**Groups, Representations, and Quantum Theory**

*Karthik Prasad*

*5/24/2025*

These notes were written for a course given at the Ross Mathematics Program in Ohio during Summer 2025. They cover abstract linear algebra, basic principles of quantum mechanics, and basic applications of representation theory in quantum mechanics. The end goal of these notes is that the reader emerges with an idea of the algebraic perspective on quantum mechanics and a newfound appreciation and understanding of the power of symmetry in physics. There are no formal prerequisites - however, exposure to linear algebra and classical mechanics is helpful.

The first two weeks of the course (first six sections) constitute various topics in a first course in abstract linear algebra. The third week covers basic quantum axioms and key principles of symmetry in quantum mechanics, highlighted through the elementary examples of $U(1)$ and $SU(2)$. The fourth week covers the official formalization of the ideas explored in the third week via introducing Lie Algebras. The fifth week attempts to apply the methods described in the fourth week to understand the physical quantity of spin. The sxith and final week involves composite quantum systems. Given time, there are three independent elective options - an introduction to quantum computing, a conceptual overview of quantum field theory, and solving the Hydrogen atom.

# Contents

**1 Experimental Basis of Quantum Mechanics** **7**
- 1.1 Blackbody Radiation . . . 7
- 1.2 Photoelectric Effect . . . 8
- 1.3 Hydrogen Spectrum . . . 9
- 1.4 Double-Slit Experiment . . . 10
- 1.5 Stern-Gerlarch Experiment . . . 12
- 1.6 Overall Conclusions . . . 12

**2 Linear Algebra I** **13**
- 2.1 Vector Spaces . . . 13
- 2.2 Subspaces . . . 14
- 2.3 Direct Sums . . . 15
- 2.4 Bases and Dimension . . . 17

**3 Linear Algebra II** **21**
- 3.1 Linear Maps . . . 21
- 3.2 Matrices . . . 24
- 3.3 Dual Space . . . 25
- 3.4 Bilinear Forms . . . 26
- 3.5 Tensor Products . . . 28

**4 Linear Algebra III** **30**
4.1 Permutations . . . 30
4.2 Multilinear Forms . . . 31
4.3 Permutation Action on Multilinear Forms . . . 32
4.4 Alternating Multilinear Forms . . . 33

**5 Linear Algebra IV** **34**
5.1 Alternating $n$-Linear Forms . . . 34
5.2 Determinants . . . 34
5.3 Geometric Interpretation of the Determinant . . . 36
5.4 Eigenvectors and Eigenvalues . . . 36

**6 Linear Algebra V** **41**
6.1 Inner Product Spaces . . . 41
6.2 Adjoint of a Linear Operator . . . 43
6.3 Orthogonal and Unitary Operators . . . 43
6.4 The Spectral Theorem for Self-Adjoint Operators . . . 44

**7 Groups, Representations, and Quantum Axioms** **47**
7.1 Groups . . . 47
7.2 Examples of Groups . . . 47
7.3 Group Actions . . . 48
7.4 Representations . . . 49
7.5 Fundamental Axioms . . . 50
7.6 Measurement Axioms . . . 52
7.7 Relation to Experimental Foundation of QM . . . 53
7.8 Groups as "Symmetries" . . . 54

**8 $U(1)$ and Conservation of Charge** **56**
8.1 Representation Theorems . . . 56
8.2 Representations of $U(1)$: Classification . . . 58
8.3 The Charge Operator . . . 59
8.4 The Lie Picture . . . 60
8.5 Conservation of Charge . . . 61
8.6 Why "Charge" is Important . . . 62

**9 $SU(2)$ and Two-State Systems** **63**
9.1 The Pauli Matrices . . . 63
9.2 Unitary Transformations on the System . . . 65
9.3 Commutation Relations . . . 67
9.4 Dynamics of the Two-State System . . . 68

**10 Lie Algebras I** **71**
10.1 Definition of a Lie Algebra . . . 71
10.2 Examples of Lie Algebras . . . 74
10.3 Orthogonal and Unitary Groups . . . 74

10.3.1 Orthogonal Group . . . 75
10.3.2 Unitary Group . . . 75
10.4 Classification of Simple Complex Lie Algebras . . . 76

**11 Lie Algebras II** **78**
11.1 Representations of Lie Algebras and Lie Groups . . . 78
11.2 The A/adjoint Representations . . . 80
11.3 Abstract Lie Algebra . . . 81
11.4 Complexification of a Lie Algebra . . . 82

**12 Rotation and Spin I** **85**
12.1 $SO(3)$ and Rotations in Three Dimensions . . . 85
12.2 Quaternions . . . 87
12.3 The Spin Group in Four Dimensions . . . 88

**13 Rotation and Spin II** **90**
13.1 The Spin Group in Three Dimensions . . . 90
13.2 Isomorphisms of Lie Groups and Algebras . . . 90
13.3 Summary . . . 94

**14 A Spin $1/2$ Particle** **96**
14.1 The Spinor Representation . . . 96
14.2 The Spin 1/2 Particle in a Magnetic Field . . . 97
14.3 The Heisenberg Picture . . . 100
14.4 Projective Space . . . 101
14.5 The Bloch Sphere . . . 103

**15 Representations of $SU(2)$: Classification** **106**
15.1 Passing to the Algebra . . . 106
15.2 Classification . . . 109
15.3 Interpretation . . . 111

**16 Representations of $SU(2)$: Construction** **112**
16.1 Homogeneous Polynomials . . . 112
16.2 Spherical Harmonics . . . 115
16.3 The Casmir Operator . . . 119

**17 Composite Quantum Systems I** **122**
17.1 Tensor Product Systems . . . 122
17.2 Physical States and Representations of $S_n$ . . . 122
17.3 Entangled and Independent Systems . . . 124
17.4 Tensor Product of Representations . . . 125
17.5 Tensor Products of $SU(2)$ Representations . . . 126
17.6 Character Approach to Clebsch-Gordan . . . 126
17.7 Some Examples . . . 127
17.8 Symmetric and Antisymmetric Multilinear Forms . . . 129

**18 Basic Quantum Computing (Not Written) 130**
18.1 Basic Gates . . . 130
18.2 Bloch Sphere - Again! . . . 130
18.3 Periods and Factoring . . . 130
18.4 Quantum Fourier Transform . . . 130
18.5 Shor's Algorithm . . . 130

**19 Quantum Field Theory (Not Written) 131**
19.1 Double-Slit Revisited . . . 131
19.2 Feynman Path Integral . . . 131
19.3 Why We Need Fields . . . 131
19.4 Canonical Quantization . . . 131

**20 The Hydrogen Atom (Not Written) 132**
20.1 Setup . . . 132
20.2 Angular Equation . . . 132
20.3 Radial Equation . . . 132
20.4 Radial Wave Function . . . 132
20.5 Summary . . . 132

**21 Introduction to Particle Physics 133**
21.1 The Standard Model . . . 133
21.2 A Little Relativity . . . 133
21.3 Feynman Diagrams . . . 133
21.4 Particle Accelerators . . . 133
21.5 Particle Detectors . . . 133

**A Matrix Theory 134**
A.1 Row-Reduction . . . 134
A.2 Correspondence Between Linear Operators and Matrices . . . 136
A.3 Change of Basis . . . 137
A.4 Proof of the Spectral Theorem . . . 137

**References 140**

# Quick reference

2.1 Definition (Vector Space) . . . 13
2.2 Example ($\mathbf{R}^n$, $\mathbf{C}^n$) . . . 13
2.3 Example (The empty set) . . . 14
2.4 Example (The zero vector space) . . . 14
2.5 Example (Matrices) . . . 14
2.7 Example (Functions) . . . 14
2.8 Definition (Subspace) . . . 14
2.12 Definition (Linear Combination, Span) . . . 15
2.13 Definition (Sum of Subspaces) . . . 15
2.15 Definition (Linearly Independent Set) . . . 17
2.17 Definition (Linearly Independent $k$-Tuple) . . . 17
2.20 Definition (Basis) . . . 18
2.21 Definition (Ordered Basis) . . . 18
3.1 Definition (Linear transformation, isomorphism) . . . 21
3.4 Definition (Kernel and Image) . . . 22
3.13 Definition (Linear Functional) . . . 25
3.15 Definition (Dual space) . . . 25
3.16 Definition (Dual Basis) . . . 25
3.19 Definition (Transpose Transformation) . . . 26
3.20 Definition (Bilinear form) . . . 27
3.24 Definition (Matrix of a Bilinear Map) . . . 28
4.3 Definition (Multilinear form) . . . 31
4.6 Definition (Group action) . . . 32
4.8 Definition (Alternating Multilinear Form) . . . 33
5.3 Definition (Determinant of a Linear Operator) . . . 35
5.5 Proposition (Formula for the determinant of a matrix) . . . 35
5.8 Definition (Eigenvectors and Eigenvalues) . . . 37
5.9 Definition (Eigenvalue of an $n \times n$ matrix) . . . 37
5.11 Definition (Eigenspace) . . . 37
6.1 Definition (Real Inner Product Space) . . . 41
6.3 Example (Examples of Inner Product Spaces) . . . 41
6.8 Definition (Adjoint operator) . . . 43
6.9 Definition (Orthogonal Group) . . . 44
6.10 Definition (Unitary Group) . . . 44
6.11 Theorem (Spectral Theorem for Self-Adjoint Operators) . . . 45
7.12 Definition (Group action) . . . 48
7.13 Definition (Representation) . . . 49
7.14 Definition (Unitary Representation) . . . 50
Axiom (States) . . . 50
Axiom (Quantum Observables) . . . 51
Axiom (Dynamics) . . . 51
Axiom (Observables) . . . 52
Axiom (The Born Rule) . . . 52

8.1 Definition (Irreducible Representation) . . . 56
8.2 Definition (Direct Sum Representation) . . . 56
8.5 Theorem (Schur's Lemma) . . . 57
8.7 Definition ($U(1)$) . . . 58
8.9 Definition (Charge Operator) . . . 60
10.1 Definition (Lie algebra) . . . 71
10.3 Definition (Lie bracket) . . . 73
10.8 Theorem (Classification of Simple Complex Lie Algebras) . . . 77
11.1 Definition (Lie algebra representation) . . . 78
11.3 Definition (Adjoint Lie algebra representation) . . . 81
11.4 Definition (Abstract Lie Algebra) . . . 81
11.5 Definition (Complexification) . . . 82
14.1 Definition (Spinor Representation) . . . 96
15.2 Theorem (Representations of $\mathfrak{sl}(2, \mathbf{C})$) . . . 109
16.1 Definition (Homogenous Polynomials) . . . 112
16.2 Definition (Casmir Operator for $SO(3)$) . . . 119
17.3 Definition (Decomposable and Indecomposable Vectors) . . . 124
17.4 Definition (Entanglement) . . . 124
17.5 Definition (Tensor Product Representation of a Group) . . . 125
17.6 Theorem (Clebsch-Gordan Decomposition) . . . 126
17.7 Definition (Character) . . . 126
17.8 Theorem (Weyl Character Formula) . . . 127
A.6 Theorem (Spectral Theorem for Normal Operators) . . . 137
A.9 Theorem (Spectral Theorem for Self-Adjoint Operators) . . . 139

# 1 Experimental Basis of Quantum Mechanics

See [4] for more discussion. Quantum Mechanics originates at the beginning of the twentieth century due to a variety of experiments that make two facts very clear:

1. At the quantum level, observable quantities are discrete, not continuous.
2. Measurement *changes the state of an object*.
   (a) Before a measurement, the state of an object is undetermined and "spread out" like a wave. Different states exist in superposition.
   (b) After a measurement, the state is determined and is collapsed like a particle.
   (c) The outcome of the measurement is probabilistic.

These two facts are in complete contrast with what we know about the macro-scale world, and in particular, the second one has major philosophical rambifications - do we really have free will if everything inside of our head is random? Einstein famously opposed quantum mechanics, claiming "God doesn't play dice with the universe."

The mathematics of quantum mechanics are strange and varied, but it is important to remember that these rules don't appear out of nowhere, but rather from physical experiment. Therefore, before we dive into the math (the primary focus of the course), we will spend some time talking about key events in physics.

## 1.1 Blackbody Radiation

A blackbody is an idealized object that perfectly absorbs all incident radiation and re-emits it as light based solely on the blackbody's temperature at various wavelengths. Classically, the Rayleigh-Jeans law predicted that at short wavelengths, infinite energy would be emitted - a nonsensical prediction.

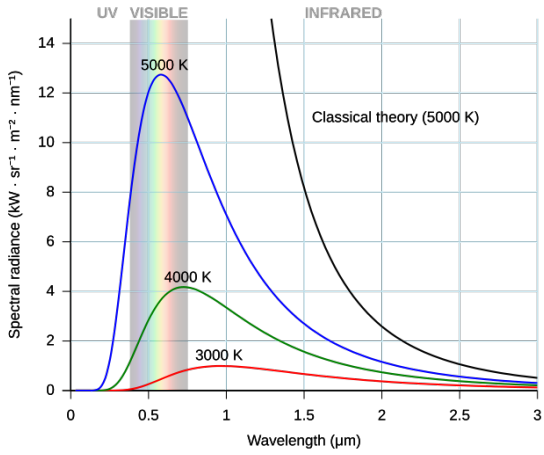


Figure 1: Intensity versus Wavelength for Different-Temperature Blackbodies (Wikipedia)

Max Planck came up with the idea that the blackbody could not emit energy at continuous intervals for a given wavelength, but rather only from multiples of some kind of

fundamental constant. That is, $E = nh\mu$ where $n$ is an integer, $\nu$ was the frequency, and $h$ was a constant. Running through the calculations, Planck's proposal matched perfectly, indicating that energy was not continuous, but rather *quantized*. However, nobody had any clue why Planck's formula worked - so it was regarded as a mathematical oddity until physicists could find more evidence.

## 1.2 Photoelectric Effect

The prevailing theory at this time was that light was a wave of energy.

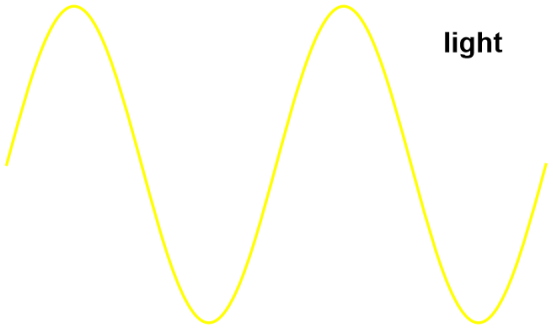


Figure 2: Depiction of a Light Wave

That means that the energy light imparts on an object it collides with should transfer energy gradually, and in particular, should depend on the intensity (brightness) of the light. It was known (due to Hertz) that shining ultraviolent light on metal would cause the metal to become negatively electrically charged, i.e the metal would emit electrons.

To test this, physicists varied the intensity and frequencies of light and shone it on metal, placing anodes and cathodes to track electric discharge:

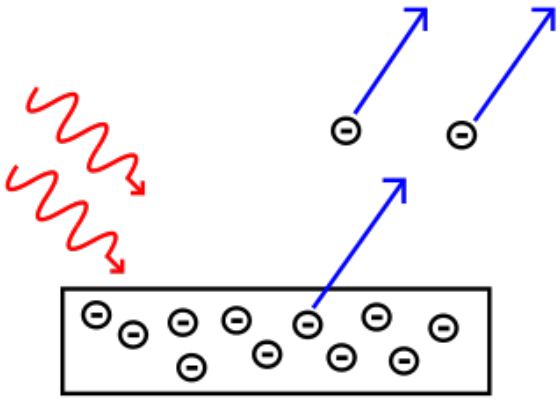

Figure 3: Depiction of the Photoelectric Effect (Thoughtco)

What they found was that at low frequencies, the metal would never discharge electrons, and at high frequencies, the metal would always discharge electrons, and that this

behavior had nothing to do with intensity. The Nobel Prize-winning solution to this problem came from Albert Einstein, who proposed that light comes in discrete packets of particles called *photons* that travel in waves.



Figure 4: Enter Caption

Each photon had energy $E = h\nu$ for some constant $h$ (exactly Planck's constant from Blackbody Radiation) where $\nu$ was the frequency.

This explained the behavior - if an individual photon was of sufficient energy (frequency), it would "knock an electron" out of the metal, and if not, it would never do so. Changing the intensity of the light merely meant changing the number of photons, not the behavior of each photon. This experiment demonstrates that light was quantized - it comes in discrete objects, not in a waveform.

## 1.3 Hydrogen Spectrum

In the almost-reverse manner of how metal emits electrons when exposed to light, when an elemental gas is excited by electricity or heat, it will emit light. We can put this light through a prism to separate it into its constituent frequencies with the following experimental setup:

When physicists did this, they observed a very interesting result.

Here, we see that the light is not made up of a continuous band of wavelengths that contribute to its color, but rather various discrete wavelengths.

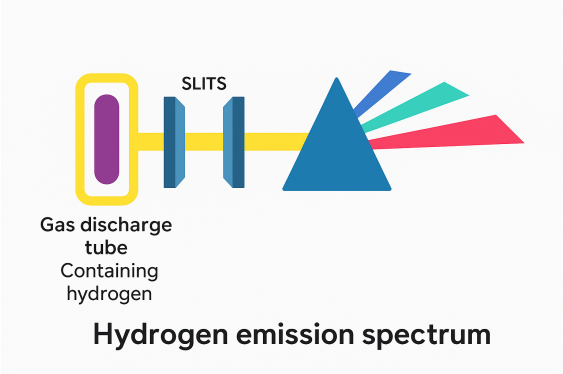


Figure 5: Experimental Setup for the Hydrogen Spectrum Experiment (ChatGPT)

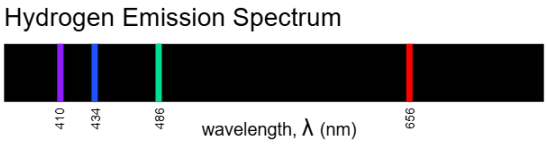


Figure 6: Hydrogen Spectrum (University of Texas)

What is going on here? The answer is that as we excite the gas, electrons in the gas are being excited. But then, these electrons must return to the ground state, and they emit the excess energy via light. This light being limited to specific frequencies tells us that the allowed energy of electrons is limited to specific frequencies, i.e, it is quantized.

## 1.4 Double-Slit Experiment

The photoelectric effect indicates that light is a particle. The double-slit experiment indicates the opposite - that light is a wave.

When two waves encounter each other, the resulting pattern is called interference. At places where crests coincide, we have constructive interference. At places where crests and troughs coincide, we have destructive interference.

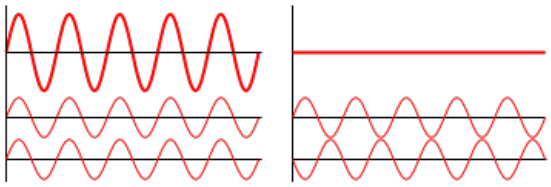

Figure 7: Depiction of Constructive/Destructive Interference (Wikipedia)

To test whether light is a wave, then, we simply have to check to see if light interferes with itself. A way to do this is the double slit experiment. Essentially, shine light through two slits. If light is a wave, the two slits will serve as light sources interfering with each other, and we will observe an interference pattern.

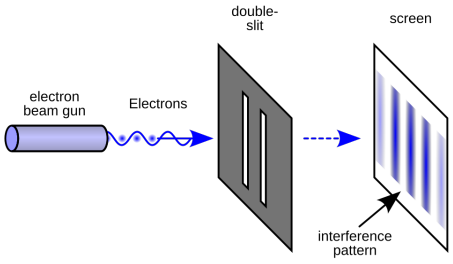


Figure 8: Experimental Setup for the Double Slit Experiment (Wikipedia)

We can calculate what the double-slit interference pattern should look like, and it turns out to look like this:

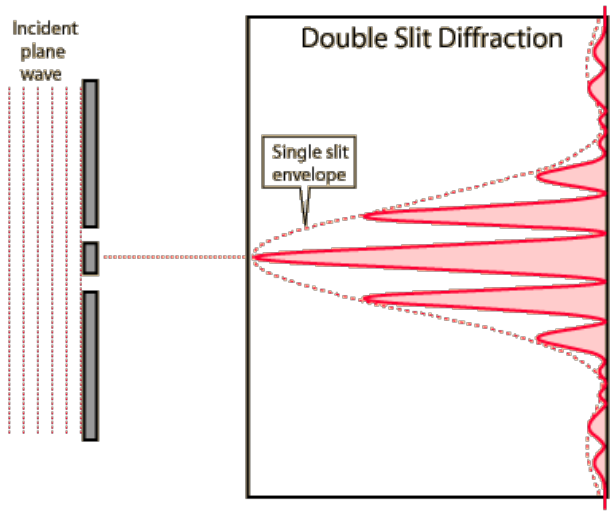


Figure 9: Double Slit Experiment Diffraction Pattern (Georgia State University)

So, physicists do this. What we find is that light behaves like a wave - contradicting the photoelectric effect's evidence that light is composed of many particles. So, to test Einstein's solution to the photoelectric effect, they send single photons at a time, at which point since there is only one photon, there can't be a wave pattern, and so they should lose the wave behavior.

To see this experiment in action, let's watch this Youtube video.

We observe wave behavior still, confirming somehow that even single photons behave like waves. But it gets even crazier. They place a detector on the slits to check which slit each photon goes through. This does not affect the trajectory of the photon. As soon as they do this, the interference pattern disappears, and they observe two bands. This is where the term *collapse* comes from - the spectrum collapses into two points. Somehow, the act of measurement forces the photon to pass through a single slit and never interfere.

The double-slit experiment has been done many times, and every time, when we measure it, we lose the interference pattern. But, it gets even more strange.

Physicists have built machines that can track which path the electron goes through, but cannot output or store it. In these cases, despite the machine "measuring' which slit it goes through, we observe an interference pattern, indicating a measurement was not made. But, when we do record what slit it goes through, suddenly we recover single bands.

It is that act of knowing, or even simply "being able to know" via recording it, that causes the collapse into the single bands. This seems to imply that somehow interaction with something macroscopic is relevant to the behavior of what is happening. Understanding how this works is broadly referred to as "the measurement problem" and is a big open question in quantum mechanics.

## 1.5 Stern-Gerlarch Experiment

In 1922, Otto Stern and Walther Gerlach wanted to test a prediction of atomic theory - that atoms have rotational momentum. If atoms have angular momentum, how do they behave in a magnetic field? They built the following machine - now called a Stern-Gerlarch machine.

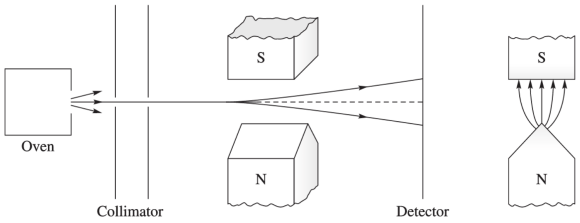


Figure 10: Diagram of a Stern-Gerlarch Machine

Their answer was to excite a beam of silver atoms and shoot them through a non-uniform magnetic field. Such a magnetic field would pull spinning atoms up or down depending on how they spun. Classically, they expected atoms could spin in any way, and so, the magnetic field should have caused a continuous spectrum of locations.

However, they found only two locations - indicating that somehow the angular momentum of the atom was quantized, and that these occurred with equal probability. Furthermore, these silver atoms contained exactly one free electron. Hence, the spin of the electron (since it was known neutral atoms had no intrinsic angular momentum) was responsible for the behavior they observed. This showed both the existence of intrinsic angular momentum of a particle, and that this quantity was quantized.

## 1.6 Overall Conclusions

It is clear from Blackbody Radiation, the Photoelectric Effect, the Hydrogen Spectrum, and the Stern-Gerlach that quantities like energy and spin of atoms are quantized, not continuous. Furthermore, the double-slit experiment and the Stern-Gerlach experiment showed that outcomes of experiments were not deterministic, but probabilistic, and in particular, the double-slit experiment showed light was somehow behaving both as a wave and a particle, with wave behavior prior to measurement and particle behavior after measurement. Mathematically analyzing the double-slit experiment also tells us that the relevant probabilities are proportional not to the amplitude of the wave itself, but rather the square of the amplitude, providing even more confusion.

Once we have built up the framework of linear algebra, the theoretical framework of quantum mechanics will naturally explain all of these experimental oddities. Keep them in mind over the next few lectures, as we will approach linear algebra from a purely-mathematical perspective, and then reconcile that abstract mathematics with the experimental ideas we saw here.

# 2 Linear Algebra I

Now that we have an idea of what experimental observations underlie quantum mechanics, we now want to find a mathematical framework that can encapsulate these. The natural choice is of course, linear algebra. We have a lot of work ahead of us.

The Linear Algebra sections are taken from a variety of resources, including [1, 3, 5],

## 2.1 Vector Spaces

**Definition 2.1 (Vector Space).** Let $\mathbf{R}$ or $\mathbf{C}$ denote the field of real numbers. A *vector space over* $\mathbf{R}$ or $\mathbf{C}$ (otherwise known as a *real vector space*) is a set $V$ endowed with two operations

$$\begin{aligned} + : V \times V &\to V \\ (v, w) &\to v + w, \end{aligned}$$

called *vector addition*, and

$$\begin{aligned} \cdot : \mathbf{R} \times V &\to V \\ (c, v) &\to cv \end{aligned}$$

called *scalar multiplication*, satisfying the following properties:

**A1.** (Associativity of Addition) $u + (v + w) = (u + v) + w$ for all $u, v, w \in V$.

**A2.** (Commutativity of Addition) $u + v = v + u$ for all $u, v \in V$.

**A3.** (Existence of an Additive Identity) There exists an element $0 \in V$ such that $v + 0 = v$ for all $v \in V$.

**A4.** (Existence of Additive Inverses) For every $v \in V$, there exists a $w \in V$ such that $v + w = 0$.

**S1.** (Associativity of Multiplication) $(xy)v = x(yv)$ for all $x, y \in \mathbf{R}$, $v \in V$.

**S2.** (Distributivity over Scalar Addition) $(x + y)v = xv + yv$ for all $x, y \in \mathbf{R}, v \in V$.

**S3** (Distributivity over vector addition) $x(v + w) = xv + xw$ for all $x \in \mathbf{R}$, $v, w \in V$.

**S4.** (Multiplication by 1 fixes each vector) $1v = v$ for all $v \in V$.

The elements of $V$ are called *vectors* and the elements of $\mathbf{R}$ or $\mathbf{C}$ are called *scalars*.

**Example 2.2 ($\mathbf{R}^n$, $\mathbf{C}^n$).** It is straightforward to check that $\mathbf{R}^n, \mathbf{C}^n$ is a vector space under the operations

$$\begin{aligned} (x_1, x_2, ..., x_n) + (y_1, y_2, ..., y_n) &= (x_1 + y_1, x_2 + y_2, ..., x_n + y_n) \\ c(x_1, x_2, ..., x_n) &= (cx_1, cx_2, ..., cx_n) \end{aligned}$$

Indeed, the definition of a vector space comes from abstracting the properties of $\mathbf{R}^n, \mathbf{C}^n$.

**Example 2.3 (The empty set).** The empty set cannot be given the structure of a vector space, since axiom $A3$ fails. Thus, every vector space is necessarily nonempty.

**Example 2.4 (The zero vector space).** Any singleton set $\{a\}$ can be given the structure of a vector space by defining $a + a = a$ and $ca = a$ for all $c \in \mathbf{R}$. Since $a$ is the 0 vector in axiom $A3$, we usually denote this by writing $a = 0$. This vector space is called the *zero vector space* or *trivial vector space*.

**Example 2.5 (Matrices).** The set of $m \times n$ matrices with real entries is a vector space under the entrywise operation

$$(A + B)_{ij} = A_{ij} + B_{ij}$$
$$(cA)_{ij} = cA_{ij}.$$

**Example 2.6.** The set $\mathcal{P})n(\mathbf{R})$ of polynomials of degree $\leq n$ with coefficients in $\mathbf{R}$ is a vector space under polynomial addition and multiplication of coefficients.

**Example 2.7 (Functions).** The set of all real-valued functions defined on a common domain $A$ is a vector space under the pointwise operations

$$(f + g)(x) = f(x) + g(x)$$
$$(cf)(x) = cf(x).$$

## 2.2 Subspaces

Now, an interesting question is "what are the substructures inside vector spaces?"

**Definition 2.8 (Subspace).** Let $V$ be a vector space. If $W$ is a nonempty subset of $V$ which is closed under vector addition and scalar multiplication, then it is straightforward to check that $W$ is itself a vector space under the same operations of $V$. In this case, we say that $W$ is a *subspace* of $V$.

The next proposition shows that we can check closure under addition and scalar multiplication all at once.

**Proposition 2.9.** A nonempty subset $W$ of a vector space $V$ is a subspace if and only if $w, w' \in W$ and $c \in \mathbf{R}$ implies $cw + w' \in W$.

*Proof.* Suppose $W$ is closed under addition and scalar multiplication and let $w, w' \in W$ and $c \in \mathbf{R}$. Then $cw \in W$ since $W$ is closed under scalar multiplication and therefore $cw + w' \in W$ since $W$ is closed under vector addition. Since $W$ is nonempty, we have $c \cdot w = 0$ for some $w \in W$, and by closure $0 \in W$.

Now suppose $W$ has the property that $w, w' \in W$ and $c \in \mathbf{R}$ implies $cw + w' \in W$. Taking $c = 1$, this implies $1 \cdot w + w' = w + w' \in W$, hence $W$ is closed under vector addition. Taking $w' = 0$ (which is in $W$ since $W$ is a subspace), $cw + 0 = cw \in W$, which shows that $W$ is closed under scalar multiplication. □

**Example 2.10.** Every vector space has at least two subspaces: $\{0\}$ and $V$ itself.

**Example 2.11.** The set $C^\infty(\mathbf{R}^n)$ of smooth functions on $\mathbf{R}^n$ is a subspace of all real-valued functions on $\mathbf{R}^n$.

**Definition 2.12 (Linear Combination, Span).** Let $V$ be a vector space. A finite sum of the form $\sum_{i=1}^k c^i v_i$, where $c^i$ are scalars and $v_i \in V$, is called a *linear combination* of the vectors $v_1, ..., v_k$.

If $S$ is an arbitrary subset of $V$, then the set of all linear combinations of elements of $S$ is caled the *span of* $S$ and is denoted by $\text{span}(S)$; it is the smallest subspace of $V$ containing $S$.

If $V = \text{span}(S)$, we say $S$ *spans* $V$. By convention, a linear combination of no elements is considered to sum to zero, and so then $\text{span}() = \{0\}$.

If $V = \text{span}(S)$ with $S$ finite, then we say that $V$ is *finite-dimensional*; otherwise we say that $V$ is infinite-dimensional.

## 2.3 Direct Sums

It also turns out to be useful to ask about sums of spaces and subspaces.

**Definition 2.13 (Sum of Subspaces).** The *sum* of two subspaces $A$ and $B$ of a vector space is the subspace

$$A + B = \{a + b \in V \mid a \in A, b \in B\}.$$

Since $c(a+b) + (a'+b') = (ca+a') + (cb+b')$ with $ca + a' \in A$ and $cb + b' \in B$, this is indeed a subspace.

If $a \cap B = \{0\}$, this sum is called an *internal direct sum* and is written $A \oplus B$. In this case, every $v$ can be written uniquely as $v = a + b$ for unique $a \in A$ and $b \in B$. For, if there are two such ways to write this, $v = a + b = a' + b'$, so $(a - a') = (b' - b)$, but $(a - a') \in A$, $b' - b \in B$, so we must have $(a - a') = (b' - b) = 0$, showing $a = a'$, $b = b'$.

In contrast to the internal direct sum, if $A$ and $B$ are any two vector spaces, then we can form the *direct sum*

$$A \oplus B = \{(a, b) : a \in A, b \in B\}$$

which is the same thing as their Cartesian set product.

If $V = A \otimes B$, $A$ is called complimentary to $B$ in $V$.

These constructions are basically the same (in fact, we can prove they are mathematically the same using some machinery we'll introduce later), which is why we use the same notation for both.

**Example 2.14.** Recall that a function $f : \mathbf{R} \to \mathbf{R}$ is *even* if $f(-x) = f(x)$ for all $x \in \mathbf{R}$ and *odd* if $f(-x) = -f(x)$ for all $x \in \mathbf{R}$. Let $V_e$ and $V_o$ be the sets of all even and odd functions, respectively.

**Exercise 2.1.** Prove that $V_e$ and $V_o$ are subspaces of $V = \mathbf{R}^\mathbf{R}$.

**Solution.** The zero function $\mathbf{0}(x) = 0$ is both even and odd, so $\mathbf{0} \in V_e$ and $\mathbf{0} \in V_o$.

Let $f_1, f_2 \in V_e$ and let $\alpha \in \mathbf{R}$. Define $f_3 = f_1 + \alpha f_2$. Then,

$$f_3(-x) = f_1(-x) + \alpha f_2(-x) = f_1(x) + \alpha f_2(x) = f_3(x),$$

so $f_3$ is even. Thus, $V_e$ is a subspace.

Similarly, if $f_1, f_2 \in V_o$, then

$$f_3(-x) = f_1(-x) + \alpha f_2(-x) = -f_1(x) - \alpha f_2(x) = -f_3(x),$$

so $f_3$ is odd, and $V_o$ is also a subspace. □

**Exercise 2.2.** Let $f(x) \in \mathbf{R}^{\mathbf{R}}$. Prove that the function $f_e(x) := f(x) + f(-x)$ is even. Construct a corresponding odd function and prove it is odd.

**Solution. Proposition.** The function $f_e(x) := f(x) + f(-x)$ is even.

*Proof.* For all $x$,

$$f_e(-x) = f(-x) + f(-(-x)) = f(-x) + f(x) = f_e(x),$$

so $f_e$ is even.

**Proposition.** The function $f_o(x) := f(x) - f(-x)$ is odd.

*Proof.* For all $x$,

$$f_o(-x) = f(-x) - f(-(-x)) = f(-x) - f(x) = -(f(x) - f(-x)) = -f_o(x),$$

so $f_o$ is odd. □

**Exercise 2.3.** Prove that $\mathbf{R}^{\mathbf{R}} = V_e + V_o$.

**Solution.** Let $f \in \mathbf{R}^{\mathbf{R}}$. Then:

$$\begin{aligned} f(x) &= \frac{2f(x)}{2} \\ &= \frac{f(x) + f(-x) + f(x) - f(-x)}{2} \\ &= \frac{f(x) + f(-x)}{2} + \frac{f(x) - f(-x)}{2} \\ &= \frac{f_e(x)}{2} + \frac{f_o(x)}{2}, \end{aligned}$$

which is the sum of an even function and an odd function. Therefore, every $f$ is the sum of an even and an odd function, so $\mathbf{R}^{\mathbf{R}} = V_e + V_o$. □

**Exercise 2.4.** Prove that $\mathbf{R}^{\mathbf{R}} = V_e \oplus V_o$.

**Solution.** We already showed $\mathbf{R}^{\mathbf{R}} = V_e + V_o$. It remains to show that $V_e \cap V_o = \{\mathbf{0}\}$.

Suppose $f \in V_e \cap V_o$. Then $f$ is both even and odd:

$$f(x) = f(-x) = -f(x) \Rightarrow 2f(x) = 0 \Rightarrow f(x) = 0 \text{ for all } x.$$

Thus, $f = \mathbf{0}$ is the only function in the intersection, and so $V_e \cap V_o = \{\mathbf{0}\}$. Therefore, $\mathbf{R}^{\mathbf{R}} = V_e \oplus V_o$. □

## 2.4 Bases and Dimension

The linear structure of vector spaces provides a lot of powerful structure. In this section, we aim to break down elements of vector spaces vector spaces into "elementary pieces" that we can use to build back up every element.

**Definition 2.15 (Linearly Independent Set).** Let $V$ be a vector space. A subset $S$ of $V$ is said to be *linearly dependent* if there exists a linear relation of the form $\sum_{i=1}^{k} a^i v_i = 0$, where $v_1, ..., v_k$ are distinct elements of $S$ and at least one of the coefficients $a^i$ is nonzero; $S$ is said to be *linearly independent* otherwise. In other words, $S$ is linearly independent if and only if the only linear combination of distinct elements of $S$ that sums to zero is the one in which all the scalar coefficients are zero.

**Remark 2.16.** Note that every set containing the zero vector is linearly dependent. By convention, the empty set is considered to be linearly independent.

It is frequently important to work with ordered $k$-tuples of vectors in $V$; such a $k$-tuple is denoted by $(v_1, ..., v_k)$ or $(v_i)$; with parentheses instead of braces to distinguish it from the (unordered) set of elements $\{v_1, ..., v_k\}$.

**Definition 2.17 (Linearly Independent $k$-Tuple).** We say that $(v_1, ..., v_k)$ is a *linearly dependent $k$-tuple* if there are scalars $(a^1, ..., a^k)$, not all zero, such that $\sum_{i=1}^{k} a^i v_i = 0$; it is a *linearly independent $k$-tuple* otherwise.

**Remark 2.18.** The only difference between a linearly independent set and a linearly independent $k$-tuple is that the latter cannot have repeated vectors. For example, if $v \in V$ is a nonzero vector, the ordered pair $(v, v)$ is linearly dependent, while the set $\{v, v\} = \{v\}$ is linearly independent. On the other hand, if $(v_1, ..., v_k)$ is any linearly independent $k$-tuple, then the set $\{v_1, ..., v_k\}$ is also linearly independent.

**Proposition 2.19.** Let $V$ be a vector space.

(a) If $S \subset V$ is linearly independent, then every subset of $S$ is linearly independent.

(b) If $S \subset V$ is linearly dependent or spans $V$, then every subset of $V$ that properly contains $S$ is linearly dependent.

(c) A subset $S \subset V$ containing more than one element is linearly dependent if and only if some element $v \in S$ can be expressed as a linear combination of elements $S\ \{v\}$.

(d) If $(v_1, ..., v_k)$ is a linearly dependent $k$-tuple in $V$ with $v_1 \neq 0$, then some $v_i$ can be expressed asa linear combination of the preceding vectors $(v_1, ..., v_{i-1})$.

*Proof.* (a) Suppose $S \subset V$ is linearly independent. Then $\sum_{i=1}^{k} a^i v_i = 0$ implies $a^i = 0$ for all $i = 1, ..., k$ whenever $v_1, ..., v_k$ are vector in $S$. If $W$ is a subset of $S$ and $\sum_{i=1}^{k} a^i v_i = 0$, then since each $v_i \in W$ is also in $S$, this is a linear combination of vectors in $S$ and therefore we have $a^i = 0$ for all $i = 1, ..., k$. hence, $W$ is linearly independent.

(b) Suppose $S \subset V$ is linearly dependent. Then there exist vectors $v_1, ..., v_k$ in $S$ and scalars $a^1, ..., a^k$ with at least one $a^j \neq 0$, such that $\sum_{i=1}^{k} a^i v_i = 0$. If $W$ is a subset that

properly contains $S$, then for any $w \in W$ and $a^{k+1} = 0$ we have $\sum_{i=1}^{k} a^i v_i + a^{k+1} w = 0$ with $a^j \neq 0$, which shows that $W$ is linearly dependent.

Suppose now that $\text{span}(S) = V$ and let $W$ be a subset of $V$ that properly contains $S$. Note that $V$ cannot be the zero vector space, since the only proper subset is , which is linearly independent. Since $S$ spans $V$, $S$ must contain a non-zero vector. Since $V$ is not the zero vector space and $S$ spans $V$, $S$ must contain a nonzero vector. If $W = S \cup \{0\}$, then $W$ is linearly dependent because it contains the zero vector. otherwise, $W$ properly contains $S$ and contains a nonzero vector $w$. Since $S$ spans $V$, there exist scalars $a^1, ..., a^k$ and vectors $v_1, ..., v_k \in S$ such that $\sum_{i=1}^{k} a^i v_i = w$. Since $w \neq 0$, at least one term $a^j v_j \neq 0$, which means that $a^j \neq 0$ and $v_j \neq 0$. It then follows that $\sum_{i=1}^{k} (-a^i) v_i + w = 0$ with $a^j \neq 0$, hence $W$ is linearly dependent.

(c) Suppose $S$ is linearly dependent. Then there exists a linear relation of the form $\sum_{i=1}^{k} a^i v_i = 0$ with $a^j \neq 0$. It follows that $v_j = \sum_{i \neq j} \left( -\frac{a^i}{a^j} \right) v_i$. Now suppose $v \in S$ can be written as $v = \sum_{i=1}^{k} a^i v_i$ for $v_1, ..., v_k \in S \setminus \{v\}$. Then $v - \sum_{i=1}^{k} a^i v_i = 0$, where the coefficient of $v$ is $1 \neq 0$, which shows that $S$ is linearly dependent.

