n} \Omega_{ji} v_i.
\] (2.95)

Consider the matrix \(\Omega_{ji} = [\langle j| \Omega |i\rangle]\), where \(\Omega_{ji}\) is the \(j\)-th row and the \(i\)-th column entry. Then, (2.95) implies

\[
\begin{bmatrix}
v'_1 \\
\vdots \\
v'_n
\end{bmatrix} =
\begin{bmatrix}
\langle 1| \Omega |1\rangle & \cdots & \langle 1| \Omega |n\rangle \\
\vdots & \ddots & \vdots \\
\langle n| \Omega |1\rangle & \cdots & \langle n| \Omega |n\rangle
\end{bmatrix}
\begin{bmatrix}
v_1 \\
\vdots \\
v_n
\end{bmatrix}
\] (2.96)

or

\[
\begin{bmatrix}
v'_1 \\
\vdots \\
v'_n
\end{bmatrix} =
\begin{bmatrix}
\Omega_{11} & \cdots & \Omega_{1n} \\
\Omega_{21} & \cdots & \Omega_{2n} \\
\vdots & \ddots & \vdots \\
\Omega_{n1} & \cdots & \Omega_{nn}
\end{bmatrix}
\begin{bmatrix}
v_1 \\
\vdots \\
v_n
\end{bmatrix}.
\] (2.97)

\(^{17}\)See the footnote on p.28 for more detail.
Because
\[
\begin{bmatrix}
\Omega_{11} & \cdots & \cdots & \cdots & \Omega_{1k} & \cdots & \cdots & \cdots & \Omega_{1n} \\
\vdots & \ddots & \ddots & \ddots & \vdots & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \vdots & \ddots & \ddots & \ddots & \vdots \\
\Omega_{k1} & \cdots & \cdots & \cdots & \Omega_{kk} & \cdots & \cdots & \cdots & \Omega_{kn} \\
\vdots & \ddots & \ddots & \ddots & \vdots & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \vdots & \ddots & \ddots & \ddots & \vdots \\
\Omega_{n1} & \cdots & \cdots & \cdots & \Omega_{nk} & \cdots & \cdots & \cdots & \Omega_{nn}
\end{bmatrix}
\begin{bmatrix}
0 \\
\vdots \\
\vdots \\
1_k \\
\vdots \\
\vdots \\
0
\end{bmatrix}
= 
\begin{bmatrix}
\Omega_{1k} \\
\vdots \\
\vdots \\
\Omega_{kk} \\
\vdots \\
\vdots \\
\Omega_{nk}
\end{bmatrix},
\tag{2.98}
\] the \(k\)-th column of \([\Omega_{ij}]\) is nothing but \(\Omega |k\rangle\).

**Example 2.3 (An Operator on \(C^2\))** The column vectors \[
\begin{bmatrix}
1 \\
0
\end{bmatrix}
\quad \text{and} \quad 
\begin{bmatrix}
0 \\
1
\end{bmatrix}
\] form an orthonormal basis for \(C^2\). If \(\Omega \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}\) and \(\Omega \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}\), we should have
\[
\Omega = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix},
\tag{2.99}
\]
because \(\Omega \begin{bmatrix} 1 \\ 0 \end{bmatrix}\) is the first column of the matrix representation of \(\Omega\), and \(\Omega \begin{bmatrix} 0 \\ 1 \end{bmatrix}\) is the second column. Now consider an arbitrary vector \(v\) given by
\[
|v\rangle = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = v_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix}.
\tag{2.100}
\]

Using the matrix representation
\[
\Omega |v\rangle = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} v_1 + v_2 \\ v_1 \end{bmatrix},
\tag{2.101}
\]
On the other hand,
\[
\Omega |v\rangle = \Omega \left( v_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) = v_1 \Omega \begin{bmatrix} 1 \\ 0 \end{bmatrix} + v_2 \Omega \begin{bmatrix} 0 \\ 1 \end{bmatrix}
= v_1 \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}
= v_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + v_2 \begin{bmatrix} 1 \\ 0 \end{bmatrix}
= \begin{bmatrix} v_1 + v_2 \\ v_1 \end{bmatrix}.
\tag{2.102}
\]
This checks.
Recall that we denoted \( \alpha |v\rangle \) by \( |\alpha v\rangle \), and so, it is natural to denote \( \Omega |v\rangle \) by \( |\Omega v\rangle \). Then, it is also natural to consider its adjoint \( \langle \Omega \Omega |v\rangle \). For a scalar \( \alpha \), \( \langle \alpha v\rangle = \langle v\rangle \alpha^* \) while \( |\alpha v\rangle = \alpha |v\rangle \). Therefore, it is not surprising if \( \langle \Omega v\rangle \neq \langle v| [\Omega_{ij}] \) though \( \langle v| [\Omega_{ij}] \) makes sense as a matrix multiplication. We will next compute the matrix elements of \( \Omega \) when it comes out of the \( \langle \Omega \rangle \) and operate on \( \langle v\rangle \) from the right. Looking ahead, let us denote this by \( \Omega^\dagger \) such that \( \langle \Omega \rangle \langle v\rangle = \langle v| \Omega^\dagger \). We have

\[
|v\rangle = \sum_{i=1}^{n} v_i |i\rangle \quad \text{and} \quad |v'\rangle = \Omega |v\rangle = |\Omega v\rangle = \sum_{j=1}^{n} v_j' |j\rangle.
\]

(2.103)

So,

\[
\langle v| = \sum_{i=1}^{n} \langle i| v_i^* \quad \text{and} \quad \langle v'| = \text{adj} (\Omega |v\rangle) = \langle \Omega v| = \sum_{j=1}^{n} \langle j| v_j'^*\),
\]

(2.104)

where \( \text{adj} \) means the adjoint. On the other hand, as

\[
\Omega |i\rangle = \sum_{j=1}^{n} \Omega_{ji} |j\rangle,
\]

\[
\langle v'| = \text{adj} (\Omega |v\rangle) = \text{adj} \left( \Omega \sum_{i=1}^{n} v_i |i\rangle \right) = \text{adj} \left( \sum_{i=1}^{n} v_i \Omega |i\rangle \right) = \text{adj} \left( \sum_{j=1}^{n} \sum_{i=1}^{n} v_i \Omega_{ji} |j\rangle \right) = \sum_{j=1}^{n} \sum_{i=1}^{n} \langle j| v_i^* \Omega_{ji}^*\rangle.
\]

(2.105)

Hence,

\[
v_j'^* = \sum_{i=1}^{n} v_i^* \Omega_{ji}^* \quad \text{for each} \quad 1 \leq j \leq n.
\]

(2.106)

This translates to the following matrix relation.

\[
\begin{bmatrix}
  v_1'^* & \ldots & v_n'^*
\end{bmatrix} =

\begin{bmatrix}
  \Omega_{11}^* & \Omega_{12}^* & \ldots & \Omega_{1n}^*
  \Omega_{21}^* & \Omega_{22}^* & \ldots & \Omega_{2n}^*
  \vdots & \vdots & \ddots & \vdots
  \Omega_{n1}^* & \Omega_{n2}^* & \ldots & \Omega_{nn}^*
\end{bmatrix}
\begin{bmatrix}
  v_1^* & \ldots & v_n^*
\end{bmatrix}
\]

(2.107)

or

\[
\begin{bmatrix}
  v_1'^* & \ldots & v_n'^*
\end{bmatrix} =

\begin{bmatrix}
  \langle 1| \Omega |1\rangle^* & \langle 2| \Omega |1\rangle^* & \ldots & \langle n| \Omega |1\rangle^*
  \langle 1| \Omega |2\rangle^* & \langle 2| \Omega |2\rangle^* & \ldots & \langle n| \Omega |2\rangle^*
  \vdots & \vdots & \ddots & \vdots
  \langle 1| \Omega |n\rangle^* & \langle 2| \Omega |n\rangle^* & \ldots & \langle n| \Omega |n\rangle^*
\end{bmatrix}
\begin{bmatrix}
  v_1^* & \ldots & v_n^*
\end{bmatrix}
\]

(2.108)
2.7. Matrix Representation of Linear Operators

Because

\[
\begin{bmatrix}
0 & \ldots & \ldots & 1 & \ldots & 0
\end{bmatrix}
\begin{bmatrix}
\Omega_{11}^* & \ldots & \cdots & \Omega_{k1}^* & \ldots & \Omega_{n1}^*
\vdots & \ddots & & \vdots & & \vdots
\vdots & & \ddots & \vdots & & \vdots
\Omega_{1k}^* & \cdots & \cdots & \Omega_{kk}^* & \cdots & \Omega_{nk}^*
\vdots & & & \ddots & & \vdots
\Omega_{1n}^* & \cdots & \cdots & \Omega_{kn}^* & \cdots & \Omega_{nn}^*
\end{bmatrix}
\]

\[=
\begin{bmatrix}
\Omega_{1k}^* & \cdots & \cdots & \Omega_{kk}^* & \cdots & \Omega_{nk}^*
\end{bmatrix},
\]

(2.109)

the \(k\)-th row of \(\Omega_{ji}^*\) is nothing but \(\langle k\Omega\rangle\).

In summary, if the matrix representation of \(\Omega\) in \(|\Omega v\rangle = \Omega |v\rangle\) is

\[
\Omega = [\Omega_{ij}]
\]

\[
\begin{bmatrix}
\Omega_{11} & \ldots & \Omega_{1n}
\Omega_{21} & \ldots & \Omega_{2n}
\vdots & \ddots & \vdots
\Omega_{n1} & \ldots & \Omega_{nn}
\end{bmatrix},
\]

(2.110)

then, the matrix representation of \(\Omega^\dagger\) such that \(\langle \Omega v\rangle = \langle v| \Omega^\dagger\) is

\[
\Omega^\dagger = [\Omega_{ji}^*]
\]

\[
\begin{bmatrix}
\Omega_{11}^* & \ldots & \Omega_{n1}^*
\Omega_{12}^* & \ldots & \Omega_{n2}^*
\vdots & \ddots & \vdots
\Omega_{1n}^* & \cdots & \Omega_{nn}^*
\end{bmatrix},
\]

(2.111)

That is, \(\Omega^\dagger\) is the transpose conjugate of \(\Omega\).

2.7.1 Matrix Representations of Operator Products

If we have matrix representations \([\Omega_{ij}]\) and \([\Gamma_{kl}]\) for two operators \(\Omega\) and \(\Gamma\), it would be convenient if the matrix representation for the operator product \(\Omega \Gamma\) is the product of the corresponding matrices. Luckily for us, this is indeed the case as shown below.

First, recall that \(\sum_{k=1}^n |k\rangle \langle k| = I\). So,

\[
(\Omega \Gamma)_{ij} = \langle i| \Omega \Gamma |j\rangle = \sum_{k=1}^n \langle i| \Omega |k\rangle \langle k| \Gamma |j\rangle = \sum_{k=1}^n \Omega_{ik} \Gamma_{kj}.
\]

(2.112)
Hence, the matrix representation of a product of two operators is the product of the matrix representations for the operators in the same order.

**Notation:** We often denote $[\Omega_{ij}]$ by $[\Omega]$ for simplicity.

### 2.8 The Adjoint of an Operator

In Section 2.7, we have already encountered this, but let us formally define the adjoint of an operator and investigate its properties.

**Definition 2.24 (The Adjoint)** Consider an operator $\Omega$ that operates in a vector space $\mathcal{V}$.

\[
\Omega |v\rangle = |\Omega v\rangle
\]  
(2.113)

If there exists an operator $\Omega^\dagger$ such that

\[
\langle \Omega v | = \langle v | \Omega^\dagger.
\]  
(2.114)

Then, $\Omega^\dagger$ is called the adjoint of $\Omega$.

We can show that the adjoint indeed exists by explicitly computing the matrix elements in a basis $\{|1\rangle, |2\rangle, \ldots, |n\rangle\}$.

\[
\left(\Omega^\dagger \right)_{ij} = \langle i | \Omega^\dagger | j \rangle = \langle \Omega i | j \rangle = \langle j | \Omega i \rangle^* = \langle j | \Omega | i \rangle^* = (\Omega_{ji})^*
\]  
(2.115)

As shown in Section 2.7, the matrix representing $\Omega^\dagger$ is the transpose conjugate of the matrix for $\Omega$.

**Example 2.4 (Adjoint of an Operator)** Consider

\[
\Gamma = \begin{bmatrix}
  i & 1 \\
  0 & 1
\end{bmatrix}.
\]

Then,

\[
\Gamma^\dagger = \begin{bmatrix}
  -i & 0 \\
  1 & 1
\end{bmatrix}.
\]

Now, for $v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$,

\[
|\Gamma v\rangle = \begin{bmatrix}
  i & 1 \\
  0 & 1
\end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} iv_1 + v_2 \\ v_2 \end{bmatrix} \Rightarrow \langle \Gamma v | = \begin{bmatrix} (iv_1 + v_2)^* \\ v_2^* \end{bmatrix}
\]
\[ = \left[ -iv_1^* + v_2^* v_2^* \right]. \tag{2.116} \]

On the other hand,

\[ \langle \psi | \Gamma^\dagger = \begin{bmatrix} v_1^* & v_2^* \end{bmatrix} \begin{bmatrix} -i & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} -iv_1^* + v_2^* v_2^* \end{bmatrix} \tag{2.117} \]

This checks.

**Fact 2.2** The adjoint of the product of two operators \((\Omega \Gamma)^\dagger\) is the product of their adjoints with the order switched; that is,

\[ (\Omega \Gamma)^\dagger = \Gamma^\dagger \Omega^\dagger. \tag{2.118} \]

**Proof**

First, regard \(\Omega \Gamma\) as one operator. Then,

\[ \langle \Omega \Gamma v | = \langle (\Omega \Gamma)v | = \langle v | (\Omega \Gamma)^\dagger. \tag{2.119} \]

Next, regard \(\Omega \Gamma v\) as \(\Omega\) acting on the vector \((\Gamma v)\) to get

\[ \langle \Omega \Gamma v | = \langle \Omega (\Gamma v) | = \langle \Gamma v | \Omega^\dagger. \tag{2.120} \]

But,

\[ \langle \Gamma v | = \langle v | \Gamma^\dagger \]

means

\[ \langle \Omega \Gamma v | = \langle v | \Gamma^\dagger \Omega^\dagger. \tag{2.121} \]

Therefore,

\[ (\Omega \Gamma)^\dagger = \Gamma^\dagger \Omega^\dagger. \tag{2.122} \]

**Note 2.1 (Taking the Adjoint)** In order to take the adjoint:

1. Reverse the order of each element of each term.
2. Take the complex conjugate of the scalars.
3. Switch bras to kets and kets to bras.
4. Take the adjoint of each operator.

For example, consider a linear expression
\[ \alpha |v\rangle = \beta |u\rangle + \gamma \langle w |q |r\rangle + \delta \Omega |s\rangle + \epsilon \Gamma \Lambda |t\rangle. \]

After Step 1
\[ |v\rangle \alpha = |u\rangle \beta + |r\rangle \langle w |q |\gamma + |s\rangle \Omega \delta + |t\rangle \Lambda \epsilon \]  \hspace{1cm} (2.123)

After Step 2
\[ |v\rangle \alpha^* = |u\rangle \beta^* + |r\rangle \langle w |q \gamma^* + |s\rangle \Omega^* \delta^* + |t\rangle \Lambda^* \epsilon^* \]
\[ = |v\rangle \alpha^* = |u\rangle \beta^* + |r\rangle \langle q |w \gamma^* + |s\rangle \Omega^* \delta^* + |t\rangle \Lambda^* \epsilon^* \] \hspace{1cm} (2.124)

After Step 3
\[ \langle v| \alpha^* = \langle u| \beta^* + \langle r| \langle q |w \gamma^* + \langle s| \Omega^* \delta^* + \langle t| \Lambda^* \epsilon^* \] \hspace{1cm} (2.125)

After Step 4
\[ \langle v| \alpha^* = \langle u| \beta^* + \langle r| \langle q |w \gamma^* + \langle s| \Omega^\dagger \delta^* + \langle t| \Lambda^\dagger \epsilon^* \] \hspace{1cm} (2.126)

However, it will be easier to take the adjoint of each term in one step and add up the terms once you get used to this operation.

2.9 Eigenvalues and Eigenvectors

Definition 2.25 (Eigenvalues, Eigenvectors, and Eigenkets)
Consider an operator \( \Omega \) operating on a vector space \( \mathbb{V} \). When a nonzero vector \( v \) or a nonzero ket \( |v\rangle \) satisfies
\[ \Omega v = \lambda v \quad \text{or} \quad \Omega |v\rangle = \lambda |v\rangle, \] \hspace{1cm} (2.127)
\( v \) is called an eigenvector, \( |v\rangle \) is called an eigenket, and \( \lambda \) is called an eigenvalue. Note that if \( |v\rangle \) (\( v \)) is an eigenket (eigenvector) with the associated eigenvalue \( \lambda \), any scalar multiple of the ket (vector) \( \alpha |v\rangle \) (\( \alpha v \)), where \( \alpha \neq 0 \), is an eigenket (eigenvector) with the same eigenvalue \( \lambda \).
Example 2.5 (Diagonal Matrices) Diagonal matrices have eigenvalues \( \{\lambda_i\} \) along the main diagonal with associated eigenvectors given by

\[
\begin{bmatrix}
0 \\
\vdots \\
\vdots \\
c \\
\vdots \\
0
\end{bmatrix}
\]

the \( i \)-th entry, \( i = 1, 2, \ldots, n \) (2.128)

where \( c \) is an arbitrary complex number. Here is a 3\( \times \)3 example. Consider

\[
\begin{bmatrix}
\lambda_1 & 0 & 0 \\
0 & \lambda_2 & 0 \\
0 & 0 & \lambda_3
\end{bmatrix}
\]

Then, we have

\[
\begin{bmatrix}
\lambda_1 & 0 & 0 \\
0 & \lambda_2 & 0 \\
0 & 0 & \lambda_3
\end{bmatrix}
\begin{bmatrix}
c \\
0 \\
0
\end{bmatrix} = \lambda_1
\begin{bmatrix}
c \\
0 \\
0
\end{bmatrix},
\]

(2.130)

\[
\begin{bmatrix}
\lambda_1 & 0 & 0 \\
0 & \lambda_2 & 0 \\
0 & 0 & \lambda_3
\end{bmatrix}
\begin{bmatrix}
c \\
0 \\
0
\end{bmatrix} = \lambda_2
\begin{bmatrix}
c \\
0 \\
0
\end{bmatrix},
\]

(2.131)

and

\[
\begin{bmatrix}
\lambda_1 & 0 & 0 \\
0 & \lambda_2 & 0 \\
0 & 0 & \lambda_3
\end{bmatrix}
\begin{bmatrix}
c \\
0 \\
0
\end{bmatrix} = \lambda_3
\begin{bmatrix}
c \\
0 \\
0
\end{bmatrix}.
\]

(2.132)

In some cases, some of the eigenvalues are the same. For example, the following 3\( \times \)3 matrix has \( \lambda_1 = \lambda_2 = 1 \) and \( \lambda_3 = 2 \).

\[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 2
\end{bmatrix}
\]

(2.133)

When this happens, we say that the multiplicity of the eigenvalue 1 is two.
Example 2.6 (Eigenvalues and Eigenvectors: A two-by-two example) Consider
\[ \Omega = \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix}. \]
Then,
\[ \Omega \begin{bmatrix} 0.6 \\ 0.4 \end{bmatrix} = \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix} \begin{bmatrix} 0.6 \\ 0.4 \end{bmatrix} = 1 \cdot \begin{bmatrix} 0.6 \\ 0.4 \end{bmatrix} \]
and
\[ \Omega \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 \\ -1 \end{bmatrix}. \] (2.134)
Hence, the eigenvalues of \( \Omega \) are 1 and \( \frac{1}{2} \) with corresponding eigenvectors which are scalar multiples of the vectors given above.

There is a standard way to find the eigenvalues of an operator \( \Omega \). Here is a sketch.

Suppose \( \lambda \) is an eigenvalue for an operator \( \Omega \) with a corresponding eigenket \( |v\rangle \).
Then,
\[ \Omega |v\rangle = \lambda |v\rangle. \] (2.136)
This means that
\[ (\Omega - \lambda I) |v\rangle = |0\rangle, \] (2.137)
which in turn implies that \( \Omega - \lambda I \) is not invertible. In order to see this, suppose \( \Omega - \lambda I \) has the inverse \( (\Omega - \lambda I)^{-1} \). Then,
\[ (\Omega - \lambda I)^{-1}(\Omega - \lambda I) |v\rangle = (\Omega - \lambda I)^{-1} |0\rangle \Rightarrow |v\rangle = |0\rangle. \] (2.138)
But, this is a contradiction because \( |v\rangle \neq |0\rangle \) if \( |v\rangle \) is an eigenvector. Now consider a matrix representation of the operator \( \Omega \) in some basis. Then, from basic matrix theory, \( \Omega - \lambda I \) is invertible if and only if \( \det(\Omega - \lambda I) \neq 0 \). If \( \Omega \) is an \( n \times n \) matrix, \( \det(\Omega - \lambda I) \), called a characteristic polynomial, is an \( n \)-th degree polynomial with \( n \) roots\(^\text{a}\). We can find all the eigenvalues by solving the characteristic equation \( \det(\Omega - \lambda I) = 0 \). In order to compute \( \det(\Omega - \lambda I) \) we need to pick a particular basis. However, the resulting eigenvalues are basis-independent because the characteristic polynomial is basis-independent. For a proof, see Theorem A.4 in Appendix A.

\(^{\text{a}}\)As described at the end of Example 2.5 on p.51, we do not always have \( n \) distinct roots. The
total number of the roots is \( n \) if multiplicity is taken into account. For example, if the polynomial has \((x - 1)^2\) when factored, 1 is a root with a multiplicity of 2, and it counts as two roots.

**Notation:** An eigenket associated with the eigenvalue \( \lambda \) is denoted by \( |\lambda\rangle \).

## 2.10 Special Types of Operators

### 2.10.1 Hermitian Operators

**Definition 2.26 (Hermitian Operators)**  A Hermitian operator is an operator which is the adjoint of itself; that is, an operator \( \Omega \) is Hermitian if and only if

\[
\Omega = \Omega^\dagger. \tag{2.139}
\]

*In particular, this means that*

\[
\langle \Omega v | w \rangle = \langle v | \Omega w \rangle
\]

*as both sides are equal to*

\[
\langle v | \Omega | w \rangle.
\]

In the linear algebra of real matrices, the matrix representation of a Hermitian operator is nothing but a symmetric matrix. But, our matrices are complex, and the matrix representing a Hermitian operator should equal its transpose conjugate.

**Definition 2.27 (Anti-Hermitian Operators)**  An operator \( \Omega \) is anti-Hermitian if

\[
\Omega^\dagger = -\Omega. \tag{2.140}
\]

It is possible to decompose every operator \( \Omega \) into its Hermitian and anti-Hermitian components; namely if we decompose \( \Omega \) as

\[
\Omega = \frac{\Omega + \Omega^\dagger}{2} + \frac{\Omega - \Omega^\dagger}{2}, \tag{2.140}
\]

These are also called self-adjoint operators. You may hear that self-adjoint is the term used by mathematicians and Hermitian is used by physicists. However, there is a subtle difference between ‘self-adjoint’ and “Hermitian”. We will not want to get bogged down worrying too much about such details, but interested readers should check a book titled “Functional Analysis” by Michael Reed and Barry Simon [Reed and Simon, 1980, p.255].
then, the first term $\Omega + \Omega^\dagger$ is Hermitian, and the second term $\Omega - \Omega^\dagger$ is anti-Hermitian.

These are also known as anti-self-adjoint operators.

Hermitian operators play a central role in quantum mechanics due to its special characteristics.

**Fact 2.3 (Eigenvalues and Eigenkets of a Hermitian Operator)**

1) **Real Eigenvalues** The eigenvalues of a Hermitian operator are real. Measurable values in physics such as position, momentum, and energy are real numbers, and they are obtained as eigenvalues of Hermitian operators in quantum mechanics.

2) **Orthonormal Eigenkets** The eigenkets of a Hermitian operator can always be chosen so that they are mutually orthogonal, and hence linearly independent. Because an eigenket can be multiplied by any nonzero number to generate another eigenket of any desired “length”, we can always choose the eigenkets which form an orthonormal set.

3) **Basis of Eigenkets** The eigenkets of a Hermitian operator is a complete set in the sense that any vector can be expressed as a linear combination of the eigenkets. Therefore, if you know the action of a Hermitian operator $\Omega$ on the eigenkets, you completely understand the action of $\Omega$ on any vector in the space.

**Verification**: Let $v = \sum \alpha_i |\lambda_i\rangle$ where $\lambda_i$’s are eigenvalues and $|\lambda_i\rangle$’s are the corresponding eigenkets. Then, $\Omega v = \Omega (\sum \alpha_i |\lambda_i\rangle) = \sum \alpha_i (\Omega |\lambda_i\rangle)$.

**Example 2.7 (A 2-by-2 Hermitian Operator/Matrix)** Let

$$\Omega = \begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}.$$

As the adjoint is the transpose conjugate, we have

$$\Omega^\dagger = \begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix} = \Omega,$$

(2.141)

---

19 Theorem 2.7 claims that eigenvectors associated with different eigenvalues of a Hermitian operator are mutually orthogonal. More generally, according to Fact A.2, eigenvectors associated with simple eigenvalues, i.e. eigenvalues of multiplicity one, are linearly independent. Therefore, the significance of this statement is that Hermiticity assures that we can always find $l$ mutually orthogonal eigenvectors if the associated eigenvalue is of multiplicity $l$. 
and \( \Omega \) is Hermitian. Observe that
\[
\Omega \begin{bmatrix} -i \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix} \begin{bmatrix} -i \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0 \cdot \begin{bmatrix} -i \\ 1 \end{bmatrix} \quad (2.142)
\]
and
\[
\Omega \begin{bmatrix} i \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix} \begin{bmatrix} i \\ 1 \end{bmatrix} = \begin{bmatrix} 2i \\ 2 \end{bmatrix} = 2 \cdot \begin{bmatrix} i \\ 1 \end{bmatrix} \quad (2.143)
\]
So, the eigenvalues are indeed real; namely, 0 and 2.

We have
\[
|0\rangle = \begin{bmatrix} -i \\ 1 \end{bmatrix} \quad \text{and} \quad |2\rangle = \begin{bmatrix} i \\ 1 \end{bmatrix}. \quad (2.144)
\]

Because
\[
\langle 0|2 \rangle = \begin{bmatrix} -i \\ 1 \end{bmatrix}^\dagger \begin{bmatrix} i \\ 1 \end{bmatrix} = \begin{bmatrix} i & 1 \end{bmatrix} \begin{bmatrix} i \\ 1 \end{bmatrix} = 0, \quad (2.145)
\]

\(|0\rangle\) and \(|2\rangle\) are mutually orthogonal.

Next note that
\[
\frac{|0\rangle + |2\rangle}{\sqrt{2}} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (2.146)
\]
and
\[
\frac{-|0\rangle + |2\rangle}{2i} = \frac{1}{2i} \begin{bmatrix} 2i \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \quad (2.147)
\]

Therefore, any ket vector \( \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \) can be expressed as a linear combination of \(|0\rangle\) and \(|2\rangle\) as follows.
\[
\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = v_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = v_1 \left( \frac{|0\rangle + |2\rangle}{\sqrt{2}} \right) + v_2 \left( \frac{-|0\rangle + |2\rangle}{2i} \right)
\]
\[
= \left( \frac{v_1}{\sqrt{2}} - \frac{v_2}{2i} \right) |0\rangle + \left( \frac{v_1}{2} + \frac{v_2}{2i} \right) |2\rangle \quad (2.148)
\]

So, any vector \( \mathbf{v} \) is a linear combination of \(|0\rangle\) and \(|2\rangle\). If we normalize \(|0\rangle\) and \(|2\rangle\), we obtain an orthonormal basis for \( \mathbb{C}^2 \). Namely,
\[
\frac{1}{\sqrt{2}} \begin{bmatrix} -i \\ 1 \end{bmatrix} \quad \text{and} \quad \frac{1}{\sqrt{2}} \begin{bmatrix} i \\ 1 \end{bmatrix} \quad (2.149)
\]
form an orthonormal basis.
Example 2.8 (3-by-3 Hermitian and Non-Hermitian Operators) This example is to demonstrate the claim made by 2) and 3) of Fact 2.3. Consider the following $3 \times 3$ Hermitian matrix $\Omega$.

$$\Omega = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \quad (2.150)$$

According to Example 2.5 on p.51 and Fact A.2 on p.276, the eigenvalue 1 is of multiplicity 2, and the eigenvalue 2 is simple, i.e. of multiplicity 1.

For the eigenvalue 2, we have

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = 2 \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} \implies \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 2a \\ 2b \\ 2c \end{bmatrix} \implies a = b = 0. \quad (2.151)$$

Hence, any nonzero vector of the form

$$\begin{bmatrix} 0 \\ 0 \\ c \end{bmatrix} \quad (2.152)$$

is an eigenvector associated with the eigenvalue 2. In particular,

$$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad (2.153)$$

is an example of a normalized eigenvector for the eigenvalue 2.

Now, for the eigenvalue 1, we have

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = 1 \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} \implies \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} a \\ b \\ 2c \end{bmatrix} \implies c = 0. \quad (2.154)$$

So, any nonzero vector of the form

$$\begin{bmatrix} a \\ b \\ 0 \end{bmatrix} \quad (2.155)$$
is an eigenvector associated with the eigenvalue 1. In particular,
\[
\begin{bmatrix}
1 \\
0 \\
0
\end{bmatrix}
\quad \text{and} \quad
\begin{bmatrix}
0 \\
1 \\
0
\end{bmatrix}
\] (2.156)
are an example of mutually orthogonal normalized eigenvectors for the eigenvalue 1. At this point, it is trivial to show that the set
\[
\begin{Bmatrix}
\begin{bmatrix}
0 \\
0 \\
1
\end{bmatrix},
\begin{bmatrix}
1 \\
0 \\
0
\end{bmatrix},
\begin{bmatrix}
0 \\
1 \\
0
\end{bmatrix}
\end{Bmatrix}
\] (2.157)
forms an orthonormal basis of $\mathbb{C}^3$. However, no claim is made as to uniqueness of such a set. For example,
\[
\begin{Bmatrix}
\frac{1}{\sqrt{10}} \begin{bmatrix}
0 \\
0 \\
1 + 3i
\end{bmatrix},
\frac{1}{\sqrt{2}} \begin{bmatrix}
1 \\
i \\
0
\end{bmatrix},
\frac{1}{\sqrt{2}} \begin{bmatrix}
1 \\
-i \\
0
\end{bmatrix}
\end{Bmatrix}
\] (2.158)
is another possible choice.

Next, consider the following $3 \times 3$ non-Hermitian matrix $\Lambda$.
\[
\Lambda =
\begin{bmatrix}
1 & 0 & 0 \\
1 & 1 & 0 \\
0 & 0 & 2
\end{bmatrix}
\] (2.159)
The eigenvalues are 1 of multiplicity 2 and 2 of multiplicity 1 as was the case with $\Omega$.\(^{20}\) For the eigenvalue 2, we have
\[
\Lambda \begin{bmatrix}
a \\
b \\
c
\end{bmatrix} = 2 \cdot \begin{bmatrix}
a \\
b \\
c
\end{bmatrix} \implies
\begin{bmatrix}
1 & 0 & 0 \\
1 & 1 & 0 \\
0 & 0 & 2
\end{bmatrix} \begin{bmatrix}
a \\
a + b \\
2c
\end{bmatrix} =
\begin{bmatrix}
a \\
2b \\
2c
\end{bmatrix} \implies a = b = 0. \quad (2.160)
\] Hence, any nonzero vector of the form
\[
\begin{bmatrix}
0 \\
0 \\
c
\end{bmatrix}
\] (2.161)
\(^{20}\)The characteristic equation mentioned on p.52 is $(1 - \lambda)^2(2 - \lambda) = 0.$
is an eigenvector associated with the eigenvalue 2.

Next, for the eigenvalue 1,

$$\Lambda \begin{bmatrix} a \\ b \\ c \end{bmatrix} = 1 \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} a + b \\ b \\ 2c \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \implies a = c = 0.$$  

(2.162)

So, any nonzero vector of the form

$$\begin{bmatrix} 0 \\ b \\ 0 \end{bmatrix}$$  

(2.163)

is an eigenvector associated with the eigenvalue 1. Hence, the set of eigenvectors for $\Lambda$ can be expressed as follows for arbitrary but nonzero complex numbers $b$ and $c$.

$$\begin{cases} \begin{bmatrix} 0 \\ b \\ 0 \end{bmatrix}, \\ \begin{bmatrix} 0 \\ b \\ c \end{bmatrix} \end{cases}$$  

(2.164)

Needless to say, we do not have a basis of $\mathbb{C}^3$.

Let us prove Fact 2.3 starting with 1).

**Theorem 2.6 (Real Eigenvalues)** *The eigenvalues of a Hermitian operator $\Omega$ are real.*

**Proof**

Let $\lambda$ be an eigenvalue of $\Omega$ and $|v\rangle$ be an associated eigenket. Then,

$$\lambda^* \langle v | v \rangle = \langle \lambda v | v \rangle = \langle \Omega v | v \rangle = \langle v | \Omega^\dagger | v \rangle = \langle v | \lambda v \rangle = \lambda \langle v | v \rangle$$

$$\implies (\lambda^* - \lambda) \langle v | v \rangle = 0 \implies \lambda^* = \lambda$$  

(2.165)

because $\langle v | v \rangle$ is nonzero if $|v\rangle$ is nonzero.\(^{21}\)

**Theorem 2.7** *Eigenkets $|v\rangle$ and $|w\rangle$ of a Hermitian operator $\Omega$ associated with different eigenvalues $\lambda_v$ and $\lambda_w$ are orthogonal.*

\(^{21}\)Recall that an eigenket/eigenvector is nonzero by Definition 2.25.
2.10. SPECIAL TYPES OF OPERATORS

Proof

\[ \langle \omega | \Omega | v \rangle = \langle w | \lambda_v | v \rangle = \lambda_w \langle w | v \rangle \]  
\hspace{1cm} (2.166)

On the other hand, as \( \Omega \) is Hermitian, and \( \lambda_v \) and \( \lambda_w \) are real,

\[ \langle w | \Omega | v \rangle = \langle w | \Omega^\dagger | v \rangle = \langle \Omega w | v \rangle = \langle v | \Omega w \rangle^* = \langle v | \lambda_w | w \rangle^* = \lambda_w \langle w | v \rangle . \]  
\hspace{1cm} (2.167)

So,

\[ \lambda_v \langle w | v \rangle - \lambda_w \langle w | v \rangle = (\lambda_v - \lambda_w) \langle w | v \rangle = 0 \Rightarrow \langle w | v \rangle = 0 \]  
\hspace{1cm} (2.168)

as \( \lambda_v \neq \lambda_w \).

We will now prove 2) and 3) of Fact 2.3 and some more.

**Lemma 2.1** Let \( \Omega \) be a linear operator on a vector space \( V^n \) and \( [\Omega_{ij}]_B \) be its matrix representation with respect to a basis \( B = \{u_1, u_2, \ldots, u_k \ldots, u_n\} \). If the \( k \)-th vector in the basis \( B \), denoted by \( u_k \) is an eigenvector with the corresponding eigenvalue \( \lambda_k \), such that \( \Omega u_k = \lambda_k u_k \), the \( k \)-th column of \( [\Omega_{ij}]_B \) is given by \( \Omega_{jk} = \lambda_k \delta_{jk} \) for \( 1 \leq j \leq n \). That is, the entries of the \( k \)-th column are zero except for the \( k \)-th entry, which is \( \lambda_k \).

the \( k \)-th column of \( [\Omega_{ij}]_B \) is

\[ \begin{bmatrix} 0 \\
\vdots \\
\vdots \\
\lambda_k \\
\vdots \\
0 \end{bmatrix}_k \]  
\hspace{1cm} (2.169)

Proof
From p. 45, we know that the $k$-th column of $[\Omega_{ij}]$ is nothing but $\Omega |u_k\rangle$. But,

$$\Omega |u_k\rangle = \lambda_k u_k = \lambda_k \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ \lambda_k \\ \vdots \\ 0 \end{bmatrix}.$$  \hfill (2.170)

In matrix form,

$$\begin{bmatrix} \Omega_{11} & \cdots & \Omega_{1k} & \cdots & \Omega_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \Omega_{k1} & \cdots & \Omega_{kk} & \cdots & \Omega_{kn} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \Omega_{n1} & \cdots & \Omega_{nk} & \cdots & \Omega_{nn} \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \Omega_{1k} \\ \vdots \\ \Omega_{kk} \\ \vdots \\ \Omega_{nk} \end{bmatrix} = \lambda_k \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ \lambda_k \\ \vdots \\ 0 \end{bmatrix}.$$

$$\Leftrightarrow \Omega_{jk} = \lambda_k \delta_{jk}. \hfill (2.171)$$

**Corollary 2.1** Let $\Omega$ be a Hermitian operator on a vector space $\mathbb{V}^n$ and $[\Omega_{ij}]_B$ be its matrix representation with respect to a basis $B = \{u_1, u_2, \ldots, u_k, \ldots, u_n\}$. If the $k$-th vector in the basis $B$, denoted by $u_k$ is an eigenvector with the corresponding eigenvalue $\lambda_k$, such that $\Omega u_k = \lambda_k u_k$, the $k$-th column of $[\Omega_{ij}]_B$ is given by $\Omega_{jk} = \lambda_k \delta_{jk}$ for $1 \leq j \leq n$. That is, the entries of the $k$-th column are zero except for the $k$-th entry, which is $\lambda_k$. Furthermore, the $k$-th row of $[\Omega_{ij}]_B$ is given by $\Omega_{kj} = \lambda_k \delta_{kj}$ for $1 \leq j \leq n$.

**Proof**

$\Omega_{jk} = \lambda_k \delta_{jk}$ follows from Lemma 2.1. $\Omega_{kj} = \lambda_k \delta_{kj}$ follows from the definition of a Hermitian operator and Theorem 2.6. The matrix representation of $\Omega^\dagger$, denoted here by $[\Omega_{ij}]_B$, is the transpose conjugate of the matrix for $\Omega$ denoted by $[\Omega_{ij}]_B$. As
the $k$-th column of $[\Omega_{ij}]_{B}$ is
\[
\begin{bmatrix}
    0 \\
    \vdots \\
    \vdots \\
    \lambda_k \\
    \vdots \\
    0
\end{bmatrix},
\]
\[
\text{(2.172)}
\]

the $k$-th row of $[\Omega_{ij}^\dagger]_{B}$ is
\[
\begin{bmatrix}
    k \\
    0 & \cdots & \cdots & \lambda_k^* & \cdots & 0
\end{bmatrix}.
\]
\[
\text{(2.173)}
\]

But, $\lambda_k^* = \lambda_k$ because $\lambda_k$ is real. In matrix form, $[\Omega_{ij}]$ looks like
\[
\begin{bmatrix}
    0 \\
    \vdots \\
    \vdots \\
    0 & \cdots & \cdots & \lambda_k \\
    \vdots \\
    0
\end{bmatrix}.
\]
\[
\text{(2.174)}
\]

**Lemma 2.2** (Block Structure of Hermitian Matrices)
Consider an ordered list of all the eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ of a Hermitian operator $\Omega$, where an eigenvalue $\lambda$ is repeated $m_\lambda$ times if its multiplicity as a root of the characteristic polynomial is $m_\lambda$. Without loss of generality, we can assume that the same eigenvalues appear next to each other without a different eigenvalue in between.\(^a\) Then, the matrix representation of $\Omega$ in the basis $\{|\lambda_1\rangle, |\lambda_2\rangle, \ldots, |\lambda_n\rangle\}$, called an eigenbasis, is block-diagonal.

**Corollary 2.2** Suppose the eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ of a Hermitian operator $\Omega$ are all distinct, such that $i \neq j \implies \lambda_i \neq \lambda_j$. Then, the matrix representation of $\Omega$ in the eigenbasis $\{|\lambda_1\rangle, |\lambda_2\rangle, \ldots, |\lambda_n\rangle\}$ is diagonal.
For simplicity, suppose we have a list of four eigenvalues consisting of two $\lambda$'s, one $\gamma$, and one $\omega$. Then, the ordered list is $\{\lambda, \lambda, \omega, \gamma\}$, $\{\lambda, \lambda, \gamma, \omega\}$, $\{\omega, \lambda, \lambda, \gamma\}$, $\{\gamma, \lambda, \lambda, \omega\}$ and not $\{\lambda, \omega, \lambda\}$.

**Theorem 2.8** If an operator $\Omega$ on $\mathbb{V}^n$ is Hermitian, there is a basis of its orthonormal eigenvectors. In this eigenbasis, the matrix representation of $\Omega$ is diagonal, and the diagonal entries are the eigenvalues. Note that this theorem only claims that there is at least one such basis. There is no claim about uniqueness as it is not true.

**Proof**

Consider the characteristic polynomial which is of $n$-th degree. If we set it equal to zero, the equation, called the characteristic equation, has at least one root $\lambda_1$ and a corresponding normalized eigenket $|\lambda_1\rangle$. This can be extended to a full basis $\mathcal{B}$ according to Theorem A.5. Now, following the Gram-Schmidt procedure, Theorem 2.4, we can find an orthonormal basis $O = \{o_1, o_2, o_3, \ldots, o_n\}$ such that $o_1 = |\lambda_1\rangle$.

In this basis, according to Corollary 2.1, $\Omega$ has the following matrix representation which we denote by $M$ or $[\Omega]$ in keeping with the notation set on p. 48.

$$M = [\Omega] = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \lambda_{n-1} \end{bmatrix}$$ (2.175)

If we denote the shaded submatrix by $N$, it can be regarded as the matrix representation of an operator acting on the subspace $O_2$ spanned by $\{o_2, o_3, \ldots, o_n\}$. The characteristic polynomial for $M$ is given by

$$\det(M - \lambda I_n) = (\lambda_1 - \lambda) \times \det(N - \lambda I_{n-1});$$ (2.176)

where $I_n$ is the $n \times n$ identity matrix, $I_{n-1}$ is the $(n - 1) \times (n - 1)$ identity matrix, and $\det(N - \lambda I_{n-1})$ is the characteristic polynomial of $N$. Because $\det(N - \lambda I_{n-1})$ is $P^{n-1}(\lambda)$, an $(n - 1)$st degree polynomial in $\lambda$, it must have at least one root $\lambda_2$ and a normalized eigenket $|\lambda_2\rangle$ unless $n = 1$. Because we are regarding $N$ as operating on $O_2$, $|\lambda_2\rangle$ is a linear combination of $o_2, o_3, \ldots, o_n$, and is orthogonal to $|\lambda_1\rangle$. As before, we can find an orthonormal basis $\{|\lambda_2\rangle, o'_3, o'_4, \ldots, o'_n\}$ of $O_2$. Note that $\lambda_2$ is an eigenvalue of $M$, $|\lambda_2\rangle$ is an eigenvector of $M$, and $O_{1,2} = \{|\lambda_1\rangle, |\lambda_2\rangle, o'_3, o'_4, \ldots, o'_n\}$ is an orthonormal basis of the original vector space $\mathbb{V}^n$.

Let us take a detour here and see how the relations between $M$ and $N$ work.
explicitly. It may help you understand what is going on better.

First, $\lambda_2$ is clearly an eigenvalue of $M$ as it is a root of the characteristic polynomial for $M$ because of the relation (2.176). Next, as $|\lambda_2\rangle$ is a linear combination of $o_2, o_3, \ldots$, and $o_n$, it is also a linear combination of $|\lambda_1\rangle, o_2, o_3, \ldots, o_n$. We can write

$$|\lambda_2\rangle = 0 \cdot |\lambda_1\rangle + \sum_{i=2}^{n} e_i o_i.$$  (2.177)

So, $|\lambda_2\rangle$ is the following column vector in the original basis $O = \{ |\lambda_1\rangle, o_2, o_3, \ldots, o_n \}$.

$$|\lambda_2\rangle = \begin{bmatrix}
0 \\
e_2 \\
\vdots \\
e_n
\end{bmatrix}$$  (2.178)

Hence,

$$M |\lambda_2\rangle = \begin{bmatrix}
\lambda_1 & 0 & \cdots & 0 \\
0 & e_2 \\
\vdots & \vdots \\
0 & e_n
\end{bmatrix} \begin{bmatrix}
0 \\
e_2 \\
\vdots \\
e_n
\end{bmatrix} = \begin{bmatrix}
0 \\
\lambda_2 e_2 \\
\vdots \\
\lambda_2 e_n
\end{bmatrix} = \lambda_2 |\lambda_2\rangle;$$  (2.179)

where the second equality in (2.179) follows from

$$\begin{bmatrix}
0 \\
\vdots \\
0
\end{bmatrix} \begin{bmatrix}
0 \\
e_2 \\
\vdots \\
e_n
\end{bmatrix} = \begin{bmatrix}
0 \\
N |\lambda_2\rangle
\end{bmatrix} = \begin{bmatrix}
0 \\
\lambda_2 |\lambda_2\rangle
\end{bmatrix}.$$  (2.180)

Note that $|\lambda_2\rangle$ is regarded as a vector in $\mathbb{V}^n$ in (2.179) and as a vector in the subspace spanned by $o_2, o_3, \ldots, o_n$ in (2.180). There is a trade-off between avoiding abuse of notation and simplicity, and our choice here is simplicity.
In the basis $O_{1,2}$, $[\Omega]$ takes the following form.

\[ M = [\Omega] = \begin{bmatrix}
\lambda_1 & 0 & 0 & \cdots & 0 \\
0 & \lambda_2 & 0 & \cdots & 0 \\
0 & 0 & \lambda_3 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & \lambda_n
\end{bmatrix} \] (2.181)

Repeating the same procedure, we will finally arrive at

\[ M = [\Omega] = \begin{bmatrix}
\lambda_1 & 0 & 0 & \cdots & 0 \\
0 & \lambda_2 & 0 & \cdots & 0 \\
0 & 0 & \lambda_3 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & \lambda_n
\end{bmatrix}. \] (2.182)

Note that some eigenvalues may occur more than once, but the above proof still works as it is.

**Example 2.9 (Diagonalization of a Hermitian Matrix) Consider**

\[ \Omega = \begin{bmatrix}
1 & 1 \\
1 & 1
\end{bmatrix}. \]

$\Omega$ is clearly Hermitian as it equals its own transpose conjugate. Observe that

\[ \Omega \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} = 2 \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \] (2.183)

and

\[ \Omega \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0 \cdot \begin{bmatrix} 1 \\ -1 \end{bmatrix}. \] (2.184)

Define normalized eigenvectors $|e_1\rangle$ and $|e_2\rangle$ as follows.

\[ |e_1\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \quad |e_2\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}. \] (2.185)
2.10. SPECIAL TYPES OF OPERATORS

Then,

\[
\Omega_{11} = \langle e_1|\Omega|e_1 \rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \frac{1}{2} \cdot 4 = 2. \quad (2.186)
\]

Similarly, we also get

\[
\Omega_{12} = \langle e_1|\Omega|e_2 \rangle = 0,
\]

\[
\Omega_{21} = \langle e_2|\Omega|e_1 \rangle = 0,
\]

\[
\Omega_{22} = \langle e_2|\Omega|e_2 \rangle = 0.
\]

(2.187)

Note that the eigenvalues are on the diagonal and all the off-diagonal entries are zero; i.e. we have diagonalized \( \Omega \).

**Fact 2.4** If \( \Omega \) is an anti-Hermitian operator:

1. The eigenvalues of \( \Omega \) are purely imaginary.
2. There is an orthonormal basis consisting of normalized eigenvectors of \( \Omega \).

2.10.2 Simultaneous Diagonalization

**Theorem 2.9 (Commuting Hermitian Operators)** If \( \Omega \) and \( \Gamma \) are commuting Hermitian operators, there exists a basis consisting of common eigenvectors, such that both \( \Omega \) and \( \Gamma \) are diagonal in that basis.

This very useful theorem is actually a corollary of the following more general theorems from matrix algebra. For details, check Appendix A.

**Theorem 2.10 (Commuting Diagonalizable Matrices)** Let \( A \) and \( B \) be diagonalizable \( n \times n \) matrices. Then, \( AB = BA \) if and only if \( A \) and \( B \) are simultaneously diagonalizable.

**Theorem 2.11 (A Family of Diagonalizable Matrices)** Let \( \mathcal{F} \) be a family of diagonalizable \( n \times n \) matrices. Then, it is a commuting family if and only if it is simultaneously diagonalizable.
2.11 Active and Passive Transformations

We will mainly deal with active transformations in this book. So, you only need to familiarize yourself with the general concept.

In Classical Mechanics
In classical mechanics, it is about a point in space and the coordinate system.

- An active transformation moves the point while leaving the coordinate axes unchanged/fixed. For example, a rotation of the point \((x_0, y_0)\) in the \(x, y\)-plane by angle \(\theta\) is effected by the matrix \(R_\theta\) given by

\[
R_\theta = \begin{bmatrix}
\cos \theta & -\sin \theta \\
\sin \theta & \cos \theta
\end{bmatrix}
\]  

(2.188)

such that the new position \((x'_0, y'_0)\) is obtained as follows.

\[
\begin{bmatrix} x'_0 \\ y'_0 \end{bmatrix} = R_\theta \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\
\sin \theta & \cos \theta
\end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} x_0 \cos \theta - y_0 \sin \theta \\ x_0 \sin \theta + y_0 \cos \theta
\end{bmatrix}
\]  

(2.189)

Under an active transformation, the value of a dynamical variable, often denoted by \(\omega(q, p)\), will not necessarily remain the same.

- A passive transformation does not move the point \((x_0, y_0)\) but only changes the coordinate system. To take a 2-dimensional rotation as an example, this corresponds to rotating the \(x-y\) coordinate system without moving the point itself. Because the only change is in the value of the point’s \(x\)- and \(y\)-coordinates relative to the new rotated axes, and the position itself does not change in this case, the physical environment of the point remains the same, and a dynamical variable \(\omega(q, p)\) will have the same value after the transformation.

Note that the new coordinates \((x'_0, y'_0)\) are numerically the same whether you rotate the axes by \(\theta\) or the point by \(-\theta\). Hence, the effect of passive transformation by \(\theta\) can be computed as below.

\[
\begin{bmatrix} x'_0 \\ y'_0 \end{bmatrix} = R_{-\theta} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} \cos(-\theta) & -\sin(-\theta) \\
\sin(-\theta) & \cos(-\theta)
\end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} x_0 \cos \theta + y_0 \sin \theta \\ -x_0 \sin \theta + y_0 \cos \theta
\end{bmatrix}
\]  

(2.190)
Again, despite this relation, the two transformations should not be confused with each other because the active transformation is the only physically meaningful transformation, in the sense that the values of dynamical variables may change.

**In Quantum Mechanics**

In quantum mechanics, it is about an operator \( \Omega \) and a basis vector \( |v\rangle \). Consider a unitary transformation \( U \) which causes a basis change \( |v\rangle \rightarrow U|v\rangle := |Uv\rangle \). Under this transformation, the matrix elements of an operator \( \Omega \) transforms as follows.

\[
\langle v'|\Omega|v \rangle \rightarrow \langle Uv'|\Omega|Uv \rangle = \langle v'|U\dagger\Omega U|v \rangle \tag{2.191}
\]

In (2.191), the change is caused to the vector \( v \) in \( \langle Uv'|\Omega|Uv \rangle \), and to the operator \( \Omega \) in \( \langle v'|U\dagger\Omega U|v \rangle \). Nevertheless, the matrix elements of \( \Omega \) are the same between the two formulations or interpretations.

- **An active transformation** corresponds to
  \[
  \langle Uv'|\Omega|Uv \rangle \tag{2.192}
  \]
  where the vector \( v \) is transformed as below.
  \[
  v \rightarrow Uv \tag{2.193}
  \]

- **A passive transformation** picture is painted by
  \[
  \langle v'|U\dagger\Omega U|v \rangle \tag{2.194}
  \]
  where the operator \( \Omega \) is transformed as follows.
  \[
  \Omega \rightarrow U\dagger\Omega U \tag{2.195}
  \]

Because the essence of quantum mechanics is the matrix elements of operators, and all of the physics of quantum mechanics lies in the matrix elements, the equality of (2.192) and (2.194) given in (2.191) indicates that these two views are completely equivalent. Hence, active and passive transformations in quantum mechanics provide two equivalent ways to describe the same physical transformation.
Exercises

1. Consider

\[ \Omega = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \]

The eigenvalues are 1 and -1.

(a) Explain why \( \Omega \) is Hermitian.

(b) Find a real normalized eigenvector for each eigenvalue. We will denote them by \( |1\rangle \) and \( |-1\rangle \).

(c) Verify that the eigenvectors are orthogonal.

(d) Express an arbitrary vector \( \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \) as a linear combination of \( |1\rangle \) and \( |-1\rangle \).

(e) Express \( \Omega \) as a matrix in the basis \( \{ |1\rangle, |-1\rangle \} \).

2. Answer the following questions about the matrix representation of a Hermitian operator.

(a) Given that the matrix

\[ M = \begin{bmatrix} 6 & 4 \\ \alpha & 0 \end{bmatrix} \]

is Hermitian, what is \( \alpha \)?

(b) The eigenvalues are 8 and -2. Find NORMALIZED and REAL eigenvectors \( |8\rangle \) and \( |-2\rangle \), remembering that \( |\lambda\rangle \) means an eigenket associated with the eigenvalue \( \lambda \).

(c) Find the matrix representation of \( M \) in the basis consisting of \( |8\rangle \) (the first basis ket) and \( |-2\rangle \) (the second basis ket).

3. Let

\[ A = \begin{bmatrix} 0 & 2 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} -1 & 2 \\ 1 & 0 \end{bmatrix}. \]

(a) Show \( [A, B] = 0 \).
(b) The eigenvalues for \( A \) are 2 and \(-1\). Find two real normalized eigenvectors, \(|2\rangle_A\) and \(|-1\rangle_A\).

(c) Show that \(|2\rangle_A\) and \(|-1\rangle_A\) are also eigenvectors of \( B \), but the eigenvalues are not the same.

(d) Verify that \{\(|2\rangle_A\), \(|-1\rangle_A\)\} forms a basis for \( \mathbb{C}^2 \).

(e) Find the matrix representations of \( A \) and \( B \) relative to the basis \{\(|2\rangle_A\), \(|-1\rangle_A\)\}.

4. Consider two Hermitian matrices

\[
M = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad \text{and} \quad N = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.
\]

(a) Prove that \( M \) and \( N \) commute.

(b) Find the eigenvalues and real normalized eigenvectors of \( M \) and \( N \).

(c) Show that \( M \) and \( N \) are simultaneously diagonalizable.
Chapter 3

Fundamental Postulates and the Mathematical Framework of Quantum Mechanics

This is where quantum mechanics markedly differs from mathematics as you know it\(^1\). The postulates of quantum mechanics cannot be proved or deduced in a purely mathematical fashion. The postulates are hypotheses, that are given to you from the beginning. If no disagreement with experiments is found, the postulates are accepted as correct hypotheses. Sometimes, they are called axioms, which means the postulates are regarded true though not provable.

You should not feel excessively uncomfortable with this. Much of what we know in physics cannot be proved. For example, consider the celebrated Newton’s Second Law of Motion; \( F = ma \). There is no purely mathematical derivation of this relationship. In physics, a theory is deemed “correct” if there is a good agreement between observations and theoretical predictions. While classical Newtonian mechanics suffices for macroscopic systems, a huge body of experimental evidence supports the view that the quantum postulates provide a consistent description of reality on the atomic scale.

It is possible and is general practice to limit the number of the postulates to a minimum of tow or three. While it may be esthetically pleasing, and mathematically

---

\(^1\)I am saying ”mathematics as you know it” because mathematics also relies on many postulates often known as axioms. However, most of us are used to encountering definitions and proofs when doing mathematics and not necessarily postulates.
CHAPTER 3. FUNDAMENTAL POSTULATES AND THE MATHEMATICAL FRAMEWORK OF QUANTUM MECHANICS

more sound, I believe it helps understand the quantum theory more easily to start with a greater number of postulates. Necessarily, not all the postulates presented here are independent of one another. However, my choice is better understanding for beginners in exchange for redundancy.

3.1 The Fundamental Postulates

1. An isolated physical system is associated with a topologically separable\(^a, b\) complex Hilbert space \(\mathcal{H}\) with inner product \(\langle \phi | \psi \rangle\). Physical states can be identified with equivalence classes of vectors of length 1 in \(\mathcal{H}\), where two vectors represent the same state if they differ only by a phase factor.

\(^a\)A topological space is separable if it contains a countable dense subset; that is, there exists a sequence \(\{x_i\}_{i=1}^{\infty}\) of elements of the space such that every nonempty open subset of the space contains at least one element of the sequence.

\(^b\)Separability means that the Hilbert space \(\mathcal{H}\) has a basis consisting of countably many vectors. Physically, this means that countably many observations suffice to uniquely determine the state.

The above is a mathematically correct and rigorous statement of Postulate 1. A more accessible simplified statement would be as follows.

Each physical system has its associated Hilbert space \(\mathcal{H}\), so that there is one-to-one correspondence between a physical state \(s\) of the system and a vector \(v_s\), or a ket \(|s\rangle\), of unit length in \(\mathcal{H}\).

2. To every observable \(q\) in classical mechanics there corresponds a Hermitian operator \(Q\) in quantum mechanics. Table 3.1 summarizes the correspondences.

In Table 3.1, I am showing the correspondence between classical observables and quantum mechanical operators for a single variable \(x\) for simplicity. If we include \(y\) and \(z\), parts of Table 3.1 should be modified as shown in Table 3.2.

\(^2\)\(H\) or \(\hat{H}\) is called Hamiltonian.
3.1. THE FUNDAMENTAL POSTULATES

<table>
<thead>
<tr>
<th>Observable</th>
<th>$q$</th>
<th>$Q$</th>
<th>Operation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position</td>
<td>$x$</td>
<td>$x$ or $\hat{x}$</td>
<td>$M_x$ : multiplication</td>
</tr>
<tr>
<td>Momentum</td>
<td>$p$</td>
<td>$p$ or $\hat{p}$</td>
<td>$i \left( -i \hbar \frac{\partial}{\partial x} \right)$</td>
</tr>
<tr>
<td>Kinetic Energy</td>
<td>$T$</td>
<td>$T$ or $\hat{T}$</td>
<td>$-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2}$</td>
</tr>
<tr>
<td>Potential Energy</td>
<td>$V(x)$</td>
<td>$V(x)$ or $\hat{V}(x)$</td>
<td>$M_{V(x)}$ : multiplication</td>
</tr>
<tr>
<td>Total Energy</td>
<td>$E$</td>
<td>$H$ or $H^2$</td>
<td>$-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x)$</td>
</tr>
<tr>
<td>Angular Momentum</td>
<td>$l_x$</td>
<td>$L_x$ or $\hat{L}_x$</td>
<td>$-i\hbar \left( y \frac{\partial}{\partial z} - z \frac{\partial}{\partial y} \right)$</td>
</tr>
</tbody>
</table>

Table 3.1: Physical observables and corresponding quantum operators

<table>
<thead>
<tr>
<th>Observable</th>
<th>$q$</th>
<th>$Q$</th>
<th>Operation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Momentum</td>
<td>$p$</td>
<td>$p$ or $\hat{p}$</td>
<td>$-i\hbar \left( \frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z} \right)$</td>
</tr>
<tr>
<td>Kinetic Energy</td>
<td>$T$</td>
<td>$T$ or $\hat{T}$</td>
<td>$-\frac{\hbar^2}{2m} \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right)$</td>
</tr>
<tr>
<td>Potential Energy</td>
<td>$V(r)$</td>
<td>$V(r)$ or $\hat{V}(r)$</td>
<td>$M_{V(r)}$ : multiplication</td>
</tr>
<tr>
<td>Total Energy</td>
<td>$E$</td>
<td>$H$ or $H^2$</td>
<td>$-\frac{\hbar^2}{2m} \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) + V(r)$</td>
</tr>
<tr>
<td>Angular Momentum</td>
<td>$l_x$</td>
<td>$L_x$ or $\hat{L}_x$</td>
<td>$-i\hbar \left( y \frac{\partial}{\partial z} - z \frac{\partial}{\partial y} \right)$</td>
</tr>
<tr>
<td></td>
<td>$l_y$</td>
<td>$L_y$ or $\hat{L}_y$</td>
<td>$-i\hbar \left( z \frac{\partial}{\partial x} - x \frac{\partial}{\partial z} \right)$</td>
</tr>
<tr>
<td></td>
<td>$l_z$</td>
<td>$L_z$ or $\hat{L}_z$</td>
<td>$-i\hbar \left( x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right)$</td>
</tr>
</tbody>
</table>

Table 3.2: Quantum operators in three dimensions

3. When a measurement is made for an observable $q$ associated with operator $Q$, the only possible outcomes are the eigenvalues of $Q$ denoted by $\{\lambda_i\}$.

4. The set of eigenstates/eigenvectors of operator $Q$, denoted by $|\lambda_i\rangle$, forms a complete$^3$ orthonormal set of the Hilbert space $\mathcal{H}$.

---

$^3$A collection of vectors $\{v_\alpha\}_{\alpha \in A}$ in a Hilbert space $\mathcal{H}$ is complete if $\langle w \mid v_\alpha \rangle = 0$ for all $\alpha \in A$ implies $w = 0$. Equivalently, $\{v_\alpha\}_{\alpha \in A}$ is complete if the span of $\{v_\alpha\}_{\alpha \in A}$ is dense in $\mathcal{H}$, that is, given $w \in \mathcal{H}$ and $\epsilon > 0$, there exists $w' \in \text{span}\{v_\alpha\}$ such that $\|w' - w\| < \epsilon$. 
5. *Suppose a physical system is in state* \( s \). *If the vector* \( v_s \in \mathcal{H} \) *associated with the physical state* \( s \) *is given by*

\[
v_s = \sum_{i=1}^{n} a_i |\lambda_i\rangle;
\]

*where* \( n \) *may go to* \( \infty \), *the possible outcomes of a measurement of the physical quantity* \( q \) *are the eigenvalues* \( \lambda_1, \lambda_2, \ldots, \lambda_k, \ldots \), *and the probability that* \( \lambda_k \) *is observed is*

\[
|\langle \lambda_k | v_s \rangle|^2 = |a_k|^2.
\]

6. *As soon as a measurement is conducted on a physical state* \( s \), *yielding* \( \lambda_k \) *as the outcome, the state vector* \( v_s \) *collapses to the corresponding eigenstate* \( |\lambda_k\rangle \). *Therefore, measurement affects the physical state of the system.*

7. *The average value of the observable* \( q \) *after a large number of measurements is given by the expectation value*

\[
\langle \bar{q} \rangle = \langle v_s | Q | v_s \rangle.
\]

8. *The state vector* \( |s\rangle \) *associated with a physical state* \( s \) *typically depends on time and the spatial coordinates; that is,* \( |s\rangle = |s(r,t)\rangle \). *The state evolves in time according to the time-dependent Schrödinger equation.*

\[
i\hbar \frac{\partial}{\partial t} |s(r,t)\rangle = \hat{H} |s(r,t)\rangle
\]

*In a typical application, our state vector is a differentiable function, called a wavefunction, and the canonical symbol used for the wavefunction is* \( \Psi \). *With this notation, we get*

\[
i\hbar \frac{\partial \Psi(r,t)}{\partial t} = \hat{H} \Psi(r,t).
\]

9. *We can use the Schrödinger Equation to show that the first derivative of the wave function should be continuous, unless the potential becomes infinite across the boundary. The wavefunction* \( \Psi(r,t) \) *and its spatial derivatives* \( \frac{\partial}{\partial x}\Psi(r,t), \frac{\partial}{\partial y}\Psi(r,t), \frac{\partial}{\partial z}\Psi(r,t) \) *are continuous if the potential* \( V(r) = V(x, y, z) \) *is finite.*

---

*This fact is often used in elaborate experimental verifications of quantum mechanics.*
10. **You can skip this for now.**

The total wavefunction must be antisymmetric with respect to the interchange of all coordinates of one fermion\(^a\) with those of another. Electron spin must be included in this set of coordinates. The Pauli exclusion principle is a direct consequence of this antisymmetry principle. Slater determinants provide a convenient means of enforcing this property on electronic wavefunctions.

\(^a\)Particles are classified into two categories, fermions and bosons. Fermions have half-integral intrinsic spins such as 1/2, 3/2, and so forth, while bosons have integral intrinsic spins such as 0, 1, and so forth. Quarks and leptons, as well as most composite particles, like protons and neutrons, are fermions. All the force carrier particles, such as photons, W and Z bosons, and gluons are bosons, as are those composite particles with an even number of fermion particles like mesons.

Of the 10 postulates, Postulates 8 is distinctly different from the other nine. The nine postulates describe the system at a given time \(t\), while Postulate 8 specifies how the system changes with time.

### 3.2 Unitary Time Evolution

Time evolution of a quantum system is always given by a unitary transformation \(\hat{U}\), such that

\[
|s(t)\rangle = \hat{U}(t) |s(0)\rangle.
\]

(3.6)

The nature of \(\hat{U}(t)\) depends on the system and the external forces it experiences. However, \(\hat{U}(t)\) does not depend on the state \(|s\rangle\). Hence, \(\hat{U}(\alpha |s_1\rangle + \beta |s_2\rangle) = \alpha \hat{U} |s_1\rangle + \beta \hat{U} |s_2\rangle\), and the time evolution operator \(\hat{U}\) is linear. Again, the key here is that the time evolution operator \(\hat{U}\) is a function of the physical system and not individual states. Sometimes \(\hat{U}(t)\) is referred to as a propagator.

The unitarity of \(\hat{U}\) follows from the time-dependent Schrödinger equation (3.4) as follows. We have

\[
\frac{i\hbar}{\partial t} |s(t)\rangle = \hat{H}(t) |s(t)\rangle \quad \text{or} \quad \frac{\partial}{\partial t} |s(t)\rangle = -\frac{i}{\hbar} \hat{H}(t) |s(t)\rangle;
\]

(3.7)

where \(\hat{H}\) is the Hamiltonian of the system and is Hermitian. Suppose

\[
|s(t)\rangle = \hat{U}(t) |s(0)\rangle
\]

(3.8)
for some operator $\hat{U}$. Plugging this into the time-dependent Schrödinger equation, we obtain
\[
\frac{\partial}{\partial t} \left( \hat{U}(t) |s(0)\rangle \right) = -\frac{i}{\hbar} \hat{H}(t) \hat{U}(t) |s(0)\rangle \implies \left( \frac{\partial}{\partial t} \hat{U}(t) \right) |s(0)\rangle = \left( -\frac{i}{\hbar} \hat{H}(t) \hat{U}(t) \right) |s(0)\rangle
\]
(3.9)
for any physical state $s$. This means that the actions of
\[
\frac{\partial}{\partial t} \hat{U}(t) \quad \text{and} \quad -\frac{i}{\hbar} \hat{H}(t) \hat{U}(t)
\]
on any set of basis vectors are the same. Therefore,
\[
\frac{\partial}{\partial t} \hat{U}(t) = -\frac{i}{\hbar} \hat{H}(t) \hat{U}(t) \tag{3.10}
\]
Taking the adjoint, and noting that $\hat{H} = \hat{H}^\dagger$,
\[
\left( \frac{\partial}{\partial t} \hat{U}(t) \right)^\dagger = \left( -\frac{i}{\hbar} \hat{H}(t) \hat{U}(t) \right)^\dagger \implies \frac{\partial}{\partial t} \hat{U}^\dagger(t) = \frac{i}{\hbar} \hat{U}^\dagger(t) \hat{H}^\dagger(t) = \frac{i}{\hbar} \hat{U}^\dagger(t) \hat{H}(t). \tag{3.12}
\]
Now, at $t = 0$, $\hat{U}(0) = I$ by necessity, and this gives $\hat{U}^\dagger(0) U(0) = I$. On the other hand,
\[
\frac{\partial}{\partial t} \left( \hat{U}^\dagger(t) \hat{U}(t) \right) = \left( \frac{\partial}{\partial t} \hat{U}^\dagger(t) \right) \hat{U}(t) + \hat{U}^\dagger(t) \left( \frac{\partial}{\partial t} \hat{U}(t) \right) = \frac{i}{\hbar} \hat{U}^\dagger(t) \hat{H}(t) \hat{U}(t) + \hat{U}^\dagger(t) \frac{-i}{\hbar} \hat{H}(t) \hat{U}(t)
\]
(3.13)
Hence, $\hat{U}^\dagger(t) \hat{U}(t) = I$ at all times $t$, and $\hat{U}(t)$ is unitary. Because $\hat{U}(t)$ is unitary,
\[
| |s(t)\rangle|^2 = \langle s(t) | s(t) \rangle = \langle s(0) | \hat{U}^\dagger(t) \hat{U}(t) | s(0) \rangle = \langle s(0) | I | s(0) \rangle = | |s(0)\rangle|^2 \tag{3.14}
\]
and the magnitude of the state vector is preserved. One way to interpret time evolution is to regard it as a rotation of the state vector in Hilbert space.
3.3 Time-Independent Hamiltonian

So far, we have considered a Hamiltonian with explicit time-dependence as is clear from the notation $\hat{H}(t)$. However, for many physical systems, the Hamiltonian is not time-dependent. For example, consider a typical Hamiltonian given by

\begin{equation}
H = -\frac{\hbar^2}{2m} \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) + V(\mathbf{r})
\end{equation}

(3.15)

in Table 3.2. In the most general case, the potential energy term $V$ can have explicit time dependence so that $V = V(\mathbf{r}, t)$. However, if $V$ does not depend on $t$, and neither does $H$ as shown above, there are simple relations among the time evolution operator $U(t)$, the system Hamiltonian $H$, and the eigenstates of $H$.

3.3.1 Propagator as a Power Series of $H$

We start with the differential relation (3.11)

\begin{equation}
\frac{d}{dt} \hat{U}(t) = -\frac{i}{\hbar} H(t) \hat{U}(t);
\end{equation}

(3.16)

where we switched to total time derivative as we will only focus on $t$ here, holding all the other variables fixed. Let us now remove time-dependence from $H$ in (3.16).

\begin{equation}
\frac{d}{dt} \hat{U}(t) = -\frac{i}{\hbar} H \hat{U}(t);
\end{equation}

(3.17)

We will compare this with the familiar case of a real-valued function $f : \mathbb{R} \rightarrow \mathbb{R}$ satisfying the differential equation

\begin{equation}
\frac{d}{dt} f(t) = -\frac{i}{\hbar} H f(t);
\end{equation}

(3.18)

where $H$ is some scalar. As the answer to this differential equation is

\begin{equation}
f(t) = f(0) \exp \left[ -\frac{iHt}{\hbar} \right].
\end{equation}

(3.19)
you may think the answer to the analogous operator differential equation (3.17) is
something like

\[ \hat{U}(t) = \hat{U}(0) \exp \left[ \frac{-iHt}{\hbar} \right] = \exp \left[ \frac{-iHt}{\hbar} \right]; \tag{3.20} \]

where the second equality holds as \( \hat{U}(0) \) is the identity operator by necessity. As it
turns out, this is the right answer with the interpretation of the right-hand side as a
power series in operator \( H \). This is explained in more detail in Appendix B. But, we
will only check it formally here to convince ourselves that the solution really works.
Maclaurin series of the exponential function gives

\[ \exp \left[ \frac{-iHt}{\hbar} \right] = \sum_{n=0}^{\infty} \frac{\left[ \frac{-iH}{\hbar} \right]^n t^n}{n!} = I + \sum_{n=1}^{\infty} \frac{\left[ \frac{-iH}{\hbar} \right]^n t^n}{n!}; \tag{3.21} \]

where \( \left[ \frac{-iH}{\hbar} \right]^0 = I \) by definition. Then,

\[
\frac{d}{dt} \left( \exp \left[ \frac{-iHt}{\hbar} \right] \right) = \frac{d}{dt} \left( I + \sum_{n=1}^{\infty} \frac{\left[ \frac{-iH}{\hbar} \right]^n t^n}{n!} \right) = \frac{d}{dt} \left( \sum_{n=1}^{\infty} \frac{\left[ \frac{-iH}{\hbar} \right]^n t^n}{n!} \right) = \sum_{n=1}^{\infty} n \frac{\left[ \frac{-iH}{\hbar} \right]^n t^{n-1}}{n(n-1)!} \\
= -\frac{iH}{\hbar} \sum_{n=1}^{\infty} \frac{\left[ \frac{-iH}{\hbar} \right]^{n-1} t^{n-1}}{(n-1)!} = -\frac{iH}{\hbar} \sum_{n=0}^{\infty} \frac{\left[ \frac{-iH}{\hbar} \right]^n t^n}{n!} \\
= -\frac{i}{\hbar} H \exp \left[ \frac{-iHt}{\hbar} \right]. \tag{3.22} \]

So,

\[ \hat{U}(t) = \exp \left[ \frac{-iHt}{\hbar} \right] \tag{3.23} \]

satisfies the differential equation (3.17).

In Section 3.2, we showed that the time evolution operator \( \hat{U}(t) \) is unitary for
any Hamiltonian \( H \). When \( H \) does not depend on time, we have an alternative
proof that \( \hat{U}(t) \) is unitary. However, note that our discussion below is somewhat
heuristic and not completely rigorous mathematically speaking.
First, taking the term-by-term adjoint of the right-hand side of (3.21), we get

\[
\left( \exp \left[ \frac{-iHt}{\hbar} \right] \right)^\dagger = \left( \sum_{n=0}^{\infty} \frac{(-iH)^n}{n!} t^n \right)^\dagger = \sum_{n=0}^{\infty} \left[ \left( \frac{-iH}{\hbar} \right)^n \right]^\dagger t^n = \sum_{n=0}^{\infty} \frac{(-i\tilde{H})^n}{n!} t^n = \exp \left[ \frac{\tilde{H}t}{\hbar} \right].
\]

It suffices to show

\[
\hat{U}(t)\hat{U}^\dagger(t) = \exp \left[ \frac{-iHt}{\hbar} \right] \exp \left[ \frac{iHt}{\hbar} \right] = I.
\]

Recall from Theorem 2.8 that a matrix representing a Hermitian operator is diagonalizable with its eigenvalues on the diagonal. Because \( H \) is Hermitian, there exists an invertible matrix \( J \) such that

\[
JHJ^{-1} = D = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_m \end{bmatrix} \quad \text{or} \quad H = J^{-1}DJ = J^{-1} \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_m \end{bmatrix} J;
\]

where \( \lambda_1, \ldots, \lambda_m \) are the eigenvalues of \( H \). We also have

\[
H^n = (J^{-1}DJ)^n = J^{-1}D^nJ = J^{-1} \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_m \end{bmatrix}^n J = J^{-1} \begin{bmatrix} \lambda_1^n & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_m^n \end{bmatrix} J.
\]

Therefore,

\[
\hat{U}(t) = e^{-\frac{iHt}{\hbar}} = \sum_{n=0}^{\infty} \left( \frac{-i}{\hbar} \right)^n H^n t^n \frac{1}{n!} = \sum_{n=0}^{\infty} \left( \frac{-i}{\hbar} \right)^n J^{-1}D^nJt^n \frac{1}{n!}
\]

\[
= \sum_{n=0}^{\infty} J^{-1} \left( \frac{-i}{\hbar} \right)^n D^n t^n \frac{1}{n!} J = J^{-1} \left( \sum_{\lambda_n=0}^{\infty} \left( \frac{-i}{\hbar} \right)^n D^n \frac{1}{n!} \right) J
\]

\[
= J^{-1} \left( \sum_{n=0}^{\infty} \left( \frac{-i}{\hbar} \right)^n \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_m \end{bmatrix}^n t^n \frac{1}{n!} \right) J
\]
\[ J = J^{-1} \left( \sum_{n=0}^{\infty} \left( \frac{-i}{\hbar} \right)^n \left[ \begin{array}{c} \lambda_1^n \\ \vdots \\ \lambda_m^n \end{array} \right] t^n \frac{1}{n!} \right) \]

\[ = J^{-1} \left[ \sum_{n=0}^{\infty} \left( \frac{-i}{\hbar} \right)^n \lambda_1^n t^n \frac{1}{n!} \right] 
\quad \cdots 
\quad \sum_{n=0}^{\infty} \left( \frac{-i}{\hbar} \right)^n \lambda_m^n t^n \frac{1}{n!} \right] \]

\[ = J^{-1} \left[ \sum_{n=0}^{\infty} \left( \frac{-i}{\hbar} \lambda_1 t \right)^n \frac{1}{n!} 
\quad \cdots 
\quad \sum_{n=0}^{\infty} \left( \frac{-i}{\hbar} \lambda_m t \right)^n \frac{1}{n!} \right] \]

\[ = J^{-1} \left[ \exp \left( \frac{-i}{\hbar} \lambda_1 t \right) 
\quad \cdots 
\quad \exp \left( \frac{-i}{\hbar} \lambda_m t \right) \right] J. \quad (3.28) \]

Similarly,

\[ \hat{U}^\dagger(t) = e^{\hat{H}t} = J^{-1} \left[ \exp \left( \frac{i}{\hbar} \lambda_1 t \right) 
\quad \cdots 
\quad \exp \left( \frac{i}{\hbar} \lambda_m t \right) \right] J. \quad (3.29) \]

We now have

\[ \hat{U}(t)\hat{U}^\dagger(t) = J^{-1} \left[ \exp \left( \frac{\mp i}{\hbar} \lambda_1 t \right) 
\quad \cdots 
\quad \exp \left( \frac{\mp i}{\hbar} \lambda_m t \right) \right] J 
\quad \exp \left( \frac{i}{\hbar} \lambda_1 t \right) 
\quad \cdots 
\quad \exp \left( \frac{i}{\hbar} \lambda_m t \right) \right] J
\]

\[ = J^{-1} \left[ \exp \left( \frac{-i}{\hbar} \lambda_1 t \right) 
\quad \cdots 
\quad \exp \left( \frac{-i}{\hbar} \lambda_m t \right) \right] \left[ \exp \left( \frac{i}{\hbar} \lambda_1 t \right) 
\quad \cdots 
\quad \exp \left( \frac{i}{\hbar} \lambda_m t \right) \right] J
\]

\[ = J^{-1} \left[ \begin{array}{c} 1 \\
\vdots \\
1 \end{array} \right] J = I. \quad (3.30) \]
Therefore, $\hat{U}(t)$ is indeed unitary for all $t$.

### 3.3.2 Eigenstate Expansion of the Propagator

Here, we will discuss how the propagator $\hat{U}(t)$ can be expressed using the normalized eigenkets of $H$. The eigenvalue problem for $H$ is

$$H |E\rangle = E |E\rangle; \quad (3.31)$$

where $E$ is an eigenvalue representing the total energy, and $|E\rangle$ is an associated normalized eigenket. From (2.48),

$$\sum |E\rangle \langle E| = I. \quad (3.32)$$

Hence,

$$|s(t)\rangle = \sum |E\rangle \langle E|s(t)\rangle, \quad (3.33)$$

and $\langle E|s(t)\rangle$ gives the coefficients when $|s(t)\rangle$ is expressed as a linear combination of $\{|E\rangle\}$. For simplicity of notation, let $c_E(t) = \langle E|s(t)\rangle$, so that

$$|s(t)\rangle = \sum c_E(t) |E\rangle. \quad (3.34)$$

From Postulate 8,

$$i\hbar \frac{\partial}{\partial t} \Psi(r, t) = \hat{H} \Psi(r, t) \implies \hbar \frac{\partial |s(t)\rangle}{\partial t} = H |s(t)\rangle$$

$$\implies \hbar \frac{\partial}{\partial t} \sum c_E(t) |E\rangle = H \sum c_E(t) |E\rangle \implies \sum \hbar \frac{\partial c_E(t)}{\partial t} |E\rangle = \sum c_E(t) E |E\rangle$$

$$\implies \sum \left( \hbar \frac{\partial c_E(t)}{\partial t} - E c_E(t) \right) |E\rangle = 0. \quad (3.35)$$

As $|E\rangle$’s are linearly independent,

$$i\hbar \frac{\partial c_E(t)}{\partial t} - E c_E(t) = 0 \quad \text{for each value of } E. \quad (3.36)$$

This implies that $c_E(t)$ is an exponential function.

$$i\hbar \frac{\partial c_E(t)}{\partial t} - E c_E(t) = 0 \implies \frac{\partial c_E(t)}{\partial t} = \frac{-iE}{\hbar} c_E(t) \implies c_E(t) = c_E(0)e^{-iEt/\hbar}. \quad (3.37)$$
CHAPTER 3. FUNDAMENTAL POSTULATES AND THE MATHEMATICAL FRAMEWORK OF QUANTUM MECHANICS

Remembering that \[ c_E(t) = \langle E | s(t) \rangle, \] and in particular \[ c_E(0) = \langle E | s(0) \rangle, \] we have

\[
|s(t)\rangle = \sum \langle E | s(0) \rangle e^{-iEt/\hbar} |E\rangle = \sum |E\rangle \langle E | s(0) \rangle e^{-iEt/\hbar}. \tag{3.38}
\]

Comparing this with (3.6), we get

\[
\hat{U}(t) = \sum_E |E\rangle \langle E | e^{-iEt/\hbar}. \tag{3.39}
\]

This is the eigenstate expansion of the propagator \( \hat{U}(t) \).

Let us double check and see if this expression of \( \hat{U}(t) \) is unitary. Following the steps described in Note 2.1,

\[
\hat{U}^\dagger(t) = \left( \sum_E |E\rangle \langle E | e^{-iEt/\hbar} \right)^\dagger = \sum_E \left( |E\rangle \langle E | e^{-iEt/\hbar} \right)^\dagger \overset{\text{Step } 1}{\longrightarrow} \sum_E e^{-iEt/\hbar} \langle E | E \rangle \overset{\text{Step } 2}{\longrightarrow} \sum_E e^{iE't/\hbar} \langle E' | E \rangle \overset{\text{Step } 3}{\longrightarrow} \hat{U}^\dagger(t) = \sum_E e^{iE't/\hbar} |E\rangle \langle E |. \tag{3.40}
\]

Hence,

\[
\hat{U}(t)\hat{U}^\dagger(t) = \left( \sum_E |E\rangle \langle E | e^{-iEt/\hbar} \right) \left( \sum_{E'} e^{iE't/\hbar} |E'\rangle \langle E' | \right)
= \sum_{E,E'} |E\rangle \langle E | E' \rangle \langle E' | e^{-i(E-E')t/\hbar} = \sum_{E=E'} |E\rangle \langle E' | e^{-i(E-E')t/\hbar} = \sum_E |E\rangle \langle E | = I. \tag{3.41}
\]

This checks.

3.4 The Uncertainty Principle

The uncertainty principle is one of the celebrated consequences of quantum mechanics, whose influence has gone beyond the realm of physics into humanities such as philosophy.

3.4.1 Uncertainty and Non-Commutation

**Theorem 3.1 (The Uncertainty Relation)** Consider two physical observables represented by Hermitian operators \( A \) and \( B \). We denote the expectation value of an
3.4. THE UNCERTAINTY PRINCIPLE

operator $\Omega$ by $\langle \Omega \rangle$; that is, if the state of the system is $|u\rangle$ after normalization,

$$\langle \Omega \rangle = \langle u|\Omega|u\rangle. \quad (3.42)$$

In addition, define $\Delta A$ by

$$\Delta A = \sqrt{\langle (A - \langle A\rangle I)^2 \rangle}. \quad (3.43)$$

Then, $[A, B] = i\alpha$ implies $\Delta A \Delta B \geq |\alpha|^2$; where $\alpha$ is a real scalar.

**Proof**

We will first define two Hermitian operators $C_A$ and $C_B$ as follows.

$$C_A = A - \langle A \rangle I \quad \text{and} \quad C_B = B - \langle B \rangle I \quad (3.44)$$

Then,

$$[C_A, C_B] = [A - \langle A \rangle I, B - \langle B \rangle I] = [A - \langle A \rangle I, B] - [A - \langle A \rangle I, \langle B \rangle I] = [A, B] = i\alpha. \quad (3.45)$$

Next define two vectors by

$$|v\rangle = C_A |u\rangle \quad \text{and} \quad |w\rangle = C_B |u\rangle. \quad (3.46)$$

Then,

$$\langle v|v\rangle = \langle u|C_A^\dagger C_A |u\rangle = \langle u|C_A C_A |u\rangle = \langle u|C_A^2 |u\rangle = \langle u|(A - \langle A \rangle I) |u\rangle = \langle (A - \langle A \rangle)^2 \rangle \quad (3.47)$$

and

$$\langle w|w\rangle = \langle u|C_B^\dagger C_B |u\rangle = \langle u|C_B C_B |u\rangle = \langle u|C_B^2 |u\rangle = \langle u|(B - \langle B \rangle I) |u\rangle = \langle (B - \langle B \rangle)^2 \rangle. \quad (3.48)$$

---

\(^5\)From the formula, you can see that $\Delta A$ is the standard deviation of the observable $A$ in the state $|u\rangle$.