(d) If $(v_1, ..., v_k)$ is a linearly dependent $k$-tuple, then there exists a $k$-tuple of scalars $(a^1, ..., a^k)$ not all zero, such that $\sum_{i=1}^{k} a^i v_i = 0$. Choose $j$ to be the largest index such that $a^j \neq 0$. Then

$$v_j = \sum_{i=1}^{j-1} \left( -\frac{a^i}{a^j} v_i \right).$$

If $j = 1$, then there are no preceding vectors, so the claim is vacuously true.

□

**Definition 2.20 (Basis).** A *basis for* $V$ is a subset $S \subset V$ that is linearly independent and spans $V$. IF $S$ is a basis for $V$, then every element of $V$ has a *unique* expression as a linear combination of elements of $S$.

If $V$ has a finite basis, then $V$ is said to be *finite-dimensional*, and otherwise it is *infinite-dimensional*. the trivial vector space $\{0\}$ is finite-dimensional, because it has the empty set as a basis.

**Definition 2.21 (Ordered Basis).** If $V$ is finite-dimensional, an *ordered basis for* $V$ is a basis endowed with a specific ordering of the basis vectors, or equivalently a linearly independent $n$-tuple $(E_i)$ that spans $V$.

For most purposes, ordered bases are more useful than unordered bases, so we always assume, often without comment, that each basis comes with a given ordering.

If $(E_1, ..., E_n)$ is an (ordered) basis for $V$, each vector $v \in V$ has a unique expression as a linear combination of basis vectors:

$$v = \sum_{i=1}^{n} v^i E_i.$$

The real numbers $v^i$ are called the *components of* $v$ with respect to this basis, and the ordered $n$-tuple $(v^1, ..., v^n)$ is called its *basis representation*. (Note that this definition requires an ordered basis.)

**Lemma 2.22.** Let $V$ be a vector space. If $V$ is spanned by a set of $n$ vectors, then every subset of $V$ containing more than $n$ vectors is linearly dependent.

*Proof.* Let $S$ be a subset of $V$ containing more than $n$ vectors. Then $S$ contains distinct vectors $w_1, ..w_m$ with $m > n$. Since $\{v_1, ..., v_m\}$ spans $V$, we can write each $w_i$ as a linear combination

$$w_i = \sum_{k=1}^{m} B_{ik} v_k.$$

Suppose there are scalars $\alpha_1, ..., \alpha_m \in F$ such that:

$$0 = \sum_{i=1}^{n} \alpha_i w_i.$$

If the set is independent, then the only solution to this system is forcing all the $w_i = 0$. Substituting the first equation into the second, we have:

$$\begin{aligned} 0 &= \sum_{i=1}^{n} \alpha_i \left( \sum_{k=1}^{m} B_{ik} v_k \right) \\ &= \sum_{i=1}^{n} \sum_{k=1}^{m} \alpha_i (B_{ik} v_k) \\ &= \sum_{i=1}^{n} \sum_{k=1}^{m} (\alpha_i B_{ik}) v_k \\ &= \sum_{k=1}^{m} \left( \sum_{i=1}^{n} \alpha_i B_{ik} \right) v_k \end{aligned}$$

Consider the linear system of equations

$$\sum_{i=1}^{n} \alpha_i B_{ik} = 0$$

for $k = 1, ..., m$. This is a system of $m$ equations in $n > m$ unknowns. It is known that such a system has multiple solutions, and in particular, some solution where all the $\alpha_i$ are not zero. This implies then that the system is not independent as desired. □

**Proposition 2.23.** If $V$ is finite-dimensional, all bases for $V$ contain the same number of elements.

*Proof.* If $\{E_1, ..., E_n\}$ is a basis for $V$ with $n$ elements, then the prior lemma implies that every set containing more than $n$ elements is linearly dependent, so no basis can have

more than $n$ elements. On the other hand, if there were a basis consisting of fewer than $n$ elements, then Lemma 4.19 would imply that $\{E-1, ..., E_n\}$ is linearly dependent, which is a contradiction.

Because of the preceding proposition, if $V$ is a finite-dimensional vector space, it makes sense to define the *dimension of* $V$, denoted by $\dim V$, to be the number of elements in any basis. □

**Proposition 2.24.** Let $V$ be a finite-dimensional vector space.

(a) Every set that spans $V$ contains a basis, and every linearly independent subset of $V$ is contained in a basis.

(b) Every subspace $S \subset V$ is finite-dimensional and satisfies $\dim S \leq \dim V$, with equality if and only if $S = V$.

(c) Show that $\dim V = 0$ if and only if $V = \{0\}$.

**Example 2.25.** The *standard basis* for $\mathbf{R}^n$ or $\mathbf{C}^n$ consists of the $n$ vectors $e_1, ..., e_n$, where $e_i = (0, ..., 1, ..., 0)$ is the vector with a 1 in the $i$th place and zeros elsewhere. This shows that $\mathbf{R}^n, \mathbf{C}^n$ hav dimension $n$ (over $\mathbf{R}, \mathbf{C} respectively$). Any element $x \in \mathbf{R}^n, \mathbf{C}^n$ can be written as $(x^1, ..., x^n) = \sum_{i=1}^n x^i e_i$ for scalars $x^i$ in $\mathbf{R}, \mathbf{C}$ respectively, so its components with respect to the standard basis are just its coordinates $(x^1, ..., x^n)$.

# 3 Linear Algebra II

## 3.1 Linear Maps

Now that we have the structure of a vector space down, we have a natural question - what are the "interesting" maps between vector spaces? In this sense, interesting means the maps preserving the vector space structure, meaning that it should preserve the operations. This naturally leads to the following definition:

**Definition 3.1 (Linear transformation, isomorphism).** Let $V$ and $W$ be vector spaces over $\mathbf{R}$ or $\mathbf{C}$. A map : $V \to W$ is called a *linear transformation*, a *vector space homomorphism*, a *linear operator*, or a *linear map* over $\mathbf{R}$ (resp $\mathbf{C}$) if for all $u, v \in V$ and $c \in \mathbf{R}$ (resp $\mathbf{C}$),

$$\begin{aligned} f(u+v) &= f(u) + f(v) \\ f(cv) &= cf(v). \end{aligned}$$

If, additionally, $f$ is a bijection, then it is called a *vector space isomorphism*, and we say that $V$ and $W$ are isomorphic, which we denote by $V \cong W$.

The natural consequence of this definition is the following:

**Remark 3.2.** The set of linear functions from $V$ to $W$ is a vector space with respect to the pointwise operations

$$\begin{aligned} (T_1 + T_2)(v) &:= T_1(v) + T_2(V) \\ (\alpha T)(v) &:= \alpha T(v). \end{aligned}$$

We denote this space $\mathrm{Hom}(V, W)$ or $\mathcal{L}(V, W)$.

What does it mean to be isomorphic? It means $V$ and $W$ we have the underlying same vector space structure, and the isomorphism is just a "renaming" of the elements. We can see this better after the next proposition.

**Proposition 3.3.** 1. If $f : V \to W$ is an isomorphism, then so is $f^{-1} : W \to V$.

2. If $f : V \to U$ and $g : U \to W$ is an isomorphism, the so is $g \circ f : V \to W$

*Proof.* 1. Since the inverse of a bijection is a bijection, all we need to show is that $f^{-1}$ is linear. Since $f$ is surjective, given $w, w' \in W$, there exist $v, v', v'' \in V$ such that $w = f(v), w' = f(v'), cw + w' = f(v'')$. It follows that

$$\begin{aligned} cw + w' &= f(v'') \\ cf(v) + f(v') &= f(v'') \\ f(cv + v') &= f(v'') \\ cv + v' &= v'' \\ cf^{-1}(w) + f^{-1}(w') &= f^{-1}(cw + w'), \end{aligned}$$

which proves that $f^{-1}$ is linear and therefore an isomorphism.

2. Since the composition of a bijection is a bijection, all we need to show is that the composition of linear maps is linear. Let $v, v' \in V$ and $c \in \mathbf{R}$. Then,

$$\begin{aligned}(g \circ f)(cv + v') &= g(f(cv + v')) \\ &= g(cf(v) + f(v')) \\ &= cg(f(v) + g(f(v')) \\ &= c(g \circ f)(v) + (g \circ f)(v'),\end{aligned}$$

which shows that $g \circ f$ is linear, and hence an isomorphism.

□

Now, what we see is that when we have an isomorphism, we have a bijection between the two spaces, so we can associate each $v \in V$ with a unique $w \in W$. The linearity conditions simply mean that the vector space operations behave the same according to this correspondence, and so really, $v$ and $w$ behave the exact same under the vector space, but have different names.

**Definition 3.4 (Kernel and Image).** If $f : V \to W$ is a linear transformation, the *kernel* of $f$ is the set

$$\ker f = \{v \in V : f(v) = 0\}$$

and the *image* of $f$ is the set

$$\operatorname{Im} f = \{w \in W : w = f(v) \text{ for some } v \in V\}.$$

**Proposition 3.5.** If $f : V \to W$ is a linear map, then $\ker f$ is a subspace of $V$ and $\operatorname{Im} f$ is a subspace of $W$.

*Proof.* Let $v, w \in \ker f$. Then $f(cv + w) = cf(v) + f(w) = c0 + 0 = 0 + 0 = 0$, so $cv + w \in \ker f$. This shows that $\ker f$ is a subspace of $V$. Let $w, w' \in \operatorname{Im} f$. Then there exist $v, v' \in V$ such that $w = f(v)$ and $w' = f(v')$. It follows that $cw + w' = cf(v) + f(v') = f(cv + v')$, which shows that $cw + w' \in \operatorname{Im} f$ and so $\operatorname{Im} f$ is a subspace of $W$. □

**Proposition 3.6.** Let $f : V \to W$ be a linear map. Then $\ker f = \{0\}$ if and only if $f$ is injective.

*Proof.* For any linear map $f(0) = f(0 + 0) = f(0) + f(0)$. Subtracting $f(0)$ from both sides gives $f(0) = 0$. hence $0 \in \ker f$ (or $\{0\} \subset \ker f$) for any linear map.

Suppose $f$ is injective. Then $f(v) = f(v')$ implies $v = v'$. If $v \in \ker f$, then $f(v) = 0 = f(0)$, so $v = 0$. This shows that $\ker f \subset \{0\}$, and therefore $\ker f = \{0\}$.

Now suppose $\ker f = \{0\}$. If $f(v) = f(v')$, then $0 = f(v) - f(v') = f(v - v')$, which says $v - v' \in \ker f$, and therefore $v - v' = 0$, which implies $v = v'$. hence, $f$ is injective. □

Now, we come across one of the most important facts in linear algebra - the fact that any finite-dimensional space is the same (up to isomorphism) as $\mathbf{R}^n$ (or $\mathbf{C}^n$).

**Proposition 3.7.** Let $V$ be an $n$-dimensional vector space and let $(v_1, ..., vn)$ be an ordered basis for $V$. Then the map

$$f : \mathbf{R}^n \to V$$
$$f(x^1, ..., x^n) = \sum_{i=1}^{n} x^i v_i$$

is an isomorphism.

*Proof.* We have:

$$\begin{aligned} f(c(x^1, ..., x^n) + (y^1, ..., y^n)) &= f(cx^1 + y^1, ..., cx^n + y^n) \\ &= \sum_{i=1}^{n} (cx^i + y^i) v_i \\ &= c \sum_{i=1}^{n} x^i + \sum_{i=1}^{n} y^i \\ &= cf(x^1, ..., x^n) + f(y^1, ..., y^n), \end{aligned}$$

which shows that $f$ is linear. Since $(v_1, ..., v_n)$ is linearly independent,

$$f(x^1, ..., x^n) = \sum_{i=1}^{n} x^i v_i = 0$$

implies $x^i = 0$ for all $i$. This proves that $\ker f = \{0\}$, hence $f$ is injective by the prior proposition.

Since $(v_1, ..., v_n)$ spans $V$, given any $v \in V$, there exist scalars $c^1, ..., c^n$ such that

$$v = \sum_{i=1}^{n} c^i v_i = f(c^1, ..., c^n),$$

which shows that $f$ is surjective. This shows that $f$ is a bijection and hence an isomorphism. □

**Proposition 3.8.** Two finite-dimensional vector spaces are isomorphic if and only if they have the same dimension.

*Proof.* Suppose $\dim V = \dim W = n$, then by the prior proposition we have an isomorphism $f : V \to \mathbf{R}^n$ and $g : \mathbf{R}^n \to W$, and therefore $g \circ f : V \to W$ is an isomorphism.

Now, suppose $V \cong W$. Then, let $f : V \to W$ be an isomorphism, and let $\{v_1, ..., v_n\}$ be a basis for $V$. we will show that $\{f(v_1), ..., f(v_n)\}$ is a basis for $W$.

First, note that

$$\begin{aligned} 0 &= \sum_{i=1}^{n} c^i f(v_i) \\ &= f\left( \sum_{i=1}^{n} c^i v_i \right) \end{aligned}$$

implies $\sum_{i=1}^n c^i v_i \in \ker f$, but by injectivity, we know $\ker f = \{0\}$, hence $\sum_{i=1}^n c^i v_i = 0$, showing independence of $f(v_1), ..., f(v_n)$.

Surjectivity follows simialrly. Let $w \in W$. Since $f$ is surjective, there exist $v \in V$ such that $w = f(v)$. Since $\{v_1, ..., v_n\}$ is a basis for $V$, there exist $c^1, ..., c^n$ such that $v = \sum_{i=1}^n$ such that $v = \sum_{i=1}^n c^i v_i$. Since $f$ is linear, it follows that

$$\begin{aligned} w &= f(v) \\ &= f\left(\sum_{i=1}^n c^i v_i\right) \\ &= \sum_{i=1}^n c^i f(v_i), \end{aligned}$$

which shows that $\{f(v_1), ..., f(v_n)\}$ spans $W$ and is therefore a basis for $W$, proving that $\dim W = \dim V$. □

## 3.2 Matrices

When we invoke a basis $(v_1, ..., v_n)$ for a vector space $V$, we can write any vector $v$ by $c_1 v_1 + c_2 v_2 + ... + c_n v_n$. Define the following:

**Definition 3.9.** Given a vector space $V$, choose an ordered basis $v_1, ..., v_n$ for $V$. Define the *column vector representation* of $V$ to be the column vector:

$$\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$$

This is essentially the same thing that we did with our isomorphism from $V \to \mathbf{R}^n$.

Now, notice the following. Given a linear map $V \to W$, given a vector $v = \sum_{i=1}^n c_i v_i$, we can write $f(v)$ in terms of the $f(v_i)$ by the following:

$$\begin{aligned} f(v) &= f\left(\sum_{i=1}^n c_i v_i\right) \\ &= \sum_{i=1}^n c_i f(v_i). \end{aligned}$$

But, the $f(v_i)$ can again be written in terms of the basis $w_1, ..., w_m$ for $W$ by $f(v_i) = \sum_{j=1}^n \alpha_{ji} w_j$. We can record this information in an $n \times n$ matrix:

$$\begin{bmatrix} \alpha_{11} & \alpha_{21} & ... & \alpha_{n1} \\ \alpha_{12} & \alpha_{22} & ... & \alpha_{n2} \\ \vdots & \vdots & & \vdots \\ \alpha_{1n} & \alpha_{2n} & ... & \alpha_{nn} \end{bmatrix}.$$

Now, notice the following:

$$\begin{bmatrix} \alpha_{11} & \alpha_{21} & ... & \alpha_{n1} \\ \alpha_{12} & \alpha_{22} & ... & \alpha_{n2} \\ \vdots & \vdots & & \vdots \\ \alpha_{1n} & \alpha_{2n} & ... & \alpha_{nn} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \alpha_{11} \\ \alpha_{12} \\ \vdots \\ \alpha_{1n} \end{bmatrix} = f(v_1).$$

More generally, we see that the matrix multiplication corresponds exactly to the action of the linear map. By invoking a basis, then, the linear map is characterized exactly by its matrix. It takes quite a large deal of matrix bashing to prove these correspondences, so we will simply take this for granted, and pass between the operator perspective and matrix perspective as necessary. Indeed, this is formalized by the following theorem:

**Theorem 3.10.** Given $V$ $m$-dimensional and $n$ $m$-dimensional over $\mathbf{R}$, the vector spaces $\mathrm{Hom}(V, W)$ and $M^{m\times n}(\mathbf{R})$ are isomorphic.

**Example 3.11.** Let $V = F^n$ and $B = (e_1, e_2, ..., e_n)$, where $e_i$ is the $n$-tuple with a 1 in the $i$th position and 0s elsewhere. Then for $x = (x_1, x_2, ..., x_n) \in V$,

$$L_B(x) = L_B(x_1, ..., x_n) = x_1e_1 + ... + x_ne_n = (x_1, ..., x_n) = x,$$

showing $L_b = \mathrm{id}_{F^n}$ (in which case we have $L_B^{-1} = \mathrm{id}_{F^n}$).

**Example 3.12.** Let $V = \mathcal{P}_n(F)$ and $B = (1, x, x^2, ..., x^n)$. Then for $p = a_0 + a_1x + ... + a_nx^n \in \mathcal{P}_n(F)$,

$$[p]_B = L_B^{-1}(a_0 + a_1x + ... + a_nx^n) = (a_0 + a_1x + ... + a_nx^n) = (a_0, a_1, ..., a_{n+1}) \in F^{n+1}.$$

More information on matrix theory can be found in the appendix.

## 3.3 Dual Space

**Definition 3.13 (Linear Functional).** Let $V$ be a vector space over $\mathbf{R}, \mathbf{C}$. A linear function $f : V \to \mathbf{R}, \mathbf{C}$, $f \in \mathrm{Hom}(V, \mathbf{R}), \mathrm{Hom}(V, \mathbf{C})$ is called a *linear functional* on $V$.

**Example 3.14.** Let $V = \mathcal{P}(\mathbf{R})$ and fix $t \in \mathbf{R}$. Show that the *evaluation map* $\mathrm{ev}_t : V \to \mathbf{R}$ defined by $\mathrm{ev}_t(p) = p(t)$ is a linear functional.

We have $\mathrm{ev}_t(\alpha p + q) = (\alpha p + q)t = \alpha p(t) + q(t) = \alpha \mathrm{ev}_t(p) + \mathrm{ev}_t(q)$, showing linearity.

**Definition 3.15 (Dual space).** Let $V$ be a vector space over $F$. We define the *dual space* of $V$, denoted by $V^*$, to be the vector space $V^* = \mathcal{L}(V, F)$.

**Definition 3.16 (Dual Basis).** Let $\alpha^i : V \to \mathbf{R}$ be the linear function picking out the ith coordinate, such that $\alpha^i(v) = c^i$ when we write $v = \sum_{i=1}^n c^i v_i$.

Note that $\alpha^i$ is characterized by:

$$\alpha^i(e_j) = \delta^i_j$$

where $\delta_{ij}$ is the Krockener delta function, which is 1 when $i = j$ and zero otherwise.

**Proposition 3.17.** The functions $\alpha^1, ..., \alpha^n$ form a basis for $V^*$

*Proof.* We first prove span. If $f$ is a linear functional, and $v = \sum_{i=1}^n c^i v_i \in V$, then we have:

$$\begin{aligned} f(v) &= f\left(\sum_{i=1}^n c^i v_i\right) \\ &= \sum_{i=1}^n c^i f(v_i) \\ &= \sum_{i=1}^n \alpha^i(v) f(v_i), \end{aligned}$$

hence the $\alpha^i$ span $V^*$.

To show linear independence, suppose $\sum_{i=1}^n c_i \alpha^i = 0$ for some $c_i \in \mathbf{R}$, where 0 here is the 0 operator. Applying this to the basis vector $v_j$ gives

$$0 = 0(e_j) = \sum_{i=1}^n c_i \alpha^i(e_j) = \sum_{i=1}^n c_i \delta_j^i = c_j$$

for all $j = 1, ..., n$, showing the linear combination is zero and hence independence. □

This basis $\alpha^1, ..., \alpha^n$ is called the *dual basis* of $v_1, ..., v_n$.

**Corollary 3.18.** The dual space $V^*$ has the same dimension as $V$ for $V$ finite-dimensional.

The space of linear functions from $V$ to the underlying field (**R** or $C$) turns out to be of key importance in linear algebra in general. Given a linear transformation $L$ acting on $V$, we can define:

**Definition 3.19 (Transpose Transformation).** The transpose of $L$ is the linear transformation

$$L^t : V^* \to V^*$$

defined by

$$(L^t \circ l)(v) = (l \circ L)(v)$$

for $l \in V^*$, $v \in V$.

It can be shown that the matrix of the transpose $L^t$ is exactly the transpose of the matrix of $L$.

## 3.4 Bilinear Forms

It is often interesting to consider a function of multiple variables that is linear in both on a vector space. So, we make the following definition:

**Definition 3.20 (Bilinear form).** Let $V, W$ be vector spaces over $\mathbf{R}$. A *blinear form* on $V \times W$ is a function $B : V \times W \to \mathbf{R}$ which is linear in each variable separately when the other variable is held constant. That is, for all $v, v_1, v_2 \in V$, $w_1, w_2, w \in W$ and $\alpha \in F$, we have

$$B(\alpha v_1 + v_2, w) = \alpha B(v_1, w) + B(v_2, w)$$

and

$$B(v, \alpha w_1 + w_2) = \alpha B(v, w_1) + B(v, w_2).$$

**Example 3.21.** Show that $B : \mathcal{P}_n(\mathbf{R}) \times \mathcal{P}_n(\mathbf{R}) \to \mathbf{R})$ defined by $B(p, q) = p(2)q'(3)$ is a bilinear form on :$P_n(\mathbf{R})$.

We have $B(\alpha p_1 + p_2, q) = (\alpha p_1 + p_2)(2)q'(3)$ at which point this follows from linearity of the evaluation map (and we use a similar argument in $q$ to show linearity in the second component).

**Proposition 3.22.** Let $V$ be a vector space over $\mathbf{R}$ and choose linear maps $f, g$ from $V \to \mathbf{R}$, $W \to \mathbf{R}$ respectively. Then, the map $B(u, v) = f(u)g(v)$ is a bilinear form on $V \times W$. More generally, if $f_1, ..., f_n, g_1, ..., g_n \in V^*$, then $B(u, v) = \sum_{i=1}^{n} f_i(u)g_i(v)$ is a bilinear form on $V$.

*Proof.* We have:

$$\begin{aligned} B(au_1 + u_2, v) &= \sum_{i=1}^{n} f_i(au_1 + u_2)g_i(v) \\ &= \sum_{i=1}^{n} af_i(u_1)g_i(v) + f_i(u_2)v_i \\ &= a\sum_{i=1}^{n} f_i(u_1)g_i) + \sum_{i=1}^{n} f_i(u_2)g(v) \\ &= aB(u_1, v) + B(u_2, v), \end{aligned}$$

and showing linearity in the second component is identical. □

**Proposition 3.23.** Define $V^{(2)}$ denote the set of all bilinear forms on $V$. Show that this is a subspace of $F^{V \times V}$, the vector space of linear maps $V \times V \to F$.

*Proof.* The zero map is a bilinear form, so it suffices to show if $B_1, B_2$ are bilinear forms, so is $\alpha B_1 + B_2$. We have:

$$\begin{aligned} (\alpha B_1 + B_2)(au_1 + u_2, v) &= \alpha B_1(au_1 + u_2, v) + B_2(au_1 + u_2, v) \\ &= \alpha a B_1(u_1, v) + \alpha B_1(u_2, v) + aB_2(u_1, v) + B_2(u_2, v) \\ &= a(\alpha B_1(u_1, v) + B_2(u_1, v)) + \alpha B_1(u_2, v) + B_2(u_2, v) \\ &= a(\alpha B_2 + B_2)(u_1, v) + (\alpha B_1 + B_2)(u_2, v), \end{aligned}$$

and the other direction again follows similarly. □

Suppose now that $V$ is a finite-dimensional vector space with $\dim V = n$. We will now prove that $V^{(2)}$ is also finite-dimensional with $\dim V^{(2)} = n^2$.

**Definition 3.24 (Matrix of a Bilinear Map).** Let $B$ be a bilinear form on $V$ and let $(e_1, ..., e_n)$ be an ordered basis of $V$. Define the *matrix* of $B$ with respect to this ordered basis to be the matrix $[B]$ whose $ij$-entry was given by

$$[B]_{ij} = B(v_i, v_j).$$

**Remark 3.25.** A different way of thinking about this is considering a "dual basis" $e^*_{i,j}$ given by $e^*_{i,j}(e_p, e_q) = \delta_{ip}\delta_{jq}$. This is really the same thing as above (since this is a basis for the space of matrices).

**Example 3.26.** Find the matrix of the bilinear form $B : \mathcal{P}_2(\mathbf{R}) \times \mathcal{P}_2(\mathbf{R}) \to \mathbf{R}$ of the prior exercise defined by $B(p, q) = p(2)q'(3)$ with respect to the standard ordered basis of $\mathcal{P}_2(\mathbf{R})$.

We have:

$$\begin{bmatrix} 0 & 1 & 6 \\ 0 & 2 & 12 \\ 0 & 4 & 24 \end{bmatrix}$$

**Theorem 3.27.** The map $T : V^{(2)} \to M^{n\times n}(F)$ defined by $T(B) = [B]$ is an isomorphism. This proves that $V^{(2)}$ is finite-dimensional with $\dim V^{(2)} = n^2 = (\dim V)^2$.

*Proof.* Using the dual basis characterization (naturally isomorphic to $M^{n\times n}$ by dimension), it is clear that the dual basis elements are bilinear forms themselves and independent (using essentially the same argument as for showing the dual basis is linear and independent), so all that suffices to show is surjectivity.

But, every bilinear map is defined by its value on the dual bases, and so the map is naturally surjective, and we are done. □

## 3.5 Tensor Products

Now, in the same way that we got the sum of two vector spaces, it turns out we can get the product of two vector spaces.

**Definition 3.28.** The tensor product $U \otimes V$ of two vector spaces is the dual vector space of the space $B(U, V)$ of bilinear forms on $U \times V$. We define $z = x \otimes y$ to be the element of $U \otimes V$ such that $z(w) = w(x, y)$ for every bilinear form $w$.

**Theorem 3.29.** If $X = \{x_1, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_m\}$ are bases in $U$ and $V$ respectively, then the set $Z = \{x_i \otimes y_j \mid 1 \le i \le n,\ 1 \le j \le m\}$ is a basis of $U \otimes V$.

*Proof.* First, we show that the set $Z = \{x_i \otimes y_j\}$ spans $U \otimes V$. Let $z \in U \otimes V$ be any element, which corresponds to a linear functional on $B(U, V)$. Since any bilinear form $w \in B(U, V)$ is determined by its values on the pairs $(x_i, y_j)$, the functionals $z_{ij}(w) = w(x_i, y_j)$ span the dual space. So every $z$ is a linear combination of the $x_i \otimes y_j$.

To show linear independence, suppose

$$\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}(x_i \otimes y_j) = 0$$

in $U \otimes V$. Then for every bilinear form $w$,

$$\sum_{i,j} a_{ij} w(x_i, y_j) = 0.$$

Define $w$ by setting $w(x_i, y_j) = \delta_{ii_0}\delta_{jj_0}$ for fixed $i_0, j_0$, and $w$ extended bilinearly. Then the sum becomes $a_{i_0 j_0} = 0$. Since this holds for all $i_0, j_0$, all $a_{ij} = 0$.

Thus $Z$ is linearly independent and spans $U \otimes V$, so it is a basis. □

It turns out that this product is associative, meaning;

**Proposition 3.30.** Given vector spaces $U, V, W$, $(U \otimes V) \otimes W \cong U \otimes (V \otimes W)$.

As such, we will drop the parentheses and freely write tensor products of $n$-dimensional spaces.

These will not really show up again until the very end of the course (when we discuss composite systems), so feel free to ignore them for now and refer back later.

# 4 Linear Algebra III

## 4.1 Permutations

Here we summarize some facts about permutations.

- A *permutation* of the set $A = \{1, 2, ..., m\}$ is a bijection $\sigma : A \to A$, which can be thought of as a reordering of the list $1, 2, ..., m$ to a new order $\sigma(1), ...\sigma(m)$. A convienient way to describe a permutation $\sigma$ is by a matrix

$$\begin{bmatrix} 1 & 2 & ... & m \\ \sigma(1) & \sigma(2) & ... & \sigma(m) \end{bmatrix}$$

  where the inputs form the first row and the corresponding outputs form the second row. For example the permutation $1 \to 3, 2 \to 1, 3 \to 2$ is represented by the matrix

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix}$$

  Note that nothing is multiplied by this matrix; it just keeps track of the inputs/outputs.

- Permutations of $A$ form a *group* under composition (which is an associative operation), denoted by $S_m$) [1]. For $\sigma, \tau \in S_m$, we write $\tau\sigma$ for the composition $\tau \circ \sigma$. The identity element of the group is the trivial permutation $\left[\begin{smallmatrix} 1 & 2 & ... & m \\ 1 & 2 & ... & m \end{smallmatrix}\right]$. We denote the inverse of a permutation $\sigma$ by $\sigma^{-1}$.

- A *cyclic* permutation or *r-cycle*, denoted $(a_1\, a_2\, ...\, a_r)$ where the $a_i$ are distinct, is the permutation $\sigma$ such that $\sigma(a_i) = a_{i+1} = a_{i+1}$ for $i = 1, ..., r$, $\sigma(a_r) = a_1$, and $\sigma(a_l) = a_l$ for $l > r$. A 2-cycle is also called a *transposition*.

- Cycles $(a_1 ... a_r)$ and $(b_1 ... b_r)$ are disjoint if the sets $\{a_1, ..., a_r\}$ and $\{b_1, ..., b_r\}$ have no elements in common.

- Any permutation $\sigma$ can be written as a product of disjoint transpositions. We say that $\sigma$ is *even* or *odd* depending on whether it is the product of an even or odd number of transpositions.

- We define the *sign* of a permutation $\sigma$, denoted by $\text{sgn}(\sigma)$, to be $+1$ or $-1$ depending on whether the permutation is even or odd. This satisfies $\text{sgn}(\tau\sigma) = \text{sgn}(\tau)\text{sgn}(\sigma)$.

**Example 4.1.** Consider the permutation

$$\sigma = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 5 & 1 & 3 \end{bmatrix} \in S_5$$

[1] A group is a set $G$ with an associative binary operation, such that $G$ has an identity element, and every element of $G$ has an inverse

To write $\sigma$ as a product of disjoint cycles, start with any element of $\{1,2,3,4,5\}$, say 1, and apply $\sigma$ to it repeatedly until we return to the initial element; this gives a cycle: $1 \to 2 \to 4 \to 1$, or $(124)$. Next, repeat this procedure beginning with any of the remaining elements, say 3, to get a second cycle: $3 \to 5 \to 3$, or $(35)$. Since all elements of $\{1,2,3,4,5\}$ are now accounted for, $\sigma = (1\,2\,4)(3\,5)$. Noting that the product $(1\,4)(1\,2)$ also sends $1 \to 2 \to 4 \to 1$, we have $(124) = (1\,4)(1\,2)$, and therefore $\sigma = (1\,4)(1\,2)(3\,5)$, so we see that $\text{sgn}(\sigma) = -1$.

**Example 4.2.** More generally, the decomposition

$$(a_1 \ldots a_r) = (a_1\, a_r)(a_1\, a_{r-1})\ldots(a_1\, a_3)(a_1\, a_2)$$

shows that an $r$-cycle is an even permutation if and only if $r$ is odd, and an odd permutation if and only if $r$ is even. Thus, a faster way to compute the sign of a permutation is to decompose it into a product of cycles and to count the number of cycles of even length. For example, the permutation $\sigma = (1\,2\,4)(3\,5)$ in Example 3.1 is odd because $(1\,2\,4)$ is odd and $(3\,5)$ is even.

## 4.2 Multilinear Forms

It turns out we can generalize the ideas and results from bilinear forms to $m$-linear forms on an $n$-dimensional vector space.

Let $V$ be a vector space over $F$. For $m$ a positive integer, let $V^m = V \times \ldots \times V$ denote the $m$-fold Cartesian product of $V$ with itself. That is, $V^m$ is the set of all ordered $m$-tuples of elements in $V$.

**Definition 4.3 (Multilinear form).** An $m$-linear form on $V$ is a map $M : V^m \to F$ which is linear in each entry when all other entries are held fixed. That is, for each $k \in \{1, \ldots, m\}$ and all $u_1, \ldots, u_m \in V$, the map

$$v \to M(u_1, \ldots, u_{k-1}, v, u_{k+1,\ldots,u_m})$$

is an element of $V^*$.

The set of $m$-linear forms on $V$ is denoted by $V^{(m)}$. An element of $V^{(m)}$ for some $m$ is called a *multilinear form*. A 1-linear form is the same as an element of $V^*$. A 2-linear form is the same as a bilinear form.

By a proof analagous to the one for $V^{(2)}$, one can prove that $V^{(m)}$ is a vector space for every positive integer $m$ with respect to the usual operations

$$(M_1 + M_2)(v_1, \ldots, v_m) := M_1(v_1, \ldots v_m) + M_2(v_1, \ldots, v_m)$$
$$(\alpha M)(v_1, \ldots, v_m) := \alpha M(v_1, \ldots, v_m).$$

**Example 4.4.** Suppose $B_1, B_2 \in V^{(2)}$. Define $M : V^4 \to F$ by letting $M(v_1, v_2, v_3, v_4)$ equal $B_1(v_1, v_2)B_2(v_3, v_4)$. Show that $B_4 \in V^{(4)}$.

*Proof.* Fixing three entries, we see that we have a map $B(u, v_i)$ which is linear since $B$ is bilinear, and multiplication by $B(u_2, u_3)$ which is a constant. □

**Example 4.5.** Define $M : (\mathcal{L}(V))^m \to F$ by $M(T_1, ..., T_m) = \text{Tr}(T_1...T_m)$. Show that $M \in V^{(m)}(\mathcal{L}(V))$.

## 4.3 Permutation Action on Multilinear Forms

**Definition 4.6 (Group action).** Let $G$ be a group and $X$ be a set. A *left group action* is a mapping

$$G \times X \to X(g, x) \to gx$$

which satisfies the following two properties:

1. $ex = x$, where $e$ is the identity element of $G$ and $x \in X$.
2. $h(gx) = (hg)x$ for all $g, h \in G$ and $x \in X$.

A right group action is defined similarly.

We can define a left action of $S_m$ on $V^{(m)}$ as

$$S_m \times V^{(m)} \to V^{(m)}(\sigma, M) = \sigma M,$$

where

$$(\sigma M)(v_1, ..., v_k) = M(v_{\sigma(1)}, ..., v_{\sigma(m)}).$$

Note this is indeed a left action of $S_m$ on $V^{(m)}$, since

1. If $\sigma$ is the trivial permutation, then

   $$(\sigma M)(v_1, ..., v_k) = M(v_{\sigma(1)}, ..., v_{\sigma(m)}) = M(v_1, ..., v_k),$$

   so $\sigma M = M$.

2. If $\sigma, \tau \in S_m$, then

$$\begin{aligned}
\tau(\sigma M)(v_1, ..., v_m) &= (\sigma f)(v_{\tau(1)}, ..., v_{\tau(m)}) \\
&= (\sigma M)(w_1, ..., w_m) \\
&= M(w_{\sigma(1)}, ..., w_{\sigma(m)}) \\
&= m(v_{\tau(\sigma(1))}, v_{\tau(\sigma(m))}) \\
&= M(v_{\tau\sigma(1)}, ...v_{\tau\sigma(m)}) \\
&= (\tau\sigma)M(v_1, ..., v_m).
\end{aligned}$$

**Proposition 4.7.** $M \in V^{(m)}$ is alternating if and only if $\sigma M = \text{sgn}(\sigma)M$ for all $\sigma \in S_m$.

*Proof.* If $\sigma M = \text{sgn}(\sigma)M$ for each $\sigma \in S_m$, then in particular, when $\sigma$ is a transposition we have $\sigma M = \text{sgn}(\sigma)M = -M$. Hence, $M$ is alternating.

Conversely, suppose now that $M$ is alternating and let $\sigma \in S_m$. As noted in the previous section, $\sigma$ can be written as a product of disjoint transpositions. Since $M$ picks up a negative sign from each transposition, we will have $\sigma M = M$ if $\sigma$ is even and $\sigma M = -M$ if $\sigma$ is odd. Hence, $\sigma M = \text{sgn}(\sigma)M$. □

## 4.4 Alternating Multilinear Forms

**Definition 4.8 (Alternating Multilinear Form).** An $m$-linear form $M$ is alternating if

$$M(v_1, ..., v_i, ..., v_j, ..., v_m) = -M(v_1, ..., v_j, ..., v_i, ..., v_m)$$

whenever we interchange two vectors $v_i$ and $v_j$. We denote the set of alternating $m$-linear forms by $V_{\text{alt}}^{(m)}$.

Since a 1-linear map is trivially alternating, $V^{(1)} = V^*$.

The following is an equivalent definition of an alternating $m$-linear form.

**Proposition 4.9.** $M \in V^{(m)}$ is alternating if and only if $M(v_1, ..., v_m) = 0$ whenever $v_i = v_j$ for two distinct $i, j \in \{1, ...m\}$.

*Proof.* ($\Rightarrow$) Suppose $M$ is alternating. Then swapping two equal arguments gives

$$M(\dots, v_i, \dots, v_j, \dots) = -M(\dots, v_j, \dots, v_i, \dots),$$

but if $v_i = v_j$, both sides are equal, so $M = -M$ and hence $M = 0$.

($\Leftarrow$) Suppose $M(v_1, \dots, v_m) = 0$ whenever $v_i = v_j$ for $i \neq j$. Consider a transposition $(i\ j)$ and define

$$f(w) := M(\dots, w, \dots, w, \dots),$$

with $w$ in positions $i$ and $j$. Since $f(w) = 0$ for all $w$, plug in $w = v_i + v_j$ and use multilinearity:

$$0 = M(\dots, v_i, \dots, v_j, \dots) + M(\dots, v_j, \dots, v_i, \dots),$$

so $M(\dots, v_j, \dots, v_i, \dots) = -M(\dots, v_i, \dots, v_j, \dots)$. Thus $M$ is alternating. □

**Proposition 4.10.** If $M \in V_{\text{alt}}^{(m)}$ and $\{v_1, .., v_m\}$ is linearly dependent, then $M(v_1, ..., v_m) = 0$.

*Proof.* If the vectors are linearly dependent, then some $v_i = \sum_{j \neq i} a_j v_j$. By multilinearity,

$$M(v_1, \dots, v_m) = \sum_{j \neq i} a_j M(\dots, v_j, \dots, v_m),$$

where each term has repeated arguments. Since $M$ is alternating, each term is 0, so $M(v_1, \dots, v_m) = 0$. □

**Corollary 4.11.** If $m > \dim V$, then $\dim V_{\text{alt}}^{(m)} = 0$.

*Proof.* All sets of vectors with $m > \dim V$ elements are dependent, so the proposition says $V_{\text{alt}}^{(m)}$ acts trivially on $V$ and so only contains the zero map $\mathbf{0}$, so it is the trivial space. □

# 5 Linear Algebra IV

## 5.1 Alternating $n$-Linear Forms

Given an $n$-dimensional vector space, we now want to consider the space $V^{(n)}$ of alternating $n$-linear forms on $V$. We have the following theorem:

**Proposition 5.1.** Let $V$ be an $n$-dimensional vector space and let $\{e_1, ..., e_n\}$ be a basis of $V$. For any $v_1, ..., v_n \in V$, there exist scalars $b_{jk}$ such that $v_k = \sum_{j=1}^{n} b_{jk} e_j$. If $M \in V^{(n)}$, then

$$M(v_1, ..., v_n) = M(e_1, ..., e_n) \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) b_{\sigma(1)1}, ....b_{\sigma(n)n}$$

*Proof.* We have:

$$M(v_1, ..., v_n])M(\sum_{j=1}^{n} b_j e_j, ...., \sum_{j=1}^{n} b_{jn} e_j).$$

By multilinearity, we can split this into a sum over $M(e_i)$ by choosing one term $e_k$ in each component of the input. The ways we can do this are precisely the permutations in $S_m$, and the $\sigma$ on the coefficients arises from which coefficient we pull out. □

As such, we see that up to a scalar multiple, there is exactly one nontrivial $n$-linear form on $V$, since every map $M$ is determined by $M(e_1, ..., e_n)$.

Choosing this form, knowing the value of $M(v_1, ..., v_n$ for $n$ independent vectors lets us know $M$ for everything. It then becomes interesting to ask what $M(Tv_1, ..., Tv_n)$ is.

## 5.2 Determinants

**Theorem 5.2.** Let $m$ be a positive integer and $M \in V_{\text{alt}}^{(m)}$. Given $T \in \mathcal{L}(V)$, define $M_T \in V_{\text{alt}}^{(m)}$ by

$$M_T(v_1, ..., v_m) = M(Tv_1, ..., Tv_m)$$

for all $(v_1, ..., v_m) \in V^m$. We have the following:

(a) $M_T \in V_{\text{alt}}^{(m)}$

(b) The map $M \to M_T$ is linear.

(c) If $m = n = \dim V$, then there exists a unique scalar $\alpha_T \in F$ such that $M_T = \alpha_T M$. That is,

$$M_T(v_1, ..., v_n) = \alpha_T M(v_1, ..., v_n)$$

for all $(v_1, ..., v_n) \in V^n$ and all $M \in V_{\text{alt}}^{(n)}$.

*Proof.* We have:

(a) We have:

$$\begin{aligned} M_T(v_1, ..., v_i, ..., v_j, ..., v_m) &= M(Tv_1, ..., Tv_i, ..., T_{v_j}, ..., T_{v_m}) \\ &= -M(Tv_1, ..., Tv_j, ...., Tv_i, ..., Tv_m) \\ &= M_T(v_1, ..., v_j, ..., v_i, ..., v_m) \end{aligned}$$

showing this for transpositions (and therefore all permutations since transpositions generate permutations and we have already shown how compositions of permutations works).

(b) Since this is linear in each component $T(\alpha v_{11} + v_{12}) = \alpha T v_{11} + T v_{12}$, this is obvious.

(c) Since the space $V_{\text{alt}}^{(n)}$ is 1-dimensional, we have that $M_T$ and $M$ must be scalar multiples of each other as vector space maps, and this scalar must clearly be unique since the maps are identical.

□

**Definition 5.3 (Determinant of a Linear Operator).** For $T \in \mathcal{L}(V)$, the *determinant of* $T$, denoted by $\det T$, is defined to be the unique scalar such that $M_T = (\det T)M$ for all $M \in V_{\text{alt}}^{(\dim V)}$.

**Example 5.4.** Let $n = \dim V$ and let $\text{id}_V$ be the identity operator on $V$. Then, for any $M \in V_{\text{alt}}^{(n)}$ and any $(v_1, ..., v_n) \in V^n$ we have

$$M_{\text{id}_V}(v_1, ..., v_n) = M(\text{id}_V v_1, ..., \text{id}_V v_n) = M(v_1, ..., v_n),$$

hence $\det(\text{id}_V) = 1$.

Notice that by applying the proposition that we used to originally conclude $\dim V^{(n)} = 1$, we had the following fact:

$$M(v_1, ..., v_n) = M(e_1, ..., e_n) \sum_{\sigma \in S_n} \text{sgn}(\sigma) b_{\sigma(1)1}, ...., b_{\sigma(n)n},$$

where the $b_{jk}$ came from writing $v_k$ in terms of the basis $e_1, ..., e_n$. But, this is exactly the same decomposition we used to translate linear maps to matrices. Therefore, we see the following:

**Proposition 5.5 (Formula for the determinant of a matrix).** Let $n$ be a positive integer and $A$ an $n \times n$ matrix. Then

$$\det(A) = \sum_{\sigma \in S_n} \text{sgn}(\sigma) A_{\sigma(1)1} ... A_{\sigma(n)n}. \tag{5.1}$$

This provides a concrete way to calculate determinants. We see the following formulas:

**Example 5.6.** Let $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$. Then:

$$\det(A) = \sum_{\sigma \in S_2} \operatorname{sgn}(\sigma) A_{\sigma(1)1} A_{\sigma(2)2} = A_{11}A_{22} - A_{21}A_{22}.$$

and:

**Example 5.7.** The permutations in $S_3$ (with sign) are:

$$(1), -(1\,2), -(2\,3), -(1\,3), (1\,2\,3), (1\,3\,2),$$

and so we have:

$$\begin{aligned} \det(A) &= \sum_{\sigma \in S_3} A_{\sigma(1)1} A_{\sigma(2)2} A_{\sigma(3)(3)} \\ &= (A_{11}A_{22}A_{33}) - (A_{21}A_{12}A_{33}) \\ &- (A_{31}A_{22}A_{13}) - (A_{11}A_{23}A_{32}) \\ &+ (A_{21}A_{32}A_{13}) + (A_{31}A_{12}A_{23}) \end{aligned}$$

There is an appendix on matrix theory and calculations.

## 5.3 Geometric Interpretation of the Determinant

The determinant seems like a perhaps meaningless construction. However, consider the linear map $M$ defined by $M(e_1, ..., e_n) = 1$. Then, given the transformation $T$ mapping $e_1 \to v_1, ..., e_n \to v_n$, the determinant of $T$ is exactly $M(Te_1, ..., Te_n)/M(e_1, ..., e_n) = M(v_1, ..., v_n)$. When the $v_i$ are independent, indicating that $T$ maps a basis to a basis (hence being an isomorphism) this is nonzero. When the $v_i$ are dependent, this is zero. Therefore, we see that a transformation has nonzero determinant if and only if it is an isomorphism.

Geometrically, the determinant can be interpreted as the $n$-volume of an object in $n$-dimensional space. In this sense, whenever we don't have $n$-independent vectors, the object created by these vectors is degenerate and has zero volume.

♣ Draw pictures to illustrate this.

## 5.4 Eigenvectors and Eigenvalues

What is the simplest way in which an operator $T$ can act on a vector? The natural answer is to have $Tv = \lambda v$ for some scalar $\lambda$, i.e $Tv$ multiplies $v$ by some scalar. When we can find a basis $e_1, ..., e_n$ of $V$ such that $T(e_i) = \lambda_i e_i$ for all $i$, it then becomes very simple to calculate quantities. Since $T(e_i) = \lambda_i e_i$, the matrix of $T$ will take the form:

$$\begin{bmatrix} \lambda_1 & 0 & ... & 0 \\ 0 & \lambda_2 & ... & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & ... & \lambda_n, \end{bmatrix}$$

making it very easy to calculate the determinant of the operator as well. So, we have the following questions:

- Can each linear operator $T : V \to V$ be represented by a diagonal matrix?
- If not, for which operators $T$ does such a basis exist?
- How can we find such a basis?

We call these vectors $v$ eigenvectors of $T$.

**Definition 5.8 (Eigenvectors and Eigenvalues).** Let $V$ be a vector space and $T : V \to V$. An *eigenvalue* of $T$ is a scalar $\lambda$ such that there is a nonzero vector $v \in V$ such that

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

The vector $v$ is called an eigenvector of $T$ associated with the eigenvalue $\lambda$.

**Definition 5.9 (Eigenvalue of an $n \times n$ matrix).** An eigenvalue of an $n \times n$ matrix $A$ is a scalar $\lambda$ such that there exists a nonzero vector $v \in \mathbf{R}^n$ such that

$$Av = \lambda v$$

The vector $v$ is called an eigenvector of $A$ associated with the eigenvalue $\lambda$.

**Example 5.10.** Let $A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$, $u = \begin{bmatrix} 6 \\ -5 \end{bmatrix}$, and $v = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$. Are $u$ and $v$ eigenvectors of $A$? If so, find the corresponding eigenvalues.

It turns out that matrices in different bases representing the same operator have the same eigenvalues. This takes some matrix theory to prove, so we'll ignore it.

We now discuss how to systematically find eigenvalues of a given matrix. A scalar $\lambda$ is an eigenvalue of $A$ if and only if $Av = \lambda v$ has a nontrivial solution. This is equivalent to the homogenous linear system

$$(A - \lambda I)v = 0, \tag{5.2}$$

where $I$ is the $n \times n$ identity matrix.

**Definition 5.11 (Eigenspace).** The set of all solutions of 1.10 is called the *eigenspace* of $A$ associated to the eigenvalue $\lambda$, which we will denote by $E_\lambda$.

**Proposition 5.12.** Suppsoe $A \in M^{n\times n}(\mathbf{R})$ has an eigenvalue $\lambda$. Then the eigenspace of $A$ corresponding to $\lambda$ is a subspace of $\mathbf{R}^n$.

*Proof.* $0$ is an eigenvector of eigenvalue $\lambda$, since $A(0) = \lambda 0$, so $0 \in E_\lambda$.

If $v_1, v_2 \in E_\lambda$, then $A(v_1 + cv_2) = Av_1 + cAv_2 = \lambda v_1 + c\lambda v_2 = \lambda(v_1 + cv_2)$, showing $v_1 + cv_2 \in E_\lambda$. □

Note that the equation $(A - \lambda I)v = \vec{0}$ has a nontrivial solution if and only if $\det(A - \lambda I) = 0$, or equivalently, if $\det(\lambda I - A) = 0$.

**Theorem 5.13.** Let $A$ be an $n \times n$ matrix. Then $\det(\lambda I - A)$ is a monic degree $n$ polynomial.

*Proof.* See the appendix on matrix theory. □

**Definition 5.14.** Given any $n \times n$ matrix $A$, the monic, degree $n$ polynomial $\det(\lambda I - A)$ is called the *characteristic polynomial* of $A$. The equation

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

is called the characteristic equation of $A$.

**Example 5.15.** Let $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. Then the characteristic equation for $A$ is

$$0 = \det(\lambda I - A) = \begin{vmatrix} \lambda & 1 \\ -1 & \lambda \end{vmatrix} = \lambda^2 + 1,$$

which has solution $\pm i$.

**Theorem 5.16.** If $V$ is a complex vector space, then every linear operator has an eigenvalue (and thus an eigenvector).

*Proof.* By the fundamental theorem of algebra, the polynomial $\det(\lambda I - A) = 0$ has a root in $\mathbf{C}$ for any matrix $A$ representing $T$. □

We can see that this theorem is false if $V$ is a real vector space via the following example.

**Example 5.17.** Let $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. Then the characteristic equation for $A$ is:

$$0 = \det(\lambda I - A) = \begin{vmatrix} \lambda & 1 \\ -1 & \lambda \end{vmatrix} = \lambda^2 + 1,$$

which has solution $\pm i$.

**Corollary 5.18.** An $n \times n$ matrix $A$ has at most $n$ distinct eigenvalues.

*Proof.* By the fundamental theorem of algebra, every non-constant polynomial has $n$ roots in $\mathbf{C}$ (counted with multiplicity). □

The next example shows how to find the eigenvalues and corresponding eigenspaces of a given $n \times n$ matrix.

**Example 5.19.** Let $A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$. The characteristic equation of $A$ is given by:

$$\begin{aligned} 0 = \begin{vmatrix} 1-\lambda & 6 \\ 5 & 1-\lambda \end{vmatrix} &= (1-\lambda)(2-\lambda) - 30 \\ &= (\lambda - 7)(\lambda + 4) \end{aligned}$$

so $A$ has two eigenvalues $\lambda = 7, -4$. To find a parametric description of $W_7$, we row-reduce the matrix $A - 7I = \begin{bmatrix} -6^6 & \\ 5 & -5 \end{bmatrix}$, giving $\begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}$ which shows that $E_7 = \text{span}\{(1,1)\}$. Geometrically, this is the line $y = x$ in the plane.

**Exercise 5.1.** Work out $E_{-4}$ from the previous example.

The row-reduction here is standard matrix theory which can be found in the appendix.

**Example 5.20.** Let $A = \begin{bmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{bmatrix}$. The characteristic equation of $A$ is given by

$$0 = \det(\lambda I - A) = \lambda^3 - 5\lambda^2 + 8\lambda - 4.$$

Using the rational root thoerem, we check to see that 1 is a root, and we geT:

$$\lambda^3 - 5\lambda^2 + 8\lambda - 4 = (\lambda - 1)(\lambda^2 - 4\lambda + 4) = (\lambda - 1)(\lambda - 2)^2 = 0,$$

and therefore the distinct eigenvalues of $A$ are $\lambda = 1, 2$.

**Example 5.21.** Show that the eigenspaces corresponding to the eigenvalues found in the previous example are

$$W_1 = \text{span}\{ \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \}, W_2 = \text{span}\{ \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \}.$$

First, we do $\lambda = 1$. In this case, we have:

$$A - 1 \cdot I = \begin{bmatrix} 2 & 1 & -1 \\ 2 & 1 & -1 \\ 2 & 2 & -1 \end{bmatrix},$$

and row-reducing this gives:

$$\begin{bmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & 0 \\ 0 & 0 & 0, \end{bmatrix}$$

so the solution set is spanned by:

$$\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}.$$

For $\lambda = 2$, we get:

$$A - 2 \cdot I = \begin{bmatrix} 1 & 1 & -1 \\ 2 & 0 & -1 \\ 2 & 2 & -2 \end{bmatrix}$$

which row-reduces to:

$$\begin{bmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 0, \end{bmatrix}$$

so the solution set is spanned by:

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

# 6 Linear Algebra V

Now, we must talk about a special kind of vector space - inner product spaces. These are the natural spaces for quantum mechanics and come with a great deal of structure. We will conclude by stating a classification of what operators have a basis of real-valued eigenvectors, which turns out to be key in quantum mechanics.

## 6.1 Inner Product Spaces

In the real case, we have:

**Definition 6.1 (Real Inner Product Space).** An real inner product space $V$ is a vector space over $\mathbf{R}$ equipped with a mapping $\langle \cdot, \cdot | \cdot, \cdot \rangle : V \times V \to \mathbf{R}$ called an inner porduct that satisfies:

1. $\langle x, y | x, y \rangle = \langle y, x | y, x \rangle$
2. $\langle \alpha x + \beta y, z | \alpha x + \beta y, z \rangle = \alpha \langle x, z | x, z \rangle + \beta \langle y, z | y, z \rangle$,
3. $\langle x, x | x, x \rangle \geq 0$.
4. $\langle x, x | x, x \rangle = 0$ implies $x = 0$.

for $x, y, z \in V$, $\alpha, \beta \in \mathbf{R}$.

**Definition 6.2.** A complex inner product space is a vector space over $\mathbf{C}$ equipped with a mapping $\langle \cdot, \cdot | \cdot, \cdot \rangle : V \times V \to \mathbf{C}$ satisfying all the properties of the inner product on the real vector space, but requiring that $\langle x, y | x, y \rangle = \langle y, \bar{x} | y, x \rangle$ instead of $\langle x, y | x, y \rangle = \langle y, x | y, x \rangle$.

Notice these definitions basically generalize the idea of the dot product in $\mathbf{R}^2, \mathbf{R}^3$.

**Example 6.3 (Examples of Inner Product Spaces).** The following are inner product spaces

(a) $\mathbf{R}^n$ with the standard dot product $(x_1, ..., x_n) \cdot (y_1, ..., y_n) = x_1 y_1 + ... + x_n y_n$.

(b) $\mathbf{C}^n$ with $\langle (x_1, ..., x_n), (y_1, ..., y_n) | (x_1, ..., x_n), (y_1, ..., y_n) \rangle = \sum_{k=1}^{n} x_k \bar{y}_k$.

(c) The space $\mathcal{C}([a, b])$ of continuous complex-valued functions on the interval $[a, b]$ together with $\langle f, g | f, g \rangle = \int_a^b f(x) \bar{g(x)} \, \mathrm{d}x$.

As we can see from $\mathbf{R}^n, \mathbf{C}^n$ being inner product spaces and any finite-dimensional real (resp. complex) vector space being isomorphic to one of these, we see that every finite-dimensional vector space is an inner product space.

An inner product provides a notion of length squared for vectors by defining $\|v\| = \langle v, v | v, v \rangle$, which satisfies all the properties of a *metric*. This isn't particularly relevant for our purposes, so we'll ignore it (ask me later if you're interested).

**Proposition 6.4.** Given an inner product space $V$, the map $v \in V \to l_v \in V^*$ defined by $l_v(w) = \langle v, w | v, w \rangle$ is an isomorphism of $V$ and $V^*$.

Notice that we showed a basis-dependent isomorphism earlier between $V$ and $V^*$ by identifying the basis and its dual basis, but this is a basis-free or *natural* isomorphism.

Physicists use a special notation called "bra-ket" notation in inner product spaces due to Dirac that makes physical calculations very clear. An element of a complex inner product space $V$ is written as a "ket" vector $|a\rangle$ where $a$ is a label for a vector in $V$.

An element of the dual space $V^*$ is written as a "bra vector" $\langle a|$ with the labelling in terms of $a$ determined by the isomorphism from the prior propisiton, i.e $\langle a| = l_{|a\rangle}$.

Evaluating $\langle a|$ in the dual space on $\beta \in V$ gives an element of $\mathbf{C}$ written by:

$$\langle \alpha| \left(|B\rangle\right) = \langle \alpha \mid \beta|\alpha \mid \beta\rangle \,,$$

where by the isomorphism from the proposition we have $\langle \alpha \mid \beta|\alpha \mid \beta\rangle = \langle |\alpha\rangle \,, |\beta\rangle||\alpha\rangle \,, |\beta\rangle\rangle$. The notation bra-ket is a play on words, since the inner product function $\langle \cdot, \cdot|\cdot, \cdot\rangle$ is called a bracket.

**Definition 6.5.** Two vectors $v, w$ in a inner product space $V$ are orthogonal if $\langle v, w|v, w\rangle = 0$.

**Definition 6.6.** An orthonormal basis $e_1, ..., e_n$ is a basis for $V$ satisfying $\langle e_i, e_j|e_i, e_j\rangle = \delta_{ij}$.

These turn out to be convenient bases, since when we write $v = \sum_{i=1}^n c_i e_i$, we see:

$$\langle v, e_j|v, e_j\rangle = \left\langle \sum_{i=1}^n c_i e_i, e_j \middle| \sum_{i=1}^n c_i e_i, e_j \right\rangle = \sum_{i=1}^n c_i \langle e_i, e_j|e_i, e_j\rangle = \sum_{i=1}^n c_i \delta_i j = c_j,$$

so we see that the dual basis function $\alpha^i$ is exactly $\langle \cdot, e_i|\cdot, e_i\rangle$.

Given an orthonormal basis $\{e_j\}$, a useful ket-notation is to let $|j\rangle = e_j$. Then, the above statement is exactly given by $\langle j \mid \alpha|j \mid \alpha\rangle$. As such, we can expand a vector $|\alpha\rangle$ in terms of the basis by

$$|\alpha\rangle = \sum_{j=1}^n |j\rangle \langle j \mid \alpha|j \mid \alpha\rangle$$

Conversely, for dual elements $\langle a| \in V^*$, we see:

$$\langle a| = \sum_{j=1}^n \langle a \mid j|a \mid j\rangle \langle j| \,.$$

The column vector for $\langle \alpha|\alpha\rangle$ is then

$$\begin{pmatrix} \langle 1 \mid \alpha|1 \mid \alpha\rangle \\ \langle 2 \mid \alpha|2 \mid \alpha\rangle \\ \vdots \\ \langle n \mid \alpha|n \mid \alpha\rangle \end{pmatrix}$$

and the matrix corresponding to the linear operator $\langle a|$ is:

$$\begin{pmatrix} \langle 1 \mid \alpha|1 \mid \alpha\rangle & \langle 2 \mid \alpha|2 \mid \alpha\rangle & ... & \langle n \mid \alpha|n \mid \alpha\rangle \end{pmatrix}$$

and the inner product is naturally the matrix product:

$$\langle a \mid b|a \mid b\rangle = \begin{pmatrix}\langle 1 \mid \alpha|1 \mid \alpha\rangle & \langle 2 \mid \alpha|2 \mid \alpha\rangle & \dots & \langle n \mid \alpha|n \mid \alpha\rangle\end{pmatrix} \begin{pmatrix}\langle 1 \mid \beta|1 \mid \beta\rangle \\ \langle 2 \mid \beta|2 \mid \beta\rangle \\ \vdots \\ \langle n \mid \beta|n \mid \beta\rangle\end{pmatrix}$$

If $L$ is a linear operator $L : V \to V$, with respect to the basis $|j\rangle$, it becomes a matrix with matrix elements $\langle k \mid L(|\, j)|k \mid L(|\, j)\rangle$.

One more important result, which we'll do as an exercise:

**Proposition 6.7.** If $v, w$ are eigenvalues of an operator $T$ on an inner product space with distinct eigenvalue, then $v, w$ are orthogonal.

## 6.2 Adjoint of a Linear Operator

When $V$ is a vector space with an inner product, we can define the adjoint of $L$ by:

**Definition 6.8 (Adjoint operator).** The adjoint of a linear operator $L : V \to V$ is the operator $L^\dagger$ satisfying

$$\langle Lv, w|Lv, w\rangle = \left\langle v, L^\dagger w \middle| v, L^\dagger w \right\rangle$$

for all $v, w \in V$.

It takes a little more matrix theory (you can see this in the appendix), but we have that in terms of matrices, $L^\dagger$ is the conjugate-transpose of the matrix of $L$. In the real case, this is just the transpose matrix.

We say a linear transformation is self-adjoint if $L^\dagger = L$, and skew-adjoint if $L^\dagger = -L$.

## 6.3 Orthogonal and Unitary Operators

A special class of linear transformations is those invertible transformations that preserve the inner product, meaning $\langle Lv, Lw|Lv, Lw\rangle = \langle v, w|v, w\rangle$ for all $v, w \in V$. These transform orthonormal bases to orthonormal bases, so they can be very powerful tools for changing perspective.

In terms of adjoints, this condition becomes:

$$\langle Lv, Lw|Lv, Lw\rangle = \left\langle v, L^\dagger Lw \middle| v, L^\dagger Lw \right\rangle = \langle v, w|v, w\rangle\,,$$

so we must have

$$L^\dagger L = \mathbf{1},$$

meaning:

$$L^\dagger = L^{-1}.$$

In the real case, such operators are called orthogonal operators, and we have the following set:

**Definition 6.9 (Orthogonal Group).** The orthogonal group $O(n)$ in $n$ dimensions is the group of invertible transformations preserving the inner product on a real vector space $V$.

This is isomorphic to the group of $n \times n$ real invertible matrices $L$ satisfying

$$L^{-1} = L^T$$

The subgroup of $O(n)$ of matrices with determinant 1 (equivalently, the subgroup preserving orientation of orthonormal bases) is called $SO(n)$.

Notice that the determinant of the transpose of a matrix is the same as the determinant of a matrix and $L^\dagger = L^T$ for real matrices, that:

$$1 = \det\{I\} = \det\left\{L^{-1}L\right\} = \det\left\{L^{-1}\right\}\det\{L\} = \det\left\{L^T\right\}\det\{L\} = \det\{L\}^2,$$

so $\det\{L\} = \pm 1$.

The simplest non-trivial example is $n = 2$, where elements of $SO(2)$ are given by matrices of the form:

$$\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

These give counter-clockwsise rotations in $\mathbf{R}^2$ by an angle $\theta$. The other matrices in $O(2)$ are given by

$$\begin{matrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{matrix}$$

and these describe a reflection and then a rotation.

Now, we move to the complex case.

**Definition 6.10 (Unitary Group).** The unitary group $U(n)$ in $n$ dimensions is the group of invertible transformations perserving the complex inner product on an $n$-dimensional complex vector space $V$. This is isomorphic to the group of $n \times n$ complex invertible matrices satisfying:

$$L^{-1} = \bar{L}^T = L^\dagger.$$

The subgroup of $U(n)$ of matrices with determinant 1 is called $SU(n)$.

the same calculation in the real case gives $\det\bar{\{L\}}\det\{L\} = |\det\{L\}|^2 = 1$, showing $\det\{L\}$ is a complex number of norm 1.

## 6.4 The Spectral Theorem for Self-Adjoint Operators

Why are we so interested in self-adjoint matrices? The answer, surprisingly, has to do with eigentheory.

**Theorem 6.11 (Spectral Theorem for Self-Adjoint Operators).** Let $V$ be a finite-dimensional complex vector space. If a linear operator $L$ is self-adjoint, then there exist eigenvectors $v_1, ..., v_n$ with real eigenvalues $\lambda_1, ..., \lambda_n$ of $L$ (not necessarily distinct) such that the $v_i$ form a basis of $V$.

This theorem holds for finite-dimensions, but there are analogous theorems in infinite-dimensions. This means in particular that given a self-adjoint operator $L$, we can label elements of a basis by eigenvalues, writing $|j\rangle = |\lambda_j\rangle = v_j$. Then, a general state is a linear combination of basis states:

$$|\psi\rangle = \sum_{j=1}^{n} |j\rangle \langle j \mid \psi | j \mid \psi\rangle$$

Why is this important? Essentially, it characterizes exactly what operators on complex $V$ have real eigenvalues such that their eigenvectors span the vector space. Once we introduce the axioms of quantum mechanics, it will become clear that this condition is necessary for our axioms to give physically-realizable quantities. Then, self-adjoint operators are precisely the "measurement" operators in quantum systems.

**Example 6.12.** Find a basis for $\mathbf{R}^3$ composed of eigenvectors of the operator $\begin{bmatrix} 1 & -2 & 2 \\ -2 & 1 & -2 \\ 2 & -2 & 1 \end{bmatrix}$.

We have:

$$A - \lambda I = \begin{bmatrix} 1-\lambda & -2 & 2 \\ -2 & 1-\lambda & -2 \\ 2 & -2 & 1-\lambda \end{bmatrix}$$

and taking the determinant gives:

$$(1-\lambda)^3 + 8 + 8 - 4(1-\lambda) - 4(1-\lambda) - 4(1-\lambda) = -\lambda^3 + 3\lambda^2 + 9\lambda + 5 = -(\lambda - 5)(\lambda + 1)^2.$$

Taking $\lambda = 5$, we get:

$$A - 5I = \begin{bmatrix} -4 & -2 & 2 \\ -2 & -4 & -2 \\ 2 & -2 & -4 \end{bmatrix},$$

and row-reduction gives:

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

so hence the eigenspace $E_5$ is spanned by:

$$\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

For $\lambda = -1$, we get:

$$A - (-1)I = \begin{bmatrix} 2 & -2 & 2 \\ -2 & 2 & -2 \\ 2 & -2 & 2 \end{bmatrix},$$

which row-reduces to:

$$\begin{bmatrix} 1 & -1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

hence the eigenspace $E_{-1}$ is spanned by:

$$\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

# 7 Groups, Representations, and Quantum Axioms

Now that we have linear algebra out of the way, we can state some basic definitions regarding symmetry. Then, we can finally state the fundamental axioms of quantum mechanics, and explain how our main observations of "quantization", "superposition", and "measurement" are answered by our linear algebra framework. We will then discuss vaguely how symmetry comes into play in QM.

From here on out, the vast majority of the material is from [6], and a lot of it is word-for-word or functionally from it.

## 7.1 Groups

**Definition 7.1.** A group is a set $G$ equipped with a binary operation $\cdot : G \times G \to G$ such that:

1. $\cdot$ is associative, meaning $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ for $a, b, c \in G$.
2. There exists an identity $e \in G$ such that $e \cdot g = g \cdot e = g$ for all $g \in G$.
3. For every $g \in G$, there exists $g^{-1} \in G$ such that $g \cdot g^{-1} = g^{-1} \cdot g = e$.

Often, we will omit the $\cdot$ and write $ab = a \cdot b$.

Notice the multiplication is not necessarily commutative.

If the set has a finite number of elements, we call this a finite group. These have applications in quantum mechanics and are the objects that you will see in a first course in abstract algebra. But, we will mostly bypass these in favor of objects called Lie groups. Defining these technically is very complicated. However, the elements of a Lie group make up a geometrical space of some dimension with a notion of continuity, and choosing coordinates on this space, the group operation will be some kind of differentiable map. Most of the Lie groups we will consider are matrix groups, meaning subgroups of the group $GL(n, \mathbf{R})$ or $GL(n, \mathbf{C})$ of $n \times n$ invertible matrix with real (resp. complex) entries. We will not define these in full abstraction or formally, since the technical dimensions are very complicated.

The structure-preserving maps of groups are called homomorphisms, and preserve the multiplication:

**Definition 7.2.** A map $\psi : G \to H$ is a group homomorphism if $\psi(g_1 g_2) = \psi(g_1)\psi(g_2)$ for all $g_1, g_2 \in G$.

## 7.2 Examples of Groups

**Example 7.3.** The set of integers $\mathbf{Z}$ under addition is a group.

**Example 7.4.** The set of integers modulo $n$, $\mathbf{Z}_n$, is a group under addition.

**Example 7.5.** The set of permutations of $n$ items, $S_n$, is a group under composition of permutations.

For some infinite groups:

**Example 7.6.** The set of real numbers $\mathbf{R}$ is a group under addition.

We have already seen the next few examples:

**Example 7.7.** The set of invertible linear transformations of a vector space $V$ forms a group under composition called $GL(V)$.

**Example 7.8.** The set of $n \times n$ invertible matrices over $\mathbf{R}$ (resp. over $\mathbf{C}$) forms a group under matrix multiplication called $GL(n, \mathbf{R})$ (resp $GL(n, \mathbf{C})$).

**Example 7.9.** The orthogonal group $O(n)$ in $n$ dimensions is the group of invertible transformations preserving the inner product on a real vector space $V$.

This is isomorphic to the group of $n \times n$ real invertible matrices $L$ satisfying

$$L^{-1} = L^T$$

The subgroup of $O(n)$ of matrices with determinant 1 (equivalently, the subgroup preserving orientation of orthonormal bases) is called $SO(n)$.

**Example 7.10.** The unitary group $U(n)$ in $n$ dimensions is the group of invertible transformations perserving the complex inner product on an $n$-dimensional complex vector space $V$. This is isomorphic to the group of $n \times n$ complex invertible matrices satisfying:

$$L^{-1} = \bar{L}^T = L^\dagger.$$

The subgroup of $U(n)$ of matrices with determinant 1 is called $SU(n)$.

And one just for fun (look at the MIT PRIMES 2025 Problem Set):

**Example 7.11.** The group $SL(2, \mathbf{Z})$ of $2 \times 2$ matrices with integer entries.

## 7.3 Group Actions

Groups often occur as "transformation groups", meaning groups of elements acting as transformations of some particular geometric object.

We have already defined these before:

**Definition 7.12 (Group action).** Let $G$ be a group and $X$ be a set. A *left group action* is a mapping

$$G \times X \to X(g, x) \to gx$$

which satisfies the following two properties:

1. $ex = x$, where $e$ is the identity element of $G$ and $x \in X$.
2. $h(gx) = (hg)x$ for all $g, h \in G$ and $x \in X$.

A right group action is defined similarly.

A good example to keep in mind is that of the space $M = \mathbf{R}^3$ with the standard inner product. There are two natural group actions preserving the inner product:

- An action of the group $G_1 = \mathbf{R}^3$ on $\mathbf{R}^3$ by translations.
- An action of the group $G_2 = O(3)$ of three-dimensional orthogonal transformations of $\mathbf{R}^3$, i.e rotations around the origin (perhaps combined with a reflection). In this case, order matters - for non-commutative groups like $O(3)$, one has $g_1 g_2 \neq g_2 g_1$ for some group elements $g_1, g_2$.

A fundamental principle in modern mathematics is to understand a space $M$ of points by the space $F(M)$ of functions on $M$ to a field $\mathbf{R}, \mathbf{C}$. This space is always a vector space, for finite $M$ being finite dimensional with $\dim F(m) =\mid m \mid$. For infinite $M$, this is more complicated and we may need to restrict to special functions.

Given a group action of $G$ on $M$, we can naturally extend this to an action on the function space $F(m)$ by:

$$(g \cdot f)(x) = f(g^{-1} \cdot x)$$

where $f$ is some function on $M$.

The order in which elements of the group act is important, so we need to have $g^{-1}$, not $g$, as we have:

$$\begin{aligned} g_1 \cdot (g_2 \cdot f)(x) &= (g_2 \cdot f)(g_1^{-1} \cdot x) \\ &= f(g_2^{-1} \cdot (g_1^{-1} x)) \\ &= f((g_2^{-1} g_1^{-1}) \cdot x) \\ &= f((g_1 g_2^{-1} x) \\ &= (g_1 g_2) \cdot f(x) \end{aligned}$$

This would not work if we defined $(g \cdot f)(x) = f(g \cdot x)$.

## 7.4 Representations

We abstract from the situation described prior by defining a representation of a group as an action of a group by linear transformations on a vector space.

**Definition 7.13 (Representation).** A representation $(\pi, V)$ of a group $G$ is a homomorphism

$$\pi : g \in G \to \pi(g) \in GL(V)$$

where $GL(V)$ is the group of invertible linear maps $V \to V$, with $V$ a vector space.

We will mostly be interested in the case of complex representations, so we will assume the representation is complex unless otherwise specified.

When $V$ is finite-dimensional and a basis of $V$ has been chosen, we can identify linear maps and matrices. This provides an isomorphism $GL(V) \cong GL(n, \mathbf{C})$ of the group of invertible linear maps of $V$ with $n \times n$ complex matrices.

A good example to understand all of this is the following:

- Let $M$ be a set of 3 elements $x_1, x_2, x_3$. So, $F(M) = \mathbf{C}^3$. For $f \in F(M)$, $f$ is a vector in $\mathbf{C}^3$ with components $(f(x_1), f(x_2), f(x_3))$.
- Take $G = S_3$, the group of permutations of 3 elements. This has 6 elements.
- Take $G$ to act on $M$ by permuting the three elements

$$(g, x_j) \to g \cdot x_j$$

- This group action provides a representation of $G$ on $F(M)$ by the linear maps

$$(\pi(g)f)(x_j) = f(g^{-1}x_j)$$

Standing the standard basis of $F(M) = \mathbf{C}^3$, the $j$th basis element corresponds to the function $f$ that is 1 on $x_j$ and 0 on the other two elements. With respect to this basis, the $(g)$ give six $3 \times 3$ complex matrices, which satisfy the same relations as the multiplication of the group.

The most interesting classes of complex representations are those for which the transformations $\pi(g)$ preserve the inner product. This is formalized by the following:

**Definition 7.14 (Unitary Representation).** A representation $(\pi, V)$ on a complex vector space $V$ with inner product $\langle \cdot, \cdot | \cdot, \cdot \rangle$ is a unitary representation if it perserves the inner product, meaning:

$$\langle \pi(g)v_1, \pi(g)v_2 | \pi(g)v_1, \pi(g)v_2 \rangle = \langle v_1, v_2 | v_1, v_2 \rangle$$

for $g \in G, v_1, v_2 \in V$.

Equivalently, this means the $\pi(g)$ live in the subgroup $U(n) \subset GL(n, \mathbf{C})$ of unitary $n \times n$ matrices.

## 7.5 Fundamental Axioms

Now, we can finally describe the axioms of quantum mechanics. To do this, we will start with an approach to classical mechanics called Hamiltonian mechanics.

In classical mechanics, the state of a system is given by a point in "phase space", which is some kind of set (not necessarily a vector space). This can be thought of as either the space of solutions of the equations of motion of the system, or (parametrizing solutions by initial values) the space of coordinates and momenta of all particles in the system. Observable quantities like momentum, energy, temperature are captured by operators on the phase space. There is one distinguished operator, the Hamiltonian operator, corresponding to energy of the system. This determines how states evolve in time through Hamilton's equations.

The basic structure of quantum mechanics is very different, and it starts by forcing a more rigid state space:

**Axiom (States).** The state of a quantum mechanical system is given by a nonzero vector in a complex vector space $\mathcal{H}$ equipped with a Hermitian inner product $\langle \cdot, \cdot | \cdot, \cdot \rangle$.

$\mathcal{H}$ can be finite or infinite-dimensional. In the infinite-dimensional case, we will require $\mathcal{H}$ to be an inner product space that is complete (meaning Cauchy sequences converge) with respect to the metric arising from the inner product. This condition is necessary so that a theory of integration makes sense in these spaces. We will concern ourselves with finite-dimensional spaces.

The state space for a single free particle on a line is $L^2(\mathbf{R})$, the space of square-integrable ($\int_{-\infty}^{\infty} f^2(x)\,\mathrm{d}x < \infty$) functions on $\mathbf{R}$. A single free particle in three-dimensional space will be $L^2(\mathbf{R}^3)$.

Note there are two major differences with classical phase space:

- The state space is always linear - hence, nontrivial linear combinations of (independent) states are also states.
- The state space is a *complex* vector space, not real; complex numbers are essential to quantum mechanics, whereas they merely serve as a convenient calculational tool in the real case.

**Axiom (Quantum Observables).** The observables of a quantum mechanical system are given by self-adjoint linear operators on $\mathcal{H}$ in finite dimensions.

In infinite dimensions, we use a kind of operator called an unbound self-adjoint operator. These are a class of self-adjoint linear operators with analytical convergence properties that allow us to prove an equivalent of the spectral theorem. This is highly complicated functional analysis, so again we will ignore it.

**Axiom (Dynamics).** There is a distinguished quantum observable, the Hamiltonian $H$. Time evolution of states $|\psi(t)\rangle \in \mathcal{H}$ is given by the Schrodinger equation

$$i\hbar\,\mathrm{d}t\,|\psi(t)\rangle = H\,|\psi(t)\rangle$$

The operator $H$ has eigenvalues that are bounded below.

We will discuss the physical interpretation of eigenvalues of operators in the next section, and we will see the statement that the observable is bounded below is necessary to ensure there is a stable "lowest energy" state of the system.

$\hbar$ is a dimensional constant called Planck's constant (dependnig on units). It has dimensions of energy times time, and holds experimental values of:

$$1.054571726 \times 10^{-34}\ \mathrm{J{\cdot}s} == 6.58211928 \times 10^{-16}\ \mathrm{eV{\cdot}s}$$

The most natural units for quantum mechanics are units such that $\hbar = 1$. Once calculations are done in these units, multiples of $\hbar$ can be used to translate to conventional unit systems.

It is sometimes useful to carry along factors of $\hbar$. Physically, it is often (but not always) that classical behavior arises in the limit where:

$$\frac{\text{energy scaletime scale}}{\hbar}$$

is large. This can be achieved by taking units like the meter or second, but can also be achieved by sending $\hbar \to 0$. Therefore, in quantum formulas, we can often extract the classical behavior by taking the limit as $\hbar \to 0$. This is not always accurate, and correctly finding classical behavior from quantum theory can be very confusing.

## 7.6 Measurement Axioms

The above principles characterize the structure of quantum mechanics, but not how we extract anything physically out of them. The main issue here is that macroscopic systems are on the order of $10^{23}$ (Avagadro's number) particles, and hence, to solve any measurable system would require solving something like $10^{23}$ quantum mechanical systems. Additionally, the measurement apparatus quantum system and the observed quantum system must be nontrivially entangled (we will talk about this when we talk about Composite Quantum Systems).

Instead of trying to formulate such a solution to this problem, we will instead propose two axioms of measurement that will allow us to clean statistical predictions from quantum theory:

**Axiom (Observables).** States for which a concrete value of an observable quantity can be measured (meaning we measure exactly one answer, not a range of answers) are eigenvectors of the corresponding self-adjoint operator. The value of the observable will then be the eigenvalue of the operator.

The Hamiltonian operator corresponds to energy. Then, the eigenvalues of the Hamiltonian operator correspond to measurable energies of the system, and the eigenvalues being bounded below means the energy of the system is bounded below.

Essentially, this principle identifies what states are "measurable" states of a system. Whenever we do a measurement, we must obtain an exact number. Therefore, if only eigenvectors return exact numbers, then only eigenvectors are measurable states.

Operators with important physical significance (such as energy, momentum, position, angular momentum, and charge) will turn out to correspond to actions of symmetry groups on the system.

**Axiom (The Born Rule).** Given an observable $\mathcal{O}$ with associated operator $A$ and two unit-norm states $|\psi_1\rangle$ and $|\psi_2\rangle$ that are eigenvectors of $O$ with distinct eigenvalues $\lambda_1$ and $\lambda_2$

$$O|\psi_1\rangle = \lambda_1 |\psi_1\rangle, O|\psi_2\rangle = \lambda_2 |\psi_2\rangle$$

the linear combination state:

$$c_1|\psi_1\rangle + c_2|\psi_2\rangle$$

will not have a well-defined value for the observable $\mathcal{O}$. If one attempts to measure this observable, one will get either $\lambda_1$ or $\lambda_2$ with probabilities:

$$\frac{\|c_1\|^2}{\|c_1\|^2 + \|c_2\|^2}$$

and:

$$\frac{\|c_2\|^2}{\|c_1\|^2 + \|c_2\|^2}$$

respectively. When we measure $\lambda_1$, the state of the system, originally $c_1 |\psi_1\rangle + c_2 |\psi_2\rangle$, will collapse to $|\psi_1\rangle$ (and respectively for $\lambda_2$).