\(^6\)This definition means $\langle \Delta A \rangle^2 = \langle (A - \langle A \rangle)^2 \rangle = \langle A^2 - 2A\langle A \rangle + \langle A \rangle^2 \rangle = \langle A^2 \rangle - 2\langle A \rangle \langle A \rangle + \langle A \rangle^2 = \langle A^2 \rangle - \langle A \rangle^2$ or $\Delta A = \sqrt{\langle A^2 \rangle - \langle A \rangle^2}$.\)
Therefore,
\[ \langle v|v \rangle = (\Delta A)^2 \quad \text{and} \quad \langle w|w \rangle = (\Delta B)^2. \]  
(3.49)

Now, by Schwarz Inequality (Theorem 2.3),
\[ (\Delta A)^2 (\Delta B)^2 = \langle v|v \rangle \langle w|w \rangle \geq |\langle v|w \rangle|^2 = |\langle u|C_A C_B|u \rangle|^2 \]
\[ = |\langle u|(C_A C_B - C_B C_A + C_A C_B + C_B C_A)/2|u \rangle|^2 = \frac{1}{4} |\langle u|[C_A, C_B] + \{C_A, C_B\}|u \rangle|^2; \]
(3.50)

where \( \{C_A, C_B\} = C_A C_B + C_B C_A \) is called an anti-commutator. Because \( C_A \) and \( C_B \) are Hermitian,
\[ \langle u|\{C_A, C_B\}|u \rangle = \langle u|C_A C_B + C_B C_A|u \rangle = \langle u|C_A C_B|u \rangle + \langle u|C_B C_A|u \rangle \]
\[ = \langle u|C_A C_B|u \rangle + \langle u|C_B C_A|u \rangle = \langle u|C_A|C_B u \rangle + \langle u|C_B|C_A u \rangle \]
\[ = \langle C_A|C_B u \rangle + \langle C_B|C_A u \rangle^* = 2 \Re \langle C_A|C_B u \rangle. \]
(3.51)

For simplicity of notation, let \( \beta = 2 \Re \langle C_A|C_B u \rangle \). Then, we have
\[ (\Delta A)^2 (\Delta B)^2 \geq \frac{1}{4} |i \alpha + \beta|^2 = \frac{1}{4} (\alpha^2 + \beta^2) \geq \frac{1}{4} \alpha^2. \]
(3.52)

Therefore,
\[ \Delta A \Delta B \geq \frac{|\alpha|}{2}. \]
(3.53)

\[ \] 3.4.2 Position and Momentum

From Table 3.1, \( \hat{p} = -i \hbar \frac{d}{dx} \) and \( \hat{x} = M_x \). In order to compute the commutator \( [\hat{x}, \hat{p}] \), we will let \( [\hat{x}, \hat{p}] \) act on a function \( f(x) \).
\[ [\hat{x}, \hat{p}] f(x) = x(-i \hbar \frac{d}{dx}) f(x) - (-i \hbar \frac{d}{dx} x) f(x) = -i \hbar x \frac{d}{dx} f(x) + i \hbar \frac{d}{dx} (xf(x)) \]
\[ = -i \hbar x \frac{d}{dx} f(x) + i \hbar f(x) + i \hbar x \frac{d}{dx} f(x) = i \hbar f(x) \quad \Rightarrow \quad [\hat{x}, \hat{p}] = i \hbar \]  
(3.54)
Therefore,
\[ \Delta x \Delta p \geq \frac{\hbar}{2}. \]  
(3.55)

### 3.4.3 Energy and Time

The momentum-position uncertainty principle \(\Delta x \Delta p \geq \frac{\hbar}{2}\) has an energy-time analog, \(\Delta E \Delta t \geq \frac{\hbar}{2}\). However, this must be a different kind of relationship because \(t\) is not a dynamical variable. This uncertainty cannot have anything to do with lack of commutation.

#### 3.4.3.1 Time rate of change of expectation values

How do expectation values change in time? That is, how can we compute

\[
\frac{d}{dt} \langle A \rangle(t) = \frac{d}{dt} \langle u(t)|A|u(t) \rangle;
\]

(3.56)

where \(|u(t)\rangle\) is the state of the system at time \(t\). From (3.4),

\[
\frac{d}{dt} |u(t)\rangle = \frac{1}{i\hbar} H |u(t)\rangle \quad \text{and} \quad \frac{d}{dt} \langle u(t) | = -\frac{1}{i\hbar} \langle u(t) | H^a \rangle.
\]

(3.57)

So,

\[
\frac{d}{dt} \langle A \rangle(t) = \left( \frac{d}{dt} \langle u(t) | \right) A |u(t)\rangle + \langle u(t) | A \left( \frac{d}{dt} |u(t)\rangle \right) = -\frac{1}{i\hbar} \langle u(t) | HA |u(t)\rangle + \langle u(t) | A \frac{1}{i\hbar} H |u(t)\rangle
\]

\[
= \frac{1}{i\hbar} \langle u(t) | - HA + AH |u(t)\rangle = \frac{1}{i\hbar} \langle [A, H] \rangle(t).
\]

(3.58)

We have

\[
\frac{d}{dt} \langle A \rangle(t) = \frac{1}{i\hbar} \langle [A, H] \rangle(t).
\]

(3.59)

The indication here is that dynamical evolution of an observable depends on its level of commutativity with \(H\), at least on the average.
3.4.3.2 Time-energy uncertainty

Our derivation of the general form of the uncertainty principle (3.53) concerns two observables. However, as mentioned already, time \( t \) is not a dynamical variable. We need a way to make a connection between \( t \) and an observable \( A \) if we are to use the general inequality (3.53) to formulate time-energy uncertainty.

We will use an observable \( A \) to characterize the change in the system in time. As \( A \) and the energy \( E \) are observables, we have

\[
\Delta A \Delta E \geq \frac{1}{2} |\langle[A, H]\rangle| \tag{3.60}
\]

from (3.53). Now, from (3.59),

\[
\frac{1}{2} |\langle[A, H]\rangle| = \frac{1}{2} \left| i\hbar \frac{d}{dt} \langle A \rangle \right| = \frac{\hbar}{2} \left| \frac{d}{dt} \langle A \rangle \right|. \tag{3.61}
\]

Hence,

\[
\Delta A \Delta E \geq \frac{\hbar}{2} \left| \frac{d}{dt} \langle A \rangle \right|. \tag{3.62}
\]

The ratio of the amount of uncertainty in \( A \), denoted by \( \Delta A \), to the rate of change in the expected value of \( A \), \( \left| \frac{d}{dt} \langle A \rangle \right| \), is one candidate for \( \Delta t \). With this definition, we have

\[
\Delta t = \frac{\Delta A}{\left| \frac{d}{dt} \langle A \rangle \right|}. \tag{3.63}
\]

Then,

\[
\Delta E \Delta t = \frac{\Delta E \Delta A}{\left| \frac{d}{dt} \langle A \rangle \right|} \geq \frac{\hbar}{2} \left| \frac{d}{dt} \langle A \rangle \right| = \frac{\hbar}{2} \implies \Delta E \Delta t \geq \frac{\hbar}{2}. \tag{3.64}
\]

Again this inequality is completely different from that for two observables. It depends on how you devise your way to measure time which is different from "the" usual time \( t \). This may be the reason why time-energy uncertainty is so notoriously controversial.
3.4. THE UNCERTAINTY PRINCIPLE

3.5 Classical Mechanics and Quantum Mechanics

According to the correspondence principle advocated by Niels Bohr, every new physical theory must contain as a limiting case the old theory. So, quantum mechanics has to be a superset of classical mechanics. Indeed, classical mechanics is a limiting case of quantum mechanics as $\hbar$ or $\hbar \to 0$. It was also shown by Ehrenfest that the expectation values of quantum mechanical observables satisfy the same equations as the corresponding variables in classical mechanics.

3.5.1 Ehrenfest’s Theorem

One embodiment of the correspondence principle are the two relations proved by Ehrenfest that connect classical mechanics and quantum mechanics.

**Theorem 3.2 (Ehrenfest Theorem)** If $\Omega$ is a quantum mechanical operator and $\langle \Omega \rangle$ is its expectation value, we have

$$\frac{d}{dt} \langle \Omega \rangle = \frac{1}{i\hbar} \langle [\Omega, H] \rangle + \left\langle \frac{\partial \Omega}{\partial t} \right\rangle.$$  \hspace{1cm} (3.65)

**Proof**

This is a generalized version of (3.59); where the operator is possibly time-dependent. We will write out the integral for the expectation values explicitly in order to see the reasons for the total derivative $\frac{d}{dt}$ and the partial derivative $\frac{\partial}{\partial t}$ clearly. We will specialize to a one-dimensional case without loss of generality and let $\Psi(x,t)$ be a normalized wavefunction representing the state. Then,

$$\frac{d}{dt} \langle \Omega \rangle = \frac{d}{dt} \int \Psi(x,t)^*\Omega(x,t)\Psi(x,t) \, dx = \int \frac{\partial}{\partial t} \left( \Psi(x,t)^*\Omega(x,t)\Psi(x,t) \right) \, dx$$

$$= \int \frac{\partial}{\partial t} \left( \Psi(x,t)^* \right) \Omega(x,t)\Psi(x,t) \, dx$$

$$+ \int \Psi(x,t)^* \frac{\partial}{\partial t} \left( \Omega(x,t) \right) \Psi(x,t) \, dx$$

$$+ \int \Psi(x,t)^*\Omega(x,t) \frac{\partial}{\partial t} \left( \Psi(x,t) \right) \, dx.$$  \hspace{1cm} (3.66)
Now, from Postulate 8 on p. 74, we have
\[ i\hbar \frac{\partial}{\partial t} |s(r, t)\rangle = \hat{H} |s(r, t)\rangle \]
\[ \implies \frac{\partial}{\partial t} |s(r, t)\rangle = \frac{1}{i\hbar} \hat{H} |s(r, t)\rangle \quad \text{and} \quad \frac{\partial}{\partial t} \langle s(r, t)| = \frac{1}{-i\hbar} \langle s(r, t)| \hat{H}^c. \] (3.67)

In our case (3.67) becomes
\[ \frac{\partial}{\partial t} \left( \Psi(x, t) \right) = \frac{1}{i\hbar} H \Psi(x, t) \quad \text{and} \quad \frac{\partial}{\partial t} \left( \Psi(x, t)^* \right) = -\frac{1}{i\hbar} \Psi(x, t)^* H. \] (3.68)

Substituting these into (3.66),
\[ \frac{d}{dt} \langle \Omega \rangle = \int \frac{-1}{i\hbar} \Psi(x, t)^* H \Omega(x, t) \Psi(x, t) \, dx + \left\langle \frac{\partial \Omega}{\partial t} \right\rangle + \int \Psi(x, t)^* \Omega(x, t) \frac{1}{i\hbar} H \Psi(x, t) \, dx \]
\[ = \frac{1}{i\hbar} \int \Psi(x, t)^* (\Omega H - H \Omega) \Psi(x, t) \, dx + \left\langle \frac{\partial \Omega}{\partial t} \right\rangle \]
\[ = \frac{1}{i\hbar} \left\langle [\Omega, H] \right\rangle + \left\langle \frac{\partial \Omega}{\partial t} \right\rangle \] (3.69)

In particular, if the Hamiltonian is time-independent as in Section 3.3, Ehrenfest's Theorem implies the following.

**Corollary 3.1 (Time-Independent Hamiltonian)** If the Hamiltonian $H$ does not have explicit time dependence, the time rate of change of the expectation value of a variable $\Omega$ represented by operator $\Omega$, $\langle \Omega \rangle$, is proportional to the expectation value of the commutator between $\Omega$ and $H$, $[\Omega, H]$, with a proportionality constant $\frac{-i}{\hbar}$. Therefore, $\langle \Omega \rangle$ is conserved if $\Omega$ commutes with a time-independent Hamiltonian $H$.

---

*One way to understand this is to regard $|u(t)\rangle$ as an $n \times 1$ matrix (a column vector) and $\langle u(t)|$ as its transpose conjugate (a row vector). With this interpretation, $\frac{d}{dt} |U(t)\rangle$ becomes entry...*
3.4. THE UNCERTAINTY PRINCIPLE

by entry differentiation of a matrix, and $(\frac{d}{dt}|u(t))\dagger = \frac{d}{dt}\langle u(t)\rangle$ and $(\frac{1}{\hbar}H|u(t))\dagger = -\frac{1}{\hbar}\langle u(t)|H$
naturally.

aOne way to understand this is to regard $|u(t)\rangle$ as an $n \times 1$ matrix (a column vector) and
$\langle u(t)\rangle$ as its transpose conjugate (a row vector). With this interpretation, $\frac{d}{dt}|U(t)\rangle$ becomes entry
by entry differentiation of a matrix, and $(\frac{d}{dt}|u(t))\dagger = \frac{d}{dt}\langle u(t)|$ and $(\frac{1}{\hbar}H|u(t))\dagger = -\frac{1}{\hbar}\langle u(t)|H$
naturally.

bWe need to switch from $\frac{d}{dt}$ to $\frac{\partial}{\partial t}$ because the independent variables are $x$ and $t$ before the
integration with respect to $x$ removes $x$ as a variable.

cNote that $\frac{\partial}{\partial t}\langle s(r, t)\rangle = -\frac{1}{\hbar}\langle s(r, t)|\hat{H}\rangle$ follows from $(\frac{\partial}{\partial t}|s(r, t))\dagger = (\frac{1}{\hbar}\hat{H}|s(r, t))\dagger$.

3.5.2 Exact Measurements and Expectation Values

We have seen in Tables 3.1 and 3.2 that each classical observable has its associated
quantum mechanical operator. This is an exact one-to-one correspondence.

However, in terms of actual measurements, a quantum mechanical operator does
not generate one exact and fixed value under repeated measurements as in the classi-
cal measurement. Instead, each quantum mechanical operator specifies a probability
distribution of possible outcomes of the measurement by way of a state function or
wavefunction. This necessitates a probabilistic interpretation of classical laws of
physics.

As we saw in Ehrenfest Theorem, the general strategy is to use expectation values
rather than exact measured values. \textit{In the quantum mechanical formulation
of classical laws of physics, the expectation values play the role of the
classical variables.}
Exercises

1. Let $Q$ be a quantum mechanical operator representing a physical observable $q$. An example of $Q$ is the momentum operator $P = -i\hbar \frac{\partial}{\partial x}$ in one dimension, representing the classical momentum $p$. For simplicity, we will only consider the cases with distinct eigenvalues.

   (a) What can you say about $Q$? In other words, what is the name of the class of operators/matrices $Q$ belongs to?

   (b) Now consider the set $\{|\lambda_i\rangle\}$ of all normalized eigenkets\(^7\) of $Q$.

      i. What is $\langle \lambda_i | \lambda_j \rangle$? This property makes $\{\lambda_i\}$ an orthonormal set.

      ii. The set $\{\lambda_i\}$ is complete. What does this mean?

2. Let $|\lambda_i\rangle$ and $|\lambda_j\rangle$ be two distinct eigenstates of $Q$, i.e. $i \neq j$. Note that $|\lambda_i\rangle$ and $|\lambda_j\rangle$ are normalized as they are states.

   (a) If $\alpha |\lambda_i\rangle + \beta |\lambda_j\rangle$ is a state, not necessarily an eigenstate, where $\alpha$ and $\beta$ are scalars, what is the relation between $\alpha$ and $\beta$?

   (b) With what probabilities will $\lambda_i$ and $\lambda_j$ be observed if a measurement is made?

   (c) A measurement has been made for $q$ and the value obtained was $\lambda_i$. What is the state the system is in after the measurement? You can only answer this up to an arbitrary phase $e^{i\theta}$.

   (d) A second measurement is made on the resulting state above. What value/values will be observed with what probability/probabilities?

---

\(^7\)Recall we use eigenket, eigenvector, and eigenfunction interchangeably.
Chapter 4
Spaces of Infinite Dimensionality

So far, we have focused mostly on finite-dimensional spaces. This was because finite dimensions make it easy to explain mathematics and physics for the author and also easy to understand mathematics and physics for the readership. However, the attentive reader may have wondered why finite dimensions will suffice.

The simplest example is the position $x$ and its quantum mechanical representation (operator) $X$ or $M_x$. The position observable is continuously distributed over the interval $(-\infty, +\infty)$, and hence, there are infinitely many distinct eigenvalues for the operator $X$, which in turn implies that there are infinitely many eigenvectors which are mutually orthogonal according to Theorem 2.7. Needless to say, the dimensionality of the Hilbert space in which the operator $X$ operates has to be infinite.

Another example is the discrete energy distribution as indicated in 4 on p.14. This states that the total energy $E$ can only take discrete values $E_1, E_2, E_3, \ldots$ and not any value $E \in [0, \infty)$ as in classical mechanics where $E$ is continuously distributed. However, it does not place an upper limit on $E$, and infinitely many discrete values of $E$ are possible.

4.1 Two Types of Infinity

Two kinds of infinite dimensional vector spaces are encountered in quantum mechanics. One infinity is known as countable or denumerable infinity, and the other infinity is called uncountable/non-denumerable infinity. Discrete energy level distribution is an example of countable infinity, and the position $x$ is an example of uncountable infinity. A canonical example of countable infinity is the number of
natural numbers $\mathbb{N} = \{1, 2, 3, \ldots\}$, and a good example of uncountable infinity is the number of points in the interval $(0, 1)$. In fact, countable infinity is far “smaller” than uncountable infinity. Namely, countably infinite number of points, when put next to each other, fit in a line segment of length 0, while uncountably many points form a line of strictly positive length. In order to understand the difference, let us see how we can show $\mathbb{N}$ is “of length 0”.

**Proof**

Let $0 < \varepsilon < 1$ be an arbitrary number. Then, we have

$$1 \in \left(1 - \frac{\varepsilon(1 - \varepsilon)}{2}, 1 + \frac{\varepsilon(1 - \varepsilon)}{2}\right), \quad 2 \in \left(2 - \frac{\varepsilon^2(1 - \varepsilon)}{2}, 2 + \frac{\varepsilon^2(1 - \varepsilon)}{2}\right),$$

$$\ldots, \quad n \in \left(n - \frac{\varepsilon^n(1 - \varepsilon)}{2}, n + \frac{\varepsilon^n(1 - \varepsilon)}{2}\right), \quad \ldots.$$  \hspace{1cm} (4.1)

This means $n$ is contained in an interval of length $\varepsilon^n(1 - \varepsilon)$. Therefore, $\mathbb{N}$ is contained in an interval of length

$$\sum_{n=1}^{\infty} \varepsilon^n(1 - \varepsilon). \quad \hspace{1cm} (4.2)$$

Noting that a geometric series $\sum_{n=1}^{\infty} ar^n$ for $0 < r < 1$ converges to $\frac{a}{1-r}$ and we have $(a, r) = (\varepsilon(1 - \varepsilon), \varepsilon)$,

$$\sum_{n=1}^{\infty} \varepsilon^n(1 - \varepsilon) = \frac{\varepsilon(1 - \varepsilon)}{1 - \varepsilon} = \varepsilon. \quad \hspace{1cm} (4.3)$$

Therefore, $\mathbb{N}$ can be contained in an interval of arbitrary length $\varepsilon$, which implies the length of an “interval” containing all natural numbers is 0.

You may wonder why the same argument cannot be used for uncountably many points, but the reason is in the name itself. If there are uncountably many points, you cannot include all of them even if you do $\sum_{n=1}^{\infty} \varepsilon^n(1 - \varepsilon)$ because this is nothing but counting, and you cannot count the uncountable. Here is an additional qualitative description of what is happening. Each point is, by definition, of length 0. If there are only countably many points, the total length is 0 even if there are infinitely many points. However, if there are uncountably many points, the total length is
strictly positive as uncountable infinity overwhelms the length of 0 of each point.

Think of the following two situations as analogues of the above example.

1. Let \( f(x) = \frac{1}{x} \) and \( g(x) = \sqrt{x} \). Then, \( \lim_{x \to \infty} f(x) = 0 \) and \( \lim_{x \to \infty} g(x) = \infty \), but
   \[ \lim_{x \to \infty} f(x)g(x) = \lim_{x \to \infty} \frac{1}{\sqrt{x}} = 0. \]

2. Let \( f(x) = \frac{1}{x} \) and \( g(x) = x^2 \). Then, \( \lim_{x \to \infty} f(x) = 0 \) and \( \lim_{x \to \infty} g(x) = \infty \), but
   \[ \lim_{x \to \infty} f(x)g(x) = \lim_{x \to \infty} x = \infty. \]

In Case 1, \( g(x) \) behaves like countable infinity, while \( g(x) \) is like uncountable infinity in Case 2.

It is now clear that countable infinity and uncountable infinity should be treated separately.

## 4.2 Countably Infinite Dimensions

If the Hilbert space is of countably infinite dimensions, we have a complete orthonormal basis

\[ \{ \lvert e_1 \rangle, \lvert e_2 \rangle, \ldots, \lvert e_n \rangle, \ldots \} \text{ or } \{ \lvert e_k \rangle \}; k = 1, 2, \ldots, \infty, \tag{4.4} \]

satisfying

\[ \langle e_i | e_j \rangle = \delta_{ij} \] (orthornormality) \tag{4.5} 

and

\[ \sum_{i=1}^{\infty} |e_i \rangle \langle e_i| = I \] (completeness). \tag{4.6} 

In this case, aside from the infinitely many components for vectors and infinitely many elements matrices carry, everything is the same as for finite dimensional Hilbert spaces.
4.3 Uncountably Infinite Dimensions

Let us consider the set of position eigenkets \{ |x \rangle \}, each labeled by the position eigenvalue x. As x is an observable and X is a Hermitian operator, \{ |x \rangle \} forms a basis. In particular, it is an orthonormal basis if the set \{ |x \rangle \} satisfies

\[ \langle x | x' \rangle = \delta(x - x') \quad \text{(orthonormality)} \]  

(4.7)

and

\[ \int^{-\infty}_{\infty} |x\rangle \langle x| \, dx = I \quad \text{(completeness)}; \]

(4.8)

where the right-hand side of (4.7) is a real-valued function called the Dirac delta function with the following definition and properties. Delta function is necessary whenever the basis kets are labeled by a continuous index including the position variable x.

Heuristic Description

Qualitatively, the delta function can be described by

\[ \delta(x) = \begin{cases} \infty & x = 0 \\ 0 & x \neq 0 \end{cases} \]  

(4.9)

and

\[ \int^{-\infty}_{\infty} \delta(x) \, dx = 1. \]  

(4.10)

Needless to say, this is an “extended” function, and no usual function has these properties. Rather than regarding the delta function as a function as we know them, we should think of it as a device to make quantum mechanics as well as other theories of physics work in a consistent fashion.

Properties of the Dirac Delta Function

Let us look at the properties of \( \delta(x - x') \) found in (4.7).

1. As before, we have

\[ \delta(x - x') = 0 \quad \text{if } x \neq x' \]
and
\[ \int_{-\infty}^{+\infty} \delta(x - x') \, dx' = 1. \] (4.11)

2. The delta function is sometimes called the sampling function as it samples the value of the function \( f(x') \) at one point \( x \); namely,
\[ \int_{-\infty}^{+\infty} \delta(x - x') f(x') \, dx' = f(x). \] (4.12)

3. The delta function is even as shown below.
\[ \delta(x - x') = \langle x | x' \rangle = \langle x' | x \rangle^* = \delta(x' - x) \] (4.13)

The last equality holds because the delta function is a real-valued function as stated on p.94.

Now, let \( f(x) \) be a differentiable function. Then,
\[ \frac{d}{dx} f(x - x') = \lim_{h \to 0} \frac{f(x - x') - f((x - h) - x')}{h} \] (4.14)

and
\[ \frac{d}{dx'} f(x - x') = \lim_{h \to 0} \frac{f(x - (x' + h)) - f(x - x')}{h} \]
\[ = \lim_{h \to 0} \frac{f(x - h - x') - f(x - x')}{h} \]
\[ = - \lim_{h \to 0} \frac{f(x - x') - f((x - h) - x')}{h} \]
\[ = - \frac{d}{dx} f(x - x') \] (4.15)

imply the following relation for the “formal” derivatives of the delta function.\(^a\)
\[ \delta'(x - x') = \frac{d}{dx} \delta(x - x') = - \frac{d}{dx'} \delta(x - x') \] (4.16)

4. We have \( \delta'(x - x') = \delta(x - x') \frac{d}{dx} \), and more generally,
\[ \frac{d^n}{dx^n} \delta(x - x') = \delta(x - x') \frac{d^n}{dx'} \] (4.17)
Therefore,
\[
\int \delta'(x - x') f(x') \, dx' = \int \delta(x - x') \frac{d}{dx'} f(x') \, dx' = \int \delta(x - x') f'(x') \, dx' = f'(x),
\]
and more generally,
\[
\int \delta^n(x - x') f(x') \, dx' = \int \delta(x - x') \frac{d^n}{dx^n} f(x') \, dx' = \int \delta(x - x') f^n(x') \, dx' = f^n(x).
\]

Next, consider \(\delta(ax)\) for any real number \(a \neq 0\). If we let \(y = ax\), then \(\frac{dy}{dx} = a \implies dx = \frac{1}{a} dy\). If \(a > 0\), \(y : -\infty \rightarrow +\infty\) if \(a > 0\), and \(y : +\infty \rightarrow -\infty\) if \(a < 0\). So,
\[
\int_{-\infty}^{+\infty} f(x) \delta(ax) \, dx = \left( \int_{-\infty}^{+\infty} f\left(\frac{y}{a}\right) \delta(y) \, dy \right) \frac{1}{a} = \frac{1}{a} f\left(\frac{0}{a}\right) = \frac{1}{a} f(0) \quad \text{if} \quad a > 0
\]
\[\text{and}\]
\[
\int_{-\infty}^{+\infty} f(x) \delta(ax) \, dx = \left( \int_{-\infty}^{+\infty} f\left(\frac{y}{a}\right) \delta(y) \, dy \right) \frac{1}{a} = \left( \int_{-\infty}^{+\infty} f\left(\frac{y}{a}\right) \delta(y) \, dy \right) \frac{1}{-a} = \frac{1}{-a} f\left(\frac{0}{a}\right) = \frac{1}{-a} f(0) \quad \text{if} \quad a < 0.
\]

5. The following rescaling property holds for the delta function.
\[
\delta(ax) = \frac{\delta(x)}{|a|} \quad \text{for any} \quad a \neq 0
\]
4.4 Delta Function as a Limit of Gaussian Distribution

The Gaussian distribution \( g_\sigma(x - x') \) labeled by the standard deviation \( \sigma \) is given by

\[
g_\sigma(x - x') = \frac{1}{\sqrt{(\pi \sigma^2)^2}} \exp \left[ -\frac{(x - x')^2}{\sigma^2} \right]. \tag{4.23}
\]

The Gaussian function \( g_\sigma \) is symmetric about \( x = x' \), has width \( \sigma \), and its peak height is \( (\pi \sigma)^{-\frac{1}{2}} \) at \( x = x' \). Furthermore, the area under the curve is unity irrespective of the value of the standard deviation \( \sigma \) or variance \( \sigma^2 \). Hence, it is at least qualitatively obvious that \( g_\sigma \) provides a better and better approximation of the delta function as \( \sigma \) approaches zero.

4.5 Delta Function and Fourier Transform

For a function \( f(x) \), its Fourier transform is

\[
f(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-ikx} f(x) \, dx, \tag{4.24}
\]

while the inverse Fourier transform is

\[
f(x') = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{ikx'} f(k) \, dk. \tag{4.25}
\]

Substituting (4.24) into (4.25), we obtain the following relation.

\[
f(x') = \int_{-\infty}^{+\infty} \left( \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{ik(x' - x)} \, dk \right) f(x) \, dx \tag{4.26}
\]

Now, compare (4.26) with (4.12). Then, we can see

\[
\frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{ik(x'-x)} \, dk = \delta(x' - x) \tag{4.27}
\]
or
\[
\frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{ik(x-x')} \, dk = \delta(x-x').
\] (4.28)

This is another characterization of the delta function.

### 4.6 \( f \) as a Vector with Uncountably Many Components

In order to understand a function \( f \) and its value at \( x \), normally denoted by \( f(x) \), we will consider vectors in a finite dimensional space, a space of countably infinite dimension, and a space spanned by uncountably many basis vectors, in this order. Without loss of generality, we will consider orthonormal bases.

#### 4.6.1 Finite Dimensions

Consider a vector \( |v\rangle \) in \( \mathbb{V}^n \) (an \( n \)-dimensional vector space). If a particular orthonormal basis \( \mathcal{B} = \{|e_1\rangle, |e_2\rangle, \ldots, |e_n\rangle \} \) is chosen, we can express \( v \) as a column vector with respect to \( \mathcal{B} \).

\[
|v\rangle = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}_{\mathcal{B}}
\] (4.29)

We have
\[
|v\rangle = \sum_{i=1}^{n} v_i |e_i\rangle
\] (4.30)

and
\[
v_i = \langle e_i | v \rangle.
\] (4.31)

#### 4.6.2 Countably Infinite Dimensions

If the vector space has a countably infinite dimension, and an orthonormal basis \( \mathcal{B} = \{|e_1\rangle, |e_2\rangle, \ldots, |e_n\rangle, \ldots \} \), we can express \( |v\rangle \) as an infinitely long column vector
\[
|v\rangle = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \\ \vdots \end{bmatrix}
\]
vector with respect to $\mathcal{B}$.

$$|v\rangle = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \\ \vdots \\
\end{bmatrix}_B$$  \hspace{1cm} (4.32)

We have

$$|v\rangle = \sum_{i=1}^{\infty} v_i |e_i\rangle$$ \hspace{1cm} (4.33)

and

$$v_i = \langle e_i | v \rangle.$$ \hspace{1cm} (4.34)

### 4.6.3 Uncountably Infinite Dimensions

When the vector space has an uncountably infinite dimension and an orthonormal basis $\mathcal{B} = \{|e_x\rangle\}_{x \in \mathbb{R}}$ or $x \in (-\infty, +\infty)$; where the kets in the basis $\mathcal{B}$ are indexed by a continuous variable $x$, each vector $v$ has uncountably many components $\{v_x\}_{x \in \mathbb{R}}$ with a continuous index $x \in \mathbb{R}$, as opposed to a discrete index $i \in \mathbb{N}$. We can still regard $|v\rangle$ as a column vector.

$$|v\rangle = \begin{bmatrix} v_{-\infty} \\ \uparrow \\ \vdots \\ v_x \\ \vdots \\ \downarrow \\ v_{+\infty} \end{bmatrix}_B$$  \hspace{1cm} (4.35)

By abuse of notation, we can write

$$|v\rangle = \sum_{x \in \mathbb{R}} v_x |e_x\rangle$$ \hspace{1cm} (4.36)

and, as before,
\[ v_x = \langle e_x | \mathbf{v} \rangle. \] (4.37)

Let us switch to a more familiar notation. Instead of \(|e_x\rangle\) and \(|\mathbf{v}\rangle\), we write \(|x\rangle\) and \(|f\rangle\). Then, we have

\[
|f\rangle = \begin{bmatrix} f_{-\infty} \\ \uparrow \\ \vdots \\ f_x \\ \vdots \\ \downarrow \\ f_{+\infty} \end{bmatrix}_B,
\] (4.38)

\[
|f\rangle = \sum_{x \in \mathbb{R}} f_x |x\rangle, \tag{4.39}
\]

and

\[
f_x = \langle x | f \rangle. \tag{4.40}
\]

Now let us write \(f(x)\) instead of \(f_x\) to get

\[
|f\rangle = \sum_{x \in \mathbb{R}} f(x) |x\rangle, \tag{4.41}
\]

and

\[
f(x) = \langle x | f \rangle. \tag{4.42}
\]

Notationally more appropriate expression for \(|f\rangle\) than (4.41) is obtained using the completeness relation (4.8).

\[
|f\rangle = I |f\rangle = \int_{-\infty}^{+\infty} |x\rangle \langle x|f \rangle \, dx = \int_{-\infty}^{+\infty} f(x) |x\rangle \, dx \tag{4.43}
\]

In this view, a function \(|f\rangle\) is a vector which is a linear combination of orthonormal basis vectors \(|x\rangle\)\(x \in \mathbb{R}\), and its coefficients, components of the column vector with
respect to the basis \{ |x\rangle \}_{x \in \mathbb{R}} , are our familiar \( f(x) \). We can rewrite (4.38) as follows.

\[
|f\rangle = \begin{bmatrix} f(-\infty) \\ \uparrow \\ \vdots \\ f(x) \\ \vdots \\ \downarrow \\ f(\infty) \end{bmatrix}_{x \in \mathbb{R}} ^{(4.44)}
\]

Finally, you can see that the inner product is the one defined for \( L^2 \)-functions in Section 2.3. All we need to do is to use the completeness property (4.8).

\[
\langle f | g \rangle = \langle f | I | g \rangle = \left< f \left| \int_{-\infty}^{+\infty} |x\rangle \langle x| \right| g \right> = \int_{-\infty}^{+\infty} \langle f | x \rangle \langle x| g \rangle \, dx \\
= \int_{-\infty}^{+\infty} \langle f | x \rangle \langle x| g \rangle \, dx = \int_{-\infty}^{+\infty} f(x)^* g(x) \, dx ^{(4.45)}
\]

4.7 Hermiticity of the Momentum Operator

Recall that the momentum operator \( P \) in one dimension is given by

\[
P = -i\hbar \frac{\partial}{\partial x} = -i\hbar \frac{d}{dx}. ^{(4.46)}
\]

This operator acts on differentiable \( L^2 \) functions \{ f(x) \} and has to be Hermitian as the momentum \( p \) is a physical observable. Let us verify this. We have the following in our notation.

\[
P |f\rangle = \left< -i\hbar \frac{df}{dx} \right>
\]

\[
\langle x | P | f\rangle = \left< x \left| -i\hbar \frac{d}{dx} \right| f \right> = \left< x \left| -i\hbar \frac{df}{dx} \right> = -i\hbar \frac{df}{dx}(x) = -i\hbar \frac{df(x)}{dx} ^{(4.48)}
\]

4.7.1 Checking \([ P_{ij} ] = [ P^*_{ji} ]\)

In this section, we will check if \( P \) is represented by a matrix which is equal to its transpose conjugate. Because our basis consists of uncountably many orthonormal
vectors \{ |x\rangle \}_{x \in \mathbb{R}}

, it is actually more appropriate to write \[ P_{xx'} = [P^*_{x'}] \]

; where \[ P_{xx'} = \langle x | P | x' \rangle \]. Using the completeness property (4.8), we can rewrite (4.48).

\[
\langle x | P | f \rangle = \langle x | P I | f \rangle = \left( x | P \right| \int_a^b |x'| \langle x' | d x' \left| f \right) \\
= \int_a^b \langle x | P | x' \rangle \langle x' | f \rangle \ dx' = \int_a^b \langle x | P | x' \rangle f(x') \ dx' = -i\hbar \frac{df(x)}{dx} \tag{4.49}
\]

Note that we used \( a \) and \( b \) as our limits of integration because the domain of the wavefunction is not always \(( -\infty, +\infty )\).

Now compare (4.18) and (4.49) to obtain

\[
\langle x | P | x' \rangle = -i\hbar \delta'(x - x') = -i\hbar \delta(x - x') \frac{d}{dx'} . \tag{4.50}
\]

As shown in the footnote on p.95, the first derivative of the real-valued \( \delta \)-function \( \delta'(y) \) is an odd function. Hence,

\[
P^*_{x'x} = \langle x' | P | x \rangle^* = (-i\hbar \delta'(x' - x))^* = (-i\hbar)^* (\delta'(x' - x))^* = i\hbar \delta'(x - x') = \delta x' - x = -\hbar \delta'(x - x') = P_{xx'} . \tag{4.51}
\]

Therefore, the matrix representation of \( P = -i\hbar \frac{d}{dx} \), denoted by \( [P_{xx'}] \), is equal to its transpose conjugate. This was a necessary and sufficient condition for an operator to be Hermitian if we have a finite dimensional Hilbert space (See Definition 2.26). However, this condition alone is not sufficient when the dimension is uncountably infinite.

### 4.7.2 Hermiticity Condition in Uncountably Infinite Dimensions

We will go back to the definition of Hermiticity. Definition 2.26 states that an operator \( \Omega \) is Hermitian if and only if

\[
\Omega = \Omega^\dagger , \tag{4.52}
\]

which is equivalent to

\[
\langle \Omega v \ | \ w \rangle = \langle v \ | \ Omega w \rangle \tag{4.53}
\]
for all \( \mathbf{v} \) and \( \mathbf{w} \). As we are interested in \( P = -i\hbar \frac{d}{dx} \), we would like to see if

\[
\langle -i\hbar \frac{df}{dx} \mid g \rangle = \langle f \mid -i\hbar \frac{dg}{dx} \rangle \iff i\hbar \langle df \mid g \rangle = -i\hbar \langle f \mid dg \rangle \\
\iff \langle \frac{df}{dx} \mid g \rangle = -\langle f \mid \frac{dg}{dx} \rangle.
\]  

(4.54)

Let us first work on the left-hand side. Completeness condition and integration by parts give us the following.

\[
\langle \frac{df}{dx} \mid g \rangle = \langle \frac{df}{dx} \mid I \rangle \langle x \rangle dx = \int_a^b \langle \frac{df(x)}{dx} \rangle^* g(x) dx = \int_a^b \frac{df^*(x)}{dx} g(x) dx \\
= f^*(x)g(x)|_{a}^{b} - \int_a^b f^*(x)g'(x) dx.
\]  

(4.55)

On the other hand, the right-hand side is

\[
-\langle f \mid \frac{dg}{dx} \rangle = -\langle f \mid \int_a^b \langle x \rangle dx \mid \frac{dg}{dx} \rangle = -\int_a^b \langle f \mid \langle x \rangle \frac{dg}{dx} dx \\
= -\int_a^b f^*(x)g'(x) dx.
\]  

(4.56)

The expressions (4.55) and (4.56) are equal if and only if

\[
f^*(x)g(x)|_{a}^{b} = 0.
\]  

(4.57)

Needless to say, this is not satisfied by all functions \( f \) and \( g \). However, there are a couple of notable cases where (4.57) holds.

**Single-Valuedness** : When the spherical coordinates \((r, \theta, \phi)\) are used as in Chapter 10, \((r, \theta, 0)\) and \((r, \theta, 2\pi)\) are the same point in the three-dimensional space. Therefore, we require \( \Psi(r, \theta, 0, t) = \Psi(r, \theta, 2\pi, t) \) due to the single-valuedness condition, Condition 2, on p.128. In this case, \( a = 0 \) and \( b = 2\pi \), and we indeed have

\[
f^*(x)g(x)|_{a}^{b} = f^*(\phi)g(\phi)|_{0}^{2\pi} = \Psi^*(r, \theta, \phi, t)\Psi(r, \theta, \phi, t)|_{0}^{2\pi} = 0
\]  

(4.58)
with respect to the $\varphi$-integral. One class of functions that satisfy (4.57) consists of periodic functions. For example, if $a = 0$ and $b = 2\pi$, trigonometric functions such as $\sin x$ and $\cos x$ and their complex counterparts including $e^{ix}$ satisfy this condition.

**Functions Vanishing at Infinity**: This is when $a = -\infty$ and $b = +\infty$, and we need

$$f^*(x)g(x)|_{-\infty}^{+\infty} = 0.$$  \hspace{1cm} (4.59)

In particular, this equality holds for $L^2$-functions, described on p.30 of Section 2.3, as $f \in L^2(-\infty, +\infty)$ implies $\lim_{|x| \to \infty} f(x) = 0$.

### 4.7.3 The Eigenvalue Problem of the Momentum Operator $P$

The eigenvalue problem for $P$ in an arbitrary basis is

$$P |\lambda\rangle = \lambda |\lambda\rangle.$$  \hspace{1cm} (4.60)

Taking the projection on $\langle x|$

$$\langle x|P|\lambda\rangle = \lambda \langle x|\lambda\rangle.$$  \hspace{1cm} (4.61)

Using completeness,

$$\langle x|P|\lambda\rangle = \langle x|P|\lambda\rangle = \langle x|P|\lambda\rangle = \langle x'|P|\lambda\rangle = \langle x|P|\lambda\rangle = \langle x'|\lambda\rangle dx' = \lambda \langle x|\lambda\rangle.$$  \hspace{1cm} (4.62)

Now, from (4.50), we have

$$\langle x|P|\lambda\rangle = -i\hbar\delta(x - x') \frac{d}{dx'}.$$  \hspace{1cm} (4.63)

Hence,

$$\langle x|P|\lambda\rangle = \int \langle x|P|x'|\rangle \langle x'|\lambda\rangle dx' = \int \langle x|P|x'|\rangle \langle x'|\lambda\rangle dx'$$
\[ = -i\hbar \frac{d}{dx} \langle x|\lambda \rangle \quad (4.64) \]

If we denote \( \langle x|\lambda \rangle \) by \( f_\lambda(x) \), we have
\[ -i\hbar \frac{d}{dx} f_\lambda(x) = \lambda f_\lambda(x). \quad (4.65) \]

Had we started with the \( X \)-basis made up of the eigenvectors of the position operator \( X \), we would have obtained \( (4.65) \) immediately as \( P = -i\hbar \frac{d}{dx} \) in this basis.

Solving the ordinary differential equation \( (4.65) \), we get
\[ f_\lambda(x) = Ce^{i\lambda x/\hbar}; \quad (4.66) \]

where \( C \) is an arbitrary constant. It is now clear that \( P \) takes any real number \( \lambda \) as its eigenvalue with the corresponding eigenfunction \( Ce^{i\lambda x} \) in the \( X \)-basis. In order to normalize the eigenfunctions, we let \( C = \frac{1}{\sqrt{2\pi \hbar}} \). Then,
\[ \langle x|\lambda \rangle = f_\lambda(x) = \frac{1}{\sqrt{2\pi \hbar}} e^{i\lambda x/\hbar}, \quad (4.67) \]

and, due to \( (4.28) \) and \( (4.22) \),
\[ \langle \lambda|\lambda' \rangle = \int_{-\infty}^{+\infty} \langle \lambda|x \rangle \langle x|\lambda' \rangle \, dx = \frac{1}{\sqrt{2\pi \hbar}} \frac{1}{\sqrt{2\pi \hbar}} \int_{-\infty}^{+\infty} e^{-i\lambda x/\hbar} e^{i\lambda' x/\hbar} \, dx \]
\[ = \frac{1}{2\pi \hbar} \int_{-\infty}^{+\infty} e^{i\lambda' (\lambda - \lambda')} x \, dx = \frac{1}{\hbar} \left( \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{i\lambda' (\lambda - \lambda')} x \, dx \right) \]
\[ = \frac{1}{\hbar} \delta \left( \frac{1}{\hbar} (\lambda - \lambda') \right) = \frac{1}{\hbar} \frac{\delta(\lambda - \lambda')}{|1/\hbar|} = \delta(\lambda - \lambda'). \quad (4.68) \]

So, \( \{ |\lambda\rangle \}_{\lambda \in \mathbb{R}} \) forms an orthonormal basis.

### 4.8 Relations Between \( X \) and \( P \)

#### 4.8.1 The Fourier Transform Connecting \( X \) and \( P \)

Given a physical system \( \mathcal{S} \), there is an associated Hilbert space \( \mathcal{H} \) according to Postulate 1 on p.72. This Hilbert space is spanned either by the eigenkets of \( P \) (the \( P \) basis) or the eigenkets of \( X \) (the \( X \) basis) as both \( X \) and \( P \) are Hermitian
operators. Given a ket $|f\rangle$, it can be expanded either in the $X$ basis $\{|x\rangle\}$ or in the $P$ basis $\{|p\rangle\}$.

Note here that we switched the notation for the $P$ basis from $\{|\lambda\rangle\}$ to $\{|p\rangle\}$ for clarity.

Recall that we have

$$|f\rangle = \sum_{x \in \mathbb{R}} \langle x|f \rangle |x\rangle \quad \text{and} \quad |f\rangle = \sum_{p \in \mathbb{R}} \langle p|f \rangle |p\rangle.$$  \quad (4.69)

So, $f(x) = \langle x|f \rangle$ is the “$x$-th” coefficient/component of $|f\rangle$, and $f(p) = \langle p|f \rangle$ is the “$p$-th” coefficient/component of $|f\rangle$. They are coefficients in view of the expansion (4.69), but they are also components of the column vector $|f\rangle$. Let us compare $f(p)$ and $f(x)$ using (4.67).

$$f(p) = \langle p|f \rangle = \int_{-\infty}^{+\infty} \langle p|x \rangle \langle x|f \rangle \, dx = \int_{-\infty}^{+\infty} \langle x|p \rangle^* \langle x|f \rangle \, dx$$

$$\Rightarrow f(p) = \frac{1}{\sqrt{2\pi \hbar}} \int_{-\infty}^{+\infty} e^{-ipx/\hbar} f(x) \, dx \quad (4.70)$$

$$f(x) = \langle x|f \rangle = \int_{-\infty}^{+\infty} \langle x|p \rangle \langle p|f \rangle \, dp$$

$$\Rightarrow f(x) = \frac{1}{\sqrt{2\pi \hbar}} \int_{-\infty}^{+\infty} e^{ipx/\hbar} f(p) \, dp \quad (4.71)$$

These are nothing but the Fourier transform and the Fourier inverse transform. Therefore, the expressions of a function $|f\rangle$ in this Hilbert space in terms of the complete $X$ basis and complete $P$ basis are related through the familiar Fourier transform.

You may note that the more familiar form of the Fourier transform can be obtained if we set $\hbar = 1/\sqrt{2\pi}$. Indeed, if we consider the operator $K = P/\hbar$ and its orthonormal eigenbasis $\{|k\rangle\}_{k \in \mathbb{R}}$, which satisfies $\langle x|k \rangle = \frac{1}{\sqrt{2\pi}} e^{ikx}$ as opposed to $\langle x|p \rangle = \frac{1}{\sqrt{2\pi \hbar}} e^{ipx/\hbar}$, $f(k)$ and $f(x)$ are related as follows.

$$f(k) = \langle k|f \rangle = \int_{-\infty}^{+\infty} \langle k|x \rangle \langle x|f \rangle \, dx = \int_{-\infty}^{+\infty} \langle x|k \rangle^* \langle x|f \rangle \, dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-ikx} f(x) \, dx \quad (4.72)$$
\[ f(x) = \langle x|f \rangle = \int_{-\infty}^{+\infty} \langle x|k \rangle \langle k|f \rangle \, dk = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{ikx} f(k) \, dk \quad (4.73) \]

As advertised, these are the Fourier transform and its inverse in their more familiar forms.

### 4.8.2 \(X\) and \(P\) in the \(X\) Basis

We already know that the position operator \(X\) in the \(X\) basis is \(x\) or \(M_x\), which is a scalar multiplication. We also know that the momentum operator \(P\) in the \(X\) basis is \(-\imath \frac{d}{dx}\). We can check the action of \(X\) in the framework of this chapter as follows.

\[ X |x\rangle = x |x\rangle \implies \langle x'|X|x\rangle = \langle x'|x\rangle = x \langle x'|x\rangle = x\delta(x' - x) \quad (4.74) \]

\[ \implies \langle x|X|f\rangle = \langle x|XI|f\rangle = \langle x|X \left( \int |x'| \langle x'| \, dx' \right) |f\rangle = \int \langle x|X|x'\rangle \langle x'|f\rangle \, dx' = \int x'\delta(x - x')f(x') \, dx' = xf(x) \quad (4.75) \]

In order to express this action in a basis-independent manner, we can adopt Shankar’s notation [Shankar, 1980, p.74].

### 4.8.3 Basis-Independent Descriptions in Reference to the \(X\) Basis

Recall that \(|f\rangle\) is a basis-independent expression for a vector (function) in our Hilbert space. One way to specify \(f\) is by giving \(\langle x|f \rangle = f(x)\) for the orthonormal basis \(\{|x\rangle\}_{x\in\mathbb{R}}\). Differently put, we cannot really specify the nature of \(f\) unless we go to some basis such as the position basis \(\{x\}_{x\in\mathbb{R}}\). With this specification, it is now possible to express \(f\) as a basis-independent ket in reference to the position orthonormal basis \(\{|x\rangle\}_{x\in\mathbb{R}}\). This potentially confusing strategy reduces to writing the basis-independent ket \(|f\rangle\) as \(|f(x)\rangle\). What does this mean? It means the following.

1. \(|f\rangle\) itself is basis-independent.
2. However, we know \(|f\rangle\) is fully characterized by the coefficients \(\{f(x)\}_{x\in\mathbb{R}}\).
3. So, we write \(|f(x)\rangle\) for an abstract basis-independent vector whose coefficients with respect to \(\{|x\rangle\}_{x\in\mathbb{R}}\) are \(\{f(x)\}_{x\in\mathbb{R}}\).
For example, from (4.75), the coefficients of \( X \left| f \right) \) with respect to \( \{ \left| x \right) \}_{x \in \mathbb{R}} \), namely \( \{ \left< x \right| X \left| f \right) \}_{x \in \mathbb{R}} \), are \( \{ x f(x) \}_{x \in \mathbb{R}} \). Therefore, the action of \( X \) can be expressed in a basis-independent manner, but in reference to the orthonormal position basis \( \{ \left| x \right) \}_{x \in \mathbb{R}} \), as below.

\[
X \left| f(x) \right) = \left| x f(x) \right)
\]

(4.76)

Similarly, a basis-independent description of the momentum operator \( P \) in reference to \( \{ \left| x \right) \}_{x \in \mathbb{R}} \) is

\[
P \left| f(x) \right) = \left| -i\hbar \frac{df(x)}{dx} \right).
\]

(4.77)

### 4.8.4 \( X \) and \( P \) in the \( P \) Basis

From (4.60), we have

\[
P \left| p \right) = p \left| p \right)
\]

(4.78)

in the \( P \) basis. Hence, the matrix element \( P_{pp'} = \left< p \right| P \left| p' \right) \) is given by

\[
\left< p \left| P \left| p' \right) = \left< p \right| p' \right) = p \left< p \right| p' \right) = p' \delta(p - p').
\]

(4.79)

It remains to show how \( X \) operates in this basis. The matrix element \( X_{pp'} = \left< p \right| X \left| p' \right) \) can be obtained as follows.

First, note that

\[
i\hbar \frac{d}{dp} \left( e^{-ipx/\hbar} \right) = i\hbar \left( -ix/\hbar \right) e^{-ipx/\hbar} = xe^{-ipx/\hbar}.
\]

(4.80)

Using (4.67) and (4.80), we get

\[
\left< p \right| X \left| p' \right) = \left< p \right| XI \left| p' \right) = \left< p \right| X \left( \int \left| x \right> \left< x \right) \left| dx \right) \left| p' \right) = \int \left< p \right| X \left| x \right) \left< x \right| p' \right) \left| dx \right.
\]

\[
= \int \left< p \right| x \right) \left< x \right| p' \right) \left| dx \right) = \int x \left< p \right| x \right) \left< x \right| p' \right) \left| dx \right) = \int x \left< x \right| p' \right) \left< x \right| p \right) \left| dx \right)
\]

\[
= \int x \left( \frac{1}{\sqrt{2\pi\hbar}} e^{ipx/\hbar} \right) \left( \frac{1}{\sqrt{2\pi\hbar}} e^{ip'x/\hbar} \right) \left| dx \right) = \frac{1}{2\pi\hbar} \int xe^{-ipx/\hbar} e^{ip'x/\hbar} \left| dx \right)
\]
\[
\frac{1}{2\pi\hbar} \int i\hbar \frac{d}{dp} \left( e^{-ipx/\hbar} e^{ip'y/\hbar} \right) dx = i \frac{d}{dp} \left( \frac{1}{2\pi} \int e^{i(p'-p)x/\hbar} dx \right) \\
= i\hbar \frac{d}{dp} \delta(p - p') = i\hbar \delta'(p - p').
\] (4.81)

The \( \hbar \) in front of the delta function arises in the following manner. If we let \( y = x/\hbar \), \( dx = \hbar dy \), and \( y : -\infty \rightarrow +\infty \) as \( x : -\infty \rightarrow +\infty \). So,

\[
\frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{i(p' - p)x/\hbar} dx = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{i(p' - p)y} \hbar dy = \hbar \left( \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{i(p' - p)y} dy \right)
\]

\[
= \hbar \delta(p' - p) = \hbar \delta(p - p').
\] (4.82)

You may realize that this can be modified to generate an alternative proof of Property 5 or (4.22) on p.96.

Hence, the action of \( X \) in the \( K \) basis is described by

\[
X f(k) = i\hbar \frac{df(k)}{dk}.
\] (4.83)

As before, the basis independent form of this with our notational convention is

\[
X \left| f(k) \right\rangle = \left| i\hbar \frac{d}{dk} f(k) \right\rangle.
\] (4.84)

Let us summarize what we have shown in Sections 4.8.2 and 4.8.4.

- In the \( X \) basis:

\[
X \leftrightarrow x \text{ or } M_x
\] (4.85)

\[
P \leftrightarrow -i\hbar \frac{d}{dx}
\] (4.86)

- In the \( P \) basis:

\[
X \leftrightarrow i\hbar \frac{d}{dp}
\] (4.87)

\[
P \leftrightarrow p \text{ or } M_p
\] (4.88)
4.8.5 The Commutator \([X, P]\)

Let \([X, P]_x\) and \([X, P]_p\) be commutators of \(X\) and \(P\) in the \(X\) and \(P\) bases, respectively. Then,

\[
[X, P]_x f(x) = \left[ M_x \left( -i\hbar \frac{d}{dx} \right) - \left( -i\hbar \frac{d}{dx} M_x \right) \right] f(x)
\]

\[
= -i\hbar x \frac{df(x)}{dx} + i\hbar \frac{d}{dx}(xf(x)) = -i\hbar x \frac{df(x)}{dx} + i\hbar \left( f(x) + x \frac{df(x)}{dx} \right)
\]

\[
= i\hbar f(x) \implies [X, P]_x = i\hbar I \quad (4.89)
\]

and

\[
[X, P]_p f(p) = \left[ \left( i\hbar \frac{d}{dp} \right) M_p - M_p \left( i\hbar \frac{d}{dp} \right) \right] f(p)
\]

\[
= i\hbar \frac{d}{dp}(pf(p)) - i\hbar \frac{df(p)}{dp} = i\hbar \left( f(p) + p \frac{df(p)}{dp} \right) - i\hbar \frac{df(p)}{dp}
\]

\[
= i\hbar f(p) \implies [X, P]_p = i\hbar I. \quad (4.90)
\]

Because the actual expression for the identity operator \(I\) is basis-independent, we do not write \(I_X\) or \(I_P\). We conclude that

\[
[X, P] = i\hbar I \quad (4.91)
\]

in either basis.

---

\(^a\)As the delta function is even, its “derivative” is an odd function. Generally speaking, let \(f(x)\) be a differentiable even function and \(y = -x\), then, for an arbitrary value \(x_0\),

\[
\left. \frac{df(x)}{dx} \right|_{x=x_0} = \left. \frac{df(-x)}{dx} \right|_{x=x_0} = -\left. \frac{df(-x)}{dx} \right|_{x=x_0} = \left. \frac{df(y)}{dy} \right|_{y=-x_0} = \left. \frac{df(x)}{dx} \right|_{x=-x_0}.
\]

So, we have \(f'(x_0) = -f'(-x_0)\) for any \(x_0\).

\(^b\)The “width” is defined as the distance between the two points of inflection of \(g_\sigma\).

\(^c\)We will see in Chapter 10 that we can separate the four variables \((r, \theta, \phi, t)\) and write \(\Psi\) as a product of functions of \(r, \theta, \phi, \) and \(t\); namely \(\Psi(r, \theta, \phi, t) = R(r) \Theta(\theta) \Phi(\phi) T(t)\).

\(^d\)This is the reason why we left \(2\pi \hbar\) as it was even though \(2\pi = \frac{2\pi \hbar}{\hbar} = \hbar\).
Chapter 5
Schrödinger’s Theory of Quantum Mechanics

In 1925 an Austrian physicist Erwin Rudolf Josef Alexander Schrödinger proposed the Schrödinger Equation which is a differential equation whose solutions are wavefunctions. He chose a differential equation to describe *new physics* since a differential equation was the most common type of equation known to the physicists with a function for a solution. In Chapter 3, the postulates of quantum mechanics were presented. However, these postulates were put together in a post hoc fashion. In this chapter, we will learn how quantum mechanical framework was initially constructed, drawing on the analogy to classical mechanics and incorporating the emerging ideas of wave-particle duality.

In the language of Chapter 3, and in retrospect, Schrödinger’s work was an attempt to associate a Hilbert space of $L^2$-functions, now known as wavefunctions, to a given physical system. The functions are called wavefunctions as they are supposed to reflect the wave nature of particles including superposition and interference.

5.1 How was it derived?

In his derivation, he was guided by the following.

1. The classical traveling wave: a simple sinusoidal traveling wave

   \[ \Psi(x, t) = \sin 2\pi \left( \frac{x}{\lambda} - \nu t \right) \]  

   \[ (5.1) \]
2. The de Broglie-Einstein postulates

\[ \lambda = \frac{h}{p}, \quad \nu = \frac{E}{\hbar} \]  

(5.2)

3. The Newtonian energy equation

\[ E = \frac{p^2}{2m} + V \]  

(5.3)

4. The equation should be linear in terms of its solutions.

If

\[ \Psi_1(x, t), \quad \Psi_2(x, t) \]  

(5.4)

are two solutions of the Schrödinger Equation, then any linear combination of the two

\[ \Psi(x, t) = c_1 \Psi_1(x, t) + c_2 \Psi_2(x, t) \]  

(5.5)

is also a solution; where \( c_1 \) and \( c_2 \) are arbitrary constants.

\[ \implies \text{This makes superposition and interference possible.} \]

Here is what we want.

- A differential equation, later named the Schrödinger Equation, which looks like the energy equation.

- Wave functions, which are the solutions of the Schrödinger Equation and behave like classical waves, including interference and superposition.

Let us first consider the energy equation and observe the following.

---

\(^1\text{You may be more familiar with }\Psi(x, t) = \sin(\xi x - \omega t). \text{ The equivalence of } \sin(2\pi (\frac{\xi}{\lambda} - \nu t)) \text{ and } \sin(kx - \omega t) \text{ is a direct consequence of the definition of the wavenumber } k \text{ and the relation between the frequency } \nu \text{ and the angular frequency } \omega; \text{ namely, } k = \frac{2\pi}{\lambda} \text{ and } \omega = 2\pi\nu. \text{ This is as shown on p.113.}\)
5.1. \textit{HOW WAS IT DERIVED?}

Let $k = \frac{2\pi}{\lambda}$ and $\omega = 2\pi \nu$. Then, we have

$$\Psi(x, t) = \sin(kx - \omega t). \quad (5.6)$$

On the other hand, the energy equation is

$$E = \frac{P^2}{2m} + V. \quad (5.7)$$

Plug $P = \frac{h}{\lambda}$ and $E = h \nu$ into (5.7) to obtain

$$\frac{h^2}{2m\lambda^2} + V(x, t) = h \nu. \quad (5.8)$$

Recall that $k = \frac{2\pi}{\lambda}$ and $\omega = 2\pi \nu$. So, we have the following relations.

$$\left(\frac{1}{\lambda}\right)^2 = \left(\frac{k}{2\pi}\right)^2 \quad \nu = \frac{\omega}{2\pi} \quad (5.9)$$

We now substitute these back into the energy equation (5.7).

$$E = \frac{P^2}{2m} + V \implies \frac{h^2}{2m\lambda^2} + V(x, t) = h \nu \implies \frac{h^2k^2}{(2\pi)^2} \cdot \frac{1}{2m} + V(x, t) = h \left(\frac{\omega}{2\pi}\right)$$

$$\implies \frac{h^2}{2\pi} k^2 \cdot \frac{1}{2m} + V(x, t) = \left(\frac{h}{2\pi}\right) \omega \quad (5.10)$$

Now we introduce a new notation.

$$\hbar = \frac{h}{2\pi} \quad (5.11)$$

When you read it aloud, it is pronounced either $h$ \textit{bar} or $h$ \textit{cross}, with $h$ \textit{bar} seeming more dominant these days. With this notation, (5.10) becomes

$$\hbar^2 k^2 \cdot \frac{1}{2m} + V(x, t) = \hbar \omega. \quad (5.12)$$

We will next look at the classical sinusoidal traveling wave given by

$$\Psi(x, t) = \sin(kx - \omega t). \quad (5.13)$$

\footnote{This $k$ is called the wavenumber or angular/circular wavenumber to make a clear distinction from another wavenumber $1/\lambda$, which is also called the spectroscopic wavenumber. By definition, $k$ is the number of wavelengths per $2\pi$ units of distance.}
Here are three of its partial derivatives which are of interest to us.

\[ \frac{\partial \Psi}{\partial x} = k \cos(kx - \omega t) \quad (5.14) \]
\[ \frac{\partial^2 \Psi}{\partial x^2} = -k^2 \sin(kx - \omega t) \quad (5.15) \]
\[ \frac{\partial \Psi}{\partial t} = -\omega \cos(kx - \omega t) \quad (5.16) \]

From these, we arrive at the implications below.

\[ \begin{aligned}
\frac{\partial^2}{\partial x^2} & \text{ gets us } k^2, \\
\frac{\partial}{\partial t} & \text{ gets us } \omega. 
\end{aligned} \quad (5.17) \]

Note that this argument is all qualitative and of story-telling type. So is the rest of the argument.

By abuse of notation, let \( \Psi(x, t) \) be the desired wave function which acts like the traveling wave.

As the energy equation, whose solution is \( \Psi(x, t) \), contains \( k^2 \) and \( \omega \), we consider the correspondences below.

\[ \frac{\hbar^2 k^2}{2m} \iff \frac{\partial^2}{\partial x^2} \Psi(x, t) \quad (5.18) \]
\[ \hbar \omega \iff \frac{\partial}{\partial t} \Psi(x, t) \quad (5.19) \]

We have used guiding assumptions (hypotheses) 1 through 3 to arrive at

\[ \alpha \frac{\partial^2}{\partial x^2} \Psi(x, t) + V(x, t) = \beta \frac{\partial \Psi(x, t)}{\partial t}. \quad (5.20) \]

Why \( \alpha \) and \( \beta \)? A very short answer is “because it works”. But, a slightly longer answer would be that our derivation is largely qualitative, and we need to incorporate some quantitative degrees of freedom in our trial formula. At any rate, whatever works!

How about the fourth assumption; the linearity assumption?
Let $\Psi_1$ and $\Psi_2$ be two solutions of the above equation. Then we have

$$\alpha \Psi_{1,xx} + V = \beta \Psi_{1,t} \quad \text{and} \quad \alpha \Psi_{2,xx} + V = \beta \Psi_{2,t}$$

$$\Rightarrow \alpha(\Psi_{1,xx} + \Psi_{2,xx}) + 2V = \beta(\Psi_{1,t} + \Psi_{2,t}).$$  \hspace{1cm} (5.21)

We are unfortunately off by the factor 2 in front of the potential $V$.

And one way around this problem is to modify the equation in the following way.

$$\alpha \frac{\partial^2}{\partial x^2} \Psi(x, t) + V(x, t)\Psi(x, t) = \beta \frac{\partial \Psi(x, t)}{\partial t}$$  \hspace{1cm} (5.22)

With a little more work, we can determine $\alpha$ and $\beta$. The easiest way is to consider $V = 0$ which gives us a free particle or a free traveling wave

$$\Psi(x, t) = e^{i(kx-\omega t)}. $$  \hspace{1cm} (5.23)

Substituting $V = 0$ and (5.23) into (5.22),

$$\alpha \frac{\partial^2}{\partial x^2} e^{i(kx-\omega t)} = \beta \frac{\partial}{\partial t} e^{i(kx-\omega t)} \quad \Rightarrow \quad -\alpha k^2 e^{i(kx-\omega t)} = -i\omega \beta e^{i(kx-\omega t)}$$

$$\Rightarrow \quad -\alpha k^2 = -i\omega \beta$$  \hspace{1cm} (5.24)

Comparing (5.24) with (5.12) when $V(x, t) = 0$, we get

$$-\alpha k^2 = -i\omega \beta \iff \frac{\hbar^2 k^2}{2m} = \hbar \omega.$$  \hspace{1cm} (5.25)

Therefore,

$$\alpha = \frac{-\hbar^2}{2m} \quad \text{and} \quad \beta = \frac{\hbar}{-i} = i\hbar.$$  \hspace{1cm} (5.26)

And our end product is

$$-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \Psi(x, t) + V(x, t)\Psi(x, t) = i\hbar \frac{\partial \Psi(x, t)}{\partial t}. $$  \hspace{1cm} (5.27)

This is known as the **Schrödinger Equation** or the **Time-Dependent Schrödinger Equation**. We often write it as below.

$$ \left[ -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x, t) \right] \Psi(x, t) = i\hbar \frac{\partial \Psi(x, t)}{\partial t}. $$  \hspace{1cm} (5.28)
The solution set \( \{ \Psi(x, t) \} \) represents the wave functions which are to be associated with the motion of a particle of mass \( m \) under the influence of the force \( F \) described by \( V(x, t) \); namely

\[
F(x, t) = -\frac{\partial V(x, t)}{\partial x}.
\]

Agreement with experiment? How do you check it? We first need a physical interpretation of \( \Psi \) to do so.

### 5.2 Born’s Interpretation of Wavefunctions

We now have

\[
-\frac{\hbar^2}{2m} \frac{\partial^2 \Psi(x, t)}{\partial x^2} + V(x, t)\Psi(x, t) = i\hbar \frac{\partial \Psi(x, t)}{\partial t}
\]

or

\[
\left[ -\frac{\hbar^2}{2m} \frac{\partial^2 \Psi(x, t)}{\partial x^2} + V(x, t) \right] \Psi(x, t) = i\hbar \frac{\partial \Psi(x, t)}{\partial t}.
\]

(5.30)

So, if \( V(x, t) \) is given, a wavefunction \( \Psi(x, t) \) can be obtained at least in principle. But, what is \( \Psi \)?

Since \( \Psi \) is a complex function, \( \Psi \) itself can not be a real physical wave. How can we extract physical reality from this?

Max Born realized that \( \Psi^* \Psi = |\Psi|^2 \) is real while \( \Psi \) itself is not and postulated that \( P(x, t) = \Psi(x, t)^* \Psi(x, t) \) can be regarded as a probability. He postulated that the probability \( P(x, t)dx \) of finding the particle at a coordinate between \( x \) and \( x + dx \) is equal to \( \Psi(x, t)^* \Psi(x, t)dx \).

However, there is a problem with this view as it is. It is too simplistic. To understand the nature of the problem simply note that \( 2\Psi(x, t) \) is also a solution if \( \Psi(x, t) \) is a solution of

\[
-\frac{\hbar^2}{2m} \frac{\partial^2 \Psi(x, t)}{\partial x^2} + V(x, t)\Psi(x, t) = i\hbar \frac{\partial \Psi(x, t)}{\partial t}.
\]

(5.31)

This means that the probability is not unique, which of course is not acceptable. Furthermore, the total probability is not necessarily one if we used an arbitrary
solution. Our necessary condition is

$$\int_{-\infty}^{+\infty} P(x,t)dx = 1.$$  \hspace{1cm} (5.32)

In other words, we need to find $\Psi$ such that

$$\int_{-\infty}^{+\infty} P(x,t)dx = \int_{-\infty}^{+\infty} \Psi^*(x,t)\Psi(x,t)dx = 1.$$  \hspace{1cm} (5.33)

This is the physically meaningful $\Psi$ called a normalized wavefunction. This process is called normalization. You should recognize that this normalized wavefunction is “the state” first mentioned in Postulate 1 on p.72 of Chapter 3.

## 5.3 Expectation Values

Postulate 7 on p.74, Chapter 3, gives

$$\bar{q} \text{ or } \langle q \rangle = \langle v_s | Q | v_s \rangle.$$  \hspace{1cm} (5.34)

For a Hilbert space of $L^2$-functions, we have

$$\langle q \rangle = \langle v_s | Q | v_s \rangle = \int v_s(x)^* Q v_s(x) dx;$$  \hspace{1cm} (5.35)

where the state vector $v_s(x)$ is an $L^2$-function of unit norm, i.e.

$$\int v_s(x)^* v_s(x) dx = 1.$$  \hspace{1cm} (5.36)

In this section, we will start from scratch and see how (5.35) is “derived”, so to speak, for the momentum $P$ and the total energy $E$.

Recall that for a die

$$\sum \text{value } \times \text{probability} = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = (1 + 2 + 3 + 4 + 5 + 6) \cdot \frac{1}{6}$$  \hspace{1cm} (5.37)

gives the expectation value.
Question: How do we calculate the “average” position of a particle?

Answer: By computing the expectation value of the position \( x \) denoted by \( \langle x \rangle \) or \( \bar{x} \).

Analogously to the die above, with \( x \) as the value, we have

\[
\langle x \rangle = \sum x \times \text{probability}.
\]  
(5.38)

The continuous version of the above is

\[
\langle x \rangle = \bar{x} = \int_{-\infty}^{+\infty} x \Psi^*(x,t)\Psi(x,t)dx = \int_{-\infty}^{+\infty} \Psi^*(x,t)x\Psi(x,t)dx.
\]  
(5.39)

How about \( P \) and \( E \)? We are tempted to write

\[
P = \int_{-\infty}^{+\infty} \Psi^*(x,t)P\Psi(x,t)dx
\]  
(5.40)

and

\[
E = \int_{-\infty}^{+\infty} \Psi^*(x,t)E\Psi(x,t)dx.
\]  
(5.41)

It turns out this is what we should indeed do.

Recall we are always to be guided by the classical wave. Instead of a sine wave, let us look at a particular solution of the Schrödinger Equation called a free particle solution. The free-ness of a free particle derives from the lack of any external force. In other words, \( V(x,t) = 0 \) for a free particle.

\[
\left[-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x,t)\right] \Psi(x,t) = i\hbar \frac{\partial \Psi(x,t)}{\partial t}
\]

\[
\implies -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \Psi(x,t) = i\hbar \frac{\partial \Psi(x,t)}{\partial t}
\]  
(5.42)

We will solve this in detail in the next chapter, but the answer is the classical traveling wave.

Previously, we used

\[
\Psi(x,t) = \sin(kx - \omega t).
\]  
(5.43)
5.3. EXPECTATION VALUES

Now we will use its complex version.

\[ \Psi(x, t) = \cos(kx - \omega t) + i \sin(kx - \omega t) = e^{i(kx - \omega t)} \]  \hspace{1cm} \text{(5.44)}

Taking the first partial derivative with respect to \( x \),

\[ \frac{\partial \Psi}{\partial x} = ik \Psi(x, t) = i \frac{P}{\hbar} \Psi(x, t); \]  \hspace{1cm} \text{(5.45)}

where the last equality follows from

\[ k = \frac{2\pi}{\lambda} = \frac{2\pi}{\hbar/p} = \frac{P}{\hbar/2\pi} = \frac{P}{\hbar}. \]  \hspace{1cm} \text{(5.46)}

Therefore,

\[ P[\Psi(x, t)] = -i\hbar \frac{\partial}{\partial x} [\Psi(x, t)]. \]  \hspace{1cm} \text{(5.47)}

Similarly,

\[ \frac{\partial \Psi(x, t)}{\partial t} = -i\omega \Psi(x, t) = -i \frac{E}{\hbar} \Psi(x, t); \]  \hspace{1cm} \text{(5.48)}

where the last equality follows from

\[ \omega = 2\pi \nu = 2\pi \frac{E}{\hbar} = \frac{E}{\hbar/2\pi} = \frac{E}{\hbar}. \]  \hspace{1cm} \text{(5.49)}

Therefore,

\[ E[\Psi(x, t)] = i\hbar \frac{\partial}{\partial t} [\Psi(x, t)]. \]  \hspace{1cm} \text{(5.50)}

On the other hand, let us compare

\[ \frac{P^2}{2m} + V(x, t) = E \]  \hspace{1cm} \text{(5.51)}

with

\[ \left[ -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x, t) \right] \Psi(x, t) = i\hbar \frac{\partial 
\Psi(x, t)}{\partial t}. \]  \hspace{1cm} \text{(5.52)}

term by term.
The first terms give us
\[
-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} = \frac{P^2}{2m}.
\]
(5.53)

Working formally,
\[
\sqrt{-\hbar^2} \sqrt{\frac{\partial^2}{\partial x^2}} = \sqrt{P^2} \implies \pm i\hbar \frac{\partial}{\partial x} = P
\]
(5.54)

We will choose the minus sign because of our previous work.
How about \( E \)? Quite straightforwardly, we obtain
\[
E = i\hbar \frac{\partial}{\partial t}.
\]
(5.55)

Hence, it is consistent, so to speak, with the Schrödinger Equation.

We have
\[
\begin{cases}
    P [\Psi(x,t)] = -i\hbar \frac{\partial}{\partial x} [\Psi(x,t)] \\
    E [\Psi(x,t)] = i\hbar \frac{\partial}{\partial t} [\Psi(x,t)]
\end{cases}
\]
(5.56)

So,
\[
\langle P \rangle = \int_{-\infty}^{+\infty} \Psi^*(x,t)P\Psi(x,t)dx = \int_{-\infty}^{+\infty} \Psi^*(x,t) \left( -i\hbar \frac{\partial}{\partial x} \right) \Psi(x,t)dx \\
= -i\hbar \int_{-\infty}^{+\infty} \Psi^* \frac{\partial}{\partial x} \Psi dx
\]
(5.57)

and
\[
\langle E \rangle = \int_{-\infty}^{+\infty} \Psi^*(x,t)E\Psi(x,t)dx = i\hbar \int_{-\infty}^{+\infty} \Psi^* \frac{\partial}{\partial t} \Psi dx.
\]
(5.58)

Incidentally, this is the same as
\[
\langle E \rangle = \int_{-\infty}^{+\infty} \Psi^*(x,t) \left( -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x,t) \right) \Psi(x,t)dx.
\]
(5.59)
**An Infinite Square Well Potential**

Consider a particle with total energy $E$ in the potential well given below.

$$V(x, t) = \begin{cases} 
0 & -\frac{a}{2} < x < \frac{a}{2} \\
\infty & |x| \geq \frac{a}{2}
\end{cases} \quad (a > 0) \quad (5.60)$$

One normalized solution to the Schrödinger Equation is

$$\Psi(x, t) = \begin{cases} 
A \cos \frac{\pi x}{a} e^{-iEt/\hbar} & -\frac{a}{2} < x < \frac{a}{2} \\
0 & |x| \geq \frac{a}{2}
\end{cases} \quad (a > 0) \quad ; \quad (5.61)$$

where $A$ is known as a normalization constant.

Let us conduct a consistency check with this function. Since

$$\frac{\partial \Psi}{\partial t} = \frac{-iE}{\hbar} e^{-iEt/\hbar} A \cos \frac{\pi x}{a} = \frac{-iE}{\hbar} \Psi(x, t), \quad (5.62)$$

we get

$$\langle E \rangle = i\hbar \int_{-\infty}^{+\infty} \Psi^*(x, t) \frac{\partial \Psi(x, t)}{\partial t} dx = i\hbar \int_{-\infty}^{+\infty} \Psi^*(x, t) \left( \frac{-iE}{\hbar} \Psi(x, t) \right)$$

$$= E \int_{-\infty}^{+\infty} \Psi^* \Psi dx. \quad (5.63)$$

However, since we have a normalized wavefunction,

$$\int_{-\infty}^{+\infty} \Psi^* \Psi dx = 1 \quad (5.64)$$

and

$$\langle E \rangle = E. \quad (5.65)$$

I said $A$ was a normalization constant, but I never told you what its value was. Let’s find out.

$$\int_{-\infty}^{+\infty} \Psi^* \Psi dx = \int_{-\frac{a}{2}}^{+\frac{a}{2}} \left( A \cos \frac{\pi x}{a} e^{-iEt/\hbar} \right)^* A \cos \frac{\pi x}{a} e^{-iEt/\hbar} dx$$

$$= \int_{-\frac{a}{2}}^{+\frac{a}{2}} A^* \cos \frac{\pi x}{a} e^{+iEt/\hbar} A \cos \frac{\pi x}{a} e^{-iEt/\hbar} dx = \int_{-\frac{a}{2}}^{+\frac{a}{2}} A^* A \cos^2 \frac{\pi x}{a} e^{+iEt/\hbar} e^{-iEt/\hbar} dx$$
\[ |A|^2 \int_{-\frac{a}{2}}^{\frac{a}{2}} \cos^2 \frac{\pi x}{a} \, dx = 2 |A|^2 \int_0^{\frac{a}{2}} \cos^2 \frac{\pi x}{a} \, dx = 1 \]  

(5.66)

If we let \( y = \frac{\pi x}{a} \), we get \( dx = \frac{a}{\pi} \, dy \) and \( y : 0 \rightarrow \frac{\pi}{2} \) as \( x : 0 \rightarrow \frac{a}{2} \). So, integration by this substitution gives us

\[
\int_0^{\frac{a}{2}} \cos^2 \frac{\pi x}{a} \, dx = \int_0^{\frac{\pi}{2}} \left( \frac{a}{\pi} \right) \cos^2 y \, dy = \left( \frac{a}{\pi} \right) \left[ \frac{y}{2} + \frac{\sin 2y}{4} \right]_0^{\frac{\pi}{2}} = \frac{a}{\pi} \cdot \frac{\pi}{4} = \frac{a}{4}.
\]

(5.67)

Therefore,

\[ 2 |A|^2 \frac{a}{4} = |A|^2 \frac{a}{2} = 1 \implies |A| = \sqrt{2} \frac{a}{a}. \]

(5.68)

Note here that the normalization constant \( A \) can not be determined uniquely. Indeed, the above relation indicates that there are infinitely many values of \( A \) that serves the purpose.

\[ |A| = \sqrt{2} \frac{a}{a} \implies A = \sqrt{2} e^{i\theta}; \]

(5.69)

where \( \theta \) is any real number. This fact that the normalization constant \( A \) can only be specified up to the argument is another reason why the wavefunction itself does not represent physical reality.
Exercises

1. Consider a classical sinusoidal traveling wave given by the equation below; where \( \lambda \) is the wavelength and \( \nu \) is the frequency so that \( \nu \lambda = v \) (the speed of its propagation).

\[
\Psi(x, t) = A \sin \left[ \frac{2\pi}{\lambda} (x - vt) \right]
\]

Show that

\[
\Psi(x, t) = \Psi(x + vt_0, t + t_0),
\]

and briefly explain what this means.\(^3\)

2. Write down, but do not derive, the time-dependent Schrödinger equation for \( V(x, t) \). This is a one-dimensional case, and the only variables are the spatial variable \( x \) and the temporal variable \( t \).

3. If \( \Psi_1(x, t) \) and \( \Psi_2(x, t) \) are both solutions of the Schrödinger equation, show that any linear combination \( \alpha \Psi_1(x, t) + \beta \Psi_2(x, t) \) is also a solution; where \( \alpha \) and \( \beta \) are scalars.

4. Define \( e^{i\theta} \) by \( e^{i\theta} = \cos \theta + i \sin \theta \) and prove the following relationships.

   (a) \( e^{i\theta}e^{i\phi} = e^{i(\theta+\phi)} \)
   (b) \( (e^{i\theta})^n = e^{in\theta} \)

5. This problem is the same as Chapter 7 Problem 1.

Consider a particle of mass \( m \) and total energy \( E \) which can move freely along the \( x \)-axis in the interval \([ -\frac{a}{2}, +\frac{a}{2} ] \), but is strictly prohibited from going outside this region. This corresponds to what is called an infinite square well potential \( V(x) \) given by \( V(x) = 0 \) for \( -\frac{a}{2} < x < +\frac{a}{2} \) and \( V(x) = \infty \) elsewhere. If we solve the Schrödinger equation for this \( V(x) \), one of the solutions is

\[
\Psi(x, t) = \begin{cases} 
A \cos \left( \frac{\pi x}{a} \right) e^{-iEt/h} & -\frac{a}{2} < x < \frac{a}{2} \\
0 & |x| \geq \frac{a}{2} \end{cases} \quad (a > 0).
\]

\(^{\text{Note that}}\)

\[
\nu \lambda = v \Rightarrow \omega = \frac{2\pi}{\lambda} v.
\]

Therefore, we have

\[
\Psi(x, t) = A \sin \left[ \frac{2\pi}{\lambda} x - \omega t \right].
\]
(a) Find \( A \) so that the function \( \Psi(x, t) \) is properly normalized.

(b) In the region where the potential \( V(x) = 0 \), the Schrödinger equation reduces to

\[
-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \Psi(x, t) = i\hbar \frac{\partial \Psi(x, t)}{\partial t}.
\]

Plug \( \Psi(x, t) = A \cos\left(\frac{\pi x}{a}\right)e^{-iEt/\hbar} \) into this relation to show \( E = \frac{\pi^2\hbar^2}{2ma^2} \).

(c) Show that \( \langle P \rangle = 0 \).

(d) Evaluate \( \langle x^2 \rangle \) for this wavefunction. You will probably need an integral table for this. (Of course, you can always try contour integration or some such. But, that is beyond this course.)

(e) Evaluate \( \langle P^2 \rangle \) for the same wavefunction. You will not need an integral table if you use the fact that the function is normalized. Of course, a brute force computation will yield the same result as well.

6. This problem is the same as Chapter 7 Problem 2.

If you are a careful student who pays attention to details, you may have realized that \( A \) can only be determined up to the argument, or up to the sign if you assume \( A \) is a real number.

(a) What is the implication of this fact as to the uniqueness of wavefunction?

(b) What is the implication of this fact as to the probability density \( \Psi^*(x, t)\Psi(x, t) \)? How about \( \langle P \rangle \), \( \langle P^2 \rangle \), and \( \langle x^2 \rangle \)?
Chapter 6

The Time-Independent
Schrödinger Equation

6.1 Separation of Time $t$ and Space $x$

In the Schrödinger Equation

$$\left[-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x,t)\right] \Psi(x,t) = i\hbar \frac{\partial \Psi(x,t)}{\partial t} \quad (6.1)$$

the potential often does not depend on time; i.e. $V(x,t) = V(x)$. An example of this would be the infinite square well potential.

$$V(x,t) = \begin{cases} 
0 & -\frac{a}{2} < x < \frac{a}{2} \\
\infty & |x| \geq \frac{a}{2} 
\end{cases} \quad (a > 0) \quad (6.2)$$

When $V(x,t) = V(x)$, we can express the multivariable function $\Psi(x,t)$ as a product of a function of $x$ and a function of $t$ as follows.

$$\Psi(x,t) = \psi(x)\phi(t) \quad (6.3)$$

This technique is called “separation of variables”. Let us substitute the above $\Psi$ into the Schrödinger Equation.

$$\left[-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x)\right] \psi(x)\phi(t) = i\hbar \frac{\partial \psi(x)\phi(t)}{\partial t}$$

$$\Rightarrow -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \psi(x)\phi(t) + V(x)\psi(x)\phi(t) = i\hbar \frac{\partial \psi(x)\phi(t)}{\partial t}$$


\[ \implies -\frac{\hbar^2}{2m}\phi(t)\frac{\partial^2}{\partial x^2}\psi(x) + V(x)\psi(x)\phi(t) = \psi(x)i\hbar\frac{\partial\phi(t)}{\partial t} \]

\[ \implies \phi(t) \left[ -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\psi(x) + V(x)\psi(x) \right] = \psi(x)i\hbar\frac{\partial\phi(t)}{\partial t} \quad (6.4) \]

Dividing through by \( \psi(x)\phi(t) \), we get

\[ \frac{1}{\psi(x)} \left[ -\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\psi(x) + V(x)\psi(x) \right] = i\hbar\frac{1}{\phi(t)} \frac{d\phi(t)}{dt}; \quad (6.5) \]

where we changed the partial derivatives to total derivatives as we now have one-variable functions.

Generally speaking,

\[ H(x) = K(t) \text{ for all pairs } (x,t) \quad (6.6) \]

means that both \( H(x) \) and \( K(t) \) are a constant. Let \( H(x) = K(t) = G \) (a constant). Then,

\[ K(t) = G \iff i\hbar\frac{1}{\phi(t)} \frac{d\phi(t)}{dt} = G \iff \frac{d\phi(t)}{dt} = \frac{1}{i\hbar}G\phi(t) = -\frac{i}{\hbar}G\phi(t). \quad (6.7) \]

Therefore,

\[ \phi(t) = Ae^{-iGt/\hbar}. \quad (6.8) \]

Compare this with the time-dependent part of the travelling wave

\[ e^{i(\omega t - kx)} = e^{ikx}e^{-i\omega t}. \quad (6.9) \]

We realize that \( \frac{G}{\hbar} = 2\pi\frac{\omega}{\hbar} \) “should be” \( \omega \).

\[ 2\pi\frac{G}{\hbar} = \omega \implies G = \frac{\omega\hbar}{2\pi} = \frac{2\pi\nu\hbar}{2\pi} = \hbar\nu = E. \quad (6.10) \]

We now have \( G = E \), and

\[ \phi(t) = e^{-iEt/\hbar}. \quad (6.11) \]
This also implies

\[ H(x) = \frac{1}{\psi(x)} \left[ -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \psi(x) + V(x)\psi(x) \right] = E \]

\[ \implies -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \psi(x) + V(x)\psi(x) = E\psi(x); \quad (6.12) \]

where \( E \) is the total energy.

\[ -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \psi(x) + V(x)\psi(x) = E\psi(x) \quad (6.13) \]

or

\[ \left[ -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x) \right] \psi(x) = E\psi(x) \quad (6.14) \]

is the Time-Independent Schrödinger Equation.