More generally, given unit-norm states $|\psi_1\rangle, ..., |\psi_n\rangle$ that are eigenvalues of $A$ with eigenvalues $\lambda_1, ..., \lambda_n$, the combination state $c_1 |\psi_1\rangle + c_2 |\psi_2\rangle + .... + c_n |\psi_n\rangle$ when measuring $\mathcal{O}$ will return $\lambda_i$ with probability:

$$\frac{\|c_i\|^2}{\|c_1\|^2 + \|c_2\|^2 + ... + \|c_n\|^2}.$$

When we measure $\lambda_i$, the state of the system, originally $c_1 |\psi_1\rangle + c_2 |\psi_2\rangle + ... + c_n |\psi_n\rangle$, will collapse to $|\psi_i\rangle$.

Notice the probabilities when added together equal 1, exactly as we expect - each of these eigenvectors is a measurable state, and the measurable states comprise all power measurements.

We have called the Born rule an axiom here, but given sufficient understanding of measurements, one can derive (or at least justify) the Born rule from the prior axioms. We will not do this here.

Notice the state $c |\psi\rangle$ has the exact same eigenvectors, eigenvalues, and probabilities as the state $\psi$ for any complex number $c$. As such, we can vary the norm of $\ker \psi$ by multiplying by a complex number. It is conventional to choose a complex number $c$ such that $\|c |\psi\rangle\| = 1$, so that we are dealing with a normalized state vector. The magnitude of $c$ is fixed, but it can have any phase $e^{i\theta}$, and therefore there is some ambiguity. For the purpose of physical results this is irrelevant, but the phase turns out to have deep mathematical implications, so we cannot simply discard it.

## 7.7 Relation to Experimental Foundation of QM

Let us restate the conclusions we recovered from experimental ideas:

1. At the quantum level, observable quantities are discrete, not continuous.
2. Measurement *changes the state of an object*.
   (a) Before a measurement, the state of an object is undetermined and "spread out" like a wave. Different states exist in superposition.
   (b) After a measurement, the state is determined and is collapsed like a particle.
   (c) The outcome of the measurement is probabilistic.

Observable quantities correspond to eigenvalues and eigenvectors of an operator. Only specific numbers are allowed to be eigenvectors of an operator, and hence, only specific numbers can be observed. This answers the first conclusion.

To explain the next conclusion, let us fix an operator $A$ and observable $\mathcal{O}$. Since $A$ is self-adjoint, the eigenvectors $|j\rangle$ of $A$ form a basis for the state space $\mathcal{H}$ and have real (hence physically-meaningful) eigenvalues. Since these form a basis, any arbitrary state $|\psi\rangle$ can be realized as a superposition of measurable states $|j\rangle$. When we measure $\mathcal{O}$, by the Born rule, we measure one of the eigenvalues $\lambda_j$ with probability given by the Born rule.

It is conventional to call $|\psi\rangle$ a wavefunction, owing to the idea that $|\psi\rangle$ is the wave-like superposition of the particle-like deterministic states in the double-slit experiment.

## 7.8 Groups as "Symmetries"

Now, we *finally* arrive to the thing that I have promised - the relationship between quantum mechanics and representation theory, and the importance of symmetry in quantum mechanics. When we have a symmetry of a quantum system corresponding to a group $G$, we really have a group action of $G$ on the state space $\mathcal{H}$. Then, the state space $\mathcal{H}$ will carry a unitary representation of $G$ (up to a phase factor).

For physicists, this allows us to use representation theory to constrain the system $G$, by applying facts about the representation to constrain the state space $\mathcal{H}$.

For mathematicians, physics provides a great deal of interesting unitary representations to study.

There is even more to this relationship than follows from first glance. Consider a Lie group $G$, which we'll imagine as a matrix group now. For a representation $\pi$ (identifying the state space $V$ as finite-dimensional and isomorphic to $\mathbf{C}^n$) and group elements $g$ "close" to the identity, we can write $\pi(g) \in GL(n, \mathbf{C})$ by:

$$\pi(g) = e^A$$

where $A$ is a matrix close to the zero matrix. $g$ lives in the Lie group $G$, $\pi(g)$ lives in $GL(n, \mathbf{C})$, and $A$ will live in $\mathfrak{g}$, an object called the Lie algebra of $G$. We will study this in much more detail and work extensively with examples.

In particular, we will show that when $\pi(g)$ is unitary (preserving the inner product), then $A$ will be a skew-adjoint matrix:

$$A^\dagger = -A$$

and defining $B = iA$, the operator $B$ is self-adjoint, meaning:

$$B^\dagger = B.$$

Therefore, in the case of finite-dimensional $\mathcal{H}$, the unitary representation $\pi$ of $G$ on $\mathcal{H}$ will *give rise* to not just unitary matrices $\pi(g)$ that can be used to transform the system, but also self-adjoint operators $B$ on $\mathcal{H}$. This therefore means every symmetry of a system gives rise to some observable quantity of the system. Many physically interesting observables (such as position, momentum, energy, and spin) arise in this manner.

We will see many examples of this soon. The fundamental is that of moving in time, i.e varying $t \in \mathbf{R}$. The group is $G = \mathbf{R}$ under addition, and we obtain a unitary representation of $\mathbf{R}$ on the state space $\mathcal{H}$. The corresponding self-adjoint operator is the Hamiltonian operator $H$ (divided by $\hbar$), and the representation is given by:

$$t \in \mathbf{R} \to \pi(t) = e^{-\frac{i}{\hbar}Ht}$$

which one can check is a group homomorphism from the additive group $\mathbf{R}$ to a group of unitary operators. This unitary representation gives the dyamics of the theory, since the Schrodinger equation is exactly the statement that $-\frac{i}{\hbar}H\Delta(t)$ is the skew-adjoint operator that is exponentiated to give the unitary translation moving $\psi(t)$ ahead in time by time $\delta(t)$.

Often, these groups are "symmetry groups", meaning that occur as transformations that preserve certain quantities of the theory. However, it is not necessary that a given group action preserves quantities (other than the inner product). In applications to physics, the term *symmetry* should be reserved for groups acting on the system preserving the equations of motion (more formally, commuting with the Hamiltonian operator).

For the case of such symmetry transformations, representation theory is a very powerful tool. But, beyond this class, representations of non-symmetry group actions are still very interesting.

# 8 $U(1)$ and Conservation of Charge

We first discuss some basic results in representation theory, and use these to classify the representations of the abelian Lie Group $U(1)$. We then use this result to derive conservation of charge, providing an example of the general principle described in the last lecture.

## 8.1 Representation Theorems

To understand the representations of a group $G$, we begin by identifying the irreducible representations:

**Definition 8.1 (Irreducible Representation).** A representation $\pi$ is called irreducible if it has no subrepresentations, meaning nonzero proper subspaces $W \subset V$ such that $(\pi_{|W}, W)$ is a representation. A representation with such a subrepresentation is called irreducible.

Given two representations, their direct sum is defined as:

**Definition 8.2 (Direct Sum Representation).** Given representations $\pi_1$ and $\pi_2$ of dimensions $n_1$ and $n_2$, there is a representation of dimension $n_1 + n_2$ called the direct sum of the two representations, denoted by $\pi_1 \oplus \pi_2$. This representation is given by the homomorphism

$$(\pi_1 \oplus \pi_2) : g \in G \to \begin{pmatrix} \pi_1(g) & \mathbf{0} \\ \mathbf{0} & \pi_2(g) \end{pmatrix}$$

In other words, representation matrices for the direct sum arise as block-diagonal matrices with $\pi_1$ and $\pi_2$ giving the blocks.

For unitary representations, we have the following structure theorem:

**Theorem 8.3.** Any unitary representation $\pi$ can be written as a direct sum

$$\pi = \pi_1 \oplus \pi_2 \oplus ... \oplus \pi_m$$

where the $\pi_i$ are irreducible.

*Proof.* If $(\pi, V)$ is not irreducible, there exists a nonzero subspace $W \subset V$ such that $(\pi_{|W}, W)$ is a representation, and

$$(\pi, V) = (\pi_{|W}, W) \oplus (\pi_{|W^\perp}, W^\perp$$

where $W^\perp$ is the orthoognal complement of $W$ in $V$ (with respect to the inner product).

$(\pi_{W^\perp}, W^\perp)$ is a subrepresentation since by unitarity, the representation matrices preserve the inner product.

More explicitly, identifying $W$ allows us to decompose $V$ by $V = W \oplus W^\perp$, so every $v \in V$ can be written as $w + w^\perp$ for $w \in W, w^\perp \in W^\perp$. Since the $\pi(g)$ are unitary operators, we have $\langle \pi(g)(w), \pi(g)(w^\perp | \pi(g)(w), \pi(g)(w^\perp \rangle = \langle w, w^\perp | w, w^\perp \rangle = 0$, hence $\pi(g)(w)$ and $\pi(g)(w^\perp)$ are orthogonal.

Then, we can apply the same argument onto $W$ and $W^\perp$ until we compose $(\pi, V)$ into a direct sum of irrreducibles. □

Notice non-unitary representations are not decomposable in this way.

**Example 8.4.** Consider the group of $2 \times 2$ upper-triangular matrices, that is, matrices of the form $\left(\begin{smallmatrix} 0 & a \\ b & c \end{smallmatrix}\right)$ acting on $V = \mathbf{C}^2$. The subspace $W \subset V$ of vectors of the form $\left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$ is a subrepresentation (since upper-triangular matrices preserve vectors of this form), but there is no complement to $W$ in $V$ that is also a subrepresentation (since upper triangular matrices do not preserve $\left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right)$).

Finding such a decomposition of an arbitrary unitary representation into irreducible components can be very nontrivial. ♣ Add woit exercise on $S_3$.

Recall that we only find matrices for $\pi(g)$ when a basis for $V$ is chosen. Hence, to find a block-diagonal form, we need to look for a basis where all the $\pi(g)$ is block-diagonal, and this can be very difficult to check.

The following irreducibility criterion is one of the most basic tools in representation theory:

**Theorem 8.5 (Schur's Lemma).** If a complex representation $(\pi, V)$ is irreducible, then the only linear maps $M : V \to V$ commuting with all the $\pi(g)$ are $\lambda\mathbf{1}$, multiplication by a scalar $\lambda \in \mathbf{C}$.

*Proof.* Assume $M$ commutes with all the $\pi(g)$. We want to show that $(\pi, V)$ irreducible implies $M = \lambda\mathbf{1}$. Since we are working over $\mathbf{C}$, we can always solve the eigenvalue equation:

$$\det(M - \lambda\mathbf{1}) = 0$$

to find the eigenvalues $\lambda$ of $M$. The eigenspaces

$$E_\lambda = \{v \in V : Mv = \lambda v\}$$

will be nonzero subspaces of $V$ that are equivalently describes as the kernel of the operator $M - \lambda\mathbf{1}$. Since this operator and the $\pi(g)$ commute, we have:

$$v \in \ker(M - \lambda\mathbf{1}) \to \pi(g)v \in \ker(M - \lambda\mathbf{1}),$$

so hence $\ker(M - \lambda\mathbf{1})$ is a representation of $G$. Since $V$ is an irreducible representation, either $\ker(M - \lambda\mathbf{1}) = V$ or $\ker(M - \lambda\mathbf{1}) = 0$. But, we know it cannot be the latter (since $\lambda$ is an eigenvalue, the space is nontrivial), hence it is the entire space, meaning $M = \lambda\mathbf{1}$ on $V$. □

When we invoke matrices, this says that for an irreducible representation, if $M$ commutes with all the representation matrices $\pi(g)$, $M$ is a scalar multiple of the unit matrix. This theorem only holds for complex representations, and for real representations, it is not true.

An important corollary is the following characterization of irreducible complex representations of $G$ when $G$ is commutative.

**Theorem 8.6.** If $G$ is commutative, all of its irreducible representations are one-dimensional.

*Proof.* For $G$ commutative, $g \in G$, any representation will satisfy:

$$\pi(g)\pi(h) = \pi(h)\pi(g)$$

for all $h \in G$. If $\pi$ is irreducible, Schur's lemma tells us that the $\pi(h)$, commuting with all the $\pi(g)$, are all multiples of the unit matrix, i.e $\pi(h) = \lambda_h \mathbf{1}$ for some $\lambda_h \in \mathbf{C}$. $\pi$ is then irreducible precisely when it is one-dimensional with $\pi(h) = \lambda_h$. □

## 8.2 Representations of $U(1)$: Classification

One might imagine the simplest possible Lie group is the real numbers $\mathbf{R}$. However, the question of what happens when $x \to \infty, -\infty$ is complicated. We obtain a much simpler group by adding a periodicity condition, resulting in the circle group of points on the unit circle. Every point is characterized by an angle, and group addition is given by addition of angles.

We could identify this with rotations of $\mathbf{R}^2$, in which case we get $SO(2)$. It is more convenient to think of it as the complex unit circle, though, by identifying $\mathbf{R}^2$ with $\mathbf{C}$ geometrically.

**Definition 8.7 ($U(1)$).** The elements of $U(1)$ are points on the unit circle, i.e unit complex numbers $e^{i\theta}$ or $\theta \in \mathbf{R}$ under the equivalence relation $\theta \sim \theta + N2\pi$ for $N \in \mathbf{Z}$.

The group operation is addition of these angles $\theta$, or equivalently complex multiplication of the numbers $e^{i\theta}$.

We call this $U(1)$ since $e^{i\theta}$ can be realized as $\left[e^{i\theta}\right]$, a $1 \times 1$ unitary matrix.

♣ Insert picture of circle.

By the theorem from the prior section, we know all representations are one-dimensional, and given by a differentiable (since the group is Lie) map:

$$\pi : U(1) \to GL(1, \mathbf{C})$$

We can use the representation and differentiaiblity of the map to show the following:

**Theorem 8.8.** All irreducible representations of the group $U(1)$ are unitary, given by:

$$\pi_k : e^{i\theta} \in U(1) \to \pi_k(\theta) = e^{ik\theta} \in U(1) \subset GL(1, \mathbf{C})$$

for $k \in \mathbf{Z}$.

*Proof.* We will write the $\pi_k$ as a function of an angle $\theta \in \mathbf{R}$, satisfying the periodicity property:

$$\pi_k(2\pi) = \pi_k(0) = 1.$$

since $\pi$ is a representation, it satisfies the homomorphism property:

$$\pi_k(\theta_1 + \theta_2) = \pi_k(\theta_1)\pi_k(\theta_2)$$

We need to show that any differentiable map

$$f : U(1) \to GL(1, \mathbf{C})$$

satisfying the homomorphism and periodicity properties is of the form $f = \pi_k$. Taking the derivative, we find:

$$\begin{aligned} f'(\theta) &= \lim_{\Delta\theta \to 0} \frac{f(\theta + \Delta\theta) - f(\theta)}{\Delta\theta} \\ &= f(\theta) \lim_{\delta\Theta \to 0} \frac{f(\Delta\theta) - 1}{\Delta\theta} \\ &= f(\theta) f'(0), \end{aligned}$$

where in the second equality we use the homomorphism property to split $f(\theta + \Delta\theta) = f(\theta) f(\Delta\Theta)$.

Denoting the constant $f'(0)$ by $c$, we get the differential equation:

$$f'(\theta) = c f(\theta),$$

which has solutions with $f(0) = 1$ by:

$$f(\theta) = e^{c\theta}$$

and requiring periodicity tells us:

$$f(2\pi) e^{c2\pi} = f(0) = 1,$$

which tells us that $c = ik$ for $k \in \mathbf{Z}$ and $f = \pi_k$ for some integer $k$. □

These are all certainly unitary matrices since the $\pi_k$ live in $U(1)$, as:

$$(e^{ik\theta})^{-1} = e^{-ik\theta} = e^{\bar{i}k\theta}$$

These representations of $U(1)$ can be extended to representations of $GL(1, \mathbf{C}) \supset U(1)$ by defining:

$$\pi_k : z \in GL(1, \mathbf{C}) \to \pi_k(z) = z^k \in GL(1, \mathbf{C})$$

These are not unitary on the full space but are unitary on the subgroup $U(1)$.

## 8.3 The Charge Operator

Last time, I claimed that when we had a unitary representation of a Lie group on state space $\mathcal{H}$, we would get an associated self-adjoint operator on $\mathcal{H}$ corresponding to a physical observable. We will illustrate this process now with the case of $G = U(1)$. We will obtain the charge operator, which we will denote $Q$.

If our representation of $U(1)$ on $\mathcal{H}$ is irreducible, it is one-dimensional with $\mathcal{H} = \mathbf{C}$ of the form $(\pi_q, \mathbf{C})$ for some integer $q$, where $\pi_q$ acts by sending $z \to z^q$. We want to find

a matrix $A$ such that $\pi_q(e^{i\theta}) = e^{A\theta}$, and this turns out to be given by $A = -iq$, and then multiplying by $i$ gives the self-adjoint matrix $B = q$. So, we get a self-adjoint operator given by mutliplying by $q$.

More generally, a unitary representation of $U(1)$ will decompose as a direct sum of irreducible representations of $U(1)$, which are one-dimensional of the form $\pi_k$. Label such a representation by $\mathcal{H}_k$. Then, by the theorems from the prior section, we have:

$$\mathcal{H} = \mathcal{H}_{q_1} \oplus \mathcal{H}_{q_2} \oplus ... \oplus \mathcal{H}_{q_n}$$

for some $q_1, ..., q_n \in \mathbf{Z}$ with $n$ the dimension of $\mathcal{H}$ (the $q_j$ are not necessarily distinct). Then, the self-adjoint operator we get is given by putting each of the self-adjoint matrices from each representation together in block-diagonal form:

**Definition 8.9 (Charge Operator).** The charge operator for the $U(1)$ representation $(\pi, \mathcal{H})$ is the self-adjoint linear operator on $\mathcal{H}$ that acts by multiplication by $q_j$ on the sub-representation $\mathcal{H}_{q_j}$. Taking basis elements in ${}_{q_j}$, it acts on $\mathcal{H}$ as the matrix:

$$\begin{pmatrix} q_1 & 0 & \dots & 0 \\ 0 & q_2 & \dots & 0 \\ \dots & \vdots & & \dots \\ 0 & 0 & \dots & q_n \end{pmatrix}$$

Thinking of $\mathcal{H}$ as a quantum-mechanical state space, $Q$ is our first example of a quantum mechanical observable, a self-adjoint operator on $\mathcal{H}$. States in the subspaces $\mathcal{H}_{q_j}$ will be eigenvectors for $Q$ and have a well-defined numerical value $q_j$ of the observable. A general state will be a linear superposition of state vectors will be a superposition of states from different $\mathcal{H}_{q_j}$ and therefore will not have a well-defined value.

The representation can be recovered from the action of $Q$ on $\mathcal{H}$, with the action of the group $U(1)$ on $\mathcal{H}$ given by multiplying by $i$ and exponentiating, to get:

$$\pi(e^{i\theta}) = e^{iQ\theta} = \begin{pmatrix} e^{iq_1\theta} & 0 & \dots & 0 \\ 0 & e^{iq_2\theta} & \dots & 0 \\ \dots & \vdots & & \dots \\ 0 & 0 & \dots & e^{iq_n\theta} \end{pmatrix} \in U(n) \subset GL(n, \mathbf{C}).$$

Physicists say that $Q$ is the "generator" of the $U(1)$ action by unitary transformations on the state space $\mathcal{H}$.

## 8.4 The Lie Picture

This is our first hint at the more general and abstract perspective, which we will discuss in detail when we talk about Lie algebras.

Essentially, a map $\pi$ is a map of geometrical spaces from the Lie group $U(1)$ to the Lie group $GL(n, \mathbf{C})$, taking the identity to the identity. But, these spaces both have a notion of differentiabiltiy, and hence we can consider the differential (infinitesmial) map $\pi'$, which

is a linear map from the tangent space at the identity of $U(1)$ to the tangent space of $GL(n, \mathbf{C})$. These are respectively $M(n, \mathbf{C})$ and $M(n, \mathbf{C})$. The tangent space at the identity of the Lie algebra is the structure I have promised, the Lie algebra. Lie algebras turn out to have very powerful structure, and so we can use the structure of the morphism of Lie algebras to determine the structure of the morphism of the Lie groups.

In this case, the differentiable map is:

$$\pi' : i\theta \in i\mathbf{R} \to \pi'(i\theta) = iQ\theta.$$

The following picture illustrates the situation:

Figure 11: Visualizing the Representation $\pi : U(1) \to U(n)$ and the Differential Map ([6])

The spherical image on the right is meant to illustrate the space $U(n) \subset GL(n, \mathbf{C})$. It has a distingusihed, point, the identity. The representation $\pi$ takes $U(1)$ a circle to a circle in $U(n)$, and $\pi(n)$ maps the tangent space line $i\mathbf{R}$ to a line in the tangent space of $U(n)$.

In the case of $G = U(1)$, this seems like a lot of unneccessary babble. However, once we ramp up to more complicated examples, this picture will be an extremely useful tool.

## 8.5 Conservation of Charge

Right now, we don't understand how measurement of a quantity at any one point in time relates to the time-dynamics of the system. If we start at $t = 0$ in a state in $H_{q_j}$, there's no reason why we should remain in this state. Recall from the axioms that time-evolution is given by the Schrodinger equation:

$$\frac{\mathrm{d}}{\mathrm{d}t} |\psi(t)\rangle = -iH |\psi(t)\rangle$$

It is possible to explore how time translation relates to the Hamiltonian operator $H$, but we will not explore that in this class. For now, just note for time-independent Hamiltonians $H$, the solution is given by exponentiating $H$, with:

$$|\psi(t)\rangle = U(t) |\psi(0)\rangle ,$$

with:

$$U(t) = e^{-itH} = 1 - itH + \frac{(-it)^2}{2!}H^2 + ...$$

The commutator of two operators $O_1, O_2$ is defined by:

$$[O_1, O_2] := O_1O_2 - O_2O_1$$

and such operators are said to commute if $[O_1, O_2] = 0$. If the Hamiltonian operator $H$ and the charge operator $Q$ commute, then $Q$ will also commute with all powers of $H$:

$$[H^k, Q] = 0,$$

and therefore with an exponential of $H$, meaning:

$$[U(t), Q] = 0$$

Now, what this means is that if we have a well-defined value $q_j$ at time $t = 0$, it will have the same value at any other time $t$, since:

$$Q\left|\psi(t)\right\rangle = QU(t)\left|\psi(0)\right\rangle = U(t)Q\left|\psi(0)\right\rangle = U(t)q_j\left|\psi_0\right\rangle = q_j\left|\psi(t)\right\rangle.$$

This is a general phenomenon - when an observable commutes with the Hamiltonian, we recover a conserved quantity of the system. For example, since the Hamiltonian commutes with itself, the energy of the system is conserved.

When $[Q, H] = 0$, the group $U(1)$ is said to act as a symmetry group of the system, since the quantity generated by it is conserved. We call the $\pi(e^{i\theta})$ the symmetry transformations of the system.

## 8.6 Why "Charge" is Important

Why do we call this operator charge? The charge operator is used to understand how particles interact with electric fields once we extend quantum mechanics to quantum field theory, but this is not really why we call it charge.

Essentially, $U(1)$ symmetries imply an observable characterizing states by integer eigenvalues. As such, this quantizes this observable by forcing it to come in integer multiple multiples of discrete values.

As a concrete example, the harmonic oscillator (perhaps the most important quantum system) has a $U(1)$ symmetry by rotations in the position-momentum plane, and here the Hamiltonian operator is a multiple of the charge operator $Q$. As such, the eigenvalues of the Hamiltonian are integers times a fixed value.

As another concrete example, rotations around any fixed axis in three-dimensional space carries a $U(1)$ symmetry. The $U(1)$ symmetry corresponds to angular momentum here. As such, measurable values of angular momentum (which we will discuss at length later) come as integer multiples of a fixed value, showing explicitly a breach from classical angular momentum.

In a very real sense, this integralization of the eigenvalues is the reason for the name quantum in quantum mechanics.

# 9 $SU(2)$ and Two-State Systems

Here, we discuss a representation of the group $SU(2)$, and use this to derive certain dynamics for a two-state quantum system, more commonly known as a qubit. We again illustrate the basic group-algebra correspondence.

The simplest nontrivial quantum systems have state spaces that are two-dimensional. The prior chapter decomposed the system into trivial one-dimensional systems. Here, we will work with a two-state quantum system, and in doing so, obtain the implications of representations of a non-commutative group, amongst more.

We will first find the possible observables by finding all possible self-adjoint matrices. Exponentiating will give the group $U(2)$ of unitary $2 \times 2$ matrices and the defining representation of $U(2)$. By restricting to the subgroup $SU(2) \subset U(2)$, we will get a representation of $SU(2)$ on $\mathbf{C}^2$, called the spin $\frac{1}{2}$ representation.

The ultimate goal of the course is to find all the irreducible representations of $SU(2)$, as it turns out, $SU(2)$ is the symmetry group corresponding to spin. These are classified by $N = 0, 1, 2, 3, ...$ and have dimension $N + 1$, being referred to as the spin $\frac{N}{2}$ representation. In the limit $N \to \infty$ we can recover classical ideas of spin.

## 9.1 The Pauli Matrices

On a two-dimensional state space $\mathcal{H} = \mathbf{C}^2$, observables are self-adjoint linear operators on $\mathbf{C}^2$. with respect to a chosen basis of $\mathbf{C}^2$, these are $2 \times 2$ complex matrices $M$ satisfying the condition $M = M^\dagger$ (with $M^\dagger$ the conjugate transpose of $M$). Any such matrix will be a (real) linear combination of four matrices:

$$M = c_0\mathbf{1} + c_1\sigma_1 + c_2\sigma_2 + c_3\sigma_3,$$

with $c_j \in \mathbf{R}$ and the standard choice of basis elements given by:

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

where the $\sigma_j$ are called Pauli matrices. This choice of basis is a convention, and in particular we have chosen $\sigma_3$ to be diagonal. This distinguishes this third direction as the up-down direction in space, which is useful when we need a distinguished direction.

The basic theory of quantum mechanics the only states with well-defined values for these observables are the eigenvectors. The first matrix is trivial (being the identity), but the third gives the two eigenvectors:

$$\sigma_3 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{9.1}$$

and:

$$\sigma_3 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = - \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

with eigenvalues $+1, -1$ respectively. In quantum computing, this is the qubit system, and these are the states $|0\rangle$ and $|1\rangle$ due to the analogy with classical information. When we talk about spin, we will see $\frac{1}{2}\sigma_3$ corresponds nontrivially to the action of the group $SO(2) = U(1)$ of rotations around the third axis in space, and the eigenvalues $-\frac{1}{2}$, $\frac{1}{2}$ of this operator will be used to label the eigenstates with:

$$\left|+\frac{1}{2}\right\rangle = \begin{pmatrix}1\\0\end{pmatrix}, \left|-\frac{1}{2}\right\rangle = \begin{pmatrix}0\\1\end{pmatrix}$$

These eigenstates provide a basis for $\mathbf{C}^2$, so an arbitrary vector in $\mathcal{H}$ can be written as:

$$|\psi\rangle = \alpha\left|\frac{1}{2}\right\rangle + \beta\left|-\frac{1}{2}\right\rangle$$

for $\alpha, \beta \in \mathbf{C}$. Only if one of these is zero do we get a well-defined measurable state, and measuring gives either $\frac{1}{2}$ or $-\frac{1}{2}$.

It is easy to check that $\left|\frac{1}{2}\right\rangle$ and $\left|-\frac{1}{2}\right\rangle$ are not eigenvectors of the $\sigma_1, \sigma_2$. We can check none of the $\sigma_j$ commute, implying that there are no simultaenous eigenvectors of more than one $\sigma_j$.

This gives rise to one of the "paradoxes" of quantum mechanics - if we are in a state where we can definitely know the value of one direction of spin $\sigma_j$, we cannot know the value of the other directions of spin. We cannot know the information simultaenously. This is completely divorced from classical physics.

While the basis vectors $\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix}$ are eigenvectors of $\sigma_3$, $\sigma_1$ and $\sigma_2$ take them to nontrivial combinations of basis vectors, and there are specific combinations of $\sigma_{1,2}$ that give a natural action on the basis vectors. Since:

$$(\sigma_1 + i\sigma_2) = \begin{pmatrix}0 & 2\\0 & 0\end{pmatrix}, (\sigma_1 - i\sigma_2) = \begin{pmatrix}0 & 0\\2 & 0\end{pmatrix},$$

we have:

$$(\sigma_1 + i\sigma_2)\begin{pmatrix}0\\1\end{pmatrix} = 2\begin{pmatrix}1\\0\end{pmatrix}, (\sigma_1 + i\sigma_2)\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}$$

and:

$$(\sigma_1 - i\sigma_2)\begin{pmatrix}1\\0\end{pmatrix} = 2\begin{pmatrix}0\\1\end{pmatrix}, (\sigma_1 - i\sigma_2)\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}$$

$(\sigma_1 + i\sigma_2)$ is called a raising operator - it either increases the eigenvalue by 2 or annihilates the vector. $\sigma_1 - i\sigma_2$ is called a lowering operator - on eigenvectors of $\sigma_3$ it decreases the eigenvalue by 2 or annihilates the vector. These are not self-adjoint or observables, but are the adjoints of each other.

We will see this idea of raising and lowering operators again when we classify representations of $SU(2)$.

## 9.2 Unitary Transformations on the System

We saw last time that knowing the observable $Q$ allowed us to determine the representation of $U(1)$ by exponentiating $i\theta Q$. Here, we will do something similar, exponentiation the two-state system observables to find the representation of the symmetry group.

Taking the identify matrix first, multiplication by $i\theta$ and exponentiation gives us:

$$e^{i\theta \mathbf{1}} = \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{i\theta} \end{pmatrix}$$

This is exactly the case studied in chapter 2 for a $U(1)$ group acting on $\mathcal{H} = \mathbf{C}^2$, with:

$$Q = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

This commutes with other matrices, so we can treat them independently.

Now, looking at the Pauli matrices, notice $\sigma_j^2 = \mathbf{1}$ for all matrices. Then, their exponential looks like the following:

$$\begin{aligned} e^{i\theta\sigma_j} &= \mathbf{1} + i\theta\sigma_j + \frac{1}{2}(i\theta)^2\sigma_j^2 + \frac{1}{3!}(i\theta)^3\sigma_j^3 + ... \\ &= \mathbf{1} + i\theta\sigma_j - \frac{1}{2}\theta^2\mathbf{1} - i\frac{1}{3!}\theta^3\sigma_j + ... \\ &= \left(1 - \frac{1}{2!}\theta^2 + ...\right)\mathbf{1} + i\left(\theta - \frac{1}{3!}\theta^3 + ...)\right)\sigma_j \\ &= \cos(\theta)\mathbf{1} + i\sin(\theta)\sigma_j \end{aligned}$$

As $\theta$ moves from $0 \to 2\pi$, this exponential traces out a circle in the space $U(2)$ of unitary $2 \times 2$ matrices, starting and ending at the unit matrix. This circle is a group, isomorphic to $U(1)$. So, we have found three different $U(1)$ subgroups inside the unitary $2 \times 2$ matrices.

Only one of these, $j = 3$, acts diagonally, with the $U(1)$ determined by:

$$Q = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Diagonalizing either $j = 1$ or $j = 2$ forces the other two matrices to not be diagonal, reflecting again the fact that we cannot understand the net action in terms of the action in each axis. We need to figure out the general group we get by exponentiating general linear combinations of Pauli matrices. To compute these, we can check the following relations:

$$[\sigma_j, \sigma_k]_+ = \sigma_j\sigma_k + \sigma_k\sigma_j = 2\delta_{jk}\mathbf{1},$$

where $[\cdot,\cdot]_+$ is the anticommutator. This relation says all $\sigma_j$ satisfy $\sigma_j^2 = \mathbf{1}$ and distinct $\sigma_j$ anticommute.

Notice that these relations impliy that when we take $\mathbf{v} = (v_1, v_2, v_3) \in \mathbf{R}^3$ and define a $2 \times 2$ matrix by:

$$\mathbf{v} \cdot \boldsymbol{\text{œ}} = v_1\sigma_1 + v_2\sigma_2 + v_3\sigma_3 = \begin{pmatrix} v_3 & v_1 - iv_2 \\ v_1 + iv_2 & -v_3 \end{pmatrix},$$

taking powers of this matrix, we find:

$$(\mathbf{v} \cdot \boldsymbol{\text{œ}})^2 = (v_1^2 + v_2^2 + v_3^2)\mathbf{1} = \|\mathbf{v}\|^2\mathbf{1}.$$

When $\mathbf{v}$ is a unit vector, we therefore see that $(\mathbf{v} \cdot \boldsymbol{\text{œ}})$ is 1 when $n$ is even and $\mathbf{v} \cdot \boldsymbol{\text{œ}}$ for $n$ odd. This is the same relation we saw for Pauli matrices, and hence, we see:

$$e^{i\theta\mathbf{v}\cdot\sigma} = (\cos\theta)\mathbf{1} + i(\sin\theta)\mathbf{v} \cdot \boldsymbol{\text{œ}}$$

The inverse of this matrix is exactly found by sending $\theta \to -\theta$, and we have:

$$(e^{i\theta\mathbf{v}\cdot\boldsymbol{\text{œ}}})^{-1} = (\cos\theta)\mathbf{1} - i(\sin\theta)\mathbf{v} \cdot \boldsymbol{\text{œ}}$$

Checking unitarity, we have:

$$\begin{aligned}(e^{i\theta\mathbf{v}\cdot\sigma}) &= ((\cos(\theta))\mathbf{1} + i(\sin(\theta))\mathbf{v} \cdot \boldsymbol{\text{œ}})^\dagger \\ &= ((\cos(\theta))\mathbf{1} - i(\sin(\theta))\mathbf{v} \cdot \boldsymbol{\text{œ}}^\dagger) \\ &= ((\cos(\theta))\mathbf{1} - i(\sin(\theta))\mathbf{v} \cdot \boldsymbol{\text{œ}}) \\ &= (e^{i\theta\mathbf{v}\cdot\boldsymbol{\text{œ}}})^{-1}\end{aligned}$$

We can also compute the determinant very quickly:

$$\begin{aligned}\det\left(e^{i\theta\mathbf{v}\cdot\boldsymbol{\text{œ}}}\right) &= \det((\cos\theta)\mathbf{1} + i(\sin\theta)(v_1 - iv_2)) \\ &= \cos^2(\theta) + (\sin^2\theta)(v_1^2 + v_2^2 + v_3^2) \\ &= 1.\end{aligned}$$

So, by exponentiating $i$ times linear combinations of the self-adjoint Pauli matrices, we get unitary matrices of determinant one. These are invertible, and form the group $SU(2)$, unitary $2 \times 2$ matrices of determinant one. If we exponentiated in addition $\phi\mathbf{1}$, we would get a unitary matrix with determinant $e^{i2\phi}$, and this forms the gorup of unitary $2 \times 2$ matrices $U(2)$.

To understand the structure of $SU(2)$ more, consider an arbitrary $2 \times 2$ complex matrix:

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$$

Unitary requires the rows are orthonormal, and orthogonality gives the relation:

$$\gamma\bar{\alpha} + \delta\bar{\beta} = 0,$$

telling us that:

$$\delta = -\gamma \frac{\bar{\alpha}}{\bar{[\beta]}}.$$

The condition that the first row is length one (normality) gives:

$$\alpha\bar{\alpha} + \beta\bar{\beta} = 1,$$

and using these two relations and computing the determinant (which should be 1) gives:

$$\alpha\delta - \beta\gamma = -\frac{\alpha\bar{\alpha}}{\bar{\beta}}\gamma - \beta\gamma = -\frac{\gamma}{\beta}(\alpha\alpha + \bar{\beta}\bar{\beta}) = -\frac{\gamma}{\beta} = 1,$$

hence $= -\bar{\beta}$, and then $\delta = \bar{\alpha}$, so we get the general form of an $SU(2)$ matrix is:

$$\begin{pmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{pmatrix}$$

where $\alpha, \beta$ are complex numbers such that their norms sum to 1. Identifying $\mathbf{C}^2$ with $\mathbf{R}^4$, these are vectors of length one in $\mathbf{R}^4$. Just as we identifies $U(1)$ with the unit circle in $\mathbf{R}^2$, then, we can identify $SU(2)$ with the unit three-sphere in $\mathbf{R}^4$.

## 9.3 Commutation Relations

An important set of relations for the Pauli matrices are their commutation relations:

$$[\sigma_j, \sigma_k] = \sigma_j\sigma_k - \sigma_k\sigma_j = 2i\sum_{l=1}^{3} \epsilon_{jkl}\sigma_l,$$

where $\epsilon_{jkl}$ satisfies $\epsilon_{123} = 1$, is antisymmetric under permutation of two of its subscripts ,and vanishes if two subscripts take the same value. More explicitly, we get:

$$[\sigma_1, \sigma_2] = 2i\sigma_3, [\sigma_2, \sigma_3] = 2i\sigma_1, [\sigma_3, \sigma_1] = 2i\sigma_2$$

These relations can be checked by direct computation. Putting these together gives a formula for the product of Pauli matrices:

$$\sigma_j\sigma_k = \delta_{jk}\mathbf{1} + i\sum_{l=1}^{3} \varepsilon_{jkl}\sigma_l.$$

Physicists prefer the self-adjoint Pauli matrices with real eigenvalues. However, mathematically, using the skew-adjoint matrices:

$$X_j = -i\frac{\sigma_j}{2}$$

is also fine, and these have simpler commutation relations:

$$[X_j, X_k] = \sum_{l=3}^{3} \varepsilon_{jkl} X_k,$$

explicitly giving;

$$[X_1, x_2] = X_3, [X_2, X_3] = X_1, [X_3, X_1] = X_2.$$

The non-triviality of these commutators reflects the non-commutativity of the group $SU(2)$ we get by exponentiating linear combinations of these $X$. Group elements $U \in SU(2)$ near the identity are of the form;

$$U \approx \mathbf{1} + c_1 X_1 + c_2 X_2 + c_3 X_3,$$

for $c_i$ small and real, just as group elements $z \in U(1)$ near the identity satisfy:

$$z \approx 1 + i\varepsilon.$$

Why is this? The answer is that the $X_i$ generate the Lie algebra of $SU(2)$. Exponentiating the Pauli matrices was exactly exponentiating the Lie algebra to generate the Lie group $SU(2)$. This linear approximation of elements close to the identity is a reflection that the Lie algebra is a tangent space to the Lie group near the identity and the tangent space is an approximation near the identity.

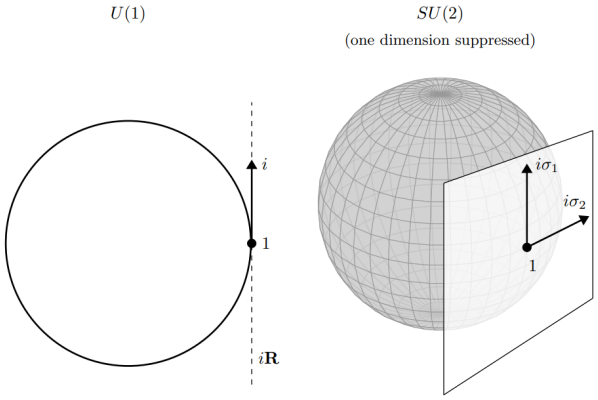


Figure 12: Comparison of geometry of $U(1)$ and $\mathfrak{u}(1)$ and $SU(2)$ and $\mathfrak{su}(2)$

## 9.4 Dynamics of the Two-State System

Recall the time-dynamics of states in quantum mechanics are given by the Schrodinger equation:

$$\frac{\mathrm{d}}{\mathrm{d}t} |\psi(t)\rangle = -iH |\psi(t)\rangle$$

with $H$ the Hamiltonian operator, a specific self-adjoint linear operator on $\mathcal{H}$. Considering the case of $H$ time-independent, the most general operator $H$ is given by:

$$H = h_0\mathbf{1} + h_1\sigma_1 + h_2\sigma_2 + h_3\sigma_3,$$

for real $h0, h_1, h_2, h_3$. The solution is given by exponentiation:

$$|\psi(t)\rangle = U(t)\,|\psi(0)\rangle$$

with:

$$U(t) = e^{-itH}$$

The $h_0\mathbf{1}$ term contributes a phase factor $e^{-ih_0t}$, with the remaining factor of $U(t)$ an element of $SU(2)$ rather than the larger group $U(2)$ of all $2 \times 2$ unitaries.

Using the equations we found before, our $U(t)$ is given by taking $\mathbf{h} = (h_1, h_2, h_3)$, $\mathbf{v} = \frac{\mathbf{h}}{\|\mathbf{h}\|}$, and $\theta = -t\|h\|$, so we get:

$$\begin{aligned}
&U(t)e^{-ih_0t}\left(\cos(-t\|\mathbf{h}\|)\mathbf{1} + i\sin(-t\|\mathbf{h}\|)\frac{h_1\sigma_1 + h_2\sigma_2 + h_3\sigma_3}{\|\mathbf{h}\|}\right)\\
&= e^{-ih_0t}\left(\cos(t\|\mathbf{h}\|)\mathbf{1} - i\sin(t\|\mathbf{h}\|)\frac{h_1\sigma_1 + h_2\sigma_2 + h_3\sigma_3}{\|\mathbf{h}\|}\right)\\
&= e^{-ih_0t}\begin{pmatrix}\cos(t\|\mathbf{h}\|) - i\frac{h_3}{\|\mathbf{h}\|}\sin(t\|\mathbf{h}\|) & -i\frac{h_1 - ih_2}{\|\mathbf{h}\|}\sin(t\|\mathbf{h}\|)\\ -i\frac{h_1 + ih_2}{\|\mathbf{h}\|}\sin(t\|\mathbf{h}\|) & \cos(t\|\mathbf{h}\|) + i\frac{h_3}{\|\mathbf{h}\|}\sin(t\|\mathbf{h}\|)\end{pmatrix}
\end{aligned}$$

In the special case $\mathbf{h} = (0, 0, h_3)$, we have:

$$U(t) = \begin{pmatrix}e^{-it(h_0+h_3)} & 0\\ 0 & e^{-it(h_0-h_3)}\end{pmatrix}$$

so if our initial state is:

$$|\psi(0)\rangle = \alpha\left|+\frac{1}{2}\right\rangle + \beta\left|-\frac{1}{2}\right\rangle$$

for $\alpha, \beta \in \mathbf{C}$, the state will evolve by:

$$|\psi(t)\rangle = \alpha e^{-it(h_0+h_3)} + \beta e^{-it(h_0-h_3)}\left|-\frac{1}{2}\right\rangle$$

In this special case, the eigenvalues of the Hamiltonian are $h_0 \pm h_3$.

We can build a physical system to test this prediction. In this case, with a spin $\frac{1}{2}$ particle (ignoring spatial motion), the Hamiltonian is given by:

$$H = \frac{ge}{4mc}(B_1\sigma_1 + B_2\sigma_2 + B_3\sigma_3)$$

where the $B_j$ are the components of the magnetic field, and the physical constants are the gyromagnetic ratio $(g)$, the electric charge $(e)$, and the mass $(m)$, and the speed of light $(c)$. By computing $U(t)$ as above, we have solved the problem of finding the time-evolution of such a system, setting $h_j = \frac{ge}{4mc} B_j$. For the special case of a magnetic field in the 3-direction, the Hamiltonian has eigenvalues $h_0 + h_3$, $h_0 - h_3$, so the difference in the measurable energies is given by:

$$2h_3 = \frac{ge}{2mc} B_3.$$

This is known as the Zeeman effect, and can be readily observed in the spectra of atoms subjected to a magnetic field.

A much more detailed discussion of the qubit can be found in Volume III of the Feynman Lectures.

# 10 Lie Algebras I

Here, we begin an exploration of Lie algebras. We state the basic definitions and then provide various examples.

## 10.1 Definition of a Lie Algebra

We have discussed numerous times the picture:

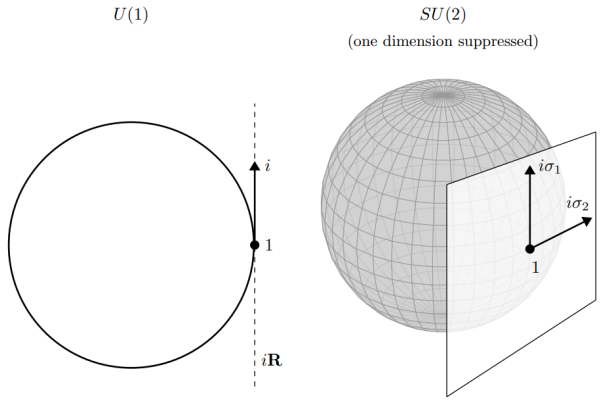


Figure 13: Tangent Space Picture

of analyzing a Lie group by considering its tangent space. In the case of the group $U(1)$, we could use the homomorphism property to determine $\pi$ in terms of its derivative $\pi'$. This is a general phenomenon. Since we are analyzing the algebraic structure of the Lie group by using the derivative to pass to the tangent space, this tangent space must too have an interesting algebraic structure.

The material in this lecture is very abstract, and so it might be difficult to understand the general principles here. But, a lot of it that should make more sense after we talk through our main examples. We will not formally prove many of the things we talk about, and will in particular not formally define the tangent space, since they have analytic complications belonging to the subject of differential geometry.

We will first define the Lie algebra of a matrix Lie group $G$, that is, a subgroup of the group $GL(n, \mathbf{C})$ of invertible matrices over $\mathbf{C}$ (or $\mathbf{R}$).

**Definition 10.1 (Lie algebra).** For $G$ a Lie group of $n \times n$ invertible matrices, the Lie algebra of $G$, written $Lie(G)$ or $\mathfrak{G}$, is the space of $n \times n$ matrices $X$ such that $e^{tX} \in G$ for $t \in \mathbf{R}$.

The exponential of a matrix here is given by the usual power series:

$$e^A = \mathbf{1} + A + \frac{1}{2}A^2 + ... + \frac{1}{n}A^n + ...$$

which can be shown to converge for any $A$.

This definition is more concrete than defining the Lie group as a tangent space, it doesn't make obvious that the Lie algebra is a real vector space. Our main interest will be using this to concretely determine the Lie algebras of certain Lie groups?

While the group $G$ determines the lie algebra $\mathfrak{g}$, the Lie algebra does not determine the group. For example, $O(n)$ and $SO(n)$ have the same tangent space at the identity, and thus the same Lie algebra, but elements at $O(n)$ not in the component of the identity can't be written in the form $e^{tX}$ (since these paths go to the identity which isn't in this connected component).

Note that for a given $X$, different values of $t$ can give the same group element, and this can happen in different ways for different groups with the same algebra. For example, consider $G = U(1)$ and $G = \mathbf{R}$ both have the same algebra $\mathbf{g} = \mathbf{R}$. In the first case, an infinite number of value gives the same group element, in the second, only one does. More subtly, we will see $SU(2)$ and $SO(3)$ are different groups with the same Lie algebra.

Now, let's look at the math more. WE have a group $G \subset GL(n, \mathbf{C})$, and $X \in M(n, \mathbf{C})$, the space of $n \times n$ complex matrices. For all $t \in \mathbf{R}$, the exponential $e^{tX}$ is an invertible matrix with inverse $e^{-tX}$, and hence lives in $GL(n, \mathbf{C})$. For each $X$, we thus have a pth of elements in $GL(n, \mathbf{C})$ going through the identity at $t = 0$, with velocity vector:

$$\frac{\mathrm{d}}{\mathrm{d}t} e^{tX} = X e^{tX}$$

which at $t = 0$ gives:

$$\frac{\mathrm{d}}{\mathrm{d}t} \left( e^{tX} \right)_{t=0} = X$$

Now we see why this is somewhat equivalent to the tangent space - taking derivatives and taking the rate of change at zero (thought of as a differential at zero) gives the matrix $X$.

To calculate this, use the power series expansion for the exponential and differentiate term-by-term.

In the case of $G = GL(n, \mathbf{C})$, we have $\mathfrak{gl}(n, \mathbf{C}) = M(n, \mathbf{C}$, which is a linear space of the right dimension to be the tangent space, so this is consistent. For subgroups of $GL(n, \mathbf{C})$ under some condition (like being unitary), we will need to identify a corresponding condition for $x \in \mathfrak{g}$, which lives in $M(n, \mathbf{C})$.

The existence of this space $\mathfrak{g} \subset M(n, \mathbf{C}$ provides a natural distingusihed representation of $\mathfrak{g}$, called the Adjoint representation:

**Definition 10.2.** The adjoint representation $(Ad, \mathfrak{g})$ is given by the homomorphism:

$$Ad : g \in G \to Ad(g) \in GL(\mathfrak{g}$$

with $Ad(g)$ acting on $X \in \mathfrak{g}$ by:

$$(Ad(g))(X) = gXg^{-1}.$$

To show this is well-defined, we need to check for $X \in \mathfrak{g}$, $gXg^{-1} \in \mathfrak{g}$. This can be shown via the identity:

$$e^{tgXg^{-1}} = ge^{tX}g^{-1},$$

(since $g, g^{-1} \in G$, we get the result via closure of multiplication).

We get this identity by expanding the exponential with a power series expansion, then using the fact:

$$(gXg^{-1})^k = (gXg^{-1})...(gXg^{-1}) = gX^kt^{-1}.$$

It is also easy to check this is a homomorphism, with:

$$Ad(g_1Ad(g_2) = Ad(g_1g_2).$$

As mentioned before, there is an interesting algebraic structure to Lie algebras.

**Definition 10.3 (Lie bracket).** The lie bracket operation on $\mathfrak{g}$ is the bilinear antisymmetric (meaning $f(x,y) = -f(y,x)$) map given by the commutator of matrices:

$$[\cdot,\cdot] : (X,Y) \in \mathfrak{g} \times \mathfrak{g} \to [X,Y] = XY - YX \in \mathfrak{g}.$$

We need to check this is well-defined, i.e given $X, Y \in \mathfrak{g}$, $Xy - YX \in \mathfrak{g}$.

**Theorem 10.4.** If $X, Y \in \mathfrak{g}$, $[X,Y] = XY - YX \in \mathfrak{g}$.

*Proof.* Since $X \in \mathfrak{g}$, we have $e^{tX} \in G$ and we can act on $Y$ by the adjoint representation:

$$Ad(e^{tX})Y = e^{tX}Ye^{-tX} \in \mathfrak{g}.$$

As $t$ varies, this gives a parametrized curve in $\mathfrak{g}$. Its velocity vector will also live in $\mathfrak{g}$, so we have:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(e^{tX}Ye^{-tX}\right) \in \mathfrak{g}.$$

By the product rule, one can show:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left(e^{tX}Ye^{-tX}\right) &= \left(\frac{\mathrm{d}}{\mathrm{d}t}\left(e^{tX}Y\right)\right)e^{-tX} + e^{tX}Y\left(\frac{\mathrm{d}}{\mathrm{d}t}e^{-tX}\right) \\ &= Xe^{tX}Ye^{-tX} - e^{tX}YXe^{-tX}\end{aligned}$$

and evaluating at $t = 0$ gives:

$$XY - YX,$$

showing $[X,Y] \in \mathfrak{g}$ as desired. □

The relation:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(e^{tX}Ye^{-tX}\right)_{|t=0} = [X,Y]$$

used here will be very important in general.

To do calculations with the Lie algebra, we can choose a basis $X_1, ..., X_n$ for the vector space $\mathfrak{g}$, and use the fact that the Lie bracket can be written in terms of this basis by:

$$[X_j, X_k] = \sum_{l=1}^{n} c_{jkl} X_l$$

where $c_{jkl}$ is a set of constants known as the structure constnats of the Lie algebra, and the $X_i$ are known as the generators of the algebra. For example, in the case of $\mathfrak{su}(2)$, as we saw last time, the Lie algebra of $SU(2)$ has a basis $X_1, X_2, X_3$ satisfying:

$$[X_j, X_k] = \sum_{l=1} 3\varepsilon_{jkl} X_l$$

with $\varepsilon_{jkl}$ the Levi-Cevita symbol (totally antisymmetric, vanishes for identical parameters).

## 10.2 Examples of Lie Algebras

Now, the last section may not have been much sense. Now, we will discuss some very concrete examples of Lie algebras.

## 10.3 Orthogonal and Unitary Groups

The groups we are most interested in are the groups of linear transformations preserving an inner product - the orthogonal and unitary groups. We have seen these are subgroups of $GL(n, \mathbf{R})$ or $GL(n, \mathbf{C})$ consisting of elements $\Omega$ satisfying the condition:

$$\Omega\Omega^* = \mathbf{1}.$$

To see what condition this becomes on the Lie algebra, write $\Omega = e^{tX}$. Then, what is $\Omega^\dagger$?

Since we know transposes reverse the order of products (meaning $(XY)^T = Y^T X^T$), and the complex conjugate of the product is the product of the complex conjugates of the matrices, we see:

$$(e^{tX})^\dagger = e^{tX^\dagger}$$

The condition:

$$QQ^\dagger = \mathbf{1}$$

then becomes:

$$e^{tX} e^{tX^\dagger} = \mathbf{1}$$

Taking the derivative gives:

$$e^{tX} X^\dagger e^{tX^\dagger} + X e^{tX} e^{tX^\dagger} = 0,$$

and evaluating at $t = 0$ gives:

$$X^\dagger + X = 0,$$

so the matrices we want will be skew-adjoint, i.e:

$$X^\dagger = -X.$$

Note that physicists will often prefer to use self-adjoint matricees by multiplying by $i$ - we will avoid this since we want to keep the real vector space structure by sticking to real numbers.

### 10.3.1 Orthogonal Group

Now, let's think geometrically with the tangent space picture for a moment. The tangent space must geometrically lie "close" to the identity, which has determinant 1. The only topologically connected part of $O(n)$ close to the identity in this sense is the subgroup $SO(n)$ (I won't explain why, since this has to do with some more complicated topology), but what this means is only elements of $SO(n) \subset O(n)$ can be exponentials of elements $X$ in the Lie algebra. These elements $\Omega$ can be written as:

$$\Omega = e^{tX},$$

and this gives a path connecting $\Omega$ to the identity. We saw that the condition $\Omega^T = \Omega^{-1}$ corresponds to skew-symmetry of $X$:

$$X^T = -X.$$

So, the Lie algebra $\mathfrak{so}(n)$ is the space of skew-symmetric $n \times n$ real matrices along with the bilinear antisymmetric product given by the commutator $[X, Y]$.

The dimension of the space of such matrices is given by:

$$1 + 2 + ... + (n-1) = \frac{n^2 - n}{2}$$

and a basis is given by the matrices $\epsilon_{jk}$ with $j, k = 1, ..., n, j < k$ defined by $e_{jk} = -1$ when $j = l, k = m$, 1 when $j = m, k = l$, and 0 otherwise.

In chapter 6 we will examine in detail the $n = 3$ case, with the Lie algebra $\mathfrak{so}(3)$ is $\mathbf{R}^3$, realized as antisymmetric $3 \times 3$ real matrices.

### 10.3.2 Unitary Group

In the case of the unitary group, we need a more general condition:

$$X^\dagger = -X$$

So, the lie algebra $\mathfrak{u}(n)$ is the space of skew-adjoint $n \times n$ complex matrices with the bilinear, antisymmetric product given by the commutator $[X, Y]$.

These matrices form a subspace of $\mathbf{C}^{n^2}$ of half the dimension, so of real dimension $n^2$.

Although we can consider $\mathfrak{u}(n)$ as a real vector space, it is not a space of real matrices - it is the space of skew-Hermitian matrices, in general complex. While these are complex, only real linear combinations of skew-Hermitian matrices are skew-Hermitian. In this space, if we look at the subspace of real matrices, we recover the sub-Lie algebra $\mathfrak{so}(n)$ of antisymmetric matrices, just as we expect (when we remove complex parts of unitary transformations, we only have the real orthogonal transformations left).

Any complex matrix $Z \in M(n, \mathbf{C})$ can be written as the sum:

$$Z = \frac{1}{2}(Z + Z^\dagger) + \frac{1}{2}(Z - Z^\dagger)$$

with the first term self-adjoint, the second skew-Hermitian. The second term can also be written as $i$ times a self-adjoint matrix:

$$\frac{1}{2}(Z - Z^\dagger) = i\left(\frac{1}{2i}(Z - Z^\dagger)\right)$$

so we see we can get all of $M(n, \mathbf{C})$ by taking complex linear combinations of self-adjoint matrices.

It is useful to use the following identity relating the determinant and the trace of a matrix:

$$\det\left(e^X\right) = e^{\mathrm{tr}(X)}$$

which can be proved by conjugating to upper-traingular form and using the fact that trace and determinant are conjugation-invariant.

Since the determinant of an $SU(n)$ matrix is 1, we see the lie algebra $\mathfrak{su}(n)$ of $SU(n)$ consists of matrices skew-Hermitian and of trace zero. The trace zero condition fixes a single entry of the matrix, and so we get a vector space of real dimension $n^2 - 1$.

One can show that $U(n)$ and $\mathfrak{u}(n)$ can be diagonalized by conjugation by a unitary matrix, and thus show any $U(n)$ matrix is an exponential of something in the Lie algebra. The corresponding theorem is true for $SO(n)$ (but requires diagonalization into $2 \times 2$ block-matrices). It is not true for $O(n)$ (we can't reach the disconnected component of the identity). It is also not true for the groups $SL(2, \mathbf{R})$ and $SL(nm\mathbf{C})$, despite the fact that these are connected.

This again illustrates the following principle: If $G$ is a Lie group with Lie algebra $\mathfrak{g}$, it is not true that every element of $G$ is an exponential of elements in $\mathfrak{g}$. However, for every algebra $\mathfrak{g}$, there exists such a group $\tilde{G}$ that is simply-connected (a topological term), called the simply-connected universal cover of $\mathfrak{g}$.

## 10.4 Classification of Simple Complex Lie Algebras

There are some other Lie algebras to mention:

**Definition 10.5.** $\mathfrak{sl}(n, \mathbf{R})$ is the set of traceless $n \times n$ real matrices.

(Notice the traceless condition is the exact condition that we need to require $\det A = 1$)

**Definition 10.6.** The symplectic group $Sp(n)$ is the subgroup:

$$Sp(n) = \{A \in GL(2n, \mathbf{R}) \mid A^T \Omega A = \Omega\}$$

with:

$$\Omega = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}$$

**Definition 10.7.** $\mathfrak{sp}(n)$ is the set of $2n \times 2n$ real matrices $X$ such that $X^T\Omega = \Omega X$.

A simple Lie algebra is an algebra with no subalgebras (hence, an irreducible algebraic object).

**Theorem 10.8 (Classification of Simple Complex Lie Algebras).** With five exceptions, every simple complex lie algebra is isomorphic to either $\mathfrak{sl}(n, \mathbf{C})$, $\mathfrak{so}(n, \mathbf{C})$, $\mathfrak{sp}(n, \mathbf{C})$ for some $n$. The exceptional algebras are called $\mathfrak{g}_2, \mathfrak{f}_4, \mathfrak{e}_6, \mathfrak{e}_7, \mathfrak{e}_8$.

The argument for this is actually combinatorial - it turns out these correspond to diagrams called Dynkin diagrams of various families. We will see parts of the ideas in this argument when we classify representations of $SU(2)$, which turns out to correspond to representations of $\mathfrak{sl}(2, \mathbf{C})$.

# 11 Lie Algebras II

Here, we describe various representations of lie algebras and basic tools for representation theory of Lie algebra. We also define abstract Lie algebras and the complexification of a Lie algebra.

We are interested in representations of Lie groups. But, we have discovered in some way we can pass between the Lie group and the Lie algebra. Now, we are interested in understanding how these representations transfer, and understanding representations of the Lie algebra.

## 11.1 Representations of Lie Algebras and Lie Groups

A group representation is a homomorphism:

$$\pi : G \to GL(n, \mathbf{C})$$

We can similarly define a Lie algebra representation as a map of Lie algebras preserving the Lie bracket:

**Definition 11.1 (Lie algebra representation).** A (complex) Lie algebra representation $(\phi, V)$ of a Lie algebra $\mathfrak{g}$ on an $n$ dimensional complex vector space $V$ is given by a real-linear map:

$$\phi : X \in \mathfrak{g} \to \phi(x) \in \mathfrak{gl}(n, \mathbf{C}) = M(n, \mathbf{C})$$

satisfying:

$$\phi([X, Y]) = [\phi(x), \phi(Y)]$$

Such a representation is called unitary if its image is in the $\mathfrak{u}(n)$, i.e, it satisfies:

$$\phi(X)^\dagger = -\phi(X).$$

More concretely, given a basis $X_1, ..., X_n$ of a Lie algebra $\mathfrak{g}$ of dimesnion $d$ with structure constnats $c_{jkl}$, a representation is given by a choice of $d$ complex matrices $\phi(X_j$ satisfying the commutation relations:

$$[\phi(X_j, \phi(X_k)] = \sum_{l=1}^{d} c_{jkl}\phi(X_l)$$

The representation is unitary when matrices are skew-adjoint.

The notion of a Lie algebra representation is motivated by the fact that the homomorphism property causes $\pi$ to be mostly determined by behavior near the identity, and thus by $\pi'$. One way to define the derivative of this map is in terms of velocity vectors of paths, and this definition i what associated to a representation $\pi : G \to GL(n, \mathbf{C})$ a linear map: where:

$$\pi'(X) = \frac{\mathrm{d}}{\mathrm{d}t}\Big(\pi(e^{tX})\Big)_{|t=0}.$$

In the case of $U(1)$, we classified all irreducible representations in this way. For general Lie groups, we can do something similar with the following theorem:

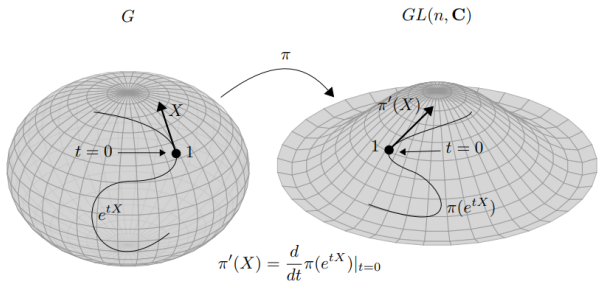


Figure 14: Derivative of a representation $\pi : G \to GL(n, \mathbf{C})$, illustrated in terms of velocity vectors along paths.

**Theorem 11.2.** If $\pi : G \to GL(n, \mathbf{C})$ is a group homomorphism, then:

$$\pi : X' \in \mathfrak{g} \to \pi'(X) = \frac{\mathrm{d}}{\mathrm{d}t}\Big(\pi(e^{tX})\Big)_{|t=0} \in \mathfrak{gl}(n, \mathbf{C}) = M(n, \mathbf{C})$$

satisfies:

1. $\pi(e^{tX}) = e^{t\pi'(X)}$
2. For $g \in G$
$$\pi'(gXg^{-1}) = \pi(g)\pi'(X)(\pi(g))^{-1}$$
3. $\pi'$ is a Lie algebra homomorphism:
$$\pi'([X, Y]) = [\pi'(X), \pi'(Y)]$$

This theorem (specifically part 3) essentially gives a way to pass a Lie group representation to a Lie algebra representation, allowing us to preserve the interesting representation information while using tools of the Lie algebra to study the representations.

*Proof.* 1. We have:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\pi(e^{tX}) &= \frac{\mathrm{d}}{\mathrm{d}s}\pi(e^{(t+s)X})_{|t=0} \\
&= \frac{\mathrm{d}}{\mathrm{d}s}\pi(e^{tX}e^{sX})_{|s=0} \\
&= \pi(e^{tX})\frac{\mathrm{d}}{\mathrm{d}s}\pi(e^{sX}_{|s=0} \\
&= \pi(e^{tX})\pi'(X)
\end{aligned}$$

We have:

$$\begin{aligned}
e^{t\pi'(gXg^{-1})} &= \pi(e^{tgXg^{-1}}) \\
&= \pi(g)\pi(e^{tX})\pi(g)^{-1} \\
&= \pi(g)e^{t\pi'(X)}\pi(g)^{-1}
\end{aligned}$$

Differentiating with respect to $t$ and $t = 0$ gives:

$$\pi'(gXg^{-1}) = \pi(g)\pi'(X)(\pi(g))^{-1}$$

Recall 5.1

$$[X, Y] = \frac{\mathrm{d}}{\mathrm{d}t}\left(e^{tX}Ye^{-tX}\right)_{|t=0}$$

so:

$$\begin{aligned}
\pi'([X, Y]) &= \pi'\left(\frac{\mathrm{d}}{\mathrm{d}t}\left(e^{tX}Ye^{-tX}\right)_{|t=0}\right) \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\pi'(e^{tX}Ye^{-tX})_{|t=0} \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\left(\pi(e^{tX})\pi'(Y)\pi(e^{-tX})\right) \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\left(e^{t\pi'(X)}\pi'(Y)e^{-t\pi'(X)}\right)_{|t=0} \\
&= [\pi'(X), \pi'(Y)].
\end{aligned}$$

□

Now, why do we want to pass to the Lie algebra? Since the Lie algebra is a vector space, the map $\pi'$ is a linear map, and so we have a lot more structure to study it. Furthermore, it is determined by its action on the basis elements, as opposed to needing to figure out the action everywhere. The $\pi'(X_j)$ will satisfy the same brackets as the $X_j$.

We will use this when we classify representations of $SU(2)$, as the lie algebra $\mathfrak{su}(2)$ is much easier to study. Notice that representations of the algebra do not ALWAYS correspond to representations of the group - those that do are called integrable (since we are "integrating" from the tangent space to the whole space). For example, looking at representations of the Lie algebra of $U(1)$ gave unitary reps for any $k$, but only integer $k$ gave representations of the group.

## 11.2 The A/adjoint Representations

For any Lie group $G$, there is a distinguished representation of $G$ on the lie algebra $\mathfrak{G}$, the Adjoint representation $(Ad, \mathfrak{g})$. The corresponding Lie algebra representatio nis called the adjoint representation, but written as $(Ad', \mathfrak{g}) = (ad, \mathfrak{g})$. From the fact that:

$$Ad(e^{tX})(Y) = e^{tX}Ye^{-tX},$$

we can differentiate and use our well-known identity to get the Lie algebra representation:

$$ad(X)(Y) = \frac{\mathrm{d}}{\mathrm{d}t}\left(e^{tX}Ye^{-tX}\right)_{t=0} = [X, Y]$$

This gives the following definition:

**Definition 11.3 (Adjoint Lie algebra representation).** $(ad, \mathfrak{g})$ is the Lie algebra representation given by:

$$X \in \mathfrak{g} \to ad(X)$$

where $ad(X)$ is defined as the linear map from $\mathfrak{g}$ to itself by:

$$Y \to [X, Y]$$

Note that this linear map $ad(X)$, which is equivalently $[X, \cdot]$ can be thought of as the infinitesmial version of the conjugation action:

$$(\cdot) \to e^{tX}(\cdot)e^{-tX}$$

## 11.3 Abstract Lie Algebra

The Lie algebra homomorphism property of *ad* says that:

$$ad([X, Y]) = ad(X) \circ ad(Y) - ad(Y) \circ ad(X),$$

where $\circ$ is composition of linear maps on $\mathfrak{g}$, so acting on $Z \in \mathfrak{g}$ we have:

$$([X, Y])(Z) = ((X) \circ ad(Y))(Z) - (ad(Y) \circ ad(X)(Z)$$

Using our expression for *ad* as a commutator, we find:

$$[[X, Y], Z] = [X, [Y, Z]] - [Y, [X, Z]]$$

This is called the Jacobi identity. We could have checked this as a blind statement about matrix multiplication, but really, what the Jacobi identity reflects is the existence of the adjoint representation. We can rewrite it in other forms using antisymmetry of the commutator, with one example:

$$[[X, Y], Z] + [[Z, X], Y] + [[Y, Z], X] = 0$$

However, it turns out as the adjoint representation implies the Jacobi identity, the Jacobi identity generates the adjoint representation, and really, the rest of the abstract Lie structure:

**Definition 11.4 (Abstract Lie Algebra).** An abstract Lie algebra over a field $\mathbf{k}$ is a vector space $A$ over $\mathbf{k}$ with a bilinear operation:

$$[\cdot, \cdot] : (X, Y) \in A \times A \to [X, Y] \in A$$

satisfying:

1. Antisymmetry:

$$[X, Y] = -[Y, X]$$

2. Jacobi identity:

$$[[X, Y], Z] + [[Z, X], Y] + [[Y, Z], X] = 0$$

These lie algebras do not need to be defined as matrices, and the operation is not in general a matrix commutator. We will not deal with these in this course, but one such group important in physics - the metapletic Lie group - has an associated interesting Lie algebra.

## 11.4 Complexification of a Lie Algebra

The Lie algebra is a real vector space. However, it is interesting in physics to take complex linear combinations of elements in the algebra (as we will see later). To do this, we will want to understand the complexification of a vector space, and extend this to the Lie algebra, to move from a real Lie algebra to a complex Lie algebra.

The way we have defined $\mathfrak{g}$, it is a real vector space, even for $G$ a group of complex matrices. For example, for $G = U(1)$ and $G = SU(2)$, $\mathfrak{u}(1) = \mathbf{1}$ and $\mathfrak{su}(2) = \mathbf{R}^3$.

While the tangent space to $GL(n, \mathbf{C})$ of all invertible complex matrices is a complex vector space $M(n, \mathbf{C})$, imposing a condition like unitarity chooses a subspace that is typically real. As such, the Adjoint representation $(Ad, \mathfrak{g})$ is in general not a complex, but rather a real representation, with:

$$Ad(g) \in GL(\mathfrak{g}) = GL(\dim \mathfrak{g}, \mathbf{R})$$

The derivative of this is the Lie algebra adjoint representation:

$$ad : X \in \mathfrak{g} \to ad(X) \in \mathfrak{gl}(\dim \mathfrak{g}, \mathbf{R})$$

and once we pick a basis of $\mathfrak{g}$, we can identify $\mathfrak{gl}(\dim \mathfrak{g}, \mathbf{R}) = M(\dim \mathfrak{g}, \mathbf{R})$. So, for each $X \in \mathfrak{g}$, we get a real linear operator on a real vector space.

But, we don't want to deal with real representations, but rather complex representations (for example, Schur's representation applies only for complex representations, amongst other things), and representation operators can be diagonalized (since the underlying field is algebraically closed). To get from a real representation to a complex representation, we will need to extend the axion of real scalars to complex scalars. For real matrices, complexification is nothing more but allowing complex entries and using the same rules for multiplying matrices as before. In general, we have:

**Definition 11.5 (Complexification).** The complexification $V_{\mathbf{C}}$ of a real vector space $V$ is the set of pairs of elements $(v_1, v_2)$ of elements of $V$ with multiplication by $a + bi \in \mathbf{C}$ given by:

$$(a + ib)(v_1, v_2) = (av_1 - bv_2, av_2 + bv_1)$$

As a real vector space, this is equivalently the sapce $\mathbf{C} \otimes V$. We interpret this as a complex vector space by treating the complex vector part $\mathbf{C}$ as a coefficient.

One should think of the complexification of $V$ as:

$$V_{\mathbf{C}} = V + iV.$$

This makes sense since:

$$\mathbf{C} = \mathbf{R} + i\mathbf{R},$$

and then:

$$C \otimes V = (\mathbf{R} + i\mathbf{R}) \otimes V = \mathbf{R} + i\mathbf{R} \otimes \mathbf{V},$$

but $R \otimes \mathbf{V} \cong \mathbf{V}$ since $V$ is a real vector space.

Given a real Lie algebra $\mathfrak{g}$, the complexification $\mathfrak{g}_\mathbf{C}$ is pairs of elements $(X, Y)$ of $\mathfrak{g}$ with the above rule for multiplication by complex scalars, which can be thought of as:

$$\mathfrak{g}_\mathbf{C} = \mathfrak{g} + i\mathfrak{g}$$

The Lie bracket on $\mathfrak{g}$ extends to a Lie bracket on $\mathfrak{g}_\mathbf{C}$ by the rule:

$$[(X_1, Y_1], (X_2, Y_2)] = ([X_1, X_2] - [Y_1, Y_2], [X_1, Y_2] + [Y_1, X_2])$$

which can be understood by the calculation:

$$[X_1 + iY_1, X_2 + iY_2] = [X_1, X_2] - [Y_1, Y_2] + i([X_1, Y_2] + [Y_1, X_2])$$

With this Lie bracket, $\mathfrak{g}_\mathbf{C}$ is a Lie algebra over the complex numbers.

For many of the cases we are interested in, this is not really necessary. When $V$ is a subspace of a complex vector space, with the property that $V \cap iV = 0$ (hence it is either entirely real or purely complex), $V_\mathbf{C}$ is just the space given by taking complex linear combinations of elements of $V$.

For example, $\mathfrak{gl}(n, \mathbf{R})$, the Lie algebra of $n \times n$ real matrices, is a subspace of $\mathfrak{gl}(n, \mathbf{C})$, the complex numbers, and we can see:

$$\mathfrak{gl}(n, \mathbf{R})_\mathbf{C} = \mathfrak{gl}(n, \mathbf{C})$$

Recalling our discussion of the unitary algebra, we see:

$$\mathfrak{u}(n)_\mathbf{C} = \mathfrak{gl}(n, \mathbf{C})$$

This example shows that two different real algebras, $\mathfrak{u}(n)$ and $\mathfrak{gl}(n, \mathbf{R})$ may have the same complexification. For another example, $\mathfrak{so}(n)$ is the Lie algebra of antisymmetric matrices, but $\mathfrak{so}(n)_\mathbf{C}$ is the Lie algebra is complex antisymmetric matrices.

For an example where the general definition is needed and the situation becomes confusing, consider the case of $\mathfrak{gl}(n, \mathbf{C})$, thinking of it as a Lie algebra and real vector space. The complexification has twice the real dimesnion, so the space:

$$\mathfrak{gl}(n, \mathbf{C})_\mathbf{C} = \mathfrak{gl}(n, \mathbf{C}) + i\mathfrak{gl}(n, \mathbf{C})_\mathbf{C},$$

is really not the same as allowing complex coefficients, but something even larger.

Given a representation $\pi'$ of a real Lie algebra, $\mathfrak{g}$, it can be extended to a representation of $\mathfrak{g}_\mathbf{C}$ by complex linearity, defining:

$$\pi'(X + iY) = \pi'(X) + i\pi'(Y).$$

If the original representation was on a complex vector space $V$, the extended one will act on the same space - if it was on a real sapce, this extended one acts on the complexification $V_\mathbf{C}$.

Some examples:

- The adjoint representation:

$$ad : \mathfrak{g} \to \mathfrak{gl}(\dim \mathfrak{g}, \mathbf{R}) = M(\dim \mathfrak{g}, \mathbf{R})$$

- Complex $n$ dimensional representations:

$$\pi' : \mathfrak{su}(2) \to M(n, \mathbf{C})$$

of $\mathfrak{su}(2)$ extend to representations:

$$\pi' : \mathfrak{su}(2)_{\mathbf{C}} = \mathfrak{sl}(2, \mathbf{C}) \to M(n, \mathbf{C}$$

Other examples include:

- Complex representations of the real Lie algebra called the Heisenberg Lie algebra play a central role in QM and QFT, as we will construct representations using annihilation and creation operators by extending the representation of the Heisenberg algebra to its complexification.
- QFTs based on complex field theories start with a Heisenberg Lie algebra that is already complex, so to get annihlation and creation operators, and we need the general theory of complexification here.

It's ok if you don't quite graph everything going on here - this is graduate-level mathematics, and has a lot of really advanced ideas spanning algebra, analysis, and geometry.

# 12 Rotation and Spin I

The group $SO(3)$ is the group of three-dimensional rotations about a point. This is one of the natural symmetry groups (as in $3D$ space, the laws of physics should be the same regardless of rotated perspective).

The observables we obtain from this group are the components of angular momentum, and understanding the state space of the quantum system in terms of representations of this group (and hence in terms of angular momentum) is hugely important for quantum mechanics.

Remarkably, it turns out quantum systems are not representations of $SO(3)$, but more naturally arise as representations of the group $Spin(3)$, which exists as a "double" cover of $SO(3)$, which we need to understand the subtle geometry of the space. In the $N = 3$ case, we have $Spin(3) \cong SU(2)$. We will determine this relationship via introducing quaternions.

## 12.1 $SO(3)$ and Rotations in Three Dimensions

Rotations in $\mathbf{R}^2$ are given by elements of $SO(2)$, with counterclockwise rotation by $\theta$ given by:

$$R(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

This can be written as an exponential $R(\theta) = e^{\theta L} = \cos(\theta)\mathbf{1} + L\sin(\theta)$ for:

$$L = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

Here, $SO(2)$ is a commutative Lie group with Lie algebra $\mathfrak{so}(2) = \mathbf{R}$. Note that we have a representation on $V = \mathbf{R}^2$ here, but it is real, not a complex representation (which is what we would have on quantum-mechanical state space).

In three-dimensions, $SO(3)$ is non-commutative and "generated" by rotations around each each axis. Choosing a unit vector $\mathbf{w}$ and angle $\theta$, an element of $SO(3)$ is in general of the form $R(\theta, \mathbf{W})$. Using standard basis vectors $\mathbf{e}_j$, rotations about coordinate axes are given by:

$$R(\theta, \mathbf{e}_1) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{pmatrix},$$

$$R(\theta, \mathbf{e}_2) = \begin{pmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{pmatrix},$$

$$R(\theta, \mathbf{e}_3) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

In general, we can write an element of $SO(3)$ in terms of three euler-angles $\phi, \theta, \psi$ with a general rotation given by:

$$R(\phi, \theta, \psi) = R(\psi, \mathbf{e}_3)R(\theta, \mathbf{e}_1)R(\phi, \mathbf{e}_3)$$

which means we rotate first about the $z$ axis by $\phi$, then rotate by $\theta$ around the new $x$-axis, then rotate by $\psi$ about the new $z$-axis.

Multiplying this out is actually quite complicated, and it's also irritating to find out the ranges for the angles.

The infinitesmial picture near the identity, given by the Lie algebra $\mathfrak{so}(3)$. Recall the Lie algebra is antisymmetric matrices, so we have:

$$l_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, l_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, l_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

which satisfy the commutation relations:

$$[l_1, l_2] = l_3, [l_2, l_3] = l_1, [l_3, l_1] = l_2$$

Note that these are exactly the same commutation relations of the basis $X_1, X_2, X_3$ of the Lie algebra $\mathfrak{su}(2)$. Therefore, we can conclude $\mathfrak{so}(3)$ and $\mathfrak{so}(2)$ are isomorphic Lie algebras.

These are both the vector space $\mathbf{R}^3$ equipped with the same Lie bracket on pairs of vectors. However, this is the same thing as something else we know:

$$\mathbf{e_1} \times \mathbf{e}_2 = \mathbf{e}_3, \mathbf{e}_2 \times \mathbf{e}_3 = \mathbf{e}_1, \mathbf{e}_3 \times \mathbf{e}_1 = \mathbf{e}_2$$

So, therefore, the Lie bracket operation $[X, Y]$ in the context of $\mathbf{R}^3$ being the Lie algebra $\mathfrak{so}(3)$ is the cross product on vectors in $\mathbf{R}^3$. So, we have three isomorphic ways of obtaining the same Lie algebra:

1. Identify $\mathbf{R}^3$ with antisymmetric real $3 \times 3$ real matrices and use the matrix commutator.

2. Identify $\mathbf{R}^3$ with skew-adjoint, traceless, complex $2 \times 2$ matrices and use the matrix commutator:

3. Take $\mathbf{R}^3$ as is and use the cross-product on vectors in $\mathbf{R}^3$ to get a Lie bracket.

It is special in $n = 3$ that the vector representation (the defining representation of $SO(n)$ matrices) is isomorphic to the adjoint representation. Recall any Lie group $G$ has a representation $(Ad, \mathfrak{g})$ on its Lie algebra $\mathfrak{g}$. For $\mathfrak{so}(n)$, its algebra is of real dimension $\frac{n^2-n}{2}$. Only for $n = 3$ does this equal $n$ itself. Geometrically, this corresponds to the fact that only in three dimensions, a plane (which rotations exist in) is determined uniquely by a vector. Equivalently, only in 3 dimesnions, there exists a cross product $\mathbf{v} \times \mathbf{w}$ which takes two vectors determining a plane to a unique perpendicular vector.

The isomorphism between the vector representation $(\pi_{vector}, \mathbf{R}^3)$ on column vectors and the adjoint representation $(Ad, \mathfrak{so}(3))$ on antisymmetric matrices is given by:

$$\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \longleftrightarrow v_1 l_1 + v_2 l_2 + v_3 l_3 = \begin{pmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{pmatrix}$$

equivalently exchanging the bases by:

$$\mathbf{e}_j \longleftrightarrow l_j$$

For the vector representation on column vectors, $\pi_{vector}(g) = g$ and $\pi'_{vector}(X) = X$, with $X$ an antisymmetrix $e \times 3$ real matrix, and $g = e^x$ an orthogonal $3 \times 3$ matrix. Both act by the usual multiplication.