Note here that the full wavefunction is given by

\[ \Psi(x,t) = \psi(x)e^{-iEt/\hbar}. \quad (6.15) \]

In the rest of this course, we will solve the Time-Independent Schrödinger Equation for various systems, culminating in a solution for the hydrogen atom.

### 6.2 Conditions on the Wavefunction \( \psi(x) \)

Previously, I mentioned that only normalized solutions are the physically meaningful solutions in order for the Born’s interpretation to make sense. However, this restriction alone would not eliminate all the mathematically correct solutions which are not acceptable on physical grounds. Hence, we need to put further restrictions on the nature of the solutions. We will place three types of restrictions on both \( \psi(x) \) and \( \frac{d\psi(x)}{dx} \), which can be extended naturally to higher dimensions such as \( \psi(x,y) \) and \( \psi(x,y,z) \). The following are ad-hoc and post-hoc conditions albeit physically very reasonable.

#### Physical Sense Revisited:

1. Finiteness and Integrability: The solution and its first derivative have to take a finite value for all \( x \). Furthermore, the wavefunction has to be integrable in order to be normalizable; i.e. \( \int_{-\infty}^{+\infty} \psi^*\psi dx < \infty \). This is due to Born’s probabilistic interpretation.
2. Single-valuedness: The solution and the first derivative should have one value at each point in space.

3. Continuity: In principle, both the solution and its first derivative have to be continuous everywhere.

We will find solutions of this type. Incidentally, Condition 3 on the derivative can not be imposed if the potential \( V(x) \) does not remain finite. We will see such an example in Section 7.5 where the potential blows up to \(+\infty\). Else, both \( \psi(x) \) and \( \frac{d\psi(x)}{dx} \) are continuous. In particular, this means that \( \psi(x) \) and \( \frac{d\psi(x)}{dx} \) are continuous even if the potential \( V(x) \) itself is discontinuous. However, the second derivative \( \frac{d^2\psi(x)}{dx^2} \) is discontinuous if \( V(x) \) is, which is clear from the Schrödinger equation itself as shown in (6.16). Note that the right-hand side of (6.16) is discontinuous if \( V(x) \) is discontinuous because \( \psi(x) \) is continuous.

\[
-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \psi(x) + V(x)\psi(x) = E\psi(x) \implies -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \psi(x) = (E - V(x)) \psi(x)
\]

\[
\implies \frac{d^2}{dx^2} \psi(x) = \frac{2m}{\hbar^2} (V(x) - E) \psi(x) \quad (6.16)
\]

One can also see from (6.16) that \( \frac{d\psi(x)}{dx} \) has to be continuous if \( V(x) \) is finite because a function has an infinite first derivative at a discontinuity and the first derivative in this case is \( \frac{d}{dx} \left( \frac{d\psi(x)}{dx} \right) = \frac{d^2\psi(x)}{dx^2} \).

In case \( V(x) \) is continuous, we can prove continuity of \( \frac{d\psi(x)}{dx} \) as follows. Let \( \delta \) be an arbitrary positive number. We will show continuity of \( \frac{d\psi(x)}{dx} \) at \( x = x_0 \).

\[
\int_{x_0}^{x_0+\delta} \frac{d^2}{dx^2} \psi(x) \, dx = \frac{2m}{\hbar^2} \int_{x_0}^{x_0+\delta} (V(x) - E) \psi(x) \, dx
\]

\[
\implies \int_{x_0}^{x_0+\delta} \frac{d}{dx} \left( \frac{d\psi(x)}{dx} \right) \, dx = \frac{2m}{\hbar^2} \int_{x_0}^{x_0+\delta} (V(x) - E) \psi(x) \, dx
\]

\[
\implies \left[ \frac{d\psi(x)}{dx} \right]_{x_0}^{x_0+\delta} = \frac{2m}{\hbar^2} \int_{x_0}^{x_0+\delta} (V(x) - E) \psi(x) \, dx
\]

\[
\implies \left[ \frac{d\psi(x)}{dx} \right]_{x_0}^{x_0+\delta} - \left[ \frac{d\psi(x)}{dx} \right]_{x_0} = \frac{2m}{\hbar^2} \int_{x_0}^{x_0+\delta} (V(x) - E) \psi(x) \, dx \quad (6.17)
\]
Now, when the potential $V(x)$ is continuous, $|V(x) - E|$ and $|\psi(x)|$ are bounded by a finite number $M$ from above on the interval $[x_0, x_0 + \delta]$.$^a$ And hence, we have

$$\left| \frac{d\psi(x)}{dx} \right|_{x_0 + \delta} - \left| \frac{d\psi(x)}{dx} \right|_{x_0} = \frac{2m}{\hbar^2} \int_{x_0}^{x_0 + \delta} (V(x) - E) \psi(x) \, dx$$

$$\leq \frac{2m}{\hbar^2} \int_{x_0}^{x_0 + \delta} |V(x) - E| |\psi(x)| \, dx \leq \frac{2m}{\hbar^2} \cdot M^2 \cdot \delta. \quad (6.18)$$

Now let $\delta \to 0$. Noting that the relation 6.18 holds for smaller $\delta$ with the same $M$, we can conclude

$$\left| \frac{d\psi(x)}{dx} \right|_{x_0 + \delta} - \left| \frac{d\psi(x)}{dx} \right|_{x_0} \xrightarrow[\delta \to 0]{} 0. \quad (6.19)$$

This means $\frac{d\psi(x)}{dx}$ is continuous from the right or right-continuous at $x = x_0$. Likewise,

$$\int_{x_0 - \delta}^{x_0} \frac{d^2\psi(x)}{dx^2} \, dx = \frac{2m}{\hbar^2} \int_{x_0 - \delta}^{x_0} (V(x) - E) \psi(x) \, dx$$

$$\Rightarrow \left| \frac{d\psi(x)}{dx} \right|_{x_0 - \delta} - \left| \frac{d\psi(x)}{dx} \right|_{x_0} \xrightarrow[\delta \to 0]{} 0. \quad (6.20)$$

And, $\frac{d\psi(x)}{dx}$ is also left-continuous at $x = x_0$. Hence, the first derivative is continuous at $x = x_0$.

$^a$This is from Theorem D.1 called the Extreme Value Theorem.

See Appendix D for a mathematical definition of continuity and a mathematically more rigorous proof of continuity of $\frac{d\psi(x)}{dx}$.
Exercises

1. Write down, but do not derive, the time-independent Schrödinger equation for \( V(x) \), that is, \( V \) is now a function of \( x \) only. This is a one-dimensional case, and the only spatial variable is \( x \).

2. The time-dependent Schrödinger Equation is

\[
-\frac{\hbar^2}{2m} \frac{\partial^2 \Psi(x,t)}{\partial x^2} + V(x,t)\Psi(x,t) = i\hbar \frac{\partial \Psi(x,t)}{\partial t}.
\]

Consider a case where the potential does not depend on time; i.e. \( V(x,t) = V(x) \). Assume \( \Psi(x,t) = \psi(x)e^{-iEt/\hbar} \) and derive the time-independent Schrödinger Equation.

3. Discuss in concrete terms and sufficient detail why \( \Phi(\phi) = \sin \phi \) is an acceptable wavefunction while \( \Phi(\phi) = \phi \) is not. We are using the spherical polar coordinates here.
Chapter 7

Solutions of Time-Independent Schrödinger Equations in One Dimension

Though you may, perhaps naively, think our world is three-dimensional, there are many physical systems which are of lower dimensions. For example, a thin film forms a two-dimensional system, and a particle moving freely without any external force is the simplest one-dimensional system. Furthermore, the Schrödinger equations for one-dimensional systems are easier to solve. Hence, it makes good mathematical and physical sense to start with one dimensional cases.

7.1 The Zero Potential

This is the case where $V(x) = 0$ for all $x$.

Our Time-Independent Schrödinger Equation is

$$-\frac{\hbar^2}{2m} \frac{d^2 \psi(x)}{dx^2} = E \psi(x). \quad (7.1)$$

The most general solution for this second order linear ordinary differential equation is of the form

$$\psi(x) = A_1 \sin \frac{\sqrt{2mE}}{\hbar} x + A_2 \cos \frac{\sqrt{2mE}}{\hbar} x = C_1 e^{ikx} + C_2 e^{-ikx}, \quad (7.2)$$
where \( A_1, A_2, C_1, \) and \( C_2 \) are arbitrary constants, and \( k = \frac{\sqrt{2mE}}{\hbar} \).

On the other hand, we always have
\[
\phi(t) = e^{-iEt/\hbar} = e^{-ih\nu t/(\hbar/2\pi)} = e^{-i2\pi\nu t} = e^{-i\omega t}
\]
for time dependence.

Therefore, the full time-dependent wavefunction is
\[
\Psi(x,t) = \psi(x)\phi(t) = \left( C_1 e^{ikx} + C_2 e^{-ikx} \right) e^{-i\omega t}
\]
(7.4)

Let us take
\[
\Psi(x,t) = C_1 e^{i(kx-\omega t)} \quad (= C_1 e^{ikx} e^{-iEt/\hbar}).
\]
(7.5)

Then,
\[
\mathcal{P} = \langle P \rangle = \int_{-\infty}^{+\infty} \Psi^* P \Psi \, dx = \int_{-\infty}^{+\infty} C_1^* e^{-ikx} e^{iEt/\hbar} (-i\hbar) \frac{\partial}{\partial x} \Psi_1 e^{ikx} e^{-iEt/\hbar} \, dx
\]
\[
= \int_{-\infty}^{+\infty} C_1^* e^{-ikx} e^{iEt/\hbar} (-i\hbar)(ik) C_1 e^{ikx} e^{-iEt/\hbar} \, dx = \left( \int_{-\infty}^{+\infty} \Psi^* \Psi \, dx \right)
\]
\[
= \hbar k \int_{-\infty}^{+\infty} \psi^* \psi \, dx = \hbar k = \frac{\hbar}{2\pi} \cdot \frac{2mE}{\hbar} = \sqrt{2mE}.
\]
(7.6)

This makes sense as
\[
\frac{P^2}{2m} = E \implies P = \sqrt{2mE}
\]
(7.7)
classically, and (7.6) serves as yet another consistency checking device.

As simple as the concept and the wavefunctions of a free particle are, there is one serious complication regarding normalization. To see this, consider an eigenfunction \( \psi(x) = Ae^{ikx} \). In order for this function to be normalizable, it has to be integrable to begin with. However,
\[
\int_{-\infty}^{+\infty} (Ae^{ikx})^* (Ae^{ikx}) \, dx = |A|^2 \int_{-\infty}^{+\infty} e^{-ikx} e^{ikx} \, dx = |A|^2 \int_{-\infty}^{+\infty} 1 \, dx = \infty,
\]
(7.8)
and $\psi(x)$ is not normalizable in the usual way.

In order to circumvent this problem, three normalization schemes have been proposed; namely, the Born normalization, Dirac normalization, and unit-flux normalization. As these normalization procedures are not crucial for the rest of this book, their discussions are relegated to Appendix G.

### 7.2 The Step Potential ($E < V_0$)

We now consider a step potential given by

$$V(x) = \begin{cases} V_0 & x > 0 \\ 0 & x \leq 0 \end{cases}$$

(7.9)

for $E < V_0$.

In Newtonian (classical) mechanics, the particle can not enter the region $(0, \infty)$. We have

$$E = K.E. + P.E. = K.E. + V_0 \text{ in } (0, \infty) \implies K.E. = E - V_0 < 0.$$  

(7.10)

This does not make any sense of any kind classically. Is this also the case in quantum mechanics? Let us just plunge in and solve it!

The time-independent Schrödinger Equation is

$$-\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + V(x)\psi(x) = E\psi(x).$$

(7.11)

In our case, we have the following.

$$-\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} = E\psi(x) \quad x \leq 0$$  

(7.12)

$$-\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + V_0\psi(x) = E\psi(x) \quad x > 0$$

(7.13)

Equation 7.12 is that for a free particle. Therefore, the general solution is

$$\psi(x) = Ae^{ik_1x} + Be^{-ik_1x};$$

(7.14)
where $k_1 = \sqrt{\frac{2mE}{\hbar}}$, and $A$ and $B$ are arbitrary constants.

On the other hand, (7.13) can be rewritten as follows.

$$
\frac{-\hbar^2}{2m} \frac{d^2 \psi(x)}{dx^2} = (E - V_0)\psi(x) \iff \frac{\hbar^2}{2m} \frac{d^2 \psi(x)}{dx^2} = (V_0 - E)\psi(x)
$$

$$
\iff \frac{d^2 \psi(x)}{dx^2} = \frac{2m(V_0 - E)}{\hbar^2} \psi(x) \quad (7.15)
$$

The general solution for $x > 0$ is

$$
\psi(x) = Ce^{k_2x} + De^{-k_2x}; \quad (7.16)
$$

where $k_2 = \sqrt{\frac{2m(V_0 - E)}{\hbar}}$, and $C$ and $D$ are arbitrary constants. Note that $V_0 - E$ is positive in this region.

As stated above, these are indeed the solutions, or the (most) general solutions, since we have linear second order ordinary differential equations.

We have

$$
\begin{cases}
\psi(x) = Ae^{ik_1x} + Be^{-ik_1x} & (k_1 = \sqrt{\frac{2mE}{\hbar}}) \quad x \leq 0 \\
\psi(x) = Ce^{k_2x} + De^{-k_2x} & (k_2 = \sqrt{\frac{2m(V_0 - E)}{\hbar}}) \quad x > 0.
\end{cases} \quad (7.17)
$$

Recall the conditions on $(\psi, \frac{d\psi}{dx})$.

1. Finite
2. Single-valued
3. Continuous

We will determine $A$, $B$, $C$, and $D$ using these conditions.

Let us begin with the region $x > 0$. In this region,

$$
\psi(x) = Ce^{k_2x} + De^{-k_2x}; \text{ where } k_2 = \sqrt{\frac{2m(V_0 - E)}{\hbar}}. \quad (7.18)
$$

Therefore,

$$
\psi^*(x)\psi(x) = (C^*e^{k_2x} + D^*e^{-k_2x})(Ce^{k_2x} + De^{-k_2x})
$$
\[ 7.2. \text{THE STEP POTENTIAL (} E < V_0 \text{)} \]

\[ E = C|e^{2k_2x} + \frac{2\text{Re}C^*D}{D^*C + C^*D}\left|D\right|^2 e^{-2k_2x}. \]  

(7.19)

Since the wavefunction has to be normalizable, it has to satisfy the finiteness condition

\[ \int_{-\infty}^{+\infty} \psi^*(x)\psi(x)dx = \int_{-\infty}^{+\infty} \left|\psi(x)\right|^2dx < \infty. \]  

(7.20)

Noting that this \( \psi \) is the wavefunction only in the region \( x > 0 \), we actually need

\[ \int_{0}^{+\infty} \left|\psi(x)\right|^2dx < \infty. \]  

(7.21)

Hence, we have to have

\[ C = 0. \]  

(7.22)

Though this may seem intuitively obvious, let us conduct a complete verification of this for once. We will start with (7.19) and note that

\[ \int_{0}^{\infty} |D|^2 e^{-2k_2x} dx = |D|^2 \frac{1}{-2k_2} e^{-2k_2x}\big|_0^{\infty} = |D|^2 \frac{1}{-2k_2} (-1) = \frac{|D|^2}{2k_2} < \infty. \]  

(7.23)

So, we need only to deal with the first two terms \( |C|^2 e^{2k_2x} + \underbrace{2\text{Re}C^*D}_{D^*C + C^*D} \). Set \( R = 2\text{Re}C^*D \) and consider \( |C|^2 e^{2k_2x} + D^*C + C^*D = |C|^2 e^{2k_2x} + R \). If \( R \geq 0 \),

\[ \int_{0}^{\infty} |C|^2 e^{2k_2x} + R dx \]  

(7.24)

is clearly not finite unless \( C = R = 0 \). So, consider \( R < 0 \) and \( C \neq 0 \), and let \( L = -R > 0 \). We have \( |C|^2 e^{2k_2x} + R = |C|^2 e^{2k_2x} - L \), and the following inequality holds.

\begin{align*}
    x > \frac{\ln \frac{2L}{|C|^2}}{2k_2} & \Rightarrow 2k_2x > \ln \frac{2L}{|C|^2} \Rightarrow e^{2k_2x} > e^{\ln \frac{2L}{|C|^2}} = \frac{2L}{|C|^2} \Rightarrow |C|^2 e^{2k_2x} > 2L \\
    & \Rightarrow |C|^2 e^{2k_2x} - L > L > 0 \Rightarrow |C|^2 e^{2k_2x} + R > L > 0
\end{align*}  

(7.25)

Therefore,

\[ \int_{0}^{\infty} |C|^2 e^{2k_2x} + \underbrace{2\text{Re}C^*D}_{D^*C + C^*D} dx = \int_{0}^{\infty} |C|^2 e^{2k_2x} + R dx \]
CHAPTER 7. SOLUTIONS OF TIME-INDEPENDENT SCHRÖDINGER EQUATIONS IN ONE DIMENSION

\[ \int_{-2k_2}^{2k_2} |C|^2 e^{2k_2 x} + R \, dx + \int_{-\infty}^{\infty} |C|^2 e^{2k_2 x} + R \, dx \]

\[ \geq \int_{0}^{\ln \frac{2L}{|C|^2}} |C|^2 e^{2k_2 x} + R \, dx + \int_{0}^{\ln \frac{2L}{|C|^2}} L \, dx \quad (7.26) \]

is not finite and we need to require \( C = 0 \). Note that this also means \( R = \frac{2Re C^*D}{D^*C + C^*D} = 0 \).

At this point, we have

\[
\begin{cases}
\psi(x) = Ae^{ik_1 x} + Be^{-ik_1 x} & k_1 = \frac{\sqrt{2mE}}{\hbar} \quad x \leq 0 \\
\psi(x) = De^{-k_2 x} & (k_2 = \frac{\sqrt{2m(V_0 - E)}}{\hbar}) \quad x > 0
\end{cases}
\quad (7.27)
\]

Next, we will use the continuity condition on \( \psi(x) \) and \( \frac{d\psi(x)}{dx} \).

Continuity of \( \psi(x) \) at \( x = 0 \) gives

\[ De^{-k_2 0} = Ae^{ik_1 0} + Be^{-ik_1 0} \implies D = A + B. \quad (7.28) \]

Since the first derivative is

\[
\begin{cases}
\frac{d\psi(x)}{dx} = A(ik_1)e^{ik_1 x} + B(-ik_1)e^{-ik_1 x} & k_1 = \frac{\sqrt{2mE}}{\hbar} \quad x \leq 0 \\
\frac{d\psi(x)}{dx} = D(-k_2)e^{-k_2 x} & (k_2 = \frac{\sqrt{2m(V_0 - E)}}{\hbar}) \quad x > 0
\end{cases}
\quad (7.29)
\]

continuity of \( \frac{d\psi(x)}{dx} \) at \( x = 0 \) gives

\[ -k_2 De^{-k_2 0} = ik_1 Ae^{ik_1 0} - ik_1 Be^{-ik_1 0} \implies -k_2 D = ik_1(A - B) \]

\[ \implies -\frac{k_2}{ik_1} D = A - B. \quad (7.30) \]

We now have a simultaneous equation

\[
\begin{cases}
\frac{ik_2}{k_1} D = A - B \\
D = A + B
\end{cases}
\quad (7.31)
\]

for \( A, B, \) and \( D \) with the solution

\[ A = \frac{1}{2} \left( 1 + \frac{ik_2}{k_1} \right) D, \]

\[ B = \frac{1}{2} \left( 1 - \frac{ik_2}{k_1} \right) D. \quad (7.32) \]
This gives
\[
\psi(x) = \begin{cases}
\frac{D}{2} (1 + ik_2/k_1) e^{ik_1x} + \frac{D}{2} (1 - ik_2/k_1) e^{-ik_1x} & x \leq 0 \\
De^{-k_2x} & x > 0
\end{cases}.
\] (7.33)

So, the full solution is
\[
\Psi(x,t) = \psi(x)\phi(t) = \psi(x)e^{-iEt/\hbar} = \begin{cases}
\frac{D}{2} (1 + ik_2/k_1) e^{i(k_1x-Et/\hbar)} + \frac{D}{2} (1 - ik_2/k_1) e^{-i(k_1x+Et/\hbar)} & x \leq 0 \\
De^{-k_2x}e^{-iEt/\hbar} & x > 0
\end{cases}.
\] (7.34)

Recall that
\[
e^{-iEt/\hbar} = e^{-i\omega t}.
\] (7.35)

This is because
\[
Et/\hbar = h\nu t/\hbar = h\nu t/\left(\frac{\hbar}{2\pi}\right) = 2\pi\nu t = \omega t.
\] (7.36)

Therefore, in the region \(x \leq 0\),
\[
\Psi(x,t) = Ae^{i(k_1x-\omega t)} + Be^{i(-k_1x-\omega t)}.
\] (7.37)

Note that
\[
Ae^{i(k_1x-\omega t)}
\] (7.38)
is traveling to the right while
\[
Be^{i(-k_1x-\omega t)}
\] (7.39)
is traveling to the left. In other words,
\[
Ae^{i(k_1x-\omega t)}
\] (7.40)
is the incident wave, incident on the step from the left, while
\[
Be^{i(-k_1x-\omega t)}
\] (7.41)
is the wave reflected by the step.
CHAPTER 7. SOLUTIONS OF TIME-INDEPENDENT SCHRÖDINGER EQUATIONS IN ONE DIMENSION

What is the reflection coefficient? Since the reflection coefficient is the ratio of the amplitude of the reflected wave and the amplitude of the incident wave, we have

\[\text{the reflection coefficient} = \frac{\text{the amplitude of the reflected wave}}{\text{the amplitude of the incident wave}}\]

\[= \frac{|B e^{i(-k_1 x - \omega t)}|^2}{|A e^{i(k_1 x - \omega t)}|^2} = \frac{|B|^2 e^{i(-k_1 x - \omega t)}|^2}{|A|^2 e^{i(k_1 x - \omega t)}|^2} = \frac{|B|^2}{|A|^2} = \frac{B^* B}{A^* A}\]

\[= \frac{D^*}{D} \left( 1 - \frac{i k_2}{k_1} \right)^* \frac{D}{D} \left( 1 - \frac{i k_2}{k_1} \right) = \frac{1 + \frac{i k_2}{k_1}}{1 - \frac{i k_2}{k_1}} = 1 \quad (7.42)\]

And, we have a total reflection.

Indeed, we have a standing wave in the region \(x \leq 0\). To see this, plug \(e^{ik_1 x} = \cos k_1 x + i \sin k_1 x\) into

\[\psi(x) = \begin{cases} \frac{D}{D} \left[ (1 + \frac{i k_2}{k_1}) e^{ik_1 x} + (1 - \frac{i k_2}{k_1}) e^{-ik_1 x} \right] & x \leq 0 \\ De^{-k_2 x} & x > 0 \end{cases} \quad (7.43)\]

to obtain

\[\psi(x) = \begin{cases} D \cos k_1 x - D \frac{k_2}{k_1} \sin k_1 x & x \leq 0 \\ De^{-k_2 x} & x > 0 \end{cases} \quad (7.44)\]

So, for \(x \leq 0\),

\[\Psi(x, t) = D \left( \cos k_1 x - \frac{k_2}{k_1} \sin k_1 x \right) e^{-iEt/\hbar}. \quad (7.45)\]

Remember the following from high school math,

\[\alpha \cos kx - \beta \sin kx = \left( \frac{\alpha}{\sqrt{\alpha^2 + \beta^2}} \cos kx - \frac{\beta}{\sqrt{\alpha^2 + \beta^2}} \sin kx \right) \sqrt{\alpha^2 + \beta^2} = \sqrt{\alpha^2 + \beta^2} \cos(kx + \theta); \quad (7.46)\]

where \(\theta\) is such that

\[\cos \theta = \frac{\alpha}{\sqrt{\alpha^2 + \beta^2}} \quad \text{and} \quad \sin \theta = \frac{\beta}{\sqrt{\alpha^2 + \beta^2}}. \quad (7.47)\]
Therefore, the nodes of $\Psi(x,t)^*\Psi(x,t)$ do not move, and we indeed have a standing wave.

Now, in the region $x > 0$,

$$\Psi^*(x,t)\Psi(x,t) = D^*De^{-2kx}.$$  \hspace{1cm} (7.48)

While

$$\lim_{x \to \infty} \Psi^*\Psi = 0,$$ \hspace{1cm} (7.49)

$$\Psi^*\Psi \neq 0$$ \hspace{1cm} (7.50)

for $\forall x > 0$. This means that the probability of finding the particle to the right of the step is not zero even if the total energy $E$ is smaller than the step height $V_0$. This phenomenon is definitely non-classical and is known as “barrier penetration”.

### 7.3 The Step Potential ($E > V_0$)

We will now consider the case where the total energy $E$ is greater than the step height.

Classically, we have the following.

1. Not a total reflection
2. Has to be a total penetration into $(0, \infty)$

But, quantum mechanically, the Schrödinger Equation predicts the following.

1. Not a total reflection (agreement with the classical theory)
2. There is partial reflection at the boundary $x = 0$ (disagreement)

Let us see how it works.

As usual, $\Psi(x,t) = \psi(x)e^{-iEt/\hbar}$, and the Time-Independent Schrödinger Equation is

$$\left\{ \begin{array}{ll}
-\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} &= E\psi(x) & x < 0 \\
-\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} &= (E - V_0)\psi(x) & x > 0 
\end{array} \right.$$  \hspace{1cm} (7.51)
As $E - V_0 > 0$ this time, the two equations are basically the same. They are both free particles, and the solutions are

$$
\psi(x) = \begin{cases} 
A e^{ik_1x} + B e^{-ik_1x} & x < 0 \\
C e^{ik_2x} + D e^{-ik_2x} & x > 0 
\end{cases} \quad (7.52)
$$

where $k_1 = \sqrt{\frac{2mE}{\hbar}}$ and $k_2 = \sqrt{\frac{2m(E-V_0)}{\hbar}}$.

Suppose the particle is in the region $x < 0$ or $(-\infty, 0)$ at $t = 0$. In other words, the particle is incident on the step from the left. Then, since the wave is travelling to the right in the region $(0, \infty)$, we should set $D = 0$. This is simply because $e^{-ik_2x} e^{-iEt/\hbar}$ is a wave travelling to the left.

Imposing the continuity condition on $\psi$ and $\frac{d\psi}{dx}$ at $x = 0$, we can express the constants $B$ and $C$ in terms of $A$ as follows.

$$
\begin{align*}
\left. \frac{\psi(x)}{dx} \right|_{x=0^-} &= \left. \frac{\psi(x)}{dx} \right|_{x=0^+} \\
\frac{A+B}{k_1(A-B)} &= k_2C
\end{align*}
\implies \begin{cases} 
B = A \frac{k_1-k_2}{k_1+k_2} \\
C = A \frac{2k_1}{k_1+k_2}
\end{cases} \quad (7.53)
$$

Therefore,

$$
\psi(x) = \begin{cases} 
A e^{ik_1x} + A \frac{k_1-k_2}{k_1+k_2} e^{-ik_1x} & x < 0 \\
A \frac{2k_1}{k_1+k_2} e^{ik_2x} & x > 0
\end{cases} \quad (7.54)
$$

The reflection coefficient $R$ is given by

$$
R = \frac{B^*B}{A^*A} = \left( \frac{k_1-k_2}{k_1+k_2} \right)^2 \quad (7.55)
$$

Since the transmission coefficient $T$ is related to $R$ via $R + T = 1$,

$$
T = 1 - R = 1 - \left( \frac{k_1-k_2}{k_1+k_2} \right)^2 = \frac{(k_1+k_2+k_1-k_2)(k_1+k_2-k_1+k_2)}{(k_1+k_2)^2} = \frac{4k_1k_2}{(k_1+k_2)^2}. \quad (7.56)
$$
As advertised, we do not have a total reflection, and we do not have a total penetration into \((0, \infty)\), either.

We now solve the same problem for a wave incident on the potential boundary at \(x = 0\) from the right. In other words, the particle is in the region \((0, \infty)\) at \(t = 0\) and traveling to the left.

Our general solution is

\[
\psi(x) = \begin{cases} 
A e^{i k_1 x} + B e^{-i k_1 x} & x < 0 \\
C e^{i k_2 x} + D e^{-i k_2 x} & x > 0 
\end{cases}
\]  

(7.57)

where \(k_1 = \sqrt{2mE/\hbar}\) and \(k_2 = \sqrt{2m(E-V_0)/\hbar}\).

Since there is no boundary to reflect the wave in the region \((-\infty, 0)\), the wave should be traveling to the left in this region. We set \(A = 0\).

\[
\psi(x) = \begin{cases} 
B e^{-i k_1 x} & x < 0 \\
C e^{i k_2 x} + D e^{-i k_2 x} & x > 0 
\end{cases}
\]  

(7.58)

The first derivative with respect to \(x\) is

\[
\psi'(x) = \begin{cases} 
-i k_1 B e^{-i k_1 x} & x < 0 \\
i k_2 C e^{i k_2 x} + (-i k_2) D e^{-i k_2 x} & x > 0 
\end{cases}
\]  

(7.59)

Imposing the continuity condition on \(\psi(x)\) and \(\frac{d\psi(x)}{dx}\) at \(x = 0\), we get

\[
\begin{cases}
B = C + D \\
-B = i k_2 (C - D)
\end{cases} \implies \begin{cases}
C + D = B \\
C - D = -\frac{k_1}{k_2} B
\end{cases} \implies \begin{cases}
C = \frac{k_2 - k_1}{k_2 + k_1} B \\
D = \frac{k_2 + k_1}{k_2 + k_1} B
\end{cases}
\]  

(7.60)

What are the reflection coefficient and the transmission coefficient in this case?

\[
R = \frac{C^* C}{D^* D} = \left(\frac{k_2 - k_1}{2k_2}\right)^2 B^* B = \left(\frac{k_1 - k_2}{k_1 + k_2}\right)^2 
\]  

(7.61)

and

\[
T = 1 - R = \frac{4k_1 k_2}{(k_1 + k_2)^2}
\]  

(7.62)
Note that these are the same as before when we computed $R$ and $T$ for the particle moving in from the left.

Incidentally, $B \neq 0$ in (7.52) can be proved on purely mathematical grounds as follows.

If $B = 0$,

$$
\psi(x) = \begin{cases} A e^{ik_1 x} & x < 0 \\ C e^{ik_2 x} & x > 0 \end{cases} \implies \begin{cases} A = C \\ k_1 A = k_2 C \end{cases} \quad (7.63)
$$

But, it is obvious that this can not be solved for $A$ and $C$ unless we accept $A = C = 0$ which is physically meaningless. Hence, $B \neq 0$ on this ground.

### 7.4 The Barrier Potential ($E < V_0$)

The potential $V(x)$ is $V_0$ in the region $(0,a)$ and 0 elsewhere.

$$
V(x) = \begin{cases} V_0 & 0 < x < a \\ 0 & x < 0, \ a < x \end{cases} \quad (7.64)
$$

Our Time-Independent Schrödinger Equations are

$$
\begin{align*}
\frac{-\hbar^2}{2m} \frac{d^2 \psi(x)}{dx^2} + V_0 \psi(x) &= E \psi(x) & x < 0 \\
\frac{-\hbar^2}{2m} \frac{d^2 \psi(x)}{dx^2} &= E \psi(x) & 0 < x < a \\
\frac{-\hbar^2}{2m} \frac{d^2 \psi(x)}{dx^2} &= E \psi(x) & a < x
\end{align*} \quad (7.65)
$$

and the general solutions are

$$
\psi(x) = \begin{cases} A e^{ik_1 x} + Be^{-ik_1 x} & x < 0 \\ Ce^{-k_2 x} + De^{k_2 x} & 0 < x < a \\ Fe^{ik_1 x} + Ge^{-ik_1 x} & a < x \end{cases} \quad (7.66)
$$

where $k_1 = \frac{\sqrt{2mE}}{\hbar}$ and $k_2 = \frac{\sqrt{2m(V_0-E)}}{\hbar}$.

Assuming $x < 0$ at $t = 0$, that is, the particle originated at $x = -\infty$ and is moving to the right, we have to set $G = 0$ as there is nothing to reflect the wave in the region $a < x$. 
On the other hand, our boundary conditions are

\[
\begin{cases}
\psi(x)|_{x=0} = \psi(x)|_{x=0+} \\
\psi(x)|_{x=a} = \psi(x)|_{x=a+} \\
\frac{d\psi(x)}{dx}|_{x=0} = \frac{d\psi(x)}{dx}|_{x=0+} \\
\frac{d\psi(x)}{dx}|_{x=a} = \frac{d\psi(x)}{dx}|_{x=a+}
\end{cases}
\]  

(7.67)

We have the above four equations and the five unknowns \(A, B, C, D,\) and \(F\). But, we also have a normalization condition. So, we can solve for the five unknowns.

As it turned out, \(F \neq 0\), and the probability density \(\Psi^*(x,t)\Psi(x,t) = \psi^*(x)\psi(x)\) has a (nonzero) tail in the region \(x > a\). Of course, it is nonzero also inside the barrier. Therefore, we have both barrier penetration and tunneling.

### 7.5 The Infinite Square Well Potential

The potential is given by

\[
V(x) = \begin{cases}
\infty & |x| > \frac{a}{2} \\
0 & |x| < \frac{a}{2}
\end{cases}
\]  

(7.68)

and we get, as we have already seen,

\[
\psi(x) = \begin{cases}
0 & |x| > \frac{a}{2} \\
A \sin kx + B \cos kx & |x| < \frac{a}{2}
\end{cases}
\]  

(7.69)

where \(k = \frac{\sqrt{2mE}}{\hbar}\). As usual, the full wavefunction is

\[
\Psi(x,t) = \psi(x)e^{-iEt/\hbar}.
\]  

(7.70)

It turned out the continuity condition of the first derivative is too restrictive for this system as the potential blows up to \(\infty\) unlike any known physical system.

The continuity condition on \(\psi(x)\) at \(|x| = \frac{a}{2}\) gives

\[
\begin{align*}
A \sin \frac{ka}{2} + B \cos \frac{ka}{2} &= 0 \\
- A \sin \frac{ka}{2} + B \cos \frac{ka}{2} &= 0
\end{align*}
\]  

\[\Rightarrow \begin{cases}
2B \cos \frac{ka}{2} = 0 \\
2A \sin \frac{ka}{2} = 0
\end{cases}
\]  

\[\Rightarrow \begin{cases}
A = 0 \quad \text{and} \quad \cos \frac{ka}{2} = 0 \\
B = 0 \quad \text{and} \quad \sin \frac{ka}{2} = 0
\end{cases}
\]  

(7.71)
Therefore,
\[
\begin{cases}
\psi(x) = B \cos kx & \text{where } \cos \frac{ka}{2} = 0 \\
\psi(x) = A \sin kx & \text{where } \sin \frac{ka}{2} = 0
\end{cases}
\]
(7.72)

We have two families of solutions.
\[
\begin{cases}
\psi_n(x) = B_n \cos k_n x & \text{where } k_n = \frac{n\pi}{a} \quad n = 1, 3, 5, \ldots \\
\psi_n(x) = A_n \sin k_n x & \text{where } k_n = \frac{n\pi}{a} \quad n = 2, 4, 6, \ldots
\end{cases}
\]
(7.73)

Note that \(n\) is a positive integer in both cases because \(k_n > 0\). We can also note that negative \(n\)’s give redundant solutions and \(n = 0\) gives a physically meaningless trivial solution. Indeed,
\[
n = 0 \implies \psi_0(x) = A_0 \sin 0 = 0.
\]
(7.74)

Let us compute the normalization constants \(A_n\) and \(B_n\). We will obtain real and positive constants.
\[
\int_{-\frac{a}{2}}^{+\frac{a}{2}} (B_n \cos k_n x)^* (B_n \cos k_n x) \, dx = |B_n|^2 \int_{-\frac{a}{2}}^{+\frac{a}{2}} \cos^2 k_n x \, dx
\]
\[
= |B_n|^2 \int_{-\frac{a}{2}}^{+\frac{a}{2}} \frac{1 + \cos 2k_n x}{2} \, dx
\]
\[
= \frac{|B_n|^2}{2} \left[ x + \frac{1}{2k_n} \sin 2k_n x \right]_{-\frac{a}{2}}^{+\frac{a}{2}}
\]
\[
= \frac{|B_n|^2}{2} \left[ \left( \frac{a}{2} + \frac{a}{2n\pi} \sin n\pi \right) - \left( \frac{a}{2} - \frac{a}{2n\pi} \sin n\pi \right) \right] = \frac{|B_n|^2}{2} \cdot a = 1
\]
\[
\implies B_n = \sqrt{\frac{2}{a}}
\]
(7.75)

\[
\int_{-\frac{a}{2}}^{+\frac{a}{2}} (A_n \sin k_n x)^* (A_n \sin k_n x) \, dx = |A_n|^2 \int_{-\frac{a}{2}}^{+\frac{a}{2}} \sin^2 k_n x \, dx
\]
\[
= |A_n|^2 \int_{-\frac{a}{2}}^{+\frac{a}{2}} \frac{1 - \cos 2k_n x}{2} \, dx
\]
\[
= \frac{|A_n|^2}{2} \left[ x - \frac{1}{2k_n} \sin 2k_n x \right]_{-\frac{a}{2}}^{+\frac{a}{2}}
\]
\[
= \frac{|A_n|^2}{2} \left[ x - \frac{a}{2n\pi} \sin \frac{2n\pi}{a} x \right]_{-\frac{a}{2}}^{+\frac{a}{2}}
\]
\[ = \frac{|A_n|^2}{2} \left[ \left( \frac{a}{2} - \frac{a}{2n\pi} \sin n\pi \right) - \left( -\frac{a}{2} + \frac{a}{2n\pi} \sin n\pi \right) \right] = \frac{|A_n|^2}{2} \cdot a = 1 \]

\[ \Rightarrow A_n = \sqrt{\frac{2}{a}} \quad (7.76) \]

Hence, (7.73) becomes

\[ \begin{cases} 
\psi_n(x) = \sqrt{\frac{2}{a}} \cos k_n x \quad \text{where} \quad k_n = \frac{n\pi}{a} \quad n = 1, 3, 5, \ldots \\
\psi_n(x) = \sqrt{\frac{2}{a}} \sin k_n x \quad \text{where} \quad k_n = \frac{n\pi}{a} \quad n = 2, 4, 6, \ldots .
\end{cases} \quad (7.77) \]

We can also write (7.77) as follows.

\[ \psi_n(x) = \sqrt{\frac{2}{a}} \sin \left[ n\pi \left( \frac{x + \frac{1}{2}}{a} \right) \right] = \sqrt{\frac{2}{a}} \sin \left[ n\pi \left( \frac{x}{a} + \frac{1}{2} \right) \right] \quad (7.78) \]

Furthermore, if we shift the \(x\)-axis so that \(V(x) = 0\) for \(0 < x < a\), we will get

\[ \psi_n(x) = \sqrt{\frac{2}{a}} \sin \left( \frac{n\pi x}{a} \right). \quad (7.79) \]

Because \(k_n = \sqrt{\frac{2mE_n}{\hbar}}\), the total energy \(E_n\) can only take discrete values.

\[ k_n = \frac{\sqrt{2mE_n}}{\hbar} \quad \Rightarrow \quad 2mE_n = \hbar^2 k_n^2 \]

\[ \Rightarrow \quad E_n = \frac{\hbar^2 k_n^2}{2m} = \frac{\pi^2 \hbar^2 n^2}{2ma^2} \quad n = 1, 2, 3, 4, 5, \ldots \quad (7.80) \]

Note that \(E_n \propto n^2\) and the spacing between neighboring energy levels

\[ E_{n+1} - E_n = \frac{\pi^2 \hbar^2 (n + 1)^2}{2ma^2} - \frac{\pi^2 \hbar^2 n^2}{2ma^2} = \frac{\pi^2 \hbar^2 (2n + 1)}{2ma^2} \quad (7.81) \]

is proportional to \(2n + 1\). This is the reason for the widening of the gap as \(n \to \infty\). More significant implication of this is that the particle can not have zero total energy. Indeed, the lowest energy level

\[ E_1 = \frac{\pi^2 \hbar^2}{2ma^2} > 0. \quad (7.82) \]

This is one reason why absolute zero degree temperature can not be achieved. While the infinite square well potential is a mathematical idealization, it approximates the potential well experienced by atoms or nuclei in a crystal for example.

Another type of bound system frequently encountered in nature is a diatomic molecule.
CHAPTER 7. SOLUTIONS OF TIME-INDEPENDENT SCHRÖDINGER EQUATIONS IN ONE DIMENSION

7.6 The Simple Harmonic Oscillator Potential

7.6.1 Classic Solution

Consider a typical spring encountered in high school physics with a spring constant $k$ such that the restoring force $F$ is given by $F = -kx$ for the displacement $x$ from the equilibrium position. For no other reason than convention, we will use $c$ instead of $k$ for the spring constant.

Because

$$V(x) = \frac{1}{2}cx^2, \quad (7.83)$$

the Time-Independent Schrödinger Equation is

$$-\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + \frac{c}{2}x^2\psi(x) = E\psi(x). \quad (7.84)$$

We will need some fancy footwork to solve this equation. As it turns out, it is convenient to introduce a new variable $\nu$ analogously to the classical theory. Hence,

$$\nu = \frac{1}{2\pi} \omega = \frac{1}{2\pi} \sqrt{\frac{c}{m}}. \quad (7.85)$$

After substituting $c = (2\pi)^2 \nu^2 m$ into the equation and dividing through by $-\frac{\hbar^2}{2m}$, we get

$$\frac{d^2\psi}{dx^2} + \left[ \frac{2mE}{\hbar^2} - \left( \frac{2\pi m \nu}{\hbar} \right)^2 \right] x^2 \psi = 0. \quad (7.86)$$

\(^1\)To see this, we start with the classical equation $F = ma$. For us,

$$-cx = m \frac{d^2x}{dt^2} \implies \frac{d^2x}{dt^2} = -\frac{c}{m}x \implies x(t) = A \sin \left( \sqrt{\frac{c}{m}} t \right) + B \cos \left( \sqrt{\frac{c}{m}} t \right)$$

$$= \sqrt{A^2 + B^2} \sin \left( \sqrt{\frac{c}{m}} t + \theta \right);$$

where $\cos \theta = \frac{A}{\sqrt{A^2 + B^2}}$ and $\sin \theta = \frac{B}{\sqrt{A^2 + B^2}}$. Therefore, we have

$$\omega = \sqrt{\frac{c}{m}}.$$
Now, let $\alpha = 2\pi m\nu/\hbar$ and $\beta = 2mE/\hbar^2$ to simplify the equation to
\[
\frac{d^2\psi}{dx^2} + (\beta - \alpha^2x^2)\psi = 0.
\]
(7.87)

Here are further manipulations.
\[
u = \sqrt{\alpha}x = \left[\frac{2\pi m}{\hbar^2} \left(\frac{c}{m}\right)^{1/2}\right]^{1/2} x = \frac{(cm)^{1/4}}{\hbar^{1/2}} x
\]
(7.88)
\[
\frac{d^2\psi}{dx^2} = \frac{d}{du} \left( \frac{d\psi}{du} \right) = \frac{d}{dx} \left( \frac{d\psi}{dx} \right) = \frac{du}{dx} \frac{d\psi}{du} \left( \sqrt{\alpha} \frac{d\psi}{du} \right)
\]
(7.90)
\[
\frac{d^2\psi}{dx^2} = \frac{\sqrt{\alpha}}{\sqrt{\alpha}} \frac{d\psi}{du} \left( \frac{d\psi}{du} \right) = \alpha \frac{d^2\psi}{du^2}
\]
(7.92)

You may be more comfortable with the following.
\[
du = \sqrt{\alpha}dx \implies dx = \frac{1}{\sqrt{\alpha}} du
\]
(7.91)
\[
\frac{d^2\psi}{dx^2} = \frac{d}{\sqrt{\alpha} du} \left( \frac{d\psi}{du} \right) = \alpha \frac{d^2\psi}{du^2}
\]
(7.92)

Either way is fine, and we now have
\[
\frac{d^2\psi}{dx^2} + (\beta - \alpha^2x^2)\psi = \alpha \frac{d^2\psi}{du^2} + (\beta - \alpha^2u^2)\psi = 0
\]
\[
\implies \frac{d^2\psi}{du^2} + \left( \frac{\beta}{\alpha} - u^2 \right) \psi = 0.
\]
(7.93)

Here comes a very inexact argument.

As $|u| \to \infty$, the differential equation behaves like
\[
\frac{d^2\psi}{du^2} - u^2 \psi = 0
\]
(7.94)
Chapter 7. Solutions of Time-Independent Schrödinger Equations in One Dimension

or

\[ \frac{d^2 \psi}{du^2} = u^2 \psi \]  

(7.95)

as \( \beta/\alpha \) is a constant.

The (mathematical) general solution of this differential equation is approximately

\[ \psi = Ae^{-u^2/2} + Be^{u^2/2} \quad \text{as} \quad |u| \to \infty. \]  

(7.96)

But, we have to set

\[ B = 0 \]  

(7.97)

to satisfy the finiteness condition. So,

\[ \psi(u) = Ae^{-u^2/2} \quad \text{as} \quad |u| \to \infty. \]  

(7.98)

We will look for a solution of the form

\[ \psi(u) = Ae^{-u^2/2}H(u); \]  

(7.99)

where \( H(u) \) must be slowly varying in comparison to \( e^{-u^2/2} \).

Compute \( \frac{d^2 \psi(u)}{du^2} \), then plug it back into the equation

\[ \frac{d^2 \psi(u)}{du^2} + \left( \frac{\beta}{\alpha} - u^2 \right) \psi(u) = 0 \]  

(7.100)

to obtain

\[ \frac{d^2 H}{du^2} - 2u \frac{dH}{du} + \left( \frac{\beta}{\alpha} - 1 \right) H = 0 \]  

(7.101)

Hence, (7.96) holds.

\[ \frac{d^2}{du^2} \left( e^{-u^2/2} \right) = \frac{d}{du} \left( -ue^{-u^2/2} \right) = -e^{-u^2/2} - u \left( -ue^{-u^2/2} \right) = (u^2 - 1)e^{-u^2/2} \xrightarrow{|u| \to \infty} u^2e^{-u^2/2} \]

\[ \frac{d^2}{du^2} \left( e^{u^2/2} \right) = \frac{d}{du} \left( ue^{u^2/2} \right) = e^{u^2/2} + u \left( ue^{u^2/2} \right) = (u^2 + 1)e^{u^2/2} \xrightarrow{|u| \to \infty} u^2e^{u^2/2} \]
as follows.

Without loss of generality, set $A = 1$, and consider

$$\psi(u) = e^{-u^2/2}H(u).$$  \hspace{1cm} (7.102)

Then, we have

$$\frac{d\psi}{du} = -ue^{-u^2/2}H(U) + e^{-u^2/2}H'(u)$$ \hspace{1cm} (7.103)

and

$$\frac{d^2\psi}{du^2} = -e^{-u^2/2}H(u) - u\left(-ue^{-u^2/2}\right)H(u) - ue^{-u^2/2}H'(u)$$

$$-ue^{-u^2/2}H'(u) + e^{-u^2/2}H''(u)$$

$$= -e^{-u^2/2}H(u) + u^2e^{-u^2/2}H(u) - 2ue^{-u^2/2}H'(u) + e^{-u^2/2}H''(u).$$ \hspace{1cm} (7.104)

Therefore,

$$\frac{d^2\psi(u)}{du^2} + \left(\frac{\beta}{\alpha} - u^2\right)\psi(u) = -e^{-u^2/2}H(u) + u^2e^{-u^2/2}H(u) - 2ue^{-u^2/2}H'(u)$$

$$+ e^{-u^2/2}H''(u) + \left(\frac{\beta}{\alpha} - u^2\right)e^{-u^2/2}H(u)$$

$$= -e^{-u^2/2}H(u) - 2ue^{-u^2/2}H'(u) + e^{-u^2/2}H''(u) + \frac{\beta}{\alpha}e^{-u^2/2}H(u) = 0.$$ \hspace{1cm} (7.105)

Dividing through by $e^{-u^2/2}$, we get

$$-H(u) - 2uH'(u) + H''(u) + \frac{\beta}{\alpha}H(u)$$

$$= H''(u) - 2uH'(u) + \left(\frac{\beta}{\alpha} - 1\right)H(u) = 0,$$ \hspace{1cm} (7.106)

which is (7.101).

We will try a power series solution for $H(u)$.

Let

$$H(u) = \sum_{l=0}^{\infty} a_l \cdot u^l = a_0 + a_1 u + a_2 u^2 + a_3 u^3 + \cdots.$$ \hspace{1cm} (7.107)
Then,
\[
u \frac{dH(u)}{du} = (a_1 + 2a_2 u + 3a_3 u^2 + 4a_4 u^3 + \cdots) u = \left( \sum_{l=0}^{\infty} (l + 1)a_{l+1} u^l \right) \cdot u
\]
\[= \sum_{l=0}^{\infty} (l + 1)a_{l+1} u^{l+1} = \sum_{l=1}^{\infty} la_l u^l \quad (7.108)
\]
and
\[
d^2H(u) \frac{du^2}{du} = 2a_2 + 2 \cdot 3a_3 u + 3 \cdot 4a_4 u^2 + \cdots = \sum_{l=0}^{\infty} (l + 1)(l + 2)a_{l+2} u^l. \quad (7.109)
\]
Plugging the above into
\[
\frac{d^2H}{du^2} - 2u \frac{dH}{du} + \left( \frac{\beta}{\alpha} - 1 \right) H = 0 \quad (7.110)
\]
gives us
\[
\sum_{l=0}^{\infty} (l + 1)(l + 2)a_{l+2} u^l - 2 \sum_{l=0}^{\infty} la_l u^l + \left( \frac{\beta}{\alpha} - 1 \right) \sum_{l=0}^{\infty} a_l u^l = 0 \quad (7.111)
\]
or
\[
\sum_{l=0}^{\infty} \left[ (l + 1)(l + 2)a_{l+2} - 2la_l + \left( \frac{\beta}{\alpha} - 1 \right) a_l \right] u^l = 0. \quad (7.112)
\]
For a power series to be zero, each term has to be zero; that is to say all the coefficients have to be zero.
We now have the following recursion relation.
\[
a_{l+2} = -\frac{(\beta/\alpha - 1 - 2l)}{(l + 1)(l + 2)}a_l \quad (7.113)
\]
The general solution is a sum of even and odd series; namely,
\[
H(u) = a_0 \left( 1 + \frac{a_2}{a_0} u^2 + \frac{a_4}{a_0} u^4 + \frac{a_6}{a_0} u^6 + \cdots \right) + a_1 \left( u + \frac{a_3}{a_1} u^3 + \frac{a_5}{a_1} u^5 + \frac{a_7}{a_1} u^7 + \cdots \right) \quad (7.114)
\]
\[
= a_0 \left( 1 + b_1 u^2 + b_2 u^4 + b_3 u^6 + \cdots \right) + a_1 u \left( 1 + c_1 u^2 + c_2 u^4 + c_3 u^6 + \cdots \right)
\]
\[
= a_0 \sum_{l \text{ even}} b_{l+2} u^l + a_1 u \sum_{l \text{ even}} c_{l+2} u^l. \quad (7.115)
\]
where
\[ b_{l/2} = \frac{a_l}{a_0} \quad \text{and} \quad c_{l/2} = \frac{a_{l+1}}{a_1} \quad \text{for} \quad l = 2, 4, 6, \ldots, \tag{7.116} \]
and the reason for the awkward notations \( b_{l/2} \) and \( c_{l/2} \) for the coefficients will become clear when compared with (7.119). What (7.114) means is that one can get all even coefficients using (7.113) if \( a_0 \) is given, and all odd coefficients can be computed if \( a_1 \) is chosen. However, there is no relation between the even and odd coefficients, and the even series and odd series of (7.114) or (7.115) are completely independent.

As \( l \to \infty \),
\[ \frac{a_{l+2}}{a_l} = \frac{(\beta/\alpha - 1 - 2l)}{(l+1)(l+2)} \to \frac{2l}{l^2} \to \frac{2}{l}. \tag{7.117} \]
So, we can see immediately that
\[ \frac{b_{l+1}}{b_{l/2}} = \frac{a_{l+2}}{a_l} \xrightarrow{l \to \infty} \frac{2}{l} \quad \text{and} \quad \frac{c_{l+1}}{c_{l/2}} = \frac{a_{l+3}}{a_{l+1}} \xrightarrow{l \to \infty} \frac{2}{l}. \tag{7.118} \]
Now compare this with the Taylor series expansion (about 0) of \( e^{u^2} \) where we change the running index from \( i \) to \( l = 2i \) and denote \( \frac{1}{(l/2)!} \) by \( d_{l/2} \).
\[ e^{u^2} = \sum_{i=0}^{\infty} \frac{(u^2)^i}{i!} = \sum_{i=0}^{\infty} \frac{u^{2i}}{i!} = \sum_{l=0}^{\text{even}} \frac{u^l}{l!} = \sum_{l=0}^{\text{even}} d_{l/2} u^l \tag{7.119} \]
Observe that
\[ \frac{d_{l+1}}{d_{l/2}} = \frac{\frac{1}{(l+1)!}}{\frac{1}{(l/2)!}} = \frac{(l/2)!}{(l/2+1)!} = \frac{(l/2)!}{(l/2+1)(l/2)!} = \frac{1}{l+1} \xrightarrow{l \to \infty} \frac{1}{l/2} = \frac{2}{l}. \tag{7.120} \]
Therefore, both the even series and the odd series of (7.115) behave like the Taylor expansion of \( e^{u^2} \) for higher order terms, or as \( l \to \infty \). Guided by this, we assume \( H(u) \) behaves like
\[ H(u) = a_0 Ke^{u^2} + a_1 K'u^{u^2}. \tag{7.121} \]

\(^3\)For example, \( \frac{a_6}{a_0} = \frac{a_6}{a_4} \frac{a_4}{a_2} \frac{a_2}{a_0} \).
So, \( e^{-u^2/2}H(u) \) of (7.102) and (7.99) behaves like
\[
a_0Ke^{u^2/2} + a_1Ke^{u^2/2} \quad \text{as} \quad |u| \to \infty. \tag{7.122}
\]
In order to satisfy the integrability condition, Condition 1 on p.127, \( H(u) \) has to terminate after some \( n \) to prevent the exponential terms in (7.122) from blowing up. This means we have to set \( \beta/\alpha = 2n + 1 \) for \( n = 0, 1, 2, 3, 4, 5, \ldots \) due to (7.113).

When \( \beta/\alpha = 2n + 1 \) we get a Hermite polynomial denoted by \( H_n(u) \). This gives us a series of eigenfunctions
\[
\psi_n(u) = A_ne^{-u^2/2}H_n(u). \tag{7.123}
\]

The first six functions look like this.
\[
\begin{align*}
n &= 0 & \psi_0 &= A_0e^{-u^2/2} \\
n &= 1 & \psi_1 &= A_1ue^{-u^2/2} \\
n &= 2 & \psi_2 &= A_2(1-2u^2)e^{-u^2/2} \\
n &= 3 & \psi_3 &= A_3(3u-2u^3)e^{-u^2/2} \\
n &= 4 & \psi_4 &= A_4(3-12u^2+4u^4)e^{-u^2/2} \\
n &= 5 & \psi_5 &= A_5(15u-20u^3+4u^5)e^{-u^2/2}
\end{align*}
\tag{7.124}
\]
Recalling that \( u = \sqrt{\alpha x}, \alpha = 2\pi m\nu/\hbar, \) and \( \beta = 2mE/\hbar^2, \) we get
\[
\beta/\alpha = \frac{2mE/\hbar^2}{2\pi m\nu/\hbar} = \frac{E}{\hbar \nu} = \frac{E}{\hbar \nu} = \frac{2E}{\hbar \nu} \tag{7.125}
\]
Setting this equal to \( 2n + 1, \)
\[
\beta/\alpha = 2n + 1 \implies \frac{2E_n}{\hbar \nu} = 2n + 1 \implies E_n = \left( n + \frac{1}{2} \right) \hbar \nu \quad n = 0, 1, 2, 3, 4, 5, \ldots \tag{7.126}
\]
Recall that
\[
\nu = \frac{1}{2\pi} \sqrt{\frac{c}{m}}. \tag{7.127}
\]
Therefore, if the force constant \( c \) and the mass \( m \) are known, we know the energy levels of this harmonic oscillator.

These are the two important facts about the simple harmonic oscillator.

\(^4\)As \( |u| \to \infty, \) higher order terms with large values of \( l \) become dominant, and this is where the series in (7.122) begin to behave like the Taylor expansion (7.119).
1. The energy levels are equally spaced. This is markedly different from the
infinite square well. This means many higher energy levels can be achieved
more easily.

2. The lowest energy possible is not zero but \( E_0 = \frac{1}{2} \hbar \nu \) for \( n = 0 \). Since vibrations
of diatomic molecules can be closely approximated by the simple harmonic
oscillator, this is another reason why the absolute zero degree temperature can
not be achieved.

When a molecule drops from one energy level to a lower level, a photon carrying
that much energy is released. On the other hand, a molecule moves to a higher
energy level if it absorbs a photon carrying the energy corresponding to the difference between the two levels. These phenomena are called \textit{emission} and \textit{absorption}
respectively.

### 7.6.2 Raising and Lowering Operators

Our primary goal in this section is to simplify the computation of the energy levels for
a simple harmonic oscillator. The beauty of this alternative approach is that we will
be able to obtain the energy levels without explicitly computing the wavefunctions.
The concept of raising and lowering operators has applications to angular momentum discussed in Chapter 9 and electron spin, a special type of angular momentum, described in Chapter 11.

From (7.84), we have

\[
-\frac{\hbar^2}{2m} \frac{d^2 \psi(x)}{dx^2} + \frac{c}{2} x^2 \psi(x) = E \psi(x).
\]  

This equation is often expressed in terms of \( \omega = 2\pi \nu \).

\[
\nu = \frac{1}{2\pi} \sqrt{\frac{c}{m}} \implies \omega = 2\pi \nu = \sqrt{\frac{c}{m}} \implies c = m\omega^2
\]

Substituting \( m\omega^2 \) for \( c \), we get
Our first step is to modify and rescale the variables. Let

\[ \hat{X} = \sqrt{\frac{m\omega}{\hbar}} x \]  

(7.131)

and

\[ \hat{P} = \frac{1}{\sqrt{m\hbar\omega}} p \]  

(7.132)

where \( x \) is the usual multiplication operator \( M_x \) and \( p = -i\hbar \frac{\partial}{\partial x} \) as given in Table 3.1. We will not use capital letters \( X \) and \( P \) in this section in order to make a clear distinction between the derived \( (\hat{X}, \hat{P}) \) and the original \( (X, P) \). \( \hat{X} \) and \( \hat{P} \) are dimensionless variables satisfying a simpler commutation relation. Indeed,

\[ \hat{X}\hat{P} - \hat{P}\hat{X} = \sqrt{\frac{m\omega}{\hbar}} x \frac{1}{\sqrt{m\hbar\omega}} p - \frac{1}{\sqrt{m\hbar\omega}} p \sqrt{\frac{m\omega}{\hbar}} x = \frac{1}{\hbar} (xp - px) \]  

(7.133)

and we have

\[ [\hat{X}, \hat{P}] = i. \]  

(7.134)

We also introduce a rescaled Hamiltonian \( \hat{H} \) defined by

\[ \hat{H} = \frac{1}{\hbar\omega} H. \]  

(7.135)

Now note that

\[ \hat{X}^2 + \hat{P}^2 = \frac{m\omega}{\hbar} x^2 + \frac{1}{m\hbar\omega} \left( -i\hbar \frac{d}{dx} \right) \left( -i\hbar \frac{d}{dx} \right) = \frac{m\omega^2}{\hbar\omega} x^2 - \frac{\hbar^2}{m\hbar\omega} \frac{d^2}{dx^2} \]

\[ = \frac{2}{\hbar\omega} \left( \frac{-\hbar^2}{2m} \frac{d^2}{dx^2} + \frac{1}{2} m\omega^2 x^2 \right) = 2 \frac{1}{\hbar\omega} H. \]  

(7.136)

Therefore,
\[ \hat{H} = \frac{1}{2} \left( \hat{X}^2 + \hat{P}^2 \right). \]  
(7.137)

We will now define a new operator \( a \) and its adjoint \( a^\dagger \), which will turn out to be "raising" and "lowering" operators in the title of this section.

\[ a = \frac{1}{\sqrt{2}} \left( \hat{X} + i \hat{P} \right) \]  
(7.138)

\[ a^\dagger = \frac{1}{\sqrt{2}} \left( \hat{X} - i \hat{P} \right) \]  
(7.139)

Going backwards, we get the following if the above relations are solved for \( \hat{X} \) and \( \hat{P} \).

\[ \hat{X} = \frac{1}{\sqrt{2}} \left( a^\dagger + a \right) \]  
(7.140)

\[ \hat{P} = \frac{i}{\sqrt{2}} \left( a^\dagger - a \right) \]  
(7.141)

Next, we will compute the commutator of \( a \) and \( a^\dagger \).

\[
\left[ \frac{1}{\sqrt{2}} (\hat{X} + i \hat{P}), \frac{1}{\sqrt{2}} (\hat{X} - i \hat{P}) \right] = \frac{1}{2} \left[ \hat{X} + i \hat{P}, \hat{X} - i \hat{P} \right]
= \frac{i}{2} \left\{ \left[ \hat{P}, \hat{X} \right] - \left[ \hat{X}, \hat{P} \right] \right\} = \frac{i}{2}(-i - i) = 1^5
\]  
(7.142)

Therefore,

\[ [a, a^\dagger] = 1. \]  
(7.143)

Now,

\[
a^\dagger a = \frac{1}{2} \left( \hat{X} - i \hat{P} \right) \left( \hat{X} + i \hat{P} \right) = \frac{1}{2} \left( \hat{X}^2 + \hat{P}^2 + i \hat{X} \hat{P} - i \hat{P} \hat{X} \right)
= \frac{1}{2} \left( \hat{X}^2 + \hat{P}^2 - 1 \right) = \hat{H} - \frac{1}{2}.
\]  
(7.144)

So,
\[ \hat{H} = a^\dagger a + \frac{1}{2}. \]  

(7.145)

On the other hand,

\[ aa^\dagger = [a, a^\dagger] + a^\dagger a = 1 + \frac{1}{2} = \hat{H} + \frac{1}{2}. \]  

(7.146)

And,

\[ \hat{H} = aa^\dagger - \frac{1}{2}. \]  

(7.147)

At this point, let us introduce another self-adjoint operator \( N \) defined by

\[ N = a^\dagger a. \]  

(7.148)

Note that \( (a^\dagger a)^\dagger = a^\dagger (a^\dagger)^\dagger = a^\dagger a \), which proves that \( N \) is self-adjoint. The operator \( N \) satisfies the following commutation relations.

\[ [N, a] = [a^\dagger a, a] = a^\dagger aa + (a^\dagger aa - a^\dagger aa) - aa^\dagger a = [a^\dagger, a]a = -a \]  

(7.149)

\[ [N, a^\dagger] = [a^\dagger a, a^\dagger] = a^\dagger aa^\dagger + (a^\dagger aa - a^\dagger aa) - a^\dagger a^\dagger a = a^\dagger [a, a^\dagger] = a^\dagger \]  

(7.150)

We have now replaced the initial eigenvalue problem for \( \hat{H} \), which is a function of \( x \) and \( p \), with that of \( N \), which is a product of \( a \) and \( a^\dagger \). Because we have

\[ H = \hbar \omega \hat{H}, \quad \hat{H} = N + \frac{1}{2} \implies H = \hbar \omega \left(N + \frac{1}{2}\right), \]  

(7.151)

\[ N \ket{n} = n \ket{n} \implies H \ket{n} = \hbar \omega \left(n + \frac{1}{2}\right) \ket{n}; \]  

(7.152)

where we are looking ahead and already using suggestive notations \( n \) and \( \ket{n} \), as opposed to the usual \( \lambda \) and \( \ket{\lambda} \). However, it is only on p.158 that we actually show that the eigenvalues of \( N \) are nonnegative integers. Relations (7.151) and (7.152) show that the eigenvalue problem for \( N \) is equivalent to the eigenvalue problem for \( \hat{H} \). When the eigenvalue problem for \( N \) is solved, we will get \( E_n = \left(n + \frac{1}{2}\right) \hbar \omega \) as
the total energy for the simple harmonic oscillator.

We will need the following three facts before solving the eigenvalue problem for $N$.

**Fact 7.1** Each eigenvalue $n$ of the operator $N$ is nonnegative.

**Proof**
Suppose $N|n\rangle = n|n\rangle$. Then,

$$\langle n|N|n\rangle = n\langle n|n\rangle. \quad (7.153)$$

On the other hand,

$$\langle n|N|n\rangle = \langle n|a^\dagger a|n\rangle = \|a|n\rangle\|^2. \quad (7.154)$$

Therefore,

$$n\langle n|n\rangle = \|a|n\rangle\|^2 \implies n = \frac{\|a|n\rangle\|^2}{\langle n|n\rangle} \geq 0. \quad (7.155)$$

**Fact 7.2** If the eigenvalue $n = 0$ for $N$, $a|0\rangle = 0$. Conversely, if $a|v\rangle = 0$, $N|v\rangle = 0|v\rangle$. For other values of $n$, i.e. $n > 0$ from Fact 7.1, $a|n\rangle$ is an eigenvector of $N$ with the eigenvalue $n - 1$.

**Proof**
Suppose $N|0\rangle = 0|0\rangle$. Then,

$$\|a|0\rangle\|^2 = \langle 0|a^\dagger a|0\rangle = \langle 0|N|0\rangle = \langle 0|0|0\rangle = 0 \implies a|0\rangle = 0. \quad (7.156)$$

Now suppose $a|v\rangle = 0$, then,

$$a|v\rangle = 0 \implies a^\dagger a|v\rangle = N|v\rangle = 0|v\rangle. \quad (7.157)$$

So, any nonzero vector $|v\rangle$ with $a|v\rangle = 0$ is an eigenvector of $N$ with the eigenvalue $n = 0$.

Next, consider a strictly positive eigenvalue $n$ of the operator $N$. Then, from (7.149),

$$[N,a]|n\rangle = Na|n\rangle - aN|n\rangle = Na|n\rangle - an|n\rangle = Na|n\rangle - na|n\rangle = -a|n\rangle$$
\[ \Rightarrow N(a|n\rangle) = na|n\rangle - a|n\rangle = (n - 1)(a|n\rangle). \]  
(7.158)

Therefore, \( a|n\rangle \) is an eigenvector of \( N \) with the eigenvalue \( n - 1 \).

**Fact 7.3** If \( |n\rangle \) is an eigenvector of \( N \) with the eigenvalue \( n \). Then, \( a^{\dagger}|n\rangle \) is an eigenvector of \( N \) with the eigenvalue \( n + 1 \).

**Proof**

We will first show \( a^{\dagger}|n\rangle \) is nonzero. From (7.143),

\[
\|a^{\dagger}|n\rangle\|^2 = \langle n|a a^{\dagger}|n\rangle = \langle n|[a, a^{\dagger}] + a^{\dagger}a|n\rangle = \langle n|1 + N|n\rangle \\
= \langle n|n\rangle + n \langle n|n\rangle = (n + 1)\||n\rangle\|^2.
\]

As \( |n\rangle \) is an eigenvector of \( N \), it is nonzero, and \( a^{\dagger}|n\rangle \) is nonzero.

Now from (7.143),

\[
[N, a^{\dagger}]|n\rangle = Na^{\dagger}|n\rangle - a^{\dagger}N|n\rangle = Na^{\dagger}|n\rangle - a^{\dagger}n|n\rangle = Na^{\dagger}|n\rangle - na^{\dagger}|n\rangle = a^{\dagger}|n\rangle \\
\Rightarrow N(a^{\dagger}|n\rangle) = (n + 1)(a^{\dagger}|n\rangle).
\]

(7.160)

Therefore, \( a^{\dagger}|n\rangle \) is an eigenvector of \( N \) with the eigenvalue \( n + 1 \).

We are now ready to show that the eigenvalues of \( N \) are non-negative integers. Suppose an eigenvalue \( \lambda \) of \( N \) is not an integer, and let \( |\lambda\rangle \) denote an associated eigenvector. From Fact 7.1, we already know that \( \lambda \) is strictly positive. From Fact 7.2, we know that \( a|\lambda\rangle \) is an eigenvector of \( N \) with the eigenvalue \( \lambda - 1 \). After repeated applications of \( a \), on \( |\lambda\rangle \), \( k \) times, we get \( a^k|\lambda\rangle \) which is an eigenvector of \( N \) with the associated eigenvalue \( \lambda - k \). As \( k \) can be as large as we wish it to be, this implies that \( N \) has strictly negative eigenvalues, contradicting Fact 7.1. Therefore, the eigenvalues of \( N \) are all nonnegative integers.

Consider an eigenvalue \( n \), which is a positive integer, and an associated eigenvector denoted by \( |n\rangle \). Then, due to Fact 7.2, \( a^k|n\rangle > \) is an eigenvector for \( k = 1, 2, \ldots, n \). When \( k = n \), the eigenvalue associated with \( a^n|n\rangle \) is 0. Then, from Fact 7.2, \( a\ (a^n|n\rangle) = 0 \), terminating the chain of eigenvectors \( |n\rangle, a|n\rangle, a^2|n\rangle, \ldots \) at \( k = n \). So, an eigenvalue \( n \) of \( N \) can only be a nonnegative integer.
On the other hand, given an eigenvector $|0\rangle$ of $N$, we can apply $a^\dagger$ to the vector any number of times, with each application generating another eigenvector whose associated eigenvalue is greater than the previous eigenvalue by 1. This is guaranteed by Fact 7.3. Therefore, the eigenvalues of $N$, $\hat{H}$, and $H$ are all nonnegative integers.

We have

$$E_n = \left(n + \frac{1}{2}\right)\hbar \omega \text{ for } n = 0, 1, 2, \ldots$$

(7.161)

as before.

### 7.6.2.1 Actions of $a$ and $a^\dagger$

Let $\{|0\rangle, |1\rangle, |2\rangle, \ldots\}$ be an orthonormal basis for the Hilbert space for $H$, $\hat{H}$, or $N$. Recall that the eigenvectors are the same for these three operators. We would like to find exactly what $a$ and $a^\dagger$ do to these basis vectors.

We already know

$$a \, |n\rangle = c_n \, |n-1\rangle$$

(7.162)

for some constant $c_n$. Taking the adjoint,

$$\langle n| a^\dagger = \langle n-1| c_n^*.$$

(7.163)

So,

$$\langle n|a^\dagger a|n\rangle = \langle n-1|n-1\rangle c_n^* c_n = c_n^* c_n \Rightarrow \langle n|N|n\rangle = c_n^* c_n$$

$$\Rightarrow \langle n|n|n\rangle = c_n^* c_n \Rightarrow n = |c_n|^2.$$  

(7.164)

We have

$$c_n = \sqrt{n} e^{i\phi}.$$  

(7.165)

By convention, we normally set $\phi = 0$. Then,

$$c_n = \sqrt{n} \text{ and } a \, |n\rangle = \sqrt{n} \, |n-1\rangle.$$  

(7.166)

---

6Because these are eigenvectors of a Hermitian operator $H$, they form a basis automatically. The only additional condition imposed here is normalization of each ket.
Similarly, we also know
\[ a^\dagger |n\rangle = d_n |n + 1\rangle \] (7.167)
for some constant \( d_n \). Taking the adjoint,
\[ \langle n| a = \langle n + 1| d_n^* . \] (7.168)
This gives us
\[ \langle n|aa^\dagger|n\rangle = \langle n + 1|d_n^*d_n|n + 1\rangle . \] (7.169)
The left-hand side can be transformed as below using the relation (7.143).
\[ \langle n|aa^\dagger|n\rangle = \langle n|[a,a^\dagger] + a^\dagger a|n\rangle = \langle n|1 + N|n\rangle = 1 + n \] (7.170)
Therefore,
\[ |d_n|^2 = n + 1. \] (7.171)
Setting the phase equal to 0 as before, we get
\[ d_n = \sqrt{n + 1} \text{ and } a^\dagger |n\rangle = \sqrt{n + 1} |n + 1\rangle . \] (7.172)
We can represent the actions of \( a \) and \( a^\dagger \) as matrices with infinitely many entries.
\[ a \leftrightarrow \begin{bmatrix} 0 & \sqrt{1} & 0 & 0 & \ldots \\ 0 & 0 & \sqrt{2} & 0 & \ldots \\ 0 & 0 & 0 & \sqrt{3} & \ldots \\ 0 & 0 & 0 & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \] (7.173)
\[ a^\dagger \leftrightarrow \begin{bmatrix} 0 & 0 & 0 & 0 & \ldots \\ \sqrt{1} & 0 & 0 & 0 & \ldots \\ 0 & \sqrt{2} & 0 & 0 & \ldots \\ 0 & 0 & \sqrt{3} & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \] (7.174)
I hope it is now clear to you why \( a \) is the lowering operator and \( a^\dagger \) is the raising operator.\(^7\)
\(^7\)Note that \( a \) and \( a^\dagger \) are also called destruction and creation operators, respectively.
Exercises

1. **This problem is the same as Chapter 5 Problem 5.**
   Consider a particle of mass $m$ and total energy $E$ which can move freely along the $x$-axis in the interval $[-\frac{a}{2}, +\frac{a}{2}]$, but is strictly prohibited from going outside this region. This corresponds to what is called an infinite square well potential $V(x)$ given by $V(x) = 0$ for $-\frac{a}{2} < x < +\frac{a}{2}$ and $V(x) = \infty$ elsewhere. If we solve the Schrödinger equation for this $V(x)$, one of the solutions is

   \[
   \Psi(x, t) = \begin{cases} 
   A \cos \left( \frac{\pi x}{a} \right) e^{-iEt/\hbar} & -\frac{a}{2} < x < \frac{a}{2} \\
   0 & |x| \geq \frac{a}{2} 
   \end{cases} \quad (a > 0).
   \]

   (a) Find $A$ so that the function $\Psi(x, t)$ is properly normalized.

   (b) In the region where the potential $V(x) = 0$, the Schrödinger equation reduces to

   \[
   -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \Psi(x, t) = i\hbar \frac{\partial \Psi(x, t)}{\partial t}.
   \]

   Plug $\Psi(x, t) = A \cos \left( \frac{\pi x}{a} \right) e^{-iEt/\hbar}$ into this relation to show $E = \frac{\pi^2 \hbar^2}{2ma^2}$.

   (c) Show that $\langle \hat{P} \rangle = 0$.

   (d) Evaluate $\langle x^2 \rangle$ for this wavefunction. You will probably need an integral table for this. (Of course, you can always try contour integration or something such. But, that is beyond this course.)

   (e) Evaluate $\langle P^2 \rangle$ for the same wavefunction. You will not need an integral table if you use the fact that the function is normalized. Of course, a brute force computation will yield the same result as well.

2. **This problem is the same as Chapter 5 Problem 6.**
   If you are a careful student who pays attention to details, you may have realized that $A$ in Problem 1 can only be determined up to the argument, or up to the sign if you assume $A$ is a real number.

   (a) What is the implication of this fact as to the uniqueness of wavefunction?

   (b) What is the implication of this fact as to the probability density $\Psi^*(x, t)\Psi(x, t)$? How about $\langle \hat{P} \rangle$, $\langle P^2 \rangle$, and $\langle x^2 \rangle$?
3. Solve the time-independent Schrödinger equation for a free particle. (Note that a free particle is a particle with no force acting on it. In terms of the potential $V(x)$, this means that it is a constant. For simplicity, assume that $V(x) = 0$ in this problem.)

4. Consider a particle of mass $m$ which can move freely along the $x$-axis anywhere from $x = -a/2$ to $x = +a/2$ for some $a > 0$, but is strictly confined to the region $[-a/2, +a/2]$. This corresponds to what is called an infinite square well potential $V(x)$ given by $V(x) = 0$ for $-a/2 < x < +a/2$ and $V(x) = \infty$ elsewhere. Where the potential is infinite, the wavefunction is zero. Use the continuity condition on $\psi(x)$ at $|x| = a/2$, but not on $\psi'(x)$, to find at least two of the solutions.

5. Give a rigorous mathematical proof that “$H(x) = K(t)$ for all pairs of real numbers $(x, t)$” implies that “$H(x)$ and $K(t)$ are a constant”.

6. Given a potential $V(x)$ with the following profile

$$V(x) = \begin{cases} 
0 & x < 0 \\
V_0 & x > 0 
\end{cases}$$

solve the Time-Independent Schrödinger Equation for a particle with the total energy $E < V_0$ traveling from $x = -\infty$ to $x = +\infty$. In fact, we have already solved this problem in this chapter. I just want you to review the procedure.

7. Show that we have total reflection in 6 above by direct computation using probability current described in Appendix J.

8. Consider the potential $V(x)$ given by $V(x) = 0$ for $x < 0$ and $V(x) = V_0 > E$ for $x > 0$, and show that we have a standing wave in the region $x < 0$ for a particle incident on the discontinuity at $x = 0$ from the left, that is, from the region characterized by negative values of $x$.

9. Consider the potential $V(x)$ given by $V(x) = 0$ for $x < 0$ and $V(x) = V_0 < E$.

   (a) Study the complete solution of the time-independent Schrödinger equation we solved in this chapter.

   (b) Show that the transmission coefficient is given by $\frac{4k_1k_2}{(k_1+k_2)^2}$. 
10. Consider a step potential given by

\[ V(x) = \begin{cases} 
0 & x < 0 \\
V_0 & x > 0 
\end{cases} \]

(a) Write down, but do not derive or solve, the time-independent Schrödinger equation for a particle of mass \( m \) and total energy \( E > V_0 \) for each region.

(b) For a particle moving toward the step from the left, the solutions are

\[ \psi(x) = Ae^{ik_1x} + Be^{-ik_1x} \quad (x < 0) \]

and

\[ \psi(x) = C e^{ik_2x} \quad (x > 0), \]

where \( k_1 = \sqrt{\frac{2mE}{\hbar}} \) and \( k_2 = \sqrt{\frac{2m(E-V_0)}{\hbar}} \). Use the two boundary conditions at \( x = 0 \) to express \( C \) in terms of \( A, k_1, \) and \( k_2 \).

(c) As it turns out, we also get \( B = \frac{k_1-k_2}{k_1+k_2} A \). Compute the reflection coefficient \( R \); that is, express it in terms of \( k_1 \) and \( k_2 \).

11. In order to solve the Time-Independent Schrödinger Equation for a particle incident on the potential barrier given by

\[ V(x) = \begin{cases} 
V_0 > E & 0 < x < a \\
0 & \text{elsewhere} 
\end{cases} \]

from the left, that is, \( x < 0 \) at \( t = 0 \), you need to consider the wavefunction given by

\[ \psi(x) = \begin{cases} 
Ae^{ik_1x} + Be^{-ik_1x} & x < 0 \\
Fe^{-k_2x} + Ge^{k_2x} & 0 < x < a \\
Ce^{ik_1x} + De^{-ik_1x} & x > a 
\end{cases} \]

with the following set of boundary conditions.

\[ \begin{align*}
\left. \psi(x) \right|_{x=0-} &= \left. \psi(x) \right|_{x=0+} \\
\left. \psi(x) \right|_{x=a-} &= \left. \psi(x) \right|_{x=a+} \\
\left. \frac{d\psi(x)}{dx} \right|_{x=0-} &= \left. \frac{d\psi(x)}{dx} \right|_{x=0+} \\
\left. \frac{d\psi(x)}{dx} \right|_{x=a-} &= \left. \frac{d\psi(x)}{dx} \right|_{x=a+}
\end{align*} \]

(a) What are \( k_1 \) and \( k_2 \) in (7.175)?
(b) Why should \( D \) in (7.175) be 0?

(c) Write down the boundary conditions (7.176) after setting \( D = 0 \).

12. Consider a particle of energy \( E > V_0 \) moving from \( x = +\infty \) to the left in the step potential:

\[
V(x) = \begin{cases} 
V_0 & x > 0 \\
0 & x < 0 
\end{cases}.
\]

(a) Write the time-independent Schrödinger equations in the regions \( x < 0 \) and \( x > 0 \).

(b) The general solutions are of the form

\[
\begin{cases}
A \times (\text{wave traveling to the right}) + B \times (\text{wave traveling to the left}) & x < 0 \\
C \times (\text{wave traveling to the right}) + D \times (\text{wave traveling to the left}) & x > 0 
\end{cases}.
\]

Complete the general solutions; i.e. substitute actual formulas for (wave traveling to the right) and (wave traveling to the left). For simplicity of notation, let \( k_1 = \sqrt{\frac{2mE}{\hbar}} \) and \( k_2 = \sqrt{\frac{2m(E-V_0)}{\hbar}} \).

(c) Determine the value of one of the constants based on the fact that there is no wave coming back from \( x = -\infty \) and traveling to the right.

(d) Use the boundary conditions at \( x = 0 \) to get the relations among the constants.

(e) Compute the reflection coefficient \( R \).

13. Consider a barrier potential given by

\[
V(x) = \begin{cases} 
0 & (x < -a) \\
V_0 & (-a < x < 0) \\
0 & (x > 0) 
\end{cases}.
\]

A stream of particles are originating at \( x = -\infty \) and traveling to the right. The total energy \( E \) is less than the barrier height \( V_0 \).

(a) Write down, but do not derive or solve, the time-independent Schrödinger equation for a particle of mass \( m \) and total energy \( E < V_0 \) for each region.

(b) For a particle moving toward the step from the left, the solutions are

\[
\psi(x) = \begin{cases} 
Ae^{ik_1x} + Be^{-ik_1x} & (x < -a) \\
Ce^{ik_2x} + De^{-k_2x} & (-a < x < 0) \\
Ge^{ik_1x} & (x > 0) 
\end{cases}.
\]
where \( k_1 = \frac{\sqrt{2mE}}{\hbar} \) and \( k_2 = \frac{\sqrt{2m(V_0 - E)}}{\hbar} \).

i. Explain in words why we have \( \psi(x) = Ge^{ik_1x} \) and not \( \psi(x) = Ge^{ik_1x} + He^{-ik_1x} \) in the region \( x > 0 \).

ii. Give the two continuity conditions at \( x = 0 \).

iii. On physical grounds, we also need continuity of \( \psi^\ast \psi \) as well as its first derivative at the boundaries. Use this continuity condition on \( \frac{d\psi^\ast \psi}{dx} \) at \( x = 0 \) to show neither \( C \) nor \( D \) is 0.

14. Show the following for the infinite square well potential described in the class.

\[
V(x) = \begin{cases} 
0 & -\frac{a}{2} < x < \frac{a}{2} \\
\infty & \text{elsewhere}
\end{cases}
\]

(a) Two general solutions \( \psi(x) = A \sin(kx) + B \cos(kx) \) and \( \psi(x) = Ce^{ikx} + De^{-ikx} \) are equivalent. Here, \( k = \sqrt{2mE}/\hbar \), and \( A, B, C, \) and \( D \) are arbitrary constants.

(b) The usual boundary conditions at \( x = 0 \) and \( a \) allow only discrete values of \( k \), which in turn implies that the total energy is discretized.

15. The potential for the infinite square well is given by

\[
V(x) = \begin{cases} 
\infty & |x| > \frac{a}{2} \\
0 & |x| < \frac{a}{2}
\end{cases}
\]

and the solutions for the time-independent Schrödinger equation given in the class were

\[
\begin{align*}
\psi_n(x) &= B_n \cos k_n x \quad \text{where} \quad k_n = \frac{n\pi}{a} \quad n = 1, 3, 5, \ldots \\
\psi_n(x) &= A_n \sin k_n x \quad \text{where} \quad k_n = \frac{n\pi}{a} \quad n = 2, 4, 6, \ldots
\end{align*}
\]

(a) Find a normalized solution for \( n = 2 \) for which the normalization constant \( A_2 \) is purely imaginary.

(b) What is the expectation value of the position \( x \) for this solution? Use the full wavefunction to solve this problem.

(c) Plug the above solution into the lefthand side of the time-independent Schrödinger equation to determine \( E_2 \), the total energy associated with this state.
16. The potential for the infinite square well is given by

\[ V(x) = \begin{cases} \infty & |x| > \frac{a}{2} \\ 0 & |x| < \frac{a}{2} \end{cases} \]

(a) Write down, but do not derive or solve, the time-independent Schrödinger equation for a particle of mass \( m \) and total energy \( E \) in the region \( -\frac{a}{2} < x < \frac{a}{2} \).

(b) Find a solution of the form \( \psi(x) = A\sin(Bx) \) for \( -\frac{a}{2} < x < \frac{a}{2} \), where \( A \) and \( B \) are constants, and compute the total energy following the steps below.

i. Express \( B \) as a function of \( E \) so that \( \psi(x) = A\sin(Bx) \) is a solution of the Shrödinger equation. (Choose the positive square root.)

ii. Find the smallest value of \( B \) that satisfies the boundary conditions.

iii. Compute the total energy corresponding to the \( B \) above.

iv. Determine \( A \) so that the wavefunction is properly normalized.

17. Consider an infinite square well whose potential is given by

\[ V(x) = \begin{cases} \infty & x < 0 \\ 0 & 0 < x < L \\ \infty & L < x \end{cases} \]

(a) Write down, but do not derive or solve, the time-independent Schrödinger equation for a particle of mass \( m \) and total energy \( E \) in the region \( 0 < x < L \).