The adjoint representation acts by conjugation:

$$Ad(g) \begin{pmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{pmatrix} = g \begin{pmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{pmatrix} g^{-1}$$

and the corresponding Lie algebra representation si given by:

$$ad(X) \begin{pmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{pmatrix} = \left[ X, \begin{pmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{pmatrix} \right]$$

One can explicitly check that these representations are isomorphic, for instance by checking the isomorphisms on the basis.

## 12.2 Quaternions

The quaternions are a four-dimensional "generalization" of the complex numbers, discovered by Hamilton. However, they do not have a commutative multiplication. Famously, Hamilton discovered them while walking by a bridge, and when realizing the right multiplication for quaternions, carved it into the bridge.

It is a well-known result that "generalizations" of the complex numbers (known as real division algebras) exist only in dimensions $1, 2, 4, 8$, with dimension 8 being known as the octonions, and losing not just commutativity of multiplication, but associativity as well.

The quaternions are defined as elements $q \in \mathbf{H}$ satisfying:

$$q = q_0 + q_1 \mathbf{i} + q_2 \mathbf{j} + q_3 \mathbf{k}$$

for $q_j \in \mathbf{R}$ and $\mathbf{i}, \mathbf{j}, \mathbf{k} \in \mathbf{H}$ satisfying:

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1, \mathbf{ij} = -\mathbf{ji} = k, \mathbf{ki} = -\mathbf{ik} = \mathbf{j}, \mathbf{jk} = -\mathbf{kj} = \mathbf{i},$$

and a conjugaton operation taking:

$$q \to \bar{q} = q_0 - q_1\mathbf{i} - q_2\mathbf{j} - q_3\mathbf{k}.$$

This operation satisfies (for $u, v \in \mathbf{H}$):

$$\bar{uv} = \bar{v}\bar{u}$$

As a vector space, $\mathbf{H}$ is isomorphic to $\mathbf{R}^4$, and has a natural norm function given by:

$$\|q\|^2 = q\bar{q} = q_0^2 + q_1^2 + q_2^2 + q_3^2$$

and this is multiplicative since:

$$\|uv\|^2 = uv(\bar{uv}) = uv\bar{v}\bar{u} = \|u\|^2\|v\|^2$$

Using:

$$\frac{q\bar{q}}{\|q\|^2} = 1$$

one has the formula for the inverse:

$$q^{-1} = \frac{\bar{q}}{\|q\|^2}$$

The length one quaternions form a group under multiplication (since two norm 1 elements multiply to norm 1), caled $Sp(1)$. There are also Lie groups called $Sp(n)$ for larger values of $n$, consisting of matrices with quaternionic vectors that act on quaternionic vectors preserving the norm, but these play little role in quantum mechanics. $Sp(1)$ can be identified with the three-dimensional sphere since the length one condition is:

$$q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1,$$

which is the equation of the sphere $S^3 \subset \mathbf{R}^4$.

## 12.3 The Spin Group in Four Dimensions

Pairs $(u, v)$ of unit quaternions give the product group $Sp(1) \times Sp(1)$. An element $(u, v)$ acts in $q \in \mathbf{H} = \mathbf{R}^4$ by left and right(inverse) multiplication:

$$q \to uqv^{-1}$$

This action preserves lengths of vectors while being linear in $q$ - hence, it is a linear norm-preserving maps and lives in $SO(4)$. One can see that pairs $(u, v)$ and $(-u, -v)$ give the same linear transformation of $\mathbf{R}^4$, so give the same element of $SO(4)$. As such, we can show that $SO(4)$ is actually the group $Sp(1) \times SP(1)$, with the two elements $(u, v)$ and

$(-u, -v)$ identified. We call the Lie group $Sp(1) \times SP(1)$ $Spin(4)$, and it is a "double cover" of $SO(4)$ in this manner, with the covering map:

$$\Phi : (u, v) \in SP(1) \times Sp(1) = SPin(4) \to \{q \to uqv^{-1}\} \in SO(4).$$

We will talk about the details of this "double cover" more formally later. Topologically, however, the reason this exists is that there exists a non-contractible loop in the geometry of $SO(n)$, but it is contractible when you go around twice. $Spin(n)$ arises naturally as a simply-connected double cover of $SO(n)$, choosing the covering map (a topological concept) to be a group homomorphism.

# 13 Rotation and Spin II

Here, we generalize the double-cover construction from the last lecture to rotations in four-dimensions. We then observe how these can be realized as isomorphisms of symmetry groups and algebras.

## 13.1 The Spin Group in Three Dimensions

Consider the subgroup of $Spin(4)$ acting by pairs $(u, u)$ of unit 1 quaternions, acting by conjugation:

$$q \to uqu^{-1}$$

an action which is trivial on real quaternions (since they commute with everything). It however will preserve and act nontrivially on the space of "pure imaginary" quaternions of the form:

$$q = \vec{v} = v_1\mathbf{1} + v_2\mathbf{j} + v_3\mathbf{k}$$

which can be identified with the vector space $\mathbf{R}^3$. An element $u \in Sp(1)$ acts on $\vec{v}$ by:

$$\vec{v} = u\vec{v}u^{-1}$$

This is linear, preserving $\|v\|$, so it corresponds to a element of $SO(3)$. We thus have a map (clearly a homomorphism):

$$\Phi : u \in Sp(1) \to \{\vec{v} \to u\vec{v}u^{-1}\} \in SO(3)$$

Both $u$ and $-u$ act in the same way on $\vec{v}$, so we have two elements in $Sp(1)$ correspodning to the same element in $SO(3)$. We can check this is a surjective map, so it is a "two-fold" cover. But, the space of length-1 quaternions is exactly $S^3$, so if we are identifying $u$ and $-u$ to obtain $SO(3)$, the space $SO(3)$ is geometrically the three-sphere with opposite points identified. This space is known as $\mathbf{RP}^3$, real projective three-space, which can also be thought of as the space of lines through the origin in $\mathbf{R}^4$ (since these intersect $S^3$ in opposite points).

The covering map $\Phi$ being one-directional reflects the fact that topologically, $S^3$ is not simply $\mathbf{RP}^3 \times \{1, -1\}$. This tells us that there is no inverse map going from $SO(3) \to Spin(3)$.

## 13.2 Isomorphisms of Lie Groups and Algebras

Identifying $\mathbf{C}$ with $\mathbf{R}^2$ allowed us to represent elements of the unit circle group $U(1)$ as exponentials $e^{i\theta}$, with $i\theta$ living in the Lie algebra $\mathfrak{u}(1) = i\mathbf{R}$ of $U(1)$. $Sp(1)$ behaves in the same way, with $\mathfrak{sp}(1)$ the space of pure imaginary quaternions, which can be identified with $\mathbf{R}^3$ as before.

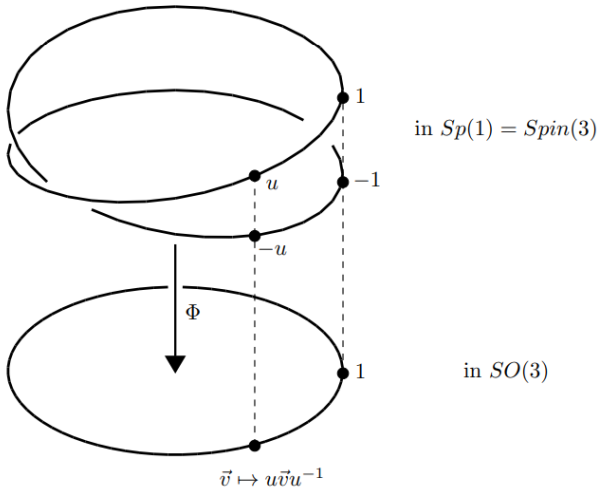


Figure 15: Double Cover $Sp(1) \to SO(3)$.

Unlike the $U(1)$ case, we have a nontrivial Lie bracket, the quaternion commutator. Elements of the group $Sp(1)$ are given by exponentiating such Lie algebra elements, which we will write in the form:

$$u(\theta, \mathbf{w}) = e^{\theta \vec{w}}$$

with $\theta \in \mathbf{R}$ and $\vec{w}$ an purely imaginary quaternion of length 1. We have:

$$\vec{w}^2 = (w_1\mathbf{i} + w_2\mathbf{j} + w_3\mathbf{k})^2 = -(w_1^2 + w_2^2 + w_3^2) = -1$$

the exponential can be expanded to get:

$$e^{\theta \vec{w}} = \cos(\theta) + \vec{w}\sin(\theta)$$

like before for $SU(2)$. Taking $\theta$ as a parameter, the $u(\theta, \mathbf{w})$ give paths in $Sp(1)$ going through the identity at $\theta = 0$, with velocity vector $\vec{w}$ since:

$$\frac{\mathrm{d}}{\mathrm{d}\theta} u(\theta, \mathbf{w})_{|\theta=0}(-\sin(\theta) + \vec{w}\cos(\theta))_{|\theta=0} = \vec{w}$$

We can explicitly evaluate the homomorphism $\Phi$ on such elements $u(\theta, \mathbf{w}) \in Sp(1)$, with the result that $\Phi$ takes $u(\theta, \mathbf{w})$ to a rotation by an angle $2\theta$ around the axis $\mathbf{w}$:

**Theorem 13.1.** $\Phi(u(\theta, \mathbf{w})) = R(2\theta, \mathbf{w})$.

*Proof.* First consider the case $\mathbf{w} = \mathbf{e}_3$ of rotations around the $z$-axis:

$$ui(\theta, \mathbf{e}_3) = e^{\theta \mathbf{k}} = \cos(\theta) + \mathbf{k}\sin(\theta)$$

and:

$$u(\theta, \mathbf{e}_3)^{-1} = e^{-\theta \mathbf{k}} = \cos(\theta) - \mathbf{k}\sin(\theta)$$

so $\Phi(u(\theta, \mathbf{e}_3))$ is the rotation taking $\mathbf{v}$ (identified with the quaternion $\vec{v} = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}$) to:

$$\begin{aligned} u(\theta, \mathbf{e}_3)\vec{v}u(\theta, \mathbf{e}_3)^{-1} &= (\cos(\theta) + \mathbf{k}\sin(\theta))(v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k})(\cos(\theta) - \mathbf{k}\sin(\theta)) \\ &= (v_1(\cos^2(\theta) - \sin^2(\theta)) - v_2(2\sin(\theta)\cos()))\mathbf{i} \\ &+ (2v_1\sin(\theta)\cos(\theta) + v_2(\cos^2(\theta) - \sin^2(\theta)))\mathbf{j} + v_3\mathbf{k} \\ &= (v_1\cos(2\theta) - v_2\sin(2\theta))\mathbf{i} + (v_1\sin(2\theta) + v_2\cos(2\theta))\mathbf{j} + v_3\mathbf{k} \end{aligned}$$

This is the orthogonal transformation of $\mathbf{R}^3$ given by:

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \to \begin{pmatrix} \cos(2\theta) & -2\sin(\theta) & 0 \\ \sin(2\theta) & \cos(2\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$$

The same calculation can readily be done for the case of $\mathbf{e}_1$, and then using the Euler angle parametrization allows us to finish. □

Notice as $\theta$ goes from $0 \to pi$, $u(\theta, \mathbf{w})$ traces out a circle in $Sp(1)$. The homomorphism $\Phi$ takes this to a circle in $SO(3)$, one traced out twice as $\theta$ goes from $0 \to 2\pi$, explicitly showing the nature of the double cover.

Now, we want to pass this analysis of the group structure to the Lie algebras. The derivative of $\Phi$ will be a Lie algebra homomorphism, a linear map:

$$\Phi' : \mathfrak{sp}(1) \to \mathfrak{so}(3)$$

It takes the Lie algebra $\mathfrak{sp}(1)$ of pure imaginary quaternions to the Lie algebra $\mathfrak{so}(3)$ of $3 \times 3$ antisymmetric real matrices. One can compute it on basis vectors, as such:

$$\begin{aligned} \Phi'(\mathbf{k}) &= \frac{\mathrm{d}}{\mathrm{d}\theta}\Phi(\cos(\theta) + k\sin(\theta))_{|\theta=0} \\ &= \begin{pmatrix} -2\sin(2\theta) & -2\cos(2\theta) & 0 \\ 2\cos(2\theta) & -2\sin(2\theta) & 0 \\ 0 & 0 & 0 \end{pmatrix}_{|\theta=0} \\ &= \begin{pmatrix} 0 & -2 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = 2l_3. \end{aligned}$$

Repeating this on other basis vectors, we find:

$$\Phi'(\mathbf{i}) = 2l_1, \Phi'(\mathbf{j}) = 2l_2, \Phi'(\mathbf{k}) = 2l_3.$$

Thus, $\Phi'$ is an isomorphism of $\mathfrak{sp}(1)$ and $\mathfrak{so}(3)$ identifying the bases:

$$\frac{\mathfrak{i}}{2}, \frac{\mathfrak{j}}{2}, \frac{\mathfrak{k}}{2}, \text{ and }, l_1, l_2, l_3$$

Note that here, the commutation relations are:

$$\left[\frac{\mathbf{i}}{2}, \mathbf{j}_2\right] = \frac{\mathbf{k}}{2}, \left[\frac{\mathbf{j}}{2}, \mathbf{k}_2\right] = \frac{\mathbf{i}}{2}, \left[\frac{\mathbf{k}}{2}, \mathbf{i}_2\right] = \frac{\mathbf{j}}{2},$$

Now, we want to identify $Spin(3)$ and $SU(2)$, as calculations with quaternions are complicated, as opposed to traditional matrix calculations. Furthermore, we want not real matrices (like above), but rather complex matrices, so that they act naturally on complex state space. The Pauli matrices can be used to give such an isomorphism, by:

$$\begin{aligned} 1 &\to \mathbf{1} && = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ \mathbf{i} &\to -i\sigma_1 = \begin{pmatrix} 0 & -1 \\ -i & 0 \end{pmatrix} \\ \mathbf{j} &\to -i\sigma_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \\ \mathbf{k} & && \to -i\sigma_3 = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix} \end{aligned}$$

The correspondence between $\mathbf{H}$ and $2 \times 2$ complex matrices is then given by:

$$q = q_0 + q_1\mathbf{i} + q_2\mathbf{j} + q_3\mathbf{k} \longleftrightarrow \begin{pmatrix} q_0 - iq_3 & -q_2 - iq_1 \\ q_2 - iq_1 & q_0 + iq_3 \end{pmatrix}$$

Since:

$$\det \begin{pmatrix} q_0 - iq_3 & -q_2 - q_1 \\ q_2 - iq_1 & q_0 + iq_3 \end{pmatrix} = q_0^2 + q_1^2 + q_2^2 + q_3^2.$$

As such, we see determinants correspond to the norm function, and as such norm 1 quaternions (elements of $Spin(3)$) correspond to determinant 1 matrices in $SU(2)$.

Under this identification of the vector space $\mathbf{H}$ with $2 \times 2$ complex matrices (giving an identification of the Lie group), we have an identifciation of Lie algebras $\mathfrak{sp}(1) = \mathfrak{su}(2)$ between pure imaginary quaternions (the lie algebra of $(3)$) and skew-Hermitian trace-zero $2 \times 2$ complex matrices (the Lie algebra of $SU(2)$):

$$\vec{w} = w_1\mathbf{i} + w_2\mathbf{j} + w_3\mathbf{k} \longleftrightarrow \begin{pmatrix} -iw_3 & -w_2 - iw_1 \\ w_2 - iw_1 & iw_3 \end{pmatrix} = i\mathbf{w} \cdot \mathbf{\text{œ}}.$$

We identify the basis $\frac{\mathbf{i}}{2}, \frac{\mathbf{j}}{2}, \frac{\mathbf{k}}{2}$ of $\mathfrak{sp}(1)$ with the basis $\mathfrak{su}(2)$ given by the matrices $X_i$, which are defined by:

$$X_j = -i\frac{\sigma_j}{2}$$

with the $X_j$ satisfying commutation relations:

$$[X_1, X_2] = X_3, [X_2, X_3] = X_1, [X_3, X_1] = X_2$$

which are precisely the same commutation relations as for $\mathfrak{so}(3)$:

$$[l_1, l_2] = l_3, [l_2, l_3] = l_1, [l_3, l_1] = l_2.$$

## 13.3 Summary

We now have three isomorphic Lie algebras $\mathfrak{sp}(1) \cong \mathfrak{su}(2_{\cong \mathfrak{so}(3)}$, with basis vectors identified by:

$$(\mathbf{i}, \mathbf{j}, \mathbf{k}) \longleftrightarrow (X_1, X_2, X_3) = -\frac{i}{2}(\sigma_1, \sigma_2, \sigma_3) \longleftrightarrow (l_1, l_2, l_3)$$

On each of these isomorphic Lie algebras, we have adjoint Lie group $(Ad)$ and Lie algebra $(ad)$ representations. $Ad$ is given by conjugation with the corresponding group elements in $Sp(1)$, $SU(2)$, and $SO(3)$. $ad$ is given by takign commutators in the respective Lie algebras of imaginary quaternions, skew-Hermitian trace-zero $2 \times 2$ complex matrices, and $3 \times 3$ real antisymmetric matrices.

Note that these are all real representations (since they are on real vector spaces). The easiest way to extend to a complex representation is to work in $\mathfrak{su}(2)$, since complex combinations of skew-Hermitian trace-zero $2 \times 2$ complex matrices gives all trace-zero $2 \times 2$ matrices, the Lie algebra $\mathfrak{sl}(2, \mathbf{C})$.

In addition, recall that there is a fourth isomorphic version of this given by the defining representation of $SO(3)$ on column vectors. This is also real, but can straightforwardly be complexified. Since $\mathfrak{so}(3)$ and $\mathfrak{su}(2$ are isomorphic Lie algebras, their complexification $\mathfrak{so}(3)_{\mathbf{C}}$ and $\mathfrak{sl}(2, \mathbf{C})$ will also be isomorphic.

In terms of $2 \times 2$ complex matrices, Lie algebra elements can be exponentiated to find group elements in $SU(2)$, and we can define:

$$\begin{aligned}\Omega(\theta, \mathbf{w}) &= e^{\theta(w_1 X_1 + w_2 X_2 + w_3 X_3} \\ &= e^{-i\frac{\theta}{2}\mathbf{w}\cdot\boldsymbol{\text{œ}}} \\ &= \mathbf{1} + \cos(\theta/2) - i(\mathbf{w}\cdot\boldsymbol{\text{œ}})\sin(\theta/2).\end{aligned}$$

Transposing the argument from $\mathbf{H}$ to complex matrices, we see identifying:

$$\mathbf{v} \longleftrightarrow \mathbf{v}\cdot\boldsymbol{\text{œ}} = \begin{pmatrix} v_3 & v_1 - iv_2 \\ v_1 + iv_2 & -v_3 \end{pmatrix}$$

one has:

$$\Phi(\Omega(\theta, \mathbf{w})) = R(\theta, \mathbf{W})$$

with $\Omega(\theta, \mathbf{w})$ acting by conjugation, taking:

$$\mathbf{v}\cdot\boldsymbol{\text{œ}} \to \Omega(\theta, \mathbf{w})(\mathbf{v}\cdot\boldsymbol{\text{œ}})\Omega(\theta, \mathbf{w})^{-1} = (R(\theta, \mathbf{w})\mathbf{v})\cdot\boldsymbol{\text{œ}}$$

Note that in changing from the quaternionic to the complex case, the factor of 2 is treated differently, since we want $\Omega(\theta, \mathbf{w})$ to rotate by $\theta$, not $2\theta$. In terms of the identification $SU(2) = Sp(1)$, we have $\Omega(\theta, \mathbf{w}) = u(\frac{\theta}{2}, \mathbf{w})$.

We can write an explicit map $\Phi : SU(2) \to SO(3)$ in terms of $2 \times 2$ $SU(2)$ matrices of the form:

$$\begin{pmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{pmatrix}$$

by:

$$\Phi\begin{pmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{pmatrix} = \begin{pmatrix} \operatorname{Re}(\alpha^2 - \beta^2) & \operatorname{Im}(\alpha^2 + \beta^2) & -2\operatorname{Re}(\alpha\beta) \\ -\operatorname{Im}(\alpha^2 - \beta^2) & \operatorname{Re}(\alpha^2 + \beta^2) & 2\operatorname{Im}(\alpha\beta) \\ 2\operatorname{Re}(\alpha\bar{\beta}) & 2\operatorname{Im}(\alpha\bar{\beta}) & \|a\|^2 - \|b\|^2 \end{pmatrix}$$

We can see from all we have done here is that there are two interesting irreducible representations of the groups $SO(3)$ and $SU(2)$:

- A representation on $\mathbf{R}^3$ which can be constructed as the adjoint of $(3)$ or $SO(3)$, and the defining representation of $SO(3)$. This is known as the ”spin 1” representation.
- A representation of $Spin(3)$ on $\mathbf{C}^2$, which is most easily seen as the defining representation of $SU(2)$. This is not a representation of $SO(3)$ due to the double-cover nature. This is called the spin $\frac{1}{2}$ or the spinor representation and will be studied in more detail next time.

# 14 A Spin $1/2$ Particle

Here, we write the dynamics of a Spin $1/2$ particle, utilizing the knowledge of spin from the last two section. We then describe how the dynamics of such a system (again, key in quantum computing) can be realized from geometric and algebraic perspectives.

## 14.1 The Spinor Representation

Last time, we examined in great deal various ways of looking at irreducible three-dimensional real representations of the groups $SO(3), SU(2)$, and $Sp(1)$. This was the adjoint representation of those groups, isomorphic to the vector representation of $SO(3)$. In the $SU(2)$ and $Sp(1)$ cases, there is an even simpler nontrivial irreudcible representation than the adjoint - the representation of $2 \times 2$ complex matrices in $SU(2)$ on column vectors $\mathbf{C}^2$ by matrix multiplication, or unit quaternions in $Sp(1)$ on $\mathbf{H}$ by scalar multiplication. Identifying $\mathbf{C}^2 = \mathbf{H}$, we recove an isomorphic representation on $\mathbf{C}^2$ or isomorphic groups, and we will use $SU(2)$ for convenience.

**Definition 14.1 (Spinor Representation).** The spinor representation of $Spin(3) = SU(2)$ is the representation on $\mathbf{C}^2$ given by:

$$g \in SU(2) \to \pi(g) = g$$

Elements of the representation space $\mathbf{C}^2$ are called "spinors".

The spin representation of $SU(2)$ is not a representation of $SO(3)$. The double cover map $\Phi : SU(2) \to SO(3)$ is a homomorphism, so given a representation $(\pi, V)$ of $OS(3)$ one gets a representation $(\pi \circ \Phi, V)$ of $SU(2)$ by composition. One cannot go in the other direction - there is no map $SO(3) \to SU(2)$ that allows us to make the spinor representation an $SO(3)$ representation.

We could try to define a representation of $SO(3)$ by:

$$g \in SO(3) \to \pi(g) = \pi(\tilde{g}) \in SU(2),$$

where $\tilde{g}$ is one of the two elements $\tilde{g} \in SU(2)$ satisfying $\Phi(\tilde{g}) = g$. The problem is this is that it isn't exactly a homomorphism, but only up to sign, since changing the choice of $\tilde{g}$ gives a minus sign:

$$\pi(g_1)\pi(g_2) = \pm\pi(g_1 g_2)$$

The nontrivial nature of the double covering map $\Phi$ implies that there is no way to completely eliminate all minus signs, no matter how we choose $\tilde{g}$ (due to the nontrivial topological nture of the double cover). Examples like this, which are representations up to sign, are known as "projective" representations (more generally, instead of $\pm$, a scalar $\phi \in U(1)$). So, the spinor representation of $SU(2)$ gives rise to a projective representation of $SO(3)$.

Physicists normally deal with this by arguing that multiplication by scalars doesn't really matter for the purpose of the observable states, and hence, the sign ambiguity has no

effect on the physics. It is more straightforward to mathematically to work with $Spin(3)$ and accept this is the correct group reflecting rotations in three=dimensional quantum systems, however.

The spinor representation is more fundamental than the vector representation in the sense that we cannot find the spin rep only knowing the factor rep, but we can find the vector rep of $SO(3)$ knowing the spin rep of $SU(2)$. Taking spinors as fundamental however, means that we must abandon describing quantum geometry purely in terms of real numbers and embracing the fact that the quantum system involves a complex representation - justifying the "need' for complex numbers in quantum mechanics.

## 14.2 The Spin $1/2$ Particle in a Magnetic Field

A long time ago, we say that a general quantum system with $\mathcal{H} = \mathbf{C}^2$ can be understood in terms of the action of $U(2)$, and self-adjoint observables correspond (up to a factor of $i$) of the corresponding Lie algebra representation. We have already looked at $U(1)$ subgroups, so now we want to look at the $SU(2)$ subgroup. For an arbitrary system, there is no such particular geometric significance. However, when it is the double cover of the rotation group, we say it carries spin - in particular, for the two-dimensional representation, we say it carries spin $\frac{1}{2}$.

As before, we take a standard basis for $\mathfrak{su}(2)$ by $X_j$ with $j = 1, 2, 3$, with:

$$X_j = -i\frac{\sigma_j}{2}$$

which satisfies the commutation relations

$$[X_1, X_2] = X_3, [X_2, X_3] = X_1, [X_3, X_1] = X_2$$

To make contact with the physics formalism, we'll use self-adjoint operators:

$$S_j = iX_j = \frac{\sigma_j}{2}$$

In general, to a skew-adjoint operator (which is what one gets from a unitary Lie algebra representation and what exponentiates to unitary operators), we associate self-adjoint operators by multiplying by $i$. Self-adjoint operators have real eigenvalues and so are preferred by physicists. We can go in the reverse by multiplying by $-i$.

Note that the conventional definition of these operators in physics books includes a factor of $\hbar$:

$$S_j = i\hbar X_j = \frac{\hbar\sigma_j}{2}$$

A compensating factor of $\frac{1}{\hbar}$ is then introduced when exponentiated to get group elements:

$$\Omega(\theta, \mathbf{w}) = e^{-i\frac{\theta}{\hbar}\mathbf{w}\cdot\mathbf{S}} \in SU(2)$$

which do not depend on $\hbar$. This is natural since $\hbar$ appears in the momentum operator. Our definitions are not quite the same, but they are fine when we choose to work in units such that $\hbar = 1$.

States in $\mathcal{H} = \mathbf{C}^2$ that have a well-defined value of the observable $S_j$ will be eigenvectors of $S_j$, with eigenvalue $\pm\frac{1}{2}$. Measurement theory ways that if we perform the measurement corresponding to $S_j$ on an arbitrary state $|\psi\rangle$, then we will;

- with probability $c_+$ get a value of $+\frac{1}{2}$ and leave the state in an eigenvector $\left|j, +\frac{1}{2}\right\rangle$ of $S_j$ with eigenvalue $+\frac{1}{2}$.
- With probability $c_-$ get a value of $-\frac{1}{2}$ and leave the state in an eigenvector $\left|j, -\frac{1}{2}\right\rangle$ with eigenvalue $-\frac{1}{2}$ where if:

  $$|\psi\rangle = \alpha\left|j, +\frac{1}{2}\right\rangle + \beta\left|j, -\frac{1}{2}\right\rangle$$

  we have:

  $$c_+ = \frac{\|\alpha\|^2}{\|\alpha\|^2 + \|\beta\|^2}, c_- = \frac{\|\beta\|^2}{\|\alpha\|^2 + \|\beta\|^2}.$$

  After such a measurement, any attempt will measure another $S_k$ with $k \neq j$ will give $\pm\frac{1}{2}$ with equal probability (since the inner product of $\left|j, \pm\frac{1}{2}\right\rangle$ and $\left|k, \pm\frac{1}{2}\right\rangle$ are equal up to a phase) and put the system in a corresponding eigenvector.

  If a quantum system is in the arbitrary state $|\psi\rangle$ it may not have a well-defined value for some observable $A$, but the expected value of $A$ can be calculated. This is the sum over a basis of $\mathcal{H}$ consisting of eigenvectors (all orthogonal) of the corresponding eigenvalues, weighted by the probability of their occurence. The calcultion of this sum in this case ($A = Sj$) using expansion in eigenvectors of $S_j$ gives:

$$\begin{aligned}\frac{\langle\psi \mid A \mid \psi|\psi \mid A \mid \psi\rangle}{\langle\psi \mid \psi|\psi \mid \psi\rangle} &= \frac{\left(\bar{\alpha}\left\langle j, +\frac{1}{2}\right| + \bar{\beta}\left\langle j, -\frac{1}{2}\right|\right) A\left(\alpha\left|j, +\frac{1}{2}\right\rangle + \beta\left|j, -\frac{1}{2}\right\rangle\right)}{\|\alpha\|^2 + \|\beta\|^2} \\ &= \frac{\|\alpha\|^2\left(+\frac{1}{2}\right) + \|\beta\|^2\left(-\frac{1}{2}\right)}{\|\alpha\|^2 + \|\beta\|^2} \\ &= c_+\left(+\frac{1}{2}\right) + c_-\left(-\frac{1}{2}\right),\end{aligned}$$

We often aim to simplfy these calculations by normalizing states so the denominator $\langle\psi \mid \psi|\psi \mid \psi\rangle$ is 1. This works in general for the probability of measuring different eigenvalues, as long as we have orthogonality and completeness of eigenvectors. So, we can find expected value by taking $\langle\psi \mid A \mid \psi|\psi \mid A \mid \psi\rangle$.

In the case of a spin 1/2 particle, the group $Spin(3) = SU(2)$ acts on states by the spinor representation with the element $\Omega(\theta, \mathbf{w}) \in SU(2)$ acting as:

$$|\psi\rangle \to \Omega(\theta, \mathbf{w}) |\psi\rangle$$

As we saw in chapter 6, the $\Omega(\theta, \mathbf{w})$ act on self-adjoint matrices by conjugation, and this corresponds to rotation of vectors when we make the identification:

$$\mathbf{v} \longleftrightarrow \mathbf{v} \cdot \mathbf{œ}$$

Under this identification the $S_j$ correspond up to a factor of 2 to the basis vectors $\mathbf{e}_j$.

Recalling the discussion on $SU(2)$ systems earlier, the spin-degree of freedom we are describing by invoking a two-state system $\mathcal{H} = \mathbf{C}^2$, we have dynamics described by the Hamiltonian:

$$H = -^{-} \cdot \mathbf{B}$$

with $\mathbf{B}$ the magnetic field and:

$$\mathbf{u} = g\frac{-e}{2mc}\mathbf{S}$$

is an operator called the magnetic moment operator. The constants that appear are: $-e$ the electric charge, $c$ the speed of light, $m$ the mass of the particle, and $g$ a dimensionless number called the gryomagnetic ratio.

The Schrodinger equation waS:

$$\frac{\mathrm{d}}{\mathrm{d}t} |\psi(t)\rangle = -i(-^{-} \cdot \mathbf{B}) |\psi(t)\rangle$$

with solution:

$$|\psi(t)\rangle = U(t) |\psi(0)\rangle$$

where:

$$U(t) = e^{it^{-}\cdot\mathbf{B}} = e^{it-\frac{ge}{2mc}\mathbf{S}\cdot\mathbf{B}} = e^{t\frac{ge}{2mc}\mathbf{X}\cdot\mathbf{B}} = e^{t\frac{ge\mathbf{B}}{2mc}\mathbf{X}\cdot\frac{\mathbf{B}}{\|\mathbf{B}\|}}$$

The time evolution is thus given at time $t$ by the same $SU(2)$ element that, acting on vectors, gies a rotation about the axis $\mathbf{w} = \frac{\mathbf{B}}{\|\mathbf{B}\|}$ by an angle:

$$\frac{ge\|\mathbf{B}\|t}{2mc}$$

so is a rotation around $\mathbf{w}$ taking place with angular velocity $\frac{ge\|\mathbf{B}\|}{2mc}$.

Out of the simple dynamics of this system, we can understand a great deal of physics:

- The Zeeman effect. As discusses earlier, the eigenvalues of the Hamiltonian here are:

$$\pm\frac{ge\|\mathbf{B}\|}{4mc}$$

and so there is a splitting in energy levels proportional to the magnetic field, and indeed this is experimentally observable.

- Earlier, we discusses the Stern-Gerlarch machine, where we passed a beam of spin $\frac{1}{2}$ particles into a varying magnetic field. We haven't discussed particle motion, so this is more complicated than we think, but it turns out we can arrange this in such a way as to pick out a specific direction $\mathbf{w}$, and then we split the beam into components with eigenvalue $\frac{1}{2}$ and $-\frac{1}{2}$, representing the eigenvalues of the operator $\mathbf{w}\cdot\mathbf{S}$.

- Nuclear magnetic resonance spectroscopy: A spin $\frac{1}{2}$ particle can be subject to a time-varying magnetic field $\mathbf{B}(t)$. Solving the Schrodinger equation is more complicated than just exponentiation, but the nuclei of atoms provide spin $1/2$ systems that we can probe in this manner, allowing imaging of the material they make up.

- Quantum Computers are built out of composite two-state systems (as we will discuss later in the course), keeping them isolated from environmental changes, but still allowing interaction in the system that allows us to take measurements.

## 14.3 The Heisenberg Picture

We have treated time-dependence so far as a function of the state - that is, the states in $\mathcal{H}$ are functions of time, obeying the Schrodinger equation determined by $H$, while the observable self-adjoint operators $\mathcal{O}$ are time-independent. Time evolution is given by a unitary transformation:

$$U(t) = e^{-itH}, |\psi(t)\rangle = U(t)\,|\psi(0)\rangle$$

This is called the Schrodinger picture. However, we can instead place the time-dependence on the observables in something called the Heisenberg picture, by making the following exchanges:

$$|\psi(t)\rangle \to |\psi(t)\rangle_H = U^{-1}\,|\psi(t)\rangle = \psi 0,$$

and:

$$\mathcal{O} \to \mathcal{O}_H(t) = U^{-1}(t) = \mathcal{O}U(t)$$

where the $H$ subscripts indicate the Heisenberg picture choice of time-dependence. It is easily seen that the physical quantities given by eigenvalues and expectation values are the same in the two pictures - as an example, we compute:

$$\begin{aligned} {}_H\langle\psi(t) \mid \mathcal{O}_H \mid \psi(t)|\psi(t) \mid \mathcal{O}_H \mid \psi(t)\rangle_H &= \left|\psi(t) \mid U(t)(U^{-1}(t)\mathcal{O}U(t)) \mid \psi(t)\right\rangle \\ &= \langle\psi(t) \mid \mathcal{O} \mid \psi(t)|\psi(t) \mid \mathcal{O} \mid \psi(t)\rangle \end{aligned}$$

In the Heisenberg picture, the dynamics is given by a differential equation not for the states, but for the operators. Recall the following formula from our discussion of the adjoint representation:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left(e^{tX}Ye^{-tX}\right) &= \frac{\mathrm{d}}{\mathrm{d}t}\left(e^{tX}Y)\right) + e^{tX}Y\left(\frac{\mathrm{d}}{\mathrm{d}t}e^{-tX}\right) \\ &= Xe^{tX}Ye^{-tX} - e^{tX}Ye^{-tX}X\end{aligned}$$

Using this with:

$$Y = \mathcal{O}, X = iH,$$

we get:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{O}_h(t) = [iH, \mathcal{O}_H(t)] = i[H, \mathcal{O}_H(t)]$$

and this equation determines the time evolution of the observables in the Heisenberg picture. Notice that this makes clear why commuting with $H$ gives rise to conserved quantities - if $\mathcal{O}_H$ commutes with $H$, the rate of change is zero.

Applying this to the spin $1/2$ system with a magnetic field, we have:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{S}_H(t) = i[H, \mathbf{S}_H(t)] = i\frac{ge}{2mc}[\mathbf{S}_H(t)\cdot\mathbf{B}, \mathbf{S}_H(t)]$$

We know from the discussion above that the solution will be:

$$\mathbf{S}_H(t) = U(t)\mathbf{S}_H(0)U(t)^{-1}$$

for:

$$U(t) = e^{-it\frac{ge\|\mathbf{B}\|}{2mc}\mathbf{S}\cdot\frac{\mathbf{B}}{\|\mathbf{B}\|}}$$

With the discussion of the correspondences of Lie algebras and Lie groups, we again see that this tells us that the spin vector observable evoles in the Heisenberg picture by rotating around the vector $\mathbf{B}$ with angular velocity $\frac{ge\|\mathbf{B}\|}{2mc}$.

## 14.4 Projective Space

There is a different approach we can use to characterize the states of a two-state system besides using spin. Multiplication of vectors in $\mathcal{H}$ by a nonzero complex number does not change eigenvectors, eigenvalues, and expectation values, and so has no physical effect. So, what is relevant is the quotient space $(\mathbf{C}^2 - \mathbf{0})/\mathbf{C}^*$, which is constructed by taking all non-zero elements of $\mathbf{C}^2$ and identifying those related by multiplication by a nonzero complex number. For some insight into this, let's look at the real analog, where $(\mathbf{R}^2 - \mathbf{0})/\mathbf{R}^*$ can be thought of as the space of all lines in the plane going through the origin. We see that each such line hits the unit circle in two opposite points, so this set

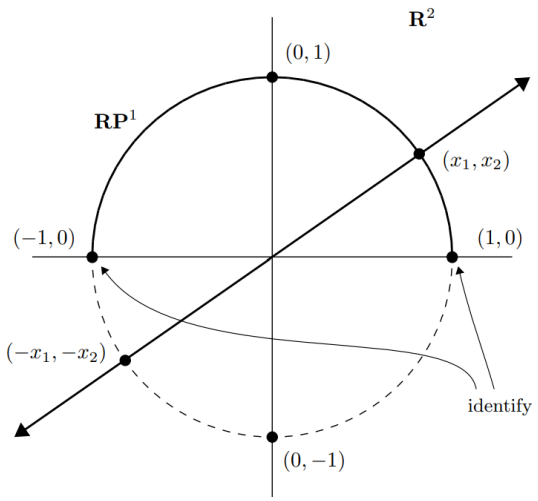


Figure 16: The Real Projective Line $\mathbf{RP}^1$

is parametrized by a semicircle, identifying points at two ends. This space is given the name $\mathbf{RP}^1$ and called the "real projective line". In higher dimensions, the space of lines through the origin in $\mathbf{R}^n$ is called $\mathbf{RP}^{n-1}$ and can be thought of as the unit sphere in $\mathbf{R}^n$ with opposite points identified.

What we are interested in is the complex analog $\mathbf{CP}^1$, which is quite a bit harder to visualize since as a real space, it is a space of two-dimensional planes through the origin of a four-dimesnional space. A standard way to do this is to associate to the vector:

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \in \mathbf{C}^2$$

the complex number $\frac{z_1}{z_2}$. Overall multiplication by a complex number drops out in this ratio, so we get different values for the coordinate $\frac{z_1}{z_2}$ for each distinct representative of the quotient, and elements of $\mathbf{CP}^1$ correspond to points on the complex plane. There is however one problem with this coordinate - the point on the plane corresponding to:

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

does not have a well-defined value - as we approach this point, the quotient $\frac{z_1}{z_2} \to \infty$ in the complex plane. In some sense, $\mathbf{CP}^1$ is the complex plane with a point at infinity added.

$\mathbf{CP}^1$ is best thought of not as a plane together with a point, but as a sphere (often called the "Riemann sphere") with the relation to the plane and point given by stereographic projection. Here, one creates a one-to-one mapping by considering the lines that go from a point on the sphere to the north pole of the sphere. Such lines intersect the plane at a point and give a one-to-one mapping everywhere except the north pole, which we identify with the point at infinity. As such, we identify $\mathbf{CP}^1$ with $S^2$. The picture is as follows: and the equations relating coordinates $(x_1, x_2, x_3)$ on the sphere and the complex coordinate

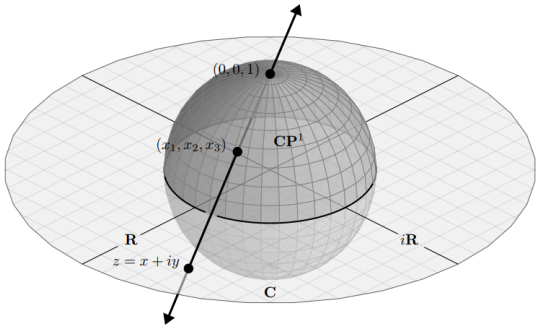


Figure 17: The Complex Projective Line $\mathbf{CP}^1$.

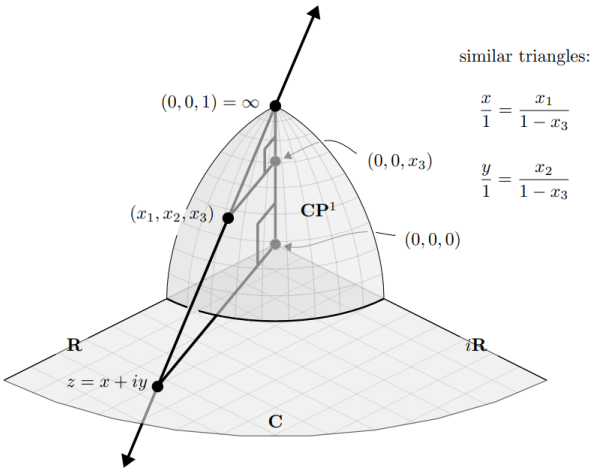


Figure 18: Complex-valued Coordinates on $\mathbf{CP}^1 - \{\infty\}$ via Stereographic Projection.

$z_1/Z_2 = z = x + iy$ on the plane are given by:

$$x = \frac{x_1}{1 - x_3}, y = \frac{x_2}{1 - x_3}$$

and:

$$x_1 = \frac{2x}{x^2 + y^2 + 1}, x_2 = \frac{2y}{x^2 + y^2 + 1}, x_3 = \frac{x^2 + y^2 - 1}{x^2 + y^2 + 1}.$$

## 14.5 The Bloch Sphere

For yet another point of view on the relation between the two-state system with $\mathcal{H} = \mathbf{C}^2$ and the geometry of the sphere, the unit sphere $S^2 \subset \mathbf{R}^3$ can be mapped to operators by:

$$\mathbf{x} \to \text{œ} \cdot \mathbf{x}.$$

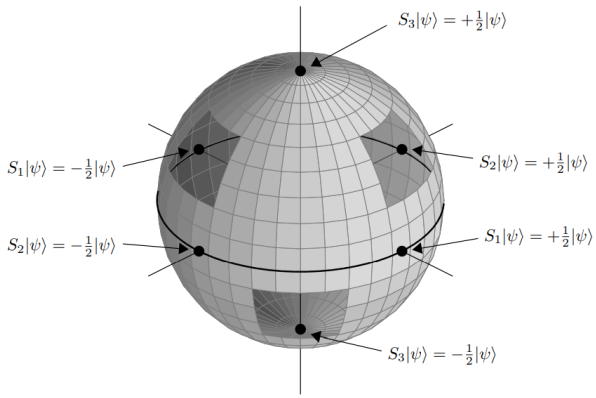


Figure 7.4: The Bloch sphere.

Figure 19: The Bloch sphere.

For each $\mathbf{x} \in S^2$, $\boldsymbol{\text{œ}} \cdot \mathbf{x}$ has eigenvalues $\pm 1$. Eigenvectors with eigenvalue $+1$ are the solutions to the equation:

$$\boldsymbol{\text{œ}} \cdot \mathbf{x} \, |\psi\rangle = |\psi\rangle$$

and give a subspace $\mathbf{C} \subset \mathcal{H}$, giving another parametrization of points ni $\mathbf{CP}(1)$.

For a more physical interpretation of this in terms of spin operators, we can multiply by $1/2$ and characterize the resulting $\mathbf{C} \subset \mathcal{H}$ corresponding to $\mathbf{x} \in S^2$ as the solutions to:

$$\mathbf{S} \cdot \mathbf{x} \, |\psi\rangle = \frac{1}{2} \, |\psi\rangle \, .$$

Then, the North pole of the sphere is a spin-up state, and the South pole is a spin-down state. Along the equator, there are two points corresponding to states with definite values for $S_1$, as well as two for states with definite values for $S_2$.

We would like to make for each $\mathbf{x}$ a choice of solution, getting a map:

$$u_+ : \mathbf{X} \in S^2 \to |\psi\rangle = u_+(\mathbf{x}) \in \mathcal{H} = \mathbf{C}^2,$$

such that:

$$(\boldsymbol{\text{œ}} \cdot \mathbf{x}) u_+(\mathbf{x}) = u_+(\mathbf{x})$$

This determines the map $u_+$ only up to multiplication by a $\mathbf{x}$-dependent scalar.

A standard choice is:

$$u_+(\mathbf{X}) = \frac{1}{\sqrt{2(1+x_3)}} \begin{pmatrix} 1 + x_3 \\ x_1 + i x_2 \end{pmatrix} = \begin{pmatrix} \cos(\theta/2) \\ e^{i\phi} \sin(\theta/2) \end{pmatrix}$$

where $\theta, \phi$ are standard spherical coordinates (we'll dicuss these later). This has some noteworthy characteristics:

- We have:

$$u_+(R\mathbf{x}) = \Omega u_+(x)$$

with $R = \Phi(\Omega)$ the rotation corresponding to the given $SU(2)$ matrix $\Omega$ (i.e, it "commutes" with the group homomorphism).

- From the above, we can determine $u_+(\mathbf{x})$ by setting it to be $\begin{pmatrix}1\\0\end{pmatrix}$ at the North pole and defining at other points $\mathbf{x}$ on the sphere by acting by $\Omega$.
- With the specific choices made, $u_+(\mathbf{x})$ is discontinous at the South pole, where $x_3 = -1$ and $\phi$ is not uniquely defined. This reflects the fact that we cannot make an everywhere continuous choice of $u_+(\mathbf{x})$.

We can see then that looking at different states in the Bloch sphere amounts to rotations of the sphere (encapsulating the action of $SU(2)$ on the state space). The Bloch sphere gives a simple geometrical interpretation, but does not preserve the inner product. The North and South poles of the sphere correspond to the orthogonal eigenstates of $S_3$, but $(0,0,1)$ and $(0,0,-1)$ are not orthogonal in $\mathbf{R}^3$.

We will discuss more about the Bloch sphere later when (if) we talk about quantum computing.

# 15 Representations of $SU(2)$: Classification

We finally begin our main goal of classifying the different spins of particles by classifying the representations of the *Spin* group.

We have seen three representations of $SU(2)$ and $SO(3)$ so far. For the case of $U(1)$, we were able to classify all complex representations, and label them by an integer in $\mathbf{Z}$. We would like to classify all complex representations of $SU(2)$ and $SO(3)$ similarly. The end result is that representations of $SU(2)$ will be classified by a non-negative integer $0, 1, 2, 3, ...$ having dimension $n+1$. For even $n$, these will correspond to irreducible representations $\rho_n$ of $SO(3)$ in the sense that:

$$\pi_n = \rho_n \circ \Phi$$

It is common to label these representations by $s = \frac{n}{2} = 0, \frac{1}{2}, 1, ...$, and call the representation labelled by $s$ the spin $s$ representation.

We already know three examples:

- Spin 0: The spin-zero representation is the trivial representation (multiplication by scalars).
- Spin $\frac{1}{2}$: This is the spinor representation we discussed last time.
- Spin 1: This is the defining representation of $SO(3)$, which we discussed a while back.

## 15.1 Passing to the Algebra

It turns out that via using the $U(1)$ subgroup of $SU(2)$, we can use the representation theory of $U(1)$ to make some powerful decompositions of the representation theory of $SU(2)$. Since we know the classification of irreducibles of $U(1)$, we know that:

$$(\pi_{|U_1}, V) = \mathbf{C}_{q_1} \oplus \mathbf{C}_{q_2} \oplus ... \oplus \mathbf{C}_{q_m}$$

for some $q_1, ..., q_m \in \mathbf{Z}$. These $q_j$ are the weights of the representation $V$.

Since our standard choice of coordinates distinguishes the z-direction and diagonalizes the action of $U(1)$ subgroup generated by this in $\mathbf{C}$, we will diagonalize the action of $SU(2)$ using this $U(1)$ subgroup to decompose into weights.

This is the subgroup of elements of $SU(2)$ of the form: $\begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}$

So, consider some representation $(\pi, V)$ of $SU(2)$, and restrict to the representation of the group $U(1)$ inside of this. Then, our decomposition of an $SU(2)$ representation $(\pi, V)$ into irreducible representations of this $U(1)$ subgroup equivalently means that we can choose a basis of $V$ such that:

$$\pi \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix} = \begin{pmatrix} e^{i\theta q_1} & 0 & ... & 0 \\ 0 & e^{i\theta q_2} & ... & 0 \\ ... & \vdots & & ... \\ 0 & 0 & ... & e^{i\theta q_m} \end{pmatrix}$$

An important property of the set of integers $q_j$ is the following:

**Theorem 15.1.** If $q$ is in the set $\{q_j\}$, so is $-q$.

*Proof.* When diagonalizing a matrix, the diagonal entries are the eigenvalues, but their order is undetermined - acting by permutations on these eigenvalues, we get different diagonalizations. In the case of $SU(2)$, we have:

$$P = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

has the property that conjugating by it permutes diagonal entries, in particular:

$$P \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix} P^{-1} = \begin{pmatrix} e^{-i\theta} & 0 \\ 0 & e^{i\theta} \end{pmatrix},$$

so:

$$\pi(P)\pi\left(\begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}\right)\pi(P)^{-1} = \pi\left(\begin{pmatrix} e^{-i\theta} & 0 \\ 0 & e^{i\theta} \end{pmatrix}\right),$$

and so $\pi(P)$ gives a change of basis $B$ such that the representation matrices in the $U(1)$ subgroup are before, but with the angles exchanging $\theta \to -\theta$. Changing this in the representation matrices is equivalent to changing the sign of the weights $q_j$. The elements of the set $\{q_j\}$ are independent of the basis, so the additional symmetry under sign change implies that for each non-zero element in the set there is another with the opposite sign. □

What does this look like for our three examples?

Restricted to $U(1)$, the spin 0 representation is one-dimensional (being scalar) and of weight 0:

$$(\pi_0, \mathbf{C}) = \mathbf{C}_0.$$

The spin 1/2 representation decomposes into $U(1)$ irreducibles of weights $\pm 1$:

$$(\pi_1, \mathbf{C}^2) = \mathbf{C}_{-1} \oplus \mathbf{C}_1$$

For the Spin 1 representation, remember the double cover map takes:

$$\begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix} \in SU(2) \to \begin{pmatrix} \cos(2\theta) & \sin(2\theta) & 0 \\ -\sin(2\theta) & \cos(2\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix} \in SO(3)$$

Acting with this matrix on $\mathbf{C}^3$ will give a unitary transformation of $\mathbf{C}^3$, which lives in $U(3)$. One can show the upper left diagonal block acts on $\mathbf{C}^2$ with weights $-2, +2$, wheras the bottom right element acts trivially on the remaining part of $\mathbf{C}^3$, which is one-dimensional of weight 0. So, restricted to $U(1)$, the spin 1 representation decomposes as:

$$(\pi_2, \mathbf{C}^3) = \mathbf{C}_{-2} \oplus \mathbf{C}_0 \oplus \mathbf{C}_{+2}$$

Recall the spin 1 representation of $SU(2)$ is often called the vector representation, since it factors in this way through the representation of $SO(3)$ by rotations on three-dimensional vectors.

To proceed further in decomposing the representation $(\pi, V)$ of $SU(2)$, we need to consider the action of the other two $U(1)$ subgroups in the other directions away from the identity. Non-commutativity of the group prevents us from simply diagonalizing the actions and assigning weights. However, passing to the Lie algebra representation $(\pi', V)$ of $\mathfrak{su}(2)$ allows us to linearize the problem. We can also exploit the complexification $\mathfrak{sl}(2, \mathbf{C})$ of $\mathfrak{su}(2)$ to further analyze the possible weights.

Recall $\mathfrak{su}(2)$ is the tangent space $\mathbf{R}^3$ to $SU(2)$ at the identity, with a basis given by the skew-adjoint $2 \times 2$ matrices:

$$X_j = -i\frac{1}{2}\sigma_j$$

which satisfy commutation relations:

$$[X_1, X_2] = X_3, [X_2, X_3] = X_1, [X_3, X_1] = X_2$$

We will often use the self-adjoint matrices $S_j = iX_j$ to extract physical quantities (in this case, that of spin in each axis).

A unitary representation $(\pi, V)$ of $SU(2)$ of dimesnion $m$ is given by a homomorphism:

$$\pi : SU(2) \to U(m)$$

Taking the derivative gives a map between the tangent space of $SU(2)$ and $U(m)$ at the identity of the groups, giving a Lie algebra rperesentation:

$$\pi' : \mathfrak{su}(2) \to \mathfrak{u}(m)$$

which takes skew-adjoint $2 \times 2$ matrices to skew-adjoint $m \times m$ matrices, preserving the commutation relations.

We have sseen that restricting to the diagonal $U(1)$ subgroup and decomposing into irreducibles tells us we can choose a basis such that:

$$(\pi, V) = (\pi_{q_1}, \mathbf{C}) \oplus (\pi_{q_2}, \mathbf{C}) \oplus ... \oplus (\pi_{q_m}, \mathbf{C})$$

For our choice of $U(1)$ as matrices of the form:

$$e^{i2\theta S_3} = \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{i\theta} \end{pmatrix}$$

with $e^{i\theta}$ going around $U(1)$ once as $\theta$ goes from 0 to $2\pi$, this means we can choose a basis of $V$ such that:

$$\pi(e^{i2\theta S_3}) = \begin{pmatrix} e^{i\theta q_1} & 0 & ... & 0 \\ 0 & e^{i\theta q_2} & ... & 0 \\ ... & \vdots & & ... \\ 0 & 0 & ... & e^{i\theta q_m} \end{pmatrix}$$

Taking the derivative to get a Lie algebra representation, using:

$$\pi'(X) = \frac{\mathrm{d}}{\mathrm{d}\theta}\pi(e^{\theta X})_{|\theta=0}$$

we find for $X = i2S_3$

$$\pi'(i2S_3) = \frac{\mathrm{d}}{\mathrm{d}\theta}\begin{pmatrix} e^{i\theta q_1} & 0 & ... & 0 \\ 0 & e^{i\theta q_2} & ... & 0 \\ ... & & & ... \\ 0 & 0 & ... & e^{i\theta q_m} \end{pmatrix}_{|\theta=0}$$
$$= \begin{pmatrix} iq_1 & 0 & ... & 0 \\ 0 & iq_2 & ... & 0 \\ ... & & & .. \\ 0 & 0 & ... & iq_m \end{pmatrix}$$

Recall that $\pi'$ is a real linear map from a real vector space ($\mathfrak{su}(2)$) into another real vector space ($\mathfrak{u}(n)$). We want to extend this via complex-linearity to a map between the complexification $\mathfrak{sl}(2,\mathbf{C})$ of $\mathfrak{su}(2)$ and the complexification $\mathfrak{gl}(m,\mathbf{C})$ or $\mathfrak{u}_m$. As an example, multiplying $X = i2S_3 \in \mathfrak{su}(2)$ by $-\frac{i}{2}$, we have $S_3 \in \mathfrak{sl}(2,\mathbf{C})$ and the diagonal elements in the matrix $\pi'(i2S_3)$ get also multiplied by $-\frac{i}{2}$ (since $\pi'$ is now complex-linear), giving:

$$\pi'(S_3) = \begin{pmatrix} \frac{q_1}{2} & 0 & ... & 0 \\ 0 & \frac{q_2}{2} & ... & 0 \\ ... & \vdots & & ... \\ 0 & 0 & ... & \frac{q_m}{2} \end{pmatrix}$$

We see then that $\pi'(S3)$ has half-integral eigenvalues.

Now that we have this in mind, we will pass towards dealing with $\mathfrak{sl}(2,\mathbf{C})$ abstractly.

## 15.2 Classification

**Theorem 15.2 (Representations of $\mathfrak{sl}(2,\mathbf{C})$).** The irreducible finite-dimensional representations of $\mathfrak{sl}2,\mathbf{C}$ are labeled by non-negative integers $n$ (of dimension $n+1$) where $n$ is the highest weight space, such that it decomposes as $V_{-n} \bigoplus V_{-n+2} \bigoplus ... V_{n-2} \bigoplus V_n$ when restricted to the $U(1)$ subgroup.

*Proof.* To start, notice that $\mathfrak{sl}(2,\mathbf{C})$ is generated by:

$$H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, X = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, Y = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

since any matrix in $\mathfrak{sl}(2,\mathbf{C})$ is of the form:

$$\begin{pmatrix} a & b \\ c & -a \end{pmatrix},$$

with commutators $[H, X] = 2X$, $[H, Y] = -2Y$, and $[X, Y] = H$. The key idea in this proof is going to be to look at the eigenspaces of $H$, which we label with weights. The $X$ and $Y$ operators will act as "raising" and "lowering" operators between these eigenspaces. We will then find a highest-weight $n$, such that the weights of the eigenspaces are $-n, -n+2, ..., n-2, n$, and therefore $V$ (as a $C^{n+1}$ vector space) decomposes as $V = C_{-n} \bigoplus C_{-n+2} \bigoplus C_{-n+4} ... C_{-2} \bigoplus C_0 \bigoplus C_2 ... \bigoplus C_{n-2} \bigoplus C_n$.

So, consider any representation of $\mathfrak{sl}(2, \mathbf{C}$ on a vector space $V$. Define $V_\lambda$ to be the eigenspace of $H$ with eigenvalue $\lambda$, concretely $V_\lambda = \{v \in V \mid hv = \lambda v\}$.

**Lemma 15.3.** $X$ is a raising operator on the eigenspaces $V_\lambda$ in the sense that if $v \in V_\lambda$, $X(v) \in V_{\lambda+2}$. Similarly, $Y$ is a lowering operator sending $v \in V_\lambda$ to $Y(v) \in V_{\lambda-2}$.

*Proof.* We have:

$$\begin{aligned} H(X(V)) &= X(H(V)) + [H, X](v) \\ &= X(\lambda v) + 2X(v) \\ &= (\lambda + 2)X(v), \end{aligned}$$

showing that $X(v)$ is an eigenvector of $H$ with eigenvalue $\lambda + 2$, so $X(v) \in V_{\lambda+2}$. The proof of the corresponding claim for $Y$ is somewhat identical. □

Now, choosing any $\lambda$ an actual eigenvector of $H$, consider the vector space $\bigoplus_{k \in \mathbf{Z}} V_{\lambda+2k}$. Given any $v$ in this space, $H, X, Y$ all send $v \in V_\alpha \subset \bigoplus_{k \in \mathbf{Z}} V_{\lambda+2k}$ to one of $V_\alpha, V_{\alpha+2}, V_{\alpha-2} \in \bigoplus_{k \in \mathbf{Z}} V_{\lambda+2k}$, so this is a subspace of $V$ invariant under our representation of $\mathfrak{sl}(2, \mathbf{C})$. Therefore, since the representation was assumed to be irreducible, we must have that this is all of $V$ itself.

Since $V$ is finite-dimensional, we can choose some $\lambda$ such that this set takes the form $V_\lambda \bigoplus V_{\lambda+2} \bigoplus ... \bigoplus V_{\lambda+2k}$ for some $k$, with $n = \lambda + 2k$ the last number in this sequence such that all other eigenspaces are trivial. Keep in mind that $n$ so far can be any complex number.

Notice that since this is the last eigenspace, $v \in V_n$ implies $X(v) \in V_{n+2}$, so $X(v) = 0$. So, this is our highest-weight vector, and we call $V_n$ the highest weight space. Now, what happens when we lower our way through the $V_\alpha$?

**Lemma 15.4.** Let $v$ be a non-zero vector in $V_n$ the highest weight space. The vectors $\{v, Y(v), ...\}$ span $V$.

*Proof.* It suffices to show that the representation of $\mathfrak{sl}(2, \mathbf{C})$ preserves this, as then it is a subrepresentation and must be all of $V$ itself by irreducibility.

So, we must check how $H, Y, V$ act on this. $H$ and $Y$ clearly preserve this (the $Y^i(v)$ live in eigenspaces of $H$ and $Y$ sends $Y^i$ to $Y^{i+1}$). So, we must check the effect of $X$.

We have $X(Y(v)) = [X, Y](v) + Y(X(v)) = H(v) + Y(0) = nv$. We have:

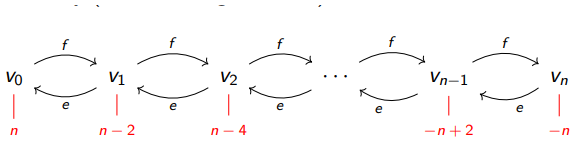


Figure 20: Diagram of Weight Space Decomposition of $\mathfrak{sl}(2, \mathbf{C})$.

$$\begin{aligned} X(Y^2(v)) &= X(Y(Y(v)) \\ &= [X, Y](Y(v)) + Y(X(Y(V))) \\ &= H(Y(v)) + Y(nv) \\ &= (n-2)Y(v) + nY(v) = 2(n-1)Y^1(v), \end{aligned}$$

using the fact that $Y(v)$ lives in $V_{n-2}$. Now, the pattern is clear, and we claim $X(Y^m(v)) = m(n-m+1)Y^{m-1}(v)$. This can be quickly verified with induction. □

Since $V$ is finite-dimensional, we must have that $Y^k(v) = 0$ for sufficiently large $k$. Take $m$ to be the least such integer, then we have:

$$X(Y^m(v)) = m(n-m+1)Y^{m-1}(v).$$

Since $m$ was the least such integer, $Y^{m-1} \neq 0$, so we must have $n = m-1$, i.e $n$ is a non-negative integer, so the weight space is labeled by a non-negative integer. Furthermore, since $m-1 = n$, and $Y^{m-1}(v)$ takes $v$ to the last non-trivial weight space, we must have that the weight spaces are $V_n, V_{n-2}, ..., V_{n-2(n)} = V_{-n}$. Visually, we get this:

We now see that any representation can be labelled by its highest weight $n$, and have dimension $n+1$. □

## 15.3 Interpretation

So, how does this relate to our prior presentation? The operator $H$ corresponds to $S_3$, except $S_3$ is $H/2$, and hence has $1/2$ the eigenvalues of $H$.

Physicists call the representation with highest weight $n$ the spin $n$ representation. Since $SU(2)$ is the symmetry group of angular momentum (with the $S_j$ the spin observables), we have now found the meaningful values for the "intrisnic" angular momentum of a quantum system. Physicists choosing $n/2$ means the values differ by 1 - for $n = 5$, we have:

$$-\frac{5}{2}, -\frac{3}{2}, -\frac{1}{2}, -\frac{1}{[}2], \frac{3}{2}, \frac{5}{2}.$$

In the case of $n = 1$ we have $\pm\frac{1}{2}$. you may have seen these numbers before as the values of the magnetic quantum number $m_s$ for an electron. This is one of the four quantum numbers in an atom, where the Pauli Exclusion Principle dictates that every electron in an atom has not all four quantum numbers identical. We will see the azimuthal and orbital magnetic quantum numbers soon.

# 16 Representations of $SU(2)$: Construction

Now that we have classified the allowed representations of the spin group, we must show they actually exist. we construct various representations and derive interesting associated operators.

## 16.1 Homogeneous Polynomials

The previous argument showed us how to construct representations given a highest weight vector, but does not provide a way to construct them. We wish to find such a method to construct an irreducible $(\pi_n, V^n)$ for each highest weight $n$.

Recall that if we have an action of $G$ on a set $M$, we have a representation on functions on $M$ by:

$$(\pi(g)f)(x) = f(g^{-1} \cdot x)$$

For $G = GL(2, \mathbf{C})$, we have by definition an action on $M = \mathbf{C}^2$, and we look at a specific class of functions on this space, the polynomials. We can break up the infinite-dimensional space of all polynomials into finite-dimensional subspaces as follows:

**Definition 16.1 (Homogenous Polynomials).** The complex vector space of homogenous polynomials of degree $n$ in two complex variables $z_1, z_2$ on $\mathbf{C}^2$ of the form:

$$f(z_1, z_2) = a_0 z_1^n + a_1 z_1^{n-1} z_2 + ... + a_{n-1} z_2^{n-1} + a_n z_2^n.$$

The space of such functions is a complex vector space of dimension $n + 1$.

This space of functions is exactly the representation $V^n$ that we need to get the spin $n/2$ irreducible representation of $SU(2) \subset GL(2, \mathbf{C})$.

If we choose a basis $\mathbf{e}_1, \mathbf{e}_2$ of $\mathbf{C}^2$, then we can write $g$ as the matrix:

$$g = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in GL(2, \mathbf{C})$$

The coordinates $z_2, z_2$ are the dual basis of the linear functions on $\mathbf{C}^2$, and $g$ will act on them by:

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \to g^{-1} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$$

The representation $\pi_n(g)$ on homogenous polynomial functions will be given by this action on $z_1, z_2$ in the expression for the polynomial.

Taking the derivative, the Lie algebra representation is given by:

$$\pi_n'(X)f = \frac{\mathrm{d}}{\mathrm{d}t}\pi_n(e^{tX} f_{|t=0} = \frac{\mathrm{d}}{\mathrm{d}t} f(e^{-tX} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix})_{|t=0}$$

where $X \in \mathfrak{gl}(2, \mathbf{C})$ is any invertible $2 \times 2$ complex matrix. By the chain rule, for:

$$\begin{pmatrix} z_1(t) \\ z_2(t) \end{pmatrix} = e^{-tX} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$$

this is:

$$\pi'_n(X)f = \left( \frac{\partial f}{\partial z_1}, \frac{\partial f}{\partial z_2} \right) \left( \frac{\mathrm{d}}{\mathrm{d}t} e^{-tX} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \right)_{|t=0}$$

$$-\frac{\partial f}{\partial z_1} - \frac{\partial f}{\partial z_2}$$

where the $X_{jk}$ are the components of the matrix $X$.

Computing this for $X_j = -i\frac{\sigma_j}{2}$, we get:

$$\pi'(X_3)f(z_1, z_2) = \frac{i}{2}\left( \frac{\partial f}{\partial z_1} z_1 - \frac{\partial f}{\partial z_2} z_2 \right)$$

so:

$$\pi'_n(X_3) = \frac{i}{2}\left( z_1 \frac{\partial}{\partial z_1} - z_2 \frac{\partial}{\partial z_2} \right)$$

and similarly:

$$\pi'_n(X_1) = \frac{i}{2}\left( z_1 \frac{\partial}{\partial z_2} + z_2 \frac{\partial}{\partial z_1} \right), \pi'_n(X_2) = \frac{1}{2}\left( z_2 \frac{\partial}{\partial z_1} - z_1 \frac{\partial}{\partial z_2} \right)$$

The $z_1^k z_2^{n-k}$ for $k = 0, ..., n$ are eigenvectors of $S_3 = iX_3$ with eigenvalue $\frac{1}{2}(n - 2k)$, since:

$$\pi'_n(S_3) z_1^k z_2^{n-k} = \frac{1}{2}(-k z_1^k z_2^{n-k} + (n-k) z_1^k z_2^{n-k}) = \frac{1}{2}(n-2k) z_1^k z_2^{n-k}.$$

Then, $z_2^n$ is an explicit highest weight vector for the representation $(\pi_n, V^n)$.

An important note here is that the formulas are not in terms of matrices, but rather a differential operator, independent of $n$, and the operator $\pi'(X)$ is the same for all the $V^n$. This has to do with the fact that the representation is on the infinite dimensional space of polynomials in $\mathbf{C}^2$.

Restricting the differential operators $\pi'(X)$ to $V^n$, the homogenous polynomials of degree $n$, they become linear operators on a finite-dimensional space, giving a highest-weight vector and a explicit construction. If we choose a basis, we get an $n + 1 \times n + 1$ matrix. Clearly, though, the differential operator is simpler to work with - this is a common principle, as in general we will obtain a partial differential equation from Lie algebra representations.

We haven't stated that the representation is unitary yet, which is important to extract physics out of the siutation. We need a inner product on $V^n$ (and more generally the

polynomial space) that is preserved by the $SU(2)$ action. In general, for complex functions on a space $M$, the following product:

$$\langle f, g | f, g \rangle = \int_M f \bar{g}$$

will be an inner product. For $M = \mathbf{C}^2$, this gives an $SU(2)$ invariant inner product on functions (it is not true for the full group $GL(2, \mathbf{C})$), it is useless for $f, g$ polynomials since the integral will diverge. In this case, a better inner product can be defined by:

$$\langle f, g | f, g \rangle = \frac{1}{\pi^2} \int_{\mathbf{C}^2} f(\bar{z_1}, z_2) g(z_1, z_2) e^{-(\|z_1\|^2 + z_2^2)} \, \mathrm{d}x_1 \, \mathrm{d}x_2 \, \mathrm{d}y_1 \, \mathrm{d}y_2$$

With $z_1 = x_1 + iy_1$, $z_2 = x_2 + iy_2$. Integrals of this time can be done fairly easily since they factorize into separate integrals over $z_1$and $z_2$, which can be treated using polar coordinates and standard calculus methods, although it is farily tedious. One can check the polynomials:

$$\frac{z_1^j z_2^k}{\sqrt{j!k!}}$$

will be an orthonormal basis of the space of polynomial functions with respect to this inner product, and the operators $\pi'(X), X \in \mathfrak{su}(2)$ will be skew-adjoint.

Working out what happens for the first few irreducible representations, we find orthonormal bases by:

- For $n = s = 0$

$$1$$

- For $n = 1, s = 1/2$:

$$z_1, z_2$$

- For $n = 2, s = 1$:

$$\frac{1}{\sqrt{2}} z_1^2, z_1 z_2, \frac{1}{\sqrt{2}} z_2^2$$

- For $n = 3, s = 3/2$:

$$\frac{1}{\sqrt{6}} z_1^3, \frac{1}{\sqrt{2}} z_2^2 z_2, \frac{1}{[} \sqrt{2}] z_1 z_2^2, \frac{1}{\sqrt{6}} z_2^3.$$

## 16.2 Spherical Harmonics

We now want to use representations of $SU(2)$ to get representations of $SO(3)$. For any representation $(\rho, V)$ of $SO(3)$, we can use the double covering homomorphism $\Phi : SU92) \to SO(3)$ to get a representation:

$$\pi = \rho \circ \Phi$$

of $SU(2)$. It can be shown for $\rho$ irreducible, $\pi$ will be too, so we must have $\pi = \rho \circ \Phi = \pi_n$, one of the irreducible representations of $SU(2)$. Using the fact that $\Phi(-\mathbf{1}) = \mathbf{1}$, we have:

$$\pi_n(-\mathbf{1}) = \rho \circ \Phi(-\mathbf{1}) = \mathbf{1}$$

From knowing the weights of $\pi_n$ are $-n, -n+2, ..., n-2, n$, we know that:

$$\pi_n(-\mathbf{1}) = \pi_n \begin{pmatrix} e^{i\pi} & 0 \\ 0 & e^{-i\pi} \end{pmatrix} = \begin{pmatrix} e^{in\pi} & 0 & ... & 0 \\ 0 & e^{i(n-2)\pi} & ... & 0 \\ ... & \vdots & & ... \\ 0 & 0 & ... & e^{-in\pi} \end{pmatrix}$$

and this is only true for $n$ even, not odd. Since the Lie algebra of $SO(3)$ is $\mathfrak{su}(2)$, the same argument applies to classify representations (but only those for $n$-even lift to representations of the group). Therefore, the representations of $SO(3)$ will be $(\rho_l, V = \mathbf{C}^{2l+1})$ for $l = 0, 1, 2, ...$ of dimension $2l+1$ and satisfying:

$$\rho_l \circ \Phi = \pi_{2l}$$

Just like $SU(2)$, we can explicitly construct these using functions on a space with an $SO(3)$ action. The natural space is $\mathbf{R}^3$, and the induced representation is as usual:

$$(\rho(g)f)(x) = f(g^{-1} \cdot x)$$

and by the same argument as in $SU(2)$, once we choose a basis, $g \in SO(3)$ is an orthogonal $3 \times 3$ matrix that acts on the coordinates $x_1, x_2, x_3$ (a basis of the dual of $\mathbf{R}^3$) by:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \to g^{-1} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

Taking the derivative, the Lie algebra representation is given by:

$$\rho'(X)f = \frac{\mathrm{d}}{\mathrm{d}t}\rho(e^{tX})f_{|t=0} = \frac{\mathrm{d}}{\mathrm{d}t}f(e^{-tX}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix})_{|t=0}$$

where $X \in \mathfrak{so}(3)$. Recall that a basis for $\mathfrak{so}(3)$ is given by the antisymmetric:

$$l_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, l_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, l_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

which satisfy the commutation relations:

$$[l_1, l_2] = l_3, [l_2, l_3] = l_1, [l_3, l_1] = l_2$$

Computing the Lie algebra representations of these, we find:

$$\begin{aligned}
&= \frac{\mathrm{d}}{\mathrm{d}tf\left(\begin{pmatrix}1 & 0 & 0\\ 0 & \cos t & \sin t\\ 0 & -\sin t & \cos t\end{pmatrix}\begin{pmatrix}x_1\\ x_2\\ x_3\end{pmatrix}\right)\Big|_{t=0}}\\
&= \frac{\mathrm{d}}{\mathrm{d}tf\left(\begin{pmatrix}x_1\\ x_2\cos t + x_3\sin t\\ -x_2\sin t + x_3\cos t\end{pmatrix}\right)\Big|_{t=0}}\\
&= \begin{pmatrix}\frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \frac{\partial f}{\partial x_3}\end{pmatrix}\cdot\begin{pmatrix}0\\ x_3\\ -x_2\end{pmatrix} = x_3\frac{\partial f}{\partial x_2} - x_2\frac{\partial f}{\partial x_3}
\end{aligned}$$

so, we have

$$\rho'(l_1) = x_3\frac{\partial}{\partial x_2} - x_2\frac{\partial}{\partial x_3}$$

and similar calculations give

$$\rho'(l_2) = x_1\frac{\partial}{\partial x_3} - x_3\frac{\partial}{\partial x_1}, \quad \rho'(l_3) = x_2\frac{\partial}{\partial x_1} - x_1\frac{\partial}{\partial x_2}$$

The space of all functions on $\mathbf{R}^3$ is much too big - it will give an infinity of copies of each finite-dimensional representation we want. When $SO(3)$ acts on $\mathbf{R}^3$ , it leaves the distance to the origin invariant. We can paraemtrize points using distance to the origin in spherical coordinates: we will have:

$$\begin{aligned}
x_1 &= r\sin(\theta)\cos(\phi)\\
x_2 &= r\sin(\theta)\sin(\phi)\\
x_3 &= r\cos(\theta)
\end{aligned}$$

Acting on $f(r, \phi, \theta)$, $SO(3)$ will leave $r$ invariant, only acting nontrivially on $\theta, \phi$. It turns out that we can cut down the space of functions to something that will only contain one copy of there representation we want.

One way to do this is to restrict our functions to the unit sphere, i.e looking at $f(\theta, \phi)$. We will see the representations we want are trigonometric functions of these two angular variables.

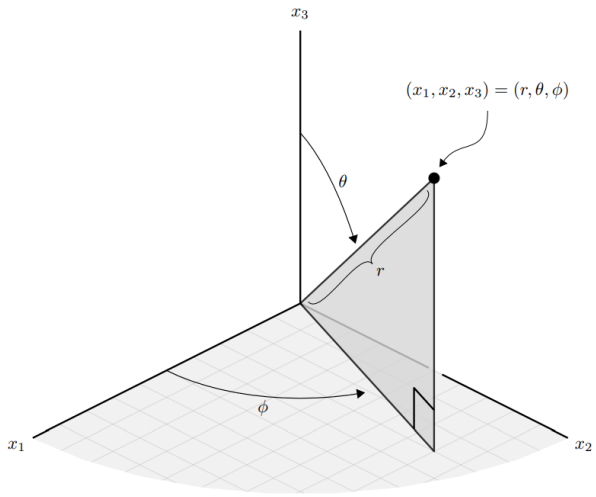


Figure 21: Spherical Coordinates.

We can construct our irreducible representations $\rho'_l$ by explicitly constructing a function we will call $Y_l^l(\theta, \phi)$ that will be a highwest weight vector of weight $l$. The weight $l$ condition and the highest weight condition give two differential equations for $Y_l^l(\theta, \phi)$:

$$L_3 Y_l^l = l Y_l^l, L_+ Y_l^l = 0,$$

where $L_+$ is the raising operator constructed similarly to as before.

These will turn out to have a unique solution up to scalars. Changing coordinates from rectangular to sphereical, we can transform $L_3, L_\pm$ using the chain rule to compute thigns like:

$$\frac{\partial}{\partial r} f(x_1(r, \theta, \phi), x_2(r, \theta, \phi), x_3(r, \theta, \phi))$$

we find:

$$\begin{pmatrix} \frac{\partial}{\partial r} \\ \frac{\partial}{\partial \theta} \\ \frac{\partial}{\partial \phi} \end{pmatrix} = \begin{pmatrix} \sin(\theta)\cos(\phi) & \sin(\theta)\sin(\phi) & \cos(\theta) \\ r\cos(\theta)\cos(\phi) & r\cos(\theta)\sin(\phi) & -r\sin(\theta) \\ -r\sin(\theta)\sin(\phi) & r\sin(\theta)\cos(\phi) & 0 \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \frac{\partial}{\partial x_3} \end{pmatrix},$$

so:

$$\begin{pmatrix} \frac{\partial}{\partial r} \\ \frac{1}{r}\frac{\partial}{\partial \theta} \\ \frac{\partial}{\partial x_3} \end{pmatrix} = \begin{pmatrix} \sin(\theta)\cos(\phi) & \sin(\theta)\sin(\phi) & \cos(\theta) \\ \cos(\theta)\cos(\phi) & r\cos(\theta)\sin(\phi) & -\sin(\theta) \\ -\sin(\phi) & r & \\ cos(\phi) & 0 & \end{pmatrix}$$

This is orthogonal, so we can invert by taking the transpose, and therefore recover:

The fucntions $Y_l^m$ are called spherical harmonics, and span the space of complex functions on the sphere in much the same way that the $e^{in\theta}$ span the space of complex-valued functions on the circle. Unlike polynomials in $\mathbf{C}^2$, for fucntions on the sphere, one gets finite numbers by integrating such functions over the sphere. so, an inner product can be defined naturally by:

$$\langle f, g | f, g \rangle = \int_{S^2} \bar{f} g \sin(\theta) \, \mathrm{d}\theta \, \mathrm{d}\phi \, .$$

It is possible (abliet tedious) to show these are variously orthogonal.

We can derive various (annoying) formulas for these in terms of Legendre polynomials, but here we'll just compute the first few examples, with the proper constants that give norm 1 with respect to the chosen inner product:

- For the $l = 0$ representation:

$$Y_0^0(\theta, \phi) = \sqrt{\frac{1}{4\pi}}$$

- For the $l = 1$ representation:

$$Y_1^1 = -\sqrt{\frac{3}{8\pi}} e^{i\phi} \sin(\pi), Y_1^0 = \sqrt{\frac{3}{4\pi}} \cos(\theta), Y_1^{-1} = \sqrt{\frac{3}{8\pi}} e^{-i\phi} \sin(\theta)$$

- For $l = 2$ representation, one has:

$$\begin{aligned} Y_2^2 &= \sqrt{\frac{15}{32\pi}} e^{i2\phi} \sin^2(\theta), \\ Y_2^l &= -\sqrt{\frac{15}{8\pi}} e^{i\phi} \sin(\theta) \cos(\theta) Y_2^0 \qquad\qquad = \sqrt{\frac{5}{16\pi}} (3\cos^2(\theta) - 1) \\ Y_2^{-1} &= \sqrt{\frac{15}{8\pi}} e^{-i\phi} \sin(\theta) \cos(\theta) \\ Y_2^{-2} &= \sqrt{\frac{15}{32\pi}} e^{-i2\phi} \sin^2(\theta). \end{aligned}$$

These turn out to be the functions giving the angular dependence of the state of a particle in a spherically symmetric potential. One such potential is the hydrogen atom:

More generally, any atom is a spherically symmetric potential. In such a case, the $SO(3)$ symmetry of the system (due to the spherical potential) implies the state space has a representation of $SO(3)$, and the action of the Hamiltonian $H$ commutes with $L_3, L_\pm$. As a result, all states in an irreducible representation of $\pi$ have the same energy. States are thus organized into "orbitals", with singlet states called "s" orbitals ($l = 0$), triplet states called "p" orbitals ($l = 1$), multiplicity 5 states called "d" orbitals ($l = 2$), etc.