(b) The most general solution of the time-independent Schrödinger equation in the region \( 0 < x < L \) is of the form

\[ \psi(x) = A\sin(kx) + B\cos(kx); \]

where \( A \) and \( B \) are arbitrary constants and \( k = \frac{\sqrt{2mE}}{\hbar} \).

i. Using one of the boundary conditions on \( \psi(x) \), determine the value of \( B \).

ii. Find all possible positive values \( k \) can take by applying one of the boundary conditions on \( \psi(x) \). Express \( k \) in terms of \( \pi \), \( L \), and an arbitrary natural number \( n \). We will refer to each of these values as \( k_n \). With this, we can write our wavefunction as

\[ \psi_n(x) = A_n\sin(k_n x) + B_n\cos(k_n x) \]

in order to make (potential) dependence on \( n \) more explicit.
iii. Impose the normalization condition to find the positive real value of $A_n$ given the above $k_n$.

(c) Now, use the relation
\[ k_n = \frac{\sqrt{2mE_n}}{\hbar} \]
to determine all possible values the total energy can take. In other words, express $E_n$ in terms of $m$, $\hbar$, $L$, and $n$.

18. Try a series solution
\[ f(x) = \sum_{j=0}^{\infty} a_j x^j \]
for the differential equation
\[ \frac{df(x)}{dx} + f(x) = 3x^2 + 8x + 3, \]
and show that
\[ a_j + (j + 1)a_{j+1} = 0 \quad \text{for} \quad j \geq 3. \]

19. For a simple harmonic oscillator, show that the wavefunction corresponding to $n = 1$ is given by
\[ \psi_1 = A_1 u e^{-u^2/2}, \]
and the wavefunction for $n = 2$ takes the form
\[ \psi_2 = A_2 (1 - 2u^2) e^{-u^2/2}. \]

20. A time-independent ground state wavefunction of the simple harmonic oscillator is given by $\Phi_0 = A_0 \exp[-u^2/2]$, where $u = [(Cm)^{\frac{1}{2}}/\hbar^{\frac{1}{2}}]x$. Find the expectation value of $x$ as a function of the normalization constant $A_0$. Use the full wavefunction with the total energy $E_0$.

21. This question concerns the simple harmonic oscillator.

(a) The full ground state wavefunction $\Psi_0$ can be expressed as
\[ \psi_0 e^{-iEt/\hbar} \]
with
\[ \psi_0 = A_0 e^{-u^2/2}; \]
where $A_0$ is a normalization constant, and $u = [(Cm)^{\frac{1}{2}}\hbar^{\frac{1}{2}}]x$. Compute the expectation value $\langle x \rangle$ of the position for this state.
(b) Show by explicit computation that the first two wavefunctions of a simple harmonic oscillator are orthogonal. Use the functions given below, where \( A_0 \) and \( A_1 \) are normalization constants.

\[
\psi_0 = A_0 e^{-u^2/2} \\
\psi_1 = A_1 u e^{-u^2/2}
\]

(c) Why do you know the third and the fifth eigenfunctions, denoted by \( \psi_2 \) and \( \psi_4 \) as the first eigenfunction is \( \psi_0 \), are orthogonal to each other without conducting explicit integration?

\[
\psi_2 = A_2 (1 - 2u^2) e^{-u^2/2} \\
\psi_4 = A_4 (3 - 12u^2 + 4u^4) e^{-u^2/2}
\]

(d) Explain in fewer than 50 words why a solid made of diatomic molecules cannot be cooled to absolute zero. Here, I am only asking you to reproduce the crude argument I gave you in the class as incomplete as it was.

22. Suppose \( |n\rangle \) is an eigenket of \( a^\dagger a \) with the eigenvalue \( n \). Show that \( a^\dagger |n\rangle \) is an eigenket of \( a^\dagger a \) with the eigenvalue \( n + 1 \).

23. Describe, in a sentence or two, one of the differences between classical mechanics and quantum mechanics. Your answer should be about an actual physical phenomenon or its interpretation. Hence, for example, the difference in form between the Schrödinger equation and the Newton’s equation is not an acceptable answer.
Chapter 8

Higher Spatial Dimensions

So far, we have only dealt with one dimensional cases partly for simplicity. However, most physical systems have multiple spatial dimensions, and there are a few new phenomena which you can observe only if you have more than one dimension. We will discuss two of such new features; degeneracy and angular momentum.

8.1 Degeneracy

In a typical dictionary, degeneracy is defined as a state of being degenerate, and the primary or ordinary meanings of the adjective degenerate are “Having declined, as in function or nature, from a former or original state” and “Having fallen to an inferior or undesirable state, especially in mental or moral qualities.” But, degeneracy in quantum mechanics does not have much to do with declining or becoming inferior. According to Wikipedia [Wikipedia contributors, nd], degeneracy in quantum mechanics has the following definition.

In quantum mechanics, a branch of physics, two or more different states of a system are said to be degenerate if they are all at the same energy level. It is represented mathematically by the Hamiltonian for the system having more than one linearly independent eigenstate with the same eigenvalue. Conversely, an energy level is said to be degenerate if it contains two or more different states. The number of different states at a particular energy level is called the level’s degeneracy, and this phenomenon is generally known as a quantum degeneracy.

From the perspective of quantum statistical mechanics, several degenerate states at the same level are all equally probable of being filled.
One example of this is the hydrogen atom discussed in Chapter 10. In this chapter, we will give a general proof that there is no degeneracy for “bound” states in one dimension, leaving discussions of degeneracy for specific cases to other chapters.

**Definition 8.1 (Bound States)** Bound states occur whenever the particle cannot move to infinity. That is, the particle is confined or bound at all energies to move within a finite and limited region of space which is delimited by two classical turning points.

As the particle cannot move to infinity, we need to have

\[ \psi(x \to \pm \infty) \to 0 \]  

(8.1)

for any bound state.

**Theorem 8.1** There is no degeneracy for one-dimensional bound states.

**Proof**

Suppose there is degeneracy for total energy \( E \). This means there are two unit vectors \( |u\rangle \) and \( |w\rangle \), representing distinct physical states, such that

\[ H |u\rangle = E |u\rangle \quad \text{and} \quad H |w\rangle = E |w\rangle. \]  

(8.2)

Let \( x \) be the only position variable in this one-dimensional space. Then, writing \( \psi_u(x) \) and \( \psi_w(x) \) for \( |u\rangle \) and \( |w\rangle \) respectively, we get

\[ -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \psi_u(x) + V \psi_u(x) = E \psi_u(x) \]  

(1)

and

\[ -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \psi_w(x) + V \psi_w(x) = E \psi_w(x). \]  

(2)

Multiplying (1) by \( \psi_w(x) \) and (2) by \( \psi_u(x) \) gives us

\[ \left( -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \psi_u(x) \right) \psi_w(x) + V \psi_u(x) \psi_w(x) = E \psi_u(x) \psi_w(x) \]  

(3)

and

\[ \left( -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \psi_w(x) \right) \psi_u(x) + V \psi_w(x) \psi_u(x) = E \psi_w(x) \psi_u(x). \]  

(4)
Now, subtracting (4) from (3) leads to
\[- \frac{\hbar^2}{2m} \left( \frac{d^2}{dx^2} \psi_u(x) \right) \psi_w(x) - \left( - \frac{\hbar^2}{2m} \right) \left( \frac{d^2}{dx^2} \psi_w(x) \right) \psi_u(x) = 0\]

\[\Rightarrow \left( \frac{d^2}{dx^2} \psi_u(x) \right) \psi_w(x) - \left( \frac{d^2}{dx^2} \psi_w(x) \right) \psi_u(x) = 0.\tag{5}\]

Noting that
\[
\left( \frac{d^2}{dx^2} \psi_u(x) \right) \psi_w(x) - \left( \frac{d^2}{dx^2} \psi_w(x) \right) \psi_u(x) = \left( \frac{d}{dx} \psi_u(x) \right) \frac{d}{dx} \psi_w(x) + \frac{d}{dx} \psi_u(x) \frac{d}{dx} \psi_w(x) - \left( \frac{d}{dx} \psi_w(x) \right) \frac{d}{dx} \psi_u(x) \]
\[= \frac{d}{dx} \left[ \left( \frac{d}{dx} \psi_u(x) \right) \psi_w(x) \right] - \frac{d}{dx} \left[ \left( \frac{d}{dx} \psi_w(x) \right) \psi_u(x) \right]\]
\[= \frac{d}{dx} \left[ \left( \frac{d}{dx} \psi_u(x) \right) \psi_w(x) - \left( \frac{d}{dx} \psi_w(x) \right) \psi_u(x) \right].\tag{6}\]

(5) implies
\[
\left( \frac{d}{dx} \psi_u(x) \right) \psi_w(x) - \left( \frac{d}{dx} \psi_w(x) \right) \psi_u(x) = c\tag{7}\]

where \(c\) is some constant. Because \(\psi_u\) and \(\psi_w\) are bound states,
\[
\psi_u(x) \xrightarrow{x \to \pm \infty} 0 \quad \text{and} \quad \psi_w(x) \xrightarrow{x \to \pm \infty} 0.\tag{8}\]

Therefore, the constant \(c\) in (7) is 0.\(^1\) Naively, i.e. assuming \(\psi_u\) and \(\psi_w\) are positive for now, we have
\[
\frac{1}{\psi_u(x)} \frac{d}{dx} \psi_u(x) = \frac{1}{\psi_w(x)} \frac{d}{dx} \psi_w(x) \Rightarrow \frac{d}{dx} \ln \psi_u(x) = \frac{d}{dx} \ln \psi_w(x)\]

\(^{1}\)A more rigorous proof also requires an examination of the behavior of \(\frac{d}{dx} \psi_u(x)\) and \(\frac{d}{dx} \psi_w(x)\) as \(x\) tends to \(\pm \infty\). For now, just think of exponential functions \(e^{\mp kx}\) with \(k > 0\) and their first derivatives as \(x \to \pm \infty\).
\[ \int \frac{d\ln \psi_u(x)}{dx} \, dx = \int \frac{d\ln \psi_w(x)}{dx} \, dx \implies \ln \psi_u(x) = \ln \psi_w(x) + k \]
\[ \implies \ln \psi_u(x) = \ln e^k \psi_w(x) \implies \psi_u(x) = e^k \psi_w(x). \] (9)

So, in this case, \( \psi_u(x) \) is a scalar multiple of \( \psi_w(x) \), and they represent the same physical state. This in turn implies that there is no degeneracy. If you think there is too much hand-waving in this “proof”, you are extremely right! I gave this naive argument first in order to make this proof more accessible for, say, freshmen.

Having said that, a more rigorous proof can be given resorting to the Wronskian argument described in Appendix E. According to Appendix E, two solutions, \( \psi_u \) and \( \psi_w \), of a linear second-order ordinary differential equation of the form
\[ \psi''(x) + p(x)\psi'(x) + q(x)\psi(x) = 0 \] (10)
where \( p(x) \) and \( q(x) \) are continuous, are linearly dependent if and only if the Wronskian, \( W(\psi_u, \psi_w) \), given by
\[ W(\psi_u, \psi_w)(x) = \psi_u(x)\psi'_w(x) - \psi'_u(x)\psi_w(x) \] (11)
is 0 for all \( x \) in the domain. As this is the case for us, we have \( \psi_u(x) = K\psi_w(x) \) for some constant \( K \), and this proves the theorem.

We have now shown that degeneracy, which requires two linearly independent eigenfunctions with the same energy eigenvalue \( E \), is a phenomenon specific to more than one dimension.

In passing, let us make a note of the following theorem found on p.217 of “Quantum Mechanics: Concepts and Applications (second edition)” by Nouredine Zettili [Zettili, 2009, p.217], which includes discreteness of energy levels for bound states.

**Theorem 8.2 (Discrete and Nondegenerate)** In a one-dimensional problem the energy levels of a bound state system are discrete and not degenerate.

The condition (8.1) together with Condition 3 (continuity conditions) on p.128 can be satisfied only for certain special values of energy, making allowed total energy values discrete. We saw how this works repeatedly in Chapter 7. Zettili also provides the following theorem.

**Theorem 8.3** The wave function \( \psi_n(x) \) of a one-dimensional bound state system has \( n \) nodes (i.e., \( \psi_n(x) \) vanishes \( n \) times) if \( n = 0 \) corresponds to the ground state and \((n - 1)\) nodes if \( n = 1 \) corresponds to the ground state.
Remark 8.1 (Unbound One-Dimensional States) Note that we do have degeneracy if we have unbound states. A good example is a free particle traveling wave where we have

\[ e^{i(kx-\omega t)} \text{ and } e^{-i(kx+\omega t)} \]  

(8.12)

according to (7.4).

These are linearly independent states corresponding to particles/waves traveling to the right \( e^{i(kx-\omega t)} \) and to the left \( e^{-i(kx+\omega t)} \). Because \( k = \sqrt{\frac{2mE}{\hbar}} \) for both waves, they have the same energy, and we have two-fold degeneracy.

8.2 Angular Momentum

The only momentum we encounter in a one-dimensional system is the linear momentum \( p = mv \) as we only have a linear motion in one-dimension. However, in higher dimensions, rotational motions are also possible, and we have another variety of momentum called the angular momentum. The classical definition of the angular momentum \( L \) of a particle of mass \( m \) with respect to a chosen origin is given by

\[ L = r \times p; \]

(8.13)

where \( r \) is the position vector connecting the origin and the particle, \( p = mv \) is the linear momentum, and \( \times \) is the usual vector cross product. If we denote the angle formed by \( r \) and \( p \) by \( \theta \), we have

\[ \|L\| = \|p\|\|r\| \sin \theta = m\|v\|\|r\| \sin \theta \quad \text{or} \quad L = mvr \sin \theta. \]

(8.14)

\[ ^a \text{We can choose either angle for this purpose as } \sin \theta = \sin (\pi - \theta). \]

The angular momentum is subject to the fundamental constraints of the conservation of angular momentum principle if there is no external torque, rotation-causing force, on the object. Using the usual substitutions
$x_i \leftrightarrow M_{x_i}$ and $p_i \leftrightarrow -i\hbar \frac{\partial}{\partial x_i}$; \hspace{1cm} (8.15)

where $x_1 = x$, $x_2 = y$, $x_3 = z$, and likewise for $p_i$'s, we can “derive” quantum mechanical angular momentum operators $L_x$, $L_y$, and $L_z$ drawing on the classical relation (8.13).

Because the study of the angular momentum is a large and thriving industry, we will devote the entire Chapter 9 to a detailed discussion of quantum mechanical angular momentum.
Chapter 9

Angular Momentum

As we move from one-dimensional systems to higher dimensionality, more and more interesting physics emerges. In this chapter, we will discuss how angular momentum is dealt with in quantum mechanics and the consequences of quantum mechanical computations of angular momenta. The angular momentum relations are very important in the study of atoms and nuclei. Like the total energy $E$, the magnitude of angular momentum, usually denoted by $L$, as well as its $z$-component, $L_z$ are conserved. Furthermore, angular momentum is quantized like energy levels.

9.1 Angular Momentum Operators

Quantum mechanical operators representing the $x$-, $y$-, and $z$-component of angular momentum were already presented in Table 3.2 without derivation. Now, we will see how we can “derive” quantum mechanical counterpart drawing on the classical theory of angular momentum.

The angular momentum of a particle about the origin is given by

$$L = r \times p \implies l_x \hat{i} + l_y \hat{j} + l_z \hat{k} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ x & y & z \\ p_x & p_y & p_z \end{vmatrix} \implies \begin{bmatrix} l_x \\ l_y \\ l_z \end{bmatrix} = \begin{bmatrix} yp_z - zp_y \\ zp_x - xp_z \\ xp_y - yp_x \end{bmatrix}; \quad (9.1)$$

where the expression $L = r \times p$ is coordinate-independent, but the component-by-component expressions on the right are in Cartesian coordinates. The quantum mechanical operator $L = L_x \hat{i} + L_y \hat{j} + L_z \hat{k}$ is obtained as follows.
First, noting that \( p_x \leftrightarrow -\hbar \frac{\partial}{\partial z} \), \( p_y \leftrightarrow -\hbar \frac{\partial}{\partial y} \), and \( p_z \leftrightarrow -\hbar \frac{\partial}{\partial x} \), we get:

\[
L_x = y \left( -\hbar \frac{\partial}{\partial z} \right) - z \left( -\hbar \frac{\partial}{\partial y} \right) = -\hbar \left( y \frac{\partial}{\partial z} - z \frac{\partial}{\partial y} \right) \\
L_y = z \left( -\hbar \frac{\partial}{\partial x} \right) - x \left( -\hbar \frac{\partial}{\partial z} \right) = -\hbar \left( z \frac{\partial}{\partial x} - x \frac{\partial}{\partial z} \right) \\
L_z = x \left( -\hbar \frac{\partial}{\partial y} \right) - y \left( -\hbar \frac{\partial}{\partial x} \right) = -\hbar \left( x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right)
\] (9.2)

This book culminates in a full solution of the Schrödinger equation for the hydrogen atom; where we only have a central force, and the spherical coordinate system is employed to simplify the computation. In spherical coordinates, we get:

\[
L_x = \hbar \left( \sin \phi \frac{\partial}{\partial \theta} + \cot \theta \cos \phi \frac{\partial}{\partial \phi} \right) \\
L_y = \hbar \left( -\cos \phi \frac{\partial}{\partial \theta} + \cot \theta \sin \phi \frac{\partial}{\partial \phi} \right) \\
L_z = -\hbar \frac{\partial}{\partial \phi}
\] (9.3)

Note that \( L_z \) is particularly simple. As for the magnitude of angular momentum, we will work with the squared magnitude, denoted by \( L^2 \).

\[
L^2 = L_x^2 + L_y^2 + L_z^2 = -\hbar^2 \left[ \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right]
\] (9.4)

We have the following commutation relations.

\[
[L_x, L_y] = i\hbar L_z \\
[L_y, L_z] = i\hbar L_x \\
[L_z, L_x] = i\hbar L_y \\
[L^2, L_x] = [L^2, L_y] = [L^2, L_z] = 0
\] (9.5, 9.6, 9.7, 9.8)

The relations (9.5), (9.6), and (9.7) can be summarized as

\[
L \times L = i\hbar L.
\] (9.9)
Let us verify (9.5) only, as (9.6) and (9.7) can be verified in the same manner. We will verify (9.5) in three different ways for the sake of comparison. The first verification uses (9.2). For simplicity of notation, we will denote \( \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \) and \( \frac{\partial}{\partial z} \) by \( \partial_x, \partial_y, \partial_z \), respectively. You might as well familiarize yourself with this notational scheme at this point, unless you are already comfortable with it, because it is widely used in physics and engineering.

\[
[L_x, L_y] = (-i\hbar)^2 \left[ (y\partial_z - z\partial_y)(z\partial_x - x\partial_z) - (z\partial_x - x\partial_z)(y\partial_z - z\partial_y) \right] \\
= (-i\hbar)^2 \left[ y\partial_z z\partial_x - y\partial_x z\partial_z - z\partial_y z\partial_x + z\partial_y x\partial_z \right] \\
= (-i\hbar)^2 \left[ y\partial_x + yz\partial_z - yx\partial_x^2 - z\partial_y \partial_x + zx\partial_y \partial_z - zy\partial_x \partial_z \right. \\
\left. + z^2 \partial_x \partial_y + xy\partial_x^2 - x\partial_y - xz\partial_x \partial_y \right] \\
= i\hbar(-i\hbar)(x\partial_y - y\partial_x) = i\hbar(-i\hbar) \left( x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x} \right) = i\hbar L_z \quad (9.10)
\]

Next, we will verify (9.5) in spherical coordinates.

\[
[L_x, L_y] = L_x L_y - L_y L_x \\
= i\hbar (\sin \phi \partial_\theta + \cot \theta \cos \phi \partial_\phi) i\hbar (-\cos \phi \partial_\theta + \cot \theta \sin \phi \partial_\phi) \\
\quad - i\hbar (-\cos \phi \partial_\theta + \cot \theta \sin \phi \partial_\phi) i\hbar (\sin \phi \partial_\theta + \cot \theta \cos \phi \partial_\phi) \\
= \left( i\hbar \right)^2 \left[ \sin \phi \partial_\theta ( -\cos \phi \partial_\theta ) + \sin \phi \partial_\theta (\cot \theta \sin \phi \partial_\phi) \right. + \left. \cot \theta \cos \phi \partial_\phi (-\cos \phi \partial_\theta ) + \cot \theta \cos \phi \partial_\phi (\cot \theta \sin \phi \partial_\phi) \right] \\
\quad - \left( i\hbar \right)^2 \left[ -\cos \phi \partial_\theta (\sin \phi \partial_\theta ) - \cos \phi \partial_\theta (\cot \theta \cos \phi \partial_\phi) \right. + \left. \cot \theta \sin \phi \partial_\phi (\sin \phi \partial_\theta ) + \cot \theta \sin \phi \partial_\phi (\cot \theta \cos \phi \partial_\phi) \right] \\
= \left( i\hbar \right)^2 \left[ -\sin \phi \cos \phi \partial_\theta^2 + \sin^2 \phi \frac{-1}{\sin^2 \theta} \partial_\phi + \sin^2 \phi \cot \theta \partial_\theta \partial_\phi \right. + \left. \cot \theta \cos \phi \sin \phi \partial_\theta - \cot \theta \cos^2 \phi \partial_\phi \partial_\theta + \cot^2 \theta \cos^2 \phi \partial_\phi \right. + \left. \cot^2 \theta \sin \phi \cos \phi \partial_\phi^2 \right] \\
\quad - \left( i\hbar \right)^2 \left[ -\sin \phi \cos \phi \partial_\theta^2 - \cos^2 \phi \frac{-1}{\sin^2 \theta} \partial_\phi - \cos^2 \phi \cot \theta \partial_\theta \partial_\phi \right. + \left. \cot \theta \sin \phi \cos \phi \partial_\phi + \sin^2 \phi \cot \theta \partial_\phi \partial_\theta - \cot^2 \theta \sin^2 \phi \partial_\phi \right. + \left. \cot^2 \theta \sin \phi \cos \phi \partial_\phi^2 \right] \\
= \left( i\hbar \right)^2 \left[ \sin^2 \phi \frac{-1}{\sin^2 \theta} \partial_\phi + \cot^2 \theta \cos^2 \phi \partial_\phi + \cos^2 \phi \frac{-1}{\sin^2 \theta} \partial_\phi \right. + \left. \cot^2 \theta \sin^2 \phi \partial_\phi \right]
\[ = (i\hbar)^2 \left( -\frac{1}{\sin^2 \theta} \partial \phi + \cot^2 \theta \partial \phi \right) = i\hbar(-i\hbar) \left( \frac{1}{\sin^2 \theta} - \cot^2 \theta \right) \partial \phi \]

\[ = i\hbar(-i\hbar)\frac{1-\cos^2 \theta}{\sin^2 \theta} \partial \phi = i\hbar(-i\hbar) \frac{\partial}{\partial \phi} = i\hbar L_z \quad (9.11) \]

Finally, we will use the commutator identity (I.5).

\[ [L_x, L_y] = [yp_z - zp_y, zp_x - xp_z] \]
\[ = [yp_z, zp_x] - [yp_z, xp_z] - [zp_y, zp_x] + [zp_y, xp_z] \]
\[ = y[p_z, z]p_x + z[y, p_z]p_x + zy[p_z, p_x] + [y, z]p_zp_x \]
\[ - (y[p_z, x]p_z + x[y, p_z]p_z + xy[p_z, p_z] + [y, x]p_zp_z) \]
\[ - (z[p_y, z]p_x + z[z, p_x]p_y + zz[p_y, p_x] + [z, z]p_yp_x) \]
\[ + [z, p_y, x]p_z + x[z, p_z]p_y + xz[p_y, p_z] + [z, x]p_yp_z \]
\[ = y(-i\hbar I)p_x + z \cdot 0 \cdot p_z + zy \cdot 0 + 0 \cdot p_zp_z \]
\[ - (y \cdot 0 \cdot p_z + x \cdot 0 \cdot p_z + xy \cdot 0 + 0 \cdot p_zp_z) \]
\[ - (z \cdot 0 \cdot p_x + z \cdot 0 \cdot p_y + zz \cdot 0 + 0 \cdot p_yp_x) \]
\[ + z \cdot 0 \cdot p_z + x(i\hbar I)p_y + xz \cdot 0 + 0 \cdot p_yp_z \]
\[ = i\hbar(xp_y - yp_x) = i\hbar L_z \quad (9.12) \]

Now, we will verify (9.8), using (1.4) of Appendix I; namely, \([AB, C] = [A, C]B + A[B, C].\)

\[ [L^2, L_x] = [L^2_x + L^2_y + L^2_z, L_x] = [L^2_x, L_x] + [L^2_y, L_x] + [L^2_z, L_x] \]
\[ = 0 + [L_y, L_x]L_y + L_y[L_y, L_x] + [L_z, L_x]L_z + L_z[L_z, L_x] \]
\[ = -i\hbar L_z L_y - i\hbar L_y L_z + i\hbar L_y L_z + i\hbar L_z L_y = 0 \quad (9.13) \]

Likewise for \(L_y\) and \(L_z.\)

### 9.2 Quantum Mechanical Rotation Operator \(U_R\)

Let \(R(\phi k)\) be a physical rotation, a counterclockwise rotation around the \(z\)-axis by \(\phi\), and \(U_R\) be the quantum mechanical operator associated with the physical rotation \(R.\) Then, the action of \(U_R\) is characterized by

\[ U_R |x, y, z\rangle = |x \cos \phi - y \sin \phi, x \sin \phi + y \cos \phi, z\rangle \quad (9.14) \]

In order to derive \(U_R\), we will first consider an infinitesimal physical rotation
$R(\varepsilon \mathbf{k})$ and its quantum mechanical counterpart $U_{R\varepsilon}$ given to first order in $\varepsilon$ by

$$U_{R\varepsilon} = I + \varepsilon \Omega. \quad (9.15)$$

As it turns out, the correct form is

$$U_{R\varepsilon} = I - \frac{i\varepsilon}{\hbar} L_z; \quad (9.16)$$

where $L_z$ is the quantum mechanical angular momentum operator around the $z$-axis (9.2). We can convince ourselves that (9.16) is indeed the desired quantum mechanical operator as follows. Note that this is not a very rigorous proof, but a rough sketch of how it should be shown.

By definition, we have

$$U_{R\varepsilon} |x, y, z\rangle = |x - y\varepsilon, y + x\varepsilon, z\rangle. \quad (9.17)$$

It follows from (9.17) that

$$U_{R\varepsilon} |\psi\rangle = U_{R\varepsilon} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |x, y, z\rangle \langle x, y, z| \psi\rangle \, dx \, dy \, dz$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} U_{R\varepsilon} |x, y, z\rangle \langle x, y, z| \psi\rangle \, dx \, dy \, dz$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |x - y\varepsilon, y + x\varepsilon, z\rangle \langle x, y, z| \psi\rangle \, dx \, dy \, dz. \quad (9.18)$$

Now, let $x' = x - y\varepsilon$, $y' = y + x\varepsilon$, and $z' = z$. Then, the Jacobian is

$$\begin{vmatrix}
\frac{\partial x'}{\partial x} & \frac{\partial x'}{\partial y} & \frac{\partial x'}{\partial z} \\
\frac{\partial y'}{\partial x} & \frac{\partial y'}{\partial y} & \frac{\partial y'}{\partial z} \\
\frac{\partial z'}{\partial x} & \frac{\partial z'}{\partial y} & \frac{\partial z'}{\partial z}
\end{vmatrix}^{-1} = \begin{vmatrix}
1 & \varepsilon & 0 \\
-\varepsilon & 1 & 0 \\
0 & 0 & 1
\end{vmatrix}^{-1} \approx \frac{1}{1 + \varepsilon^2} \approx 1 - \varepsilon^2 = 1 \quad (9.19)$$
to first order in $\varepsilon$. So,

$$U_{R\varepsilon} |\psi\rangle = \iiint_{-\infty}^{+\infty} |x', y', z'\rangle \langle x, y, z' | \psi\rangle \frac{1}{1 + \varepsilon^2} dx' dy' dz'$$

$$= \iiint_{-\infty}^{+\infty} |x', y', z'\rangle \langle x, y, z | \psi\rangle dx' dy' dz'$$

$$= \iiint_{-\infty}^{+\infty} |x', y', z'\rangle \langle x' + y\varepsilon, y' - x\varepsilon, z' | \psi\rangle dx' dy' dz'$$

(9.20)

to first order in $\varepsilon$. Hence,

$$\langle x, y, z | U_{R\varepsilon} |\psi\rangle = \langle x, y, z | \iint_{-\infty}^{+\infty} |x', y', z'\rangle \langle x' + y\varepsilon, y' - x\varepsilon, z' | \psi\rangle dx' dy' dz'$$

$$= \iint_{-\infty}^{+\infty} \langle x, y, z | x', y', z'\rangle \langle x' + y\varepsilon, y' - x\varepsilon, z' | \psi\rangle dx' dy' dz'$$

$$= \int_{-\infty}^{+\infty} \delta(x - x')\delta(y - y')\delta(z - z') \langle x' + y\varepsilon, y' - x\varepsilon, z' | \psi\rangle dx' dy' dz'$$

$$= \langle x + y\varepsilon, y - x\varepsilon, z | \psi\rangle .$$

(9.21)

So, we have

$$\Rightarrow \langle x, y, z | U_{R\varepsilon} |\psi\rangle = (U_{R\varepsilon} |\psi\rangle) (x, y, z) = \psi(x + y\varepsilon, y - x\varepsilon, z).$$

(9.22)

Expanding $\psi(x + y\varepsilon, y - x\varepsilon, z)$ in a Taylor series to order $\varepsilon$, we obtain

$$\psi(x + y\varepsilon, y - x\varepsilon, z) = \psi(x, y, z) + \frac{\partial \psi}{\partial x}(y\varepsilon) + \frac{\partial \psi}{\partial y}(-x\varepsilon).$$

(9.23)

On the other hand,

$$\langle x, y, z | U_{R\varepsilon} |\psi\rangle = \langle x, y, z | I + \varepsilon\Omega |\psi\rangle = \langle x, y, z |\psi\rangle + \varepsilon \langle x, y, z |\Omega |\psi\rangle$$

$$= \psi(x, y, z) + \varepsilon \langle x, y, z |\Omega |\psi\rangle .$$

(9.24)
Therefore,

\[ \langle x, y, z | U_{R\varepsilon} | \psi \rangle = \psi(x + y\varepsilon, y - x\varepsilon, z) \]

\[ \Rightarrow \psi(x, y, z) + \varepsilon \langle x, y, z | \Omega | \psi \rangle = \psi(x, y, z) + \frac{\partial \psi}{\partial x}(y\varepsilon) + \frac{\partial \psi}{\partial y}(-x\varepsilon) \]

\[ \Rightarrow \varepsilon \langle x, y, z | \Omega | \psi \rangle = \varepsilon \left[ \frac{\partial \psi}{\partial x} + (-x) \frac{\partial \psi}{\partial y} \right] \]

\[ \Rightarrow (\Omega | \psi ) (x, y, z) = \left[ \left( \frac{\partial}{\partial x} - x \frac{\partial}{\partial y} \right) \psi \right] (x, y, z) \]

\[ \Rightarrow \Omega = y \frac{\partial}{\partial x} - x \frac{\partial}{\partial y} \Rightarrow i\hbar \Omega = i\hbar \left( y \frac{\partial}{\partial x} - x \frac{\partial}{\partial y} \right) \]

\[ \Rightarrow -i\hbar \left( x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right) = L_z \Rightarrow \Omega = \frac{L_z}{i\hbar} = -\frac{i}{\hbar} L_z \]  \hspace{1cm} (9.25)

This verifies (9.16).

\[ U_{R\varepsilon} = I - \frac{i\varepsilon}{\hbar} L_z \]  \hspace{1cm} (9.26)

The quantum mechanical operator \( U_R \) for a finite rotation \( R(\phi_0 k) \) can be constructed from (9.16) as follows.

\[ U_R = \lim_{N \to \infty} \left( I - \frac{i \phi_0}{\hbar N} L_z \right)^N = \exp \left( -i\phi_0 L_z / \hbar \right) \]  \hspace{1cm} (9.27)

The exponential here is to be regarded as a convergent power series.

\[ \exp \left( -i\phi_0 L_z / \hbar \right) = \sum_{n=0}^{\infty} \frac{(-i\phi_0 / \hbar)^n}{n!} L_z^n \]  \hspace{1cm} (9.28)

Now, from (9.3),

\[ L_z = -i\hbar \frac{\partial}{\partial \phi} \]  \hspace{1cm} (9.29)

So, in spherical coordinates,

\[ U_R = \exp \left( -i\phi_0 L_z / \hbar \right) = \exp \left( -i\phi \left( -i\hbar \frac{\partial}{\partial \phi} \right) / \hbar \right) = \exp \left( -\phi \frac{\partial}{\partial \phi} \right) \]  \hspace{1cm} (9.30)
9.3 Rotationally Invariant Hamiltonian

Suppose the only force in the system is a central force \( F(r) \), which derives from a central potential \( V(r) \) where \( r \) is the distance from the origin, so that

\[
F = -\nabla V(r) = -\frac{dV(r)}{dr} \hat{r}.
\]  

(9.31)

Then, the Hamiltonian

\[
H = \frac{p^2}{2m} + V(r)
\]  

(9.32)

is invariant under arbitrary rotations, and it commutes with \( L_x, L_y, L_z \), and \( L^2 \). From (I.20) and (I.21), we have the following.

\[
[H, L_x] = [H, L_y] = [H, L_z] = 0 \quad \text{and} \quad [H, L^2] = 0
\]  

(9.33)

Therefore, \( L^2, L_x, L_y, \) and \( L_z \) are constants of motion and conserved. We know from Theorem 2.9 that commuting Hermitian operators \( \Omega \) and \( \Gamma \) have common eigenvectors which diagonalize \( \Omega \) and \( \Gamma \) simultaneously.

Because \( L_x, L_y, \) and \( L_z \) do not commute with each other, we cannot diagonalize \( H, L^2, L_x, L_y, \) and \( L_z \) simultaneously. The best we can do is to diagonalize \( H, L^2 \), and only one of the components of \( L \). It is customary to choose \( L_z \). So, we will diagonalize \( H, L^2 \), and \( L_z \) simultaneously \(^1\), which amounts to finding all eigenfunctions \( \{\psi(r, \theta, \phi)\} \) which are, by definition, stationary under \( H, L^2, \) and \( L_z \).

The explicit forms of the eigenfunctions will be found in Chapter 10. But, there is a way to solve the eigenvalue problem of \( L^2 \) and \( L_z \) using raising and lowering operators as for the simple harmonic oscillator discussed in Section 7.6.2.

\(^1\) We can show that \( \frac{dL}{dt} = 0 \) if \( V = V(r) \) within the framework of classical mechanics as well. Namely, we have

\[
L = r \times p \implies \frac{dL}{dt} = \frac{dr}{dt} \times p + r \times \frac{dp}{dt} = \frac{dr}{dt} \times p + r \times m \frac{dv}{dt} = v \times p + r \times m \alpha = v \times p + r \times F
\]

\[
= v \times m v + r \times \left( -\frac{dV(r)}{dr} \frac{r}{|r|} \right) = m(v \times v) - \frac{dV(r)}{dr} \frac{1}{|r|} (r \times r) = 0.
\]  

(9.34)

\(^2\) We can draw a contradiction if we assume there is a common eigenstate \( |l\rangle \) for \( L_z \) and \( L_x \) such that

\[
L_x |l\rangle = l_x |l\rangle \quad \text{and} \quad L_z |l\rangle = l_z |l\rangle.
\]
9.4 Raising and Lowering Operators: $L_+$ and $L_-$

In this section, we will find the eigenvalues of $L^2$ and $L_z$ without finding the eigenfunctions explicitly. Looking ahead, we will denote the eigenvalues of $L^2$ by $\hbar^2(l+1)$ and those of $L_z$ by $m_l\hbar$. We will write $\{|l, m_l\rangle\}$ for the orthonormal eigenstates with associated eigenvalues $\hbar^2(l+1)$ and $m_l\hbar$.

$$L^2 |l, m_l\rangle = \hbar^2(l+1) |l, m_l\rangle$$
$$L_z |l, m_l\rangle = m_l\hbar |l, m_l\rangle$$
$$\langle l, m_l|m_l'|l'\rangle = \delta_{ll'}\delta_{m_l,m_{l'}}$$

(9.35)

We will show that $l$ is a nonnegative integer and $m_l = -l, -l+1, \ldots, 0, \ldots, l-1, l$.

The number $l$ is referred to as an orbital quantum number, and $m_l$ is called either a magnetic quantum number or an azimuthal quantum number. We first define the raising and lowering operators as follows. The reason for the naming will become clear shortly.

$$L_+ = L_x + iL_y$$
$$L_- = L_x - iL_y$$

(9.36)
(9.37)

Because $L_x$ and $L_y$ are Hermitian, it follows that

$$L_+^\dagger = (L_x + iL_y)^\dagger = L_x^\dagger - iL_y^\dagger = L_x - iL_y = L_-$$

(9.38)

This implies

$$[L_z, L_x] |l\rangle = i\hbar L_y |l\rangle = L_z L_x |l\rangle - L_x L_z |l\rangle = l_z l_x |l\rangle - l_x l_z |l\rangle = 0.$$

This in turn implies

$$[L_x, L_y] |l\rangle = i\hbar L_z |l\rangle = L_x L_y |l\rangle - L_y L_x |l\rangle = 0 - L_y (l_x |l\rangle) = 0 - 0 = 0.$$

We now have

$$L_y |l\rangle = L_z |l\rangle = 0 \implies L_x |l\rangle = 0.$$

Therefore, only possible common states are those for zero angular momentum. Needless to say, we cannot find a set of orthonormal eigenstates that simultaneously diagonalizes $L_x, L_y$, and $L_z$.

$^3$As we will see later, the energy levels among the states with different $m_l$’s are the same in the absence of an external magnetic field.

$^4$It is so called as the angle $\theta$ in spherical coordinate system is called an azimuth.
and
\[ L_\pm = (L_x - iL_y)^\pm = L_x^\pm + iL_y^\pm = L_x + iL_y = L_\pm. \] (9.39)

Incidentally, using (9.3), we can express \( L_+ \) and \( L_- \) in spherical coordinates.

\[ L_\pm = \hbar e^{\pm i\phi} \left( \pm \frac{\partial}{\partial \theta} + i \cot \theta \frac{\partial}{\partial \phi} \right) \] (9.40)

As we already know \( L^2 \) commutes with \( L_x \) and \( L_y \), we have

\[ [L^2, L_+] = [L^2, L_-] = 0. \] (9.41)

Let us also compute other commutators.

\[ [L_+, L_-] = [L_x + iL_y, L_x - iL_y] = [L_x, L_x] + [L_x, -iL_y] \]
\[ + [iL_y, L_x] + [iL_y, -iL_y] \]
\[ = 0 + (-i)\hbar L_z + i(-i\hbar L_z) + 0 = 2\hbar L_z \] (9.42)

\[ [L_+, L_-] = [L_x + iL_y, L_x - iL_y] = [L_x, L_z] + i[L_y, L_z] = -i\hbar L_y + i(i\hbar L_x) \]
\[ = -\hbar L_x - \hbar L_x = -\hbar(L_x + iL_y) = -\hbar L_+ \] (9.43)

\[ [L_-, L_z] = [L_x - iL_y, L_z] = [L_x, L_z] - i[L_y, L_z] = -i\hbar L_y - i(i\hbar L_x) \]
\[ = \hbar L_x - i\hbar L_y = \hbar(L_x - iL_y) = \hbar L_- \] (9.44)

We also have the following.

\[ L_+ L_- = (L_x + iL_y)(L_x - iL_y) = L_x^2 - iL_x L_y + iL_y L_x + L_y^2 \]
\[ = L^2 - L_z^2 - i[L_x, L_y] = L^2 - L_z^2 - i(i\hbar L_z) = L^2 - L_z^2 + \hbar L_z \] (9.45)

\[ L_- L_+ = (L_x - iL_y)(L_x + iL_y) = L_x^2 + iL_x L_y - iL_y L_x + L_y^2 \]
\[ = L^2 - L_z^2 + i[L_x, L_y] = L^2 - L_z^2 + i(i\hbar L_z) = L^2 - L_z^2 - \hbar L_z \] (9.46)

Because the operators \( L_x, L_y, \) and \( L_z \) are all Hermitian,

\[ \langle l, m_l|L^2|l, m_l\rangle = \langle l, m_l|L_x^2 + L_y^2 + L_z^2|l, m_l\rangle \]
\[ = \langle l, m_l|L_x^2|l, m_l\rangle + \langle l, m_l|L_y^2|l, m_l\rangle + \langle l, m_l|L_z^2|l, m_l\rangle \]
\[ = \langle l, m_l|L_x^2|l, m_l\rangle + \langle l, m_l|L_y^2|l, m_l\rangle + \langle l, m_l|L_z^2|l, m_l\rangle \]
9.4. RAISING AND LOWERING OPERATORS: \( L_+ \) AND \( L_- \)

\[
\begin{align*}
&= \langle L_x(l, m_l) | L_x(l, m_l) \rangle + \langle L_y(l, m_l) | L_y(l, m_l) \rangle \\
&\quad + \langle L_z(l, m_l) | L_z(l, m_l) \rangle \\
&= \| L_x(l, m_l) \| ^2 + \| L_y(l, m_l) \| ^2 + \| L_z(l, m_l) \| ^2 \geq 0
\end{align*}
\]

On the other hand,

\[
\langle l, m_l | L^2 | l, m_l \rangle = \hbar^2 l(l + 1) \langle l, m_l | l, m_l \rangle = \hbar^2 l(l + 1).
\]

Therefore,

\[
\hbar^2 l(l + 1) \geq 0 \implies l \geq 0 \text{ or } l \leq -1.
\]

Because the parabola \( y = x(x + 1) \) is symmetric about the axis \( x = -\frac{1}{2} \), we can choose \( l \geq 0 \) and still get all possible nonnegative values \( l(l + 1) \) can take.

Now, let

\[
|+\rangle := L_+ | l, m_l \rangle;
\]

where \( | l, m_l \rangle \) is a normalized eigenstate of \( L^2 \) with the eigenvalue \( \hbar^2 l(l + 1) \) and also a normalized eigenstate of \( L_z \) with the eigenvalue \( m_l \) from (9.35). Then,

\[
L^2 |+\rangle = L^2 L_+ | l, m_l \rangle = L_+ L^2 | l, m_l \rangle = L_+ \left( \hbar^2 l(l + 1) | l, m_l \rangle \right) \\
= \hbar^2 l(l + 1) L_+ | l, m_l \rangle = \hbar^2 l(l + 1) |+\rangle.
\]

Hence, \( |+\rangle \) is still an eigenstate of \( L^2 \) with the same eigenvalue \( \hbar^2 l(l + 1) \). On the other hand, due to (9.35) and (9.43),

\[
L_z |+\rangle = L_z L_+ | l, m_l \rangle = ([L_z, L_+] + L_+ L_z) | l, m_l \rangle \\
= (\hbar L_+ + m_l \hbar L_+) | l, m_l \rangle = (m_l + 1) L_+ | l, m_l \rangle = (m_l + 1) \hbar |+\rangle.
\]

Hence, \( |+\rangle \) is still an eigenstate of \( L_z \) but with an eigenvalue which is \( \hbar \) greater than the original \( m_l \hbar \). Likewise, if we let

\[
|\rangle := L_- | l, m_l \rangle,
\]

we get
\[ L^2 |−\rangle = L^2 L_- |l, m_t\rangle = L_- L^2 |l, m_t\rangle = L_- (\hbar^2 l(l + 1) |l, m_t\rangle) \]
\[ = \hbar^2 l(l + 1)L_- |l, m_t\rangle = \hbar^2 l(l + 1) |−\rangle \]  \hfill (9.54)

\[ L_z |−\rangle = L_z L_- |l, m_t\rangle = ([L_z, L_-] + L_- L_z) |l, m_t\rangle \]
\[ = (−\hbar L_- + m_t \hbar L_-) |l, m_t\rangle = (m_t − 1)L_- |l, m_t\rangle \]
\[ = (m_t − 1)\hbar |−\rangle . \]  \hfill (9.55)

This shows that |−\rangle is an eigenstate of \( L^2 \) with an eigenvalue \( \hbar^2 l(l + 1) \) and is an eigenstate of \( L_z \) with an eigenvalue that is \( \hbar \) smaller than the original \( m_t \hbar \). In sum, we have shown that \( L_+ / L_- \) raises/lowers the magnetic quantum number \( m_t \) by one, but preserves the total angular momentum \( l \). In other words, \( L_+ |l, m_t\rangle \) is proportional to \( |l, m_t + 1\rangle \), and \( L_- |l, m_t\rangle \) is proportional to \( |l, m_t - 1\rangle \).

Let us now compute the square of the norm for |+\rangle and |−\rangle, respectively.
\[ \| |−\rangle \|^2 = \langle −|−\rangle = \langle L_- (l, m_t)|L_- (l, m_t) \rangle \]
\[ = \langle l, m_t|L_- L_- |l, m_t\rangle = \langle l, m_t|L^2 - L_z^2 - \hbar L_z |l, m_t\rangle \]
\[ = (\hbar^2 l(l + 1) - \hbar^2 m_t^2 - \hbar^2 m_t) \langle l, m_t|l, m_t\rangle = \hbar^2(l(l + 1) - m_t(m_t + 1)) \]
\[ = \hbar^2(l^2 + l - m_t^2 - m_t) = \hbar^2((l + m_t)(l - m_t) + (l - m_t)) \]
\[ = \hbar^2(l - m_t)(l + m_t + 1) \]  \hfill (9.56)

\[ \| |−\rangle \|^2 = \langle −|−\rangle = \langle L_- (l, m_t)|L_- (l, m_t) \rangle \]
\[ = \langle l, m_t|L_- L_- |l, m_t\rangle = \langle l, m_t|L^2 - L_z^2 + \hbar L_z |l, m_t\rangle \]
\[ = (\hbar^2 l(l + 1) - \hbar^2 m_t^2 + \hbar^2 m_t) \langle l, m_t|l, m_t\rangle = \hbar^2(l(l + 1) - m_t(m_t - 1)) \]
\[ = \hbar^2(l^2 + l - m_t^2 + m_t) = \hbar^2((l + m_t)(l - m_t) + (l + m_t)) \]
\[ = \hbar^2(l + m_t)(l - m_t + 1) \]  \hfill (9.57)

Because we have to have \( \| |+\rangle \|^2 \geq 0 \) and \( \| |−\rangle \|^2 \geq 0 \), from (9.56) and (9.57), we can see that
\[ \hbar^2(l - m_t)(l + m_t + 1) \geq 0 \]  \hfill (9.58)
9.4. RAISING AND LOWERING OPERATORS: $L_+$ AND $L_-$

\[ h^2(l + m)(l - m + 1) \geq 0. \]  

(9.59)

The inequality (9.58) requires either

I. \( l \geq m \) and \( l \geq -m - 1 \)

or

II. \( l \leq m \) and \( l \leq -m - 1 \);

while the other inequality (9.59) implies either

III. \( l \geq -m \) and \( l \geq m - 1 \)

or

IV. \( l \leq -m \) and \( l \leq m - 1 \).

Therefore, we should have either condition I or condition II holding true, and at the same time, one of conditions III and IV to be true.

First, suppose \( m \geq 0 \). As specified on p.185, we chose \( l \geq 0 \). If \( m \) is nonnegative, \( l \leq -m - 1 \) is impossible. Hence, condition I should be satisfied. In particular, we need \( l \geq m \) as \( l \geq -m - 1 \) is automatically satisfied in this case. Consider conditions III and IV next. Condition III reduces to \( l \geq m - 1 \), and condition IV is impossible because \( l \leq m \) and \( l \leq m - 1 \) implies \( 2l \leq -1 \) which contradicts \( l \geq 0 \). So, we have to satisfy \( l \geq m \) and \( l \geq m - 1 \) simultaneously. Therefore,

\[ m \leq l \quad \text{if} \quad m \geq 0. \]  

(9.60)

Next, suppose \( m < 0 \). Condition I reduces to \( l \geq -m - 1 \), and condition II is impossible as \( l > 0 > m \). Now, condition III reduces to \( l \geq -m \) as \( l \geq m - 1 \) is automatically satisfied, and condition IV is impossible because \( l \geq 0 > m > m - 1 \). Hence, we require \( l \geq -m - 1 \) and \( l \geq -m \). Therefore,

\[ m \geq -l \quad \text{if} \quad m < 0. \]  

(9.61)

The inequalities (9.60) and (9.61) together with the fact that \( L_+/L_- \) raises/lowers the magnetic quantum number \( m \) by one imply
\[ m_l = -l, -l + 1, \ldots, l - 1, +l; \text{ where } l \geq 0. \] (9.62)

As the spacing for each pair of neighboring \( m_l \) values is 1, it is necessary that \( l - (-l) = 2l \) is a nonnegative integer. But, this is possible only for an integral or half-integral \( l \). It turns out \( l \) is an integer for the orbital angular momentum.

Before closing this section, we will briefly discuss a simple implication of (9.51), (9.52), and (9.56), which will be useful later. A similar result follows from (9.54), (9.55), and (9.57). From (9.51), we can see that \( L_+|l, m_l\rangle \) is an eigenvector of \( L^2 \) with the eigenvalue \( \hbar^2 l(l + 1) \). Furthermore, (9.52) indicates that \( L_+|l, m_l\rangle \) is an eigenvector of \( L_z \) with the eigenvalue \( (m_l + 1)\hbar \). Hence, we have

\[
L_+|l, m_l\rangle = C|l, m_l + 1\rangle
\] (9.63)

for some scalar \( C \). However, we already know from (9.56) that

\[
||L_+|l, m_l\rangle||^2 = \langle l, m_l + 1|C^*C|l, m_l + 1\rangle = |C|^2 = \hbar^2(l - m_l)(l + m_l + 1),
\] (9.64)

and so,

\[
C = e^{i\theta}\hbar \sqrt{(l - m_l)(l + m_l + 1)}.
\] (9.65)

According to standard convention, we choose \( \theta = 0 \) or \( e^{i\theta} = 1 \) [Shankar, 1980, p.336] to get

\[
C = \hbar \sqrt{(l - m_l)(l + m_l + 1)} \implies
\]

\[
L_+|l, m_l\rangle = \hbar \sqrt{(l - m_l)(l + m_l + 1)}|l, m_l + 1\rangle.
\] (9.66)

Similarly, we can also show

\[
L_-|l, m_l\rangle = \hbar \sqrt{(l + m_l)(l - m_l + 1)}|l, m_l - 1\rangle.
\] (9.67)

Finally, noting that \(|l, -l\rangle, |l, -l + 1\rangle, \ldots, |l, l - 1\rangle, \text{ and } |l, l\rangle\) are the only eigenvectors of \( L_+\) from (9.62), we have to have

\(^5\)Half-integral \( l \)'s correspond to an internal or intrinsic angular momentum called the spin of a particle. We will discuss electron spin in Chapter 11.
\[ L_+ |l, l\rangle = 0 \]  
and  
\[ L_- |l, -l\rangle = 0 \]

(9.68)  
(9.69)

as in Section 7.6.2.

### 9.5 Matrix Representations of \( L^2 \) and \( L_z \)

In the infinite-dimensional basis of the orthonormal eigenstates \( \{|l, m_l\} \), where \( l = 0, 1, 2, 3, \ldots \) and \( m_l = -l, -l + 1, -l + 2, \ldots, 0, \ldots, l - 2, l - 1, l \) for each \( l \), \( L^2 \) and \( L_z \) are simultaneously diagonal. Needless to say, the matrices have infinitely many rows and columns.

\[
L^2 = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
0 & 0 & 2h^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
0 & 0 & 2h^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
0 & 0 & 0 & 6h^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
0 & 0 & 0 & 6h^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
0 & 0 & 0 & 0 & 6h^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
0 & 0 & 0 & 0 & 6h^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots
\end{bmatrix}
\]

(9.70)
The diagonal blocks correspond to $l = 0, l = 1, l = 2, \ldots$ starting from the trivial 1-by-1 block in the upper left-hand corner. For $L^2$, the diagonal entries for the $l$-th block are the same and given by $l(l + 1)\hbar^2$, while they range from $-lh$ to $+lh$ for $L_z$. Here, $L^2$ is exhibiting “degeneracy” discussed in Section 8.1 on p. 169. We have $(2l + 1)$-fold degeneracy for each diagonal. To the contrary, $L_z$ is not degenerate in each diagonal as the eigenvalues range from $-lh$ to $+lh$. For example, consider the $l = 2$ block. The eigenstates, $|l, m_l\rangle$’s $= |2, m_l\rangle$’s with $m_l = -2, -1, 0, +1, +2$, are given by

\[ |2, -2\rangle = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix}, \quad |2, -1\rangle = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix}, \quad |2, 0\rangle = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix}, \quad |2, +1\rangle = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix}, \quad |2, +2\rangle = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix}. \]
For these states, we have

\begin{align}
L^2 |2, -2\rangle &= L^2 |2, -1\rangle = L^2 |2, 0\rangle = L^2 |2, +1\rangle = L^2 |2, +2\rangle = 2(2 + 1)\hbar^2 = 6\hbar^2, \\
\end{align}

and

\begin{align}
L_z |2, -2\rangle &= -2\hbar, L_z |2, -1\rangle = -\hbar, L_z |2, 0\rangle = 0, L_z |2, +1\rangle = +\hbar, \\
\text{and } L_z |2, +2\rangle &= +2\hbar.
\end{align}

9.6 Generalized Angular Momentum \( J \)

As shown on p.176, the three operators associated with the components of an arbitrary classical angular momentum, \( L_x, L_y, \) and \( L_z \), satisfy the commutation relations (9.5), (9.6), and (9.7). These relations originate from the geometric properties of rotations in three-dimensional space. However, in quantum mechanics it is sometimes necessary to consider more abstract “angular momentum” with no counterpart in classical mechanics. What we do is to go backwards, from the commutation relations to “rotations”, and define quantum mechanical angular momentum \( J \) as any set of three observables/Hermitian operators \( J_x, J_y, \) and \( J_z \) which satisfies the same commutation relations as \( L_x, L_y, \) and \( L_z \). Namely,

\[ [J_x, J_y] = i\hbar J_z, \]
\[ [J_y, J_z] = i\hbar J_x, \]
\[ [J_z, J_x] = i\hbar J_y, \]
\[ \text{and} \]
\[ [J^2, J_x] = [J^2, J_y] = [J^2, J_z] = 0; \]

where (9.78) follows from (9.75), (9.76), and (9.77) as in (9.13) on p.178. Note that \( J^2 = J_x^2 + J_y^2 + J_z^2 \) (a scalar) by definition, and (9.78) implies \([J^2, J] = 0\). With this theoretical framework, quantum mechanical angular momentum is based entirely on the commutation relations (9.75), (9.76), and (9.77).

If you examine the computations we conducted for \( L, L^2, L_x, L_y, \) and \( L_z \), you can see that we only used the commutation relations (9.5), (9.6), and (9.7), which are equivalent to (9.75), (9.76), and (9.77). Therefore, all the results we obtained for \( L \) apply to \( J \) without modification. In particular, we can define raising and lowering operators by
\[ J_+ = J_x + iJ_y \]  
and  
\[ J_- = J_x - iJ_y, \]  
which satisfy

\[ J_+ |j, m_j\rangle = \hbar \sqrt{(j - m_j)(j + m_j + 1)} |j, m_j + 1\rangle \]  
and  
\[ J_- |j, m_j\rangle = \hbar \sqrt{(j + m_j)(j - m_j + 1)} |j, m_j - 1\rangle \]  
as well as

\[ J_+ |j, j\rangle = 0 \]  
and  
\[ J_- |j, -j\rangle = 0. \]  
So, we have

\[ -j \leq m_j \leq +j \text{ for } j \geq 0, \]  
\[ J^2 |j, m_j\rangle = \hbar^2 j(j + 1) |j, m_j\rangle, \]  
and  
\[ J_z |j, m_j\rangle = m_j \hbar |j, m_j\rangle. \]  
Finally, recall that one immediate implication of this theory for experimental physics is that simultaneous measurements of two components of \( J \) is impossible while \( J^2 \) and one component of \( J \), typically chosen to be \( J_z \), can be determined at the same time. We will revisit the concept of generalized angular momentum in Chapter 11, where we discuss the electron spin.
1. Show that

\[ [L_z, L_x] = i\hbar L_y. \]
Chapter 10

The Hydrogen Atom

This is the first “real” system with multiple variables. So, let us take a deep breath here and briefly summarize what we have established in terms of the Schrödinger Equation with one variable.

So far

1. TDSE (the Time-Dependent Schrödinger Equation)

\[
\left[ -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x, t) \right] \Psi(x, t) = i\hbar \frac{\partial \Psi(x, t)}{\partial t} \tag{10.1}
\]

2. TISE (the Time-Independent Schrödinger Equation)

\[
\left[ -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x) \right] \psi(x) = E\psi(x); \tag{10.2}
\]

where \(\psi(x)\) is an eigenfunction, and \(E\) is an eigenvalue.

3. The Relation between \(\Psi\) and \(\psi\)

\[
\Psi(x, t) = \psi(x)e^{-iEt/\hbar} \tag{10.3}
\]

We have not really tried any real system yet. But, we need to carry out a real computation for a real system for an experimental confirmation of the correctness of the Schrödinger Equation

One of the simplest systems is a hydrogen atom. However, this poses a couple of difficulties since (1) two particles are involved now and also since (2) it is no longer one-dimensional, but three-dimensional.
10.1 2 Particles Instead of 1

Let $M$ be the mass of the nucleus and $m$ the mass of the electron. The nucleus and the electron are both moving about their fixed center of mass. But, we definitely prefer a system where the nucleus is standing still and the electron is the only object in motion. Introduction of the reduced mass $\mu$ given by

$$\mu = \frac{Mm}{M + m} \quad (10.4)$$

solves this problem. It literally reduces the original two-body problem to a one-body problem, where a body of mass $\mu$ experiences the same force that is at work between the two bodies. $^1$ This force is the Coulombic force for us and is inversely proportional to the square of the distance between the two bodies. For the hydrogen atom, $M = 1.67262178 \times 10^{-27}$ kilograms and $m = 9.1093891 \times 10^{-31}$ kilograms. So,

---

$^1$The reduced mass $\mu$ is an effective inertial mass in the following sense. First, let $F$ be the force exerted on the electron by the nucleus. Then, we know from Newton’s third law that the force exerted by the electron on the nucleus is $-F$. Now, denote the accelerations of the nucleus and electron by $a_M$ and $a_m$. We have

$$F = ma_m \quad \text{and} \quad -F = Ma_M \Rightarrow ma_m = -Ma_M. \quad (10.5)$$

Hence, the relative acceleration $a$, the acceleration of the electron relative to the nucleus, is given by

$$a = a_m - a_M = \frac{F}{m} - \frac{-F}{M} = \left(\frac{1}{m} + \frac{1}{M}\right)F = \frac{M + m}{Mm}F. \quad (10.6)$$

Dividing (10.6) through by $\frac{M+m}{Mm}$ to make the expression look like Newton’s second law, we get

$$F = \frac{Mm}{M + m}a. \quad (10.7)$$

This can be interpreted as force $F$ acting on an object of mass $\mu = \frac{Mm}{M+m}$. Finally, note that the relation between $\mu$, $M$, and $m$ can be expressed as follows.

$$\frac{1}{\mu} = \frac{1}{M} + \frac{1}{m} \quad (10.8)$$
\[ \mu = \left( \frac{M}{m + M} \right) m \approx 0.999455679m \approx m, \]  
\hspace{1cm} (10.9)

and we can simply use \( m \) instead of \( \mu \) in this case.

## 10.2 Three-Dimensional System

The classical energy equation is

\[ \frac{p^2}{2\mu} + V = E \]  
\hspace{1cm} (10.10)

or

\[ \frac{1}{2\mu} \left( p_x^2 + p_y^2 + p_z^2 \right) + V(x, y, z) = E. \]  
\hspace{1cm} (10.11)

We want to transform this into a three-dimensional Schrödinger Equation. Our solution, the wavefunction, now depends on four variables \( x, y, z, \) and \( t \).

\[ \Psi(x, t) \implies \Psi(x, y, z, t) \]  
\hspace{1cm} (10.12)

Here are the operator correspondences we are going to use.

\[ p_x \leftrightarrow -i\hbar \frac{\partial}{\partial x} \hspace{0.5cm} p_y \leftrightarrow -i\hbar \frac{\partial}{\partial y} \hspace{0.5cm} p_z \leftrightarrow -i\hbar \frac{\partial}{\partial z} \hspace{0.5cm} E \leftrightarrow i\hbar \frac{\partial}{\partial t} \]  
\hspace{1cm} (10.13)

Of course, this means we have the following.

\[ p_x^2 = -\hbar^2 \frac{\partial^2}{\partial x^2} \hspace{0.5cm} p_y^2 = -\hbar^2 \frac{\partial^2}{\partial y^2} \hspace{0.5cm} p_z^2 = -\hbar^2 \frac{\partial^2}{\partial z^2} \]  
\hspace{1cm} (10.14)

Therefore,

\[ \frac{p^2}{2\mu} = \frac{p_x^2 + p_y^2 + p_z^2}{2\mu} = \frac{1}{2\mu} \left[ -\hbar^2 \frac{\partial^2}{\partial x^2} + \left( -\hbar^2 \frac{\partial^2}{\partial y^2} \right) + \left( -\hbar^2 \frac{\partial^2}{\partial z^2} \right) \right] \]
\[
\frac{-\hbar^2}{2\mu} \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right).
\] (10.15)

On the other hand, the potential energy is now a function of \( x, y, \) and \( z \) as below.

\[
V = V(x, y, z) = \frac{-\epsilon^2}{4\pi\varepsilon_0 \sqrt{x^2 + y^2 + z^2}}
\] (10.16)

We now have the following Time-Dependent Schrödinger Equation in three dimensions.

\[
\frac{-\hbar^2}{2\mu} \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \Psi(x, y, z, t) + V(x, y, z)\Psi(x, y, z, t) = i\hbar \frac{\partial \Psi(x, y, z, t)}{\partial t}
\] (10.17)

Let us introduce a Laplacian operator or “\textit{del squared}” \( \nabla^2 \) defined by

\[
\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.
\] (10.18)

This simplifies the equation to the following form.

\[
\frac{-\hbar^2}{2\mu} \nabla^2 \Psi + V\Psi = i\hbar \frac{\partial \Psi}{\partial t};
\] (10.19)

where

\[
V = V(x, y, z) \quad \text{and} \quad \Psi = \Psi(x, y, z, t).
\] (10.20)

Now let

\[
\Psi(x, y, z, t) = \psi(x, y, z)e^{-iEt/\hbar}.
\] (10.21)

Then, we get the Time-Independent Schrödinger Equation.
The equation for the three-dimensional system is

\[ -\frac{\hbar^2}{2\mu} \nabla^2 \psi(x, y, z) + V(x, y, z) \psi(x, y, z) = E \psi(x, y, z) \]  

(10.22)

As it turns out, it is easier to solve this equation in spherical (polar) coordinates, where \((x, y, z)\) are replaced by \((r, \theta, \phi)\). Since a hydrogen atom possesses spherical symmetry, this is a better coordinate system. Figure 10.1 shows the spherical coordinates used in physics, while Figure 10.2 is the convention used in mathematics. It is unfortunate that the roles of \(\theta\) and \(\phi\) are switched between the two systems, but there is not much chance of confusion once you become used to the spherical coordinate system. In this class, we will use the physicists’ convention.

\[ V(x, y, z) = -\frac{e^2}{4\pi \varepsilon_0 \sqrt{x^2 + y^2 + z^2}} = -\frac{e^2}{4\pi \varepsilon_0 r} = V(r, \theta, \varphi) \]  

(10.23)

In fact, the potential energy only depends on the distance from the origin.

\(^2\)Note that in pure math books \(\theta\) and \(\phi\) are usually switched.
\[ V(r, \theta, \varphi) = -\frac{e^2}{4\pi\varepsilon_0 r} = V(r) \]  \hspace{1cm} (10.24)

Of course,

\[ \psi(x, y, z) \implies \psi(r, \theta, \varphi). \]  \hspace{1cm} (10.25)

This means that we have to convert \( \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \) to another equivalent operator expressed as derivatives with respect to \( r, \theta, \) and \( \varphi. \)

Then, we will have

\[ -\frac{\hbar^2}{2\mu} \nabla^2 \psi(r, \theta, \varphi) + V(r)\psi(r, \theta, \varphi) = E\psi(r, \theta, \varphi). \]  \hspace{1cm} (10.26)

The answer is

\[ \nabla^2 = \frac{1}{r^2 \frac{\partial}{\partial r}} \left( r^2 \frac{\partial}{\partial r} \right) + \frac{1}{r^2 \sin \theta \frac{\partial}{\partial \theta}} \left( \sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta \frac{\partial^2}{\partial \varphi^2}}, \]  \hspace{1cm} (10.27)
and we can see that $\nabla^2$ appears far more complicated in the spherical coordinates.

Here is how it is done. We begin with the three relations between $(x, y, z)$ and $(r, \theta, \varphi)$.

\[
\begin{align*}
x &= r \sin \theta \cos \varphi \\
y &= r \sin \theta \sin \varphi \\
z &= r \cos \theta
\end{align*}
\] (10.28)

We can see how $\frac{\partial^2}{\partial x^2}$ etc. can be converted. What we should do is to keep using chain rules and relations like $\frac{\partial x}{\partial r} = \sin \theta \cos \varphi$. But, the actual operation is quite tedious.

A full and complete proof of (10.27) can be found in Appendix L. But, let us prove only a part of this conversion here. Incidentally, it will be worth your while to read Sections K.1, K.2, and K.3 of Appendix K because the rules involving partial differentiation is often confusing.

Consider a one-variable function $\psi(r)$ for $r = \sqrt{x^2 + y^2 + z^2}$.

As (K.8) gives

\[
\frac{\partial \psi}{\partial x} = \frac{\partial \psi}{\partial r} \frac{\partial r}{\partial x} + \frac{\partial \psi}{\partial \theta} \frac{\partial \theta}{\partial x} + \frac{\partial \psi}{\partial \varphi} \frac{\partial \varphi}{\partial x} = \frac{x}{\sqrt{x^2 + y^2 + z^2}} \frac{\partial \psi}{\partial r} = \frac{x \partial \psi}{r \partial r}
\] (10.29)

and

\[
\frac{\partial r}{\partial x} = \frac{1}{2} \left( x^2 + y^2 + z^2 \right)^{-1/2} \cdot 2 = \frac{x}{r},
\] (10.30)

we have

\[
\frac{\partial^2 \psi}{\partial x^2} = \frac{\partial}{\partial x} \left( \frac{x \partial \psi}{r \partial r} \right) = \frac{\partial}{\partial x} \left( x \cdot \frac{1}{r \partial r} \right) = \frac{\partial x}{\partial x} \left( \frac{1}{r \partial r} \right) + x \frac{\partial}{\partial x} \left( \frac{1}{r \partial r} \right) = \frac{1}{r \partial r} + x \frac{r \partial \partial r}{x \partial \partial x} \left( \frac{1}{r \partial r} \right) = \frac{1}{r \partial r} + x \frac{r \partial \partial r}{x \partial \partial x} \left( \frac{1}{r \partial r} \right) = \frac{1}{r \partial r} + x^2 \frac{\partial}{\partial r} \left( \frac{1}{r \partial r} \right).
\] (10.31)

Similarly,

\[
\frac{\partial^2 \psi}{\partial y^2} = \frac{1}{r \partial r} + \frac{y^2}{r \partial r} \left( \frac{1}{r \partial r} \right)
\] (10.32)
and
\[
\frac{\partial^2 \psi}{\partial z^2} = \frac{1}{r} \frac{\partial \psi}{\partial r} + \frac{z^2}{r} \frac{\partial}{\partial r} \left( \frac{1}{r} \frac{\partial \psi}{\partial r} \right). \tag{10.33}
\]

Therefore,
\[
\nabla^2 \psi = \frac{3 \partial \psi}{r \partial r} + \frac{x^2 + y^2 + z^2}{r} \frac{\partial}{\partial r} \left( \frac{1}{r} \frac{\partial \psi}{\partial r} \right) = \frac{3 \partial \psi}{r \partial r} + r \frac{\partial}{\partial r} \left( \frac{1}{r} \frac{\partial \psi}{\partial r} \right)
= \frac{3 \partial \psi}{r \partial r} + r \left( -\frac{1}{r^2} \frac{\partial \psi}{\partial r} + \frac{1}{r} \frac{\partial^2 \psi}{\partial r^2} \right) = \frac{2 \partial \psi}{r \partial r} + \frac{\partial^2 \psi}{r^2 \partial r^2} + \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial \psi}{\partial r} \right). \tag{10.34}
\]

In order to get other parts, try \( \psi = \psi(\varphi) \) and \( \psi = \psi(\theta) \).

In preparation for the rest of the course as well as your career as a physicist or just as someone who needs to understand and use physics, please do get used to the spherical coordinates. For example, the volume element
\[
dx dy dz
\]
becomes
\[
r^2 dr \sin \theta d\theta d\varphi \quad \text{or} \quad -r^2 dr d(\cos \theta) d\varphi \tag{10.35}
\]
in spherical coordinates as shown in Figure 10.3. Needless to say, the limits of integration should be adjusted accordingly. For example, the normalization integrals are
\[
\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Psi^*(x,y,z,t) \Psi(x,y,z,t) \, dx \, dy \, dz,
\]
\[
\int_{0}^{2\pi} \int_{0}^{2\pi} \int_{0}^{\pi} \Psi^*(r,\theta,\phi,t) \Psi(r,\theta,\phi,t) \, r^2 \, dr \, d\theta \, d\phi,
\]
and
\[
\int_{0}^{+1} \int_{0}^{2\pi} \Psi^*(r,\theta,\phi,t) \Psi(r,\theta,\phi,t) \, r^2 \, dr \, d(\cos \theta) \, d\varphi;
\]
where the last integral (10.38) is particularly useful when we have a wavefunction whose \( \theta \)-dependence is via \( \cos \theta \) or \( \sin \theta \). As you can see in Table (10.3), this is indeed...
the case for the wavefunctions of the hydrogen atom. We sometimes write $dV$ for the volume element and $d\Omega$ for $\sin\theta d\theta d\phi$, so that

$$dV = r^2drd\Omega. \quad (10.39)$$

The quantity $\Omega$ is known as solid angle.

Let us pause for a minute here, take a deep breath, and summarize what we have achieved so far to clearly understand where we are now, as well as, "hopefully", where we are heading.

1. Due to the spherical symmetry inherent in the hydrogen atom, we made a decision to use the spherical coordinates as opposed to the familiar Cartesian coordinates. In particular, this simplifies the expression for the potential energy greatly. It is now a function only of the radial distance between the
electron and the proton nucleus.\textsuperscript{a}

\[ V(x, y, z) = V(r) = \frac{-e^2}{4\pi \varepsilon_0 r} \quad (10.40) \]

2. However, a rather heavy trade-off was the conversion of \( \nabla^2 \) to the equivalent expression in the spherical coordinates. The conversion process is cumbersome, and the resulting expression is far less palatable.

\[ \nabla^2 = \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \quad (10.41) \]

\textsuperscript{a}If the charge on the nucleus is Ze rather than just e, we have \( V(x, y, z) = V(r) = \frac{-Ze^2}{4\pi \varepsilon_0 r} \) instead. This makes extending the results obtained for the hydrogen atom to hydrogen-like atoms quite straightforward as we will see in Section 10.9.

We are now ready to solve the Time-Independent Schrödinger Equation in spherical coordinates given by

\[ -\frac{\hbar^2}{2\mu} \nabla^2 \psi(r, \theta, \varphi) + V(r)\psi(r, \theta, \varphi) = E\psi(r, \theta, \varphi). \quad (10.42) \]

We will resort to the separation of variables technique yet one more time. So, consider \( \psi \) as a product of three single variable functions of \( r, \theta, \) and \( \varphi \) respectively.

\[ \psi(r, \theta, \varphi) = R(r)\Theta(\theta)\Phi(\varphi) \quad (10.43) \]

Then, the Time-Independent Schrödinger Equation is

\[ \left[ -\frac{\hbar^2}{2\mu} \nabla^2_{(r, \theta, \varphi)} + V(r) \right] R(r)\Theta(\theta)\Phi(\varphi) = E R(r)\Theta(\theta)\Phi(\varphi). \quad (10.44) \]

Let us write it out in its full glory and start computing!

\[ \left[ -\frac{\hbar^2}{2\mu} \left( \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial (R\Theta\Phi)}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial (R\Theta\Phi)}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 (R\Theta\Phi)}{\partial \varphi^2} \right) 
+ V(r)R(r)\Theta(\theta)\Phi(\varphi) = E R(r)\Theta(\theta)\Phi(\varphi) \right. \]
\[ \Rightarrow -\frac{\hbar^2}{2\mu} \left[ \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial R}{\partial r} \Theta \Phi \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial \Theta}{\partial \theta} R \Phi \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 \Phi}{\partial \varphi^2} R \Theta \right] + V(r) R \Theta \Phi \]

\[
= -\frac{\hbar^2}{2\mu} \left[ \Theta \Phi \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial R}{\partial r} \right) + \frac{1}{r^2 \sin \theta} R \Phi \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial \Theta}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} R \Theta \frac{\partial^2 \Phi}{\partial \varphi^2} \right] + V(r) R \Theta \Phi \]

\[
= -\frac{\hbar^2}{2\mu} \left[ \Theta \Phi \frac{1}{r^2} \frac{d}{dr} \left( r^2 \frac{dR}{dr} \right) + \frac{1}{r^2 \sin \theta} R \Phi \frac{d}{d\theta} \left( \sin \theta \frac{d\Theta}{d\theta} \right) + \frac{1}{r^2 \sin^2 \theta} R \Theta \frac{d^2 \Phi}{d\varphi^2} \right] + V(r) R \Theta \Phi = ER \Theta \Phi (\varphi) \quad (10.45) \]

Now, multiply through by

\[
-\frac{2\mu}{\hbar^2} \cdot r^2 \sin^2 \theta \cdot \frac{1}{R \Theta \Phi},
\]

\[
-\frac{\hbar^2}{2\mu} \cdot \frac{-2\mu}{\hbar^2} r^2 \sin^2 \theta \cdot \frac{1}{R \Theta \Phi} \left[ \Theta \Phi \frac{1}{r^2} \frac{d}{dr} \left( r^2 \frac{dR}{dr} \right) + \frac{1}{r^2 \sin \theta} R \Phi \frac{d}{d\theta} \left( \sin \theta \frac{d\Theta}{d\theta} \right) 
+ \frac{1}{r^2 \sin^2 \theta} R \Theta \frac{d^2 \Phi}{d\varphi^2} \right] + V(r) R \Theta \Phi \frac{-2\mu}{\hbar^2} r^2 \sin^2 \theta \cdot \frac{1}{R \Theta \Phi}
\]

\[
= ER \Theta \Phi \cdot \frac{-2\mu}{\hbar^2} r^2 \sin^2 \theta \cdot \frac{1}{R \Theta \Phi} \quad (10.46) \]

\[
\Rightarrow \frac{\sin^2 \theta}{R} \frac{d}{dr} \left( r^2 \frac{dR}{dr} \right) + \frac{\sin \theta}{\Theta} \frac{d}{d\theta} \left( \sin \theta \frac{d\Theta}{d\theta} \right) + \frac{1}{\Phi} \frac{d^2 \Phi}{d\varphi^2} + \frac{-2\mu}{\hbar^2} r^2 \sin^2 \theta V(r)
\]

\[
= \frac{-2\mu}{\hbar^2} r^2 \sin^2 \theta E
\]

\[
\Rightarrow \frac{1}{\Phi} \frac{d^2 \Phi}{d\varphi^2} = -\frac{\sin^2 \theta}{R} \frac{d}{dr} \left( r^2 \frac{dR}{dr} \right) - \frac{\sin \theta}{\Theta} \frac{d}{d\theta} \left( \sin \theta \frac{d\Theta}{d\theta} \right) - \frac{2\mu}{\hbar^2} r^2 \sin^2 \theta \left[ E - V(r) \right]
\quad (10.47)
The LHS (lefthand side) is a function of $\varphi$, and the RHS (righthand side) is a function of $r$ and $\theta$. And the equality holds for all values of $r$, $\theta$, and $\varphi$. Therefore, both sides have to equal a constant. For later convenience, we denote this by $-m_l^2$. So, we have

$$\frac{1}{\Phi} \frac{d^2 \Phi}{d\varphi^2} = -m_l^2 \tag{10.48}$$

or

$$\frac{d^2 \Phi}{d\varphi^2} = -m_l^2 \Phi. \tag{10.49}$$

According to (10.47), this also means

$$-\sin^2 \theta \frac{d}{R} \left( r^2 \frac{dR}{dr} \right) - \frac{\sin \theta}{\Theta} \frac{d}{d\theta} \left( \sin \theta \frac{d\Theta}{d\theta} \right) - \frac{2\mu}{\hbar^2} r^2 \sin^2 \theta [E - V(r)] = -m_l^2. \tag{10.50}$$

Dividing through by $\sin^2 \theta$,

$$-\frac{1}{R} \frac{d}{dr} \left( r^2 \frac{dR}{dr} \right) - \frac{1}{\Theta \sin \theta} \frac{d}{d\theta} \left( \sin \theta \frac{d\Theta}{d\theta} \right) - \frac{2\mu}{\hbar^2} r^2 [E - V(r)] = -\frac{m_l^2}{\sin^2 \theta}$$

$$\Rightarrow \frac{1}{R} \frac{d}{dr} \left( r^2 \frac{dR}{dr} \right) + \frac{2\mu}{\hbar^2} r^2 [E - V(r)] = \frac{m_l^2}{\sin^2 \theta} - \frac{1}{\Theta \sin \theta} \frac{d}{d\theta} \left( \sin \theta \frac{d\Theta}{d\theta} \right). \tag{10.51}$$

The LHS of (10.51) depends on $r$, and the RHS is a function of $\theta$. As the equality holds for any values of $r$ and $\theta$, both sides have to be a constant. Denote that constant by $l(l + 1)^3$, for the reason that will become clear shortly, to get

$$\frac{-1}{\sin \theta} \frac{d}{d\theta} \left( \sin \theta \frac{d\Theta}{d\theta} \right) + \frac{m_l^2 \Theta}{\sin^2 \theta} = l(l + 1) \Theta \tag{10.52}$$

and

---

3We are not placing any restriction on the values $l$ can take. Because $l$ can be a complex number, any scalar can be represented as $l(l + 1)$. However, we will later show that $l$ is in fact a nonnegative integer.
10.3. THE SOLUTIONS FOR $\Phi$

We have now reduced the problem to that of solving the following three equations, each in one variable.

\begin{align}
\frac{1}{r^2} \frac{d}{dr} \left( r^2 \frac{dR}{dr} \right) + \frac{2\mu}{\hbar^2} [E - V(r)] R &= l(l + 1) \frac{R}{r^2}. \quad (10.53)
\end{align}

The easiest to solve is equation (10.54).

\begin{align}
\frac{d^2 \Phi}{d\varphi^2} &= -m_l^2 \Phi \quad (10.54)
\end{align}

So,

\begin{align}
\Phi(\varphi) &= e^{im_l \varphi}. \quad (10.58)
\end{align}

To be precise, we have

\begin{align}
A e^{im_l \varphi} \quad (10.59)
\end{align}

for an arbitrary scalar $A$. However, we can adjust the constant when we normalize the total wavefunction $\Psi(r, \varphi, \theta, t)$. So, we have set $A = 1$ for now.

Consider single-valuedness.
\[ \Phi(0) = \Phi(2\pi) \implies e^{im\cdot0} = e^{im\cdot2\pi} \]
\[ \implies 1 = e^{im\cdot2\pi} = \cos(m_l2\pi) + i \sin(m_l2\pi) \quad (10.60) \]

Now, let \( m_l = a + bi \) for \( (a, b \in \mathbb{R}) \).
\[ |1| = |e^{im\cdot2\pi}| = |e^{i(a+bi)2\pi}| = |e^{ia2\pi}| \cdot |e^{-b2\pi}| \quad (10.61) \]
\[ |e^{-b2\pi}| = e^{-b2\pi} = 1 \quad (10.62) \]
\[ 2\pi b = 0 \implies b = 0 \quad (10.63) \]

It is now obvious that \( m_l \) is a real number and
\[ |m_l| = 0, 1, 2, 3, 4, \ldots \quad (10.64) \]

For each value of \( m_l \), there is a corresponding solution
\[ \Phi_{m_l}(\varphi) = e^{im_l\varphi}. \quad (10.65) \]

The numbers \( m_l \) are known as magnetic quantum numbers.

Now we need to solve equations (10.55) and (10.56).

**10.4 The Solutions for \( \Theta \)**

Equation (10.55) is known as the angular equation.
\[ -\frac{1}{\sin \theta} \frac{d}{d\theta} \left( \sin \theta \frac{d\Theta}{d\theta} \right) + \frac{m_l^2\Theta}{\sin^2 \theta} = l(l+1)\Theta \quad (10.66) \]

Note that
\[ d \cos \theta = -\sin \theta d\theta. \quad (10.67) \]

So,
\[ -\frac{1}{\sin \theta} \frac{d}{d\theta} = \frac{d}{d \cos \theta}. \quad (10.68) \]
and

\[
\sin \theta \frac{d}{d\theta} = \sin \theta \frac{d}{d\cos \theta} = -\sin^2 \theta \frac{d}{d\cos \theta} = -(1 - \cos^2 \theta) \frac{d}{d\cos \theta},
\]

(10.69)

This suggests that (10.55) can be simplified if we let \( z = \cos \theta \). We have

\[
\frac{d}{d\cos \theta} \left( -(1 - \cos^2 \theta) \frac{d\Theta}{d\cos \theta} \right) + \frac{m_l^2 \Theta}{1 - \cos^2 \theta} = l(l + 1)\Theta,
\]

(10.70)

which leads to

\[
\frac{d}{dz} \left( -(1 - z^2) \frac{d\Theta}{dz} \right) + \frac{m_l^2 \Theta}{1 - z^2} = l(l + 1)\Theta \Rightarrow
\]

\[
\frac{d}{dz} \left( 1 - z^2 \right) \frac{d\Theta}{dz} + \left( l(l + 1) - \frac{m_l^2}{1 - z^2} \right) \Theta = 0
\]

(10.71)

At the time Schrödinger was solving his equation for the hydrogen atom, this was already a well-known differential equation, and the solutions are called the associated Legendre functions denoted by \( \Theta_{lm}(z) \). This is related to a better known set of functions called the Legendre polynomials denoted by \( P_l(z) \). The relation between \( \Theta_{lm}(z) \) and \( P_l(z) \) is as follows.

\[
\Theta_{lm}(z) = (1 - z^2)^{|m_l|/2} \frac{d^{|m_l|} P_l(z)}{dz^{|m_l|}}
\]

(10.72)

Our first job is to prove this.

Now \( P_l(z) \) is a solution to

\[
(1 - z^2) \frac{d^2 P_l}{dz^2} - 2z \frac{dP_l}{dz} + l(l + 1)P_l = 0.
\]

(10.73)

Differentiate both sides with respect to \( z \) to obtain

\[
-2z \frac{d^2 P_l}{dz^2} + (1 - z^2) \frac{d^3 P_l}{dz^3} - 2z \frac{dP_l}{dz} - 2z \frac{d^2 P_l}{dz^2} + l(l + 1) \frac{dP_l}{dz}
\]

\[
= (1 - z^2) \frac{d^3 P_l}{dz^3} - 4z \frac{d^2 P_l}{dz^2} + \left( (l(l + 1) - 2) \frac{dP_l}{dz} \right) = 0.
\]

(10.74)
Try \( \frac{d}{dz} \) again.

\[
-2 \frac{d^2 P_l}{dz^2} - 2z \frac{d^3 P_l}{dz^3} - 2z \frac{d^3 P_l}{dz^3} + (1 - z^2) \frac{d^4 P_l}{dz^4} - 2 \frac{d^2 P_l}{dz^2} - 2 \frac{d^2 P_l}{dz^2} \\
-2z \frac{d^3 P_l}{dz^3} + l(l + 1) \frac{d^2 P_l}{dz^2} \\
= (1 - z^2) \frac{d^4 P_l}{dz^4} - 6z \frac{d^3 P_l}{dz^3} + (l(l + 1) - 6) \frac{d^2 P_l}{dz^2} = 0
\] (10.75)

Suppose we get

\[
(1 - z^2) \frac{d^{k+2} P_l}{dz^{k+2}} - 2(k + 1)z \frac{d^{k+1} P_l}{dz^{k+1}} + (l(l + 1) - k(k + 1)) \frac{d^k P_l}{dz^k} = 0
\] (10.76)

after differentiating \( k \) times. Relation (10.76) holds for \( k = 0, 1, \) and \( 2 \) according to (10.73), (10.74), and (10.75).