The number $l$ is called the azimuthal quantum number, and the number $m$ is the magnetic quantum number $m_l$. These make up three of the four quantum numbers of an

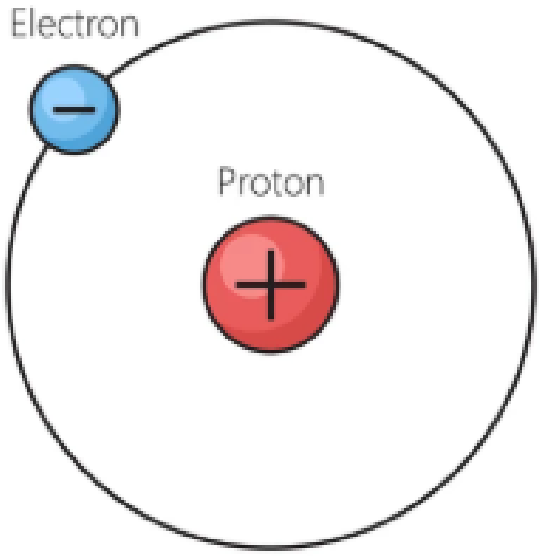


Figure 22: The Hydrogen Atom

electron. The last is the principal quantum number $n$, the number of energy levels of the atom. We will not discuss this here, but a solution for the Hydrogen atom can be found in any standard quantum mechanics textbook, such as [4] or [2].

We must always have $l \leq n$, since each irreducible representation must have different energies, so the number of irreducible representations is less than the number of energy states. This hopefully, for those of you who have taken AP Chemistry, explains a great deal of what is going on there.

## 16.3 The Casmir Operator

For both $SU(2)$ and $SO(3)$, we found representations come out of function spaces, with the Lie algebra representation acting as a first-order differential operator. IT turns out there is a second-order differential operator coming from the representation, called the Casmir operator. For the case of $SO(3)$:

**Definition 16.2 (Casmir Operator for $SO(3)$).** The Casmir operator for the representation of $SO(3)$ on functions on $S^2$ is the second-order differential operator:

$$L^2 = L_1^2 + L_2^2 + L_3^2$$

(the $^2$ here does not mean this is the square of an operator)

A straightforward (abliet tedious) calculation shows that:

$$[L^2, \rho'(X)] = 0$$

for any $X \in \mathfrak{so}(3)$. Applying a similar argument as to Schur's lemma, we see $L^2$ acts on an irreducible representation as a scalar (meaning all vectors in the representation are eigenvectors of $L^2$ with the same eigenvalue). This can be used to characterize the representation.

The easiest way to compute this turns out to act with $L^2$ on a highest-weight vector. We rewrite $L^2$ in terms of raising and lowering operators:

$$\begin{aligned} L_- L_+ &= (L_1 - iL_2)(L_1 + iL_2) \\ &= L_1^2 + L_2^2 + i[L_1, L_2] \\ &= L_1^2 + L_2^2 - L_3, \end{aligned}$$

so:

$$L^2 = L_1^2 + L_2^2 + L_3^2 = L_- L_+ + L_3 + L_3^2.$$

For the representation $\rho$ on $SO(3)$ on functions of $S^2$ ocnstructed aboe, we know on a highest weight vector of the representation $\rho_l$, we have the eigenvalue equations:

$$L_+ f = 0, L_3 f = lf$$

with the solution functions multiplies of $Y_l^l(\theta, \phi)$. From this, we can immediately plug in these eigenvalues to the expression of the Casmir operator to find:

$$L^2 f = L_- L_+ f + (L_3 + L_3^2) f = (0 + l + l^2) f = l(l+1) f$$

We have thus shown that the representation $p_l$ is equivalently the representation on which $L^2$ acts by the scalar $l(l+1)$.

In summary, we have two different sets of partial differential equations whose solutions provide a highest weight vector for as such determine the irreducible representation $\rho_l$:

- $$L_+ f = 0, L_3 f = lf$$

  which are first-order equations with the first using complexification and something like a Cauchy-Riemann equation, and:

- $$L^2 f = l(l+1) f, L_3 f = lf$$

  where the first equation is a second-order equation, something like a Laplace equation.

That a solution of the first set gives a solution of the second set (as we have just seen) is clear. It is harder to show the reverse direction. A rough argument involves taking the space of solutions to $L^2 f = l(l+1) f$ for $l$ a non-negative integer, which turns out to

be the span of the $Y_l^m(\theta, \phi)$, and since the action of $L^2$ commutes with the $L_i$, this $2l+1$ dimensional space will provide the irreducible representation of spin $l$.

Physically, this tells us something very interesting. Physically, $L^2$ represents the "orbital angular momentum" squared of the system, and $L_1, L_2, L_3$ are the $1/2/3$-components of orbital angular momentum. This tells us we can know simultaenously the value of one component of angular momentum $l$ (the $L_3$ or 3-component) and also the entire angular momentum, but we cannot know $L_1$ and $L_2$, i.e the 1-component and 2-components of orbtal angular momentum. All we know is that their squares sum to $l$.This is important and useful for physicists looking to extract physical data from a system.

Even more interesting, when we measure $L^2$ in a spherically-symmetric potential, we see $L^2$ determines the value of $l$, and hence applying the vector $L_3$ will return the highest-weight vector $l$ since we are in this irreducible representation. But, when we apply the operator $L_1$ (doing a change of basis to make the 1-direction the distinguished direction), then, we must measure the value $l$ (since we are in a highest weight space). But, we then destroy our information of $l_2$ and $l_3$, as the state changes to the highest weight vector in $L_1$.

In the $\rho$ representation on functions (using spherical coordinates), we can compute:

$$\begin{aligned} L^2 &= L_1^2 + L_2^2 + L_3^2 \\ &= \left( i \left( \sin(\phi) \frac{\partial}{\partial \theta} + \cot(\theta) \cos(\phi) \frac{\partial}{\partial \phi} \right) \right)^2 \\ &+ \left( i \left( -\cos(\phi) \frac{\partial}{\partial \theta} + \cot(\theta) \sin(\phi) \frac{\partial}{\partial \phi} \right) \right)^2 + \left( -i \frac{\partial}{\partial \phi} \right)^2 \\ &= -\left( \frac{1}{\sin(\theta)} \frac{\partial}{\partial \theta} \left( \sin(\theta) \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2(\theta)} \frac{\partial^2}{\partial \phi \partial 2} \right) \end{aligned}$$

This is the angular part of the Laplace operator on $\mathbf{R}^3$, which arises in writing down a spherical potential.

# 17 Composite Quantum Systems I

We now turn to discussing quantum systems of multiple particles, leading us to a discussion of the statistics for multi-particle systems and entanglement.

## 17.1 Tensor Product Systems

Consider two quantum systems, one defined by a state space $\mathcal{H}_1$ with operators $\mathcal{O}_1$ on it, and the second given by $\mathcal{H}_2$ and $\mathcal{O}_2$. One can describe composite quantum systems corresponding to the two quantum systems as a single one, with no interaction between them, by taking as a new state space:

$$\mathcal{H}_T = \mathcal{H}_1 \otimes \mathcal{H}_\in,$$

with operators of the form:

$$A \otimes \mathbf{Id} + \mathbf{Id} \otimes B$$

with $A \in \mathcal{O}_1, B \in \mathcal{O}_2$. The state space $\mathcal{H}_T$ can be used to describe an interacting system, but the operators are more complex.

If $\mathcal{H}$ describes a single particle quantum system, then a system of $N$ particles is given by the multiple tensor product:

$$\mathcal{H}^{\otimes n} = \mathcal{H} \otimes \mathcal{H} \otimes ...\mathcal{H}$$

## 17.2 Physical States and Representations of $S_n$

There is a clear action of the symmetric group $S_n$ on this state space, and one has a representation $(\pi, \mathcal{H}^{\otimes n})$ as follows. For $\sigma \in S_n$, a permutation of the set $\{1, 2, ..., n\}$ acts by:

$$\pi(\sigma)(v_1 \otimes v_2 \otimes ... \otimes v_n) = v_{\sigma(1)} \otimes v_{\sigma(2)} \otimes ... \otimes v_{\sigma(n)}$$

The resulting representation is in general irreducible, containing various components with different irreducible representations of the group $S_n$.

**Axiom.** If the state space $\mathcal{H}^{\otimes n}$ describes $n$ identical particles, all physical states are one-dimensional representations of $S_n$.

Now, we aim to classify the one-dimensional representations of $S_n$.

**Theorem 17.1.** The one-dimensional representations of $S_n$ are:

1. The trivial or symmetric representation given by $\pi(\sigma) = 1$.
2. The sign or antisymmetry representation given by $\pi(\sigma) = \text{sgn}(\sigma)$.

*Proof.* Consider the transpositions $(i\,j)$ which generate the group $S_n$. To determine the representation on the group, we then only have to determine the representation on the generators. We must have:

$$\pi((i\,j))^2 = \pi((i\,j)^2) = \pi(e) = 1,$$

so hence:

$$\pi((i\,j)) = 1, -1.$$

Furthermore, all transpositions are conjugate to each other (in general, the conjugacy clases in $S_n$ are exactly the $k$-cycles). But then, if $(a\,b)$ conjugates $(c\,d)$ to $(e\,f)$, then:

$$\begin{aligned}\pi((e\,f)) &= \pi((a\,b)(c\,D)(a\,b)) \\ &= \pi((a\,b))\pi((c\,d))\pi((a\,b)) \\ &= \pi((a\,b))^2\pi((c\,d)) \\ &= \pi((c\,d)),\end{aligned}$$

so we see the value of the representation of all transpositions is equal. Then, we have two cases - everything is 1, in which case we have the trivial representation. In the second case, we have $-1$, in which case the sign of an arbitrary permutation $\sigma$ is exactly the number of transpositions in its transposition decomposition, exactly the sign of $\sigma$. Hence $\pi(\sigma) = \text{sgn}(\sigma)$, and this is the sign representation. □

This axiom can actually be proved in quantum field theory, using a celebrated theorem called the spin-statistics theorem. Furthermore, there is even more information here - for spin $n/2$ particles with $n$ odd, otherwise known as fermions, the representation will be the sign representation. For integer spin particles, otherwise known as bosons, the representation will be the symmetric representation. This has important implications on the physics of the theory. However, it is not arbitrarily true in quantum mechanics, and hence it is elevated to a level of an axiom of the theory.

**Definition 17.2.** A state $v \in \mathcal{H}^{\otimes n}$ is called

- symmetric (or bosonic) if for every $\sigma \in S_n$:

  $$\pi(\sigma)v = v$$

  The space of such states is denoted $S^n(\mathcal{H})$.

- antisymmetric, or fermionic if for every $\sigma \in S_n$:

  $$\pi(\sigma)v = \text{sgn}(\sigma)(v)$$

  The space of such states is denoted $\Lambda^n(\mathcal{H})$.

In the fermionic case, for $\sigma$ a transposition interchanging two particles, $\pi$ acts on $\mathcal{H} \otimes \mathcal{H}$ by interchanging vectors, taking:

$$w \otimes w \in \mathcal{H} \otimes \mathcal{H}$$

to itself for any $w \in \mathcal{H}$, since these are identical. Antisymmetry however requires that $\pi$ sends $w \otimes w$ to $-w \otimes w$. Hence, we must have $w = 0$. As such, we cannot have identical states in a fermionic system. This is the Pauli Exclusion principle.

Now, let's look back at our exploration of magnetic numbers. The energy level $n$ and the azithumal quantum number $l$ determined the representation we were in. The magnetic quantum number $m_l$ determined which weight space we were in, and the spin quantum number $m_s$ determined the eigenstate we were in. But, an $n$-particle electron system (in a hydrogen potential) is a tensor product of two-state systems, which are entirely determined by the representation of $SU(2)$. So, the magnetic numbers entirely determine states, and the statement in chemistry that two electrons cannot have the same quantum numbers reflects this inability to have identical states.

## 17.3 Entangled and Independent Systems

If one is given a function $f$ on $X$ and a function $g$ on $Y$, we can define $fg$ on the product space $X \times Y$ can be defined by taking:

$$(fg)(x, y) = f(x)g(y)$$

However, most functions on $X \times Y$ are not decomposable in this manner. Similarly, for a tensor product of vector spaces:

**Definition 17.3 (Decomposable and Indecomposable Vectors).** A vector in $V \otimes W$ is called decomposable if it is of the form $v \otimes w$ for some $v \in V, w \in W$. If it cannot be put in this form, it is called indecomposable.

Notice that the basis vectors of $V \otimes W$ are all decomposable since they are products of basis vectors of $V$ and $W$. In particular, since the bases for $V$ and $W$ are eigenstates for $V, W$ quantum systems, this means the measurable states (hence the basis vectors) for $V \otimes W$ are precisely products of measurable states for $V$ and $W$.

However, what happens for linear combinations of basis states? In the physics context, the language is:

**Definition 17.4 (Entanglement).** An indecomposable state in the tensor product state space $\mathcal{H}_T = \mathcal{H}_1 \otimes \mathcal{H}_2$.

The phenomenon of entanglement is responsible for some of the most important quantum behavior. The Einstein-Podolosky-Rosen paradox concerns the behavior of an entangled state of two quantum systems when one moves them far apart. Then, performing a measurement on one system can give one information about what will happen when we do a measurement on the far-removed system, introducing an unexpected non-locality.

In particular, consider two entangled electrons. They can either be $\left|+\frac{1}{2}\right\rangle$ or $\left|-\frac{1}{2}\right\rangle$. But, the Pauli exclusion principle says that they cannot be the same state. Upon measurement,

each electron is forced to be one of these. So, if we separate them, and then measure one electron to be $\left|\frac{1}{2}\right\rangle$, we know the other is $\left|-\frac{1}{2}\right\rangle$. However, at far enough distances, this seems to break relativity, since relativity says nothing (not even information) travels faster than the speed of light.

However, since the outcome of the measurement is random, we cannot actually transmit usable information (since any such information would functionally be random). If the state was pre-prepared to give a guaranteed value, we would already know the value of the other particle. So, we aren't transmitting information (not breaking relativity), but we still have the problem of "non-local action", that is, the problem of the one particle being able to act on the other from extremely far distances through no obvious mechanism.

Measurement theory itself crucially involves entanglement of the state of a system being measured, $\mathcal{H}_{system}$ and the state of the measurement apparatus, $\mathcal{H}_{apparatus}$. The laws of quantum mechanics should apply to the system $\mathcal{H}_{system} \otimes \mathcal{H}_{apparatus}$, but since $\mathcal{H}_{apparatus}$ is macroscopic, we can only use approximate terms, and so it is very difficult to understand the full nature of measurement.

These kinds of questions lay at the center of measurement theory. Things like Bell's theorem, and the 2022 Nobel Prize in Physics have made progress towards this (feel free to ask me for more, although I will have to do some background reading to explain these well), but we do not yet have a full understanding of entanglement.

## 17.4 Tensor Product of Representations

Given two representations of a group, a new representation can be defined, by:

**Definition 17.5 (Tensor Product Representation of a Group).** For $(\pi_V, V)$ and $(\pi_W, W)$ representations of a group $G$, there is a tensor product representation $(\pi_{V\otimes W}, VW)$ defined by:

$$(\pi_{V\otimes W}(g))(v \otimes w) = \pi_V(g) \otimes \pi_W(g)w$$

One can easily check this is a homomorphism

To see what happens to the Lie algebra representation, computing:

$$\begin{aligned}
\pi'_{V\otimes W}(X)(v \otimes w) &= \frac{\mathrm{d}}{\mathrm{d}t}\pi_{V\otimes W}(e^{tX})(v \otimes w)_{t=0} \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\Big(\pi_V(e^{tX})v \otimes \pi_W(e^{tX})w\Big)_{t=0} \\
&= \left(\left(\frac{\mathrm{d}}{\mathrm{d}t}\pi_V(e^{tX})v\right) \otimes \pi_W(e^{tX})w\right)_{t=0} \\
&+ \left(\pi_V(e^{tX})v \otimes \left(\frac{\mathrm{d}}{\mathrm{d}t}\pi_W(e^{tX})w\right)\right)_{t=0} \\
&= (\pi'_V(X)v) \otimes w + v \otimes (\pi'_W(X)w)
\end{aligned}$$

which is equivalently:

$$\pi'_{V\otimes W}(X) = (\pi'_V(X) \otimes \mathbf{1}_W) + (\mathbf{1}_V \otimes \pi'_W(X))$$

## 17.5 Tensor Products of $SU(2)$ Representations

Given two representations $(\pi_V, V)$ and $(\pi_W, W)$ of a group $G$, we can decompose each into irreducibles. To do the same for the tensor product of the two representations, we need to know how to decompose the tensor product of irreducibles.

This is in-general nontrivial, and the answer for $SU(2)$ is as follows:

**Theorem 17.6 (Clebsch-Gordan Decomposition).** The tensor product $(\pi_{V^{n_1}\otimes V^{n_2}}, V^{n_1} \otimes V_{n_2}$ decomposes into irreducibles as:

$$(\pi_{n_1+n_2}, V^{n_1+n_2}) \oplus (\pi_{n_1+n_2-2}, V^{n_1+n_2-2}) \oplus ... \oplus (\pi_{|n_1-n_2}, V^{|n_1-n_2|}.$$

*Proof.* We will sketch a proof with highest weight theory here, and then later will show an argument using characters (an important tool in finite-dimensional representation theory).

To start, if $v_{n_1} \in V_{n_1}, v_{n_2} \in V_{n_2}$ are highest weight vectors, $v_{n_1} \otimes v_{n_2}$ will be a highest weight vector in the tensor product representation, annhilated by $\pi'_{n_1+n_2}(S_+)$ of weight $n_1 + n_2$. So, $(\pi_{n_1+n_2}, V^{n_1+n_2})$ will occur in the decomposition. Applying $\pi'_{n_1+n_2}(S_-)$ to this state gives bases for the rest of the vectors in the weight space. At the weight $n_1 + n_2 - 2$, we find another kind of vector, a highest weight vector orthogonal to the vectors in $(\pi_{n_1+n_2-2}, V^{n_1+n_2-2})$. Doing this again gives more orthogonal height weight vectors at $n_1 + n_2 - 4$, and so on, until this terminates at weight $|n_1 - n_2|$. □

## 17.6 Character Approach to Clebsch-Gordan

A standard tool for dealing with representations is that of associating to a representation an invariant called its character. This is a conjugation-invariant function on the group that only depends on the equivalence class of the representation. Given two representations constructed in different ways, often their characters can show whether they are isomorphic or not.

The problem of identifying the possible irreducible representations of a group can be studied by analyzing characters of irreducible representations. We will not enter the general theory in this course, but will see some basic examples for $G = SU(2)$. These provide a simple argument for the Clebsch-Gordan decomposition of the product of $SU(2)$ representations.

**Definition 17.7 (Character).** The character of a representation $(\pi, V)$ of group $G$ is the function on $G$ given by:

$$\chi_V(g) = \mathrm{Tr}(\pi(g))$$

Since the trace of a matrix is invariant under conjugation, $\chi_V$ will be a complex-valued, conjugation-invariant function on $G$. One can easily check that it will satisfy the relations:

$$\chi_{V\oplus W} = \chi_V + \chi_W, \chi_{V\otimes W} = \chi_V\chi_W.$$

For the case of $G = SU(2)$, any element can be conjugated to be in the $U(1)$ subgroup of diagonal matrices. Knowing the weights of the irreducible representations $(\pi_n, V^n)$ of $SU(2)$, we know the characters to be the functions:

$$\chi_{V^n}\begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix} = e^{in\theta} + e^{i(n-2)\theta} + ... + e^{-i(n-2)\theta} + e^{-in\theta}.$$

As $n$ increases, this is complicated, but there is a useful formula:

**Theorem 17.8 (Weyl Character Formula).**

$$\chi_{V^n}\begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix} = \frac{e^{i(n+1)\theta} - e^{-i(n+1)\theta}}{e^{i\theta} - e^{-i\theta}} = \frac{\sin((n+1)\theta)}{\sin(\theta)}$$

*Proof.* One just needs to recognize this is a geometric series to find the identity:

$$e^{in\theta} + e^{i(n-2)\theta} + ... + e^{-i(n-2)\theta} + e^{-in\theta}(e^{i\theta} - e^{-i\theta}) = e^{i(n+1)\theta} - e^{-i(n+1)\theta}.$$

□

To get a Proof of 9.1, compute the character of the tensor product on the diagonal matrices using the Weyl character formula for the second factor (ordering things with $n_2 > n_1$):

$$\begin{aligned} \chi_{V^{n_1}\otimes V^{n_2}} &= \chi_{V^{n_1}V^{n_2}} \\ &= (e^{in_1\theta} + e^{i(n_1-2)\theta} + ... + e^{-i(n_1-2)\theta} + e^{-in\theta}) \cdot \frac{e^{i(n_2+1)\theta} - e^{-i(n_2+1)\theta}}{e^{i}\theta - e^{-i\theta}} \\ &= \frac{(e^{i(n_1+n_2+1)\theta} - e^{-i(n_1+n_2+1)\theta) + ... + (e^{i(n_2-n_1+1)\theta} - e^{-i(n_2-n_1+1)\theta}}}{e^{i\theta} - e^{-i\theta}} \\ &= \chi_{V^{n_1+n_2}} + \chi_{V^{n_1+n_2-2}} + ... + \chi_{V^{n_2-n_1}} \end{aligned}$$

So, when we decompose the tensor product of irreducibles into a direct sum of irreducibles, the irreducibles that arise are exactly those in the Clebsch-Gordan decomposition, as desired.

## 17.7 Some Examples

Some simple examples of how these work are:

- Tensor product of two spinors:

$$V^1 \otimes V^1 = V^2 \oplus V^0$$

This says that the four dimensional tensor product of two spinor representations (which are each two complex dimensional) decomposes into irreducibles as the sum of three-dimesnional vector representation and a one-dimensional trivial (scalar) representation.

Using the basis $\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}$ for $V^1$, the tensor product $V^1 \otimes V^1$ has a basis:

$$\begin{pmatrix}1\\0\end{pmatrix} \otimes \begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix} \otimes \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix} \otimes \begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix} \otimes \begin{pmatrix}0\\1\end{pmatrix}.$$

The vector:

$$\frac{1}{\sqrt{2}}\left(\begin{pmatrix}1\\0\end{pmatrix} \otimes \begin{pmatrix}0\\1\end{pmatrix} \begin{pmatrix}0\\1\end{pmatrix} \otimes \begin{pmatrix}1\\0\end{pmatrix} \in V^1 \otimes V^1\right.$$

is clearly antisymmetric under permutation of the two factors of $V^1 \otimes V^1$. One can show that this vector is invariant under $SU(2)$ by computing either the action of $SU(2)$ or of its Lie algebra $\mathfrak{su}(2)$. So, this vector is a basis for the component $V^0$ in the Clebsch-Gordan decomposition of $V^1 \otimes V^1$.

The other component, $V^2$, is three-dimensional, and has a basis:

$$\begin{pmatrix}1\\0\end{pmatrix} \otimes \begin{pmatrix}1\\0\end{pmatrix}, \frac{1}{[}\sqrt{2}]\left(\begin{pmatrix}1\\0\end{pmatrix} \otimes \begin{pmatrix}0\\1\end{pmatrix} + \begin{pmatrix}0\\1\end{pmatrix} \otimes \begin{pmatrix}1\\0\end{pmatrix}\right), \begin{pmatrix}0\\1\end{pmatrix} \otimes \begin{pmatrix}0\\1\end{pmatrix}.$$

These three vectors span one-dimensional complex subspaces of weights $q = 2, 0, -2$ under the $U(1) \subset SU(2)$ subgroup:

$$\begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}$$

They are symmetric under permutation of the two factors of $V^1 \otimes V^1$.

We see that if we take two identical quantum systems with $\mathcal{H} = V^1 = \mathbf{C}^2$ and make a composite system out of them, if they were bososn we would get a three-dimensional state space $V^2 = S^2(V^1)$, transforming as a vector (spin one) under $SU(2)$. If they were fermions, we would get a one-dimensional state space $V^0 = \Lambda^2(V^1)$ of spin zero (invariant under $SU(2)$)). In this second case, we automatically get an entangled state, one that cannot be written as a decomposable product.

This is one of the reasons why qubits are complicated - in order to do interesting mathematics, we must include both the antisymmetric and the symmetric parts of the state space. This means that qubits cannot be purely fermionic like electrons, making them exceedingly difficult to prepare.

- Tensor product of three or more spinors:

$$V^1 \otimes V^1 \otimes V^1 = (V^2 \oplus V^0) \otimes V^1 = (V^2 \otimes V^1) \oplus (V^0 \otimes V^1) = V^3 \oplus V^1 \oplus v^1.$$

This says that the tensor product of three spinor representations decomposes as a four-dimensional ("spin 3/2") representation plus two copies of the spinor representation.

This can be generalized by considering $N$-fold tensor products $(V^1)^{\otimes N}$ of the spinor representation. This will be a sum of irreducible representations including one copy of the irreducible $V^N$. This is an alternative (and perhaps more quantum-mechanically natural) way to construct the representations of $SU(2)$, but we need to project out the desired $V^n$ component, which can be done by using the action of the symmetric group $S_n$ on $(V^1)^{\otimes n}$ and an understanding of irreducible reps of $S_N$ This relationship between irreducible reps of $SU(2)$ and those of $S_n$ coming from looking at their actions on $(V^1)^{\otimes n}$ is known as Schur-Weyl duality. This generalizes to the case of $SU(n)$ for arbitrary $n$, where one can consider $N$-fold tensor products of the defining representtion of $SU(n)$ matrices on $\mathbf{C}^n$. For $SU(n)$ this is perhaps the easiest construction of all irreducible representations.

## 17.8 Symmetric and Antisymmetric Multilinear Forms

Earlier when we talked about determinants, we defined symmetric and alternating multilinear forms. These provide multilinear functions on the part of the tensor product space that decomposes into symmetric representations. These have a basis by taking $\alpha_1 \otimes ... \otimes \alpha_n$ with $\alpha_i$ dual basis elements. We can multiply these functions by:

$$\begin{aligned}(\alpha_1 \otimes ... \otimes \alpha_j)(\alpha_{j+1} \otimes ... \otimes \alpha_n &= P^+(\alpha_1 \otimes ... \otimes \alpha_n) \\ &= \frac{1}{n!}\sum_{\sigma \in S_n} \alpha_{\sigma(1)} \otimes ... \otimes \alpha_{\sigma(n)}\end{aligned}$$

One can show that the algebra of these symmetric multilinear forms is isomorphic to the algebra of polynomials (indeed, telling us the construction outlined above of representations of $SU(2)$ is actually the same construction as from homogenous polynomials in 2 dimensions).

Antisymmetric forms similarly act on the part of the tensor product space decomposing into antisymmetric representationsm, with a multiplication:

$$\begin{aligned}(\alpha_1 \otimes ... \otimes \alpha_j) \wedge (\alpha_{j+1} \otimes ... \otimes \alpha_n) &= P^-(\alpha_1, ..., \alpha_n) \\ &= \frac{1}{n!}\sum_{\sigma \in S_n} \mathrm{sgn}(\sigma)\alpha_{\sigma(1)} \otimes ... \otimes \alpha_{\sigma(n)}\end{aligned}$$

This gives a product on the space of antisymmetric multilinear forms of different degrees, giving us a space called the exterior algebra. This turns out to be the correct algebraic structure for doing fermionic statistics (since it encapsulates antisymmetry). It also, when we take $V$ to take a geometric space, turns out to be the right space to define the generalization of derivatives to $n$-dimensions, called differential forms.

# 18 Basic Quantum Computing (Not Written)

We conclude the course by discussing basic definitions and principles in quantum computing, perhaps the most relevant modern application of quantum mechanics. In particular, we discuss Shor's algorithm, a quantum factoring algorithm significantly beating current alternatives.

## 18.1 Basic Gates

## 18.2 Bloch Sphere - Again!

## 18.3 Periods and Factoring

## 18.4 Quantum Fourier Transform

## 18.5 Shor's Algorithm

# 19 Quantum Field Theory (Not Written)

## 19.1 Double-Slit Revisited

## 19.2 Feynman Path Integral

## 19.3 Why We Need Fields

## 19.4 Canonical Quantization

# 20 The Hydrogen Atom (Not Written)

## 20.1 Setup

## 20.2 Angular Equation

## 20.3 Radial Equation

## 20.4 Radial Wave Function

## 20.5 Summary

# 21 Introduction to Particle Physics

## 21.1 The Standard Model

## 21.2 A Little Relativity

## 21.3 Feynman Diagrams

## 21.4 Particle Accelerators

## 21.5 Particle Detectors

# A Matrix Theory

This appendix was generated with ChatGPT, since I unfortunately did not have time to write all of it up in full detail. If anything seems suspicious, talk to me directly and I can explain things directly. Apologies in advance...Chat-GPT is really bad at math.

## A.1 Row-Reduction

Row-reduction is a method used to simplify matrices and solve linear systems by applying a sequence of *elementary row operations*:

- (Type I) Swap two rows.
- (Type II) Multiply a row by a nonzero scalar.
- (Type III) Replace a row by the sum of itself and a scalar multiple of another row.

The goal is typically to bring the matrix to *Reduced Row Echelon Form (RREF)*, where:

1. Each leading entry is 1.
2. Each leading 1 is the only nonzero entry in its column.
3. Each leading 1 is to the right of the leading 1 in the row above.
4. Rows with all zeros are at the bottom.

**Example A.1.** Solve the system:

$$\begin{cases} x + 2y + z = 3 \\ 2x + 4y + 5z = 12 \\ 3x + 6y + 6z = 18 \end{cases}$$

We write the augmented matrix:

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 2 & 4 & 5 & 12 \\ 3 & 6 & 6 & 18 \end{array}\right]$$

Apply row operations:

$$R_2 \leftarrow R_2 - 2R_1 \quad \Rightarrow \quad \left[\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 0 & 0 & 3 & 6 \\ 3 & 6 & 6 & 18 \end{array}\right]$$

$$R_3 \leftarrow R_3 - 3R_1 \quad \Rightarrow \quad \left[\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 0 & 0 & 3 & 6 \\ 0 & 0 & 3 & 9 \end{array}\right]$$

$$R_3 \leftarrow R_3 - R_2 \quad \Rightarrow \quad \left[\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 0 & 0 & 3 & 6 \\ 0 & 0 & 0 & 3 \end{array}\right]$$

The final row implies $0 = 3$, a contradiction, so the system is inconsistent.

## Pivots, Free Variables, and Extracting a Basis for the Solution Set

Given a homogeneous system of linear equations $A\vec{x} = \vec{0}$, where $A \in \mathbb{F}^{m\times n}$, the general solution can be described in terms of pivot and free variables after row-reducing $A$ to its row echelon or reduced row echelon form (RREF).

**Definitions.**

- A **pivot position** is a position in the matrix that corresponds to the leading 1 in a row of the RREF.
- A **pivot column** is any column of the matrix that contains a pivot position.
- The variables corresponding to pivot columns are called **basic variables**.
- The remaining variables are called **free variables** and can be assigned arbitrary values.

**Procedure to Find a Basis for the Solution Set.**

1. Row-reduce the matrix $A$ to its RREF.
2. Identify pivot columns and the corresponding basic variables.
3. Express each basic variable in terms of the free variables.
4. Write the general solution vector as a linear combination of vectors, one for each free variable.
5. The collection of these vectors forms a basis for the solution space (null space) of $A$.

**Example.** Solve $A\vec{x} = \vec{0}$ where:

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

We reduce $A$ to RREF:

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

Let $\vec{x} = (x_1, x_2, x_3)^T$. The system becomes:

$$\begin{cases} x_1 + x_3 = 0 \\ x_2 + x_3 = 0 \end{cases} \quad \Rightarrow x_1 = -x_3, \quad x_2 = -x_3$$

Letting $x_3 = t \in \mathbb{F}$, we get:

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = t \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$$

So a basis for the solution space is:

$$\left\{ \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} \right\}$$

This vector spans the null space of $A$.

## A.2 Correspondence Between Linear Operators and Matrices

Let $V$ be a finite-dimensional vector space over a field $\mathbb{F}$, with ordered basis $B = v_1, \ldots, v_n$. Let $T : V \to V$ be a linear operator. Then we can represent $T$ by a matrix $[T]_B \in \mathbb{F}^{n \times n}$ as follows.

**Theorem A.2.** There is a bijective correspondence between linear operators on $V$ and $n \times n$ matrices over $\mathbb{F}$, relative to a fixed basis.

*Proof.* 1. For each linear operator $T : V \to V$, define the matrix $A = [T]B$ by:

$$T(v_j) = \sum i = 1^n a_{ij} v_i \quad \text{for } j = 1, \ldots, n.$$

Then for any $v = \sum x_j v_j$, we have:

$$[T(v)]_B = A[v]_B.$$

2. Conversely, given $A \in \mathbb{F}^{n \times n}$, define a linear operator $T$ by:

$$T(v_j) = \sum_{i=1}^{n} a_{ij} v_i,$$

which extends linearly to all of $V$.

3. This gives a bijective, linear correspondence between $\mathrm{End}(V)$ and $\mathbb{F}^{n \times n}$.

□

We see that this correspondence extends to operations on these operators. It is clear that addition and scalar mutliplication are preserved by this correspondence. The following theorem shows us it holds with respect to multiplication.

Let $S, T : V \to V$ be linear operators, and let $A = [T]_B$, $B = [S]_B$ be their matrix representations with respect to basis $B$. Then:

**Theorem A.3.**

$$[S \circ T]_B = [S]_B \cdot [T]_B = BA.$$

*Proof.* Let $v \in V$. Then:

$$S \circ T = S(T(v)),$$

and taking coordinates with respect to basis $B$:

$$S \circ T_B = [S]_B \cdot [T(v)]_B = B(A[v]_B) = (BA)[v]_B.$$

Thus,

$$[S \circ T]_B = BA.$$

□

♣ This proof is not right...use my proof from Trimm Linear Algebra to correct this.

This confirms that the algebra of linear operators (under addition and composition) is mirrored by the algebra of matrices (under addition and multiplication).

## A.3 Change of Basis

♣ Put proofs here...GPT is really stupid.

Let $V$ be a finite-dimensional vector space, and let $B = v_1, \ldots, v_n$ and $B' = v'_1, \ldots, v'_n$ be two ordered bases of $V$.

The *change of basis matrix* from $B$ to $B'$, denoted $P_{B' \leftarrow B}$, is the matrix whose columns are the coordinate vectors of $v_i$ in basis $B'$. That is,

$$[v_i]B' = PB' \leftarrow B \cdot e_i$$

Given a linear operator $T$, the matrix representation changes under basis transformation as:

$$[T]_{B'} = P^{-1}[T]BP$$

where $P = PB' \leftarrow B$ and $[T]B, [T]B'$ are matrix representations of $T$ in bases $B$ and $B'$, respectively.

## A.4 Proof of the Spectral Theorem

**Definition A.4.** An operator $T : V \to V$ is normal if $TT^\dagger = T^\dagger T$.

Notice in particular that every self-adjoint matrix is normal, and that the operator $TT^\dagger - T^\dagger T$ is self-adjoint since it is the zero operator.

Now, the following lemma is true about normal operators.

**Proposition A.5.** Given a norma operator $T$, $v$ is an eigenvector of $T$ with eigenvalue $\lambda$ if and only if $T^\dagger v = \bar{\lambda} v$.

A proof of this fact can be found in any standard linear algebra book, such as [1]. We don't prove this here as it takes too much machinery of inner product spaces to establish.

**Theorem A.6 (Spectral Theorem for Normal Operators).** Let $V$ be a finite-dimensional complex inner product space, and let $T$ be a normal operator on $V$. Then there exists an orthonormal basis of $V$ consisting of eigenvectors of $T$. In particular, $T$ is unitarily diagonalizable.

*Proof.* We proceed by induction on $\dim V$.

If $\dim V = 1$, then any linear operator is multiplication by a scalar $\lambda \in \mathbb{C}$, and any nonzero vector is an eigenvector. Any orthonormal basis (just one unit vector) satisfies the conclusion.

Now, we show the inductive step. Assume the theorem holds for all complex inner product spaces of dimension less than $n$, and let $\dim V = n$.

Since $T$ is a linear operator on a finite-dimensional complex vector space, it has at least one eigenvalue $\lambda \in \mathbb{C}$, with corresponding eigenvector $v \neq 0$. Normalize $v$ so that $\|v\| = 1$.

Let $W = \{v\}^{\perp} \subset V$ be the orthogonal complement of $v$, that is, the set of vectors $w$ such that $\langle v, w|v, w\rangle = 0$. We claim that $W$ is invariant under $T$, i.e., $T(W) \subseteq W$.

To prove this, let $w \in W$. Then $\langle w, v\rangle = 0$. We compute:

$$\langle Tw, v\rangle = \langle w, T^{\dagger}v\rangle.$$

Since $Tv = \lambda v$, we have:

$$T^{\dagger}Tv = T^{\dagger}\lambda v = \lambda T^{\dagger}v.$$

Thus,

$$\langle Tw, v\rangle = \langle w, \overline{\lambda}v\rangle = \overline{\lambda}\langle w, v\rangle = 0.$$

So $Tw \in W$, and hence $W$ is $T$-invariant.

Now restrict $T$ to the subspace $W$, defining $T' = T|_W$. Note that $T'$ is also normal: for all $w \in W$, since $W$ is invariant under both $T$ and $T^{\dagger}$, the restrictions commute:

$$T'(T')^{\dagger} = T|_W \cdot T^{\dagger}|_W = (TT^{\dagger})|_W = (T^{\dagger}T)|_W = T^{\dagger}|_W \cdot T|_W = (T')^{\dagger}T'.$$

By the inductive hypothesis, $T'$ has an orthonormal basis $\{v_2, \ldots, v_n\}$ of eigenvectors in $W$.

Since $v \perp W$, we may prepend $v_1 = v$ to the list to get an orthonormal basis $\{v_1, \ldots, v_n\}$ for $V$, consisting of eigenvectors of $T$. This completes the inductive step.

Thus, $T$ is diagonalizable via a unitary matrix, and the proof is complete. □

We now build toward the spectral theorem for self-adjoint operators by first noting that self-adjoint operators are normal.

**Lemma A.7.** Every self-adjoint operator $T$ on a finite-dimensional complex inner product space is normal.

*Proof.* If $T$ is self-adjoint, then $T = T^{\dagger}$, so

$$TT^{\dagger} = TT = T^2 = T^{\dagger}T.$$

Hence, $T$ is normal. □

**Lemma A.8.** Let $T : V \to V$ be a self-adjoint operator on a finite-dimensional complex inner product space. Then every eigenvalue of $T$ is real.

*Proof.* Let $v \in V$ be a nonzero eigenvector of $T$ with eigenvalue $\lambda \in \mathbb{C}$. Then

$$Tv = \lambda v.$$

Taking the inner product of both sides with $v$, we get:

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

But since $T$ is self-adjoint,

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

Comparing both expressions and using $\langle v, v \rangle > 0$, we get:

$$\lambda = \overline{\lambda},$$

so $\lambda \in \mathbb{R}$. □

We now state and prove the spectral theorem for self-adjoint operators.

**Theorem A.9 (Spectral Theorem for Self-Adjoint Operators).** Let $V$ be a finite-dimensional complex inner product space. If a linear operator $L : V \to V$ is self-adjoint, then there exist eigenvectors $v_1, \ldots, v_n \in V$ and real eigenvalues $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$, not necessarily distinct, such that $\{v_1, \ldots, v_n\}$ is an orthonormal basis of $V$ and $Lv_i = \lambda_i v_i$ for each $i$.

*Proof.* By the previous lemma, a self-adjoint operator $L$ is normal. By the Spectral Theorem for Normal Operators, $L$ has an orthonormal basis $\{v_1, \ldots, v_n\}$ consisting of eigenvectors. Since $L$ is self-adjoint, each corresponding eigenvalue $\lambda_i$ is real by the previous lemma. Hence, the conclusion follows. □

# References

[1] Sheldon Axler. *Linear Algebra Done Right*. Springer, Cham, 3 edition, 2015.

[2] David J. Griffiths and Darrell F. Schroeter. *Introduction to Quantum Mechanics*. Cambridge University Press, Cambridge, 3 edition, 2018.

[3] Paul R. Halmos. *Finite-Dimensional Vector Spaces*. Undergraduate Texts in Mathematics. Springer, 2nd edition, 1974.

[4] J. J. Sakurai and Jim Napolitano. *Modern Quantum Mechanics*. Cambridge University Press, 3rd edition, 2020.

[5] D. Anderson Trimm. An Introduction to De Rham Cohomology. For IMSA Intersession Course.

[6] Peter Woit. *Quantum Theory, Groups and Representations: An Introduction*. Springer, 2017.