Now apply \( \frac{d}{dz} \) one more time.

\[
-2z \frac{d^{k+2} P_l}{dz^{k+2}} + (1 - z^2) \frac{d^{k+3} P_l}{dz^{k+3}} - 2(k + 1) \frac{d^{k+1} P_l}{dz^{k+1}} - 2(k + 1)z \frac{d^{k+2} P_l}{dz^{k+2}} \\
+ (l(l + 1) - k(k + 1)) \frac{d^{k+1} P_l}{dz^{k+1}} \\
= (1 - z^2) \frac{d^{(k+1)+2} P_l}{dz^{(k+1)+2}} - 2((k + 1) + 1)z \frac{d^{(k+2)+1} P_l}{dz^{(k+2)+1}} \\
+ (l(l + 1) - (k + 1)((k + 1) + 1)) \frac{d^{k+1} P_l}{dz^{k+1}} = 0
\] (10.77)

So, by mathematical induction,

\[
(1 - z^2) \frac{d^{k+2} P_l}{dz^{k+2}} - 2(k + 1)z \frac{d^{k+1} P_l}{dz^{k+1}} + (l(l + 1) - k(k + 1)) \frac{d^k P_l}{dz^k} = 0
\] (10.78)

when we apply \( \frac{d}{dz^k} \) to both sides of (10.73).

On the other hand, consider

\[
\Theta_{lm} = (1 - z^2)^{|m|/2} \Gamma(z)
\] (10.79)

and substitute this into
\[
\frac{d}{dz} \left( (1 - z^2) \frac{d\Theta}{dz} \right) + \left( l(l+1) - \frac{m^2}{1-z^2} \right) \Theta = 0. \quad (10.80)
\]

We have the following lengthy computation.

\[
\frac{d}{dz} \left( (1 - z^2) \frac{d}{dz} (1 - z^2)^{k/2} \Gamma \right) + \left( l(l+1) - \frac{k^2}{1-z^2} \right) (1 - z^2)^{k/2} \Gamma \\
= -2z \frac{d}{dz} ((1 - z^2)^{k/2} \Gamma) + (1 - z^2) \frac{d^2}{dz^2} ((1 - z^2)^{k/2} \Gamma) \\
+ \left( l(l+1) - \frac{k^2}{1-z^2} \right) (1 - z^2)^{k/2} \Gamma \\
= -2z \left( \frac{k}{2} (1 - z^2)^{k/2-1} (-2z) \Gamma + (1 - z^2)^{k/2} \frac{d\Gamma}{dz} \right) \\
+ (1 - z^2) \frac{d}{dz} \left( \frac{k}{2} (1 - z^2)^{k/2-1} (-2z) \Gamma + (1 - z^2)^{k/2} \frac{d\Gamma}{dz} \right) \\
+ \left( l(l+1) - \frac{k^2}{1-z^2} \right) (1 - z^2)^{k/2} \Gamma \\
= -2z \left( k z (1 - z^2)^{k/2-1} \Gamma + (1 - z^2)^{k/2} \frac{d\Gamma}{dz} \right) \\
+ (1 - z^2) \left[ \frac{k}{2} (k/2 - 1) (1 - z^2)^{k/2-2} (-2z) (2z) \Gamma + \frac{k}{2} (1 - z^2)^{k/2-1} (-2) \Gamma \\
+ \frac{k}{2} (1 - z^2)^{k/2-2} (-2z) \frac{d\Gamma}{dz} + \frac{k}{2} (1 - z^2)^{k/2-1} (-2z) \frac{d\Gamma}{dz} + (1 - z^2)^{k/2} \frac{d^2\Gamma}{dz^2} \right] \\
+ \left( l(l+1) - \frac{k^2}{1-z^2} \right) (1 - z^2)^{k/2} \Gamma \\
= 2k z^2 (1 - z^2)^{(k/2-1)} \Gamma - 2z (1 - z^2)^{k/2} \frac{d\Gamma}{dz} \\
+ (1 - z^2) \left[ 2kz^2 \left( \frac{k}{2} - 1 \right) (1 - z^2)^{(k/2-2)} \Gamma - k (1 - z^2)^{(k/2-1)} \Gamma \right]
\[-kz(1 - z^2)^{(k/2 - 1)} \frac{d\Gamma}{dz} - kz(1 - z^2)^{(k/2 - 1)} \frac{d\Gamma}{dz} + (1 - z^2)\frac{d^2\Gamma}{dz^2} \]
\[+ \left( l(l + 1) - \frac{k^2}{1 - z^2} \right) (1 - z^2)^{k/2} \Gamma = 0 \] \hspace{1cm} (10.81)

Multiplying through by \((1 - z^2)^{-k/2}\),
\[\frac{2kz^2}{1 - z^2} \frac{d\Gamma}{dz} - 2z \frac{d\Gamma}{dz} + (1 - z^2) \left[ 2kz^2 \left( \frac{k}{2} - 1 \right) (1 - z^2)^{-2} \Gamma - k(1 - z^2)^{-1} \Gamma \right] \]
\[-kz(1 - z^2)^{-1} \frac{d\Gamma}{dz} - kz(1 - z^2)^{-1} \frac{d\Gamma}{dz} - kz \frac{d\Gamma}{dz} \]
\[+ (1 - z^2) \frac{d^2\Gamma}{dz^2} + \left( l(l + 1) - \frac{k^2}{1 - z^2} \right) \Gamma \]
\[= (1 - z^2) \frac{d^2\Gamma}{dz^2} + (-2kz - 2z) \frac{d\Gamma}{dz} + \left[ l(l + 1) + \frac{2kz^2 + 2kz^2 \left( \frac{k}{2} - 1 \right) - k^2}{1 - z^2} - k \right] \Gamma \]
\[= (1 - z^2) \frac{d^2\Gamma}{dz^2} - 2(k + 1)z \frac{d\Gamma}{dz} + \left[ l(l + 1) + \frac{k^2z^2 - k^2}{1 - z^2} - k \right] \Gamma \]
\[= (1 - z^2) \frac{d^2\Gamma}{dz^2} - 2(k + 1)z \frac{d\Gamma}{dz} + \left[ l(l + 1) - k^2 - k \right] \Gamma = 0. \] \hspace{1cm} (10.82)

So, we have shown
\[(1 - z^2) \frac{d^2\Gamma}{dz^2} - 2(k + 1)z \frac{d\Gamma}{dz} + [l(l + 1) - k^2 - k] \Gamma = 0 \] \hspace{1cm} (10.83)

or
\[(1 - z^2) \frac{d^2\Gamma}{dz^2} - 2(|m_l| + 1)z \frac{d\Gamma}{dz} + [l(l + 1) - |m_l| (|m_l| + 1)] \Gamma = 0. \] \hspace{1cm} (10.84)

Comparing this with (10.76), which is given below again for reference,
\[(1 - z^2) \frac{d^{k+2}}{dz^{k+2}} P_l - 2(k + 1)z \frac{d^{k+1} P_l}{dz^{k+1}} + (l(l + 1) - k(k + 1)) \frac{d^k P_l}{dz^k} = 0, \] \hspace{1cm} (10.85)
we conclude

\[ \Theta_{l;m_l} = (1 - z^2)^{|m_l|/2} \frac{d^{|m_l|} P_l}{dz^{|m_l|}}. \]  

It remains to solve (10.73)

\[ (1 - z^2) \frac{d^2 P_l}{dz^2} - 2z \frac{dP_l}{dz} + l(l + 1)P_l = 0 \]  

for \( P_l \). Try

\[ P_l(z) = \sum_{k=0}^{\infty} a_k z^k. \]  

We get a recursion relation as we did before on p.150 in Section 7.6.1 for the simple harmonic oscillator.

\[ a_{j+2} = \frac{j(j + 1) - l(l + 1)}{(j + 2)(j + 1)} a_j \]  

Also as before, we do not want an infinite series; that is, we want the series to terminate after a finite number of terms, so that the function is integrable/normalizable. We need

\[ j(j + 1) - l(l + 1) = j^2 + j - l^2 - l = (j + l)(j - l) + j - l \]
\[ = (j - l)(j + l + 1) = 0 \iff l = j \text{ or } -j - 1. \]  

As \( j \) is a nonnegative integer, \( l \) can be any integer, positive, 0, or negative. However, the only meaningful quantity is \( l(l + 1) \). Noting that the parabola

\[ y(l) = l(l + 1) = \left( l + \frac{1}{2} \right)^2 - \left( \frac{1}{2} \right)^2 \]  

is symmetric about \( l = -\frac{1}{2} \), it is sufficient to consider \( l = 0, 1, 2, 3, \ldots \). So, \( l = 0, 1, 2, 3, \ldots \) give all acceptable solutions. In fact, \( y(-l - 1) = (-l - 1)(-l - 1 + 1) = (-l - 1)(-l) = l(l + 1) \) proves that the pairs \( (0, -1), (1, -2), (2, -3), (3, -4) \ldots \) give the same value for \( l(l + 1) \), and negative \( l \) values are clearly redundant. We get
\[ P_0 = 1, \quad P_1 = z, \quad P_2 = 1 - 3z^2, \quad P_3 = 3z - 5z^3, \ldots \quad (10.92) \]

So, the corresponding \( \Theta_{lm} \)'s are

\[
\begin{align*}
\Theta_{00} &= 1, & \Theta_{10} &= z, & \Theta_{1\pm 1} &= (1 - z^2)^{1/2}, & \Theta_{20} &= 1 - 3z^2, \\
\Theta_{2\pm 1} &= (1 - z^2)^{1/2}z, & \Theta_{2\pm 2} &= 1 - z^2, \ldots
\end{align*} \quad (10.93) \]

Note that \( P_l(z) \) is an \( l \)th degree polynomial. Therefore, for each \( l \), we have

\[ m_l = -l, \quad -l + 1, \quad \ldots, \quad 0, \quad \ldots, \quad l - 1, \quad +l. \quad (10.94) \]

### 10.5 Associated Legendre Polynomials and Spherical Harmonics

Recall that \( P_l(z) \)'s are called the Legendre polynomials, and \( \Theta_{lm} \)'s are called associated Legendre functions/polynomials as explained on p.209. The associated Legendre polynomials are also denoted by \( P_l^m(\cos \theta) \) or simply by \( P_l^m(\cos \theta) \). So,

\[ \Theta_{lm}(\cos \theta) = P_l^m(\cos \theta). \quad (10.95) \]

We now introduce a new set of functions known as the spherical harmonics, which are the products of \( \Phi \) and \( \Theta \) up to the normalization constant. It is customary to denote each spherical harmonic by \( Y_l^m \) or simply by \( Y_l^m \). So, we have

\[ Y_l^m(\theta, \phi) \propto \Phi_m(\phi)\Theta_{lm}(\cos \theta) = \Phi_m(\phi)P_l^m(\cos \theta) = e^{im\phi}P_l^m(\cos \theta). \quad (10.96) \]

With the appropriate normalization constant\(^4\),

\(^4\)Different definitions for spherical harmonics are used in different fields such as geodesy, magnetics, quantum mechanics, and seismology. The definition given here is commonly adopted in the quantum mechanics community.
10.5. ASSOCIATED LEGENDRE POLYNOMIALS AND SPHERICAL HARMONICS

\[ Y_l^{m_l}(\theta, \phi) = (-1)^{m_l} \sqrt{\frac{2l + 1}{4\pi} \frac{(l - m_l)!}{(l + m_l)!}} P_l^{m_l}(\cos \theta) e^{im_l\phi}, \quad (10.97) \]

where the normalization is chosen such that

\[
\int_0^{2\pi} \int_0^\pi (Y_l^{m_l}(\theta, \phi))^* Y_l^{m_l}(\theta, \phi) \sin \theta d\theta d\phi = 1, \quad (10.98)
\]

or

\[
\int \int (Y_l^{m_l}(\theta, \phi))^* Y_l^{m_l}(\theta, \phi) d\Omega = 1. \quad (10.99)
\]

In fact, \(Y_l^{m_l}\)'s are orthonormal such that

\[
\int_0^{2\pi} \int_0^\pi (Y_l^{m_l}(\theta, \phi))^* Y_l^{m_l'}(\theta, \phi) \sin \theta d\theta d\phi = 1, \quad (10.100)
\]

We will see how this comes about in Section 10.6. In passing let us make a note of the fact that the normalization constant from the \(\phi\)-dependent portion of the integral is \(\sqrt{\frac{1}{2\pi}}\), and the normalization constant from the \(\theta\)-dependent portion of the integral is \(\sqrt{\frac{(2l+1)(l-m_l)!}{2(l+m_l)!}}\).

Spherical harmonics for \(m_l < 0\) has the following alternative formula

\[ Y_l^{-m_l}(\theta, \phi) = (-1)^{-m_l} \sqrt{\frac{2l + 1}{4\pi} \frac{(l + |m_l|)!}{(l - |m_l|)!}} P_l^{-m_l}(\cos \theta) e^{im_l\phi}. \quad (10.101) \]

This is a direct consequence of the following identity satisfied by the associated Legendre polynomials.\(^5\) We will write \(m\) instead of \(m_l\) for simplicity.

\[ P_l^{-m}(x) = (-1)^m \frac{(l - |m|)!}{(l + m)!} P_l^m(x) \quad (10.102) \]

\(^5\)See Fact M.1 of Appendix M on p.342 for a proof.
Indeed, for \( m = -|m| < 0 \), we have

\[
Y_l^m(\theta, \phi) = Y_l^{-|m|}(\theta, \phi) = (-1)^{-|m|} \sqrt{\frac{2l + 1}{4\pi}} \frac{(l + |m|)!}{(l - |m|)!} P_l^{-|m|}(\cos \theta) e^{-i|m|\phi}
\]

\[
= (-1)^{-|m|} \sqrt{\frac{2l + 1}{4\pi}} \frac{(l - |m|)!}{(l + |m|)!} \left[ (-1)^{|m|} \frac{(l - |m|)!}{(l + |m|)!} P_l^{|m|}(\cos \theta) \right] e^{i(|m|)\phi}
\]

\[
= \sqrt{\frac{2l + 1}{4\pi}} \frac{(l - |m|)!}{(l + |m|)!} P_l^{|m|}(\cos \theta) e^{i|m|\phi}.
\] (10.103)

We can combine (10.97) and (10.103) in the following manner.

\[
Y_l^{mi}(\theta, \phi) = (-1)^{(m_l+|m|)/2} \sqrt{\frac{2l + 1}{4\pi}} \frac{(l - |m|)!}{(l + |m|)!} P_l^{|m|}(\cos \theta) e^{im_l\phi}. \] (10.104)

Note that the phase factor \((-1)^{m_l}\) or \((-1)^{(m_l+|m|)/2}\) does not change the physical observables, and it is more or less a mathematical book keeping device.

### 10.6 \( L^2 \), \( L_z \), and the Spherical Harmonics

The orthonormality relation (10.100) is a consequence of the following fact.

**Fact 10.1** The spherical harmonics are simultaneous eigenstates of \( L^2 \) and \( L_z \) corresponding to the eigenvalues \( l(l + 1)\hbar^2 \) and \( m_l\hbar \).

\[
L^2 Y_l^{mi} (\theta, \phi) = l(l + 1)\hbar^2 Y_l^{mi} (\theta, \phi)
\] (10.105)

and

\[
L_z Y_l^{mi} (\theta, \phi) = m_l\hbar Y_l^{mi} (\theta, \phi)
\] (10.106)

Because \( L^2 \) and \( L_z \) are Hermitian operators, and also because \( Y_l^{mi} \)'s are normalized, the set

\[
\{Y_l^{mi}\}_{l=0,1,2,\ldots}^{m_l \leq l \leq +l}(10.107)
\]

forms an orthonormal basis. Hence, \( Y_l^{mi} \)'s are not only orthonormal, but also complete, and any function of \( \theta \) and \( \phi \) can be represented as a superposition (linear combination) of spherical harmonics.
Recall that for each of \( l = 0, 1, 2, 3, \ldots \) we have \( m_l = -l, -l+1, -l+2, \ldots, 0, \ldots, l-2, l-1, l \). All the spherical harmonics for \( l = 0, 1, \) and \( 2 \) are listed in Table 10.1.

<table>
<thead>
<tr>
<th>( l )</th>
<th>( m_l )</th>
<th>( Y_0^m(\theta, \phi) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>( Y_0^0 = \frac{1}{\sqrt{\pi}} )</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>( Y_{1}^{-1} = \frac{1}{2} \sqrt{\frac{3}{2\pi}} e^{-i\phi} \sin \theta = \frac{1}{2} \sqrt{\frac{3}{2\pi}} \frac{x-iy}{r} )</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>( Y_{1}^{0} = \frac{1}{2} \sqrt{\frac{3}{\pi}} \cos \theta = \frac{1}{2} \sqrt{\frac{3}{\pi}} \frac{z}{r} )</td>
</tr>
<tr>
<td></td>
<td>+1</td>
<td>( Y_{1}^{1} = -\frac{1}{2} \sqrt{\frac{3}{2\pi}} e^{i\phi} \sin \theta = -\frac{1}{2} \sqrt{\frac{3}{2\pi}} \frac{x+iy}{r} )</td>
</tr>
<tr>
<td>2</td>
<td>-2</td>
<td>( Y_{2}^{-2} = \frac{1}{4} \sqrt{\frac{15}{2\pi}} e^{-2i\phi} \sin^2 \theta = \frac{1}{4} \sqrt{\frac{15}{2\pi}} \frac{(x-iy)^2}{r^2} )</td>
</tr>
<tr>
<td></td>
<td>-1</td>
<td>( Y_{2}^{-1} = \frac{1}{2} \sqrt{\frac{15}{2\pi}} e^{-i\phi} \sin \theta \cos \theta = \frac{1}{2} \sqrt{\frac{15}{2\pi}} \frac{(x-iy)z}{r^2} )</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>( Y_{2}^{0} = \frac{1}{4} \sqrt{\frac{5}{\pi}} (3 \cos^2 \theta - 1) = \frac{1}{4} \sqrt{\frac{5}{\pi}} \frac{2x^2 - y^2}{r^2} )</td>
</tr>
<tr>
<td></td>
<td>+1</td>
<td>( Y_{2}^{1} = +\frac{1}{2} \sqrt{\frac{15}{2\pi}} e^{i\phi} \sin \theta \cos \theta = +\frac{1}{2} \sqrt{\frac{15}{2\pi}} \frac{(x+iy)z}{r^2} )</td>
</tr>
<tr>
<td></td>
<td>+2</td>
<td>( Y_{2}^{2} = \frac{1}{4} \sqrt{\frac{15}{2\pi}} e^{2i\phi} \sin^2 \theta = \frac{1}{4} \sqrt{\frac{15}{2\pi}} \frac{(x+iy)^2}{r^2} )</td>
</tr>
</tbody>
</table>

Table 10.1: Spherical Harmonics for \( l = 0, 1, \) and \( 2 \)

Next on our agenda is the radial function \( R \).

### 10.7 The Radial Function \( R \)

The general radial equation for a one-electron atom is

\[
\frac{1}{r^2} \frac{d}{dr} \left( r^2 \frac{dR}{dr} \right) + \frac{2\mu}{\hbar^2} \left[ E + \frac{Ze^2}{4\pi\varepsilon_0 r} \right] R = \frac{l(l+1)}{r^2};
\]  

where \( Ze \) is the charge on the nucleus. We will only solve this for \( Z = 1 \) or for the hydrogen atom.
Let us first invoke some substitutions to put the radial equation in a more manageable form.

\[ \beta^2 = \frac{2\mu E}{\hbar} \quad (\beta > 0) \]  

(10.109)

\[ \rho = 2\beta r \]  

(10.110)

\[ \gamma = \frac{\mu e^2}{4\pi \varepsilon_0 \hbar^2 \beta} \]  

(10.111)

With these substitutions, we get

\[ \frac{1}{\rho^2} \frac{d}{d\rho} \left( \rho^2 \frac{dR}{d\rho} \right) + \left[ -\frac{1}{4} - \frac{l(l+1)}{\rho^2} + \frac{\gamma}{\rho} \right] R = 0. \]  

(10.112)

As \( \rho \to \infty \)

\[ R(\rho) \approx e^{-\rho/2}. \]  

(10.113)

So, let us write

\[ R(\rho) = e^{-\rho/2} F(\rho), \]  

(10.114)

and substitute this into the radial equation to get

\[ \frac{d^2 F}{d\rho^2} + \left( \frac{2}{\rho} - 1 \right) \frac{dF}{d\rho} + \left[ \frac{\gamma - 1}{\rho} - \frac{l(l+1)}{\rho^2} \right] F = 0. \]  

(10.115)

Once again, we will try a series solution with

\[ F(\rho) = \rho^s \sum_{k=0}^{\infty} a_k \rho^k \quad (a_0 \neq 0, s \geq 0); \]  

(10.116)

where the condition on \( s \) is in place to prevent \( F \) from blowing up at the origin, while \( a_0 \neq 0 \) assures that our solution is not trivial as we will see in (10.119).

After substituting this trial series solution into the differential equation (10.115), we get

\[ [s(s+1) - l(l+1)]a_0 \rho^{s-2} + \sum_{j=0}^{\infty} \{ [(s + 1)(s + j + 2) - l(l+1)]a_{j+1} \]


10.7. THE RADIAL FUNCTION $R$

\[-(s + j + 1 - \gamma)a_j \rho^{s+j-1} = 0. \tag{10.117}\]

For the left-hand side to be zero for any value of $\rho$, the following two relations should hold.

\[
\begin{cases}
  s(s+1) - l(l+1) = 0 \tag{10.118} \\
  a_{j+1} = \frac{s+j+1-\gamma}{(s+j+1)(s+j+2) - l(l+1)} a_j \tag{10.119}
\end{cases}
\]

The first equality (10.118) gives $s = l$ and $s = -(l+1)$. But, $s = -(l+1)$ should be rejected as $s \geq 0$. The second relation (10.119)

\[
a_{j+1} = \frac{j+l+1-\gamma}{(j+l+1)(j+l+2) - l(l+1)} a_j \tag{10.120}
\]

indicates that

\[
\gamma = j + l + 1 = n \text{ (Call this } n. \text{)} \tag{10.121}
\]

assures termination of the series after a finite number of terms; namely, after the $j$th term. Hence,

\[
n = l + 1, l + 2, l + 3, \ldots \tag{10.122}
\]

as $j$ runs from 0 to $\infty$ with $l = 0, 1, 2, 3, \ldots$.

Now recall from (10.109) on p.218

\[
\beta^2 = \frac{2\mu E}{\hbar^2} \tag{10.123}
\]

and from (10.111) on p.218

\[
\gamma = \frac{\mu e^2}{4\pi \varepsilon_0 \hbar^2 \beta} \tag{10.124}
\]

So,
\[ E_n = -\frac{\mu e^4}{(4\pi \varepsilon_0)^2 2\hbar^2 n^2} = -\left( \frac{\mu e^4}{32\pi^2 \varepsilon_0^2 \hbar^2} \right) \frac{1}{n^2} = -\left( \frac{\hbar^2}{2\mu a_0^2} \right) \frac{1}{n^2} \]

for \( n = 1, 2, 3, \ldots \). \( \text{(10.125)} \)

Here,

\[ a_0 = \frac{4\pi \varepsilon_0 \hbar^2}{\mu e^2} \]

is analogous to the Bohr radius \( a_B \) defined by

\[ a_B = \frac{4\pi \varepsilon_0 \hbar^2}{m_e e^2} \]

where \( m_e \) is the rest mass of an electron. The numerical value of \( a_B \) is about 5.29177 \( \times 10^{-11} \) m, 52.9177 pm, 0.0529177 nm, or 0.529177 angstroms.

Because the radial function \( R(r) \) depends on two parameters, \( n \) and \( l \), we write \( R_{nl}(r) \) for the radial part of the wavefunction.

### 10.8 The Full Wavefunction

Putting it all together, we get

\[ \psi_{nlm_l}(r, \theta, \varphi) = R_{nl}(r) Y_{lm_l}^m(\theta, \varphi) = R_{nl}(r) \Theta_{lm_l}(\theta) \Phi_{m_l}(\varphi); \]

\( \text{(10.128)} \)

where

\[ \Phi_{m_l}(\varphi) = e^{im_l\varphi} \quad |m_l| = 0, 1, 2, 3, \ldots \]

\( \text{(10.129)} \)

\[ \Theta_{lm_l}(\theta) = \Theta_{lm_l}(\cos \theta) = P_{l}^{m_l}(\cos \theta) = \sin^{[m_l]} \theta F_{l|m_l}(\cos \theta), \]

\( \text{(10.130)} \)

and
\[ R_{nl}(r) = e^{-r/na_0} \left( \frac{r}{a_0} \right)^2 G_{nl} \left( \frac{r}{a_0} \right). \]  \hfill (10.131)

The functions \( F \) and \( G \) are both polynomials;

\[ F_{l|m|}(\cos \theta) \]  \hfill (10.132)

is a polynomial in \( \cos \theta \), and

\[ G_{nl} \left( \frac{r}{a_0} \right) \]  \hfill (10.133)

is a polynomial in \( r/a_0 \).

More specifically, the normalized radial wavefunctions are\(^6,7\)

\[ R_{nl}(r) = \sqrt{\left( \frac{2}{na_0} \right)^3 (n - l - 1)!} e^{-r/na_0} \left( \frac{2r}{na_0} \right)^l L_{n-l-1}^{2l+1} \left( \frac{2r}{na_0} \right) \]  \hfill (10.134)

and the normalized full wavefunctions are

\[
\psi_{nlm} = R_{nl}(r)Y_{lm}^m(\theta, \phi)
= \sqrt{\left( \frac{2}{na_0} \right)^3 (n - l - 1)!} e^{-r/na_0} \left( \frac{2r}{na_0} \right)^l L_{n-l-1}^{2l+1} \left( \frac{2r}{na_0} \right)
\times \left[ \frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!} P_l^{m} \left( \cos \theta \right) e^{im\phi} \right].
\]  \hfill (10.135)

Let us now summarize the relations among \( n, l, \) and \( m_l \). We have obtained

\[ |m_l| = 0, 1, 2, 3, \cdots, \]  \hfill (10.136)

\[ l = |m_l|, |m_l| + 1, \cdots, |m_l| + 3, \cdots, \]  \hfill (10.137)

\(^6\) The appropriate normalization scheme for \( R_{nl} \) is discussed in Note 10.1.
\(^7\) \( L \) in (10.134) is an associated Laguerre polynomial described in Appendix N.
in this order, but this can be reorganized as follows.

\[ n = l + 1, l + 2, l + 3, \cdots \]  \hspace{1cm} (10.138)

\[ n = 1, 2, 3, 4, \cdots \]  \hspace{1cm} (10.139)

\[ l = 0, 1, 2, \cdots, n - 1 \]  \hspace{1cm} (10.140)

\[ m_l = -l, -l + 1, \cdots, 0, \cdots, +l \]  \hspace{1cm} (10.141)

The first number \( n \) is associated with \( R(r) \) and is called the principal quantum number or the radial quantum number. The second number \( l \) is associated with \( \Theta(\theta) \) and is called the orbital quantum number or the angular momentum quantum number. Finally, the number \( m_l \) is called the magnetic quantum number which is associated with \( \Phi(\varphi) \). In addition to these, we also have a spin quantum number \( m_s \), which we will discuss in Chapter 11.

**Remark 10.1 (Degeneracy for the Hydrogen Atom)** The total energy \( E \) depends only on the principal quantum number \( n \) for the hydrogen atom. This means all the distinct states \( \{\psi_{nlm}\} \) for \( l = 0, 1, \ldots, n - 1 \) and \( -l \leq m_l \leq +l \) share the same energy level. Noting that there are \( 2l + 1 \) distinct states for each \( l \), we have

\[
\sum_{l=0}^{n-1} 2l + 1 = 2 \sum_{l=0}^{n-1} l + \sum_{l=0}^{n-1} 1 = 2 \cdot (n - 1 + 0) \cdot \frac{n}{2} + n = n^2 - n + n = n^2. \hspace{1cm} (10.142)
\]

So, the \( n \)-th energy level of the Hydrogen atom is \( n^2 \)-fold degenerate.

As it turned out, this is a special property of a \( \frac{1}{r} \) potential. In order to see the special nature of the \( \frac{1}{r} \) potential, we will compare the expectation values \( \langle r^k \rangle \) for \( k = -3, -2, -1, 1, \) and 2. According to the normalization scheme described in Note 10.1, we have the following equality.

\[
\iiint \psi_{nlm}(r, \theta, \phi)^* r^k \psi_{nlm}(r, \theta, \phi) \, dV = \iiint (R_{nl}(r)\Theta_{lm}(\theta)\Phi_{m_l}(\varphi))^* r^k (R_{nl}(r)\Theta_{lm}(\theta)\Phi_{m_l}(\varphi)) \, dV = \int_0^\infty \iiint (R_{nl}(r)\Theta_{lm}(\theta)\Phi_{m_l}(\varphi))^* r^k (R_{nl}(r)\Theta_{lm}(\theta)\Phi_{m_l}(\varphi)) \, r^2 \, dr \, d\Omega = \int_0^\infty r^{k+2} R_{nl}(r)^* R_{nl}(r) \, dr \cdot \int_0^\pi \Theta_{lm}(\theta)^* \Theta_{lm}(\theta) \sin \theta \, d\theta \cdot \int_0^{2\pi} \Phi_{m_l}(\phi)^* \Phi_{m_l}(\phi) \, d\phi = \int_0^\infty r^{k+2} R_{nl}(r)^* R_{nl}(r) \, dr \times 1 \times 1 = \int_0^\infty r^{k+2} R_{nl}(r)^* R_{nl}(r) \, dr \hspace{1cm} (10.143)
\]
Therefore,
\[ \langle r^k \rangle = \int_0^\infty r^{k+2} R_{nl}(r)^* R_{nl}(r) \, dr \quad \text{or} \quad \int_0^\infty r^{k+2} \left| R_{nl}(r) \right|^2 \, dr. \] (10.144)

Actual computations tend to be quite tedious, but the following are the results.
\[ \langle r^2 \rangle = \frac{a_0^2 n^2}{2} \left[ 5n^2 + 1 - 3l(l + 1) \right] \] (10.145)
\[ \langle r \rangle = \frac{a_0}{2} \left[ 3n^2 - l(l + 1) \right] \] (10.146)
\[ \langle \frac{1}{r} \rangle = \frac{1}{n^2 a_0} \] (10.147)
\[ \langle \frac{1}{r^2} \rangle = \frac{1}{(l + \frac{1}{2}) n^3 a_0^2} \] (10.148)
\[ \langle \frac{1}{r^3} \rangle = \frac{1}{l (l + \frac{1}{2}) (l + 1)n^3 a_0^3} \] (10.149)

One can note that only \( \langle \frac{1}{r} \rangle \) does not depend on the angular momentum quantum number \( l \).

The first 6 normalized radial wavefunctions for a hydrogen-like atom, where the positive charge on the nucleus is \( Ze \) rather than \( e \), are presented in Table 10.2. Likewise, the first 14 normalized full wavefunctions are listed in Table 10.3.

**Note 10.1 (Individual Normalization for \( R_{nl}, \Theta_{lm}, \) and \( \Phi_m \))** We know that a normalized time-independent wavefunction \( \psi(r, \theta, \varphi) \) and time-dependent wavefunction \( \Psi(r, \theta, \varphi, t) = \psi(r, \theta, \varphi) e^{-iEt/\hbar} \) should satisfy
\[ \int \psi(r, \theta, \varphi, t)^* \psi(r, \theta, \varphi, t) \, dV = \int \left( \psi(r, \theta, \varphi) e^{-iEt/\hbar} \right)^* \psi(r, \theta, \varphi) e^{-iEt/\hbar} \, dV \\
= \int \psi(r, \theta, \varphi)^* \psi(r, \theta, \varphi) \, dV = 1; \] (10.150)

where \( dV = r^2 dr \sin \theta d\theta d\phi \) is the volume element. More specifically, we have
\[ \int \psi_{nlm}(r, \theta, \varphi)^* \psi_{nlm}(r, \theta, \varphi) \, dV = \int_0^\infty \int_0^{2\pi} \int_0^\pi \psi_{nlm}(r, \theta, \varphi)^* \psi_{nlm}(r, \theta, \varphi) r^2 \sin \theta \, d\theta d\phi \\
= \int_0^\infty \int_0^{2\pi} \int_0^\pi (R_{nl}(r) \Theta_{lm}(\theta) \Phi_m(\varphi))^* (R_{nl}(r) \Theta_{lm}(\theta) \Phi_m(\varphi)) r^2 \sin \theta \, d\theta d\phi \]
\[ = \int_0^\infty R_{nl}(r)^* R_{nl}(r) r^2 dr \int_0^\pi \Theta_{lm}(\theta)^* \Theta_{lm}(\theta) \sin \theta d\theta \int_0^{2\pi} \Phi_{ml}(\varphi)^* \Phi_{ml}(\varphi) d\varphi = 1. \]  

(10.151)

Now, it is absolutely NOT necessary that we have
\[ \int_0^\infty R_{nl}(r)^* R_{nl}(r) r^2 dr = \int_0^\pi \Theta_{lm}(\theta)^* \Theta_{lm}(\theta) \sin \theta d\theta = \int_0^{2\pi} \Phi_{ml}(\varphi)^* \Phi_{ml}(\varphi) d\varphi = 1 \]

(10.152)

separately for the \( r \)-, \( \theta \)-, and \( \varphi \)-dependent parts of the wavefunction \( \psi \). However, it is often convenient to adjust the scaling factors so that (10.152) is indeed satisfied. Normalization for \( \Phi \) is assured by
\[ \Phi_{ml}(\varphi) = \frac{1}{\sqrt{2\pi}} e^{im\varphi} \]

(10.153)

because
\[ \int_0^{2\pi} \Phi(\varphi)^* \Phi(\varphi) d\varphi = \int_0^{2\pi} \frac{1}{\sqrt{2\pi}} e^{-im\varphi} \frac{1}{\sqrt{2\pi}} e^{im\varphi} d\varphi = \frac{1}{2\pi} \int_0^{2\pi} 1 d\varphi = 1. \]  

(10.154)

On the other hand, from p. 215, the normalization constant for \( \Theta \) is
\[ \sqrt{\frac{(2l+1)(l-m)!}{2(l+m)!}}, \]

(10.155)

and, from (10.97), we have the following normalized \( \Theta_{lm} \) as well as \( Y_i^{ml}(\theta, \varphi) = \Theta_{lm}(\theta) \times \Phi_{ml}(\varphi) \).
\[ \Theta_{lm} = (-1)^m \sqrt{\frac{(2l+1)(l-m)!}{2(l+m)!}} P_i^{ml}(\cos \theta) \]

(10.156)

Finally, we need
\[ \int_0^\infty R_{nl}(r)^* R_{nl}(r) r^2 dr = 1. \]

(10.157)

This is satisfied by the \( R_{nl}(r) \) given in (10.134).
\[ R_{nl}(r) = \left( \frac{2}{na_0} \right)^3 \left( \frac{n-l-1}{2n[(n+l)!]^3} \right) e^{-r/na_0} \left( \frac{2r}{na_0} \right)^{l-1/2} L_{n-l-1}^{2l+1} \left( \frac{2r}{na_0} \right) \]

(10.158)
Definition 10.1 (Radial Probability Density \( P(r) \)) The radial probability density \( P(r) \) is defined by

\[
P(r) = \int \psi^*(r, \theta, \phi) \psi(r, \theta, \phi) r^2 d\Omega = \int_0^{2\pi} \int_0^\pi \Psi^*(r, \theta, \phi) \Psi(r, \theta, \phi) r^2 \sin \theta d\theta d\phi
\]

so that \( P(r_0) dr \) gives the probability of finding the electron in the range \([r_0, r_0 + dr]\), which is the spherical shell of thickness \( dr \) bounded by the surfaces of the spheres of radii \( r_0 \) and \( r_0 + dr \). Needless to say, \( P(r) \) satisfies

\[
\int_0^\infty P(r) dr = 1.
\]

10.9 Hydrogen-Like Atoms

A hydrogen-like atom, or hydrogen-like ion, is an atomic nucleus with one electron. Some examples other than the hydrogen atom itself are \( \text{He}^+, \text{Li}^{2+}, \text{Be}^{3+}, \text{and B}^{4+} \). If we denote the atomic number by \( Z \), the charge carried by hydrogen-like ions are \( e(Z - 1) \). With this notation, we can derive all the wavefunctions for the hydrogen-like ions from the wavefunctions for the hydrogen atom by replacing the charge \( e \) on proton by \( Ze \). You can see this from the time-independent Hamiltonian.

\[
\left\{ -\frac{\hbar^2 c}{2\mu} \left[ \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \varphi^2} \right] - \frac{Ze^2}{4\pi \varepsilon_0 r} \right\} \psi(r, \theta, \varphi) = E\psi(r, \theta, \varphi)
\]

The only difference is \( Z \) appearing once in the potential term. The resulting wavefunctions are

\[
\psi_{nlm_l,Z} = R_{nl,Z}(r) Y_l^{m_l}(\theta, \phi)
\]

\[
= \left\{ \frac{(2Z)^3}{(n_0)^3} \frac{1}{(n_0)(n_0 + l)!} e^{-Zr/n_0} \left( \frac{2Zr}{n_0} \right)^l \prod_{l=0}^{l} \frac{2l + 1}{4\pi} \frac{(l-m_l)!}{(l+m_l)!} \right\}
\]

And, the total energy \( E_Z \) is given by
\\[ E_{Z,n} = -\frac{Z^2\mu e^4}{(4\pi\varepsilon_0)^22\hbar^2n^2} = -\left(\frac{Z^2\mu e^4}{32\pi^2\varepsilon_0^2\hbar^2}\right)\frac{1}{n^2} = -\left(\frac{Z^2\hbar^2}{2\mu a_0^3}\right)\frac{1}{n^2} \tag{10.163} \]

\[
\text{for } n = 1, 2, 3, \ldots.
\]

<table>
<thead>
<tr>
<th>Quantum Numbers</th>
<th>Eigenfunctions (R_{nl}(r))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n)</td>
<td>(l)</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 10.2: Some Normalized Radial Eigenfunctions for the One-Electron Atom

### 10.10 Simultaneous Diagonalization of \(H\), \(L^2\), and \(L_z\)

Let us examine the big picture now. In Section 10.6, we saw that the spherical harmonics

\[
\{Y_l^{m_l}\}_{l=0,1,2,...} \tag{10.164}
\]

form an orthonormal basis that simultaneously diagonalizes \(L^2\) and \(L_z\). Likewise, the set of normalized eigenfunctions or eigenkets

\[
\{\psi_{n_l m_l}\} \text{ or } \{|n, l, m_l\}\} \tag{10.165}
\]
### Quantum Numbers

<table>
<thead>
<tr>
<th>$n$</th>
<th>$l$</th>
<th>$m_l$</th>
<th>Eigenfunctions $\psi_{nlm_l}(r, \theta, \phi)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>$\psi_{100} = \frac{1}{\sqrt{\pi}} \left( \frac{Z}{a_0} \right)^{3/2} e^{-Zr/a_0}$</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>$\psi_{200} = \frac{1}{4 \sqrt{2\pi}} \left( \frac{Z}{a_0} \right)^{3/2} \left( 2 - \frac{Zr}{a_0} \right) e^{-Zr/2a_0}$</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>$\psi_{210} = \frac{1}{4 \sqrt{2\pi}} \left( \frac{Z}{a_0} \right)^{3/2} \frac{Zr}{a_0} e^{-Zr/2a_0} \cos \theta$</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>$\pm 1$</td>
<td>$\psi_{21\pm 1} = \frac{1}{8 \sqrt{\pi}} \left( \frac{Z}{a_0} \right)^{3/2} \frac{Zr}{a_0} e^{-Zr/2a_0} \sin \theta e^{\pm i\phi}$</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>$\psi_{300} = \frac{1}{18 \sqrt{3\pi}} \left( \frac{Z}{a_0} \right)^{3/2} \left( 6 - \frac{4Zr}{a_0} + \frac{4Z^2r^2}{9a_0^2} \right) e^{-Zr/3a_0}$</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>$\psi_{310} = \frac{1}{27 \sqrt{2\pi}} \left( \frac{Z}{a_0} \right)^{3/2} \left( 4 - \frac{2Zr}{3a_0} \right) \frac{Zr}{a_0} e^{-Zr/3a_0} \cos \theta$</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>$\pm 1$</td>
<td>$\psi_{31\pm 1} = \frac{1}{54 \sqrt{\pi}} \left( \frac{Z}{a_0} \right)^{3/2} \left( 4 - \frac{2Zr}{3a_0} \right) \frac{Zr}{a_0} e^{-Zr/3a_0} \sin \theta e^{\pm i\phi}$</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0</td>
<td>$\psi_{320} = \frac{1}{81 \sqrt{6\pi}} \left( \frac{Z}{a_0} \right)^{3/2} \frac{Z^2r^2}{a_0^3} e^{-Zr/3a_0} (3 \cos^2 \theta - 1)$</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>$\pm 1$</td>
<td>$\psi_{32\pm 1} = \frac{1}{81 \sqrt{\pi}} \left( \frac{Z}{a_0} \right)^{3/2} \frac{Z^2r^2}{a_0^3} e^{-Zr/3a_0} \sin \theta \cos \theta e^{\pm i\phi}$</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>$\pm 2$</td>
<td>$\psi_{32\pm 2} = \frac{1}{162 \sqrt{\pi}} \left( \frac{Z}{a_0} \right)^{3/2} \frac{Z^2r^2}{a_0^3} e^{-Zr/3a_0} \sin^2 \theta e^{\pm 2i\phi}$</td>
</tr>
</tbody>
</table>

Table 10.3: Some Normalized Full Eigenfunctions for the One-Electron Atom
for \( n = 1, 2, 3, \ldots, l = 0, 1, \ldots, n - 1, \) and \( m_l = -l, -l + 1, \ldots 0 \ldots, l - 1, +l \) form an orthonormal basis that simultaneously diagonalizes the Hamiltonian \( H \), the magnitude of the orbital angular momentum squared \( L^2 \), and the z-component of the orbital angular momentum \( L_z \). We have the following set of simultaneous eigenvalue problems with a common solution \( \{|n,l,m_l\} \).

\[
\begin{align*}
H |\psi\rangle &= E_n |\psi\rangle \\
L^2 |\psi\rangle &= l(l + 1)\hbar^2 |\psi\rangle \\
L_z |\psi\rangle &= m_l \hbar |\psi\rangle
\end{align*}
\] (10.166)

The reason why such common eigenkets can be found is that the commutators \([H, L^2]\) and \([H, L_z]\) are both zero in addition to \([L^2, L_z] = 0\) which we know from (9.8). As you can see in Appendix L, there are alternative forms of \( \nabla^2 \). The form given in (10.27) is

\[
\nabla^2 = \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2},
\] (10.167)

and one alternative form is

\[
\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2} + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2},
\] (10.168)

or

\[
\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} \left[ \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right].
\] (10.169)

From (9.4) and (10.169), we have

\[
\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} \left[ \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right]
\]

\[
= \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} - \frac{1}{\hbar^2 r^2} (-\hbar^2) \left[ \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right]
\]

\[
= \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} - \frac{L^2}{\hbar^2 r^2}.
\] (10.170)

We will use this form of \( \nabla^2 \) in order to show \([H, L^2] = 0\) and \([H, L_z] = 0\).\(^8\)

\(^8\)Both \([H, L^2] = 0\) and \([H, L_z] = 0\) are proved in a different manner in Appendix I.
Fact 10.2 \([H, L^2] = 0\) The full Hamiltonian

\[
H = -\frac{\hbar^2}{2\mu} \nabla^2 + V(r) = -\frac{\hbar^2}{2\mu} \left[ \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} - \frac{L^2}{\hbar^2 r^2} \right] + V(r)
\]  
(10.171)

commutes with the squared magnitude of the angular momentum

\[
L^2 = -\hbar^2 \left[ \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right].
\]  
(10.172)

**Proof**

The important point here is that the expression of \(H\) contains only the radial variable \(r\) and \(L^2\), while the expression of \(L^2\) contains the azimuthal and polar variables \(\phi\) and \(\theta\). With this in mind, the commutator \([H, L^2]\) can be computed as follows.

\[
\begin{align*}
[H, L^2] &= \left[ -\frac{\hbar^2}{2\mu} \left[ \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} - \frac{L^2}{\hbar^2 r^2} \right] + V(r), L^2 \right] \\
&= \left[ -\frac{\hbar^2}{2\mu} \left[ \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} - \frac{L^2}{\hbar^2 r^2} \right], L^2 \right] + [V(r), L^2] \\
&= \left[ -\frac{\hbar^2}{2\mu} \left[ \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} \right], L^2 \right] + \left[ \frac{L^2}{\hbar^2 r^2}, L^2 \right] + [V(r), L^2].
\end{align*}
\]  
(10.173)

It is now clear that each term of (10.173) is zero, and we have verified

\[
[H, L^2] = 0.
\]  
(10.174)

Fact 10.3 \([H, L_z] = 0\) The full Hamiltonian

\[
H = -\frac{\hbar^2}{2\mu} \nabla^2 + V(r) = -\frac{\hbar^2}{2\mu} \left[ \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} - \frac{L^2}{\hbar^2 r^2} \right] + V(r)
\]  
(10.175)

commutes with the \(z\)-component of the angular momentum

\[
L_z = -i\hbar \frac{\partial}{\partial \phi}.
\]  
(10.176)
Proof

\[ [H, L_z] = \left[ -\frac{\hbar^2}{2\mu} \left( \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} - \frac{L_z^2}{\hbar^2 r^2} \right) + V(r), L_z \right] \]

\[ = \left[ -\frac{\hbar^2}{2\mu} \left( \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} \right), L_z \right] + \left[ \frac{L_z^2}{2\mu r^2}, L_z \right] + [V(r), L_z] = 0. \quad (10.177) \]

10.11 Revisiting the Fundamental Postulates

Let us recall the Fundamental Postulates presented in Section 3.1. The following is the set of guiding principles that connects the physical reality with mathematical abstraction.

- Each physical system \( S \) has an associated Hilbert space \( \mathcal{H} \).
- Each physical state \( s \) of the physical system \( S \) has an associated normalized ket \( |s\rangle \), called a state ket, in \( \mathcal{H} \).
- Every physical observable \( q \) is represented by a Hermitian Operator \( Q \) whose domain is \( \mathcal{H} \).
- Measurement of \( q \) in the physical system \( S \) is mirrored in the abstract quantum mechanical system by the action of the corresponding operator \( Q \) on the state ket \( |s\rangle \).
- The only possible outcomes of quantum mechanical measurements are the eigenvalues of the corresponding operator \( Q \).
- A state ket is initially a linear combination of normalized eigenkets of \( Q \). However, after the measurement, the state ket becomes the eigen ket associated with the measured eigenvalue.

You can note here that the quantum mechanical system is composed of the threesome \( (\mathcal{H}, |s\rangle, Q) \), which are purely mathematical and abstract constructs. In particular, no reference is made to any coordinate system or a function defined relative to such
Therefore, the generic simultaneous eigenvalue problems (10.166) and their solutions (10.165) are actually coordinate free. In other words, the eigenkets \(|n, l, m_\ell\rangle\) “live” in some abstract unspecified Hilbert space \(\mathcal{H}\). From Section 4.6, we can expand \(|n, l, m_\ell\rangle\) in the coordinate basis \(|r, \theta, \phi\rangle\) to get

\[
|n, l, m_\ell\rangle = I |n, l, m_\ell\rangle = \left( \sum_{0 \leq r < +\infty} \sum_{0 \leq \theta \leq \pi} \sum_{0 \leq \phi < 2\pi} |r, \theta, \phi\rangle \langle r, \theta, \phi| \right) |n, l, m_\ell\rangle
\]

or more appropriately

\[
|n, l, m_\ell\rangle = \left( \int_0^{\infty} \int_0^{\pi} \int_0^{2\pi} |r, \theta, \phi\rangle \langle r, \theta, \phi| \ r^2 \sin \theta d\theta d\phi \right) |n, l, m_\ell\rangle
\]

where we used the identities

\[
\sum_{0 \leq r < +\infty} \sum_{0 \leq \theta \leq \pi} \sum_{0 \leq \phi < 2\pi} |r, \theta, \phi\rangle \langle r, \theta, \phi| = I
\]

and

\[
\int_0^{\infty} \int_0^{\pi} \int_0^{2\pi} |r, \theta, \phi\rangle \langle r, \theta, \phi| \ r^2 \sin \theta d\theta d\phi = I.
\]

When we write

\[
\psi_{nlm_l}(r, \theta, \phi) = \langle r, \theta, \phi|n, l, m_\ell\rangle
\]

for the coefficients in (10.178) and (10.180), we are taking the projection of \(|n, l, m_\ell\rangle\) onto the infinite-dimensional coordinate system whose axes are defined by \(|r, \theta, \phi\rangle\).
to use a geometric language. In a more common language, this amounts to solving the set of eigenvalue problems (10.166) using the spherical coordinate system \((r, \theta, \phi)\) in order to obtain the eigenvectors as functions of \(r, \theta, \) and \(\phi\). The operators \(H, L^2,\) and \(L_z\) are also abstract and coordinate-independent generic objects whose representations in the spherical coordinate system are

\[
H = -\frac{\hbar^2}{2\mu} \left[ \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right] + V(r), \quad (10.183)
\]

\[
L^2 = -\hbar^2 \left[ \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right], \quad (10.184)
\]

and

\[
L_z = -i\hbar \frac{\partial}{\partial \phi}. \quad (10.185)
\]

Similarly, we can project \(|n, l, m_l\rangle\) onto the Cartesian system defined by the familiar \(|\{x, y, z\}\rangle\).

\[
\psi_{nlm_l}(x, y, z) = \langle x, y, z| n, l, m_l \rangle \quad (10.186)
\]

In this case, we already know

\[
H = -\frac{\hbar^2}{2\mu} \left[ \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right] + V(x, y, z) \quad (10.187)
\]

and

\[
L_z = -i\hbar \left( x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right). \quad (10.188)
\]

---

\(^9\)This is not to be confused with the usual projection onto the \(x, y, \) and \(z\)-axes of a vector \(v,\) which allows us to write \(v = (v_x, v_y, v_z).\) Our coordinate system here is not the usual three-dimensional system, but it is of infinite dimensions, each axis of which is defined by one of \(|\{x, y, z\}\rangle;\) where \(-\infty < x, y, z < +\infty.\)

\(^{10}\)As we know \(L_x = -i\hbar \left( y \frac{\partial}{\partial z} - z \frac{\partial}{\partial y} \right)\) and \(L_y = -i\hbar \left( z \frac{\partial}{\partial x} - x \frac{\partial}{\partial z} \right)\) as well as \(L^2 = L_x^2 + L_y^2 + L_z^2,\) we can also express \(L^2\) in terms of \(x, y, z\) and \(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}.\) However, the resulting messy formula is not of much instructional value or other use to us. That is the reason why \(L^2\) is not included in this list.
Exercises

1. Consider a one-variable function $\psi(r)$ for $r = \sqrt{x^2 + y^2 + z^2}$, and derive the relation below.

$$\frac{\partial^2 \psi}{\partial x^2} = \frac{1}{r} \frac{\partial \psi}{\partial r} + \frac{x^2}{r} \frac{\partial}{\partial r} \left( \frac{1}{r} \frac{\partial \psi}{\partial r} \right)$$

2. Without showing detailed computations, simply sketch the proof that

$$\Theta_{lm} = (1 - z^2)^{m_l/2} q^{m_l} P_l \frac{d}{dz} l l$$

3. Hydrogen, deuterium, and singly ionized helium are all examples of one-electron atoms. The deuterium nucleus has the same charge as the hydrogen nucleus, and almost exactly twice the mass. The helium nucleus has twice the charge of the hydrogen nucleus, and almost exactly four times the mass. Make an approximate prediction of the ratios of the ground state energies of these atoms.

4. Consider $n = 3$ state of a hydrogen atom.

   (a) What $l$ values are possible?

   (b) For each value of $l$, what $m_l$ values are possible?

   (c) Finally, how many degenerate states do we have for $n = 3$?

5. What is the energy of a photon emitted when the electron drops from the 3rd highest energy level ($n = 3$) to the ground state ($n = 1$)? Leave $\mu$, $e$, $\pi$, $\varepsilon_0$, and $\hbar$ as they are.

6. Answer the following questions about the hydrogen atom.

   (a) The differential equation of the $\phi$-dependent function $\Phi(\phi)$ is given by $\frac{d^2 \Phi}{d\phi^2} = -k^2 \Phi$ with a solution $\Phi(\phi) = e^{ik\phi}$. Find all values of $k = a + bi$ ($a, b$ real) consistent with the single-valuedness condition. Show all your work.

   (b) For each principal quantum number $n$, there are $n^2$ degenerate energy levels. Prove this.

7. Answer the following questions about the hydrogen atom.

   (a) Consider $n = 4$. 
i. What is the largest allowed value of \( l \)?

ii. What is the magnitude of the corresponding angular momentum? Leave \( \hbar \) as it is.

iii. How many different \( z \)-components may this angular momentum vector have?

iv. What is the magnitude of the largest \( z \) component? Leave \( \hbar \) as it is.

(b) Consider an electron in the ground state of the hydrogen atom characterized by \( n = 0 \). The normalized ground state wavefunction is given by

\[
\psi_{000} = \frac{1}{\sqrt{\pi}} \left( \frac{1}{a_0} \right)^{3/2} e^{-r/a_0}.
\]

You may use

\[
\int x^2 e^{bx} \, dx = e^{bx} \left( \frac{x^2}{b} - \frac{2x}{b^2} + \frac{2}{b^3} \right) + C
\]

and

\[
\int x^3 e^{bx} \, dx = e^{bx} \left( \frac{x^3}{b} - \frac{3x^2}{b^2} + \frac{6x}{b^3} - \frac{6}{b^4} \right) + C.
\]

i. What is the probability \( P(0 \leq r < a_0) \) that the electron lies inside a sphere of radius \( a_0 \) centered at the origin? Leave \( e \) as it is.

ii. Find the average distance \( <r> \) of the electron from the nucleus.

(c) For a hydrogen-like atom with \( Z = 3 \), what is the energy of a photon emitted when the electron drops from the 3rd highest energy level (\( n = 3 \)) to the ground state (\( n = 1 \))? Leave \( \mu, e, \pi, \varepsilon_0, \) and \( \hbar \) as they are. Do not use \( a_0 \) in your answer.
Chapter 11
Electron Spin

There are quantum mechanical phenomena that do not fit in the framework based on the postulates given in Section 3.1 in a straightforward manner. While the postulates draw heavily on the analogy based on the corresponding classical mechanical system, these phenomena have no classical analogues. They require an additional variable called intrinsic spin and a special Hilbert space on which the spin operators act. The intrinsic spin $S$ is a very important signature for many particles. We will devote this chapter to the discussion of electron spin as it is the most representative example. Intrinsic spin can be characterized and understood best if we regard it as another kind of angular momentum that requires a separate treatment different from that for the familiar orbital angular momentum represented by $L_x$, $L_y$, and $L_z$.

11.1 What is the Electron Spin?

The electron was discovered in 1897 by Joseph John Thompson. It was the first subatomic and fundamental particle discovered. Physicists knew from the beginning that an electron had charge and mass. But, the concept of electron spin was proposed by Samuel Abraham Goudsmit and George Uhlenbeck much later in 1925. Because the electron was usually regarded as a point particle with no internal structure and no physical dimension, it was actually impossible for the electron to spin like a top. Therefore, it was necessary to assume that the electron was born with intrinsic angular momentum, called electron spin, which is not associated with its orbital motion. The first sign of electron spin came from an experiment conducted by Otto Stern and Walther Gerlach in 1921.
11.1.1 Stern-Gerlach Experiment

In this experiment, electrically neutral silver atoms passed through a nonuniform magnetic field $B = B\hat{z}$, oriented along the $z$-axis, which deflected the atoms. The atoms, after passing through the magnetic field, hit a photographic plate generating small visible dots. The deflection is along the $z$-axis and the magnitude of the force on the atoms is proportional to $\frac{dB}{dz}$. In the classical picture, the orbiting electron produces magnetic dipole moments whose magnitudes are continuously distributed. If this is the case, the extent of vertical deflection should also be continuously distributed forming a “blob” on the plate. On the other hand, in the quantum mechanical picture, there are $2l + 1$ $z$-components, $m_l$'s, for $l = 1, 2, 3, \ldots$. In particular, note that $m_l = 0$ is always possible. When $m_l = 0$, there is no force on the atom and the atoms with $m_l = 0$ will not be deflected. Hence, if $l = 0$, no atom will be deflected.

Figure 11.1: Stern-Gerlach Experiment (Diagram drawn by en.wikipedia Theresa Knott.)

The results of Stern-Gerlach experiment presented in their original paper are reproduced in Figure 11.2 [Gerlach and Stern, 1922, p.350]. The picture on the left shows only one line as there was no external magnetic field. When an external non-uniform magnetic field was turned on, there were two lines shown on the right due to the force proportional to $\frac{dB}{dz}$. However, this was inconsistent with their expectations in two ways.
1. There was no line at the center corresponding to $m_l = 0$.

2. There should be an odd number of lines; namely $2l + 1$. But, they observed two lines.

Later in 1927, T. E. Phipps and J. B. Taylor reproduced this effect using hydrogen atoms in the ground state ($l = 0$ and $m_l = 0$). This was a convincing piece of evidence that there was “something” other than the orbital angular momentum because $l = 0$ implies the orbital angular momentum is 0.

### 11.1.2 Fine Structure of the Hydrogen Spectrum: Spin-Orbit Interaction

Generally speaking, the term fine structure refers to the splitting of spectral lines due to first order relativistic effects. These relativistic effects add three fine structure terms to the Hamiltonian known as kinetic, spin-orbit interaction, and Darwin terms. And it is the spin-orbit splitting that provided one of the first pieces of evidence for electron spin.

We relegate the discussion of these relativistic effects to Appendix O.
11.2 Spin and Pauli Matrices

One way to account for these apparent discrepancies is to assume the existence of a fourth quantum number known as a spin quantum number denoted by \( s \). We can easily see that if we assume the magnitude of the spin \( S \) is given by

\[
S = \sqrt{s(s+1)}
\]

and \( s = 1/2 \), the number of values its \( z \)-component \( S_z = m_s \hbar \) can take is \( 2s + 1 = 2 \), and we can successfully explain the two-lines observed. In a nutshell, this is what we are going to do. While we will specialize to the electron shortly, spin itself is an intrinsic property possessed by any fundamental particle. Every fundamental particle has its own unique signature value of \( s \) that characterizes the particle. While \( m_s \) varies between \(-s\) and \(+s\), \( s \) itself does not.

As mentioned on p.235, **spin is best understood as intrinsic angular momentum**. According to Section 9.6, this means that the three components of the spin satisfy the same commutation relations as \( L_x, L_y, \) and \( L_z \); or their generalized counterparts \( J_x, J_y, \) and \( J_z \). Because spin is a generalized angular momentum, we can understand spin if we understand the quantum mechanical properties of angular momentum. Here are the basic properties of spin as a quantum mechanical angular momentum.

\[
S = S_x \hat{x} + S_y \hat{y} + S_z \hat{z} \tag{11.1}
\]

\[
S^2 = S_x^2 + S_y^2 + S_z^2 \tag{11.2}
\]

\[
[S_x, S_y] = i \hbar S_z \tag{11.3}
\]

\[
[S_y, S_z] = i \hbar S_x \tag{11.4}
\]

\[
[S_z, S_x] = i \hbar S_y \tag{11.5}
\]

\[
[S^2, S_x] = [S^2, S_y] = [S^2, S_z] = [S^2, S] = 0 \tag{11.6}
\]

From these properties and the results of actual measurements, we can construct the following framework to describe and understand the electron spin. We will find an orthonormal basis that diagonalizes \( S^2 \) and \( S_z \) simultaneously and find (non-diagonal) matrix representations for \( S_x \) and \( S_y \) in that basis.

First, based on the outcomes of various experiments, we know \( s = 1/2 \) and \( m_s = \pm 1/2 \). This means

\[
S^2 = s(s+1)\hbar^2 I = 1/2 \left( \frac{1}{2} + 1 \right) \hbar^2 I = \frac{3}{4} \hbar^2 I,
\]

where \( I \) is the Identity matrix.

\(^1\)In the framework of our quantum theory based on the Schrödinger equation, spin is an a posteriori consideration. However, spin appears naturally when the relativistic analog of the Schrödinger equation, known as the Dirac equation, is used. Spin is a consequence of demanding that the Dirac equation be first order in time and space.
11.2. SPIN AND PAULI MATRICES

identity operator, and the eigenvalues of $S_z$ are $\pm \frac{\hbar}{2}$. Now, due to Properties 2) and 3) on p.54 as well as Theorem 2.7 on p.58, two normalized eigenvectors of $S_z$ form an orthonormal basis for the corresponding Hilbert space. We will denote a normalized eigenvector for the eigenvalue $+\frac{\hbar}{2}$ by $|\uparrow\rangle$ and that for the eigenvalue $-\frac{\hbar}{2}$ by $|\downarrow\rangle$. Using the notation of Section 9.6 on generalized angular momentum, $|j, m_j\rangle$, we have

$$|\uparrow\rangle = |s = 1/2, m_s = 1/2\rangle = |1/2, 1/2\rangle$$
(11.7)

and

$$|\downarrow\rangle = |1/2, -1/2\rangle.$$  
(11.8)

Because our Hilbert space is two-dimensional, all operators can be represented by 2-by-2 matrices. Let us compute the matrix elements for $S^2$, $S_z$, $S_x$, and $S_y$.

Let us arrange $|\uparrow\rangle$ and $|\downarrow\rangle$ in this order, so that

$$|\uparrow\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$
(11.9)

and

$$|\downarrow\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$
(11.10)

and express the $i$-th row and $j$-th column component of the (2-by-2) matrix representation of an operator $\Omega$ by $\Omega_{ij}$. For $S^2$,

$$S^2_{11} = \langle \uparrow | S^2 | \uparrow \rangle = \frac{3}{4} \hbar^2 \langle \uparrow | \uparrow \rangle = \frac{3}{4} \hbar^2$$
(11.11)

$$S^2_{12} = \langle \uparrow | S^2 | \downarrow \rangle = \frac{3}{4} \hbar^2 \langle \uparrow | \downarrow \rangle = 0$$
(11.12)

$$S^2_{21} = \langle \downarrow | S^2 | \uparrow \rangle = \frac{3}{4} \hbar^2 \langle \downarrow | \uparrow \rangle = 0$$
(11.13)

$$S^2_{22} = \langle \downarrow | S^2 | \downarrow \rangle = \frac{3}{4} \hbar^2 \langle \downarrow | \downarrow \rangle = \frac{3}{4} \hbar^2$$
(11.14)
\[ S^2 = \begin{bmatrix} \frac{3}{4} \hbar^2 & 0 \\ 0 & \frac{3}{4} \hbar^2 \end{bmatrix} = \frac{3}{4} \hbar^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (11.15) \]

For \( S_z \),
\[
S_{z11} = \langle \uparrow | S_z | \uparrow \rangle = \left\langle \uparrow \right| \frac{\hbar}{2} \left| \uparrow \right\rangle = \frac{\hbar}{2} \langle \uparrow | \uparrow \rangle = \frac{\hbar}{2} \\
S_{z12} = \langle \uparrow | S_z | \downarrow \rangle = \left\langle \uparrow \right| -\frac{\hbar}{2} \left| \downarrow \right\rangle = -\frac{\hbar}{2} \langle \uparrow | \downarrow \rangle = 0 \\
S_{z21} = \langle \downarrow | S_z | \uparrow \rangle = \left\langle \downarrow \right| \frac{\hbar}{2} \left| \uparrow \right\rangle = \frac{\hbar}{2} \langle \downarrow | \uparrow \rangle = 0 \\
S_{z22} = \langle \downarrow | S_z | \downarrow \rangle = \left\langle \downarrow \right| -\frac{\hbar}{2} \left| \downarrow \right\rangle = -\frac{\hbar}{2} \langle \downarrow | \downarrow \rangle = -\frac{\hbar}{2} \quad (11.19)
\]

\[ S_z = \begin{bmatrix} +\frac{\hbar}{2} & 0 \\ 0 & -\frac{\hbar}{2} \end{bmatrix} = \frac{\hbar}{2} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \quad (11.20) \]

In order to compute the matrix representations of \( S_x \) and \( S_y \), it is convenient to use raising and lowering operators \( S_+ = S_x + iS_y \) and \( S_- = S_x - iS_y \). Solving for \( S_x \) and \( S_y \), we get

\[ S_x = \frac{1}{2} (S_+ + S_-) \quad (11.21) \]

and

\[ S_y = \frac{1}{2i} (S_+ - S_-). \quad (11.22) \]

From Section 9.6, we have

\[ S_+ | s, m_s \rangle = \hbar \sqrt{(s - m_s)(s + m_s + 1)} | s, m_s + 1 \rangle \quad (11.23) \]

and

\[ S_- | s, m_s \rangle = \hbar \sqrt{(s + m_s)(s - m_s + 1)} | s, m_s - 1 \rangle. \]

\[ a \]
Recall that these relations hold for any generalized momentum, including, of course, electron spin. All we require are the commutation relations
\[
[J_x, J_y] = i\hbar J_z, \\
[J_y, J_z] = i\hbar J_x, \\
and \\
[J_z, J_x] = i\hbar J_y;
\]
where \( J \) is \( S \) for us.

These relations give the following.

\[
S_+ |\uparrow\rangle = S_+ |s = 1/2, m_s = 1/2\rangle = 0 
\]
\[
S_+ |\downarrow\rangle = S_+ |1/2, -1/2\rangle = \hbar \sqrt{\left(\frac{1}{2} - \left(-\frac{1}{2}\right)\right) \left(\frac{1}{2} + \left(-\frac{1}{2}\right) + 1\right)} |1/2, 1/2\rangle \\
= \hbar |1/2, 1/2\rangle = \hbar |\uparrow\rangle 
\]
\[
S_- |\uparrow\rangle = S_- |1/2, 1/2\rangle = \hbar \sqrt{\left(\frac{1}{2} + \frac{1}{2}\right) \left(\frac{1}{2} - \frac{1}{2} + 1\right)} |1/2, -1/2\rangle \\
= \hbar |1/2, -1/2\rangle = \hbar |\downarrow\rangle 
\]
\[
S_- |\downarrow\rangle = S_- |1/2, -1/2\rangle = 0
\]

For \( S_x = \frac{1}{2} (S_+ + S_-) \),

\[
S_{x11} = \langle \uparrow |S_x| \uparrow \rangle = \frac{1}{2} (\langle \uparrow \mid (S_+ + S_-) \mid \uparrow \rangle) = \frac{1}{2} (\langle \uparrow \mid S_+ \mid \uparrow \rangle + \langle \uparrow \mid S_- \mid \uparrow \rangle) \\
= \frac{1}{2} (0 + \langle \uparrow \mid \hbar \mid \downarrow \rangle) = \frac{1}{2} (0 + 0) = 0 
\]
\[
S_{x12} = \langle \uparrow |S_x| \downarrow \rangle = \frac{1}{2} (\langle \uparrow \mid (S_+ + S_-) \mid \downarrow \rangle) = \frac{1}{2} (\langle \uparrow \mid S_+ \mid \downarrow \rangle + \langle \uparrow \mid S_- \mid \downarrow \rangle) \\
= \frac{1}{2} (\langle \uparrow \mid \hbar \mid \downarrow \rangle + 0) = \frac{\hbar}{2}  
\]
\[
S_{x21} = \langle \downarrow |S_x| \uparrow \rangle = \frac{1}{2} (\langle \downarrow \mid (S_+ + S_-) \mid \uparrow \rangle) = \frac{1}{2} (\langle \downarrow \mid S_+ \mid \uparrow \rangle + \langle \downarrow \mid S_- \mid \uparrow \rangle) \\
= \frac{1}{2} (0 + \langle \downarrow \mid \hbar \mid \uparrow \rangle) = \frac{\hbar}{2}  
\]
\[
S_{x22} = \langle \downarrow |S_x| \downarrow \rangle = \frac{1}{2} (\langle \downarrow \mid (S_+ + S_-) \mid \downarrow \rangle) = \frac{1}{2} (\langle \downarrow \mid S_+ \mid \downarrow \rangle + \langle \downarrow \mid S_- \mid \downarrow \rangle)
\]
\[
S_x = \begin{bmatrix}
0 & \frac{\hbar}{2} \\
\frac{\hbar}{2} & 0
\end{bmatrix} = \frac{\hbar}{2} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.
\]  
\[(11.33)\]

For \(S_y = \frac{1}{\sqrt{2}} (S_+ - S_-)\),

\[
S_{y11} = \langle \uparrow | S_y | \uparrow \rangle = \left\langle \uparrow \left| \frac{1}{2i} (S_+ - S_-) \right| \uparrow \right\rangle = \frac{1}{2i} (\langle \uparrow | S_+ | \uparrow \rangle - \langle \uparrow | S_- | \uparrow \rangle)
\]
\[
= \frac{1}{2i} \left( 0 - \langle \uparrow | \hbar | \downarrow \rangle \right) = \frac{1}{2i} (0 - 0) = 0
\]  
\[(11.34)\]

\[
S_{y12} = \langle \uparrow | S_y | \downarrow \rangle = \left\langle \uparrow \left| \frac{1}{2i} (S_+ - S_-) \right| \downarrow \right\rangle = \frac{1}{2i} (\langle \uparrow | S_+ | \downarrow \rangle - \langle \uparrow | S_- | \downarrow \rangle)
\]
\[
= \frac{1}{2i} \left( \langle \uparrow | \hbar | \uparrow \rangle + 0 \right) = \frac{\hbar}{2i} = -i \frac{\hbar}{2}
\]  
\[(11.35)\]

\[
S_{y21} = \langle \downarrow | S_y | \uparrow \rangle = \left\langle \downarrow \left| \frac{1}{2i} (S_+ - S_-) \right| \uparrow \right\rangle = \frac{1}{2i} (\langle \downarrow | S_+ | \uparrow \rangle - \langle \downarrow | S_- | \uparrow \rangle)
\]
\[
= \frac{1}{2i} \left( 0 - \langle \downarrow | \hbar | \downarrow \rangle \right) = -i \frac{\hbar}{2i} = \frac{\hbar}{2}
\]  
\[(11.36)\]

\[
S_{y22} = \langle \downarrow | S_y | \downarrow \rangle = \left\langle \downarrow \left| \frac{1}{2i} (S_+ - S_-) \right| \downarrow \right\rangle = \frac{1}{2i} (\langle \downarrow | S_+ | \downarrow \rangle - \langle \downarrow | S_- | \downarrow \rangle)
\]
\[
= \frac{1}{2i} \left( \langle \downarrow | \hbar | \uparrow \rangle + 0 \right) = \frac{1}{2i} (0 + 0) = 0
\]  
\[(11.37)\]

\[
S_y = \begin{bmatrix}
0 & -i \frac{\hbar}{2} \\
i \frac{\hbar}{2} & 0
\end{bmatrix} = \frac{\hbar}{2} \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}.
\]  
\[(11.38)\]

**Definition 11.1 (Pauli Matrices)** The Pauli matrices \(\sigma\) is defined by

\[
S = \frac{\hbar}{2} \sigma;
\]  
\[(11.39)\]

where \(S = (S_x, S_y, S_z)\) and \(\sigma = (\sigma_x, \sigma_y, \sigma_z)\). Component by component, we have the following matrix representations of \(\sigma_i\)'s in the orthonormal basis consisting of normalized eigenvectors of \(S_z\).
\[ \sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \text{and} \quad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \] 

(11.40)

In addition, we often define \( \sigma_0 \) as follows; namely, \( \sigma_0 \) is the identity matrix.

\[ \sigma_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \] 

(11.41)

Noting that

\[ i\sigma_y = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \] 

(11.42)

we can see

\[ \frac{1}{2}(\sigma_x + i\sigma_y) = \frac{1}{2}\left( \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \right) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \] 

(11.43)

Similarly,

\[ \frac{1}{2}(\sigma_x - i\sigma_y) = \frac{1}{2}\left( \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \right) = \frac{1}{2} \begin{bmatrix} 0 & 0 \\ 2 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}. \] 

(11.44)

\[ \frac{1}{2}(\sigma_0 + \sigma_z) = \frac{1}{2}\left( \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \right) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \] 

and

\[ \frac{1}{2}(\sigma_0 - \sigma_z) = \frac{1}{2}\left( \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \right) = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}. \] 

(11.45)

(11.46)

Therefore, \( \{\sigma_0, \sigma_x, \sigma_y, \sigma_z\} \) forms a basis for the vector space spanned by

\[ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}. \] 

(11.47)

Properties of Pauli Matrices

1. \( |\uparrow\rangle \) and \( |\downarrow\rangle \) are normalized eigenvectors of \( \sigma_z \), and also \( S_z \). Similarly, \( |\uparrow\rangle_x \) and \( |\downarrow\rangle_x \) are normalized eigenvectors of \( \sigma_x \), and \( |\uparrow\rangle_y \) and \( |\downarrow\rangle_y \) are normalized
CHAPTER 11. ELECTRON SPIN

eigenvectors of $\sigma_y$. Here, $|\uparrow\rangle_x$ and $|\downarrow\rangle_x$ are the states corresponding to spin up and spin down along the $x$-axis; and likewise for $|\uparrow\rangle_y$ and $|\downarrow\rangle_y$.\(^2\)

2. Pauli spin matrices are Hermitian. For example,

$$\sigma_y^\dagger = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}^\dagger = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} = \sigma_y, \quad (11.48)$$

and similarly for $\sigma_x$ and $\sigma_z$.

3. Pauli matrices are unitary. For example,

$$\sigma_y^\dagger \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}^\dagger \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} = \begin{bmatrix} (-i)i & 0 \\ 0 & i(-i) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (11.49)$$

Likewise for $\sigma_x$ and $\sigma_z$.

4. Pauli matrices anti-commute; i.e. the anti-commutator

$$[\sigma_i, \sigma_j]_+ = \sigma_i \sigma_j + \sigma_j \sigma_i = 0 \quad \text{if } i \neq j \quad \text{and} \quad i, j \neq 0. \quad (11.50)$$

5. We have the following cyclic relations.

$$\sigma_x \sigma_y = i\sigma_z, \quad \sigma_y \sigma_z = i\sigma_x, \quad \text{and} \quad \sigma_z \sigma_x = i\sigma_y. \quad (11.51)$$

6. Pauli matrices are traceless.

$$\text{Tr } \sigma_x = \text{Tr } \sigma_y = \text{Tr } \sigma_z = 0 \quad (11.52)$$

7. Pauli matrices are idempotent, which is a direct consequence of Properties 2 and 3.

$$\sigma_i^2 = I. \quad (11.53)$$

\(^2\)The eigenvalues of $S_x$, $S_y$, and $S_z$ are $\pm \frac{\hbar}{2}$; while the eigenvalues are $\pm 1$ for the rescaled $\sigma_x$, $\sigma_y$, and $\sigma_z$.

\(^3\)Some authors use $\{\}$ for the anti-commutator. However, $\{\}$ is used for the Poisson bracket in classical mechanics. So, we will use $[\ ]_+$ instead.
11.3 Sequential Measurements

We will now consider consecutive Stern-Gerlach type experiments. This will demonstrate how Postulates 3, 5, and 6 from Chapter 3 work.

We need to do some preliminary work first. In the $S_z$ basis $\{|\uparrow\rangle, |\downarrow\rangle\}$, we have

\[
\sigma_x \frac{|\uparrow\rangle + |\downarrow\rangle}{\sqrt{2}} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \left( \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\
= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1 \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 1 \cdot \frac{|\uparrow\rangle + |\downarrow\rangle}{\sqrt{2}},
\]

and

\[
\sigma_x \frac{|\uparrow\rangle - |\downarrow\rangle}{\sqrt{2}} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \left( \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\
= \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = -1 \cdot \frac{1}{\sqrt{2}} \left( \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) = -1 \cdot \frac{|\uparrow\rangle - |\downarrow\rangle}{\sqrt{2}}.
\]

Therefore,

\[
|\uparrow\rangle_x = \frac{1}{\sqrt{2}} (|\uparrow\rangle + |\downarrow\rangle) \\
\text{and} \\
|\downarrow\rangle_x = \frac{1}{\sqrt{2}} (|\uparrow\rangle - |\downarrow\rangle).
\]

Similarly,

\[
\sigma_y \frac{|\uparrow\rangle + i|\downarrow\rangle}{\sqrt{2}} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \left( \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ i \end{bmatrix} \right) = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \begin{bmatrix} 1 \\ i \end{bmatrix} \\
= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ i \end{bmatrix} = 1 \cdot \frac{1}{\sqrt{2}} \left( \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ i \end{bmatrix} \right) = 1 \cdot \frac{|\uparrow\rangle + i|\downarrow\rangle}{\sqrt{2}},
\]

and

\[
\sigma_y \frac{|\uparrow\rangle - i|\downarrow\rangle}{\sqrt{2}} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \left( \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ i \end{bmatrix} \right) = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -i \end{bmatrix} \\
= \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ i \end{bmatrix} = -1 \cdot \frac{1}{\sqrt{2}} \left( \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ i \end{bmatrix} \right) = -1 \cdot \frac{|\uparrow\rangle - i|\downarrow\rangle}{\sqrt{2}}.
\]
Therefore,

\[
|\uparrow\rangle_y = \frac{1}{\sqrt{2}}(|\uparrow\rangle + i|\downarrow\rangle), \quad (11.60)
\]

and

\[
|\downarrow\rangle_y = \frac{1}{\sqrt{2}}(|\uparrow\rangle - i|\downarrow\rangle). \quad (11.61)
\]

Note that (11.56), (11.57), (11.60), and (11.61) also mean

\[
|\uparrow\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle_x + |\downarrow\rangle_x) \quad \text{and} \quad |\downarrow\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle_x - |\downarrow\rangle_x), \quad (11.62)
\]

\[
|\uparrow\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle_y + |\downarrow\rangle_y) \quad \text{and} \quad |\downarrow\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle_y - |\downarrow\rangle_y). \quad (11.63)
\]

Now we are ready to embark on a study of sequential Stern-Gerlach experiments/measurements. Stern-Gerlach device does not only allow us to detect two components, spin up and spin down, but also enables us to isolate one variety of electrons, be it spin up or spin down. This is the feature we will need.

We start with a state

\[
\alpha|\uparrow\rangle + \beta|\downarrow\rangle \quad \text{with} \quad |\alpha|^2 + |\beta|^2 = 1. \quad (11.64)
\]

The condition on \(\alpha\) and \(\beta\) assures that the linear combination \(\alpha|\uparrow\rangle + \beta|\downarrow\rangle\) is of unit length; i.e. it is indeed a state. The fact that we can only observe \(S_z = +\frac{h}{2}\) or \(S_z = -\frac{h}{2}\) is a consequence of Postulate 3 and the matrix representation of \(S_z\) given by (11.20), which implies that the eigenvalues of \(S_z\) are indeed \(+\frac{h}{2}\) and \(-\frac{h}{2}\). After the electrons in this state go through the first Stern-Gerlach device, you will observe spin up \(|\uparrow\rangle\) and spin down \(|\downarrow\rangle\) states with probabilities \(|\alpha|^2\) and \(|\beta|^2\), respectively according to Postulate 5.

If an electron with the spin wavefunction (11.64) is found to have an up-spin, Postulate 6 forces its wavefunction to be \(|\uparrow\rangle\) after the measurement. So, we have
11.3. SEQUENTIAL MEASUREMENTS

\[ \alpha |\uparrow\rangle + \beta |\downarrow\rangle \xrightarrow{\text{first measurement}} |\uparrow\rangle. \quad (11.65) \]

Having isolated \(|\uparrow\rangle\), we now have a pure eigenstate with an up-spin, and the second Stern-Gerlach device find \(|\uparrow\rangle\) with a probability 1.0. This obvious consequence is described in (11.66).

\[ \alpha |\uparrow\rangle + \beta |\downarrow\rangle \xrightarrow{\text{first measurement}} |\uparrow\rangle \xrightarrow{\text{second measurement}} |\uparrow\rangle \quad (11.66) \]

Next, let us see what will happen if the second Stern-Gerlach device is rotated by 90° so that it is now in the \(x\)-direction. With what probabilities will we observe \(|\uparrow\rangle_x\) and \(|\downarrow\rangle_x\)? The answer lies in (11.62), which gives

\[ |\uparrow\rangle = \frac{1}{\sqrt{2}} (|\uparrow\rangle_x + |\downarrow\rangle_x). \quad (11.67) \]

According to Postulate 5, the probabilities of observing each eigenstate is given by \(\left| \frac{1}{\sqrt{2}} \right|^2\). Therefore, we will observe the up-spin 50% of the time and the down-spin 50% of the time.

\[ \alpha |\uparrow\rangle + \beta |\downarrow\rangle \xrightarrow{\text{first measurement}} |\uparrow\rangle \xrightarrow{\text{second measurement in the } x\text{-direction}} |\uparrow\rangle_x \text{ (50\%) and } |\downarrow\rangle_x \text{ (50\%) } \quad (11.68) \]

Similarly, we have

\[ \alpha |\uparrow\rangle + \beta |\downarrow\rangle \xrightarrow{\text{first measurement to choose the spin-down state}} |\downarrow\rangle \xrightarrow{\text{second measurement in the } x\text{-direction}} |\uparrow\rangle_x \text{ (50\%) and } |\downarrow\rangle_x \text{ (50\%) } \quad (11.69) \]

as

\[ |\downarrow\rangle = \frac{1}{\sqrt{2}} (|\uparrow\rangle_x - |\downarrow\rangle_x) \quad (11.70) \]
from (11.63).

What if $|\uparrow\rangle_x$ goes through a Stern-Gerlach device oriented in the $y$-direction? By symmetry, we should observe the up-spin and down-spin states with a probability of 0.5, respectively. Let us double check this by an explicit computation. By (11.56), (11.62), and (11.63),

$$|\uparrow\rangle_x = \frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle) = \frac{1}{\sqrt{2}} \left( \frac{1}{\sqrt{2}}(|\uparrow\rangle_y + |\downarrow\rangle_y) + \frac{1}{\sqrt{2}i}(|\uparrow\rangle_y - |\downarrow\rangle_y) \right)$$

$$= \frac{1}{2}(|\uparrow\rangle_y + |\downarrow\rangle_y) + \frac{1}{2i}(|\uparrow\rangle_y - |\downarrow\rangle_y) = \frac{1}{2}(|\uparrow\rangle_y + |\downarrow\rangle_y) - \frac{1}{2i}(|\uparrow\rangle_y - |\downarrow\rangle_y)$$

$$= \frac{1-i}{2} |\uparrow\rangle_y + \frac{1+i}{2} |\downarrow\rangle_y. \quad (11.71)$$

Because

$$\left| \frac{1-i}{2} \right|^2 = \left( \frac{\sqrt{2}}{2} \right)^2 = \frac{1}{2} \quad \text{and} \quad \left| \frac{1+i}{2} \right|^2 = \left( \frac{\sqrt{2}}{2} \right)^2 = \frac{1}{2}, \quad (11.72)$$

our intuition is supported mathematically according to Postulate 5.

### 11.4 Spin States in an Arbitrary Direction

As our system possesses spherical symmetry, the choice of the $z$-axis is arbitrary. In other words, we should observe the same physics despite different choices of the $z$-direction. The following is a mathematical treatment of this fact.

#### 11.4.1 Spin Matrix $S_u$ and Its Eigenvectors

Recall the spin operator $S$ given in (11.1).

$$S = S_x \hat{x} + S_y \hat{y} + S_z \hat{z}. \quad (11.73)$$

This is so designed that the following “inner products” provide the spin matrices in the $x$-, $y$-, and $z$-directions.

$$S \cdot \hat{x} = S_x, \quad S \cdot \hat{y} = S_y, \quad S \cdot \hat{z} = S_z \quad (11.74)$$
11.4. Spin States in an Arbitrary Direction

Analogously, the spin matrix in the direction defined by a unit vector \( \hat{u} \) extending from the origin to the point \((r, \theta, \phi)\), \( \hat{u} = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \phi) \), is given by

\[
S \cdot \hat{u} = (S_x, S_y, S_z) \cdot (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \phi) \\
= \sin \theta \cos \phi S_x + \sin \theta \sin \phi S_y + \cos \theta S_z \\
= \sin \theta \cos \phi \left[ 0 \quad \frac{\hbar}{2} \quad 0 \right] + \sin \theta \sin \phi \left[ 0 \quad -i\frac{\hbar}{2} \quad 0 \right] + \cos \theta \left[ +\frac{\hbar}{2} \quad 0 \quad -\frac{\hbar}{2} \right] \\
= \frac{\hbar}{2} \left[ \cos \theta \sin \theta(\cos \phi - i \sin \phi) \quad -\cos \theta \right] = \frac{\hbar}{2} \left[ \cos \theta \sin e^{i\phi} - \cos \theta \right] \\
\tag{11.75}
\]

Let us compute the eigenvalues and eigenvectors of \((11.75)\). It suffices to consider

\[
\begin{bmatrix}
\cos \theta & \sin \theta e^{-i\phi} \\
\sin \theta e^{i\phi} & -\cos \theta
\end{bmatrix}
\]

instead. If we denote the eigenvalues by \(\lambda\) and \(\lambda'\), we have, by \((A.8)\),

\[
\lambda + \lambda' = \text{trace} \begin{bmatrix}
\cos \theta & \sin \theta e^{-i\phi} \\
\sin \theta e^{i\phi} & -\cos \theta
\end{bmatrix} = 0 \implies \lambda' = -\lambda \tag{11.77}
\]

and

\[
\lambda \lambda' = -\lambda^2 = \det \begin{bmatrix}
\cos \theta & \sin \theta e^{-i\phi} \\
\sin \theta e^{i\phi} & -\cos \theta
\end{bmatrix} = -\cos^2 \theta - \sin^2 \theta = -1. \tag{11.78}
\]

So, the two eigenvalues of \(\begin{bmatrix}
\cos \theta & \sin \theta e^{-i\phi} \\
\sin \theta e^{i\phi} & -\cos \theta
\end{bmatrix}\) are +1 and −1, which in turn implies that the two eigenvalues of \(\frac{\hbar}{2} \begin{bmatrix}
\cos \theta & \sin \theta e^{-i\phi} \\
\sin \theta e^{i\phi} & -\cos \theta
\end{bmatrix}\) are \(+\frac{\hbar}{2}\) and \(-\frac{\hbar}{2}\). You can clearly see here that the electron spin is nothing like the orbital angular momentum, either classically or quantum mechanically, as the values along any axis are \(\pm \frac{\hbar}{2}\), irrespective of \(\theta\) and \(\phi\). Recall that quantum mechanical angular momentum has \(2l + 1\) values for \(L_z\). This is an odd number while the electron spin has an even number of components; namely, two. Also, compare this with a classical orbital angular momentum vector \(L\), for which \(L_z\) depends on the angle \(\theta\) formed by \(L\) and the \(z\)-axis.

Next, we will determine the eigenvectors. As usual, this is only up to an arbitrary scaling factor. Let \(\begin{bmatrix} a \\ b \end{bmatrix}\) be an eigenvector of \(\begin{bmatrix}
\cos \theta & \sin \theta e^{-i\phi} \\
\sin \theta e^{i\phi} & -\cos \theta
\end{bmatrix}\). Then,

\[
\begin{bmatrix}
\cos \theta & \sin \theta e^{-i\phi} \\
\sin \theta e^{i\phi} & -\cos \theta
\end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a \cos \theta + b \sin \theta e^{-i\phi} \\ a \sin \theta e^{i\phi} - b \cos \theta \end{bmatrix} = \pm 1 \cdot \begin{bmatrix} a \\ b \end{bmatrix}. \tag{11.79}
\]
Consider

\[
\begin{bmatrix}
  a \cos \theta + b \sin \theta e^{-i\phi} \\
  a \sin \theta e^{i\phi} - b \cos \theta
\end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}
\]  

(11.80)

first.

\[
\begin{align*}
  a \cos \theta + b \sin \theta e^{-i\phi} &= a \\
  a \sin \theta e^{i\phi} - b \cos \theta &= b
\end{align*}
\]

\[
\implies \begin{cases}
  b \sin \theta e^{-i\phi} = (1 - \cos \theta)a \\
  a \sin \theta e^{i\phi} = (1 + \cos \theta)b
\end{cases}
\]

\[
\implies \begin{cases}
  a = \frac{\cos \theta}{\sin \theta} e^{-i\phi} \\
  b = \frac{\sin \theta e^{i\phi}}{\sin \theta e^{i\phi}}
\end{cases}
\]

(11.81)

Next, consider

\[
\begin{bmatrix}
  a \cos \theta + b \sin \theta e^{-i\phi} \\
  a \sin \theta e^{i\phi} - b \cos \theta
\end{bmatrix} = -1 \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} -a \\ -b \end{bmatrix}.
\]  

(11.82)

\[
\begin{align*}
  a \cos \theta + b \sin \theta e^{-i\phi} &= -a \\
  a \sin \theta e^{i\phi} - b \cos \theta &= -b
\end{align*}
\]

\[
\implies \begin{cases}
  b \sin \theta e^{-i\phi} = -(1 + \cos \theta)a \\
  a \sin \theta e^{i\phi} = -(1 - \cos \theta)b
\end{cases}
\]

\[
\implies \begin{cases}
  a = \frac{\sin \theta e^{-i\phi}}{\sin \theta e^{i\phi}} \\
  b = \frac{-\cos \theta}{\sin \theta} e^{i\phi}
\end{cases}
\]

(11.83)

This gives us the following set of eigenvectors.

\[
|\uparrow\rangle_u = \begin{bmatrix} \cos \frac{\theta}{2} e^{-i\phi} \\ \sin \frac{\theta}{2} \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} \cos \frac{\theta}{2} e^{-i\phi/2} \\ \sin \frac{\theta}{2} e^{i\phi/2} \end{bmatrix}
\]  

(11.84)

\[
|\downarrow\rangle_u = \begin{bmatrix} \sin \frac{\theta}{2} e^{-i\phi} \\ -\cos \frac{\theta}{2} \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} \sin \frac{\theta}{2} e^{-i\phi/2} \\ -\cos \frac{\theta}{2} e^{i\phi/2} \end{bmatrix}
\]  

(11.85)
11.4.2 Sequential Measurements Revisited

Recalling that we are in the $S_z$-basis, (11.84) and (11.85) mean

$$|\uparrow\rangle_u = \cos \frac{\theta}{2} e^{-i\phi} |\uparrow\rangle + \sin \frac{\theta}{2} |\downarrow\rangle$$  \hspace{1cm} (11.86)

and

$$|\downarrow\rangle_u = \sin \frac{\theta}{2} e^{-i\phi} |\uparrow\rangle - \cos \frac{\theta}{2} |\downarrow\rangle.$$  \hspace{1cm} (11.87)

Suppose we have made a measurement of the spin in the $\hat{u}$ direction and found an up-spin state, the ket after the measurement is, according to Postulate 6 on p.74, Chapter 3,

$$|\uparrow\rangle_u = \cos \frac{\theta}{2} e^{-i\phi} |\uparrow\rangle + \sin \frac{\theta}{2} |\downarrow\rangle.$$  \hspace{1cm} (11.88)

If we further measure the spin in the $z$-direction, the probability of finding an up-spin is

$$|\langle \uparrow | \uparrow \rangle_u|^2 = \left| \cos \frac{\theta}{2} e^{-i\phi} \langle \uparrow | \uparrow \rangle + \sin \frac{\theta}{2} \langle \uparrow | \downarrow \rangle \right|^2 = \left| \cos \frac{\theta}{2} e^{-i\phi} \right|^2 = \cos^2 \frac{\theta}{2},$$  \hspace{1cm} (11.89)

and the probability for the down-spin is

$$|\langle \downarrow | \uparrow \rangle_u|^2 = \left| \cos \frac{\theta}{2} e^{-i\phi} \langle \downarrow | \uparrow \rangle + \sin \frac{\theta}{2} \langle \downarrow | \downarrow \rangle \right|^2 = \left| \sin \frac{\theta}{2} \right|^2 = \sin^2 \frac{\theta}{2},$$  \hspace{1cm} (11.90)

according to Postulate 5 on p.74, Chapter 3.

11.4.3 Comparison with the Classical Theory

As mentioned on p.89 in Chapter 3, the expectation values play the role of the classical variables in the quantum mechanical formulation of classical laws of physics. Let us see how this plays out for electron spin.

If you have the state $|\uparrow\rangle_u$, the angular momentum vector of length $\frac{\hbar}{2}$ is pointing in the direction of $\hat{u} = (r, \theta, \phi)$. Hence, the $z$-component is $\frac{\hbar}{2} \cos \theta$ in the classical picture. Let us compute the expectation value of $S_z$ for the state $|\uparrow\rangle_u$. Our computation will be in the $S_z$-basis.

$$u \langle \uparrow | S_z | \uparrow \rangle_u = \left[ \cos \frac{\theta}{2} e^{i\phi} \sin \frac{\theta}{2} \right] \begin{bmatrix} + \frac{\hbar}{2} & 0 \\ 0 & - \frac{\hbar}{2} \end{bmatrix} \begin{bmatrix} \cos \frac{\theta}{2} e^{-i\phi} \\ \frac{\hbar}{2} \sin \frac{\theta}{2} \end{bmatrix}$$
\[ \begin{align*}
&= \left[ \cos \frac{\theta}{2} e^{i\phi} \sin \frac{\theta}{2} \right] + \frac{\hbar}{2} \cos \frac{\theta}{2} e^{-i\phi} \\
&= \frac{\hbar}{2} \left( \cos \frac{\theta}{2} - \sin \frac{\theta}{2} \right) = \frac{\hbar}{2} \cos \left( 2 \cdot \frac{\theta}{2} \right) = \frac{\hbar}{2} \cos \theta
\end{align*} \]

This confirms that the quantum mechanical expectation value indeed plays the role of the corresponding classical variable in this case as well.
1. Consider an electron in the initial spin state

\[ \frac{1}{\sqrt{2}} |\uparrow\rangle + \frac{1}{\sqrt{2}} |\downarrow\rangle. \]

(a) If two Stern-Gerlach devices are sequentially arranged so that the first device measures the spin in the \( z \)-direction and the second in the \( y \)-direction, what is the probability that the first device measures an up-spin in the \( z \)-direction and the second device measures a down-spin in the \( y \)-direction?

(b) In the above, what is the final spin state of the electron?
Chapter 12

Addition of Orbital and Spin Angular Momenta
Chapter 13

Molecular Rotation

There is only one three-dimensional motion for a point mass such as an electron; namely translation or simply the motion of its center of mass, i.e. its position, in the three-dimensional space. However, for entities with more complex structures such as molecules, there are three types of motion; translation, vibration, and rotation. Translation is the change of the position of its center of mass and vibration is as in a simple harmonic oscillator. For rotation, think of a top spinning while remaining at the same spot. The point is that translation is the only motion that involves a change in the center of gravity, and three dimensional motions of molecules can be decomposed into translation, vibration, and rotation. As we have already discussed translational motion of particles and the simple harmonic oscillator as an example of vibration of diatomic molecules, rotation is our final frontier. We will only consider a rigid body in this chapter.

13.1 Rotational Kinetic Energy

The classical formula for the rotational kinetic energy $E$ of a freely rotating molecule is

$$E = \frac{1}{2} I_x \omega_x^2 + \frac{1}{2} I_y \omega_y^2 + \frac{1}{2} I_z \omega_z^2,$$

(13.1)

where $x$, $y$, and $z$ are principal axes of rotation, and $I_x$, $I_y$, and $I_z$ are the moments of inertia about the $x$-, $y$-, and $z$-axes. As the angular momenta about the principal

\footnote{There is no such thing as rotational potential energy, and we only concern ourselves with rotational kinetic energy.}
axes are given by

\[ L_x = I_x \omega_x, \quad L_y = I_y \omega_y, \quad \text{and} \quad L_z = I_z \omega_z, \]  

(13.2)

we can also express the rotational kinetic energy as follows in terms of the angular momenta about the principal axes.

\[ E = \frac{L_x^2}{2I_x} + \frac{L_y^2}{2I_y} + \frac{L_z^2}{2I_z}; \]  

(13.3)

The quantum mechanical Hamiltonian is obtained by the usual substitution of classical observables by the corresponding quantum mechanical operators. Namely,

\[ H = \frac{L_x^2}{2I_x} + \frac{L_y^2}{2I_y} + \frac{L_z^2}{2I_z}; \]  

(13.4)

where

\[ L_x = -i\hbar \left( y \frac{\partial}{\partial z} - z \frac{\partial}{\partial y} \right) = i\hbar \left( \sin \phi \frac{\partial}{\partial \theta} + \cot \theta \cos \phi \frac{\partial}{\partial \phi} \right) \]

\[ L_y = -i\hbar \left( z \frac{\partial}{\partial x} - x \frac{\partial}{\partial z} \right) = i\hbar \left( -\cos \phi \frac{\partial}{\partial \theta} + \cot \theta \sin \phi \frac{\partial}{\partial \phi} \right) \]  

(13.5)

\[ L_z = -i\hbar \left( x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right) = -i\hbar \frac{\partial}{\partial \phi}. \]

from (9.2) and (9.3).

### 13.2 Moment of Inertia for a Diatomic Molecule

Moment of inertia is defined relative to a chosen axis of rotation. For a point mass, the moment of inertia is defined as the product of the mass and the square of perpendicular distance to the axis of rotation. Therefore,

\[ I = mr^2, \]  

(13.6)

where \( m \) is the mass and \( r \) is the distance between the point mass and the axis of rotation. To a first approximation, a diatomic molecule can be regraded as two pint masses \( M \) and \( m \) connected by a rigid massless rod of length \( r \). Now introduce
a canonical coordinate axes, call them the $x$- and $y$-axes, so that the massless rod lies on the $x$-axis with the center of mass of the molecule at the origin. Suppose the coordinate of the point mass $M$ is $(r_1, 0)$, and that of $m$ is $(-r_2, 0)$, where the positive numbers $r_1$ and $r_2$ satisfy $r_1 + r_2 = r$. Our axis of rotation is the $y$-axis that goes through the center of mass. Then, the moment of inertia $I$ for this diatomic molecule is the sum of the moment for the two point masses to a first approximation.

$$I = Mr_1^2 + mr_2^2$$

(13.7)

Now, because the center of mass is at the origin, we have

$$\frac{Mr_1 - mr_2}{M + m} = 0 \implies Mr_1 = mr_2.$$  

(13.8)

We can use this relation to express $r_1$ and $r_2$ in terms of $r$.

$$Mr_1 = mr_2 \implies r_1 = \frac{m}{M} r_2 = \frac{m}{M} (r - r_1) \implies \left(1 + \frac{m}{M}\right) r_1 = \frac{m}{M} r$$

$$\frac{M + m}{M} r_1 = \frac{m}{M} r \implies r_1 = \frac{m}{M + m} r \implies r_2 = \frac{M}{m} r_1 = \frac{M}{M + m} r$$

(13.9)

Plugging these into (13.7),

$$I = Mr_1^2 + mr_2^2 = M \frac{m^2}{(M + m)^2} r^2 + m \frac{M^2}{(M + m)^2} r^2$$

$$= \frac{Mm(m + M)}{(M + m)^2} r^2 = \frac{Mm}{M + m} r^2.$$  

(13.10)

Recall that $\mu = \frac{Mm}{M+m}$ is the reduced mass. Therefore, we get

$$I = \mu r^2.$$  

(13.11)

### 13.3 Two-Dimensional Rotation Confined to the $x, y$-Plane

The angular momentum vector is always parallel to the $z$-axis; i.e. $J = J_z \hat{k}^2$. But, the difference from the classical counterpart is the quantization of $J_z$. Without loss of generality, we will specialize to rotations about the $z$-axis. As $L_x$ and $L_y$ are 0 in (13.4) and (13.5), we get

\[ \text{In the literature about molecular rotation, the angular momentum vector is typically denoted by } J \text{ rather than } L. \]
\[ H = \frac{L_z^2}{2I} = -\frac{i\hbar}{2I} \frac{\partial}{\partial \phi} \left( -\frac{i\hbar}{2I} \frac{\partial}{\partial \phi} \right) = -\frac{\hbar^2}{2I} \frac{\partial^2}{\partial \phi^2}; \]  

(13.12)

where we dropped the subscript \( z \) from \( I_z \) because \( I_z \) is the only nonzero moment of inertia in our situation. Our Schrödinger equation is

\[ H\psi = E\psi \implies -\frac{\hbar^2}{2I} \frac{\partial^2 \psi}{\partial \phi^2} = E\psi. \]  

(13.13)

Because \( \psi \) here is a function of \( \phi \) alone, and also because it is basically the same as \( \Phi \) encountered in Chapter 10 “The Hydrogen Atom”, it is customary to denote this wavefunction by \( \Phi \). With this convention, we now have

\[-\frac{\hbar^2}{2I} \frac{\partial^2 \Phi(\phi)}{\partial \phi^2} = E\Phi(\phi). \]  

(13.14)

The solutions are

\[ \Phi_m(\phi) = \sqrt{\frac{1}{2\pi}} e^{im\phi} \quad \text{for} \quad m = 0, \pm 1, \pm 2, \ldots. \]  

(13.15)

As in the case of the hydrogen atom, possible values for \( m \) are determined by the single-valuedness condition; namely, \( \Phi(\phi) = \Phi(\phi + 2n\pi) \) for any integer \( n \). For the total energy, we have

\[ m = \pm \sqrt{2IE} \quad \implies \quad E = \frac{m^2 \hbar^2}{2I} \quad \text{for} \quad m = 0, \pm 1, \pm 2, \ldots. \]  

(13.16)

The term \( m^2 \) in the expression for the energy implies that we have two-fold degeneracy for each value of \( m \) except for \( m = 0 \), which means there is no rotation. In the classical picture, we can regard this as deriving from the sense of rotation, clockwise or counter-clockwise.
13.4 Three-Dimensional Rotation

In this section, we still consider a rotational motion confined to a plane. However, unlike in Section 13.3 where the rotation was confined to the \(x, y\)-plane, we now allow the plane of rotation to be tilted relative to the \(z\)-axis. In other words, it is necessary to give the magnitude of angular momentum \(\|\mathbf{J}\|\) and its tilt from the \(z\)-axis, measured by \(J_z\) as before, to completely specify the rotation.

From (10.41) the full Hamiltonian in spherical coordinates is

\[
\frac{-\hbar^2}{2\mu} \left[ \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \varphi^2} \right]
\]

\[
= \frac{-\hbar^2}{2\mu r^2} \left[ \frac{\partial}{\partial r} \left( r^2 \frac{\partial}{\partial r} \right) + \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \varphi^2} \right].
\]  

(13.17)

Because we are dealing with a rigid molecule, \(r\) is fixed, and there is no kinetic energy of radial motion. So, the term

\[
\frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial}{\partial r} \right)
\]

vanishes. Also, we know from (13.11) that \(I = \mu r^2\). Therefore, the Hamiltonian for three-dimensional rotation of a rigid body is given by

\[
\frac{-\hbar^2}{2I} \left[ \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \varphi^2} \right].
\]

(13.19)

This is basically the same as the \(\theta\)- and \(\varphi\)-dependent parts of the Hamiltonian for the hydrogen atom given in Chapter 10. The time-independent Schrödinger equation is

\[
\frac{-\hbar^2}{2I} \left[ \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \varphi^2} \right] \psi(\theta, \varphi) = E \psi(\theta, \varphi).
\]

(13.20)

Separation of variables work here as well. So, let

\[
\psi(\theta, \varphi) = \Theta(\theta)\Phi(\varphi).
\]

(13.21)

Then, we will get
where $P_j^{m_J}(\cos \theta)$'s denote the associated Legendre polynomials first encountered in (10.95) on p.214. We usually use the symbol $\hbar$ for $\frac{\hbar}{2I}$, and so, we have

\[ E = BJ(J + 1). \] (13.28)

As for $m_J$, it is proportional to the $z$-component $J_z$ of the rotational angular momentum; namely,

\[ J_z = m_J \hbar. \] (13.29)

This is another example of degeneracy in higher dimensions discussed in Section 8.1. The total rotational energy depends only on $J$, but there are $2J + 1$ distinct states.
with \( m_J = -J, -J + 1, \ldots, 0, \ldots, J - 1, J \) which share the same \( J \), and hence, have the same total energy \( E \).

In the most general case, we have \( I_x, I_y, \) and \( I_z \) which are different from each other. For a rigid body like this, we need to introduce three constants \( A = \frac{\hbar^2}{2I_x}, B = \frac{\hbar^2}{2I_y}, \) and \( C = \frac{\hbar^2}{2I_z} \) rather than just \( B = \frac{\hbar^2}{2I} \). Interested readers should consult “Microwave Molecular Spectra (3rd edition)” by Gordy and Cook [Gordy and Cook, 1984].
Appendices
Appendix A

Matrix Algebra

A.1 Invertible Matrices

Fact A.1 (Invertible Matrix) Let $A$ be a square $n \times n$ matrix. Then the following statements are equivalent. Hence, they are either all true or all false.

1. $A$ is an invertible matrix.
2. $A$ is bijective; that is, $A$ is one-to-one and onto.
3. $A$ is row equivalent to the $n \times n$ identity matrix.
4. $A$ has $n$ pivot positions.
5. The equation $Ax = 0$ has only the trivial solution $0$.
6. The null space of $A$ is $\{0\}$.
7. The dimension of the null space of $A$ is $0$.
8. The columns of $A$ form a linearly independent set.
9. The linear transformation $x \mapsto Ax$ is one-to-one.
10. The equation $Ax = b$ has at least one solution for each $b \in \mathbb{R}^n$.
11. The columns of $A$ span $\mathbb{R}^n$.
12. The linear transformation $x \mapsto Ax$ maps $\mathbb{R}^n$ onto $\mathbb{R}^n$.
13. There is an $n \times n$ matrix $L$ such that $LA = I$.
14. There is an \( n \times n \) matrix \( R \) such that \( AR = I \).

15. \( A^t \) is an invertible matrix.

16. The columns of \( A \) form a basis of \( \mathbb{R}^n \).

17. The column space \( C(A) = \mathbb{R}^n \).

18. The dimension of \( C(A) = n \).

19. \( \text{rank } A = n \)

20. The number 0 is not an eigenvalue of \( A \).

21. \( \det A \neq 0 \)

### A.2 Rank of a Matrix

**Definition A.1 (Rank of a Matrix)** The column rank of a matrix \( A \) is the maximum number of linearly independent column vectors of \( A \). The row rank of \( A \) is the maximum number of linearly independent row vectors of \( A \). Equivalently, the column rank of \( A \) is the dimension of the column space of \( A \), while the row rank of \( A \) is the dimension of the row space of \( A \). The column rank and the row rank are always equal, and we call this number simply the rank of \( A \), denoted by \( \text{rank } A \).

**Theorem A.1 (Column Rank = Row Rank)** The column rank and the row rank of a matrix \( A \) are equal.

The following proofs are basically from Wikipedia.

**Proof**

Consider an \( m \times n \) matrix \( A \) whose \( i \)-th row and \( j \)-th column entry is \( A_{ij} \) and column rank is \( r \). Pick an arbitrary basis \( \{v_1, v_2, \ldots, v_r\} \) of the column space of \( A \). Each basis vector is an \( m \times 1 \) column vector, and we can form an \( m \times r \) matrix

\[
C = [v_1 \ v_2 \ \cdots \ v_r] = [v_{1k}^{1 \leq k \leq r}]_{1 \leq i \leq r}.
\]  

(A.1)

\(^1\)Alternatively, the row rank of \( A \) is the number of non-zero rows in the row reduced form of \( A \).
Note that we are writing the $k$-th component of the $m \times 1$ column vector $v_l$ as $v_{kl}$ with $1 \leq k \leq m$ such that

$$ v_l = \begin{bmatrix} v_{1l} \\ v_{2l} \\ \vdots \\ v_{ml} \end{bmatrix} \quad (A.2) $$

for bookkeeping purposes. The $j$-th column $[A_j]$ of $A$ can be expressed as a linear combination of $\{v_1, v_2, \ldots, v_r\}$, or $\{v_i\}$ for short, with coefficients $b_{ij}$ such that

$$ [A_j] = \sum_{p=1}^{r} b_{pj} v_p \quad \text{for} \quad 1 \leq j \leq n. \quad (A.3) $$

In its full glory, (A.3) can be written as follows.

$$ \begin{bmatrix} A_{1j} \\ A_{2j} \\ \vdots \\ A_{mj} \end{bmatrix} = \begin{bmatrix} \sum_{p=1}^{r} v_{1p} b_{pj} \\ \sum_{p=1}^{r} v_{2p} b_{pj} \\ \vdots \\ \sum_{p=1}^{r} v_{mp} b_{pj} \end{bmatrix} \quad (A.4) $$

Hence,

$$ A_{ij} = \sum_{p=1}^{r} v_{ip} b_{pj} \implies A = CB; \quad (A.5) $$

where $A = [A_{ij}]_{1 \leq i \leq n}^{1 \leq j \leq m}$ is the original $m \times n$ matrix, $C = [v_{kl}]_{1 \leq k \leq m}^{1 \leq l \leq r}$ is the $m \times r$ matrix whose columns are the basis vectors $\{v_1, v_2, \ldots, v_r\}$, and $B = [b_{pj}]_{1 \leq p \leq r}^{1 \leq j \leq n}$ is the $r \times n$ coefficient matrix defined by (A.3). Now consider the $i$-th row of $A$ denoted here by $[A^i]$.

$$ [A^i] = [A_{i1} \quad A_{i2} \quad \ldots \quad A_{im}] = \left[ \sum_{p=1}^{r} v_{ip} b_{p1} \sum_{p=1}^{r} v_{ip} b_{p2} \ldots \sum_{p=1}^{r} v_{ip} b_{pn} \right] $$

$$ = \sum_{p=1}^{r} v_{ip} \begin{bmatrix} b_{p1} \\ b_{p2} \\ \vdots \\ b_{pn} \end{bmatrix} \quad (A.6) $$

Because $[b_{p1} \quad b_{p2} \quad \ldots \quad b_{pn}]$ is the $p$-th row of $B$, the rows $\{A^i\}$ of the $m \times n$ matrix $A$ are linear combinations of the rows of $B$, we have $\text{row rank } A \leq \text{row rank } B \leq \text{row rank } B$. 
r = column rank A. We have now shown that row rank A ≤ column rank A. Furthermore, if we subject \( A^t \) to the same procedure, we will arrive at row rank \( A^t \) ≤ column rank \( A \) \( \implies \) column rank \( A \leq \) row rank \( A \). This proves the theorem.

An Alternative Proof for Real Matrices

Let \( A \) be an \( m \times n \) real matrix whose row rank is \( r \); i.e., the dimension of the row space of \( A \) is \( r \). Choose a basis \( \{v_1, v_2, \ldots, v_r\} \) of the row space of \( A \). Note that \( v_i \)'s are \( 1 \times n \) row vectors, \( (v_i)^t \)'s are \( n \times 1 \) column vectors, and \( A(v_i)^t \)'s are \( m \times 1 \) column vectors. Now, consider a linear combination

\[
\sum_{i=1}^{r} c_i A(v_i)^t = A \left[ \sum_{i=1}^{r} c_i (v_i)^t \right] = A w^t = 0; \tag{A.7}
\]

where \( c_i \)'s are some scalars, and \( w := \sum_{i=1}^{r} c_i v_i \) is some vector in the row space of \( A \). Because \( A w = 0 \) and the \( i \)-th component of the \( m \times 1 \) column vector \( A w^t \) is given by \( \sum_{j=1}^{n} A_{ij} w_j \), which is the inner product between the \( i \)-th row of \( A \) and \( w \), (A.7) implies that \( w \) is orthogonal to each row vector of \( A \). Hence, \( w \) is orthogonal to all the vectors in the row space of \( A \). But, we know \( w \) belongs to the row space of \( A \).

Therefore, \( w = \sum_{i=1}^{r} c_i v_i = 0 \), which in turn implies that all \( c_i \)'s are 0 because \( v_i \)'s are linearly independent. It now follows that \( A(v_1)^t, A(v_2)^t, \ldots, A(v_r)^t \) are linearly independent. Noting that \( [A(v_i)^t]_k = \sum_{l=1}^{n} A_{kl} (v_i)^t \), we have

\[
A(v_i)^t = \begin{bmatrix}
\sum_{l=1}^{n} A_{1l} (v_i)^t \\
\sum_{l=1}^{n} A_{2l} (v_i)^t \\
\vdots \\
\sum_{l=1}^{n} A_{ml} (v_i)^t
\end{bmatrix} = \begin{bmatrix}
A_{11} v_1^t + A_{12} v_2^t + \ldots + A_{1n} v_n^t \\
A_{21} v_1^t + A_{22} v_2^t + \ldots + A_{2n} v_n^t \\
\vdots \\
A_{m1} v_1^t + A_{m2} v_2^t + \ldots + A_{mn} v_n^t
\end{bmatrix} = (v_i)_1^t A_{11} + (v_i)_2^t A_{21} + \ldots + (v_i)_n^t A_{mn}, \tag{A.8}
\]

which is a linear combination of the column vectors of \( A \). Hence, we have \( r \) linearly independent vectors \( A(v_1)^t, A(v_2)^t, \ldots, A(v_r)^t \) in the column space of \( A \). We conclude that the dimension of the column space is at least \( r \), and we have shown that column rank \( A \geq \) row rank \( A \). We next consider \( A^t \) and go through the same procedure to show that column rank \( A^t \geq \) row rank \( A^t \). But, column rank \( A^t \) =
A.3. CHANGE OF BASIS

row rank $A$ and row rank $A' = column rank A$ give us row rank $A \geq column rank A$. This concludes the proof. 

A.3 Change of Basis

Theorem A.2 (Change of Basis: Transition Matrix) [Anton, 2010, p.218] Consider two bases $B = \{u_1, u_2, \ldots, u_n\}$ and $B' = \{u'_1, u'_2, \ldots, u'_n\}$ of an $n$-dimensional vector space $\mathbb{V}^n$; where we regard $u_i$'s and $u'_i$'s as $n \times 1$ column vectors. If $v_B$ is the $n \times 1$ column matrix representation of a vector $v \in \mathbb{V}^n$ in $B$, and $v_{B'}$ the same in $B'$, then, the relation between $v_B$ and $v_{B'}$ can be described by

$$v_B = P_{B' \rightarrow B} v_{B'}, \tag{A.9}$$

where $P_{B' \rightarrow B}$, the transition matrix from $B'$ to $B$, is an $n \times n$ matrix whose columns are $u'_1, u'_2, \ldots, u'_n$. Note that with our notation $u'_i$ is the $i$-th basis vector of $B'$ expressed as a column vector in the basis $B$. Similarly, due to symmetry, with another transition matrix $P_{B \rightarrow B'}$ whose columns are $u_j$, we have

$$v_{B'} = P_{B \rightarrow B'} v_B. \tag{A.10}$$

In fact, we can show

$$P_{B' \rightarrow B} P_{B \rightarrow B'} = P_{B \rightarrow B'} P_{B' \rightarrow B} = I \tag{A.11}$$

and $P_{B \rightarrow B'}$ is the inverse of $P_{B' \rightarrow B}$.

Proof

As $B$ and $B'$ are bases, each vector in $B$ can be expressed as a unique linear combination of the vectors in $B'$, and each vector in $B'$ can be expressed as a unique linear combination of the vectors in $B$. Hence, we can find the unique coefficients $S_{ij}$ and $T_{ij}$, for $1 \leq j, i \leq n$, or differently put, unique matrices $S = [S_{ij}]$ and $T = [T_{ij}]$ such that

$$u_i = \sum_{j=1}^{n} S_{ji} u'_j \quad \text{and} \quad u'_i = \sum_{j=1}^{n} T_{ji} u_j \tag{A.12}$$

or in matrix form

$$\begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} = \begin{bmatrix} u'_1 & u'_2 & \cdots & u'_n \end{bmatrix} \begin{bmatrix} S_{11} & \cdots & S_{1n} \\ \vdots & \ddots & \vdots \\ S_{n1} & \cdots & S_{nn} \end{bmatrix} \tag{A.13}$$
and
\[
\begin{bmatrix}
\mathbf{u}'_1 & \mathbf{u}'_2 & \cdots & \mathbf{u}'_n
\end{bmatrix} =
\begin{bmatrix}
\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n
\end{bmatrix}
\begin{bmatrix}
T_{11} & \cdots & T_{1n} \\
\vdots & \ddots & \vdots \\
T_{n1} & \cdots & T_{nn}
\end{bmatrix}.
\] (A.14)

If we combine the two relations of (A.12), we get
\[
\mathbf{u}_i = \sum_{j=1}^{n} S_{ji} \mathbf{u}'_j = \sum_{j=1}^{n} S_{ji} \sum_{k=1}^{n} T_{kj} \mathbf{u}_k = \sum_{k=1}^{n} \left( \sum_{j=1}^{n} T_{kj} S_{ji} \right) \mathbf{u}_k
\] (A.15)

and
\[
\mathbf{u}'_i = \sum_{j=1}^{n} T_{ji} \mathbf{u}_j = \sum_{j=1}^{n} T_{ji} \sum_{k=1}^{n} S_{kj} \mathbf{u}'_k = \sum_{k=1}^{n} \left( \sum_{j=1}^{n} S_{kj} T_{ji} \right) \mathbf{u}'_k.
\] (A.16)

However, because \( B = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\} \) and \( B' = \{\mathbf{u}'_1, \mathbf{u}'_2, \ldots, \mathbf{u}'_n\} \) are bases, \( \mathbf{u}_i \)'s and \( \mathbf{u}'_i \)'s are linearly independent. Hence,
\[
\sum_{j=1}^{n} T_{kj} S_{ji} = \sum_{j=1}^{n} S_{kj} T_{ji} = \delta_{ki}. \] (A.17)

Recall that
\[
\delta_{ki} = \begin{cases} 
1 & i = k \\
0 & i \neq k
\end{cases} \] (A.18)

If we form the matrices \( S = [S_{ij}] \) and \( T = [T_{ij}] \), \( \sum_{j=1}^{n} S_{kj} T_{ji} \) is the \( k \)-th row and the \( i \)-th column entry of the matrix product \( ST \), and \( \sum_{j=1}^{n} T_{kj} S_{ji} \) is the \( k \)-th row and the \( i \)-th column entry of the matrix product \( TS \). Therefore,
\[
ST = TS = I, \] (A.19)

and \( T \) is the inverse of \( S \). Getting back to (A.12),
\[
\mathbf{u}_i = \sum_{j=1}^{n} S_{ji} \mathbf{u}'_j = S_{1i} \mathbf{u}'_1 + S_{2i} \mathbf{u}'_2 + \cdots + S_{ni} \mathbf{u}'_n = \begin{bmatrix}
S_{1i} \\
S_{2i} \\
\vdots \\
S_{ni}
\end{bmatrix}_{B'}
\] (A.20)
and

\[ \mathbf{u}'_i = \sum_{j=1}^{n} T_{ji} \mathbf{u}_j = T_{1i} \mathbf{u}_1 + T_{2i} \mathbf{u}_2 + \ldots + T_{ni} \mathbf{u}_n = \begin{bmatrix} T_{1i} \\ T_{2i} \\ \vdots \\ T_{ni} \end{bmatrix}_B; \quad (A.21) \]

where \([\cdot]_B\) and \([\cdot]_{B'}\) mean they are with respect to \(B\) and \(B'\), respectively. This proves that the \(j\)-th column of \([S_{ij}]\) is \(\mathbf{u}_{jB'}\), and that the \(j\)-th column of \([T_{ij}]\) is \(\mathbf{u}'_{jB}\). It remains to show that

\[ [T_{ij}] = P_{B' \rightarrow B} \quad \text{and} \quad [S_{ij}] = P_{B \rightarrow B'} \quad (A.22) \]

as described in the theorem. Take any vector \(\mathbf{v} \in \mathbb{V}^n\) and express it as linear combinations of \(\mathbf{u}_i\)’s and \(\mathbf{u}'_i\)’s, respectively.

\[ \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}_B \quad \text{and} \quad \mathbf{v}' = \begin{bmatrix} v'_1 \\ v'_2 \\ \vdots \\ v'_n \end{bmatrix}_B' \quad (A.23) \]

We get

\[ \mathbf{v} = \sum_{i=1}^{n} v'_i \mathbf{u}'_i = \sum_{i=1}^{n} v'_i \sum_{j=1}^{n} T_{ji} \mathbf{u}_j = \sum_{i=1}^{n} \left( \sum_{j=1}^{n} T_{ji} v'_i \right) \mathbf{u}_j \]

\[ = \left( \sum_{i=1}^{n} T_{1i} v'_i \right) \mathbf{u}_1 + \left( \sum_{i=1}^{n} T_{2i} v'_i \right) \mathbf{u}_2 + \ldots + \left( \sum_{i=1}^{n} T_{ni} v'_i \right) \mathbf{u}_n = \begin{bmatrix} \sum_{i=1}^{n} T_{1i} v'_i \\ \sum_{i=1}^{n} T_{2i} v'_i \\ \vdots \\ \sum_{i=1}^{n} T_{ni} v'_i \end{bmatrix}_B \]

\[ \implies [\mathbf{v}]_B = [T_{ij}] [\mathbf{v}]_{B'} \quad (A.24) \]

and

\[ \mathbf{v} = \sum_{i=1}^{n} v_i \mathbf{u}_i = \sum_{i=1}^{n} v_i \sum_{j=1}^{n} S_{ji} \mathbf{u}'_j = \sum_{i=1}^{n} \left( \sum_{j=1}^{n} S_{ji} v_i \right) \mathbf{u}'_j \]

\[ = \left( \sum_{i=1}^{n} S_{1i} v_i \right) \mathbf{u}'_1 + \left( \sum_{i=1}^{n} S_{2i} v_i \right) \mathbf{u}'_2 + \ldots + \left( \sum_{i=1}^{n} S_{ni} v_i \right) \mathbf{u}'_n = \begin{bmatrix} \sum_{i=1}^{n} S_{1i} v_i \\ \sum_{i=1}^{n} S_{2i} v_i \\ \vdots \\ \sum_{i=1}^{n} S_{ni} v_i \end{bmatrix}_{B'} \]

\[ \quad \text{as described in the theorem.} \]
\[ [v]_{B'} = [S_{ij}] [v]_B \]  \hspace{1cm} (A.25)

This proves (A.22), and we have now proved the theorem by an explicit construction of the transition matrices \( P_{B' \rightarrow B} \) and \( P_{B \rightarrow B'} \).

**Theorem A.3 (Change of Basis and Linear Transformations)**  Let \( \Omega \) be a linear operator operating on a vector space \( V^n \). In other words, \( \Omega \) is a linear map from \( V^n \) into \( V^n \). Consider two bases \( B = \{ u_1, u_2, \ldots, u_n \} \) and \( B' = \{ u'_1, u'_2, \ldots, u'_n \} \) of \( V^n \) and the corresponding matrix representations of \( \Omega \) in those bases, denoted by \( [\Omega]_{ij} \) and \( [\Omega']_{ij} \) respectively. Then,

\[ \Omega'_{ij} = \sum_{k,l=1}^{n} T^{-1}_{ik} \Omega_{kl} T_{lj} \text{ or } [\Omega']_{ij} = [T^{-1}_{ij}] [\Omega]_{ij} [T_{ij}] \text{ or } \Omega' = T^{-1} \Omega T; \]  \hspace{1cm} (A.26)

where \( T = P_{B' \rightarrow B} \).

**Proof**

From (2.91) on p.43,

\[ \Omega u_i = \sum_{j=1}^{n} \Omega_{ji} u_j \]  \hspace{1cm} (A.27)

and

\[ \Omega' u'_i = \sum_{j=1}^{n} \Omega'_{ji} u'_j. \]  \hspace{1cm} (A.28)

Next, from (A.12), we have

\[ u'_i = \sum_{j=1}^{n} T_{ji} u_j. \]  \hspace{1cm} (A.29)

Substituting (A.29) into the right-hand side of (A.28),

\[ \sum_{j=1}^{n} \Omega'_{ji} u'_j = \sum_{j=1}^{n} \Omega'_{ji} \left( \sum_{k=1}^{n} T_{kj} u_k \right) = \sum_{k=1}^{n} \left( \sum_{j=1}^{n} T_{kj} \Omega'_{ji} \right) u_k. \]  \hspace{1cm} (A.30)

Now, substituting (A.29) and (A.27) successively into the left-hand side of (A.28),

\[ \Omega u'_i = \Omega \left( \sum_{j=1}^{n} T_{ji} u_j \right) = \sum_{j=1}^{n} T_{ji} \Omega u_j = \sum_{j=1}^{n} T_{ji} \left( \sum_{k=1}^{n} \Omega_{kj} u_k \right) = \sum_{k=1}^{n} \left( \sum_{j=1}^{n} \Omega_{kj} T_{ji} \right) u_k. \]  \hspace{1cm} (A.31)
Because \( u_k \)'s are linearly independent, (A.30) and (A.31) imply that
\[
\sum_{j=1}^{n} T_{kj} \Omega'_{ji} = \sum_{j=1}^{n} \Omega_{kj} T_{ji} \quad (A.32)
\]
for all \( 1 \leq i, k \leq n \). This means that the \((k, i)\)-entry of the matrix product \( T \Omega' \) equals the \((k, i)\)-entry of the matrix product \( \Omega T \) for all \( 1 \leq i, k \leq n \). Therefore,
\[
T \Omega' = \Omega T \implies \Omega' = T^{-1} \Omega T; \quad (A.33)
\]
where we know \( T = P_{B' \rightarrow B} \) from Theorem A.2.

**Theorem A.4 (Basis Independence of Characteristic Polynomials)** The characteristic polynomial of a linear transformation \( \Omega \) is well defined. That is, it is basis-independent and does not depend on the choice of a basis.

**Proof**
Consider two bases \( B = \{u_i\} \) and \( B' = \{v'_i\} \), and denote the matrix representations of a linear transformation \( \Omega \) with respect to these bases by \( \Omega \) and \( \Omega' \). If \( T := P_{B' \rightarrow B} \) so that \( T v'_k = u_k \), then, from Theorem A.3 we have \( \Omega' = T^{-1} \Omega T \). This gives us
\[
det(\Omega' - \lambda I) = det(T^{-1} \Omega T - \lambda I) = det(T^{-1} \Omega T - T^{-1}(\lambda I)T) \\
= det(T^{-1}(\Omega - \lambda I)T) = det(T^{-1})det(\Omega - \lambda I)detT = \frac{1}{detT}det(\Omega - \lambda I)detT \\
= det(\Omega - \lambda I), \quad (A.34)
\]
which is the desired result.

**Theorem A.5** If there is a linearly independent set \( S \), it can be extended to a basis of the vector space \( V \); that is, if \( u_1, u_2, \ldots, u_m \) (\( m < n \)) are linearly independent, we can find \( u_{m+1} \ldots, u_n \) such that \( \{u_1, u_2, \ldots, u_{m+1} \ldots u_n\} \) is a basis for \( V \).

**Proof by Zorn’s Lemma**

### A.4 Multiplicity of an Eigenvalue

If \( M \) is an \( n \times n \) matrix, then, \( det(A - \lambda I) \) is an \( n \)-th degree polynomial, called the characteristic polynomial, and the eigenvalues are the roots of the characteristic polynomial. If we have a factor of the form \((\lambda - c)^l\) in the factorization of \( det(A - \lambda I) \), we say the eigenvalue \( c \) is of multiplicity \( l \).
**Fact A.2 (Multiplicity and Linear Independence)** Let \( \{\lambda_1, \lambda_2, \ldots, \lambda_k\} \) be the list of all eigenvalues, including repeated roots of the characteristic polynomial, i.e. \( \lambda_0 \) of multiplicity \( l \) appears in the list \( \{\lambda_1, \lambda_2, \ldots, \lambda_k\} \) \( l \)-times.

1. An eigenvalue/a root that occurs only once in the list \( \{\lambda_1, \lambda_2, \ldots, \lambda_k\} \) is called **simple**.
2. If \( \lambda_0 \) occurs \( l \) times, \( \lambda_0 \) is of **multiplicity** \( l \).
3. Eigenvectors associated with simple eigenvalues are linearly independent.
4. An eigenvalue of multiplicity \( l \) may have 1, 2, \ldots, or \( l \) linearly independent eigenvectors.

**A.5 Diagonalizability and Simultaneous Diagonalization**

In the following, we will denote the set of all \( n \times n \) matrices by \( M_n \).

**Definition A.2 (Similarity)** A matrix \( B \in M_n \) is similar to another matrix \( A \in M_n \) if there exists an invertible matrix \( S \in M_n \) such that

\[
B = S^{-1}AS. \quad (A.35)
\]

**Definition A.3 (Diagonalizability)** A matrix \( A \in M_n \) is diagonalizable if it is similar to a diagonal matrix.

**Theorem A.6** A matrix \( A \in M_n \) is diagonalizable if and only if there exists a set of \( n \) linearly independent eigenvectors of \( A \).

**Proof**

(\( \Rightarrow \))

Let \( S \) be an invertible matrix such that \( S^{-1}AS = D \), where \( D \) is a diagonal matrix. We have \( S^{-1}AS = D \Rightarrow S(S^{-1}AS) = SD \Rightarrow AS = SD \). Now, denote the \( j \)-th column of \( S \) by \( s_j \), which can also be regarded as an \( n \times 1 \) column vector such that

\[
s_j = [s_{ij}] \quad \text{with} \quad 1 \leq i \leq n, \quad (A.36)
\]
where \( s_{ij} \) is the \( i \)-th row and \( j \)-th column entry of \( S \). With this notation,

\[
A s_j = 
\begin{bmatrix}
\sum_i A_{1i} s_{ij} \\
\vdots \\
\sum_i A_{ki} s_{ij} \\
\vdots \\
\sum_i A_{ni} s_{ij}
\end{bmatrix}.
\tag{A.37}
\]

So, \( A s_j \) is the \( j \)-th column of \( AS \), and we have

\[
AS = [A s_1 \ A s_2 \ldots \ A s_n].
\tag{A.38}
\]

Next, let \( D = [d_{ii}] \), where \( d_{ii} \) is the \( i \)-th diagonal entry of \( D \). Then, the \( i \)-th row and the \( j \)-th column entry of \( SD \) denoted by \((SD)_{ij}\) is given by

\[
(SD)_{ij} = \sum_k s_{ik} d_{kj} = s_{ij} d_{jj} = d_{jj} s_{ij}.
\tag{A.39}
\]

Hence, the \( j \)-th column of \( SD \) is given by

\[
\begin{bmatrix}
d_{jj} s_{1j} \\
\vdots \\
d_{jj} s_{kj} \\
d_{jj} s_{nj}
\end{bmatrix} = d_{jj} s_j,
\tag{A.40}
\]

which in turn indicates

\[
SD = [d_{11} s_1 \ d_{22} s_2 \ldots \ d_{nn} s_n].
\tag{A.41}
\]

Putting these all together,

\[
AS = SD \implies [A s_1 \ A s_2 \ldots \ A s_n] = [d_{11} s_1 \ d_{22} s_2 \ldots \ d_{nn} s_n] \implies A s_j = d_{jj} s_j
\]

for \( 1 \leq j \leq n \).

\[
\tag{A.42}
\]

Therefore, each \( s_j \) is an eigenvector of \( A \) with corresponding eigenvalue \( d_{jj} \). Because \( S \) is invertible, \( \text{rank } S = n \), and \( s_j \)'s are linearly independent.

\((\implies)\)

Conversely, suppose there is a linearly independent set \( \{s_1, s_2, \ldots, s_n\} \) of eigenvectors of \( A \) with associated eigenvalues \( s_1, s_2, \ldots, s_n \), and form \( S = [s_1 \ s_2 \ldots \ s_n] \). As before,
\[ AS = [A s_1 \ A s_2 \ \ldots \ A s_n] = [\lambda_1 s_1 \ \lambda_2 s_2 \ \ldots \ \lambda_n s_n]. \]

If we let \( D \) be the diagonal matrix whose \( k \)-th diagonal entry is \( \lambda_k \). Then, \( SD = [\lambda_1 s_1 \ \lambda_2 s_2 \ \ldots \ \lambda_n s_n] = AS \). Because \( \{s_j\} \) is a linearly independent set, \( S \) is of full rank, and \( S^{-1} \) exists. We now have \( AS = SD \implies S^{-1}AS = S^{-1}SD = D \), and \( A \) is diagonalizable. \( \blacksquare \)

**Definition A.4 (Simultaneous Diagonalization)** Two matrices \( A, B \in M_n \) are said to be simultaneously diagonalizable if there exists an invertible matrix \( S \) such that both \( S^{-1}AS \) and \( S^{-1}BS \) are diagonal matrices.

**Theorem A.7 (Simultaneous Diagonalization and Commutativity)** Let \( A, B \in M_n \) be diagonalizable. Then \( AB = BA \) if and only if \( A \) and \( B \) are simultaneously diagonalizable.

In order to prove Theorem A.7, we need the following two lemmas.

**Lemma A.1** Two matrices \( A \in M_n \) and \( B \in M_m \) are diagonalizable if and only if the composite block matrix

\[ C = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \in M_{n+m} \quad (A.43) \]

is diagonalizable.

**Proof**

\( \implies \)

Suppose \( A \in M_n \) and \( B \in M_m \) are diagonalizable. Then, we can find two invertible matrices \( S_A \in M_n \) and \( S_B \in M_m \) such that \( S_A^{-1}AS_A \) and \( S_B^{-1}BS_B \) are diagonal matrices. Define \( S \in M_{n+m} \) by

\[ S := \begin{bmatrix} S_A & 0 \\ 0 & S_B \end{bmatrix}. \quad (A.44) \]

Then,

\[ S^{-1} = \begin{bmatrix} S_A^{-1} & 0 \\ 0 & S_B^{-1} \end{bmatrix}, \quad (A.45) \]

and

\[ S^{-1}CS = S^{-1} \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} S = \begin{bmatrix} S_A^{-1} & 0 \\ 0 & S_B^{-1} \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \begin{bmatrix} S_A & 0 \\ 0 & S_B \end{bmatrix} \]
\[ S^{-1} A S = \begin{bmatrix} \mathbf{0} & 0 \\ 0 & S^{-1}_B B S_B \end{bmatrix} \] (A.46)

is a diagonal matrix, and \( C \) is diagonalizable.

\[ \text{(⇐)} \]

Conversely, assume that \( C \) is diagonalizable. Then, there exists an invertible matrix \( S \in M_{n+m} \) such that \( D := S^{-1} C S \) is diagonal. Now, denote the \( j \)-th column of \( S \) by \( s_j \in \mathbb{C}^{n+m} \). We can write \( S = [s_1 \ s_2 \ \ldots \ s_{n+m}] \), and because \( CS = SD \), with the usual representation \( D = [d_i] \), where \( d_i \) is the \( i \)-th diagonal entry, we have

\[ [Cs_1 \ Cs_2 \ \ldots \ Cs_{n+m}] = [d_1 s_1 \ d_2 s_2 \ \ldots \ d_{n+m} s_{n+m}] \] (A.47)

as in (A.42). This shows that \( Cs_k = d_k s_k \), and \( s_k \)'s are eigenvectors of \( C \) with the associated eigenvalues \( d_k \). Now, we write each \( s_k \) in composite form

\[ s_k = \begin{bmatrix} x_k \\ y_k \end{bmatrix}, \] (A.48)

where \( x_k \in \mathbb{C}^n \) and \( y_k \in \mathbb{C}^m \). Due to the block structure of \( S \), \( Cs_k = d_k s_k \) requires \( Ax_k = d_k x_k \) and \( By_k = d_k y_k \). Denoting \([x_1 \ x_2 \ \ldots \ x_{n+m}]\) by \( X \) and \([y_1 \ y_2 \ \ldots \ y_{n+m}]\) by \( Y \), we have

\[ S = \begin{bmatrix} X \\ Y \end{bmatrix}. \] (A.49)

Because \( X \) is an \( n \times (n + m) \) matrix and \( Y \) is an \( m \times (n + m) \) matrix,

\[ \text{rank } X \leq n \quad \text{and} \quad \text{rank } Y \leq m. \] (A.50)

On the other hand,

\[ \text{rank } S = n + m \] (A.51)

as \( S \) is invertible. We have the following implications.

\[ n + m = \text{rank } S \leq \text{rank } X + \text{rank } Y \leq n + m \]

\[ \implies \text{rank } X + \text{rank } Y = n + m \implies \text{rank } X = n, \ \text{rank } Y = m \] (A.52)

Therefore, \( X \) contains \( n \) linearly independent columns and \( Y \) contains \( m \) linearly independent columns. Recalling that \( A \in M_n \) operates on vectors in \( \mathbb{C}^n \) and the columns of \( X \) are eigenvectors of \( A \), we can now see that \( A \) is diagonalizable by Theorem A.6. Likewise, \( B \) is also diagonalizable as \( B \in M_m \) operates on vectors in \( \mathbb{C}^m \) and the columns of \( Y \) are eigenvectors of \( B \).
Lemma A.2 For a diagonal matrix

\[
D = \begin{bmatrix}
    d_1 & 0 & \ldots & 0 & 0 \\
    0 & d_2 & \ldots & 0 & 0 \\
    \vdots & \vdots & \ddots & \vdots & \vdots \\
    0 & 0 & \ldots & d_{n-1} & 0 \\
    0 & 0 & \ldots & 0 & d_n \\
\end{bmatrix} \in M_n, \tag{A.53}
\]

there exists an invertible matrix \( P \in M_n \) such that \( P^{-1}DP \) switches the positions of \( d_i \) and \( d_j \) for any \( 1 \leq i, j \leq n \). In fact, such \( P \) can be obtained by switching the \( i \)-th row and the \( j \)-th row of the identity matrix \( I \in M_n \).

Before embarking on a general proof of this lemma. Let us see how \( P \) works with a \( 4 \times 4 \) matrix

\[
D = \begin{bmatrix}
    d_1 & 0 & 0 & 0 \\
    0 & d_2 & 0 & 0 \\
    0 & 0 & d_3 & 0 \\
    0 & 0 & 0 & d_4 \\
\end{bmatrix}. \tag{A.54}
\]

We will try to switch \( d_1 \) and \( d_3 \) to obtain

\[
D_{d_1 \leftrightarrow d_3} = \begin{bmatrix}
    d_3 & 0 & 0 & 0 \\
    0 & d_2 & 0 & 0 \\
    0 & 0 & d_1 & 0 \\
    0 & 0 & 0 & d_4 \\
\end{bmatrix}. \tag{A.55}
\]

According to our program, \( P \) is obtained by switching the first and the third rows of \( I \in M_4 \). Hence,

\[
P = \begin{bmatrix}
    0 & 0 & 1 & 0 \\
    0 & 1 & 0 & 0 \\
    1 & 0 & 0 & 0 \\
    0 & 0 & 0 & 1 \\
\end{bmatrix} \quad \text{and} \quad P^{-1} = \begin{bmatrix}
    0 & 0 & 1 & 0 \\
    0 & 1 & 0 & 0 \\
    1 & 0 & 0 & 0 \\
    0 & 0 & 0 & 1 \\
\end{bmatrix}. \tag{A.56}
\]

With this \( P \), we have

\[
P^{-1}DP = \begin{bmatrix}
    0 & 0 & 1 & 0 \\
    0 & 1 & 0 & 0 \\
    1 & 0 & 0 & 0 \\
    0 & 0 & 1 & 0 \\
\end{bmatrix} \begin{bmatrix}
    d_1 & 0 & 0 & 0 \\
    0 & d_2 & 0 & 0 \\
    0 & 0 & d_3 & 0 \\
    0 & 0 & 0 & d_4 \\
\end{bmatrix} \begin{bmatrix}
    0 & 0 & 1 & 0 \\
    0 & 1 & 0 & 0 \\
    1 & 0 & 0 & 0 \\
    0 & 0 & 0 & 1 \\
\end{bmatrix}
\]
\[ \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & d_1 & 0 \\ 0 & d_2 & 0 & 0 \\ d_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & d_4 \end{bmatrix} = \begin{bmatrix} d_3 & 0 & 0 & 0 \\ 0 & d_2 & 0 & 0 \\ 0 & 0 & d_1 & 0 \\ 0 & 0 & 0 & d_4 \end{bmatrix} = D_{d_1 \leftrightarrow d_3} \quad (A.57) \]

**Proof of Lemma A.2**

Our proof is by way of an explicit computation. Let us see how the matrix \( P \) given by (A.58) switches \( d_i \) and \( d_j \). In (A.58), ",", ";", and a blank space signify 0, while ";;" stands for 1.

\[
\begin{pmatrix}
i & j
\end{pmatrix}
\begin{bmatrix}
i \\
0 & 1 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 0 & 0 \\
0 & 1 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 0 & 0 \\
0 & 0 & \ldots & \ldots & 0 & \ldots & \ldots & \ldots & 1 & \ldots & \ldots & 0 & 0 \\
0 & 0 & \ldots & \ldots & 0 & 1 & \ldots & \ldots & \ldots & 0 & \ldots & \ldots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 1 & \ldots & \ldots & 0 & 0 \\
0 & 0 & \ldots & \ldots & 1 & \ldots & \ldots & \ldots & 0 & \ldots & \ldots & 0 & 0 \\
0 & 0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 1 & \ldots & \ldots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 1 & \ldots & \ldots & 0 & 0 \\
0 & 0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 0 & 1 \\
\end{bmatrix}
\quad (A.58)
\]

Note first that \( P \) has the following block diagonal structure; where \( I_{i-1} \) is the \((i - 1) \times (i - 1)\) identity matrix, \( I_{n-j} \) is the \((n - j) \times (n - j)\) identity matrix, ",;" stands
for 0, and "..." denotes 1.

\[
P_{ij} = \begin{bmatrix}
I_{i-1} & \cdots & \cdots & \cdots & \cdots & \cdots \\
\cdots & 0 & \cdots & \cdots & \cdots & \cdots \\
\cdots & 0 & 1 & \cdots & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & 1 & 0 \\
\cdots & \cdots & \cdots & 0 & \cdots & \cdots \\
\cdots & \cdots & \cdots & 0 & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\end{bmatrix} = \begin{bmatrix}
I_{i-1} & \cdots & \cdots \\
\cdots & \cdots & \cdots \\
\cdots & \cdots & I_{n-j} \\
\end{bmatrix}; \quad (A.59)
\]

where \( Q \in M_{j-i+1} \) is defined by

\[
Q = \begin{bmatrix}
0 & \cdots & \cdots & \cdots & \cdots & 1 \\
0 & 1 & \cdots & \cdots & \cdots & 0 \\
\vdots & \cdots & \cdots & \cdots & \cdots & \vdots \\
\vdots & \cdots & \cdots & \cdots & \cdots & \vdots \\
0 & \cdots & \cdots & 1 & 0 \\
1 & \cdots & \cdots & \cdots & 0 \\
\end{bmatrix}. \quad (A.60)
\]

It is easy to see \( Q^2 = I \) or \( Q^{-1} = Q \). Due to the block diagonal structure of \( P \), it suffices to consider the action of \( Q \) on the following submatrix \( D \) of \( D \).

\[
D := \begin{bmatrix}
d_i & \cdots & \cdots & \cdots & 0 \\
0 & d_{i+1} & \cdots & \cdots & 0 \\
\vdots & \cdots & \cdots & \cdots & \vdots \\
\vdots & \cdots & \cdots & \cdots & \vdots \\
0 & \cdots & \cdots & d_{j-1} & 0 \\
0 & \cdots & \cdots & d_j & 0 \\
\end{bmatrix} \quad (A.61)
\]

We have

\[
Q^{-1}DQ = \begin{bmatrix}
0 & \cdots & \cdots & \cdots & 1 \\
0 & 1 & \cdots & \cdots & 0 \\
\vdots & \cdots & \cdots & \cdots & \vdots \\
\vdots & \cdots & \cdots & \cdots & \vdots \\
0 & \cdots & \cdots & 1 & 0 \\
1 & \cdots & \cdots & \cdots & 0 \\
\end{bmatrix}
\]
There is an obvious corollary to Lemma A.2.

**Corollary A.1** Given a diagonal matrix \( D \in M_n \) and \( D'_{\text{permutation}} \in M_n \), where \( D'_{\text{permutation}} \) is another diagonal matrix whose diagonal entries are a permuted version of the original \( D \), there exists an invertible matrix \( Q \in M_n \) such that \( D'_{\text{permutation}} = Q^{-1}DQ \). In other words, one can permute the diagonal entries by some similarity transformation \( Q^{-1}DQ \).

**Proof of Theorem A.7**
(simultaneous diagonalizability \( \implies \) commutativity)
As \( S^{-1}AS \) and \( S^{-1}BS \) are both diagonal, and diagonal matrices commute with each other, we get
\[
AB = S \left( S^{-1}AS \right) \left( S^{-1}BS \right) S^{-1} = S \left( S^{-1}BS \right) \left( S^{-1}AS \right) S^{-1} = BA. \tag{A.63}
\]
(commutativity $\implies$ simultaneous diagonalizability)

As $A$ is diagonalizable, there is an invertible matrix $S \in M_n$ such that $D = S^{-1}AS$ is diagonal. By Corollary A.1, we can permute the diagonal entries of $D$ by multiplying $S$ by an appropriate invertible matrix. Hence, we may assume that

$$D = \begin{bmatrix}
\lambda_1 I_{m_1} & 0 & \cdots & 0 \\
0 & \lambda_2 I_{m_2} & & 0 \\
\vdots & \ddots & \ddots & \\
0 & \cdots & 0 & \lambda_r I_{m_r}
\end{bmatrix};$$

(A.64)

where $\lambda_i$’s are distinct, $m_j$ is the multiplicity of $\lambda_j$, and $I_{m_j}$ is the $m_j \times m_j$ identity matrix. As $AB = BA$,

$$\left(S^{-1}AS\right) \left(S^{-1}BS\right) = S^{-1}ABS = S^{-1}BAS = \left(S^{-1}BS\right) \left(S^{-1}AS\right).$$

(A.65)

Denoting $(S^{-1}BS)$ by $C$, we have $DC = CD$. Now, let $c_{ij}$, $d_{ij}$, $(cd)_{ij}$, and $(dc)_{ij}$ signify the $i$-th row and the $j$-th column entries of $C$, $D$, $CD$, and $DC$ respectively. As $D$ is diagonal, we can write $d_{ij} = \delta_{ij}d_{jj}$. So,

$$(cd)_{ij} = \sum_k c_{ik}d_{kj} = \sum_k c_{ik}(\delta_{kj}d_{jj}) = c_{ij}d_{jj}$$

and

$$(dc)_{ij} = \sum_k d_{ik}c_{kj} = \sum_k (\delta_{ik}d_{ii})c_{kj} = d_{ii}c_{ij}.$$  

(A.66) (A.67)

Because $CD = DC$ means $(cd)_{ij} = (dc)_{ij}$ for all $1 \leq i, j \leq n$, we have

$$(cd)_{ij} = (dc)_{ij} \implies c_{ij}d_{jj} = d_{ii}c_{ij} \implies (d_{ii} - d_{jj})c_{ij} = 0 \text{ for } 1 \leq i, j \leq n. \quad (A.68)$$

Therefore, $c_{ij} = 0$ unless $d_{ii} = d_{jj}$. But, $d_{ii} = d_{jj}$ only if $d_{ii}$ and $d_{jj}$ belong to the same diagonal block of (A.64). This implies that $c_{ij} = 0$ unless $d_{ii}$ and $d_{jj}$ belong to the same block of (A.64). Hence, the matrix $C$ is block diagonal whose $i$-th block is an $m_i \times m_i$ submatrix. In other words, $C$ has the same block structure as $D$, namely

$$C = \begin{bmatrix}
C_1 & 0 & \cdots & 0 \\
0 & C_2 & & 0 \\
\vdots & \ddots & \ddots & \\
0 & \cdots & 0 & C_r
\end{bmatrix}; \text{ where } C_l \in M_{m_l} \text{ for each } l. \quad (A.69)$$

As $B$ is also diagonalizable, there is an invertible matrix $R$ such that $R^{-1}BR$ is diagonal. So,

$$R^{-1}SCS^{-1}R = R^{-1}SS^{-1}BSS^{-1}R = R^{-1}BR,$$  

(A.70)
and \( C \) is diagonalizable. From Lemma A.1, all the blocks of \( C \) are diagonalizable; that is, for each \( j \), there exists an invertible matrix \( T_j \in M_{m_j} \) such that \( T_j^{-1}C_jT_j \) is diagonal. Next, define a block diagonal matrix \( T \in M_n \) by

\[
T = \begin{bmatrix}
T_1 & 0 & \ldots & 0 \\
0 & T_2 & 0 \\
\vdots & \ddots & \ddots \\
0 & \ldots & 0 & T_r
\end{bmatrix}.
\]  
(A.71)

Then, in a straightforward manner,

\[
T^{-1} = \begin{bmatrix}
T_1^{-1} & 0 & \ldots & 0 \\
0 & T_2^{-1} & 0 \\
\vdots & \ddots & \ddots \\
0 & \ldots & 0 & T_r^{-1}
\end{bmatrix}.
\]  
(A.72)

We now have

\[
(ST)^{-1}B(ST) = T^{-1}(S^{-1}BS)T = T^{-1}CT
\]

\[
= \begin{bmatrix}
T_1^{-1}C_1T_1 & 0 & \ldots & 0 \\
0 & T_2^{-1}C_2T_2 & 0 \\
\vdots & \ddots & \ddots \\
0 & \ldots & 0 & T_r^{-1}C_rT_r
\end{bmatrix}
\]  
(A.73)

and

\[
(ST)^{-1}A(ST) = T^{-1}(S^{-1}AS)T = T^{-1}DT
\]

\[
= \begin{bmatrix}
T_1^{-1}\lambda_1I_{m_1}T_1 & 0 & \ldots & 0 \\
0 & T_2^{-1}\lambda_2I_{m_2}T_2 & 0 \\
\vdots & \ddots & \ddots \\
0 & \ldots & 0 & T_r^{-1}\lambda_rI_{m_r}T_r
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\lambda_1I_{m_1} & 0 & \ldots & 0 \\
0 & \lambda_2I_{m_2} & 0 \\
\vdots & \ddots & \ddots \\
0 & \ldots & 0 & \lambda_rI_{m_r}
\end{bmatrix} = D.
\]  
(A.74)

As both (A.73) and (A.74) are diagonal, \( A \) and \( B \) are simultaneously diagonalizable by \( ST \in M_n \).
A.6 Some Useful Definitions, Facts, and Theorems

Definition A.5 (The Characteristic Polynomial)  

The characteristic polynomial of a square matrix $M$ is the function

$$f(\lambda) = \det(\lambda I - M).$$ (A.75)

Fact A.3  

If a square matrix $M$ has eigenvalues $\lambda_i$ ($i = 1, \ldots, n$), we can write the characteristic polynomial $f(\lambda)$ as the following product.

$$f(\lambda) = \det(\lambda I - M) = \prod_{i} (\lambda - \lambda_i)$$ (A.76)

Theorem A.8 (Trace, Determinant, and Eigenvalues)  

If $M$ is an $n \times n$ square matrix with eigenvalues $\lambda_i$ ($i = 1, 2, \ldots, n$), the following equalities hold.

$$\text{trace } M = \sum_{i=1}^{n} \lambda_i$$ (A.77)

$$\det M = \prod_{i=1}^{n} \lambda_i$$ (A.78)

Proof

We can expand the product (A.76) as follows.

$$f(\lambda) = \prod_{i} (\lambda - \lambda_i) = \lambda^n - \lambda^{n-1} \sum_{i=1}^{n} \lambda_i + (-1)^n \prod_{i=1}^{n} \lambda_i$$ (A.79)
Appendix B

Holomorphic Functional Calculus

Consider a function $f: \mathbb{C} \to \mathbb{C}$ and an operator $\Omega$. If $f$ can be expanded in a Taylor series about $z = 0^1$ such that

$$f(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} z^n,$$

we can define a map $f: \Omega \mapsto f(\Omega)$ by

$$f(\Omega) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} \Omega^n.$$  

(B.2)

One way to understand this formalism is to consider a matrix representation of $\Omega$. Then, the sum is also a square matrix and convergence of the series is nothing but an entry-by-entry convergence.

Fact B.1

Fact B.2 (Basis Independence of the Characteristic Polynomial) The characteristic polynomial is basis independent. Hence, the characteristic equation, the eigenvalues, and their multiplicities are basis-independent.

Proof

1The general form of Taylor series about point $z_0$ is

$$f(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(z_0)}{n!} (z - z_0)^n.$$  

287
Consider two bases $\mathcal{B} = \{b_1, b_2, \ldots, b_n\}$ and $\mathcal{D} = \{d_1, d_2, \ldots, d_n\}$, and let $M = [M_{kl}]$ be the matrix of basis change, such that

$$b_i = \sum_{j=1}^{n} M_{ij} d_j$$  \hspace{1cm} (B.3)
Appendix C

Traveling Waves and Standing Waves

When a wave is not confined to a finite portion of space, we observe a traveling wave. On the other hand, it is possible to have a wave confined to a given space producing a regular pattern, and we call this a standing wave. Traveling waves transport energy from one area of space to another, whereas standing waves can not transport energy.

In this section again, we will straddle over the line between classical mechanics and quantum mechanics.

C.1 The Wave Equation

The classical wave equation for a plane wave $y(x,t)$ traveling on a string in the $x$-direction is

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 y}{\partial t^2}. \quad \text{(C.1)}$$

In (C.1), $v$ is the propagation speed (the phase velocity\(^1\)) of the wave satisfying $\nu \lambda = v$, where $\nu$ is the frequency and $\lambda$ is the wavelength.

\(^1\) There are phase velocity, group velocity, and signal velocity. For example, you can find an extensive discussion at http://www.mathpages.com/home/kmath210/kmath210.htm.
C.1.1 Traveling Wave Solutions

For an ideal string extending from $x = -\infty$ to $x = +\infty$, two useful solutions are

$$y(x, t) = A \sin \frac{2\pi}{\lambda} (x \pm vt) \quad \text{or} \quad A \sin k(x \pm vt)$$

or $$A \sin (kx \pm \omega t) \quad \text{or} \quad A \sin 2\pi \left( \frac{x}{\lambda} \pm \frac{t}{T} \right);$$

where $A$ is an arbitrary amplitude, $k = \frac{2\pi}{\lambda}$ is the wave number, $\omega = 2\pi\nu$ is the angular frequency, and $T$ is the period.

In (C.2), $\sin k(x - vt)$ represents a wave traveling to the right, and $\sin k(x + vt)$ is a wave moving to the left. In order to see this, just consider why the propagation speed $v$ is known as the "phase" velocity. The angle $k(x \pm vt)$ is the phase which specifies each point of the traveling wave. Hence, we should follow the points with the same phase as the time $t$ progresses in order to trace the movement of the wave.

Consider a point on the traveling wave located at position $x_0$ at time $t_0$ denoted by $(x_0, t_0)$. The phase that identifies this point is $k(x_0 \pm vt_0)$. Now, consider a later time $t > t_0$. Where is the point? Let the point be at $x$. Because it is the same point, the phase indicated by $(x_0, t_0)$ should equal the phase for $(x, t)$. Therefore, we have

$$x_0 \pm vt_0 = x \pm vt \implies x = x_0 \pm vt_0 \mp vt \implies x = x_0 \mp v(t - t_0) \quad \text{(C.3)}$$

or

$$\begin{cases} 
  x_0 - vt_0 = x - vt \implies x = x_0 + v(t - t_0) \\
  x_0 + vt_0 = x + vt \implies x = x_0 - v(t - t_0)
\end{cases} \quad \text{and} \quad \text{(C.4)}$$

Now, $x_0 + v(t - t_0)$ means the point has moved to the right by $v(t - t_0)$, and $x_0 - v(t - t_0)$ indicates that the point is now displaced by $v(t - t_0)$ to the left. Therefore, the wave $y(x, t) = A \sin(kx - \omega t)$ travels to the right, and the wave $y(x, t) = A \sin(kx + \omega t)$ travels to the left as claimed above.

C.1.1.1 Real and Complex Expressions for the Traveling Wave

In Section C.1.1, we discussed representative solutions $y(x, t) = A \sin(kx \pm \omega t)$. However, $\cos(kx \pm \omega t)$ are also solutions, and the most general solution for a given $k$ or $\lambda$ is of the form\(^2\)

$$A \sin(kx + \omega t) + B \cos(kx + \omega t) + C \sin(kx - \omega t) + D \cos(kx - \omega t).$$

\(^2\)We fixed the speed $v$ when we wrote down the wave equation (C.1).
A complex form
\[ A e^{i(kx - \omega t)} + B e^{i(kx + \omega t)} \]  
(C.6)
can also be used. With the definition
\[ e^{i\theta} = \cos \theta + i \sin \theta, \]  
(C.7)
we can show that (C.5) and (C.6) are equivalent.

We generally use \( \sin(kx \pm \omega t) \) or \( \cos(kx \pm \omega t) \) for classical waves which have to be real. On the other hand, the complex form \( e^{i(kx \pm \omega t)} \) may be preferable in quantum mechanics as it often simplifies the computation.

**C.1.2 Standing Wave Solutions**

Generally speaking, standing waves can be produced by two identical waves traveling in opposite directions. Due to the superposition principle, these waves form an interference pattern that does not seem to move.

**C.1.2.1 Standing Waves on a String**

For a string of length \( L \) fixed at both ends, the solutions can be expressed as follows.

\[ y(x, t) = A \sin \omega_n t \sin \frac{n \pi x}{L} \]  
(C.8)

If \( T \) represents a constant horizontal tension, we have

\[ \omega_n = \sqrt{\frac{T}{\rho \ L}} \quad \text{for} \quad n = 1, 2, 3, \ldots \]  
(C.9)

For the string, the propagation speed \( v = \sqrt{\frac{T}{\rho}} \), and hence the frequency

\[ \nu_n = \frac{\omega_n}{2\pi} = \sqrt{\frac{T}{\rho \ L}} \frac{n \pi}{2\pi} = \frac{v}{L} \frac{n \pi}{2\pi} = \frac{n v}{2L}. \]  
(C.10)
C.1.2.2 Standing Waves in Quantum Mechanics

A typical standing wave in quantum mechanics is generated when an incident wave interferes with a reflected wave. Such reflection occurs at a potential boundary. For example, on p.138 in Section 7, an incident wave originating at $x = -\infty$ interferes with the reflected wave bouncing back from the potential discontinuity at $x = 0$. The full wavefunction given in (7.45)

$$\Psi(x, t) = D \left( \cos k_1 x - \frac{k_2}{k_1} \sin k_1 x \right) e^{-iEt/h} \quad (C.11)$$

can be rewritten as follows.

$$\Psi(x, t) = \sqrt{\alpha^2 + \beta^2} \cos(kx + \theta)e^{-iEt/h} \quad (C.12)$$

Therefore, as claimed on p.139, the wavefunction $\Psi(x, t)$ has fixed nodes, which in turn implies that it is a standing wave by definition.
Appendix D
Continuous Functions

It is neither necessary nor feasible to conduct all derivations in physics in a mathematically rigorous manner. This is particularly so when a new theory is being developed. More often than not, however, sloppy mathematics when learning an established theory becomes an obstruction to going beyond the theory and doing own physics. It is for this reason that mathematically rigorous definitions and proofs are presented here.

D.1 Definition of Continuity

We specialize to a one-variable function \( f(x) \) with \( x \in \mathbb{R} \) here.

**Definition D.1 (Continuous Functions: \( \delta-\epsilon \))** A function \( f(x) \) is continuous at \( x = c \in \mathbb{R} \) if for an arbitrarily chosen \( \epsilon > 0 \) there exists \( \delta_\epsilon > 0 \) such that \( |x - c| < \delta_\epsilon \implies |f(x) - f(c)| < \epsilon \).

In words, this means that \( f(x) \) is arbitrarily close to \( f(c) \) if \( x \) is sufficiently close to \( c \). Sometimes this is referred to as the delta-epsilon (\( \delta-\epsilon \)) definition and is one of several equivalent definitions of continuity. Some readers may prefer another definition in terms of the convergence of a sequence.

**Definition D.2 (Convergence of a Sequence)** A sequence \( \{x_i\}_{i \in \mathbb{N}} \) converges to \( c \) if for an arbitrary \( \epsilon > 0 \) there exists \( n_\epsilon \in \mathbb{N} \) such that \( i > n_\epsilon \implies |x_i - c| < \epsilon \). When this happens, we write

\[
\lim_{i \to \infty} x_i = c.
\] (D.1)
This means that \( x_i \) is arbitrarily close to \( c \) if \( i \) is sufficiently large.

**Definition D.3 (Continuous Functions: Convergence of Sequences)**  A function \( f(x) \) is continuous at \( x = c \in \mathbb{R} \) if for any sequence \( \{x_i\}_{i \in \mathbb{N}} \) converging to \( c \) the sequence \( \{f(x_i)\}_{i \in \mathbb{N}} \) converges to \( f(c) \). That is,

\[
\lim_{i \to \infty} x_i = c \implies \lim_{i \to \infty} f(x_i) = f(c). \tag{D.2}
\]

**D.2 Extreme and Intermediate Value Theorems**

There are two powerful and convenient theorems for a continuous function defined on a closed interval \([a, b]\).

**Theorem D.1 (Extreme Value Theorem)** If a continuous function \( f \) is defined on a closed interval \([a, b]\), then the function attains both its maximum and minimum values on the interval. That is to say there are numbers \( c, d \in [a, b] \) such that \( f(c) \geq f(x) \) and \( f(d) \leq f(x) \) for all \( x \in [a, b] \).

Note that, this does not necessarily apply to open endpoints. For example, consider \( e^x \) on the interval \((0, 1]\). While the maximum value of \( e^1 = e \) is indeed taken, the minimum does not even exist. All we can say is that \( e^x \) is bounded from below by 1, but \( e^x \) is always strictly greater than 1 on \((0, 1]\).

**Theorem D.2 (Intermediate Value Theorem)** If a continuous function \( f \) is defined on a closed interval \([a, b]\) and \( f(a) \leq k \leq f(b) \) or \( f(a) \geq k \geq f(b) \), one can find \( c \in [a, b] \) such that \( f(c) = k \).

This means that any value between \( f(a) \) and \( f(b) \) is actually taken by \( f \) at least once on the interval \([a, b]\).

**D.3 Continuity Condition on the First Derivative**

We are now ready to give a more rigorous proof of the continuity condition shown on p.128 when the potential \( V(x) \) is continuous. We will show that the first derivative \( \frac{d\psi(x)}{dx} \) is continuous at \( x = x_0 \).

First note that it suffices to consider \( \delta < 1 \) in the \( \delta - \epsilon \) definition (Definition D.1) of continuity. Because both \( V(x) \) and \( \psi(x) \) are continuous, \( |V(x) - E| \) and \( |\psi(x)| \)
are continuous and bounded from above by some finite number $N$ on the interval $[x_0 - 1, x_0 + 1]$. Now, given $\epsilon > 0$, let $\delta = \min\{1, \frac{h^2 \epsilon}{2mN^2}\}$. Then,

$$
|x - x_0| < \delta \implies \left| \left[ \frac{d\psi(x)}{dx} \right]_x - \left[ \frac{d\psi(x)}{dx} \right]_{x_0} \right| = \frac{2m}{\hbar^2} \left| \int_{x_0}^x (V(x) - E) \psi(x) \, dx \right|
$$

$$
\leq \frac{2m}{\hbar^2} \int_{x_0}^x |V(x) - E| |\psi(x)| \, dx \quad (x \geq x_0) \quad \text{or} \quad \frac{2m}{\hbar^2} \int_x^{x_0} |V(x) - E| |\psi(x)| \, dx \quad (x \leq x_0)
$$

$$
< \frac{2m}{\hbar^2} \cdot N^2 \cdot \delta \leq \frac{2m}{\hbar^2} \cdot N^2 \cdot \frac{h^2 \epsilon}{2mN^2} = \epsilon.
$$

(D.3)

Therefore, we have shown

$$
|x - x_0| < \delta \implies \left| \left[ \frac{d\psi(x)}{dx} \right]_x - \left[ \frac{d\psi(x)}{dx} \right]_{x_0} \right| < \epsilon.
$$

(D.4)

And, $\frac{d\psi(x)}{dx}$ is continuous at $x = x_0$. 
Appendix E

Wronskian and Linear Independence

The following theorems were taken from lecture notes prepared by Kiam Heong Kwa [Kwa, nd] except for the last one.

Definition E.1 (The Wronskian) Let $y_1$ and $y_2$ be two differentiable functions of $x$. Then, the Wronskian $W(y_1, y_2)$, associated to $y_1$ and $y_2$, is the function

$$W(y_1, y_2)(x) = \begin{vmatrix} y_1(x) & y_2(x) \\ y_1'(x) & y_2'(x) \end{vmatrix} = y_1(x)y_2'(x) - y_1'(x)y_2(x).$$

Theorem E.1 [Bocher, 1901] If $f(x)$ and $g(x)$ are differentiable functions on an open interval $I$ such that $W(f, g)(x_0) \neq 0$ for some point $x_0$ in $I$, then $f(x)$ and $g(x)$ are linearly independent on $I$. Equivalently, if $f(x)$ and $g(x)$ are linearly dependent on $I$, then $W(f, g)(x) = 0$ for all $x \in I$.

Remark E.1 The converse of Theorem E.1 does not hold without additional conditions on the pair of functions involved. That is to say, it is not true in general that two differentiable functions $f(x)$ and $g(x)$ with a vanishing Wronskian on an open interval must be linearly dependent on the same interval.

Theorem E.2 If $f(x)$ and $g(x)$ are differentiable functions on an open interval $I$ such that $f(x) \neq 0$ and $W(f, g)(x) = 0$ for all $x \in I$, then $f(x)$ and $g(x)$ are linearly dependent. In particular, $g(x) = kf(x)$ for a constant $k$. 

297
Theorem E.3 (Abel’s Theorem) Consider a linear second-order ordinary differential equation
\[ y'' + p(x)y' + q(x)y = 0; \]
(E.2)
where the coefficients \( p(x) \) and \( q(x) \) are continuous on an open interval \( I \).
The Wronskian \( W(y_1, y_2) \) is given by the following integral on \( I \) for some constant that depends on \( y_1(x) \) and \( y_2(x) \).
\[ W(y_1, y_2)(x) = c \exp \left[ -\int p(x) \, dx \right] \]  
(E.3)
To be more explicit, if \( x_0 \) is any given point in \( I \),
\[ W(y_1, y_2)(x) = W(y_1, y_2)(x_0) \exp \left[-\int_{x_0}^{x} p(s) \, ds \right] \]  
(E.4)
for all \( x \in I \).

Remark E.2 By (E.3) and (E.4), the Wronskian of any two solutions of (E.2) is either always negative, always zero, or always positive on the interval \( I \).

Theorem E.4 [Bostan and Dumas, 2010] Let \( y_1(x) \) and \( y_2(x) \) be solutions of (E.2). Then \( y_1(x) \) and \( y_2(x) \) are linearly independent if and only if \( W(y_1, y_2)(x) \neq 0 \) for all \( x \in I \). Equivalently, \( y_1(x) \) and \( y_2(x) \) are linearly dependent if and only if \( W(y_1, y_2)(x) = 0 \) for all \( x \in I \).

Definition E.2 (Analytic Functions) A function \( f(z) \) is said to be analytic about a point \( z_0 \) if it has a convergent power series representation in a neighborhood of \( z_0 \); that is,
\[ f(z) = \sum_{k=0}^{\infty} f_k(z - z_0)^k \]  
(E.5)
is convergent in an open set, or interval if \( z \) is real, containing \( z_0 \). In addition, the coefficients \( \{f_k\} \) are uniquely determined by
\[ f_k = \frac{f^{(k)}(z_0)}{k!} \]  
(E.6)
for each \( k = 0, 1, 2, \ldots \).
Therefore, a function is (globally) analytic if and only if its Taylor series about \( z_0 \) converges to the function in some neighborhood for every \( z_0 \) in its domain.
Theorem E.5 Let \( f(x) \) and \( g(x) \) be analytic functions on an open interval \( I \). If 
\( W(f, g)(x) = 0 \) on \( I \), then \( f(x) \) and \( g(x) \) are linearly dependent.

Fact E.1 (Principle of Superposition) Let \( y_1 \) and \( y_2 \) be solutions of a homogeneous linear differential equation of the general form

\[
A_0(x) \frac{d^n y(x)}{dx^n} + A_1(x) \frac{d^{n-1} y(x)}{dx^{n-1}} + \ldots \\
+ A_{n-1}(x) \frac{dy(x)}{dx} + A_n(x) y(x) = 0
\]

(E.7)

or

\[
\sum_{i=0}^{n} A_i(x) \frac{d^{n-i} y(x)}{dx^{n-i}} = \left[ \sum_{i=0}^{n} A_i(x) \frac{d^{n-i}}{dx^{n-i}} \right] y(x) = 0;
\]

(E.8)

where \( \frac{d^n}{dx^n} y(x) \) is to be regarded as \( y(x) \) by definition. Then,

\[
c_1 y_1(x) + c_2 y_2(x)
\]

(E.9)

is a solution of the same differential equation for any constants \( c_1 \) and \( c_2 \).

Theorem E.6 (Fundamental Set) If \( y_1(x) \) and \( y_2(x) \) are two linearly independent solutions of the equation

\[
y''(x) + p(x)y'(x) + q(x)y(x) = 0,
\]

(E.10)

then, any solution \( y(x) \) is given by

\[
y(x) = c_1 y_1(x) + c_2 y_2(x)
\]

(E.11)

for some constants \( c_1 \) and \( c_2 \). In this case, the set \( \{y_1(x), y_2(x)\} \) is called a fundamental set of solutions.

Corollary E.1 Two polynomials \( p(x) \) and \( q(x) \) are linearly dependent if and only if 
\( W(p, q)(x) = 0 \) everywhere.
Appendix F

Newtonian, Lagrangian, and Hamiltonian Mechanics

Newtonian mechanics, Lagrangian mechanics, and Hamiltonian mechanics are three equivalent formulations of classical mechanics. Therefore, any of these formulations can be used to solve a problem in classical mechanics. However, depending on the nature of the problem, one of these formulations leads to equations which are much easier to solve. One major difference between Newtonian mechanics and the other two is that Newtonian mechanics describes the system in terms of force, while Lagrangian and Hamiltonian mechanics do so in terms of energy. The following is a very brief review and description of the three formulations.

F.1 An Overview

Here is a summary of the qualitative differences among Newtonian, Lagrangian, and Hamiltonian formulations.

Newtonian formulation considers all the forces acting in/on the system. If you know the forces and the mass, you can get a differential equation for the position of each particle.

Lagrangian formulation is based on an action principle. In a conservative field, action should be minimized for the actual trajectory (path of motion) of a particle. The action in this case is what is known as the Lagrangian \( L = T - V \), total energy minus potential energy, and its minimization leads to a description of the behavior of each particle by way of Lagrange’s (differential) equations. The variables involved
are generalized positions (often denoted by \( q \)), generalized velocities (often denoted by \( \dot{q} \)), and time \( t \); i.e. \( L = L(q, \dot{q}, t) \).

Hamiltonian formulation is usually derived from the Lagrangian. When the total energy is conserved, which is the case in this book, the Hamiltonian \( H \) is nothing but the total energy of the system, \( H = T + V \). The variables are generalized positions \( q \), generalized momenta \( p \), and time \( t \); i.e. \( H = H(p, q, t) \). This is the formulation employed almost exclusively in quantum mechanics.

More quantitative/mathematical descriptions will follow.

**F.2 Newtonian Mechanics**

Newton’s second law is

\[
F_j = ma_j \quad \text{or} \quad F_j = m \frac{d^2 x_j}{dt^2} \quad \text{for} \quad j = 1, \ldots, n.
\]

(F.1)

The position \( x_j(t) \) is obtained by solving this second-order ordinary differential equation with some boundary condition such as the initial position.

The Lagrangian and Hamiltonian are formulated in an \( n \)-dimensional configuration space.

**F.3 Lagrangian Mechanics**

The Lagrangian, \( L \), is defined as the difference between the kinetic energy \( T \) and the potential energy \( V \).

\[
L = T - V
\]

(F.2)

The motion is described by Lagrange’s equations of motion

\[
\frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}_j} \right) = \frac{\partial L}{\partial q_j} \quad \text{for} \quad j = 1, \ldots, n;
\]

(F.3)

where \( q_j \)’s are generalized coordinates (positions) and \( \dot{q}_j \)’s are generalized velocities.\(^1\)

---

\(^1\)Generalized coordinates refer to the parameters that describe the configuration of the system relative to some reference configuration. Their time derivatives are the generalized velocities.
**F.4 Hamiltonian Mechanics**

The Hamiltonian $H$ is defined as the sum of the kinetic and potential energies.

$$H = T + V$$ \hspace{1cm} (F.4)

Hamilton’s equations of motion are as follows.

$$\frac{\partial H}{\partial q_j} = -\dot{p}_j \quad \text{and} \quad \frac{\partial H}{\partial p_j} = \dot{q}_j \quad \text{for} \quad j = 1, \ldots, n;$$ \hspace{1cm} (F.5)

where $q_j$’s are generalized coordinates (positions) and $p_j$’s are generalized momenta. We also have the following time derivatives connecting $H$ and $L$.

$$\frac{\partial H}{\partial t} = -\frac{\partial L}{\partial t}$$ \hspace{1cm} (F.6)

One apparent difference between the Lagrangian formulation and Hamiltonian formulation is that Lagrange’s equations describe the system by a set of $n$ second-order differential equations, while Hamilton’s equations are $n$ sets of coupled first-order differential equations. Hence, Hamiltonian mechanics requires $2n$ first-order differential equations while Lagrangian mechanics requires $n$ second-order differential equations. As mentioned in the beginning on p.301, one can also note that “force” enters the equation of motion explicitly only for Newtonian mechanics.
Appendix G

Normalization Schemes for a Free Particle

We will only consider $\psi(x) = Ae^{i k x}$ for $k = \sqrt{\frac{2mE}{\hbar}}$. Computations for $\psi(x) = Ae^{-i k x}$ are completely analogous.

G.1 The Born Normalization

This highly controversial but widely used procedure involves restricting the range of integration to a finite length $L$ and letting $L$ to infinity afterwards. If we only consider real and positive normalization constants,

$$\int_{-L/2}^{+L/2} (Ae^{i k x})^* (Ae^{i k x}) \ dx = |A|^2 \int_{-L/2}^{+L/2} e^{-i k x} e^{i k x} \ dx = |A|^2 L = 1$$

$$\implies A = \frac{1}{\sqrt{L}}. \quad (G.1)$$

Hence, the Born-normalized wavefunction is

$$\psi(x) = \frac{1}{\sqrt{L}} e^{i k x}. \quad (G.2)$$

G.2 The Dirac Normalization

Dirac normalization is based on the Dirac delta function $\delta(x-x')$ described in Section 4.3. In particular, we will use the relation between Fourier transform and the delta function.
G.2.1 Fourier Transform and the Delta Function

The derivations found in Section 4.5 is reproduced below with slightly different notations more fitting for this argument.\(^1\) For a function \(f(x)\), we have

\[
F(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x) e^{-ikx} \, dx \text{ the Fourier transform (G.3)}
\]

and

\[
f(x') = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(k) e^{ikx'} \, dk \text{ the inverse Fourier transform (G.4)}
\]

Substituting (G.3) for \(F(k)\) in (G.4),

\[
f(x') = \int_{-\infty}^{+\infty} f(x) \left[ \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{ik(x-x')} \, dk \right] \, dx. \tag{G.5}
\]

This leads to one representation of Dirac delta function

\[
\delta(x' - x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{ik(x-x')} \, dk = \frac{1}{2\pi\hbar} \int_{-\infty}^{+\infty} e^{ip(x'-x)/\hbar} \, dp; \tag{G.6}
\]

where we used the relation \(k = \frac{\sqrt{2mE}}{\hbar} = \sqrt{\frac{2mE}{\hbar^2}} = \frac{p}{\hbar}\). Similarly,

\[
\frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-ik'x} e^{ikx} \, dx = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{ix(k-k')} \, dx = \delta(k - k'). \tag{G.7}
\]

G.2.2 Wavefunctions as Momentum or Position Eigenfunctions

For a free particle, we have

\[
H = \frac{P^2}{2m} \tag{G.8}
\]

and the commutator

\[
[H, P] = 0. \tag{G.9}
\]

\(^1\)In the author’s modest and personal opinion, seeing/understanding the same derivation and/or relation expressed with different notations is one good way to solidify your grasp of a mathematical and physical concept.
Therefore, the stationary solutions, or the eigenfunctions, of the Hamiltonian $H$ are also eigenfunctions of the momentum operator $P$. Indeed,

$$PAe^{ikx} = -i\hbar \frac{d}{dx} Ae^{ikx} = -i\hbar \left(\frac{p}{\hbar}\right) Ae^{ikx} = pAe^{ikx}. \tag{G.10}$$

In the case of a free particle, the distribution of the momentum eigenvalues \{p\} is continuous ranging from $-\infty$ to $+\infty$ just like the position \{x\}. This is the situation described in Section 4.8. We will label each normalized eigenket of $P$ by its associated momentum $p$ to obtain \{\ket{p}\} or \{A_p e^{i\frac{p}{\hbar}x}\}; where $A_p$’s are normalization constants. Now, from (4.87), the representation of $X$ in the $P$ basis is given by

$$X = i\hbar \frac{d}{dp}. \tag{G.11}$$

Noting that

$$XAe^{-ikx} = i\hbar \frac{d}{dp} Ae^{-i\frac{p}{\hbar}x} = i\hbar \left(-i\frac{x}{\hbar}\right) Ae^{-i\frac{p}{\hbar}x} = xAe^{-ikx}, \tag{G.12}$$

we will label each normalized eigenket of $X$ by its associated position eigenvalue $x$ to obtain \{\ket{x}\} or \{A_x e^{-i\frac{p}{\hbar}x}\}; where $A_x$’s are normalization constants.

It is important to understand the reason behind writing $\ket{p}$ as $A_p e^{i\frac{p}{\hbar}x}$ and $\ket{x}$ as $A_x e^{-i\frac{p}{\hbar}x}$. When we interpret $p$ as a fixed value and $x$ as the index in the $X$ space, we write $e^{i\frac{p}{\hbar}x}$, while we write $e^{-i\frac{p}{\hbar}x}$ when $x$ is fixed and $p$ is the index in the $P$ space. It is all a matter of what is interpreted as fixed and what is interpreted as a running index.

In the braket notation, we get

$$\langle x|p \rangle = A_p e^{i\frac{p}{\hbar}x} \tag{G.13}$$

and

$$\langle p|x \rangle = A_x e^{-i\frac{p}{\hbar}x}. \tag{G.14}$$

Needless to say, $A_x = A_p^*$ is a necessary condition because $\langle p|x \rangle = \langle x|p \rangle^*$ gives

$$A_x e^{-i\frac{p}{\hbar}x} = A_p^* e^{-i\frac{p}{\hbar}x} \implies A_x = A_p^*. \tag{G.15}$$

Based on (4.7), the appropriate normalization scheme suggests itself at this point.
Definition G.1 (The Dirac Normalization) The free particle eigenfunctions \{\langle p | \rangle\}

are normalized a la Dirac if

\[ \langle p' | p \rangle = \int_{-\infty}^{+\infty} (A_p e^{ik'_x})^* (A_p e^{ik_x}) \, dx = \delta(p - p'). \tag{G.16} \]

We have

\[ \int_{-\infty}^{+\infty} (A_p e^{ik'_x})^* (A_p e^{ik_x}) \, dx = |A_p|^2 \int_{-\infty}^{+\infty} e^{-ik'_x} e^{ik_x} \, dx \]

\[ = |A_p|^2 \int_{-\infty}^{+\infty} e^{-i\pi x} e^{i\pi x} \, dx = |A_p|^2 \int_{-\infty}^{+\infty} e^{-i\pi x (p' - p) / \hbar} \, dx \]

\[ = 2\pi\hbar |A_p|^2 \left[ \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-iy(p' - p)} \, dy \right] = 2\pi\hbar |A_p|^2 \delta(p - p'); \tag{G.17} \]

where we let \( y = x / \hbar \). The expression (G.17) becomes \( \delta(p - p') \) if we let \( A_p = \frac{1}{\sqrt{2\pi\hbar}} \).

This gives us \( |p\rangle = \psi_p \) with

\[ \langle x | p \rangle = \psi_p(x) = \frac{1}{\sqrt{2\pi\hbar}} e^{ikx} \text{ or } \frac{1}{\sqrt{2\pi\hbar}} e^{ix \hbar}. \tag{G.18} \]

with the proposed normalization (G.16)

\[ \langle p | p' \rangle = \int_{-\infty}^{+\infty} \psi_p(x)^* \psi_{p'}(x) \, dx = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-iy(p' - p)} \, dy = \delta(p - p'). \tag{G.19} \]

We also have

\[ \int_{-\infty}^{+\infty} (A_x e^{-ik'_x})^* (A_x e^{-ik_x}) \, dp = |A_x|^2 \int_{-\infty}^{+\infty} e^{ik'_x} e^{-ik_x} \, dp \]

\[ = |A_x|^2 \int_{-\infty}^{+\infty} e^{i\pi p} e^{-i\pi p} \, dp = |A_x|^2 \int_{-\infty}^{+\infty} e^{i\pi (x' - x) / \hbar} \, dp \]

\[ = 2\pi\hbar |A_x|^2 \left[ \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{i\pi (x' - x)} \, dy \right] = 2\pi\hbar |A_x|^2 \delta(x - x'); \tag{G.20} \]

where we let \( y = p / \hbar \). The expression (G.20) becomes \( \delta(x - x') \) if we let \( A_x = \frac{1}{\sqrt{2\pi\hbar}} \).

This gives us \( |x\rangle = \psi_x \) with

\[ \langle p | x \rangle = \psi_x(p) = \frac{1}{\sqrt{2\pi\hbar}} e^{ikx} \text{ or } \frac{1}{\sqrt{2\pi\hbar}} e^{ix \hbar}. \tag{G.21} \]

with the normalization

\[ \langle x | x' \rangle = \int_{-\infty}^{+\infty} \psi_x(p)^* \psi_{x'}(p) \, dp = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{iy(x' - x)} \, dy = \delta(x - x'). \tag{G.22} \]
We have made sure $A_p = A^*_x$ as required by (G.15).

Finally, note again that there is symmetry between the Dirac normalization condition $\langle p | p' \rangle = \delta (p - p')$ for the eigenkets $\{| p \rangle \}$ and the completeness (or closure) condition $\langle x | x' \rangle = \delta (x - x')$ for the position kets $\{| x \rangle \}$.

### G.3 The Unit-Flux Normalization

This is based on the concept of particle current density discussed in Appendix J. A wavefunction is normalized if it has a unit current density. Namely, the following functions are normalized in the unit flux sense.

$$
\psi_{\pm}(x) = \frac{1}{\sqrt{\hbar k/m}} e^{\pm ikx}; \text{ where } k = \frac{\sqrt{2mE}}{\hbar}
$$

Let us verify that we indeed have a unit current density.

$$
\dot{j} = \frac{i\hbar}{2m} \left( \psi \frac{\partial}{\partial x} \psi^* - \psi^* \frac{\partial}{\partial x} \psi \right)
= \frac{i\hbar}{2m} \left( \frac{1}{\sqrt{\hbar k/m}} e^{\pm ikx} \frac{\partial}{\partial x} \frac{1}{\sqrt{\hbar k/m}} e^{\mp ikx} - \frac{1}{\sqrt{\hbar k/m}} e^{\mp ikx} \frac{\partial}{\partial x} \frac{1}{\sqrt{\hbar k/m}} e^{\pm ikx} \right)
= \frac{i\hbar}{2m (\hbar k/m)} \left( e^{\pm ikx} (\mp ik) e^{\mp ikx} - e^{\mp ikx} (\pm ik) e^{\pm ikx} \right)
= \frac{i}{2k} (\mp 2ik) = \pm 1
$$

The $\pm$ sign signifies the direction of the flow relative to the way the $x$-axis is set up. Now note that

$$
\hbar k/m = \left( \hbar \frac{P}{m} \right) / m = \frac{P}{m} = \frac{mv}{m} = v.
$$

Hence, the probability density is given by

$$
\psi_{\pm}(x)^* \psi_{\pm}(x) = \left( \frac{1}{\sqrt{\hbar k/m}} e^{\pm ikx} \right)^* \frac{1}{\sqrt{\hbar k/m}} e^{\mp ikx} = \frac{1}{\hbar k/m} = \frac{1}{v},
$$

which provides a classical explanation for the unit flux because the “amount” of probability that flows through any point $x = x_0$ per unit time, say one second, is $\frac{1}{v} \times v = 1$. 

Appendix H
Symmetries and Conserved Dynamical Variables

The description and consequences of symmetry are important and beautiful components of both classical mechanics and quantum mechanics. In the context of physics, symmetry means that a certain observable remains constant under some transformation. In other words, symmetry is always in reference to a given transformation.

H.1 Poisson Brackets and Constants of Motion

Definition H.1 (Poisson Bracket) Let \( p \) and \( q \) denote sets of \( n \) generalized momenta and generalized coordinates: \( p = \{ p_1, p_2, \ldots, p_n \} \) and \( q = \{ q_1, q_2, \ldots, q_n \} \). The Poisson bracket between \( \omega(p, q) \) and \( \lambda(p, q) \), denoted by \( \{ \omega, \lambda \} \), is defined as below.

\[
\{ \omega, \lambda \} = \sum_{i=1}^{n} \left( \frac{\partial \omega}{\partial q_i} \frac{\partial \lambda}{\partial p_i} - \frac{\partial \omega}{\partial p_i} \frac{\partial \lambda}{\partial q_i} \right)
\]

Poisson bracket arises naturally as the total time derivative of a function \( \omega(p, q) \), which does not depend on time \( t \) explicitly, as follows.

\[
\frac{d\omega}{dt} = \sum_{i=1}^{n} \left( \frac{\partial \omega}{\partial q_i} \dot{q}_i + \frac{\partial \omega}{\partial p_i} \dot{p}_i \right)
= \sum_{i=1}^{n} \left( \frac{\partial \omega}{\partial q_i} \frac{\partial H}{\partial p_i} + \frac{\partial \omega}{\partial p_i} \frac{\partial H}{\partial q_i} \right)
= \{ \omega, H \}
\]
Fact H.1 From (H.2), we can see that a variable $\gamma$ whose Poisson bracket with $\mathcal{H}$ vanishes, $\{\gamma, \mathcal{H}\} = 0$, is a constant of motion (COM), or preserved.

Theorem H.1 (Invariant $\mathcal{H}$ and Conservation of a Dynamical Variable) Consider a dynamical variable $g(p, q)$ which generates the following infinitesimal transformation

$$q_i \rightarrow \bar{q}_i = q_i + \varepsilon \frac{\partial g}{\partial p_i} = q_i + \delta q_i$$

$$p_i \rightarrow \bar{p}_i = p_i - \varepsilon \frac{\partial g}{\partial q_i} = p_i + \delta p_i.$$  \hspace{1cm} (H.3)

If $\mathcal{H}$ is invariant under the infinitesimal transformation (H.3), then $g$ is conserved; i.e. $g$ is a constant of motion (COM).

Proof

The Taylor series $T_f(x_1, x_2, \ldots, x_n)$ of a multivariate function $f(x_1, x_2, \ldots, x_n)$ about a point $(a_1, a_2, \ldots, a_n)$ is given by

$$T_f(x_1, x_2, \ldots, x_n) = \sum_{i_1=0}^{\infty} \sum_{i_2=0}^{\infty} \cdots \sum_{i_n=0}^{\infty} \frac{(x_1 - a_1)^{i_1}(x_2 - a_2)^{i_2} \cdots (x_n - a_n)^{i_n}}{i_1!i_2! \cdots i_n!} \frac{\partial^1 f}{\partial x_1^{i_1} \partial x_2^{i_2} \cdots \partial x_n^{i_n}}(a_1, a_2, \ldots, a_n).$$  \hspace{1cm} (H.4)

Therefore, the Taylor series expansion $T_{\mathcal{H}}(p_1, p_2, \ldots, p_n, q_1, q_2, \ldots, q_n)$ of our Hamiltonian

$$\mathcal{H}(p, q) = \mathcal{H}(p_1, p_2, \ldots, p_n, q_1, q_2, \ldots, q_n)$$  \hspace{1cm} (H.5)

about the point $(a_1, a_2, \ldots, a_{2n})$ is given by

$$T_{\mathcal{H}}(p_1, p_2, \ldots, p_n, q_1, q_2, \ldots, q_n) = \sum_{i_1=0}^{\infty} \sum_{i_2=0}^{\infty} \cdots \sum_{i_{2n}=0}^{\infty} \frac{(p_1 - a_1)^{i_1} \cdots (p_n - a_{2n})^{i_{2n}}}{i_1!i_2! \cdots (i_{2n})!} \frac{\partial^{i_1 + i_2 + \cdots + i_{2n}} \mathcal{H}}{\partial p_1^{i_1} \partial p_2^{i_2} \cdots \partial q_n^{i_{2n}}}(a_1, a_2, \ldots, a_{2n}).$$  \hspace{1cm} (H.6)

Now we let $p_i = \bar{p}_i$, $q_i = \bar{q}_i$, $a_i = p_i$ for $1 \leq i \leq n$, and $a_j = q_j$ for $n + 1 \leq j \leq 2n$ in (H.6).

$$T_{\mathcal{H}}(\bar{p}_1, \bar{p}_2, \ldots, \bar{p}_n, \bar{q}_1, \bar{q}_2, \ldots, \bar{q}_n) = \sum_{i_1=0}^{\infty} \sum_{i_2=0}^{\infty} \cdots \sum_{i_{2n}=0}^{\infty} \frac{(\bar{p}_1 - p_1)^{i_1} \cdots (\bar{q}_n - q_n)^{i_{2n}}}{i_1!i_2! \cdots (i_{2n})!} \frac{\partial^{i_1 + i_2 + \cdots + i_{2n}} \mathcal{H}}{\partial p_1^{i_1} \partial p_2^{i_2} \cdots \partial q_n^{i_{2n}}}(a_1, a_2, \ldots, a_{2n}).$$

\[1\] We say $g$ is the generator of the infinitesimal transformation (H.3).
\[
\left( \frac{\partial^{i_1 + i_2 + \cdots + i_{2n}} \mathcal{H}}{\partial p_1^{i_1} \partial p_2^{i_2} \cdots \partial q_n^{i_{2n}}} \right) (p_1, p_2, \ldots, q_n). \tag{H.7}
\]

Note that \( \bar{p}_i - p_i = -\varepsilon \frac{\partial g}{\partial q_i} \) and \( \bar{q}_i - q_i = \varepsilon \frac{\partial g}{\partial p_i} \) from (H.3). Substituting these into (H.7), we get

\[
T \mathcal{H}(\bar{p}_1, \bar{p}_2, \ldots, \bar{p}_n, \bar{q}_1, \bar{q}_2, \ldots, \bar{q}_n) = \sum_{i_1=0}^{\infty} \sum_{i_2=0}^{\infty} \cdots \sum_{i_{2n}=0}^{\infty} \frac{(-\varepsilon \frac{\partial g}{\partial q_i})^{i_1} \cdots (\varepsilon \frac{\partial g}{\partial p_i})^{i_{2n}}}{i_1! i_2! \cdots (i_{2n})!} \left( \frac{\partial}{\partial p_1^{i_1}} \frac{\partial}{\partial p_2^{i_2}} \cdots \frac{\partial}{\partial q_n^{i_{2n}}} \right) (p_1, p_2, \ldots, q_n) = \mathcal{H}(p_1, p_2, \ldots, q_n) + \sum_{i=1}^{n} \left[ \frac{\partial \mathcal{H}}{\partial p_i} \left( -\varepsilon \frac{\partial g}{\partial q_i} \right) + \frac{\partial \mathcal{H}}{\partial q_i} \left( \varepsilon \frac{\partial g}{\partial p_i} \right) \right] + \mathcal{O}(\varepsilon^2). \tag{H.8}
\]

As \( \mathcal{H} \) is invariant under (H.3), we have

\[
\mathcal{H}(\bar{p}_1, \bar{p}_2, \ldots, \bar{q}_n) - \mathcal{H}(p_1, p_2, \ldots, q_n) = \sum_{i=1}^{n} \left[ \frac{\partial \mathcal{H}}{\partial p_i} \left( -\varepsilon \frac{\partial g}{\partial q_i} \right) + \frac{\partial \mathcal{H}}{\partial q_i} \left( \varepsilon \frac{\partial g}{\partial p_i} \right) \right] + \mathcal{O}(\varepsilon^2) = \varepsilon \mathcal{H}(\mathcal{H}, g) + \mathcal{O}(\varepsilon^2) = 0 \tag{H.9}
\]

Suppose \( \{\mathcal{H}, g\}(p, q) \), which is \( \{\mathcal{H}, g\} \) evaluated at \((p, q)\), is \( \gamma > 0 \) or \( \delta < 0 \). As \( \mathcal{O}(\varepsilon) \xrightarrow{\varepsilon \to 0} 0 \), \( |\{\mathcal{H}, g\} + \mathcal{O}(\varepsilon)| > 0 \) if \( \varepsilon \) is sufficiently small. But, this contradicts (H.9). Therefore, \( \{\mathcal{H}, g\} = 0 \), and this in turn implies that the dynamical variable \( g \) is conserved due to (H.2).

**H.2 Lz As a Generator**

Let \( g \) in (H.3) be the \( z \)-component of angular momentum; i.e. \( g = l_z = xp_y - yp_x \) from (9.1). Then, we have the following.

\[
x \rightarrow \bar{x} = x + \varepsilon \frac{\partial g}{\partial p_x} = x - y\varepsilon \tag{H.10}
\]
Comparing this with (H.20), we can see that this is an infinitesimal rotation around the z-axis. Hence, the angular momentum around the z-axis is the generator of rotations around the z-axis. Therefore, \( l_z \) is conserved if \( \mathcal{H} \) is invariant under rotations of a physical system around the z-axis. For example, if we only have a central force \( F(r) \), and hence a central potential \( V(r) \), there is no force in the tangential direction as we move around a circle \( \|\vec{r}\| = r_0, \theta = \theta_0 \), whose center is on the z-axis. Because we have no torque in this case, it follows that \( l_z \) is a constant of motion (COM).

**H.3 Rotation Around The Origin**

One simple example of such an observable-transformation pair is the distance from the origin, \( f(x, y, z) = \sqrt{x^2 + y^2 + z^2} \), of a particle under a rotation about the origin denoted by \( \mathcal{R} \). In this case, if the rotation is by \( \phi \) about the x-axis, followed by a rotation by \( \theta \) about the y-axis, followed by a rotation by \( \psi \) about the z-axis, all in the counter-clockwise direction, the transformation \( \mathcal{R} \) is given by the following matrix.

\[
\mathcal{R} = \begin{bmatrix}
\cos \theta \cos \phi & \cos \phi \sin \psi + \sin \phi \sin \theta \cos \psi & \sin \phi \sin \psi - \cos \phi \sin \psi - \cos \phi \sin \theta \cos \psi \\
-\cos \theta \sin \psi & \cos \phi \cos \psi - \sin \phi \sin \theta \sin \psi & \sin \phi \cos \psi + \cos \phi \sin \theta \sin \psi \\
\sin \theta & -\sin \phi \cos \theta & \cos \phi \cos \theta
\end{bmatrix}
\] (H.13)

As it is very cumbersome to carry on a computation with the full three-dimensional rotation matrix, we will only consider a counter-clockwise rotation \( \mathcal{R}_z \) about the z-axis given by the matrix below.

\[
\mathcal{R}_z = \begin{bmatrix}
\cos \theta & -\sin \theta & 0 \\
\sin \theta & \cos \theta & 0 \\
0 & 0 & 1
\end{bmatrix}
\] (H.14)

Under the transformation, \( (x, y, z) \) becomes \( (\bar{x}, \bar{y}, \bar{z}) \).

\[
\begin{bmatrix}
\bar{x} \\
\bar{y} \\
\bar{z}
\end{bmatrix} = \mathcal{R}_z \begin{bmatrix}
x \\
y \\
z
\end{bmatrix} = \begin{bmatrix}
\cos \theta & -\sin \theta & 0 \\
\sin \theta & \cos \theta & 0 \\
0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
x \\
y \\
z
\end{bmatrix} = \begin{bmatrix}
x \cos \theta - y \sin \theta \\
x \sin \theta + y \cos \theta \\
z
\end{bmatrix}
\] (H.15)

So,

\[ f(\mathcal{R}_z[x, y, z]) = f(\bar{x}, \bar{y}, \bar{z}) = f(x \cos \theta - y \sin \theta, x \sin \theta + y \cos \theta, z) \]
H.4. INFINITESIMAL ROTATIONS AROUND THE Z-AXIS

\[ H.16 \]

\[ f = \sqrt{(x \cos \theta - y \sin \theta)^2 + (x \sin \theta + y \cos \theta)^2 + z^2} \]

\[ = \sqrt{x^2 \cos^2 \theta - 2xy \sin \theta \cos \theta + y^2 \sin^2 \theta + 2x \sin \theta \cos \theta + y^2 \cos^2 \theta + z^2} \]

\[ = \sqrt{x^2 (\sin^2 \theta + \cos^2 \theta) + y^2 (\sin^2 \theta + \cos^2 \theta) - 2xy \sin \theta \cos \theta + 2xy \sin \theta \cos \theta + z^2} \]

\[ = \sqrt{x^2 + y^2 + z^2} = f(x,y,z), \quad (H.16) \]

and the distance from the origin \( f \) is invariant under a rotation about the \( z \)-axis.

Similarly, we can also show

\[ f(R_x[x,y,z]) = f(x,y,z) \quad \text{and} \quad f(R_y[x,y,z]) = f(x,y,z); \quad (H.17) \]

where \( R_x \) and \( R_y \) are rotations about the \( x \)- and \( y \)-axes, respectively. As any rotation is a combination of \( R_x, R_y, \) and \( R_z \), this proves that the distance from the origin is invariant under any rotation about the origin as it should be.

\[ \]

H.4 Infinitesimal Rotations Around the \( z \)-Axis

We will first specialize to rotations about the origin in two dimensions for computational simplicity. The results can be extended to rotations about the \( z \)-axis in three dimensions in a straightforward manner.

In two dimensions, we only have \( R_z \). The effect of \( R_z \) by \( \theta \) is as follows.

\[
\begin{bmatrix}
  x \\
  y
\end{bmatrix}
= R_z
\begin{bmatrix}
  x \\
  y
\end{bmatrix}
= \begin{bmatrix}
  \cos \theta & -\sin \theta \\
  \sin \theta & \cos \theta
\end{bmatrix}
\begin{bmatrix}
  x \\
  y
\end{bmatrix}
= \begin{bmatrix}
  x \cos \theta - y \sin \theta \\
  x \sin \theta + y \cos \theta
\end{bmatrix}.
\]

Next, consider an infinitesimal transformation which differs from the identity transformation \( I \) by angle \( \varepsilon \). In the limit as \( \varepsilon \) approaches 0, \( \cos \varepsilon \) approaches 1 and \( \sin \varepsilon \) tends to \( \varepsilon \). Hence,

\[
\begin{bmatrix}
  \bar{x} \\
  \bar{y}
\end{bmatrix}
= \begin{bmatrix}
  x \cos \varepsilon - y \sin \varepsilon \\
  x \sin \varepsilon + y \cos \varepsilon
\end{bmatrix}
\xrightarrow{\varepsilon \to 0}
\begin{bmatrix}
  x - y\varepsilon \\
  x\varepsilon + y
\end{bmatrix}.
\]

(H.19)

This extends to three dimensions as below.

\[
\begin{bmatrix}
  \bar{x} \\
  \bar{y} \\
  \bar{z}
\end{bmatrix}
= \begin{bmatrix}
  x \cos \varepsilon - y \sin \varepsilon \\
  x \sin \varepsilon + y \cos \varepsilon \\
  z
\end{bmatrix}
\xrightarrow{\varepsilon \to 0}
\begin{bmatrix}
  x - y\varepsilon \\
  x\varepsilon + y \\
  z
\end{bmatrix}.
\]

(H.20)
Appendix I

Commutators

I.1 Commutator Identities

Here are some basic and useful commutator identities.

\[
[A, B + C] = [A, B] + [A, C] \\
[A + B, C] = [A, C] + [B, C] \\
[AB, C] = [A, C]B + A[B, C]
\] (I.1)

\[
\] (I.2)

\[
\] (I.3)

\[
\] (I.4)

\[
\] (I.5)

\[
[A, [B, C]] + [C, [A, B]] + [B, [C, A]] = 0
\] (I.6)

Let us verify (I.3), (I.4), and (I.5). It is straightforward for (I.3) and (I.4).

\[
\] (I.7)

Likewise for (I.4). Verification of (I.5) is similar.

\[
[AB, CD] = [(AB), CD] = [AB, C]D + C[AB, D]
\]

\[
\]

\[
\] (I.8)

I.2 Commutators involving \( X \) and \( P \)

From Tables 3.1 and 3.2, we have \( M_{x_i} \) (multiplication by \( x_i \)) and \( \hat{x}_i \left( -i\hbar \frac{\partial}{\partial x_i} \right) \) for \((x_1, x_2, x_3) = (x, y, z)\) and \((\hat{x}_1, \hat{x}_2, \hat{x}_3) = (\hat{i}, \hat{j}, \hat{k})\). Of 18 possible combinations for the
commutator, there are only three that are nonzero. We have:

\[
[X_i, X_j] = [M_{x_i}, M_{x_j}] = 0 \quad \text{for all } (i, j)
\]

(I.9)

\[
[P_i, P_j] = \left[ -i\hbar \frac{\partial}{\partial x_i}, -i\hbar \frac{\partial}{\partial x_j} \right] = 0 \quad \text{for all } (i, j)
\]

(I.10)

\[
[P_i, X_j] = -i\hbar \delta_{ij} \quad \text{for all } (i, j).
\]

(I.11)

These relations are very useful in computing other commutators such as those among \( L_x, L_y, L_z, L^2, \) and \( H. \)

I.3 Commutators involving \( X, P, L, \) and \( r \)

\[
[L_z, x] = i\hbar y
\]

(I.12)

\[
[L_z, y] = -i\hbar x
\]

(I.13)

\[
[L_z, z] = 0
\]

(I.14)

\[
[L_z, p_x] = i\hbar p_y
\]

(I.15)

\[
[L_z, p_y] = -i\hbar p_x
\]

(I.16)

\[
[L_z, p_z] = 0
\]

(I.17)

\[
[L_z, r^2] = 0
\]

(I.18)

\[
[L_z, p^2] = 0
\]

(I.19)

\[
[H, L_x] = [H, L_y] = [H, L_z] = 0 \quad \text{when } V = V(r)
\]

(I.20)

\[
[H, L^2] = 0
\]

(I.21)

Verification

\[
[L_z, x] = [xp_y - yp_x, x] = [xp_y, x] - [yp_x, x] = [x, x]p_y + x[p_y, x] - ([y, x]p_x + y[p_x, x])
\]

\[
= 0 + 0 - y[p_x, x] = -y(-i\hbar) = i\hbar y \quad \text{verification of I.12}
\]

\[
[L_z, p_x] = [xp_y - yp_x, p_x] = [xp_y, p_x] - [yp_x, p_x] = [x, p_x]p_y + x[p_y, p_x]
\]

\[
= 0 + 0 - y[p_x, p_x] = -y(-i\hbar) = i\hbar y
\]

One can write these as \([H, L] = 0.\)
\(-([y,p_x]p_x + y[p_x,p_x]) = i\hbar p_y + 0 - 0 = i\hbar p_y\) (verification of I.15)

Recall from (9.3) that

\[ L_z = -i\hbar \frac{\partial}{\partial \phi}. \] (1.22)

Because \(L_z\) is a partial derivative with respect to \(\phi\), it commutes with multiplication by any function of \(r\), including \(r^2\). This proves (I.18).

\[
[L_z, p^2] = [L_z, p_x^2 + p_y^2 + p_z^2] = [L_z, p_x^2] + [L_z, p_y^2] + [L_z, p_z^2] \\
= [L_z, p_x]p_x + p_x[L_z, p_x] + [L_z, p_y]p_y + p_y[L_z, p_y] + [L_z, p_z]p_z + p_z[L_z, p_z] \\
= i\hbar p_x p_x + i\hbar p_x p_y - i\hbar p_x p_y - i\hbar p_y p_x + 0 + 0 = 0 \quad \text{(verification of I.19)}
\]

\[
[H, L_z] = \left[ \frac{p^2}{2m} + V(r), L_z \right] = \frac{1}{2m} [p^2, L_z] + [V(r), L_z] = 0 + 0 = 0 \quad \text{(verification of I.20)}
\]

\[
[H, L^2] = [H, L_x^2 + L_y^2 + L_z^2] = [H, L_x^2] + [H, L_y^2] + [H, L_z^2] \\
= [H, L_x]L_x + L_x[H, L_x] + [H, L_y]L_y + L_y[H, L_y] + [H, L_z]L_z + L_z[H, L_z] \\
= 0 + 0 + 0 + 0 + 0 + 0 \quad \text{(verification of I.21)}
\]
Appendix J

Probability Current

The following is based on a lecture note of Professor Steven Carlip at UC Davis [Carlip, 2013].

Let $R$ be a region in $\mathbb{R}^3$ with a boundary $\partial R$, and consider a particle whose wavefunction is $\Psi(x,t)$. Then, the probability $P(R)$ that the particle is found in $R$ is given by

$$P(R) = \int_R \Psi^*(x,t)\Psi(x,t) \, dx dy dz. \quad (J.1)$$

If the particle is moving in the classical sense, or the magnitude of the wavefunction is changing with time in the quantum mechanical sense, $P(R)$ changes with time, and the first derivative is given by

$$\frac{d}{dt} P(R) = \int_R \frac{\partial}{\partial t} (\Psi(x,t)\Psi(x,t)) \, dx dy dz = \int_R \left( \frac{\partial \Psi^*}{\partial t} - \Psi^* \frac{\partial \Psi}{\partial t} \right) dx dy dz. \quad (J.2)$$

The time-dependent Schrödinger equation gives

$$i\hbar \frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \Psi + V\Psi \quad (J.3)$$

as well as its complex conjugate

$$-i\hbar \frac{\partial \Psi^*}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \Psi^* + V\Psi^*. \quad (J.4)$$

Therefore,

$$\frac{\partial \Psi^*}{\partial t} = \frac{1}{-i\hbar} \left( -\frac{\hbar^2}{2m} \nabla^2 \Psi^* + V\Psi^* \right) \Psi = \frac{\hbar}{i2m} \left( \nabla^2 \Psi^* \right) \Psi - \frac{V}{i\hbar} \Psi^* \Psi \quad (J.5)$$
and
\[
\frac{\Psi^* \partial \Psi}{\partial t} = \Psi^* \frac{1}{i \hbar} \left( -\frac{\hbar^2}{2m} \nabla^2 \Psi + V \Psi \right) = -\frac{\hbar}{i2m} \Psi^* \nabla^2 \Psi + \frac{V}{i \hbar} \Psi^* \Psi. \tag{J.6}
\]

So, the integrand of (J.2) is
\[
\frac{\partial \Psi^*}{\partial t} \Psi + \Psi^* \frac{\partial \Psi}{\partial t} = \frac{\hbar}{i2m} \left( \Psi \nabla^2 \Psi^* - \Psi^* \nabla^2 \Psi \right) = -\frac{i \hbar}{2m} \left( \Psi \nabla^2 \Psi^* - \Psi^* \nabla^2 \Psi \right). \tag{J.7}
\]

Now we need the following identity.

**Fact J.1** For differentiable functions \( f \) and \( g \) we have
\[
\nabla \cdot (f \nabla g - g \nabla f) = f \nabla^2 g - g \nabla^2 f. \tag{J.8}
\]

**Proof**

For simplicity, we will introduce the following notation, which is used widely in the physics literature.

\[
f_{x_i} := \frac{\partial f}{\partial x_i} \quad \text{and} \quad f_{x_i x_j} := \frac{\partial^2 f}{\partial x_j \partial x_i}; \quad \text{where} \quad x_i, x_j = x, y, \text{or} \ z \tag{J.9}
\]

Then,
\[
\nabla \cdot (f \nabla g - g \nabla f) = \nabla \cdot \left[ f \left( g_{x_i} + g_{y_j} + g_{z_k} \right) - g \left( f_{x_i} + f_{y_j} + f_{z_k} \right) \right]
\]
\[
= \left( \frac{i}{\partial x} + \frac{j}{\partial y} + \frac{k}{\partial z} \right) \cdot \left[ f \left( g_{x_i} + g_{y_j} + g_{z_k} \right) - g \left( f_{x_i} + f_{y_j} + f_{z_k} \right) \right]
\]
\[
= \frac{\partial}{\partial x} (fg_{x} - gf_{x}) + \frac{\partial}{\partial y} (fg_{y} - gf_{y}) + \frac{\partial}{\partial z} (fg_{z} - gf_{z})
\]
\[
= f_{x}g_{x} + fg_{xx} - g_{x}f_{x} - gf_{xx} + f_{y}g_{y} + fg_{yy} - g_{y}f_{y} - gf_{yy}
\]
\[
+ f_{z}g_{z} + fg_{zz} - g_{z}f_{z} - gf_{zz}
\]
\[
= f_{x}g_{xx} + fg_{yy} + f_{x}g_{x} + f_{y}g_{y} + g_{x}f_{x} + g_{y}f_{y} - gf_{xx} - gf_{yy}.
\]

\[
= f \nabla^2 g - g \nabla^2 f. \tag{J.10}
\]

Now, let \( f = \Psi \) and \( g = \Psi^* \) in (J.8) to obtain
\[
\Psi \nabla^2 \Psi^* - \Psi^* \nabla^2 \Psi = \nabla \cdot (\Psi \nabla \Psi^* - \Psi^* \nabla \Psi). \tag{J.11}
\]
If we define a vector $\mathbf{j}$ by

$$
\mathbf{j} := \frac{i\hbar}{2m} (\Psi \nabla \Psi^* - \Psi^* \nabla \Psi),
$$

(J.12)

then, from (J.11) we have

$$
\nabla \cdot \mathbf{j} = \frac{i\hbar}{2m} (\Psi \nabla^2 \Psi^* - \Psi^* \nabla^2 \Psi),
$$

(J.13)

and (J.2) becomes

$$
\frac{d}{dt}P(R) = \int_R \left( \frac{\partial \Psi^*}{\partial t} \Psi + \Psi^* \frac{\partial \Psi}{\partial t} \right) \, dxdydz
= -\int_R \frac{i\hbar}{2m} (\Psi \nabla^2 \Psi^* - \Psi^* \nabla^2 \Psi) \, dxdydz
= -\int_R \nabla \cdot \mathbf{j} \, dxdydz.
$$

(J.14)

Next, remember Stokes’ Theorem from multivariable calculus.

**Theorem J.1 (Stokes’ Theorem: the Divergence Theorem)**

$$
\int_R \nabla \cdot \mathbf{j} \, d^3x = \int_{\partial R} \mathbf{j} \cdot \hat{n} \, d^2x;
$$

(J.15)

where $\hat{n}$ is the unit normal to the boundary or surface $\partial R$.

We finally have

$$
\frac{d}{dt}P(R) = -\int_{\partial R} \mathbf{j} \cdot \hat{n} \, d^2x.
$$

(J.16)

What (J.16) is telling us is that the amount of $\mathbf{j}$ escaping through the surface $\partial R$ equals, in magnitude, the change in the amount of probability we have in the region $R$. This is why $\mathbf{j}$ is called the probability current.

**Definition J.1 (Probability Current)** Given a wavefunction $\Psi(x, t)$ we define the probability current $\mathbf{j}$ by

$$
\mathbf{j} := \frac{i\hbar}{2m} (\Psi \nabla \Psi^* - \Psi^* \nabla \Psi).
$$

(J.17)
In one dimension, this reduces to

\[ j := \frac{i\hbar}{2m} \left( \frac{\partial}{\partial x} \Psi^* - \Psi^* \frac{\partial}{\partial x} \Psi \right). \]  

(J.18)

In particular, when the potential, and hence the Hamiltonian, does not have explicit time dependence, \( j \) reduces to an expression in the spatial part of the wavefunction \( \psi(x) \) rather than the full wave function \( \Psi(x,t) \) such that

\[ j = \frac{i\hbar}{2m} \left( \psi \frac{\partial}{\partial x} \psi^* - \psi^* \frac{\partial}{\partial x} \psi \right). \]  

(J.19)

Note that this is completely analogous in form and substance to the continuity equation of fluid dynamics and electromagnetism.

\[ \frac{dQ}{dt} = -\int_{\partial R} j \cdot \hat{n} dS \quad (dS \text{ is the surface integral over } \partial R, \text{ the boundary of } R). \]  

(J.20)

In fluid dynamics, \( Q \) is the amount of the fluid, and \( j \) is the flow of the fluid. In electromagnetism, \( Q \) is the amount of electric charge, and \( j \) is the electric current. Finally, let us make a note of the fact that the equality does not hold if there is a “sink” or “source” in the region \( R \), which is the difference between the classical continuity equation and our quantum mechanical equation.

---

1 The derivation is straightforward as below.

\[ j := \frac{i\hbar}{2m} \left( \psi \frac{\partial}{\partial x} \Psi^* - \Psi^* \frac{\partial}{\partial x} \Psi \right) = \frac{i\hbar}{2m} \left( (\psi(x)e^{-i\omega t})^* \frac{\partial}{\partial x} (\psi(x)e^{-i\omega t}) \right) \]

\[ - (\psi(x)e^{-i\omega t})^* \frac{\partial}{\partial x} (\psi(x)e^{-i\omega t}) \]

\[ = \frac{i\hbar}{2m} \left( \psi(x)e^{-i\omega t} \frac{\partial}{\partial x} \psi(x)^* e^{i\omega t} - \psi(x)^* e^{i\omega t} \frac{\partial}{\partial x} \psi(x)e^{-i\omega t} \right) \]

\[ = \frac{i\hbar}{2m} \left( \psi \frac{\partial}{\partial x} \psi^* - \psi^* \frac{\partial}{\partial x} \psi \right) \]

2 This form is known as the integral form. Its differential counterpart, called the differential form quite appropriately, is given by

\[ \frac{\partial \rho(x,t)}{\partial t} = -\nabla \cdot j(x,t) \quad \text{or} \quad \frac{\partial \rho(x,t)}{\partial t} = -\frac{\partial}{\partial x} j(x,t) \quad \text{in one dimension;} \]

where \( \rho \) is the density.
Appendix K

Chain Rules in Partial Differentiation

K.1 One Independent Variable

Consider a function \( f = f(x, y, z) \), where \( x = x(t) \), \( y = y(t) \), and \( z = z(t) \). Then, we have

\[
\frac{df}{dt} = \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt} + \frac{\partial f}{\partial z} \frac{dz}{dt}.
\]  

(K.1)

K.2 Three Independent Variables

Consider a function \( f = f(x, y, z) \), where \( x = x(r, \theta, \phi) \), \( y = y(r, \theta, \phi) \), and \( z = z(r, \theta, \phi) \). Then, we have

\[
\frac{\partial f}{\partial r} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial r} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial r} + \frac{\partial f}{\partial z} \frac{\partial z}{\partial r},
\]

(K.2)

\[
\frac{\partial f}{\partial \theta} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial \theta} + \frac{\partial f}{\partial z} \frac{\partial z}{\partial \theta},
\]

(K.3)

and

\[
\frac{\partial f}{\partial \phi} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial \phi} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial \phi} + \frac{\partial f}{\partial z} \frac{\partial z}{\partial \phi}.
\]

(K.4)

Actually, the following operator forms may be more convenient in actual use.

\[
\frac{\partial}{\partial r} = \frac{\partial x}{\partial r} \frac{\partial}{\partial x} + \frac{\partial y}{\partial r} \frac{\partial}{\partial y} + \frac{\partial z}{\partial r} \frac{\partial}{\partial z}.
\]

(K.5)
\[
\frac{\partial}{\partial \theta} = \frac{\partial x}{\partial \theta} \frac{\partial}{\partial x} + \frac{\partial y}{\partial \theta} \frac{\partial}{\partial y} + \frac{\partial z}{\partial \theta} \frac{\partial}{\partial z} \tag{K.6}
\]

\[
\frac{\partial}{\partial \phi} = \frac{\partial x}{\partial \phi} \frac{\partial}{\partial x} + \frac{\partial y}{\partial \phi} \frac{\partial}{\partial y} + \frac{\partial z}{\partial \phi} \frac{\partial}{\partial z} \tag{K.7}
\]

Of course, we also have

\[
\frac{\partial f}{\partial x} = \frac{\partial f}{\partial r} \frac{\partial r}{\partial x} + \frac{\partial f}{\partial \theta} \frac{\partial \theta}{\partial x} + \frac{\partial f}{\partial \phi} \frac{\partial \phi}{\partial x}, \tag{K.8}
\]

\[
\frac{\partial f}{\partial y} = \frac{\partial f}{\partial r} \frac{\partial r}{\partial y} + \frac{\partial f}{\partial \theta} \frac{\partial \theta}{\partial y} + \frac{\partial f}{\partial \phi} \frac{\partial \phi}{\partial y}, \tag{K.9}
\]

and

\[
\frac{\partial f}{\partial z} = \frac{\partial f}{\partial r} \frac{\partial r}{\partial z} + \frac{\partial f}{\partial \theta} \frac{\partial \theta}{\partial z} + \frac{\partial f}{\partial \phi} \frac{\partial \phi}{\partial z}, \tag{K.10}
\]

or

\[
\frac{\partial}{\partial x} = \frac{\partial r}{\partial x} \frac{\partial}{\partial r} + \frac{\partial \theta}{\partial x} \frac{\partial}{\partial \theta} + \frac{\partial \phi}{\partial x} \frac{\partial}{\partial \phi}, \tag{K.11}
\]

\[
\frac{\partial}{\partial y} = \frac{\partial r}{\partial y} \frac{\partial}{\partial r} + \frac{\partial \theta}{\partial y} \frac{\partial}{\partial \theta} + \frac{\partial \phi}{\partial y} \frac{\partial}{\partial \phi}, \tag{K.12}
\]

and

\[
\frac{\partial}{\partial z} = \frac{\partial r}{\partial z} \frac{\partial}{\partial r} + \frac{\partial \theta}{\partial z} \frac{\partial}{\partial \theta} + \frac{\partial \phi}{\partial z} \frac{\partial}{\partial \phi}. \tag{K.13}
\]

What we do not have is the following.

**K.3 Reciprocal Relation is False**

Unlike the total derivative, we do not have

\[
\frac{\partial r}{\partial x} = \left(\frac{\partial x}{\partial r}\right)^{-1}. \quad \text{FALSE} \quad \tag{K.14}
\]

Instead, we generally have

\[
\frac{\partial r}{\partial x} \neq \left(\frac{\partial x}{\partial r}\right)^{-1}. \quad \text{TRUE} \quad \tag{K.15}
\]
This is easy to understand because $\frac{\partial r}{\partial x}$ means the derivative/change of $r$ with respect to $x$ where $y$ and $z$ are held constant, while $\frac{\partial x}{\partial r}$ is the derivative/change of $x$ with respect to $r$ where $\theta$ and $\phi$ are held fixed. An explicit computation follows.

\[
\left( \frac{\partial r}{\partial x} \right)_{y,z: \text{fixed}} = \left( \frac{\partial}{\partial x} \sqrt{x^2 + y^2 + z^2} \right)_{y,z: \text{fixed}} = \frac{1}{2} \frac{2x}{\sqrt{x^2 + y^2 + z^2}} = \frac{x}{r} = \sin \theta \cos \phi
\]  

(K.16)

\[
\left( \frac{\partial x}{\partial r} \right)_{\theta,\phi: \text{fixed}} = \left( \frac{\partial r \sin \theta \cos \phi}{\partial r} \right)_{\theta,\phi: \text{fixed}} = \sin \theta \cos \phi
\]  

(K.17)

Hence, the inequality (K.15) holds in this case. Needless to say, (K.15) is generally true and is not limited to spherical and Cartesian coordinates.

Compare (K.15) with the following example where we have total derivatives.

\[
\frac{dy}{dx} = 2x \quad \text{or} \quad x = \pm \sqrt{y}
\]  

(K.18)

\[
\frac{dx}{dy} = \pm \frac{1}{2 \sqrt{y}} = \pm \frac{1}{2} \pm \frac{1}{2} = \frac{1}{2x}
\]  

(K.19)

\[
\frac{dy}{dx} = \left( \frac{dx}{dy} \right)^{-1}
\]  

(K.20)

Note here that (K.20) really means

\[
\frac{dy}{dx} \bigg|_{x=x_0} = \left( \frac{dx}{dy} \bigg|_{y=y_0^2} \right)^{-1}.
\]  

(K.21)

See Theorem K.1 for a more general description.

### K.4 Inverse Function Theorems

The reciprocal relation (K.20) is nothing but a straightforward application of the Inverse Function Theorem.

**Theorem K.1 (Inverse Function Theorem: Single Variable)** If $f$ is a continuously differentiable function of $x$ with nonzero derivative at the point $a$, then, $f$ is invertible in a neighborhood of $a$, the inverse $f^{-1}$ is continuously differentiable, and

\[
(f^{-1})'(b) = \frac{1}{f'(a)},
\]  

(K.22)

where $b = f(a)$.
Theorem K.2 (Inverse Function Theorem: Multiple Variables) If the total derivative of a continuously differentiable multivariable function \( F \) defined on an open subset \( \mathcal{O} \) of \( \mathbb{R}^n \) and into \( \mathbb{R}^n \) is invertible at a point \( p \), that is, the Jacobian \( J \) of \( F \) at \( p \) is non-zero, then, \( F \) is an invertible function with the inverse \( F^{-1} \) near the point \( p \). This means that the inverse \( F^{-1} \) is defined in some neighborhood of \( F(p) \). Moreover, the inverse function \( F^{-1} \) is continuously differentiable. As for the Jacobian, we have

\[
J_{F^{-1}}(F(p)) = [J_F(p)]^{-1};
\]  

(K.23)

where \( J_G(q) \) is the Jacobian of the function \( G \) at the point \( q \), and \( [J_F(p)]^{-1} \) is the matrix inverse.
Appendix L

Laplacian Operator in Spherical Coordinates

The Laplacian operator in Cartesian coordinates is

\[ \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \]  (L.1)

We will use the following relations

\[ r = \sqrt{x^2 + y^2 + z^2} \]
\[ x = r \sin \theta \cos \phi \]
\[ y = r \sin \theta \sin \phi \]
\[ z = r \cos \theta \]
\[ \cos \theta = \frac{z}{\sqrt{x^2 + y^2 + z^2}} \]
\[ \sin \theta = \frac{\sqrt{x^2 + y^2}}{\sqrt{x^2 + y^2 + z^2}} \]
\[ \cos \phi = \frac{x}{\sqrt{x^2 + y^2}} \]
\[ \sin \phi = \frac{y}{\sqrt{x^2 + y^2}} , \]

along with the chain rules below from p.325 of Appendix K.

\[ \frac{\partial}{\partial x} = \frac{\partial r}{\partial x} \frac{\partial}{\partial r} + \frac{\partial \theta}{\partial x} \frac{\partial}{\partial \theta} + \frac{\partial \phi}{\partial x} \frac{\partial}{\partial \phi}, \]  (L.2)

\[ \frac{\partial}{\partial y} = \frac{\partial r}{\partial y} \frac{\partial}{\partial r} + \frac{\partial \theta}{\partial y} \frac{\partial}{\partial \theta} + \frac{\partial \phi}{\partial y} \frac{\partial}{\partial \phi}, \]  (L.3)

\[ \frac{\partial}{\partial z} = \frac{\partial r}{\partial z} \frac{\partial}{\partial r} + \frac{\partial \theta}{\partial z} \frac{\partial}{\partial \theta} + \frac{\partial \phi}{\partial z} \frac{\partial}{\partial \phi}. \]  (L.4)

\[ \frac{\partial}{\partial r} = \frac{\partial x}{\partial r} \frac{\partial}{\partial x} + \frac{\partial y}{\partial r} \frac{\partial}{\partial y} + \frac{\partial z}{\partial r} \frac{\partial}{\partial z}. \]  (L.5)
The partial derivatives of $r$ with respect to $x$, $y$, and $z$ are as follows.

\[
\frac{\partial r}{\partial x} = \frac{1}{2} \frac{2x}{\sqrt{x^2 + y^2 + z^2}} = \frac{x}{r} = \sin \theta \cos \phi \quad (L.8)
\]

\[
\frac{\partial r}{\partial y} = \frac{1}{2} \frac{2y}{\sqrt{x^2 + y^2 + z^2}} = \frac{y}{r} = \sin \theta \sin \phi \quad (L.9)
\]

\[
\frac{\partial r}{\partial z} = \frac{1}{2} \frac{2z}{\sqrt{x^2 + y^2 + z^2}} = \frac{z}{r} = \cos \theta \quad (L.10)
\]

Next, we will use $z = r \cos \theta$ to compute $\frac{\partial \theta}{\partial x}$, $\frac{\partial \theta}{\partial y}$, and $\frac{\partial \theta}{\partial z}$.  

\[
\frac{\partial z}{\partial x} = \frac{\partial r}{\partial x} \cos \theta + r \frac{\partial \cos \theta}{\partial x}
\]

\[
= \sin \theta \cos \phi \cos \theta + r \left( \frac{\partial r}{\partial x} \frac{\partial \partial \theta}{\partial r} + \frac{\partial \theta}{\partial x} \frac{\partial \phi}{\partial \theta} + \frac{\partial \phi}{\partial x} \frac{\partial \phi}{\partial \phi} \right) (\cos \theta)
\]

\[
= \sin \theta \cos \phi \cos \theta + r \frac{\partial \theta}{\partial x} (- \sin \theta) \implies 0 = \sin \theta \cos \phi \cos \theta - r \sin \theta \frac{\partial \theta}{\partial x}
\]

\[
\implies r \sin \theta \frac{\partial \theta}{\partial x} = \sin \theta \cos \phi \cos \theta
\]

\[1\] There is a simpler, but possibly less comfortable way to compute $\frac{\partial \theta}{\partial x}$, $\frac{\partial \theta}{\partial y}$, and $\frac{\partial \theta}{\partial z}$ using the equalities

\[
\cos \theta = \frac{z}{r} \quad \text{and} \quad \sin \theta = \frac{\sqrt{x^2 + y^2}}{\sqrt{x^2 + y^2 + z^2}}.
\]

Namely,

\[
\cos \theta = \frac{z}{r} \implies - \sin \theta d\theta = \frac{z(-x)}{r^3} \implies \frac{\partial \theta}{\partial x} = \frac{\cos \theta \cos \phi}{r},
\]

\[
\cos \theta = \frac{z}{r} \implies - \sin \theta d\theta = \frac{z(-y)}{r^3} \implies \frac{\partial \theta}{\partial y} = \frac{\cos \theta \sin \phi}{r},
\]

and

\[
\sin \theta = \frac{\sqrt{x^2 + y^2}}{\sqrt{x^2 + y^2 + z^2}} \implies \cos \theta d\theta = \frac{\sqrt{x^2 + y^2}(-z)}{r^3} \implies \frac{\partial \theta}{\partial z} = - \frac{\sin \theta}{r}.
\]
\[ \frac{\partial \theta}{\partial x} = \frac{\cos \theta \cos \phi}{r} \]  
\( (L.11) \)

\[ \frac{\partial z}{\partial y} = \frac{\partial r}{\partial y} \cos \theta + r \frac{\partial \cos \theta}{\partial y} = \sin \theta \sin \phi \cos \theta + r \frac{\partial \theta}{\partial y} \cos \theta \frac{\partial}{\partial y} \cos \theta \]  
\[ \Rightarrow 0 = \sin \theta \sin \phi \cos \theta - r \frac{\partial \theta}{\partial y} \]  
\[ \Rightarrow r \sin \theta \frac{\partial \theta}{\partial y} = \sin \theta \sin \phi \cos \theta \]

\[ \Rightarrow \frac{\partial \theta}{\partial y} = \frac{\cos \theta \sin \phi}{r} \]  
\( (L.12) \)

\[ \frac{\partial z}{\partial z} = \frac{\partial r}{\partial z} \cos \theta + r \frac{\partial \cos \theta}{\partial z} = \cos \theta \cos \theta + r \frac{\partial \theta}{\partial z} \cos \theta + \frac{\partial \phi}{\partial z} \cos \theta \]  
\[ \Rightarrow r \sin \theta \frac{\partial \theta}{\partial z} = -\sin^2 \theta \]

\[ \Rightarrow \frac{\partial \theta}{\partial z} = -\frac{\sin \theta}{r} \]  
\( (L.13) \)

Now, use \( x = r \sin \theta \cos \phi \) to compute \( \frac{\partial \phi}{\partial x}, \frac{\partial \phi}{\partial y}, \) and \( \frac{\partial \phi}{\partial z} \)^2

\[ \frac{\partial x}{\partial x} = \left( \frac{\partial r}{\partial x} \frac{\partial}{\partial r} + \frac{\partial \theta}{\partial x} \frac{\partial}{\partial \theta} + \frac{\partial \phi}{\partial x} \frac{\partial}{\partial \phi} \right) \left( r \sin \theta \cos \phi \right) \]

\( ^2 \) There is an alternative way using the relations

\[ \cos \phi = \frac{x}{\sqrt{x^2 + y^2}} \quad \text{and} \quad \sin \phi = \frac{y}{\sqrt{x^2 + y^2}}. \]

Namely,

\[ \cos \phi = \frac{x}{\sqrt{x^2 + y^2}} \implies -\sin \phi d\phi = \left( \frac{1}{\sqrt{x^2 + y^2}} + x \left( \frac{-1}{2} \right) \frac{1}{(x^2 + y^2)^{3/2}} (2x) \right) dx \]

\[ = \left( \frac{1}{\sqrt{x^2 + y^2}} - \frac{x^2}{(x^2 + y^2)^{3/2}} \right) dx = \frac{x^2 + y^2 - x^2}{(x^2 + y^2)^{3/2}} dx = \frac{y^2}{(x^2 + y^2)^{3/2}} dx \]
\[= \sin \theta \cos \phi \sin \theta \cos \phi + \frac{\cos \theta \cos \phi}{r} r \cos \theta \cos \phi - r \sin \theta \sin \phi \frac{\partial \phi}{\partial x}\]
\[= \sin^2 \theta \cos^2 \phi + \cos^2 \theta \cos^2 \phi - r \sin \theta \sin \phi \frac{\partial \phi}{\partial x}\]
\[= \cos^2 \phi - r \sin \theta \sin \phi \frac{\partial \phi}{\partial x}\]
\[\implies 1 = \cos^2 \phi - r \sin \theta \sin \phi \frac{\partial \phi}{\partial x} \implies r \sin \theta \sin \phi \frac{\partial \phi}{\partial x} = -\sin^2 \phi\]
\[\implies \frac{\partial \phi}{\partial x} = -\frac{\sin \phi}{r \sin \theta} \tag{L.14}\]

\[\frac{\partial x}{\partial y} = \left( \frac{\partial r}{\partial y} \frac{\partial}{\partial r} + \frac{\partial \theta}{\partial y} \frac{\partial}{\partial \theta} + \frac{\partial \phi}{\partial y} \frac{\partial}{\partial \phi} \right) (r \sin \theta \cos \phi)\]
\[= \sin \theta \sin \theta \cos \phi \cos \phi + \frac{\cos \theta \sin \phi}{r} r \cos \theta \cos \phi - r \sin \theta \sin \phi \frac{\partial \phi}{\partial y}\]
\[= \sin^2 \theta \sin \phi \cos \phi + \cos^2 \theta \sin \phi \cos \phi - r \sin \theta \sin \phi \frac{\partial \phi}{\partial y}\]
\[= \sin \phi \cos \phi - r \sin \theta \sin \phi \frac{\partial \phi}{\partial y}\]
\[\implies 0 = \sin \phi \cos \phi - r \sin \theta \sin \phi \frac{\partial \phi}{\partial y} \implies r \sin \theta \sin \phi \frac{\partial \phi}{\partial y} = \sin \phi \cos \phi\]
\[\implies \frac{\partial \phi}{\partial y} = \frac{\cos \phi}{r \sin \theta} \tag{L.15}\]

\[\frac{\partial x}{\partial z} = \left( \frac{\partial r}{\partial z} \frac{\partial}{\partial r} + \frac{\partial \theta}{\partial z} \frac{\partial}{\partial \theta} + \frac{\partial \phi}{\partial z} \frac{\partial}{\partial \phi} \right) (r \sin \theta \cos \phi)\]
\[= \frac{(r \sin \theta \sin \phi)^2}{(r \sin \theta)^3} dx \implies \frac{\sin^2 \phi}{r \sin \theta} dx \implies \frac{\sin^2 \phi}{r \sin \theta} \frac{1}{(- \sin \phi)} = -\frac{\sin \phi}{r \sin \theta},\]

and
\[
\sin \phi = \frac{y}{\sqrt{x^2 + y^2}} \implies \cos \phi \, d\phi = \left( \frac{1}{\sqrt{x^2 + y^2}} + y \left( \frac{1}{2} \right) \frac{1}{(x^2 + y^2)^{3/2}} \right) dy
\]
\[
= \left( \frac{1}{\sqrt{x^2 + y^2}} - \frac{y^2}{(x^2 + y^2)^{3/2}} \right) dy = \frac{x^2 + y^2 - y^2}{(x^2 + y^2)^{3/2}} dy = \frac{x^2}{(x^2 + y^2)^{3/2}} dy
\]
\[
= \frac{(r \sin \theta \cos \phi)^2}{(r \sin \theta)^3} dy = \frac{\cos^2 \phi}{r \sin \theta} dy \implies \frac{\partial \phi}{\partial y} = \frac{\cos^2 \phi}{r \sin \theta \cos \phi} = \frac{\cos \phi}{r \sin \theta}.\]
\[
= \cos \theta \sin \theta \cos \phi - \frac{\sin \theta}{r} r \cos \theta \cos \phi - r \sin \theta \sin \phi \frac{\partial \phi}{\partial z} \\
= -r \sin \theta \sin \phi \frac{\partial \phi}{\partial z} \\
\Rightarrow \frac{\partial \phi}{\partial z} = 0
\] (L.16)

Incidentally, we already knew that \( \frac{\partial \phi}{\partial z} = 0 \) from geometric considerations; that is, \( \phi \) is determined completely by the \( x \)- and \( y \)-coordinates.

Let us take a deep breath at this point and summarize what we have found so far.

\[
\begin{align*}
\frac{\partial r}{\partial x} &= \sin \theta \cos \phi \\
\frac{\partial r}{\partial y} &= \sin \theta \sin \phi \\
\frac{\partial r}{\partial z} &= \cos \theta \\
\frac{\partial \theta}{\partial x} &= \cos \theta \cos \phi \frac{r}{r} \\
\frac{\partial \theta}{\partial y} &= \cos \theta \sin \phi \frac{r}{r} \\
\frac{\partial \theta}{\partial z} &= -\sin \theta \frac{r}{r} \\
\frac{\partial \phi}{\partial x} &= -\sin \phi \frac{r}{r} \\
\frac{\partial \phi}{\partial y} &= \cos \phi \frac{r}{r} \\
\frac{\partial \phi}{\partial z} &= 0
\end{align*}
\] (L.17)

Our next step is to compute \( \frac{\partial}{\partial x} \), \( \frac{\partial}{\partial y} \), and \( \frac{\partial}{\partial z} \).

\[
\begin{align*}
\frac{\partial}{\partial x} &= \frac{\partial r}{\partial x} \frac{\partial}{\partial r} + \frac{\partial \theta}{\partial x} \frac{\partial}{\partial \theta} + \frac{\partial \phi}{\partial x} \frac{\partial}{\partial \phi} \\
&= \sin \theta \cos \phi \frac{\partial}{\partial r} + \cos \theta \cos \phi \frac{\partial}{\partial \theta} - \sin \phi \frac{\partial}{\partial \phi}
\end{align*}
\] (L.18)

\[
\begin{align*}
\frac{\partial}{\partial y} &= \frac{\partial r}{\partial y} \frac{\partial}{\partial r} + \frac{\partial \theta}{\partial y} \frac{\partial}{\partial \theta} + \frac{\partial \phi}{\partial y} \frac{\partial}{\partial \phi} \\
&= \sin \theta \sin \phi \frac{\partial}{\partial r} + \cos \theta \sin \phi \frac{\partial}{\partial \theta} + \cos \phi \frac{\partial}{\partial \phi}
\end{align*}
\] (L.19)

\[
\begin{align*}
\frac{\partial}{\partial z} &= \cos \theta \frac{\partial}{\partial r} - \frac{\sin \theta}{r} \frac{\partial}{\partial \theta} + 0 \cdot \frac{\partial}{\partial \phi} = \cos \theta \frac{\partial}{\partial r} - \frac{\sin \theta}{r} \frac{\partial}{\partial \theta}
\end{align*}
\] (L.20)

We are ready to compute \( \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} = \frac{\partial}{\partial x} \frac{\partial}{\partial x} + \frac{\partial}{\partial y} \frac{\partial}{\partial y} + \frac{\partial}{\partial z} \frac{\partial}{\partial z} \).

\[
\begin{align*}
\frac{\partial^2}{\partial x^2} &= \frac{\partial}{\partial x} \frac{\partial}{\partial x} = \left( \sin \theta \cos \phi \frac{\partial}{\partial r} + \cos \theta \cos \phi \frac{\partial}{\partial \theta} - \frac{\sin \phi}{r \sin \theta} \frac{\partial}{\partial \phi} \right) \cdot \\
&\quad \left( \sin \theta \cos \phi \frac{\partial}{\partial r} + \cos \theta \cos \phi \frac{\partial}{\partial \theta} - \frac{\sin \phi}{r \sin \theta} \frac{\partial}{\partial \phi} \right) \\
&= \left( \sin \theta \cos \phi \frac{\partial}{\partial r} \right) \left( \sin \theta \cos \phi \frac{\partial}{\partial r} \right) + \left( \sin \theta \cos \phi \frac{\partial}{\partial r} \right) \left( \cos \theta \cos \phi \frac{\partial}{\partial \theta} \right)
\end{align*}
\]
\[-\left(\sin \theta \cos \phi \frac{\partial}{\partial r}\right) \left(\frac{\sin \phi}{r \sin \theta} \frac{\partial}{\partial \phi}\right) + \left(\cos \theta \cos \phi \frac{\partial}{\partial \theta}\right) \left(\sin \theta \cos \phi \frac{\partial}{\partial r}\right) + \left(\cos \theta \cos \phi \frac{\partial}{\partial \theta}\right) \left(\cos \theta \cos \phi \frac{\partial}{\partial r}\right)\]
\[-\left(\cos \theta \sin \phi \cos \phi \frac{\partial}{\partial \theta}\right) \left(\frac{\sin \phi}{r \sin \theta} \frac{\partial}{\partial \phi}\right) - \left(\cos \theta \sin \phi \cos \phi \frac{\partial}{\partial \theta}\right) \left(\frac{\cos \theta \cos \phi}{r} \frac{\partial}{\partial \theta}\right)\]
\[+ \left(\cos \theta \sin \phi \cos \phi \frac{\partial}{\partial \phi}\right) \left(\sin \phi \frac{\partial}{\partial \phi}\right)\]
\[= \sin^2 \theta \cos^2 \phi \frac{\partial^2}{\partial r^2} + \sin \theta \cos \theta \cos^2 \phi \frac{1}{r} \frac{\partial}{\partial \theta} \left(\frac{1}{r \partial \theta}\right)\]
\[-\sin \phi \cos \phi \frac{\partial}{\partial r} \left(\frac{1}{r \partial \phi}\right) + \frac{\cos \theta \cos^2 \phi}{r} \frac{\partial}{\partial \theta} \left(\sin \frac{\partial}{\partial \theta}\right) + \frac{\cos \theta \cos^2 \phi}{r} \frac{\partial}{\partial \theta} \left(\cos \theta \frac{\partial}{\partial \theta}\right)\]
\[-\cos \theta \sin \phi \cos \phi \frac{1}{r^2} \frac{\partial}{\partial \theta} \left(\cos \phi \frac{\partial}{\partial \phi}\right) - \frac{\cos \theta \sin \phi \cos \phi}{r^2 \sin \theta} \frac{\partial}{\partial \phi} \left(\cos \phi \frac{\partial}{\partial \theta}\right)\]
\[+ \frac{\sin \phi}{r^2 \sin^2 \theta} \frac{\partial}{\partial \phi} \left(\sin \phi \frac{\partial}{\partial \phi}\right)\]
\[= \sin^2 \theta \cos^2 \phi \frac{\partial^2}{\partial r^2} + \sin \theta \cos \theta \cos^2 \phi \left[-\frac{1}{r^2} \frac{\partial^2}{\partial \theta^2} + \frac{1}{r} \frac{\partial^2}{\partial r \partial \theta}\right]\]
\[-\sin \phi \cos \phi \left[-\frac{1}{r^2} \frac{\partial^2}{\partial \phi^2} + \frac{1}{r} \frac{\partial^2}{\partial r \partial \phi}\right]\]
\[+ \frac{\cos \theta \cos^2 \phi}{r} \left[\cos \theta \frac{\partial}{\partial r} + \sin \theta \frac{\partial^2}{\partial r \partial \theta}\right]\]
\[+ \frac{\cos \theta \cos^2 \phi}{r^2} \left[-\sin \theta \frac{\partial}{\partial \theta} + \cos \theta \frac{\partial^2}{\partial \theta^2}\right]\]
\[-\frac{\cos \theta \sin \phi \cos \phi}{r^2 \sin \theta} \left[-\cos \theta \frac{\partial}{\sin \theta \partial \phi} + \frac{\partial^2}{\partial \theta \partial \phi}\right]\]
\[-\frac{\sin \phi}{r} \left[ -\sin \phi \frac{\partial}{\partial r} + \cos \phi \frac{\partial^2}{\partial r \partial \phi} \right] \]
\[-\frac{\cos \theta \sin \phi}{r^2 \sin \theta} \left[ -\sin \phi \frac{\partial}{\partial \theta} + \cos \phi \frac{\partial^2}{\partial \theta \partial \phi} \right] \]
\[+ \frac{\sin \phi}{r^2 \sin^2 \theta} \left[ \cos \phi \frac{\partial}{\partial \phi} + \sin \phi \frac{\partial^2}{\partial \phi^2} \right] \]
\[= \sin^2 \theta \cos^2 \phi \frac{\partial^2}{\partial r^2} + \frac{\cos^2 \theta \cos^2 \phi}{r^2} \frac{\partial^2}{\partial \theta^2} + \frac{\sin^2 \phi}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \theta \partial \phi^2} + \frac{\cos^2 \theta \cos^2 \phi + \sin^2 \phi}{r} \frac{\partial}{\partial r} \]
\[+ \sin \theta \cos \theta \cos^2 \phi \left[ -\frac{1}{r^2} \frac{\partial}{\partial \theta} + \frac{1}{r} \frac{\partial^2}{\partial r \partial \theta} \right] - \sin \phi \cos \phi \left[ -\frac{1}{r^2} \frac{\partial}{\partial \phi} + \frac{1}{r} \frac{\partial^2}{\partial r \partial \phi} \right] \]
\[+ \frac{\cos \theta \cos^2 \phi}{r} \left[ + \sin \theta \frac{\partial^2}{\partial r \partial \theta} \right] \]
\[+ \frac{\cos \theta \cos^2 \phi}{r^2} \left[ - \sin \phi \frac{\partial}{\partial \theta} \right] \]
\[- \frac{\cos \theta \sin \phi \cos \phi}{r^2 \sin \theta} \left[ - \cos \phi \frac{\partial}{\partial \phi} + \frac{\partial^2}{\partial \theta \partial \phi} \right] \]
\[- \frac{\sin \phi}{r} \left[ + \cos \phi \frac{\partial^2}{\partial r \partial \phi} \right] \]
\[- \frac{\cos \theta \sin \phi}{r^2 \sin \theta} \left[ - \sin \phi \frac{\partial}{\partial \theta} + \cos \phi \frac{\partial^2}{\partial \theta \partial \phi} \right] \]
\[+ \frac{\sin \phi}{r^2 \sin^2 \theta} \left[ \cos \phi \frac{\partial}{\partial \phi} \right] \]

\[
\frac{\partial^2}{\partial y^2} = \frac{\partial}{\partial y} \frac{\partial}{\partial y} = \left( \sin \theta \sin \phi \frac{\partial}{\partial r} + \frac{\cos \theta \sin \phi}{r} \frac{\partial}{\partial \theta} + \frac{\cos \phi}{r \sin \theta} \frac{\partial}{\partial \phi} \right) \cdot \]
\[
\left( \sin \theta \sin \phi \frac{\partial}{\partial r} + \frac{\cos \theta \sin \phi}{r} \frac{\partial}{\partial \theta} + \frac{\cos \phi}{r \sin \theta} \frac{\partial}{\partial \phi} \right) \]
\[= \left( \sin \theta \sin \phi \frac{\partial}{\partial r} \right) \left( \sin \theta \sin \phi \frac{\partial}{\partial r} \right) + \left( \sin \theta \sin \phi \frac{\partial}{\partial r} \right) \left( \cos \phi \frac{\partial}{\partial \phi} \right) \]
\[+ \left( \sin \theta \sin \phi \frac{\partial}{\partial r} \right) \left( \cos \phi \frac{\partial}{\partial \phi} \right) \left( \frac{\cos \theta \sin \phi}{r} \frac{\partial}{\partial \theta} \right) \]
\[+ \left( \frac{\cos \theta \sin \phi}{r} \frac{\partial}{\partial \theta} \right) \left( \sin \theta \sin \phi \frac{\partial}{\partial r} \right) + \left( \frac{\cos \theta \sin \phi}{r} \frac{\partial}{\partial \theta} \right) \left( \cos \phi \frac{\partial}{\partial \phi} \right) \]
\[+ \left( \frac{\cos \theta \sin \phi}{r} \frac{\partial}{\partial \theta} \right) \left( \cos \phi \frac{\partial}{\partial \phi} \right) \]
\[ + \left( \frac{\cos \theta \sin \phi}{r} \frac{\partial}{\partial \theta} \right) \left( \frac{\cos \phi}{r \sin \theta} \frac{\partial}{\partial \phi} \right) \\
+ \left( \frac{\cos \phi}{r \sin \theta} \frac{\partial}{\partial \phi} \right) \left( \sin \theta \sin \phi \frac{\partial}{\partial r} \right) + \left( \frac{\cos \phi}{r \sin \theta} \frac{\partial}{\partial \phi} \right) \left( \frac{\cos \theta \sin \phi}{r} \frac{\partial}{\partial \theta} \right) \\
+ \left( \frac{\cos \phi}{r \sin \theta} \frac{\partial}{\partial \phi} \right) \left( \frac{\cos \phi}{r \sin \theta} \frac{\partial}{\partial \phi} \right) \\
= \sin^2 \theta \sin^2 \phi \frac{\partial^2}{\partial r^2} + \sin \theta \cos \theta \sin^2 \phi \frac{\partial}{\partial r} \left( \frac{1}{r} \frac{\partial}{\partial \theta} \right) \\
+ \sin \phi \cos \phi \frac{\partial}{\partial r} \left( \frac{1}{r} \frac{\partial}{\partial \phi} \right) \\
+ \frac{\cos \theta \sin^2 \phi}{r} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial r} \right) + \frac{\cos \theta \sin^2 \phi}{r^2} \frac{\partial}{\partial \theta} \left( \cos \theta \frac{\partial}{\partial \theta} \right) \\
+ \frac{\cos \theta \sin \phi \cos \phi}{r^2 \sin^2 \theta} \frac{\partial}{\partial \phi} \left( \frac{1}{\sin \theta} \frac{\partial}{\partial \phi} \right) \\
+ \frac{\cos \phi}{r^2 \sin^2 \theta} \frac{\partial}{\partial \phi} \left( \cos \frac{\partial}{\partial \phi} \right) \\
= \sin^2 \theta \sin^2 \phi \frac{\partial^2}{\partial r^2} + \sin \theta \cos \theta \sin^2 \phi \left[ - \frac{1}{r^2} \frac{\partial}{\partial \theta} + \frac{1}{r} \frac{\partial^2}{\partial r \partial \theta} \right] \\
+ \sin \phi \cos \phi \left[ - \frac{1}{r^2} \frac{\partial}{\partial \phi} + \frac{1}{r} \frac{\partial^2}{\partial r \partial \phi} \right] \\
+ \frac{\cos \theta \sin^2 \phi}{r} \left[ \cos \theta \frac{\partial}{\partial r} + \sin \theta \frac{\partial^2}{\partial r \partial \theta} \right] \\
+ \frac{\cos \theta \sin^2 \phi}{r^2} \left[ - \sin \theta \frac{\partial}{\partial \theta} + \cos \theta \frac{\partial^2}{\partial \theta^2} \right] \\
+ \frac{\cos \theta \sin \phi \cos \phi}{r^2 \sin \theta} \left[ - \frac{\cos \theta}{\sin \theta} \frac{\partial}{\partial \phi} + \frac{\partial^2}{\partial \theta \partial \phi} \right] \\
+ \frac{\cos \phi}{r} \left[ \cos \frac{\partial}{\partial r} + \sin \frac{\partial^2}{\partial r \partial \phi} \right] \\
+ \frac{\cos \theta \cos \phi}{r^2 \sin \theta} \left[ \cos \frac{\partial}{\partial \theta} + \sin \phi \frac{\partial^2}{\partial \theta \partial \phi} \right] \\
+ \frac{\cos \theta \cos \phi}{r^2 \sin \theta} \left[ \cos \frac{\partial}{\partial \theta} + \sin \phi \frac{\partial^2}{\partial \theta \partial \phi} \right]
\[ \frac{\partial^2}{\partial z^2} = \frac{\partial}{\partial z} \frac{\partial}{\partial z} = \left( \cos \theta \frac{\partial}{\partial r} - \frac{\sin \theta}{r} \frac{\partial}{\partial \theta} \right) \left( \cos \theta \frac{\partial}{\partial r} - \frac{\sin \theta}{r} \frac{\partial}{\partial \theta} \right) \]

\[ + \left( \frac{\sin \theta}{r} \frac{\partial}{\partial \theta} \right) \left( \frac{\sin \theta}{r} \frac{\partial}{\partial \theta} \right) \]

\[ = \cos^2 \theta \frac{\partial^2}{\partial r^2} - \sin \theta \cos \theta \left( \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2} + \frac{1}{r} \frac{\partial^2}{\partial r \partial \theta} \right) \]

\[ + \frac{\sin \theta}{r^2} \left[ \cos^2 \theta \frac{\partial}{\partial \theta} + \sin^2 \theta \frac{\partial^2}{\partial \theta^2} \right] \]

(L.23)

Finally, we can compute \( \nabla^2 \), combining (L.21), (L.22), and (L.23). There are some nice cancellations and simplifications. But, we have to be methodical and systematic in deciding on our computational strategy, that is, which terms to combine in what order.

\[ \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \]

\[ = \left[ \sin^2 \theta \left( \sin^2 \phi + \cos^2 \phi \right) + \cos^2 \theta \right] \frac{\partial^2}{\partial r^2} \]

\[ + \left[ \sin \theta \cos \theta \left( \sin^2 \phi + \cos^2 \phi \right) - \sin \theta \cos \theta \right] \left[ -\frac{1}{r^2} \frac{\partial}{\partial \theta} + \frac{1}{r} \frac{\partial^2}{\partial r \partial \theta} \right] \]

\[ + (\sin \phi \cos \phi - \sin \phi \cos \phi) \left[ -\frac{1}{r^2} \frac{\partial}{\partial \phi} + \frac{1}{r} \frac{\partial^2}{\partial r \partial \phi} \right] \]

\[ + \frac{\cos \theta (\sin^2 \phi + \cos^2 \phi)}{r} \left[ \cos \theta \frac{\partial}{\partial r} + \sin \theta \frac{\partial^2}{\partial r \partial \theta} \right] \]

\[ + \frac{\cos \theta (\sin^2 \phi + \cos^2 \phi)}{r^2} \left[ -\sin \theta \frac{\partial}{\partial \theta} + \cos \theta \frac{\partial^2}{\partial \theta^2} \right] \]
\[ + \left( \frac{\cos \theta \sin \phi \cos \phi}{r^2 \sin \theta} - \frac{\cos \theta \sin \phi \cos \phi}{r^2 \sin \theta} \right) \left[ - \frac{\cos \theta}{\sin \theta} \frac{\partial}{\partial \phi} + \frac{\partial^2}{\partial \theta \partial \phi} \right] \]

The following three terms derive from \( \frac{\partial^2}{\partial x^2} \).

\[ - \frac{\sin \phi}{r} \left[ - \sin \phi \frac{\partial}{\partial r} + \cos \phi \frac{\partial^2}{\partial r \partial \phi} \right] \]

\[ - \frac{\cos \theta \sin \phi}{r^2 \sin \theta} \left[ - \sin \phi \frac{\partial}{\partial \theta} + \cos \phi \frac{\partial^2}{\partial \theta \partial \phi} \right] \]

\[ + \frac{\sin \phi}{r^2 \sin^2 \theta} \left[ \cos \phi \frac{\partial}{\partial \phi} + \sin \phi \frac{\partial^2}{\partial \phi^2} \right] \]

The following three terms derive from \( \frac{\partial^2}{\partial y^2} \).

\[ + \frac{\cos \phi}{r} \left[ \cos \phi \frac{\partial}{\partial r} + \sin \phi \frac{\partial^2}{\partial r \partial \phi} \right] \]

\[ + \frac{\cos \theta \cos \phi}{r^2 \sin \theta} \left[ \cos \phi \frac{\partial}{\partial \theta} + \sin \phi \frac{\partial^2}{\partial \theta \partial \phi} \right] \]

\[ + \frac{\cos \phi}{r^2 \sin^2 \theta} \left[ - \sin \phi \frac{\partial}{\partial \phi} + \cos \phi \frac{\partial^2}{\partial \phi^2} \right] \]

The following two terms derive from \( \frac{\partial^2}{\partial z^2} \).

\[ - \frac{\sin \theta}{r} \left[ - \sin \theta \frac{\partial}{\partial r} + \cos \theta \frac{\partial^2}{\partial r \partial \theta} \right] \]

\[ + \frac{\sin \theta}{r^2} \left[ \cos \theta \frac{\partial}{\partial \theta} + \sin \theta \frac{\partial^2}{\partial \theta^2} \right] \]

\[ = \frac{\partial^2}{\partial r^2} + \frac{\cos \theta}{r} \left[ \cos \theta \frac{\partial}{\partial r} + \sin \theta \frac{\partial^2}{\partial r \partial \theta} \right] + \frac{\cos \theta}{r^2} \left[ - \sin \theta \frac{\partial}{\partial \theta} + \cos \theta \frac{\partial^2}{\partial \theta^2} \right] \]

\[ + \left( \sin^2 \phi + \cos^2 \phi + \sin^2 \theta \right) \frac{1}{r} \frac{\partial}{\partial r} + \left( \frac{\cos \theta}{r^2 \sin \theta} \right) \left( \sin^2 \phi + \cos^2 \phi \right) \frac{\partial}{\partial \theta} + \frac{\sin \theta \cos \theta}{r^2} \frac{\partial}{\partial \theta} \]

\[ + \frac{1}{r^2 \sin^2 \theta} \left( \sin \phi \cos \phi - \sin \phi \cos \phi \right) \frac{\partial}{\partial \phi} + \frac{\sin^2 \theta}{r^2} \frac{\partial^2}{\partial \theta^2} - \frac{\sin \theta \cos \theta}{r} \frac{\partial^2}{\partial r \partial \theta} \]

\[ + \frac{\sin \phi \cos \phi - \sin \phi \cos \phi}{r} \frac{\partial^2}{\partial r \partial \phi} + \frac{\cos \theta}{r^2 \sin \theta} \left( \sin \phi \cos \phi - \sin \phi \cos \phi \right) \frac{\partial^2}{\partial \theta \partial \phi} \]

\[ + \frac{\sin^2 \phi + \cos^2 \phi}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \]

\[ = \frac{\partial^2}{\partial r^2} + \frac{\cos \theta}{r} \left[ \cos \theta \frac{\partial}{\partial r} + \sin \theta \frac{\partial^2}{\partial r \partial \theta} \right] + \frac{\cos \theta}{r^2} \left[ - \sin \theta \frac{\partial}{\partial \theta} + \cos \theta \frac{\partial^2}{\partial \theta^2} \right] \]
\[
\begin{align*}
&+ \left( 1 + \sin^2 \theta \right) \frac{1}{r} \frac{\partial}{\partial r} + \left( \frac{\cos \theta}{r^2 \sin \theta} \right) \frac{\partial}{\partial \theta} + \frac{\sin \theta \cos \theta}{r^2} \frac{\partial}{\partial \theta} \\
&+ \frac{\sin^2 \theta}{r^2} \frac{\partial^2}{\partial \theta^2} - \frac{\sin \theta \cos \theta}{r} \frac{\partial}{\partial r \partial \theta} \\
&+ \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \\
&= \frac{\partial^2}{\partial r^2} + \frac{\cos \theta}{r} \left[ \cos \theta \frac{\partial}{\partial r} + \sin \theta \frac{\partial^2}{\partial r \partial \theta} - \sin \theta \frac{\partial^2}{\partial \theta \partial r} \right] \\
&+ \frac{\cos \theta}{r^2} \left[ -\sin \theta \frac{\partial}{\partial \theta} + \cos \theta \frac{\partial^2}{\partial \theta^2} + \sin \theta \frac{\partial}{\partial \theta} \right] \\
&+ \left( 1 + \sin^2 \theta \right) \frac{1}{r} \frac{\partial}{\partial r} + \left( \frac{\cos \theta}{r^2 \sin \theta} \right) \frac{\partial}{\partial \theta} \\
&+ \frac{\sin^2 \theta}{r^2} \frac{\partial^2}{\partial \theta^2} + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \\
&= \frac{\partial^2}{\partial r^2} + \frac{\cos \theta}{r^2} \left[ \cos \theta \frac{\partial^2}{\partial \theta^2} \right] + \left( 1 + \sin^2 \theta \cos^2 \theta \right) \frac{1}{r} \frac{\partial}{\partial r} + \left( \frac{\cos \theta}{r^2 \sin \theta} \right) \frac{\partial}{\partial \theta} \\
&+ \frac{\sin^2 \theta}{r^2} \frac{\partial^2}{\partial \theta^2} + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \\
&= \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} + \left( \frac{\cos \theta}{r^2 \sin \theta} \right) \frac{\partial}{\partial \theta} + \frac{\sin^2 \theta + \cos^2 \theta}{r^2} \frac{\partial^2}{\partial \theta^2} + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \\
&= \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} + \left( \frac{\cos \theta}{r^2 \sin \theta} \right) \frac{\partial}{\partial \theta} + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2} + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2}. \quad \text{(L.24)}
\end{align*}
\]

Therefore, we have shown

\[
\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} + \left( \frac{\cos \theta}{r^2 \sin \theta} \right) \frac{\partial}{\partial \theta} + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2} + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2}. \quad \text{(L.25)}
\]

Alternatively, we can also write

\[
\nabla^2 = \frac{1}{r^2} \frac{\partial}{\partial r} r^2 \frac{\partial}{\partial r} + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \sin \theta \frac{\partial}{\partial \theta} + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2}. \quad \text{(L.26)}
\]
Verification is straightforward as below.

\[
\begin{align*}
\frac{1}{r^2} \frac{\partial}{\partial r} r^2 \frac{\partial}{\partial r} + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \sin \theta \frac{\partial}{\partial \theta} \\
= \frac{1}{r^2} \left( 2r \frac{\partial}{\partial r} + r^2 \frac{\partial^2}{\partial r^2} \right) + \frac{1}{r^2 \sin \theta} \left( \cos \theta \frac{\partial}{\partial \theta} + \sin \theta \frac{\partial^2}{\partial \theta^2} \right) \\
= \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} + \left( \frac{\cos \theta}{r^2 \sin \theta} \right) \frac{\partial}{\partial \theta} + \frac{1}{r^2 \sin \theta} \frac{\partial^2}{\partial \theta^2}
\end{align*}
\]

(L.27)
Appendix M

Legendre and Associated Legendre Polynomials

M.1 Legendre Polynomials

The $n$-th degree Legendre polynomial $P_n(x)$ is a solution to the following Legendre’s differential equation, where $n = 0, 1, 2, \ldots$

$$
\frac{d}{dx} \left[(1 - x^2) \frac{d}{dx} P_n(x) \right] + n(n + 1)P_n(x) = 0 \tag{M.1}
$$

Each $P_n(x)$ can be obtained using Rodrigues’ formula

$$
P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} \left[(x^2 - 1)^n\right], \tag{M.2}
$$

which can be verified by differentiating both sides of the identity

$$
(x^2 - 1) \frac{d}{dx} (x^2 - 1)^n = 2nx(x^2 - 1)^n \tag{M.3}
$$

$(n + 1)$ times. When the condition $P_n(1) = 1$ is imposed, we have

$$
\begin{align*}
P_0(x) &= 1 \\
P_1(x) &= x \\
P_2(x) &= \frac{1}{2}(3x^2 - 1) \\
P_3(x) &= \frac{1}{2}(5x^3 - 3x)
\end{align*}
$$
\[ P_4(x) = \frac{1}{8}(35x^4 - 30x^2 + 3) \]
\[ P_5(x) = \frac{1}{8}(63x^5 - 70x^3 + 15x) \]
\[ P_6(x) = \frac{1}{16}(231x^6 - 315x^4 + 105x^2 - 5) \]

### APPENDIX M. LEGENDRE AND ASSOCIATED LEGENDRE POLYNOMIALS

#### M.2 Associated Legendre Polynomials

The associated Legendre polynomial of degree \( l \) and order \( m \), denoted by \( P^m_l(x) \), satisfies the following general Legendre equation.

\[
(1 - x^2) \frac{d^2}{dx^2} P^m_l(x) - 2x \frac{d}{dx} P^m_l(x) + \left[l(l+1) - \frac{m^2}{1-x^2}\right] P^m_l(x) = 0
\]

(M.4)

\( P^m_l(x) \) can be expressed in the following Rodrigues type form.

\[
P^m_l(x) = \frac{(-1)^l}{2^l l! m!} (1 - x^2)^{m/2} \frac{d^{l+m}}{dx^{l+m}} (1 - x^2)^l
\]

(M.5)

In particular, for \( m > 0 \), we have

\[
P^m_l(x) = (1 - x^2)^{m/2} \frac{d^m}{dx^m} P_l(x).
\]

(M.6)

**Fact M.1** We have the following identity.

\[
P_l^{-m}(x) = (-1)^m \frac{(l-m)!}{(l+m)!} P^m_l(x)
\]

(M.7)

**Proof** (taken from a lecture note prepared by Professor Hitoshi Murayama of UC Berkeley)

We will first compute how

\[
\frac{d^{l+m}}{dx^{l+m}} (1 - x^2)^l
\]

is related to

\[
\frac{d^{l-m}}{dx^{l-m}} (1 - x^2)^l.
\]

We need two lemmas for this proof.

**Lemma M.1 (Combinatorial Identity)** We have the following identity.

\[
\binom{k+1}{r} = \binom{k}{r-1} + \binom{k}{r}
\]

(M.8)
Proof of Lemma M.1
It is nothing but a straightforward computation.

\[
\binom{k}{r-1} + \binom{k}{r} = \frac{k!}{(k-r+1)(r-1)!} + \frac{k!}{(k-r+1)(k-r)!} \\
= \frac{k!r}{(k+1-r)!r(r-1)!} + \frac{k!(r+k-r+1)}{(k+1-r)!r!} \\
= \frac{k!(k+1)}{(k+1-r)!r!} = \frac{(k+1)!}{(k+1-r)!r!} = \binom{k+1}{r} \tag{M.9}
\]

Lemma M.2 Let \(F(x)\) and \(G(x)\) be two functions which are at least \(n\)-times differentiable. Then,

\[
\frac{d^n}{dx^n} F(x)G(x) = \sum_{r=0}^{n} \binom{n}{r} \left[ \frac{d^r}{dx^r} F(x) \right] \left[ \frac{d^{n-r}}{dx^{n-r}} G(x) \right]. \tag{M.10}
\]

Proof of Lemma M.2
Let us introduce a short-hand notation

\[
d^r F := \frac{d^r}{dx^r} F(x); \tag{M.11}
\]

where \(\frac{d^r}{dx^r} F(x) = F(x)\) by definition. Our proof is by mathematical induction. First, when \(n = 1\), we have

\[
d(FG) = (dF)G + F(dG) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} (dF)G + \begin{pmatrix} 1 \\ 1 \end{pmatrix} F(dG) \\
= \sum_{r=0}^{1} \binom{1}{r} \left[ \frac{d^r}{dx^r} F(x) \right] \left[ \frac{d^{1-r}}{dx^{1-r}} G(x) \right], \tag{M.12}
\]

and (M.10) holds for \(n = 1\). Now, suppose (M.10) holds for \(n = k \geq 1\), so that

\[
\frac{d^k}{dx^k} F(x)G(x) = \sum_{r=0}^{k} \binom{k}{r} \left[ \frac{d^r}{dx^r} F(x) \right] \left[ \frac{d^{k-r}}{dx^{k-r}} G(x) \right] = \sum_{r=0}^{k} \binom{k}{r} d^r F d^{k-r} G. \tag{M.13}
\]

Then, for \(n = k + 1\), we get

\[
\frac{d^{k+1}}{dx^{k+1}} F(x)G(x) = \frac{d}{dx} \left[ \frac{d^k}{dx^k} F(x)G(x) \right] 
\]
\[
\frac{d}{dx} \left\{ \sum_{r=0}^{k} \binom{k}{r} \left[ \frac{d^r}{dx^r} F(x) \right] \left[ \frac{d^{k-r}}{dx^{k-r}} G(x) \right] \right\} = \frac{d}{dx} \left[ \sum_{r=0}^{k} \binom{k}{r} d^r F d^{k-r} G \right] = \sum_{r=0}^{k} \binom{k}{r} d^r F d^{k-r} G
\]

\[
= \sum_{r=0}^{k} \binom{k}{r} \left( d^{r+1} F d^{k-r} G + d^r F d^{k-r+1} G \right)
\]

\[
= \sum_{r=0}^{k} \binom{k}{r} d^{r+1} F d^{k-r} G + \sum_{r=0}^{k} \binom{k}{r} d^r F d^{k-r+1} G. \tag{M.14}
\]

Now, let \( u = r + 1 \). Then,

\[
\sum_{r=0}^{k} \binom{k}{r} d^{r+1} F d^{k-r} G = \sum_{u=1}^{k+1} \binom{k}{u-1} d^u F d^{k+1-u} G
\]

\[
= \sum_{r=1}^{k+1} \binom{k}{r-1} d^r F d^{k+1-r} G. \tag{M.15}
\]

Plug this into (M.14) and use Lemma M.1.

\[
\frac{d^{k+1}}{dx^{k+1}} F(x) G(x) = \sum_{r=0}^{k} \binom{k}{r} d^{r+1} F d^{k-r} G + \sum_{r=0}^{k} \binom{k}{r} d^r F d^{k-r+1} G
\]

\[
= \sum_{r=1}^{k+1} \binom{k}{r-1} d^r F d^{k+1-r} G + \sum_{r=0}^{k} \binom{k}{r} d^r F d^{k+1-r} G
\]

\[
= \binom{k+1}{k} d^0 F d^{k+1} G + \sum_{r=1}^{k} \left[ \binom{k}{r-1} + \binom{k}{r} \right] d^r F d^{k+1-r} G
\]

\[
+ \binom{k+1}{k+1} d^{k+1} F d^0 G
\]

\[
= \binom{k+1}{0} d^0 F d^{k+1} G + \sum_{r=1}^{k} \left[ \binom{k+1}{r} \right] d^r F d^{k+1-r} G + \binom{k+1}{k+1} d^{k+1} F d^0 G
\]

\[
= \sum_{r=0}^{k+1} \binom{k+1}{r} d^r F d^{k+1-r} G. \tag{M.16}
\]
So, M.10 holds for \( n = k + 1 \), and we have proved the lemma by mathematical induction.

Let us now resume the proof of Fact M.1. We have

\[
\frac{d^{l+m}}{dx^{l+m}}(1-x^2)^l = \frac{d^{l+m}}{dx^{l+m}}(1-x)^l(1+x)^l
\]

\[
= \sum_{r=0}^{l+m} \binom{l+m}{r} \left( \frac{d^r}{dx^r}(1-x)^l \right) \left( \frac{d^{l+m-r}}{dx^{l+m-r}}(1+x)^l \right).
\]  

(M.17)

However, because \((1-x)^l\) and \((1+x)^l\) are both \(l\)-th degree polynomials,

\[
\frac{d^r}{dx^r}(1-x)^l = 0 \text{ if } r > l
\]

and

\[
\frac{d^{l+m-r}}{dx^{l+m-r}}(1+x)^l = 0 \text{ if } r < m.
\]

(M.18)

(M.19)

Therefore, the only nonzero terms are for \( m \leq r \leq l \), and

\[
\frac{d^{l+m}}{dx^{l+m}}(1-x^2)^l = \sum_{r=m}^{l} \binom{l+m}{r} \left( \frac{d^r}{dx^r}(1-x)^l \right) \left( \frac{d^{l+m-r}}{dx^{l+m-r}}(1+x)^l \right).
\]  

(M.20)

Now, observe the following.

\[
\frac{d}{dx}(1-x)^l = \frac{d}{dx}(1-x) \frac{d}{dx}(1-x)^l = l(1-x)^{l-1}(-1),
\]

\[
\frac{d^2}{dx^2}(1-x)^l = \frac{d}{dx} \left[ \frac{d}{dx}(1-x)^l \right] = (-1)^l l \frac{d}{dx}(1-x)^{l-1}
\]

\[
= (-1)^l l(1-x)^{l-2}(-1) = (-1)^2 l(l-1)(1-x)^{l-2}(-1)
\]

\[
= (-1)^k l(l-1) \ldots (l-k+1)(1-x)^{l-k} = (-1)^k \frac{l!}{(l-k)!}(1-x)^{l-k}
\]

(M.21)

Similarly, we also have

\[
\frac{d^k}{dx^k}(1+x)^l = \frac{l!}{(l-k)!}(1+x)^{l-k}.
\]

(M.23)
Therefore, (M.20) can be expressed as follows.
\[
\frac{d^{l+m}}{dx^{l+m}}(1-x^2)^l = \sum_{r=m}^l \binom{l+m}{r} \frac{(-1)^r l!}{(l-r)!} (1-x)^{l-r} \frac{l!}{(r-m)!}(1+x)^{r-m} \quad (M.24)
\]

At this point, let \( s = r - m \) to obtain
\[
\frac{d^{l+m}}{dx^{l+m}}(1-x^2)^l = \sum_{s=0}^{l-m} \binom{l+m}{m+s} \frac{(-1)^{m+s}l!}{(l-m-s)!} (1-x)^{l-m-s} \frac{s!}{s!}(1+x)^s. \quad (M.25)
\]

Next, we will multiply the right-hand side by \( \frac{(1-x^2)^m}{(1-x^2)^m} \).
\[
\frac{d^{l+m}}{dx^{l+m}}(1-x^2)^l = \sum_{s=0}^{l-m} \binom{l+m}{m+s} \frac{(-1)^{m+s}l!}{(l-m-s)!} (1-x)^{l-m-s} \frac{s!}{s!}(1+x)^s
\]
\[
= \frac{(1-x^2)^m}{(1-x^2)^m} \sum_{s=0}^{l-m} \binom{l+m}{m+s} \frac{(-1)^{m+s}l!}{(l-m-s)!} (1-x)^{l-m-s} \frac{s!}{s!}(1+x)^s
\]
\[
= \frac{(1-x)^m(1+x)^m}{(1-x^2)^m} \sum_{s=0}^{l-m} \binom{l+m}{m+s} \frac{(-1)^{m+s}l!}{(l-m-s)!} (1-x)^{l-s} \frac{s!}{s!}(1+x)^s(1+x)^m
\]
\[
= \frac{1}{(1-x^2)^m} \sum_{s=0}^{l-m} \binom{l+m}{m+s} \frac{(-1)^{m+s}l!}{(l-m-s)!} (1-x)^{l-s} \frac{s!}{s!}(1+x)^{m+s}
\]
\[
= \frac{1}{(1-x^2)^m} \sum_{s=0}^{l-m} \frac{(l+m)!}{(l-s)!(m+s)!(l-m-s)!} (-1)^{m+s}l! (1-x)^{l-s} \frac{s!}{s!}(1+x)^{m+s}
\]
\[
= \frac{1}{(1-x^2)^m} \sum_{s=0}^{l-m} \frac{(l+m)!}{s!(l-m-s)!} (1-x)^{l-s} \frac{s!}{s!}(1+x)^{m+s}
\]
where the last equality follows from (M.17) with \( l + m \) replaced by \( l - m \).

Then, from (M.5) and (M.26), we get

\[
P_l^m(x) = \frac{(-1)^l}{2^l!} (1 - x^2)^{m/2} \frac{d^{l+m}}{dx^{l+m}} (1 - x^2)^l
\]

\[
= \frac{(-1)^l}{2^l!} (1 - x^2)^{m/2} (l + m)! \frac{d^m}{dx^m} (1 - x^2)^l
\]

\[
= \frac{(l + m)!}{(l - m)!} (-1)^l (1 - x^2)^{-m/2} \frac{d^{-m}}{dx^{-m}} (1 - x^2)^l
\]

\[
= \frac{(l + m)!}{(l - m)!} (-1)^l P_l^{-m}(x)
\]

\[
\implies P_l^{-m}(x) = (-1)^m \frac{(l - m)!}{(l + m)!} P_l^m(x)
\]
Appendix N

Laguerre and Associated Laguerre Polynomials

For a nonnegative integer \( m \), i.e. \( m = 0, 1, 2, 3, \ldots \), Laguerre’s equation is the linear second order ordinary differential equation

\[
xy''(x) + (1 - x)y'(x) + my(x) = 0, \quad (N.1)
\]

whose polynomial solutions, denoted by \( L_m(x) \) are known as Laguerre polynomials. The general term is given by the Rodrigues formula

\[
L_n(x) = \frac{e^x}{n!} \frac{d^n}{dx^n} \left( e^{-x} x^n \right), \quad (N.2)
\]

or the closed form\(^1\)

\[
L_k(x) = \sum_{i=0}^{k} \binom{k}{i} \frac{(-1)^i}{i!} x^i \quad \text{or} \quad \sum_{i=0}^{k} k C_i \frac{(-1)^i}{i!} x^i. \quad (N.3)
\]

The first few polynomials are given below.

\[
L_0(x) = 1 \\
L_1(x) = -x + 1 \\
L_2(x) = \frac{1}{2} (x^2 - 4x + 2) \\
L_3(x) = \frac{1}{6} (-x^3 + 9x^2 - 18x + 6)
\]

\(^1\)Note that \( \binom{k}{i} \) and \( k C_i \) mean the same thing. Depending on where you went to school, you may be familiar with either.
\[ L_4(x) = \frac{1}{24}(x^4 - 16x^3 + 72x^2 - 96x + 24) \]
\[ L_5(x) = \frac{1}{120}(-x^5 + 25x^4 - 200x^3 + 600x^2 - 600x + 120) \]
\[ L_6(x) = \frac{1}{720}(x^6 - 36x^5 + 450x^4 - 2400x^3 + 5400x^2 - 4320x + 720) \]

The Laguerre polynomials satisfy a three-term recurrence relation

\[ (k + 1)L_{k+1}(x) = (2k + 1 - x)L_k(x) - kL_{k-1}(x). \quad (N.4) \]

According to Favard’s theorem, the Laguerre polynomials form a family of orthogonal polynomials.

**Theorem N.1 (Favard’s Theorem)** If a system of polynomials \( \{P_n(x)\} \) of degree \( n = 0, 1, 2, \ldots \) satisfies

\[
xP_n(x) = a_nP_{n+1}(x) + b_nP_n(x) + c_nP_{n-1}(x) \quad n \geq 1, \quad (N.5)
xP_0(x) = a_0P_1(x) + b_0P_0(x) \quad (N.6)
\]

where \( a_n, b_n \) (\( n \geq 0 \)), and \( c_n \) (\( n \geq 1 \)) are real constants, and \( a_n c_{n+1} > 0 \), then, there exists a positive measure \( \mu \) such that the polynomials \( P_n(x) \) are orthogonal with respect to \( \mu \).\(^2\)

Indeed, \( L_k \)'s are orthogonal with respect to an inner product defined by

\[ \langle f|g \rangle = \int_0^\infty f(x)g(x)e^{-x} \, dx. \quad (N.7) \]

---

\(^2\)The following holds for \( \mu \).

1. The measure \( \mu \) may not be unique even if a scaling factor is ignored.
2. If \( \mu \) is unique, the polynomials are dense in \( L^2 \) with respect to \( \mu \).
3. If there is a \( \mu \) with bounded support, then, \( \mu \) is unique.
Appendix O

Fine Structure and Lamb Shift

O.1 Relativistic Effects: The Dirac Theory

The Hamiltonian $H_0$ for the hydrogen atom used in Chapter 10 was:

$$H_0 = \frac{\mathbf{p}^2}{2\mu} + V(r) = \frac{\mathbf{p}^2}{2\mu} - \frac{q^2}{4\pi\varepsilon_0 r} = \frac{\mathbf{p}^2}{2\mu} - \frac{e^2}{r};$$  \hspace{1cm} (O.1)

where $q$ is the electron charge and $e^2 := \frac{q^2}{4\pi\varepsilon_0}$. The first term is the kinetic energy in the center of mass frame, and the second term is the electrostatic interaction energy between the electron and the proton.

Now, the problem is that this Hamiltonian is not exact as it does not include relativistic effects. For example, the magnetic effects involving the electron spin as well as magnetic interactions related to the proton spin are ignored. In fact, these relativistic effects are typically about $(1/137)^2$ the size of $H_0$. Still, spectroscopic measurements are accurate enough to detect these effects. In 1928, Charles Galton Darwin\textsuperscript{2} and Walter Gordon found the following expressions for the energy of an

\textsuperscript{1}Nothing is carved in stone. However, it appears customary to express the Hamiltonian in terms of $e$ as defined here. Note that what we called $e$ in Chapter 10 is denoted by $q$ here. For those “old-timers” who are familiar with the CGS (centimeter-gram-second) System of Units, this is a device to connect CGS with SI (the International System of Units) if you may.

\textsuperscript{2}Sir Charles Galton Darwin (1987-1962), a British physicist, was a grandson of the Charles Darwin of evolution fame. Sir Charles was the first, with Gordon, to work out the exact energy levels of Hydrogen according to the Dirac equation and thereby discovered the eponymous term in the levels. He worked out the Lagrangian and Hamiltonian for classical motion of several interacting charges correct to $\mathcal{O}(v^2/c^2)$. He also worked on statistical mechanics (Darwin-Fowler method). Later in life he took part in the Manhattan project.
electron in a hydrogen atom.

\[ W = mc^2 \left( 1 + \frac{\gamma^2}{(k' + n')^2} \right)^{-\frac{3}{2}}; \tag{O.2} \]

where \( \gamma = \frac{2\pi^2}{\hbar c} \) is the fine structure constant, \( k(= 0, 1, 2, \ldots) \) is the azimuthal quantum number, \( k' = \sqrt{(k + 1)^2 - \gamma^2} \), and \( n' \) is a nonnegative integer [Darwin, 1928, p.672].

\[ E = \frac{mc^2}{\sqrt{1 + \frac{\alpha^2}{(\gamma + k')^2}}}; \tag{O.3} \]

where \( \alpha \) is the fine structure constant, \( K = \pm 1, \pm 2, \pm 3, \ldots \), and \( n = 0, 1, 2, \ldots \) [Gordon, 1928, p.14][Chen, nd, p.6].

\[ \frac{E}{mc} = \left[ \frac{1}{\alpha Z} \frac{\alpha Z}{\sqrt{\gamma - \alpha^2}} \right]^2 \tag{O.4} \]

[Ketterle, nd, p.2]

\[ E = mc^2 \left( 1 + \frac{\gamma^2}{(s + n')^2} \right)^{-\frac{1}{2}}; \tag{O.5} \]

where \( \gamma = \frac{Ze^2}{\hbar c} \), \( k = \pm 1, \pm 2, \ldots \), \( s = (k^2 - \gamma^2)^{\frac{1}{2}} \), and \( n' = 0, 1, 2, \ldots \) [Schiff, 1968, p.486].

\[ E = \mu c^2 \left( 1 + \frac{Z^2 \alpha^2}{n + \sqrt{\lambda^2 - Z^2 \alpha^2}} \right)^{-\frac{1}{2}}; \tag{O.6} \]

where \( \mu \) is the reduced mass, \( \alpha = \frac{e^2}{\hbar c} \approx 1/137 \), \( Z \) is the atomic number of the nucleus, and \( \lambda \) is an integer [Chongming and Dianyan, 1984, p.171].

\[ E_{njlm} = mc^2 \left[ 1 + \left( \frac{Z\alpha}{n - (j + 1/2) + \sqrt{(j + 1/2)^2 - (Z\alpha)^2}} \right)^2 \right]^ {-\frac{1}{2}}; \tag{O.7} \]

where \( n = 1, 2, \ldots \) are the principal quantum numbers, and \( j+1/2 \leq n \) [Murayama, nd, p.17].
Because the hydrogen atom is a weakly relativistic system\(^3\), it is useful to approximate the Hamiltonian to the lowest order in \(v/c\). We can write
\[
H = mc^2 + \frac{p^2}{2m} - \frac{e^2}{r} + H_{FS} = mc^2 + H_0 + H_{FS}; \tag{O.8}
\]
where \(mc^2\) is the electron’s rest energy, \(\frac{p^2}{2m} - \frac{e^2}{r} = H_0\) is the non-relativistic Hamiltonian, and \(H_{FS}\) is the fine structure interaction given by
\[
H_{FS} = -\frac{p^4}{8m^3c^2} + \frac{e^2}{2m^2c^2} \frac{L \cdot S}{r^3} - \frac{\hbar^2}{8m^2c^2} \nabla^2 \frac{e^2}{r}. \tag{O.9}
\]
The three terms of \(H_{FS}\) are known as the kinetic, spin-orbit interaction, and Darwinian terms, denoted by \(H_k\), \(H_{SO}\), and \(H_D\), respectively. Contribution made by each of the three terms is of the order \((v/c)^2\).
\[
H_{FS} = -\underbrace{-\frac{p^4}{8m^3c^2}}_{H_k} + \underbrace{\frac{e^2}{2m^2c^2} \frac{L \cdot S}{r^3}}_{H_{SO}} - \underbrace{\frac{\hbar^2}{8m^2c^2} \nabla^2 \frac{e^2}{r}}_{H_D}. \tag{O.10}
\]

### O.1.1 Variation of the Mass with the Velocity: \(H_k\)

The relativistic expression for the energy \(E\) of a classical particle of rest-mass \(m\) and momentum \(p\) is given by
\[
E = c \sqrt{p^2 + m^2c^2}, \tag{O.11}
\]
and the first three terms of its expansion in powers of \(p/mc\) gives
\[
E = mc^2 + \frac{p^2}{2m} - \underbrace{\frac{p^4}{8m^3c^2}}_{H_k} + \ldots \tag{O.12}
\]
\[
= \underbrace{mc^2}_{\text{rest-mass energy}} + \underbrace{\frac{p^2}{2m}}_{\text{non-relativistic kinetic energy}} - \underbrace{\frac{p^4}{8m^3c^2}}_{\text{first correction}} + \ldots. \tag{O.13}
\]
This is the origin of \(H_k\).

\(^3\)In the Bohr model, the velocity \(v\) of an electron in the orbit characterized by \(n = 1\) is given by \(v = (e^2/h)c \approx (1/137)c \ll 1\).

\(^4\)Some schools define \(S\) and \(L\) such that \(\|S\|\hbar\) and \(\|L\|\hbar\) are the magnitudes of the corresponding spin and orbital angular momenta. When this convention is adopted, the second term becomes \(\frac{\hbar^2 e^2}{2m^2c^2} \frac{1}{r} L \cdot S\).
O.1.2 Spin-Orbit Interaction: $H_{SO}$

The intrinsic spin dipole magnetic moment of an electron $\mu_S$ is given by

$$\mu_S = -g_S \mu_B \frac{S}{\hbar};$$

(O.14)

where the spin $g$-factor $g_S$ for an electron is approximately 2 ($g_S \approx 2$), $\mu_B$ is a physical constant known as Bohr magneton, and $S$ is the electron’s spin angular momentum whose magnitude $S$ is given by $S^2 = s(s+1)\hbar^2 = \frac{3}{2}\hbar^2$. As the electron is a spin one-half particle, $S = \frac{\hbar}{2}$, and the magnitude $\mu_S$ of $\mu_S$ is approximately equal to one Bohr magneton.

$$\mu_S \approx 2 \frac{e\hbar}{2m} \frac{\hbar}{2\hbar} = \mu_B$$

(O.15)

The spin-orbit interaction energy stems from the interaction between this magnetic dipole and the motional magnetic field $B$ experienced by the electron as it moves through the electric field of the proton. Naively, the motional magnetic field $B_{naive}$ is given by

$$B_{naive} = -\frac{v}{c} \times E = -\frac{v}{c} \times \frac{e}{r^3} r = \frac{e}{mc r^3} L.$$  

(O.16)

However, the relativistic transformation of a vector between two coordinate frames moving with different velocities, which are not collinear, involves a rotation in addition to a dilation. That rotation is known as the Thomas precession, which effectively cuts the motional magnetic field in half. Therefore, the true motional magnetic field $B_{motion}$ experienced by the electron is given by

$$B_{motion} = \frac{1}{2} \frac{e}{mc r^3} L.$$  

(O.17)

---

$^5$The Bohr magneton is given by

$$\mu_B = \frac{e\hbar}{2m}$$

in SI units and by

$$\mu_B = \frac{e\hbar}{2mc}$$

in Gaussian CGS units.
Therefore, the total spin-orbit interaction $H_{SO}$ is given by

$$H_{SO} = -\mathbf{\mu}_S \cdot \mathbf{B}_{\text{motion}} = \left(-gs\mu_B \frac{S}{\hbar}\right) \cdot \left(\frac{1}{2mc} \frac{e}{cr^3} \mathbf{L}\right) = 2\cdot \frac{e^2}{2mc\hbar} \frac{1}{2mc} \frac{e}{cr^3} \mathbf{L} \cdot \mathbf{S}$$

$$= \frac{e^2}{2m^2c^2} \frac{1}{r^3} \mathbf{L} \cdot \mathbf{S}. \quad (O.18)$$

### O.1.3 The Darwin Term: $H_D$

A particle has what is known as the Compton wavelength $\lambda_C$ given by

$$\lambda_C = \frac{\hbar}{mc}; \quad (O.19)$$

where the capital $C$ in $\lambda_C$ stands for “Compton”, and the small letter $c$ in the denominator of the right-hand side is the speed of light. Often, the reduced Compton wavelength $\lambda_{rC}$ defined by

$$\lambda_{rC} = \frac{\lambda_C}{2\pi} = \frac{\hbar}{mc} \quad (O.20)$$

is used instead. For the electron, the reduced Compton wavelength is roughly 386 fm (femtometers) or $3.86 \times 10^{-13}$ meters$^7$. The Compton wavelength can be thought of as a fundamental limitation on measuring the position of a particle, taking quantum mechanics and special relativity into account.

What this means is that we should use an effective potential $V_{\text{effective}}$, which is $V(r)$ averaged over a small ball whose radius is the Compton wavelength. In other words, rather than $V(r)$, we should average $V$ over all the points $r + \epsilon$ with $\|\epsilon\| \leq \lambda_C$ or $\lambda_{rC}$. Typically, we resort to the Taylor expansion of $V(r + \epsilon)$ as follows and conduct term-by-term integration to get the average. This fluctuation or uncertainty is sometimes called Zitterbewegung (trembling motion).

$$V(r + \epsilon) = V(r) + \epsilon \cdot \nabla V(r) + \sum_{i,j=1}^{3} \frac{1}{2} \epsilon_i \epsilon_j \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} V(r) + \ldots \quad (O.21)$$

When it is all said and done, we get

$$H_D = \frac{\hbar^2}{8m^2c^2} 4\pi \left(\frac{Ze^2}{4\pi\varepsilon_0}\right) \delta^3(r) = \frac{Ze^2\hbar^2}{8m^2c^2\varepsilon_0} \delta^3(r); \quad (O.22)$$

---

$^6$More detail can be found in Ketterle’s lecture notes for example [Ketterle, nd].

$^7$1 nanometer = $10^6$ femtometers
where the identity

\[ \nabla^2 \frac{1}{r} = -4\pi \delta^3(r) \]  

was used.

Now, even the exact solution of the Dirac equation still has some shortcomings. One notable problem is that it predicts degeneracy for such pairs of states as \((2S_{\frac{1}{2}}, 2P_{\frac{1}{2}}), (3S_{\frac{3}{2}}, 3P_{\frac{3}{2}}), (3P_{\frac{3}{2}}, 3D_{\frac{3}{2}}), \) and \((3D_{\frac{5}{2}}, 3F_{\frac{5}{2}})^\text{8},\) when indeed these states are not degenerate. This problem arises because states of the same values of \(n\) and \(j\) are degenerate according to the Dirac theory as shown in \((O.7)\). We need to consider an additional quantum mechanical effect sometimes called the vacuum interaction to remove this degeneracy. The effect is the largest for \(n = 2\), and the energy splitting between the \(2S_{\frac{1}{2}}\) state and the \(2P_{\frac{1}{2}}\) state is called the Lamb Shift.

### O.2 The Lamb Shift

Some physicists say "Vacuum is noisy." Others say "Vacuum is not really empty." What they typically mean is that there is quantum fluctuation which allows nonzero electromagnetic fields in vacuum. We have zero-point fluctuation of the electromagnetic modes of free space because they behave like harmonic oscillators with a strictly positive zero-point energy \(h\nu/2\), and the time averages of the squared amplitudes of electric and magnetic fields are strictly positive. This electric field causes the average potential experienced by the electron to deviate from \(V(r)\), and this is the reason behind the Lamb Shift. The observed value is 1,058 MHz.

Interested readers can find a very readable exposition in a lecture note by Wolfgang Ketterle of MIT [Ketterle, nd].

---

**Spectroscopic Notation**

The following explanation was adopted with slight modifications from Jim Branson’s notes [Branson, nd]. Spectroscopic notation is widely used in atomic physics, which provides a standard way to describe the angular momentum quantum numbers of a state. The general form is \(N^{2s+1}L_j\); where \(N\) is the principal quantum number, \(s\) signifies the total spin quantum number so that there are \(2s + 1\) spin states, \(L\) refers to the orbital angular momentum quantum number \(l\) customarily written as \(S, P, D, F, \ldots\) for \(l = 0, 1, 2, 3, \ldots\), and \(j\) is the total angular momentum quantum number. Note that \(N\) is often dropped. For the hydrogen atom, we have \(1^2S_{\frac{1}{2}}, 2^2S_{\frac{1}{2}}, 2^2P_{\frac{3}{2}}, 2^2P_{\frac{1}{2}}, 3^2S_{\frac{1}{2}}, 3^2P_{\frac{3}{2}}, 3^2P_{\frac{1}{2}}, 3^2D_{\frac{5}{2}}, 3^2D_{\frac{3}{2}}, 4^2S_{\frac{1}{2}}, 4^2P_{\frac{3}{2}}, 4^2P_{\frac{1}{2}}, 4^2D_{\frac{5}{2}}, 4^2D_{\frac{3}{2}}, 4^2F_{\frac{7}{2}}, 4^2F_{\frac{5}{2}}, \) and so on.
Bibliography


Answers to Exercises

This section is a work in progress with some dummy place-holders at this point.

Chapter 2

1. (a) $\Omega$ is Hermitian because $(\Omega^t)^* = \Omega$ as shown below.

\[
\left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^t\right)^* = \left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\right)^* = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}
\]

(b)

\[
\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = 1 \cdot \begin{bmatrix} a \\ b \end{bmatrix} \implies \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} \implies a = b
\]

\[
\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = -1 \cdot \begin{bmatrix} a \\ b \end{bmatrix} \implies \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} -a \\ -b \end{bmatrix} \implies a = -b
\]

Hence, $|1\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, and $|-1\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

(c)

\[
\langle 1| - 1\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 0
\]

(d) Noting that

\[
\frac{|1\rangle + |-1\rangle}{\sqrt{2}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}
\]
and
\[
\begin{pmatrix}
|1\rangle - |-1\rangle
\end{pmatrix} = \begin{pmatrix}
0 \\
1
\end{pmatrix},
\]
we get
\[
v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = v_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + v_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = v_1 \left( \frac{|1\rangle + |-1\rangle}{\sqrt{2}} \right) + v_2 \left( \frac{|1\rangle - |-1\rangle}{\sqrt{2}} \right)
\]
\[
= \frac{v_1 + v_2}{\sqrt{2}} |1\rangle + \frac{v_1 - v_2}{\sqrt{2}} |-1\rangle
\]
(e)
\[
\langle 1|\Omega|1 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = 1
\]
\[
\langle 1|\Omega|-1 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 0
\]
\[
\langle -1|\Omega|1 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = 0
\]
\[
\langle -1|\Omega|-1 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = -1
\]
Hence,
\[
[\Omega_{ij}] = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.
\]
2. (a) As \(M\) is Hermitian, we have to have \(M^\dagger = M\). So,
\[
\begin{pmatrix} 6 & 4 \\ \alpha & 0 \end{pmatrix}^\dagger = \begin{pmatrix} 6 & \alpha \\ 4 & 0 \end{pmatrix}^\dagger = \begin{pmatrix} 6 & \alpha^* \\ 4 & 0 \end{pmatrix} = \begin{pmatrix} 6 & 4 \\ \alpha & 0 \end{pmatrix} \implies \alpha = 4.
\]
(b)
\[
\begin{pmatrix} 6 & 4 \\ 4 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 6a + 4b \\ 4a \end{pmatrix} = 8 \begin{pmatrix} a \\ b \end{pmatrix} \implies a = 2b \implies |8\rangle = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ 1 \end{pmatrix}
\]
\[
\begin{pmatrix} 6 & 4 \\ 4 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 6a + 4b \\ 4a \end{pmatrix} = -2 \begin{pmatrix} a \\ b \end{pmatrix} \implies b = -2a \implies |-2\rangle = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ -2 \end{pmatrix}
\]
(c) 

\[ M_{11} = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 & 1 \\ 4 & 0 \end{bmatrix} \begin{bmatrix} 6 & 4 \\ 4 & 0 \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 1 \end{bmatrix} \begin{bmatrix} 16 \\ 8 \end{bmatrix} = \frac{1}{5} \cdot 40 = 8 \]

\[ M_{12} = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 & 1 \\ 4 & 0 \end{bmatrix} \begin{bmatrix} 6 & 4 \\ 4 & 0 \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ -2 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 1 \\ -2 \end{bmatrix} \begin{bmatrix} -2 \\ 4 \end{bmatrix} = 0 \]

\[ M_{21} = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & -2 \\ 4 & 0 \end{bmatrix} \begin{bmatrix} 6 & 4 \\ 4 & 0 \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 1 \end{bmatrix} \begin{bmatrix} -2 \\ 4 \end{bmatrix} = 0 \]

\[ M_{22} = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & -2 \\ 4 & 0 \end{bmatrix} \begin{bmatrix} 6 & 4 \\ 4 & 0 \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ -2 \end{bmatrix} \begin{bmatrix} -2 \\ 4 \end{bmatrix} = -2 \]

Therefore,

\[ M = [M_{ij}] = \begin{bmatrix} 8 & 0 \\ 0 & -2 \end{bmatrix}. \]

3. (a) 

\[ [A, B] = AB - BA = \begin{bmatrix} 0 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 2 \\ 1 & 0 \end{bmatrix} - \begin{bmatrix} -1 & 2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 2 \\ 1 & 1 \end{bmatrix} \]

\[ = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} - \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = 0 \]

(b) 

\[ \begin{bmatrix} 0 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = 2 \begin{bmatrix} a \\ b \end{bmatrix} \Rightarrow \begin{bmatrix} 2b \\ a + b \end{bmatrix} = \begin{bmatrix} 2a \\ 2b \end{bmatrix} \Rightarrow a = b \]

So,

\[ |2\rangle_A = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \]

\[ \begin{bmatrix} 0 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = -1 \begin{bmatrix} c \\ d \end{bmatrix} \Rightarrow \begin{bmatrix} 2d \\ c + d \end{bmatrix} = \begin{bmatrix} -c \\ -d \end{bmatrix} \Rightarrow c = -2d \]

So,

\[ |-1\rangle_A = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ -1 \end{bmatrix}. \] (13.24)
(c)\[B\ket{2}_A = \begin{bmatrix} -1 & 2 \\ 1 & 0 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1 \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}\]

\[B\ket{-1}_A = \begin{bmatrix} -1 & 2 \\ 1 & 0 \end{bmatrix} \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} -4 \\ 2 \end{bmatrix} = -2 \cdot \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ -1 \end{bmatrix}\]

Therefore, \(\ket{2}_A\) is an eigenvector of \(B\) with the associated eigenvalue 1, and \(\ket{-1}\) is an eigenvector of \(B\) with the associated eigenvalue -2.

(d)\[\frac{\sqrt{5}\ket{-1}_A + \sqrt{2}\ket{2}_A}{3} = \frac{1}{3} \left( \begin{bmatrix} 2 \\ -1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right) = \frac{1}{3} \begin{bmatrix} 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}\]

\[\frac{\sqrt{5}\ket{-1}_A - 2\sqrt{2}\ket{2}_A}{-3} = \frac{1}{-3} \left( \begin{bmatrix} 2 \\ -1 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \end{bmatrix} \right) = \frac{1}{-3} \begin{bmatrix} 0 \\ -3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}\]

Therefore, \(\{\ket{2}_A, \ket{-1}_A\}\) indeed forms a basis for \(\mathbb{C}^2\).

(e) Remembering that \(\Omega_{ij} = \langle i|\Omega|j\rangle\), we can compute the entries as follows.

\[A_{11} = \langle 2|A|2\rangle_A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 0 \\ 2 \end{bmatrix} \frac{1}{\sqrt{2}} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}\]

\[= \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \sqrt{\frac{2}{\sqrt{2}}} = 2; \text{ where } A\langle 2|\text{ is the adjoint of } \ket{2}_A\]

\[A_{12} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 0 \\ 2 \end{bmatrix} \frac{\sqrt{2}}{\sqrt{2}} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \frac{\sqrt{2}}{\sqrt{2}} = \frac{-1}{\sqrt{10}}\]

\[A_{21} = \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 0 \\ 2 \end{bmatrix} \frac{\sqrt{2}}{\sqrt{5}} = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{-1}{\sqrt{5}} \end{bmatrix} \frac{\sqrt{2}}{\sqrt{5}} = \frac{\sqrt{2}}{5}\]

\[A_{22} = \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 0 \\ 2 \end{bmatrix} \frac{\sqrt{5}}{\sqrt{5}} = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{-1}{\sqrt{5}} \end{bmatrix} \frac{\sqrt{5}}{\sqrt{5}} = -1\]

\[B_{11} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} \frac{\sqrt{2}}{\sqrt{2}} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \frac{\sqrt{2}}{\sqrt{2}} = 1\]

\[B_{12} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} \frac{\sqrt{2}}{\sqrt{2}} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \frac{\sqrt{2}}{\sqrt{2}} = \frac{-2}{\sqrt{10}}\]

\[B_{21} = \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} \frac{\sqrt{2}}{\sqrt{2}} = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{-1}{\sqrt{5}} \end{bmatrix} \frac{\sqrt{2}}{\sqrt{2}} = \frac{1}{\sqrt{10}}\]
\[B_{22} = \begin{bmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{-1}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} -1 & 2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} = -2\]

4. (a) As \(M = I - N\), \([M, N] = [I - N, N] = 0\) is obvious.

(b) 
\[
det \left( \lambda I - \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right) = det \left[ \begin{array}{cc} \lambda & -1 \\ -1 & \lambda \end{array} \right] = \lambda^2 - 1 = (\lambda + 1)(\lambda - 1) = 0
\]
\[\Rightarrow \lambda = \pm 1\]
So, the eigenvalues of \(N\) are \(\pm 1\).

\[
\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} b \\ a \end{bmatrix} = \pm 1 \cdot \begin{bmatrix} a \\ b \end{bmatrix} \Rightarrow a = \pm b
\]
\[\Rightarrow |1\rangle_N = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \text{ and } |−1\rangle_N = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}
\]
As we already know \(M\) and \(N\) commute, they share the same eigenvectors.

\[
\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0 \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix}
\]
\[
\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \end{bmatrix} = 2 \cdot \begin{bmatrix} 1 \\ -1 \end{bmatrix}
\]
So, the eigenvalues of \(M\) are 0 and 2, and corresponding normalized eigenvectors are

\[
|0\rangle_M = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \text{ and } |2\rangle_M = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.
\]

(c) 
\[
\left( \left( \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \right)^* \left( \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \right) \left( \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \right) \right)^* = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}
\]
\[
= \frac{1}{2} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ -2 & 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}
\]
and
\[
\left( \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \right)^t \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \left( \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \right)
= \frac{1}{2} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}
= \frac{1}{2} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}
= \frac{1}{2} \begin{bmatrix} -2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}
\]

Chapter 3
1. [3 pts]

(a) 0.4772 [1 pt]

(b) 0.5 - 0.4495 = 0.0505 [1 pt]

(c) \(P(|z| > 1.96) = 2 \times (0.5 - 0.4750) = 0.05\) [1 pt]

Chapter 5
1.
\[
\Psi(x + vt_0, t + t_0) = A \sin \left[ \frac{2\pi}{\lambda} \left( x + vt_0 - v(t + t_0) \right) \right]
= A \sin \left[ \frac{2\pi}{\lambda} \left( x + vt_0 - vt - vt_0 \right) \right] = A \sin \left[ \frac{2\pi}{\lambda} (x - vt) \right] = \Psi(x, t)
\]
\(\Psi(x, t)\) is the displacement at location \(x\) at time \(t\). As the speed is \(v\), this moves to the position “\(x + vt_0\)” time “\(t_0\)” later or at time “\(t + t_0\)”.

2.
\[
- \frac{\hbar^2}{2m} \frac{\partial^2 \Psi(x, t)}{\partial x^2} + V(x, t) \Psi(x, t) = i\hbar \frac{\partial \Psi(x, t)}{\partial t}
\]

3. We have
\[
- \frac{\hbar^2}{2m} \frac{\partial^2 \Psi_1(x, t)}{\partial x^2} + V(x, t) \Psi_1(x, t) = i\hbar \frac{\partial \Psi_1(x, t)}{\partial t}
\]
and
CHAPTER 5

4. (a) \( e^{i\theta} e^{i\phi} = (\cos \theta + i \sin \theta)(\cos \phi + i \sin \phi) = (\cos \theta \cos \phi - \sin \theta \sin \phi) + i(\sin \theta \cos \phi + \cos \theta \sin \phi) = \cos(\theta + \phi) + i \sin(\theta + \phi) = e^{i(\theta + \phi)} \)

(b) Our proof is by mathematical induction.
Consider \((\cos \theta + i \sin \theta)^k\). For \(k = 1\), we have \((\cos \theta + i \sin \theta)^1 = e^{i \cdot 1 \cdot \theta} = e^{i\theta}\). Hence, \((\cos \theta + i \sin \theta)^k = e^{ik\theta}\) holds for \(k = 1\).
Now suppose \((\cos \theta + i \sin \theta)^k = e^{ik\theta} = \cos(k\theta) + i \sin(k\theta)\) for some \(k\).
Then, \((\cos \theta + i \sin \theta)^{k+1} = (\cos \theta + i \sin \theta)^k (\cos \theta + i \sin \theta) = [\cos(k\theta) + i \sin(k\theta)] (\cos \theta + i \sin \theta) = \cos(k\theta) \cos \theta - \sin(k\theta) \sin \theta + i[\sin(k\theta) \cos \theta + \sin \theta \cos(k\theta)] = \cos((k+1)\theta) + i \sin((k+1)\theta) = e^{i(k+1)\theta}.
Therefore, \((e^{i\theta})^n = e^{in\theta}\) by induction.

5. (a)
\[
\int_{-\frac{a}{2}}^{\frac{a}{2}} \Psi^* \Psi \, dx = \int_{-\frac{a}{2}}^{\frac{a}{2}} A^* \cos \left( \frac{\pi x}{a} \right) e^{iEt/h} A \cos \left( \frac{\pi x}{a} \right) e^{-iEt/h} \, dx
\]
\[
= \int_{-\frac{a}{2}}^{\frac{a}{2}} A^* A \cos^2 \left( \frac{\pi x}{a} \right) \, dx = |A|^2 \int_{-\frac{a}{2}}^{\frac{a}{2}} \frac{1 + \cos \left( \frac{2\pi x}{a} \right)}{2} \, dx
\]
\[
= |A|^2 \left[ \frac{x}{2} + \frac{a}{2 \cdot 2\pi} \sin \left( \frac{2\pi x}{a} \right) \right]_{-\frac{a}{2}}^{\frac{a}{2}}
\]
\[
= |A|^2 \left[ \frac{a}{4} + \frac{a}{4\pi} \sin \pi - \frac{-a}{4} - \frac{a}{4\pi} \sin(-\pi) \right] = |A|^2 \frac{a}{2} = 1
\]
\[
\implies A = e^{i\theta} \sqrt{\frac{2}{a}}
\]

(b)
\[
-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \left[ A \cos \left( \frac{\pi x}{a} \right) e^{-iEt/h} \right] = i\hbar \frac{\partial}{\partial t} \left[ A \cos \left( \frac{\pi x}{a} \right) e^{-iEt/h} \right]
\]
\[
\implies -\frac{\hbar^2}{2m} A \left[ - \left( \frac{\pi}{a} \right)^2 \right] \cos \left( \frac{\pi x}{a} \right) e^{-iEt/h} = i\hbar A \cos \left( \frac{\pi x}{a} \right) \left( \frac{-iE}{\hbar} \right) e^{-iEt/h}
\]
\[
\implies -\frac{\hbar^2}{2m} \left[ - \left( \frac{\pi}{a} \right)^2 \right] = i\hbar \left( \frac{-iE}{\hbar} \right) = E \implies E = \frac{\pi^2 \hbar^2}{2ma^2} \quad (13.25)
\]

(c)
\[
\langle P \rangle = \int_{-\frac{a}{2}}^{\frac{a}{2}} A^* \cos \left( \frac{\pi x}{a} \right) e^{iEt/h} \left( -i\hbar \frac{\partial}{\partial x} \right) A \cos \left( \frac{\pi x}{a} \right) e^{-iEt/h} \, dx
\]
\[
= \frac{2}{a} (-i\hbar) \int_{-\frac{a}{2}}^{\frac{a}{2}} \cos \left( \frac{\pi x}{a} \right) \frac{-\pi}{a} \sin \left( \frac{\pi x}{a} \right) \, dx
\]
\[
= \frac{i\pi \hbar}{a^2} \int_{-\frac{a}{2}}^{\frac{a}{2}} 2 \sin \left( \frac{\pi x}{a} \right) \cos \left( \frac{\pi x}{a} \right) \, dx
\]
\[
\begin{align*}
&= \frac{i\pi \hbar}{a^2} \int_{-\frac{a}{2}}^{\frac{a}{2}} \sin \left( \frac{2\pi x}{a} \right) \, dx = \frac{i\pi \hbar}{a^2} \left[ -a \frac{\cos \left( \frac{2\pi x}{a} \right) }{2\pi} \right]_{-\frac{a}{2}}^{\frac{a}{2}} \\
&= \frac{i\pi \hbar}{a^2} \frac{1}{2\pi} \left[ \cos \pi - \cos (-\pi) \right] = \frac{i\pi \hbar}{a^2} \frac{1}{2\pi} \left[ -1 - (-1) \right] = 0
\end{align*}
\]

(d)

\[
\langle x^2 \rangle = \frac{2}{a} \int_{-\frac{a}{2}}^{\frac{a}{2}} x^2 \cos \left( \frac{\pi x}{a} \right) e^{+iE_t/\hbar} \, x^2 A \cos \left( \frac{\pi x}{a} \right) e^{-iE_t/\hbar} \, dx
\]

Let \( y = \frac{\pi}{a} x \), so that \( dx = \frac{a}{\pi} dy \).

\[
\begin{align*}
&= \frac{2}{a} \int_{-\frac{a}{2}}^{\frac{a}{2}} \left( \frac{a}{\pi} \right)^2 y^2 \cos^2 \left( \frac{\pi y}{a} \right) y \, dy = \frac{2}{a} \left( \frac{a}{\pi} \right)^3 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} y^2 \cos^2 y \, dy \\
&= \frac{2}{a} \left( \frac{a}{\pi} \right)^3 \left[ \frac{y^3}{6} + \left( \frac{y^2}{4} - \frac{1}{8} \right) \sin 2y + \frac{y \cos 2y}{4} \right]_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \\
&= \frac{2}{a} \left( \frac{a}{\pi} \right)^3 \left[ \frac{y^3}{6} + \frac{y \cos 2y}{4} \right]_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \\
&= \frac{2}{a} \left( \frac{a}{\pi} \right)^3 \left[ 2 \left( \frac{\pi}{2} \right)^3 + \frac{2 \pi \cos \pi}{4} \right] = \frac{2a^2}{\pi^3} \left[ \frac{\pi^3}{24} - \frac{\pi}{4} \right] \\
&= \frac{a^2}{12} \left( 1 - \frac{6}{\pi^2} \right);
\end{align*}
\]

where we used the formula

\[
\int y^2 \cos^2 y \, dy = \frac{y^3}{6} + \left( \frac{y^2}{4} - \frac{1}{8} \right) \sin 2y + \frac{y \cos 2y}{4} + C.
\]

(e)

\[
\begin{align*}
\langle P^2 \rangle &= \int_{-\frac{a}{2}}^{\frac{a}{2}} A^* \cos \left( \frac{\pi x}{a} \right) e^{+iE_t/\hbar} \left( \frac{\hbar}{a} \frac{\partial}{\partial x} \right)^2 A \cos \left( \frac{\pi x}{a} \right) e^{-iE_t/\hbar} \, dx \\
&= -\hbar^2 \left[ - \left( \frac{\pi}{a} \right)^2 \right] \int_{-\frac{a}{2}}^{\frac{a}{2}} A^* \cos \left( \frac{\pi x}{a} \right) e^{+iE_t/\hbar} A \cos \left( \frac{\pi x}{a} \right) e^{-iE_t/\hbar} \, dx.
\end{align*}
\]
But, the normalization condition gives
\[
\int_{-\frac{a}{2}}^{\frac{a}{2}} A^* \cos\left(\frac{\pi x}{a}\right) e^{+iEt/\hbar} A \cos\left(\frac{\pi x}{a}\right) e^{-iEt/\hbar} \, dx = 1.
\]
Therefore, we get
\[
\langle P^2 \rangle = \left(\frac{\pi \hbar}{a}\right)^2.
\]
6. (a) As \(e^{i\theta}A\) for any real number \(\theta\) can replace \(A\), the wavefunction is not unique in that sense.

(b) The normalization constant \(e^{i\theta}A\) enters into the computations of \(\Psi^*(x,t)\Psi(x,t)\), \(\langle P \rangle\), \(\langle P^2 \rangle\), and \(\langle x^2 \rangle\) as \((e^{i\theta}A)^\ast \left(e^{i\theta}A\right)\). However, \((e^{i\theta}A)^\ast \left(e^{i\theta}A\right) = e^{-i\theta}A^*e^{i\theta}A = A^*A\), and the value of \(\theta\) does not affect these physical observables.

**Chapter 6**

1. 
\[-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \psi(x) + V(x)\psi(x) = E\psi(x)\]

2. Plugging \(V(x,t) = V(x)\) and \(\Psi(x,t) = \psi(x)e^{-iEt/\hbar}\) into the time-dependent Schrödinger Equation, we get
\[
-\frac{\hbar^2}{2m} \frac{\partial^2 \Psi(x,t)}{\partial x^2} + V(x,t)\Psi(x,t) = i\hbar \frac{\partial \Psi(x,t)}{\partial t}
\]
\[
\Longrightarrow -\frac{\hbar^2}{2m} \frac{\partial^2 \psi(x)e^{-iEt/\hbar}}{\partial x^2} + V(x)\psi(x)e^{-iEt/\hbar} = i\hbar \frac{\partial \psi(x)e^{-iEt/\hbar}}{\partial t}
\]
\[
\Longrightarrow -\frac{\hbar^2}{2m} e^{-iEt/\hbar} \frac{\partial^2 \psi(x)}{\partial x^2} + V(x)\psi(x)e^{-iEt/\hbar} = i\hbar \left(-iE/\hbar\right) e^{-iEt/\hbar} \psi(x)
\]
\[
\Longrightarrow -\frac{\hbar^2}{2m} \frac{\partial^2 \psi(x)}{\partial x^2} + V(x)\psi(x) = E\psi(x).
\]

3. The first derivative of \(\sin \phi\) is \(\cos \phi\). Both \(\sin \phi\) and \(\cos \phi\) are continuous, and \(\sin^2 \phi\) is obviously integrable over the interval \([0,2\pi]\). Furthermore, \(\sin(\phi + 2m\pi) = \sin \phi\) and \(\cos(\phi + 2m\pi) = \cos \phi\) for all integers \(m\). Hence, the condition of single-valuedness is also satisfied.

On the other hand, if \(\psi(\phi) = \phi\), we have \(\psi(0) = 0\) and \(\psi(2\pi) = 2\pi \neq 0 = \psi(0)\) despite the fact that \(\phi = 0\) and \(\phi = 2\pi\) signify the same position in space. Hence, this function is not single-valued and cannot be a wavefunction as such.
Chapter 7

1. Same as Chapter 5 Problem 5. So, see the answers to Chapter 5 Problem 5.

2. Same as Chapter 5 Problem 6. So, see the answers to Chapter 5 Problem 6.

3. 
\[ -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \psi(x) = E \psi(x) \implies \frac{d^2}{dx^2} \psi(x) = -\frac{2mE}{\hbar^2} \psi(x) \implies \psi(x) = Ae^{\pm i\sqrt{\frac{2mE}{\hbar^2}}x} \]

4. Let \( k = \sqrt{\frac{2mE}{\hbar}} \). Then, the general solution is 
\[ \psi(x) = C \sin kx + D \cos kx. \]

Taking the boundary conditions into account,
\[ \psi \left( \pm \frac{a}{2} \right) = 0 \implies \begin{cases} C \sin \frac{ka}{2} + D \cos \frac{ka}{2} = 0 \\ -C \sin \frac{ka}{2} + D \cos \frac{ka}{2} = 0 \end{cases} \implies \begin{cases} D \cos \frac{ka}{2} = 0 \\ C \sin \frac{ka}{2} = 0 \end{cases}. \]

Because \( \frac{ka}{2} > 0 \), we have
\[ D \cos \frac{ka}{2} = 0 \iff D = 0 \text{ or } \frac{ka}{2} = \left( n + \frac{1}{2} \right) \pi \text{ for } n = 0, 1, 2, \ldots \]

and
\[ C \sin \frac{ka}{2} = 0 \iff C = 0 \text{ or } \frac{ka}{2} = n\pi \text{ for } n = 1, 2, 3, \ldots. \]

Note that \( C = D = 0 \) leads to a zero wavefunction, and \( \frac{ka}{2} = \left( n + \frac{1}{2} \right) \pi \) and \( \frac{ka}{2} = n\pi \) cannot hold simultaneously.

Hence, we have either
\[ \frac{ka}{2} = \left( n + \frac{1}{2} \right) \pi \text{ (} n = 0, 1, 2, \ldots \text{) and } C = 0 \]
or
\[ \frac{ka}{2} = n\pi \text{ (} n = 1, 2, 3, \ldots \text{) and } D = 0. \]
When, $C = n = 0$, $k = \frac{\pi}{a}$ and

$$\psi(x) = D \cos \frac{\pi x}{a}.$$ 

When $D = 0$ and $n = 1$, $k = \frac{2\pi}{a}$ and

$$\psi(x) = C \sin \frac{2\pi x}{a}.$$ 

Let us now calculate the real and positive normalization constants.

For $D$,

$$\int_{-\frac{a}{2}}^{\frac{a}{2}} \psi^*(x)\psi(x) \, dx = \int_{-\frac{a}{2}}^{\frac{a}{2}} |D|^2 \cos^2 \frac{\pi x}{a} \, dx = |D|^2 \int_{-\frac{a}{2}}^{\frac{a}{2}} \frac{1 + \cos \frac{2\pi x}{a}}{2} \, dx$$

$$= \frac{|D|^2}{2} \left[ x + \frac{a}{2\pi} \sin \frac{2\pi x}{a} \right]_{-\frac{a}{2}}^{\frac{a}{2}} = \frac{|D|^2}{2} \cdot a = 1 \implies D = \sqrt{\frac{2}{a}}.$$ 

For $C$,

$$\int_{-\frac{a}{2}}^{\frac{a}{2}} \psi^*(x)\psi(x) \, dx = \int_{-\frac{a}{2}}^{\frac{a}{2}} |C|^2 \sin^2 \frac{2\pi x}{a} \, dx = |C|^2 \int_{-\frac{a}{2}}^{\frac{a}{2}} \frac{1 - \cos \frac{4\pi x}{a}}{2} \, dx$$

$$= \frac{|C|^2}{2} \left[ x - \frac{a}{4\pi} \sin \frac{4\pi x}{a} \right]_{-\frac{a}{2}}^{\frac{a}{2}} = \frac{|C|^2}{2} \cdot a = 1 \implies C = \sqrt{\frac{2}{a}}.$$ 

So, the functions we have found are

$$\psi_1(x) = \sqrt{\frac{2}{a}} \cos \frac{\pi x}{a} \quad \text{and} \quad \psi_2(x) = \sqrt{\frac{2}{a}} \sin \frac{2\pi x}{a}.$$ 

5. Because $H(x) = K(t)$ for all $(x, t) \in \mathbb{R}^2 := \mathbb{R} \times \mathbb{R}$, $H(0) = K(t)$ and $H(x) = K(0)$ for all $(x, t) \in \mathbb{R}^2$. Combining this with $H(x) = K(t)$, we get $H(0) = K(0)$, which in turn implies $H(x) = K(t) = H(0)(= K(0))$ for all $(x, t) \in \mathbb{R}^2$.

6. Here is a summary of the full solution given in Section 7.2. The Schrödinger equations are

$$-\hbar^2 \frac{d^2 \psi(x)}{2m \, dx^2} = E \psi(x) \quad x \leq 0$$

$$-\hbar^2 \frac{d^2 \psi(x)}{2m \, dx^2} + V_0 \psi(x) = E \psi(x) \quad x > 0$$
with the corresponding general solutions

\[ \psi_-(x) = Ae^{ik_1x} + Be^{-ik_1x} \text{ where } k_1 = \frac{\sqrt{2mE}}{\hbar} \text{ and } x < 0 \]

and

\[ \psi_+(x) = Ce^{k_2x} + De^{-k_2x} \text{ where } k_2 = \frac{\sqrt{2m(V_0 - E)}}{\hbar} \text{ and } x > 0. \]

As \( \psi_+ \) has to be normalizable/integrable in the region \( x > 0 \), we set \( C = 0 \) to obtain

\[ \psi_+(x) = De^{-k_2x} \text{ where } k_2 = \frac{\sqrt{2m(V_0 - E)}}{\hbar} \text{ and } x > 0. \]

Now use two continuity conditions

\[ \psi_+(0) = \psi_-(0) \quad \text{and} \quad \psi'_+(0) = \psi'_-(0) \]

to obtain

\[
\begin{cases} 
\frac{ik_2}{k_1} D = A - B \\
D = A + B 
\end{cases}
\]

with the solution

\[ A = \frac{1}{2} \left( 1 + \frac{ik_2}{k_1} \right) D \]

\[ B = \frac{1}{2} \left( 1 - \frac{ik_2}{k_1} \right) D. \]

This gives

\[ \psi(x) = \begin{cases} 
\frac{D}{2}(1 + ik_2/k_1)e^{ik_1x} + \frac{D}{2}(1 - ik_2/k_1)e^{-ik_1x} & x \leq 0 \\
De^{-k_2x} & x > 0 
\end{cases} \]
10. (a) 
\[ -\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} = E\psi(x) \quad x < 0 \]
\[ -\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + V_0\psi(x) = E\psi(x) \quad x > 0 \]

(b) Continuity of the wavefunction \( \psi(x) \) at \( x = 0 \) gives
\[ Ae^{ik_1x} + Be^{-ik_1x} = Ce^{ik_2x} \quad (\text{at } x = 0) \quad \text{or} \quad A + B = C. \]
Continuity of the first derivative \( \frac{d\psi(x)}{dx} \) gives
\[ ik_1Ae^{ik_1x} - ik_1Be^{-ik_1x} = ik_2Ce^{ik_2x} \quad (\text{at } x = 0) \quad \text{or} \quad k_1(A - B) = k_2C. \]
The rest is Arithmetic, and we get
\[ C = \frac{2k_1}{k_1 + k_2} A. \]

(c) 
\[ R = \frac{B^*B}{A^*A} = \left(\frac{k_1 - k_2}{k_1 + k_2}\right)^2 \]

11.

12. (a) 
\[ \begin{cases} \frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} = E\psi(x) & x < 0 \\ \frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + V_0\psi(x) = E\psi(x) & x > 0 \end{cases} \]

(b) 
\[ \begin{cases} Ae^{ik_1x} + Be^{-ik_1x} & x < 0 \\ Ce^{ik_2x} + De^{-ik_2x} & x > 0 \end{cases} \]

(c) \( A = 0 \)

(d) 
\[ \begin{cases} \psi(x)|_{0-} = \psi(x)|_{0+} \\ \psi'(x)|_{0-} = \psi'(x)|_{0+} \end{cases} \quad \Rightarrow \quad \begin{cases} B = C + D \\ -ik_1B = ik_2(C - D) \end{cases} \]
\[ \Rightarrow \begin{cases} C + D = \frac{B}{k_1} \\ C - D = -\frac{k_1}{k_2}B \end{cases} \quad \Rightarrow \quad \begin{cases} C = \frac{k_2 - k_1}{2k_2}B \\ D = \frac{k_2}{2k_2}B. \end{cases} \]
(e) \[ R = \frac{C^* \, C}{D^* \, D} = \left( \frac{k_2 - k_1}{2k_2} \right)^2 \frac{B^* \, B}{\left( \frac{k_2 + k_1}{2k_2} \right)^2} = \left( \frac{k_1 - k_2}{k_1 + k_2} \right)^2 \]

13. (a) \[-\frac{\hbar^2}{2m} \frac{d^2 \psi(x)}{dx^2} = E \psi(x) \quad x < -a \quad \text{and} \quad x > 0 \]
\[-\frac{\hbar^2}{2m} \frac{d^2 \psi(x)}{dx^2} + V_0 \psi(x) = E \psi(x) \quad -a < x < 0 \]

(b) i. Since there is nothing that reflects the wave in the region \( x > 0 \), there should not be a wave traveling to the left in this region. This means \( H = 0 \) as \( e^{-ik_1 x} e^{-i\omega t} = e^{-i(kx + \omega t)} \) represents a wave traveling toward the left.

ii. Continuity of the wavefunction \( \psi(x) \) at \( x = 0 \) gives \[ Ce^{k_2 x} + De^{-k_2 x} = Ge^{ik_1 x} \quad \text{(at} \ x = 0) \quad \implies \quad C + D = G. \]

Continuity of the first derivative \( \frac{d\psi(x)}{dx} \) at \( x = 0 \) gives \[ k_2Ce^{k_2 x} - k_2De^{-k_2 x} = ik_1 Ge^{ik_1 x} \quad \text{(at} \ x = 0) \quad \implies \quad k_2(C-D) = ik_1 G. \]

iii. In the region \( x > 0 \), \( \psi^*(x) \psi(x) = G^* e^{-ik_1 x} G e^{ik_1 x} = |G|^2 \), and \( \frac{d\psi^*(x)\psi(x)}{dx} \bigg|_{0+} = 0 \). Hence, we also need \( \frac{d\psi^*(x)\psi(x)}{dx} \bigg|_{0-} = 0 \).

Suppose \( C = 0 \). Then, we have \( \psi(x) = De^{-k_2 x} \) in the region \( -a < x < 0 \). This gives \( \frac{d\psi^*(x)\psi(x)}{dx} \bigg|_{0-} = \frac{dD^* De^{-2k_2 x}}{dx} \bigg|_{0-} = (-2k_2)|D|^2 e^{-2k_2 x} \bigg|_{0-} = -2k_2|D|^2 = 0 \quad \implies \quad D = 0 \), and both \( C \) and \( D \) are zero. Similarly, \( D = 0 \) forces \( C \) to be zero.
If \( \psi(x) \) is zero inside the barrier, the continuity conditions on \( \psi(x) \) and \( \frac{d\psi(x)}{dx} \) at \( x = -a, 0 \) implies \( \psi(x) = 0 \) everywhere, which is a physically meaningless solution. Therefore, neither \( C \) nor \( D \) can be zero.

14. (a) \[ \int_{-a/2}^{+a/2} \psi^* \psi \, dx = 2|A|^2 \int_{0}^{a/2} \sin^2 kx \, dx = 1 \implies A = \pm \sqrt{\frac{2}{a}} i \]
\[ \psi_2(x) = \pm \sqrt{\frac{2}{a}} \sin \frac{2\pi x}{a} \] (one of these)

(b)
\[ \int_{-a/2}^{+a/2} \Psi^* x \Psi \, dx = \int_{-a/2}^{+a/2} A_2^* \sin (k_2 x) e^{iE_2 t/\hbar} \cdot x \cdot A_2 \sin (k_2 x) e^{-iE_2 t/\hbar} \, dx \]
\[ = |A_2|^2 \int_{-a/2}^{+a/2} x \sin^2 (k_2 x) \, dx = 0 \quad \text{(We have an odd function.)} \]

(c)
\[ \frac{-\hbar^2}{2m} \frac{d^2}{dx^2} A_2 \sin (k_2 x) = \frac{2\pi^2 \hbar^2}{ma^2} A_2 \sin (k_2 x) \implies E_2 = \frac{2\pi^2 \hbar^2}{ma^2} \]

16. (a) \[-\frac{\hbar^2}{2m} \frac{d^2 \psi(x)}{dx^2} = E \psi(x) \]
(b) i. \[\frac{d^2 \psi(x)}{dx^2} = \frac{d}{dx} (AB \cos (Bx)) = -AB^2 \sin (Bx) \]
\[\implies -\frac{\hbar^2}{2m} (AB^2) \sin (Bx) = AE \sin (Bx) \]
\[\implies B^2 = \frac{2mE}{\hbar^2} \]
\[\implies B = \sqrt{\frac{2mE}{\hbar^2}} \]
ii. \[\sin \left( \pm \frac{a^2}{2} B \right) = 0 \implies \frac{a}{2} B = \pi \implies B = \frac{2\pi}{a} \]
iii. \[B^2 = \frac{4\pi^2}{a^2} = \frac{2mE}{\hbar^2} \implies E = \frac{2\pi^2 \hbar^2}{ma^2} \]
iv. \[\int_{-a/2}^{a/2} A \sin (Bx) \cdot A \sin (Bx) \, dx = 1 \implies A = \sqrt{2/a} \]

17. (a) \[-\frac{\hbar^2}{2m} \frac{d^2 \psi(x)}{dx^2} = E \psi(x) \]
(b) i. \[\psi(0) = A \sin 0 + B \cos 0 = B = 0 \]
ii. \[\psi(L) = A \sin kL = 0 \implies kL = n\pi \implies k = \frac{n\pi}{L} \]
iii. \[\int_0^L \psi^* \psi \, dx = \int_0^L |A|^2 \sin^2 k_n x \, dx = \left[ \frac{x}{2} \sin^2 \frac{2k_n x}{4k_n} \right]_0^L = \frac{|A|^2 L}{2} = 1 \implies A_n = \sqrt{\frac{2}{L}} \]
(c) \[k_n = \frac{n\pi}{L} = \frac{\sqrt{2mE_n}}{\hbar} \implies E_n = \left( \frac{n\pi}{L} \right)^2 \frac{1}{2m} = \frac{n^2 \hbar^2}{8mL^2} \]

18. \[\frac{df(x)}{dx} = \sum_{j=1}^{\infty} ja_j x^{j-1} = \sum_{j=0}^{\infty} (j + 1) a_{j+1} x^j \]
\[\implies \frac{df(x)}{dx} + f(x) = \sum_{j=0}^{\infty} [a_j + (j + 1) a_{j+1}] x^j = 3x^2 + 8x + 3 \]
\[\implies a_j + (j + 1) a_{j+1} = 0 \quad \text{for} \quad j \geq 3. \]
19. The full wave function is $\psi_0 e^{-iE_0t/\hbar}$.

$$\langle x \rangle = \int_{-\infty}^{+\infty} (\psi_0 e^{-iE_0t/\hbar})^* x (\psi_0 e^{-iE_0t/\hbar}) dx$$

$$= \int_{-\infty}^{+\infty} \psi_0^* x \psi_0 dx$$

$$= \int_{-\infty}^{+\infty} x A_0^* e^{-u^2/2} e^{-u^2/2} dx$$

$$= |A_0|^2 \int_{-\infty}^{+\infty} x e^{-|Cm|^{1/2}/\hbar x^2} \, dx$$

as the integrand is an odd function.

20. (a) The full ground state wavefunction $\Psi_0$ is given by $\Psi_0 = \psi_0 e^{-iEt/\hbar}$. Therefore,

$$\langle x \rangle = \int_{-\infty}^{+\infty} \psi_0^* x \psi_0 \, dx = \int_{-\infty}^{+\infty} \psi_0^* e^{iEt/\hbar} x \psi_0 e^{-iEt/\hbar} \, dx = \int_{-\infty}^{+\infty} \psi_0^* x \psi_0 \, dx$$

$$= \int_{-\infty}^{+\infty} A_0^* e^{-u^2/2} x A_0 e^{-u^2/2} \, dx = A_0^* A_0 \int_{-\infty}^{+\infty} x e^{-u^2} \, dx$$

$$= A_0^* A_0 \int_{-\infty}^{+\infty} x e^{-|Cm|^{1/2}/\hbar x^2} \, dx = 0$$

as the integrand is an odd function.

(b) $\int_{-\infty}^{+\infty} \psi_0^* \psi_1 \, dx = \int_{-\infty}^{+\infty} A_0^* e^{-u^2/2} A_1 u e^{-u^2/2} \, dx = A_0^* A_1 \int_{-\infty}^{+\infty} u e^{-u^2} \, dx$

$$= A_0^* A_1 \frac{1}{\sqrt{\alpha}} \int_{-\infty}^{+\infty} u e^{-u^2} \, du = 0$$

because $ue^{-u^2}$ is an odd function.

(c) Because the Hamiltonian is a Hermitian operator, its eigenvectors/functions/kets associated with different eigenvalues are orthogonal to each other.

(d) Vibration of diatomic molecules can be approximated by the simple harmonic oscillator. As the lowest allowed energy is strictly positive, there is always some residual motion. This nonzero kinetic energy prevents the system from reaching zero degrees kelvin.
22. Some examples are as follows.

- Energy levels are continuous in classical mechanics, while they are discrete in quantum mechanics.
- There is barrier penetration only in quantum mechanics.
- The position of the particle is probabilistically determined in quantum mechanics. But, it takes one unique value in classical mechanics.

Chapter 8

1.

Chapter 9

1. Proceeding as in (9.12),

\[
[L_z, L_x] = [xp_y - yp_x, yp_z - zp_y] \\
= [xp_y, yp_z] - [xp_y, yp_z] - [yp_x, yp_z] + [yp_x, yp_z] \\
= x[p_y, y] + y[x, p_z] + yx[p_y, p_z] + [x, y]p_y p_z \\
- (x[p_y, z]p_y + z[x, p_y]p_y + xz[p_y, p_y] + [x, z]p_y p_y) \\
- (y[p_x, y]p_z + y[y, p_z]p_x + yy[p_x, p_z] + [y, y]p_x p_z) \\
+ y[p_z, z]p_y + z[y, p_y]p_z + zy[p_x, p_y] + [y, z]p_x p_y \\
= x(-i\hbar) p_z + y \cdot 0 \cdot p_y + yx \cdot 0 + 0 \cdot p_y p_z \\
- (x \cdot 0 \cdot p_y + z \cdot 0 \cdot p_y + yz \cdot 0 + 0 \cdot p_y p_y) \\
- (y \cdot 0 \cdot p_z + y \cdot 0 \cdot p_z + yz \cdot 0 + 0 \cdot p_z p_z) \\
+ y \cdot 0 \cdot p_y + z(i\hbar) p_z + zy \cdot 0 + 0 \cdot p_z p_y \\
= i\hbar(zp_x - xp_z) = i\hbar L_y.
\]

In spherical coordinates as in (9.11),

\[
[L_z, L_x] = L_z L_x - L_x L_z \\
= -i\hbar \partial_\phi (\sin \phi \partial_\theta + \cot \theta \cos \phi \partial_\phi) \\
- i\hbar (\sin \phi \partial_\theta + \cot \theta \cos \phi \partial_\phi) (-i\hbar) \partial_\phi
\]
\[-(i\hbar)^2 \left( \cos \phi \partial_\theta + \sin \phi \partial_\phi \partial_\theta - \cot \theta \sin \phi \partial_\phi + \cot \theta \cos \phi \partial_\phi^2 \right) \]
\[+ (i\hbar)^2 \left( \sin \phi \partial_\phi \partial_\theta + \cot \theta \cos \phi \partial_\phi^2 \right) \]
\[= i\hbar \left[ i\hbar \left( - \cos \phi \partial_\theta + \cot \theta \sin \phi \partial_\phi \right) \right] \]
\[= i\hbar \left[ i\hbar \left( - \cos \phi \frac{\partial}{\partial \theta} + \cot \theta \sin \phi \frac{\partial}{\partial \phi} \right) \right] = i\hbar L_y.\]

Chapter 10

1. 
2. 
3. 
4. (a) \[l = 0, 1, 2\]
   (b) \[m_l = 0 \text{ for } l = 0, \]
   \[m_l = -1, 0, 1 \text{ for } l = 1,\]
   \[m_l = -2, -1, 0, 1, 2 \text{ for } l = 2\]
   (c) 9 degenerate states
5. 
   \[E_3 - E_1 = -\frac{\mu e^4}{(4\pi \varepsilon_0)^2 2\hbar^2 3^2} - \left( -\frac{\mu e^4}{(4\pi \varepsilon_0)^2 2\hbar^2 1^2} \right) = \frac{8\mu e^4}{(4\pi \varepsilon_0)^2 2\hbar^2 9} = \frac{\mu e^4}{36(\pi \varepsilon_0)^2 \hbar^2}\]
6. (a) \(e^{ik(2\pi m)} = e^{ik(2\pi n)}\) for all \((m, n)\). Consider \((m, n) = (1, 0)\). Then, we have
   \[|e^{ik2\pi}| = |e^{i(a+b)2\pi}| = |e^{i02\pi}|e^{-b2\pi} = 1 \implies b = 0.\]
   Now consider \(m = n + 1.\)
   \[e^{ik2\pi} = 1 \implies k = 0, \pm 1, \pm 2, \cdots \text{ (necessary condition)}\]
   But, this is also sufficient.
(b) For each \( n, l = 0, 1, 2, \cdots, n - 1 \), and \( m_l = \underbrace{-l, -l + 1, \cdots, l - 1, l}_{2l+1} \)

\[
\sum_{l=0}^{n-1} (2l + 1) = 2 \sum_{l=0}^{n-1} l + \sum_{l=0}^{n-1} 1 = 2 \cdot \frac{n(n-1)}{2} + n = n^2 \quad \text{Q.E.D.}
\]

7. (a) i. \( l = 3 \)

ii. \( \|L\| = \sqrt{3(3+1)}h = 2\sqrt{3}h \)

iii. \( 2l + 1 = 2 \cdot 3 + 1 = 7 \)

iv. \( 3h \)

(b) i.

\[
\begin{align*}
P(0 \leq r < a_0) &= \int_0^{a_0} \int_0^{2\pi} \int_0^{\pi} \left( \frac{1}{\sqrt{\pi}} \left( \frac{1}{a_0} \right)^{3/2} e^{-r/a_0} \right)^* \left( \frac{1}{\sqrt{\pi}} \left( \frac{1}{a_0} \right)^{3/2} e^{-r/a_0} \right) r^2 dr d\theta d\phi \\
&= \frac{1}{\pi a_0^3} \int_0^{a_0} \int_0^{2\pi} \int_0^{\pi} e^{-2r/a_0} r^2 dr d\theta d\phi = \frac{1}{\pi a_0^3} (4\pi) \int_0^{a_0} e^{(-2/a_0)r} dr \\
&= \frac{4}{a_0^3} \left\{ e^{-2r/a_0} \left[ \frac{r^2}{-2/a_0} - \frac{2r}{(-2/a_0)^2} + \frac{2}{(-2/a_0)^3} \right] \right\}_0^{a_0} \\
&= \frac{4}{a_0^3} \left\{ e^{-2} \left[ -\frac{1}{2} a_0^3 - \frac{1}{2} a_0^3 - \frac{1}{4} a_0^3 \right] - \left[ -\frac{1}{4} a_0^3 \right] \right\} \\
&= 4e^{-2} \left[ -\frac{2 - 2 - 1}{4} \right] + 1 = 1 - 5e^{-2}
\end{align*}
\]

ii.

\[
\begin{align*}
< r > &= \int_0^{\infty} \int_0^{2\pi} \int_0^{\pi} \left( \frac{1}{\sqrt{\pi}} \left( \frac{1}{a_0} \right)^{3/2} e^{-r/a_0} \right)^* \left( \frac{1}{\sqrt{\pi}} \left( \frac{1}{a_0} \right)^{3/2} e^{-r/a_0} \right) r^2 dr d\theta d\phi \\
&= \frac{1}{\pi a_0^3} \int_0^{\infty} \int_0^{2\pi} \int_0^{\pi} e^{-2r/a_0} r^3 dr d\theta d\phi = \frac{1}{\pi a_0^3} (4\pi) \int_0^{\infty} r^3 e^{(-2/a_0)r} dr
\end{align*}
\]
\[ \frac{4}{a_0^3} \left( e^{-2r/a_0} \left[ \frac{r^3}{(-2/a_0)^2} + \frac{6r}{(-2/a_0)^3} - \frac{6}{(-2/a_0)^4} \right] \right) \left. \right|_0^\infty = \frac{4}{a_0^3} \cdot 1 \cdot \frac{6a_0^4}{4 \cdot 4} = \frac{3}{2} a_0 \]

(c)

\[ E_{Z,n} = -\frac{Z^2 \mu e^4}{(4\pi \varepsilon_0)^2 2\hbar^2 n^2} = -\left( \frac{Z^2 \mu e^4}{32\pi^2 \varepsilon_0^2 \hbar^2} \right) \frac{1}{n^2} \Rightarrow E_{n,3} = -\left( \frac{9\mu e^4}{32\pi^2 \varepsilon_0^2 \hbar^2} \right) \frac{1}{n^2} \]

Hence,

\[ E_{3,3} - E_{1,3} = -\left( \frac{9\mu e^4}{32\pi^2 \varepsilon_0^2 \hbar^2} \right) \left( \frac{1}{9} - 1 \right) = -\left( \frac{9\mu e^4}{32\pi^2 \varepsilon_0^2 \hbar^2} \right) \left( -\frac{8}{9} \right) = \frac{\mu e^4}{4\pi^2 \varepsilon_0^2 \hbar^2} \]

Chapter 11

1. (a) The probability of the first device getting an up-spin is \( \left( \frac{1}{\sqrt{2}} \right)^2 = 0.5 \). After the measurement, the state is

\[ |\uparrow\rangle = \frac{1}{\sqrt{2}} |\uparrow\rangle_y + \frac{1}{\sqrt{2}} |\downarrow\rangle_y. \]

Hence, the probability that the second device measures a down-spin in the \( y \)-direction is also \( \left( \frac{1}{\sqrt{2}} \right)^2 = 0.5 \). Therefore, the answer is \( 0.5 \times 0.5 = 0.25 \).

(b) The final spin state is \( |\downarrow\rangle_y \) due to the fundamental postulates of quantum mechanics.

Chapter 12

1.
Subject Index

This is a work in progress. A very slow progress it is.

E
Erwin Schrödinger, 111

N
Newtonian Mechanics, 13

Q
Quantum Mechanics, 13

R
Relativity, 13

T
traveling wave, 111